Compensation for Nonuniform Resolution using Penalized-Likelihood Reconstruction in Space-Variant Imaging Systems

J. Webster Stayman and Jeffrey A. Fessler, Member IEEE
4415 EECS Dept., University of Michigan, Ann Arbor, MI 48109-2122
email: stayman@eecs.umich.edu

Abstract— **Imaging systems that form estimates using a statistical approach generally yield images with nonuniform resolution properties. That is, the reconstructed images possess resolution properties marked by space-variant and/or anisotropic responses. We have previously developed a space-variant penalty for penalized-likelihood reconstruction that yields nearly uniform resolution properties [1]. In [1], we demonstrated how to efficiently calculate this penalty and apply it to an idealized positron emission tomography (PET) system whose geometric response is space-invariant. In this paper, we demonstrate the efficient calculation and application of this penalty to *space-variant* systems. (The method is most appropriate when the system matrix has been precalculated.) We apply the penalty to a large field-of-view PET system where crystal penetration effects make the geometric response space-variant, and to a 2D single photon emission computed tomography (SPECT) system whose detector responses are modeled by a depth-dependent Gaussian with linearly varying full-width half-maximum (FWHM). We perform a simulation study comparing reconstructions using our proposed penalized-likelihood approach with other reconstruction methods and demonstrate the relative resolution uniformity, and discuss trade-offs among estimators that yield nearly uniform resolution. We observe similar noise performance for the penalized-likelihood and post-smoothed maximum-likelihood approaches with carefully matched resolution, so choosing one estimator over another should be made on other factors like computational complexity and convergence rates of the iterative reconstruction. Additionally, because the post-smoothed maximum-likelihood and the proposed penalized-likelihood approach can outperform one another in terms of resolution uniformity depending on the desired reconstruction resolution, we present and discuss a hybrid approach adopting both a penalty and post-smoothing.**

Index Terms—**Tomography, Regularization, PET, SPECT, Bayesian reconstruction**

I. Introduction

Real imaging systems are subject to a number of physical effects that make the system response space-variant and image-dependent. For example, in SPECT systems, the detector response is depth-dependent, such that farther from the detector a larger region of emissions are detected. This leads to degraded intrinsic resolution in the center of the field-of-view. In PET, nonuniform sampling and crystal penetration effects also lead to space-variance [2]. Both PET and SPECT also have object-dependent attenuation effects that can affect resolution. Resolution nonuniformities can arise even in simpler imaging problems. For example, even in cases involving a simple space-invariant deblurring operation, nonuniform resolution properties can arise due to the particular form of the statistical estimator and the measurement noise model [3].

When these effects are not compensated, the reconstructed images can suffer from quantitative inaccuracies and geometric distortions due to the anisotropic responses. In many applications, such distortions are undesirable. For example, nonuniform resolution properties can complicate image registration tasks. Similarly, image-dependent resolution properties make the comparison of different images or different features in the same image more difficult.

For tomographic applications, there are a number of analytical methods that can compensate for space-variant physical effects. For SPECT, several methods have been proposed to correct for attenuation [4–6], for a depth-dependent response [7, 8], or for both [9]. Similarly, methods for PET with irregular sampling functions have been developed [10]. Under proper conditions, these techniques can yield images free of resolution nonuniformities, however generally they ignore any noise model.

An alternative is to perform maximum-likelihood (ML) reconstruction using an accurate system model. When the system model accurately incorporates the system geometry and all physical effects, it is often possible to obtain image estimates that have nearly "perfect" resolution when a pixelated object model is used. This means that the local impulse response defined in [3] is a Kronecker impulse.

Such estimates require iterating the algorithms used to maximize the ML objective until convergence. Unfortunately, such images usually appear overly noisy due to the ill-conditioned nature of inverse problems. A number of solutions have been proposed to improve the appearance of such images: (1) The ML images can simply be post-filtered with a shift-invariant blur. If the ML image has "perfect" resolution, then the blurred image will have uniform resolution. However this still requires a fully-converged solution to an unregularized problem, which may take very many iterations. (2) Iteration can begin with a uniform image and can be stopped prematurely yielding a smoother result. However such images will have nonuniform resolution properties [11, 12]. (3) The problem can be regularized, improving convergence rates and image quality. While one can use sieves [13] to regularize the problem, the appropriate kernels may not always exist

This work was supported in part by NSF grant BES-9982349 and NIH grant CA-60711.

for a desired resolution and system model. Additionally, a space-invariant sieve cannot provide uniform resolution for space-variant systems. Another form of regularization is the penalized-likelihood approach, where a penalty term is added to the objective function that discourages rough images.

Standard space-invariant penalties yield nonuniform resolution properties even for space-invariant systems due to the implicit data weighting of the penalized-likelihood objective [3]. Space-variant penalties have been developed that yield uniform resolution properties [1, 14]. In principle, these techniques may be applied to a wide range of space-variant imaging systems to correct for both the implicit data weighting and to compensate for the various physical effects that make the system space-variant. Such methods would provide for easy resolution control, where one needs only to specify the desired point spread function.

However, these space-variant penalties are data-dependent, and they must be computed for each data acquisition. Moreover, fast techniques for calculation of the penalty term have only been developed for *space-invariant* systems. [1] The main obstacle to the application of these techniques to space-variant systems, in general, remains the efficient calculation of the penalty.

The primary contribution of this paper is a new efficient procedure for the application of these penalties to *space-variant* systems. The method is particularly suited to cases where the system matrix has been precalculated. We demonstrate the use of this penalty on a 2D SPECT model, where the detector response is modeled by a depth-dependent Gaussian response whose width linearly increases with distance from the detector. We also illustrate the use of our space-variant penalty on a large field-of-view PET system where crystal penetration effects and nonuniform sampling lead to a space-variant geometric response. Such a large field-of-view model is important with the increased use of small animal PET systems.

Section II reviews the local impulse response of implicitly-defined estimators like the penalized-likelihood approach. We present a new form of the local impulse response when there is an inherent mismatch between the true continuous domain object and a discretized reconstruction model of the object. We review the penalty design approach of [1] using this formulation of the local impulse response. Section III discusses efficient procedures for calculating these penalties. We compare the resolution (Section IV) and noise performance (Section V) of penalized-likelihood with our space-variant penalty with a number of other reconstruction techniques, and discuss optimality among classes of estimators that provide uniform resolution. Section VI discusses a hybrid approach that combines elements of the post-smoothed ML approach and our proposed penalized-likelihood approach. We present an analytical comparison of penalized-likelihood and ML approaches (under a Gaussian noise model) in the appendix.

II. Background

We begin this section with a discussion of the general class of imaging models under investigation. This discussion includes both the forward measurement model and the reconstruction model used in a penalized-likelihood reconstruction. We next derive the local impulse response for this class of estimators and imaging systems. The local impulse response quantifies the local resolution properties and is the basis of our space-variant penalty design (the last topic of this section) which allows for fine resolution control.

A. Measurement Model

Let $\underline{Y} \in \mathbb{R}^N$ denote the measurement vector recorded by the imaging system. We treat $\underline{Y}$ as a random vector whose unknown mean depends on a true continuous-domain object function, $f(\boldsymbol{x})$, where $\boldsymbol{x} \in \mathbb{R}^2$ denotes the continuous spatial coordinates. We assume these means have the following form:

$$\begin{aligned} \bar{Y}_i^\dagger(f) \triangleq E_f[Y_i] &= g_i^\dagger \left(\int a_i(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} \right) \\ &= g_i^\dagger \left([\mathcal{A}f]_i \right), \end{aligned} \quad (1)$$

where $a_i(\boldsymbol{x})$ is the system "sensitivity" function for the ith measurement and $g_i^\dagger$ denotes a transformation relating the weighted integral to the mean measurements. The collection of weighted integrals for all measurements is written concisely using the continuous-to-discrete-domain operator, $\mathcal{A}$, which maps a continuous image into N (untransformed) measurements.

B. Reconstruction Model

In practice the $a_i(\boldsymbol{x})$ and $g_i^\dagger(\cdot)$ functions are rarely known exactly. Thus, there is an inherent mismatch between the measurement and the reconstruction models. Additionally, it is commonplace to use a discrete object model where the object function is represented as a linear combination of P basis functions with coefficient vector, $\underline{\theta} \in \mathbb{R}^P$. The discrete object model is selected to simplify reconstruction, display, and storage of reconstructed images (*e.g.*, standard discretizations are the pixel or voxel bases). Instead of using the (typically unknown) true relationship in (1), we adopt a reconstruction model where the mean measurements are modeled to be related to the discretized object as follows:

$$\bar{Y}_i(\underline{\theta}) = g_i \left([\boldsymbol{A}\underline{\theta}]_i \right), \quad (2)$$

where $\boldsymbol{A}$ is called the system matrix and $g_i(\cdot)$ is a function that relates weighted sums of image parameters to the mean measurements. The $N \times P$ matrix, $\boldsymbol{A}$ is meant to approximate the action of the continuous-to-discrete operator $\mathcal{A}$, and $g_i(\cdot)$ is meant to approximate the transformation $g_i^\dagger(\cdot)$.

We reconstruct $\underline{\theta}$ using penalized-likelihood estimation. The penalized-likelihood estimator is written implicitly as

the maximizer of an objective function:

$$\begin{aligned} \hat{\underline{\theta}} &= \arg\max_{\underline{\theta}} \mathbf{\Phi}(\underline{\theta}, \underline{Y}), \\ \Phi(\underline{\theta}, \underline{Y}) &\triangleq L(\underline{\theta}, \underline{Y}) - R(\underline{\theta}), \end{aligned} \tag{3}$$

where the objective is composed of two terms; the log-likelihood, $L(\underline{\theta}, \underline{Y})$, and the penalty $R(\underline{\theta})$, which discourages rough images. Under the usual assumption of independent measurements the log-likelihood has the form

$$\begin{aligned} L(\underline{\theta}, \underline{Y}) &= \sum_i h_i(Y_i, \bar{Y}_i(\underline{\theta})) \\ &= \sum_i h_i(Y_i, g_i([\boldsymbol{A}\underline{\theta}]_i)), \end{aligned} \tag{4}$$

where $h_i(\cdot, \cdot)$ is a two dimensional function of the ith measurement, Y_i, and its mean $\bar{Y}_i$. Each $h_i(u, v)$ term represents a marginal log-likelihood whose form depends on the chosen noise model. The system model enters into (4) through the reconstruction model for mean measurements in (2).

The above general framework covers a range of imaging systems and noise models. For example, one can model an emission tomography system with Poisson measurements by the following choices:

$$\begin{aligned} g_i(l) &= l + r_i \\ h_i(u, v) &= u \log(v) - v - u!, \end{aligned} \tag{5}$$

where r_i represents the mean contribution of background and scatter events. Similarly, a transmission tomography system with Gaussian measurements can be modeled using

$$\begin{aligned} g_i(l) &= b_i \exp(-l) + r_i \\ h_i(u, v) &= -\frac{1}{\sigma_i^2}(u - v)^2 - \log\sqrt{2\pi}\sigma_i, \end{aligned} \tag{6}$$

where b_i represents detector normalization factors, and σ_i represents known standard deviations for the ith measurement.

C. Quadratic Regularization

Penalized-likelihood estimators discourage overly noisy images by including the penalty term, $R(\underline{\theta})$, in (3). There are many possible choices for the penalty term. We will focus on pairwise quadratic penalties with the following form,

$$R(\underline{\theta}) = \frac{1}{2}\sum_{j=1}^{P}\sum_{k=1}^{P} w_{jk}(\theta_j - \theta_k)^2, \tag{7}$$

where the w_{jk} terms are nonnegative values that the algorithm designer must choose. Quadratic penalties can be written conveniently as $R(\underline{\theta}) = \frac{1}{2}\underline{\theta}'\boldsymbol{R}\underline{\theta}$, where the penalty matrix $\boldsymbol{R}$ has elements,

$$\boldsymbol{R}_{jk} = \begin{cases} \sum_{l=1}^{P} w_{lj}, & k = j \\ -w_{jk}, & k \neq j. \end{cases} \tag{8}$$

The definition of the elements of $\boldsymbol{R}$ represented in (8) allows for asymmetric $\boldsymbol{R}$. However, only the symmetric component of $\boldsymbol{R}$ is important for the penalty, since $\underline{\theta}'\boldsymbol{R}\underline{\theta} = (\underline{\theta}'\boldsymbol{R}\underline{\theta})' = \underline{\theta}'\boldsymbol{R}'\underline{\theta}$.

Practical roughness penalties use only a small neighborhood about each pixel, so most w_{jk} are zero. Therefore, we adopt the following parameterization of the penalty matrix. Letting $\underline{e}^j$ denote the jth unit vector, a column of the matrix $\boldsymbol{R}$ has the form

$$\boldsymbol{R}\underline{e}^j = \begin{bmatrix} -w_{1,j} \\ \vdots \\ -w_{j-1,j} \\ \sum_{l\neq j} w_{l,j} \\ -w_{j+1,j} \\ \vdots \\ -w_{P,j} \end{bmatrix} = \sum_{l\neq j} w_{lj}(\underline{e}^j - \underline{e}^l) \tag{9}$$

$$= \sum_{q=1}^{B} w_{l_{jq},j}\,\mathrm{vec}\{b_q(m - m_j, n - n_j)\} \tag{10}$$

$$= \boldsymbol{B}^j\underline{w}^j. \tag{11}$$

In (9), we simply rewrite a column of $\boldsymbol{R}$ as a weighted sum of the difference of two unit vectors. (The jth unit vector is written as $\underline{e}^j$.) These unit vector differences identify pairs of pixels in the penalty. Since most pixel pairs are not penalized (*i.e.*: $w_{jk} = 0$), one can alternately write the penalty in terms of a small number of basis functions that identify pixel pairs. For a 2D imaging problem, these basis functions are of the form:

$$b_q(m, n) = \delta(m, n) - \delta(m + m_q, n + n_q), \tag{12}$$

where (m, n) are image coordinates and (m_q, n_q) are coordinate offsets for the qth neighbor (basis). Thus, in (10) we write the parameterization as a weighted sum over the B bases, which are lexicographically reordered into vector form (as denoted by vec$\{\cdot\}$). The bases are shifted by (m_j, n_j) which represent the coordinates of the jth pixel. In (10), l_{jq} represents the vector position which corresponds to the pixel identified by position (m_j, n_j) and the offset (m_q, n_q). This sum may also be written succinctly, as in (11), in a matrix form using a $P \times B$ basis matrix $\boldsymbol{B}^j$ and a vector $\underline{w}^j$ that is composed of the weights $\{w_{l_{jq},j}\}_{q=1}^{B}$.

D. Local Impulse Response

The local impulse response is a useful tool for investigating the resolution properties of imaging systems and the reconstruction methods used to form the images. Additionally, knowing the local impulse response of a penalized-likelihood estimator prior to reconstruction allows one to customize a specific penalty function that yields user-specified resolution properties. In [3] an approximate local impulse response was derived for discrete object models. Here, we extend those derivations for a continuous object model.

The local impulse response is defined in terms of the mean reconstruction, $\mu(f) = E_f[\hat{\underline{\theta}}(\underline{Y})]$, and is the limiting

TABLE I
DERIVATIVES OF $h_i(u,v)$ UNDER VARIOUS NOISE MODELS.

(Additive constants not important for the maximization of the penalized-likelihood objective have been dropped.)

Distribution	$h_i(u,v)$	$h_i^{01}(u,v)$	$-h_i^{02}(u,v)$	$h_i^{11}(u,v)$
Gaussian	$-\frac{1}{2\sigma_i^2}(u-v)^2$	$\frac{1}{\sigma_i^2}(u-v)$	$\frac{1}{\sigma_i^2}$	$\frac{1}{\sigma_i^2}$
Generalized Gaussian*	$-\left(\frac{\lvert u-v\rvert}{\beta_i}\right)^{\alpha_i}$	$-\mathrm{sgn}(u-v)\frac{\alpha_i}{\beta_i}\left(\frac{\lvert u-v\rvert}{\beta_i}\right)^{\alpha_i-1}$	$\frac{(\alpha_i-\alpha_i^2)}{\beta_i^2}\left(\frac{\lvert u-v\rvert}{\beta_i}\right)^{\alpha_i-2}$	$\frac{(\alpha_i-\alpha_i^2)}{\beta_i^2}\left(\frac{\lvert u-v\rvert}{\beta_i}\right)^{\alpha_i-2}$
Poisson	$u\log(v)-v$	$\frac{u}{v}-1$	$\frac{u}{v^2}$	$\frac{1}{v}$
Shifted Poisson	$(u+a_i)\log(v+a_i)-v$	$\frac{u+a_i}{v+a_i}-1$	$\frac{u+a_i}{(v+a_i)^2}$	$\frac{1}{v+a_i}$

*For the generalized Gaussian distribution, $\alpha_i \in (1,\infty)$ and $\beta_i \in (0,\infty)$.

difference between mean reconstructions of an image and reconstructions of a perturbed image. The local impulse response at spatial location $\boldsymbol{x}_0$ is defined as

$$l(\boldsymbol{x}_0) \triangleq \lim_{\varepsilon\to 0}\frac{\mu\left(f(\boldsymbol{x})+\varepsilon\delta(\boldsymbol{x}-\boldsymbol{x}_0)\right)-\mu\left(f(\boldsymbol{x})\right)}{\varepsilon}, \qquad (13)$$

where $\delta_{\boldsymbol{x}_0} \triangleq \delta(\boldsymbol{x}-\boldsymbol{x}_0)$ is a Dirac impulse at position $\boldsymbol{x}_0$. As in [3], if we assume that the estimator is locally linear, then $\mu(f) \approx \hat{\underline{\theta}}(\bar{\underline{Y}}^\dagger(f))$, so

$$\begin{aligned} l(\boldsymbol{x}_0) &= \lim_{\varepsilon\to 0}\frac{\hat{\underline{\theta}}\left(\bar{\underline{Y}}^\dagger(f+\varepsilon\delta_{\boldsymbol{x}_0})\right)-\hat{\underline{\theta}}\left(\bar{\underline{Y}}^\dagger(f)\right)}{\varepsilon} \\ &= \nabla_Y\hat{\underline{\theta}}\left(\bar{\underline{Y}}^\dagger(f)\right)\cdot\left(\lim_{\varepsilon\to 0}\frac{\bar{\underline{Y}}^\dagger(f+\varepsilon\delta_{\boldsymbol{x}_0})-\bar{\underline{Y}}^\dagger(f)}{\varepsilon}\right) \\ &= \nabla_Y\hat{\underline{\theta}}\left(\bar{\underline{Y}}^\dagger(f)\right)\cdot\triangle\bar{\underline{Y}}^\dagger(f;\boldsymbol{x}_0), \end{aligned} \qquad (14)$$

where $\nabla_Y = [\frac{\partial}{\partial Y_1}\cdots\frac{\partial}{\partial Y_N}]$. We evaluate $\triangle\bar{\underline{Y}}^\dagger(f;\boldsymbol{x}_0)$ using the mean measurements in (1) and applying the chain rule. The ith element is

$$\begin{aligned} \left[\triangle\bar{\underline{Y}}^\dagger(f;\boldsymbol{x}_0)\right]_i &= \dot{g}_i^\dagger\left(\langle a_i,f\rangle\right)\lim_{\varepsilon\to 0}\frac{\langle a_i,f+\varepsilon\delta_{\boldsymbol{x}_0}\rangle-\langle a_i,f\rangle}{\varepsilon} \\ &= \dot{g}_i^\dagger\left(\langle a_i,f\rangle\right)a_i(\boldsymbol{x}_0) \\ &= \dot{g}_i^\dagger\left([\mathcal{A}f]_i\right)\mathcal{A}\delta_{\boldsymbol{x}_0}. \end{aligned} \qquad (15)$$

In [3], an equation for $\nabla_Y\hat{\underline{\theta}}(\underline{Y})$ was derived for implicitly-defined estimators of the form in (3). Assuming that $-\nabla^{20}\Phi(\underline{\theta},\underline{Y})$ is positive definite and $\Phi(\underline{\theta},\underline{Y})$ is twice differentiable, we may write

$$\nabla_Y\hat{\underline{\theta}}(\underline{Y}) = \left[-\nabla^{20}\Phi\left(\hat{\underline{\theta}}(\underline{Y}),\underline{Y}\right)\right]^{-1}\nabla^{11}\Phi\left(\hat{\underline{\theta}}(\underline{Y}),\underline{Y}\right), \qquad (16)$$

where ∇^{20} is an operator that yields a matrix whose (j,k)th element is $\frac{\partial^2}{\partial\theta_j\partial\theta_k}$, and ∇^{11} is an operator that yields a matrix whose (j,i)th element is $\frac{\partial^2}{\partial\theta_j\partial Y_i}$. Because the objective function in (3) is simply a difference of the likelihood and penalty terms, one can evaluate the above partial derivatives separately for the likelihood and penalty terms.

We will identify these partial derivative terms in the following paragraphs; however, it is helpful to first adopt some shorthand notation for the derivatives of various model components. Specifically, the derivatives of $h_i(u,v)$, $g_i(l)$, and $g_i^\dagger(l)$ are denoted as follows

$$\begin{array}{rclcrcl} h_i^{01}(u,v) &=& \frac{\partial}{\partial v}h_i(u,v) & & \dot{g}_i^\dagger(l) &=& \frac{\partial}{\partial l}g_i^\dagger(l) \\ h_i^{02}(u,v) &=& \frac{\partial^2}{\partial v^2}h_i(u,v) & & \dot{g}_i(l) &=& \frac{\partial}{\partial l}g_i(l) \\ h_i^{11}(u,v) &=& \frac{\partial^2}{\partial u\partial v}h_i(u,v) & & \ddot{g}_i(l) &=& \frac{\partial^2}{\partial l^2}g_i(l). \end{array} \qquad (17)$$

For penalties that are not a function of $\underline{Y}$, $\nabla^{11}R(\underline{\theta})$ is zero. Even though we will eventually design a penalty that is dependent on the projection data, we have found that ignoring the dependence of R on $\underline{Y}$ nevertheless leads to good estimates of the local impulse response. In other words, the derivatives of the penalty with respect to $\underline{Y}$ are sufficiently small as to be disregarded when evaluating (16). For a quadratic penalty, the Hessian of the penalty is the symmetric portion of the penalty matrix. Thus,

$$\nabla^{20}R(\underline{\theta}) = \frac{1}{2}(\boldsymbol{R}+\boldsymbol{R}') \triangleq \boldsymbol{R}_{\mathrm{sym}} \qquad (18)$$

$$\nabla^{11}R(\underline{\theta}) = \boldsymbol{0}. \qquad (19)$$

Returning to (4), we may write the associated expressions for the likelihood term. Specifically, using the chain rule and adopting a matrix notation form, we find that

$$\nabla^{20}L(\underline{\theta},\underline{Y}) = \boldsymbol{A}'\mathrm{diag}\left\{\begin{array}{c} h_i^{02}\left(Y_i,g_i([\boldsymbol{A}\underline{\theta}]_i)\right)\left[\dot{g}_i([\boldsymbol{A}\underline{\theta}]_i)\right]^2 + \\ h_i^{01}\left(Y_i,g_i([\boldsymbol{A}\underline{\theta}]_i)\right)\left[\ddot{g}_i([\boldsymbol{A}\underline{\theta}]_i)\right] \end{array}\right\}\boldsymbol{A} \qquad (20)$$

$$\nabla^{11}L(\underline{\theta},\underline{Y}) = \boldsymbol{A}'\mathrm{diag}\left\{h_i^{11}\left(Y_i,g_i([\boldsymbol{A}\underline{\theta}]_i)\right)\left[\dot{g}_i([\boldsymbol{A}\underline{\theta}]_i)\right]\right\}. \qquad (21)$$

We may now use (18)-(21) to find a more specific expression for (16). Moreover, performing this substitution and plugging (15) and (16) back into (14), yields following

expression for the local impulse response for a penalized-likelihood estimator with a quadratic penalty:

$$\underline{l}(\boldsymbol{x}_0) = [\boldsymbol{A}'\boldsymbol{D}_1\boldsymbol{A} + \boldsymbol{R}_{\text{sym}}]^{-1}\boldsymbol{A}'\boldsymbol{D}_2\mathcal{A}\delta_{\boldsymbol{x}_0} \tag{22}$$

where $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are the following $N \times N$ diagonal matrices:

$$\begin{aligned}[\boldsymbol{D}_1]_{ii} &= -h_i^{02}\left(\bar{Y}_i^\dagger(f), \bar{Y}_i(\breve{\underline{\theta}})\right)\left[\dot{g}_i([\boldsymbol{A}\breve{\underline{\theta}}]_i)\right]^2 \\ &\quad -h_i^{01}\left(\bar{Y}_i^\dagger(f), \bar{Y}_i(\breve{\underline{\theta}})\right)\left[\ddot{g}_i([\boldsymbol{A}\breve{\underline{\theta}}]_i)\right] \qquad (23)\\ [\boldsymbol{D}_2]_{ii} &= h_i^{11}\left(\bar{Y}_i^\dagger(f), \bar{Y}_i(\breve{\underline{\theta}})\right)\left[\dot{g}_i([\boldsymbol{A}\breve{\underline{\theta}}]_i)\right] \\ &\quad \cdot\left[\dot{g}_i^\dagger([\mathcal{A}f]_i)\right], \qquad (24)\end{aligned}$$

where $\breve{\underline{\theta}} \triangleq \hat{\underline{\theta}}(\bar{\underline{Y}}^\dagger(f))$ denotes the estimate of $\underline{\theta}$ using the mean data. In the typical cases where $g(\cdot)$ and $g^\dagger(\cdot)$ are invertible functions, we can write the diagonal matrices (23) and (24) as functions of the mean measurements. Specifically,

$$\begin{aligned}[\boldsymbol{D}_1]_{ii} &= -h_i^{02}\left(\bar{Y}_i^\dagger(f), \bar{Y}_i(\breve{\underline{\theta}})\right)\left[\dot{g}_i\left(g^{-1}\left(\bar{Y}_i(\breve{\underline{\theta}})\right)\right)\right]^2 \\ &\quad -h_i^{01}\left(\bar{Y}_i^\dagger(f), \bar{Y}_i(\breve{\underline{\theta}})\right)\left[\ddot{g}_i\left(g^{-1}\left(\bar{Y}_i(\breve{\underline{\theta}})\right)\right)\right] \qquad (25)\\ [\boldsymbol{D}_2]_{ii} &= h_i^{11}\left(\bar{Y}_i^\dagger(f), \bar{Y}_i(\breve{\underline{\theta}})\right)\left[\dot{g}_i\left(g^{-1}\left(\bar{Y}_i(\breve{\underline{\theta}})\right)\right)\right] \\ &\quad \cdot\left[\dot{g}_i^\dagger\left(g_i^{\dagger -1}\left(\bar{\underline{Y}}_i^\dagger(f)\right)\right)\right]. \qquad (26)\end{aligned}$$

The seasoned observer will note that (22) has a general form very similar to a linear penalized least-squares estimator. In fact, if the transformation function $g_i(l)$ is linear, it is straightforward to derive a very similar expression for the local impulse response without relying on the linearization discussed immediately before (14). Thus, in some ways this linearization is similar to making a Gaussian assumption on individual measurements, and we would expect the linearized response approximation in (22) to be very accurate for "high count" data where the central limit theorem is at work. While this is intuitively the case for "high count" data we find that (22) provides very accurate results for "low count" data as well.

Strictly speaking, to calculate the local impulse response, one must substitute (25) and (26) into (22). However, when the system model, $\boldsymbol{A}$ and $g_i(x)$, closely approximates the actual system, $\mathcal{A}$ and $g_i^\dagger(x)$, the means, $\bar{Y}_i^\dagger(f)$ and $\bar{Y}_i(\breve{\underline{\theta}})$ are often very similar to each other. That is, the mean measurements from the actual system model and object are nearly the same as the mean measurements from the modeled system and the mean object reconstruction. Typically the mean reconstruction is a slightly blurred version of the true object, whereas the mean measurements are more heavily blurred (in tomography this "blur" includes the projection operation). Thus, as long as the estimator smoothing parameter is relatively small, the measurement "blur" will dominate making $\bar{Y}_i^\dagger(f)$ and $\bar{Y}_i(\breve{\underline{\theta}})$ appear very similar. Thus, the same estimate of $\bar{Y}_i$ can be used for both arguments of the derivatives of h_i in (25) and (26). For very large smoothing parameters, one would need to form different approximations for the heavily blurred $\breve{\underline{\theta}}$ and the unblurred f, to estimate each measurement mean separately.

In cases where the mean measurements are unknown, a simple plug-in technique where we replace $\bar{Y}_i$ by Y_i often yields very good approximations [3]. This technique tends to produce good estimates, since the Y_i terms are found only "sandwiched" between blur operators that effectively average out much the noise.

Because we will generally be evaluating derivatives of $h_i(u, v)$ with $u = v$, it is interesting to note a few properties of h_i under this condition. First, $h_i^{01}(v, v)$ often equals zero. Such is the case when $h_i(v, u) \leq h_i(v, v), \forall u$. Recall that the second term of h_i represents the mean measurements, and h_i is the log-likelihood for the ith measurement. Thus, this case is satisfied when the log-likelihood attains a peak at its mean. For such noise models, the second term of (25) disappears. Similarly, for many practical noise models like those in Table I, $h_i^{11}(v, v) = -h_i^{02}(v, v)$. Thus, when $g_i(l) = g_i^\dagger(l), \forall i$, the diagonal matrices, $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, are equal, and the local impulse response simplifies to

$$\begin{aligned}\underline{l}(\boldsymbol{x}_0) &= [\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} + \boldsymbol{R}_{\text{sym}}]^{-1}\boldsymbol{A}'\boldsymbol{D}\mathcal{A}\delta_{\boldsymbol{x}_0} \\ \boldsymbol{D}_{ii} &= h_i^{11}(\bar{Y}_i, \bar{Y}_i)\left[\dot{g}_i\left(g^{-1}(\bar{Y}_i)\right)\right]^2, \qquad (27)\end{aligned}$$

where $\boldsymbol{D} \triangleq \boldsymbol{D}_1 = \boldsymbol{D}_2$. For simplicity in the following sections, we will focus on the case where $\boldsymbol{D} \triangleq \boldsymbol{D}_1 = \boldsymbol{D}_2$, although the ideas generalize readily. The efficient methods for calculating the penalty apply in the case where $\boldsymbol{D}_1 \neq \boldsymbol{D}_2$; however, the computation time will be roughly double due to the need to calculate terms for both $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$.

Using (27), one can estimate local impulse responses for many imaging systems. For example, returning to the emission and transmission models in (5) and (6), it is straightforward to calculate the diagonal matrix in (27). Specifically, we find that

$$\boldsymbol{D}_{\text{emis}} = \text{diag}\left\{\frac{1}{Y_i}\right\}, \quad \boldsymbol{D}_{\text{trans}} = \text{diag}\left\{\frac{(Y_i - r_i)^2}{\sigma_i^2}\right\}, \tag{28}$$

where we have used the simple plug-in technique for unknown means. In practice, for the emission tomography case, one must be careful in using Y_i terms that approach zero (this is common for rays that traverse the edges of the object). In these cases, we typically substitute $\max\{Y_i, t_c\}$ for Y_i, where t_c is a relatively small number that discourages overly large weightings. In our experience, this technique provides very good results; however, one might adopt more sophisticated techniques for estimating $1/Y_i$ for low-count cases as Qi has done in [15].

E. Penalty Design

Because our local impulse response approximation (27) is a function of the measurements $\underline{Y}$, but not of the object $\underline{\theta}$, we can find approximate local impulse responses *prior* to image reconstruction. Thus, we can use the impulse response estimates to generate a penalty matrix, $\boldsymbol{R}$, that yields user-specified resolution properties.

Although one can evaluate local impulse responses in (22) or (27) for any spatial coordinates denoted by the continuous variable $\boldsymbol{x}_0$, for penalty design, we would like to implement a design over a finite set of positions. For example, for a pixel basis representation of the object, we can consider a single local impulse response for each pixel. Selecting the position, $\boldsymbol{x}_j$, the center of the jth pixel, is sufficient for penalty design.

Similarly, although (22) or (27) could be evaluated using iterative techniques [3], we would like to evaluate local responses over many locations and would prefer a faster approximate technique for the purpose of penalty design. Because $\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}$ is generally *locally* space-invariant, we use the following circulant approximation (as in [1] and [15]) to the local impulse response at the jth pixel:

$$\underline{l}^j_{\text{circ}} \quad = \quad \mathcal{F}^{-1}\left\{ \frac{\mathcal{F}\{\underline{e}^j\} \odot \mathcal{F}\{\boldsymbol{A}'\boldsymbol{D}\mathcal{A}\delta_{\boldsymbol{x}_j}\}}{\mathcal{F}\{\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\underline{e}^j\} + \mathcal{F}\{\boldsymbol{R}_{\text{sym}}\underline{e}^j\}} \right\} \quad (29)$$

where $\odot$ denotes element-by-element multiplication and the division is an element-by-element division, and $\delta_{\boldsymbol{x}_j}$ denotes a Dirac impulse centered at the center of the jth pixel. The function $\mathcal{F}\{\cdot\}$ takes the 2D Fourier transform of its argument. One can calculate (29) quickly for any j using fast Fourier transforms (FFTs). This circulant approximation includes the term, $\mathcal{F}\{\underline{e}^j\}$, which includes the appropriate complex exponentials such that the response is centered at the jth position, and $\underline{l}^j_{\text{circ}} \approx \underline{l}^j$.

Using the substitution $\boldsymbol{R}_{\text{sym}}\underline{e}^j \approx \boldsymbol{B}^j\underline{w}^j$ in (29), one can consider (29) to be a function of the local weightings $\underline{w}^j$ and use this in a design objective. Strictly speaking this substitution does not yield a symmetric $\boldsymbol{R}$. However, one may calculate a symmetric $\boldsymbol{R}$ after the design, or simply apply an asymmetric $\boldsymbol{R}$, since only the symmetric portion is important for penalties of the form $\frac{1}{2}\underline{\theta}'\boldsymbol{R}\underline{\theta}$. Applying this substitution to (29), one would like to choose weights to approximate some desired response $\underline{l}^j_0$ such that

$$\underline{l}^j_{\text{circ}}(\underline{w}^j) = \mathcal{F}^{-1}\left\{ \frac{\mathcal{F}\{\underline{e}^j\} \odot \mathcal{F}\{\boldsymbol{A}'\boldsymbol{D}\mathcal{A}\delta_{\boldsymbol{x}_j}\}}{\mathcal{F}\{\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\underline{e}^j\} + \mathcal{F}\{\boldsymbol{B}^j\underline{w}^j\}} \right\} \approx \underline{l}^j_0, \quad (30)$$

where $\underline{l}^j_0$ represents the desired response centered at pixel j. As described in [1], one can perform a linearized design by Fourier transforming both sides of (30) and cross-multiplying to obtain:

$$\begin{aligned} \underline{L}^j_0 \odot \mathcal{F}\{\boldsymbol{B}^j\underline{w}^j\} \quad &\approx \quad \mathcal{F}\{\underline{e}^j\} \odot \mathcal{F}\{\boldsymbol{A}'\boldsymbol{D}\mathcal{A}\delta_{\boldsymbol{x}_j}\} - \\ &\qquad \underline{L}^j_0 \odot \mathcal{F}\{\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\underline{e}^j\}, \end{aligned} \quad (31)$$

where $\underline{L}^j_0$ represents the 2D Fourier transform of the shifted desired response. The form of (31) suggested that we could design the local penalty weights, $\underline{w}^j$, using the following constrained least-squares approach [1]:

$$\hat{\underline{w}}^j = \arg\min_{\underline{w}^j \geq \underline{0}} \left|\left| \boldsymbol{\Phi}^j\underline{w}^j - \underline{d}^j \right|\right|^2, \quad (32)$$

with

$$\boldsymbol{\Phi}^j \quad \triangleq \quad \boldsymbol{V}^j D\left[\underline{L}^j_0\right] \mathcal{F}\{\boldsymbol{B}^j\} \quad (33)$$

$$\begin{aligned} \underline{d}^j \quad &\triangleq \quad \boldsymbol{V}^j D\left[\mathcal{F}\{\underline{e}^j\}\right] \mathcal{F}\{\boldsymbol{A}'\boldsymbol{D}\mathcal{A}\delta_{\boldsymbol{x}_j}\} \\ &\qquad -\boldsymbol{V}^j D\left[\underline{L}^j_0\right] \mathcal{F}\{\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\underline{e}^j\}, \end{aligned} \quad (34)$$

where $\boldsymbol{V}^j$ represents a user-selected least-squares weighting that could possibly be space-variant. The above penalty design is a constrained because we would like to ensure a positive definite penalty matrix $\boldsymbol{R}$, so that the objective has a unique maximizer. Simply constraining the weights to be nonnegative is a straightforward way of ensuring this condition; however, alternate approaches have also been investigated in [16]. By systematically evaluating (32) for all pixel positions j, one can design a penalty that leads to increased resolution uniformity.

However, for typical applications, straightforward evaluation of (32) for every pixel generally requires significantly more computation time than it takes to solve the actual image reconstruction problem represented in (3). Therefore, for practical use, it is desirable to find an efficient procedure for computing the penalty.

III. Efficient Penalty Design

In [1], we developed a technique appropriate for efficiently evaluating (32) when the system matrix is factorable such that $\boldsymbol{A}'\boldsymbol{A} \approx \boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}$ and $\boldsymbol{G}'\boldsymbol{G}$ is approximately space-invariant. Unfortunately, such approximate factorizations are not applicable to many systems including SPECT systems and wide field-of-view PET systems. This section describes approximations and observations that reduce the computational burden in evaluating (32) to a practical level even for space-variant systems. One of the key developments that makes this new approach different from the one discussed in [1] is the realization that the the penalty design can be performed in an image domain. This new approach does not rely on any particular factorization of $\boldsymbol{A}$, but does require fast access to the elements of $\boldsymbol{A}$.

Of the factors in (33) and (34) that contribute to the high computational effort, perhaps the most important are the $\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\underline{e}^j$ and $\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\delta_{\boldsymbol{x}_j}$ terms. In tomography, this involves a projection and backprojection for every pixel position. Other factors include the need for 2D Fourier transforms for every pixel position and the necessary shifting operations on $\underline{L}^j_0$ and $\boldsymbol{B}^j$. Another subtle complication is that for asymmetric bases, such as those in (12), the Fourier transforms in (33) and (34) will yield complex numbers, making (32) a complex design problem. The following modifications to the penalty design will alleviate these problems and yield a practical implementation. Specifically, we identify three techniques including particular choice of the weighting matrix, $\boldsymbol{V}^j$, use of a local backprojection, and image subsampling.

However, before we investigate these speed-ups, recall that the formulation in (34) requires knowledge of the actual continuous-to-discrete operator, $\mathcal{A}$. While one could obtain $\mathcal{A}\delta_{\boldsymbol{x}_j}$ by placing point sources in the imaging system with careful registration, for approximate penalty design, we simply approximate these projections using $\boldsymbol{A}\underline{e}^j$. Using this substitution in (34) yields the following simplified

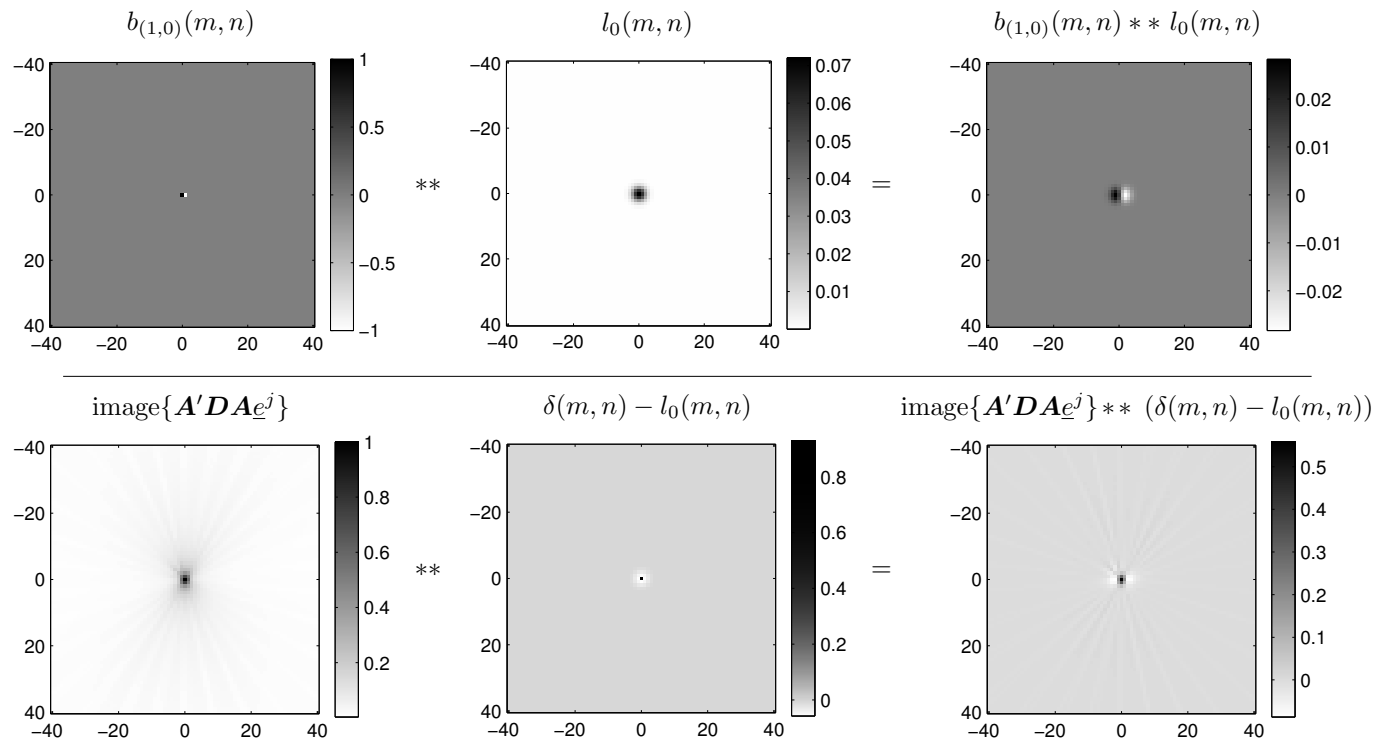

Fig. 1. An example calculation of $\boldsymbol{\Phi}$ and $\underline{d}^j$. The top row shows the convolution operation for one column of $\boldsymbol{\Phi}$, and the bottom row shows a typical convolution operation for calculation of the $\underline{d}^j$ term. While the image size for this example is 170×170 pixels, the above images are presented zoomed-in for clarity. All images have a linear colormap except for $\delta(m,n) - l_0(m,n)$ which has been windowed to show details.

form:

$$\underline{d}^j = \boldsymbol{V}^j D\left[\mathcal{F}\{\underline{e}^j\} - \underline{L}_0^j\right] \mathcal{F}\{\boldsymbol{A'DA}\underline{e}^j\}. \tag{35}$$

Generally, this form also represents a computational speed-up over (34). Thus, hereafter we will focus on (35) for practical penalty design implementations.

A. Appropriate Weighting Matrix

To speed up the penalty design, we first note that we have allowed for an arbitrary least-squares scaling $\boldsymbol{V}^j$. In [1], we considered one choice for $\boldsymbol{V}^j$ for shift-invariant systems that leads to a scalable penalty for a class of desired responses. However, there are many other choices for $\boldsymbol{V}^j$.

Consider the case where the weighting $\boldsymbol{V}^j$ is a matrix representation of the inverse 2D Fourier operator $\mathcal{F}\{\cdot\}$. (We note that this "weighting" is actually a unitary transformation that does not affect the penalty design solution; however, we will continue to describe $\boldsymbol{V}^j$ as a weighting so that a user might conduct a weighted least-squares design.) In this case the convolution property of the Fourier transform simplifies (33) as follows:

$$\boldsymbol{\Phi}^j \;=\; \mathcal{F}^{-1}\left\{D[\underline{L}_0^j]\mathcal{F}\{\boldsymbol{B}^j\}\right\} = \left[\underline{\phi}_1^j \cdots \underline{\phi}_B^j\right], \tag{36}$$

where

$$\underline{\phi}_q^j \triangleq \text{vec}\{b_q(m-m_j, n-n_j) ** l_0(m-m_j, n-n_j)\}, \tag{37}$$

and $l_0(m,n)$ is the 2D function denoting the desired impulse response, $**$ represents a 2D convolution. Similarly, (35) becomes

$$\begin{aligned} \underline{d}^j &= \mathcal{F}^{-1}\left\{D\left[\mathcal{F}\{\underline{e}^j\} - \underline{L}_0^j\right]\mathcal{F}\{\boldsymbol{A'DA}\underline{e}^j\}\right\} \\ &= \text{vec}\left\{ \begin{pmatrix} \delta(m-m_j, n-n_j) - \\ l_0(m-m_j, n-n_j) \end{pmatrix} \right. \\ &\qquad \left. ** \text{ image}\{\boldsymbol{A'DA}\underline{e}^j\} \right\}, \end{aligned} \tag{38}$$

where image$\{\cdot\}$ is the opposite of vec$\{\cdot\}$ operator, reorganizing a lexicographically ordered vector back into an image. We can simplify (36) and (38) by spatially shifting each by $(-m_j, -n_j)$, which we refer to as "centering," to obtain

$$\boldsymbol{\Phi} = \left[\text{vec}\left\{ \begin{matrix} b_1(m,n) \\ ** \\ l_0(m,n) \end{matrix} \right\} \middle| \cdots \middle| \text{vec}\left\{ \begin{matrix} b_B(m,n) \\ ** \\ l_0(m,n) \end{matrix} \right\} \right] \tag{39}$$

and

$$\underline{d}^j = \text{vec}\left\{ \begin{matrix} (\delta(m,n) - l_0(m,n)) ** \\ \text{shift}\left\{\text{image}\{\boldsymbol{A'DA}\underline{e}^j\}\right\} \end{matrix} \right\}, \tag{40}$$

where the shift$\{\cdot\}$ operator shifts an image by $(-m_j, -n_j)$. In terms of the weighting matrix, $\boldsymbol{V}^j$ can be represented by multiplying a permutation matrix by the Fourier matrix. Using (39) and (40) in (32) solves a few issues. First, the only shifts necessary appear in (40). Therefore we can precalculate (39) entirely for a given target response $l_0(\cdot,\cdot)$.

Second, since (39) and (40) are formed by convolution operations, the design is real. Additionally, the application of the inverse Fourier transform means that the design takes place in an image domain where one can make an approximation (discussed next) that reduces computation. Although the work presented in [1] allowed for this kind of least-squares weighting, it was only later that we realized that this could allow for significant speedups by performing an image domain design. This capability is central to the speedups discussed in the following sections.

B. Local Backprojections

Figure 1 illustrates (39) and (40) for a tomography system. The calculation of a single column of $\boldsymbol{\Phi}$ is represented in the upper row of images, and a typical $\underline{d}^j$ evaluation is shown in the bottom row. Because each basis in (12) only uses a small neighborhood, the columns of $\boldsymbol{\Phi}$ (of which the rightmost upper image is an example), generally will be concentrated in a region similar in size to the desired response, $l_0(m,n)$. Similarly, even though $\boldsymbol{A'DA}\underline{e}^j$ can have quite a large support region, $\underline{d}^j$ is often quite concentrated. This is because $\delta(m,n) - l_0(m,n)$ typically takes the form of a high pass filter (for standard choices of $l_0(\cdot,\cdot)$) and $\boldsymbol{A'DA}\underline{e}^j$ is generally smoothly varying.

Because columns of $\boldsymbol{\Phi}$ and $\underline{d}^j$ are highly localized about the origin, we can perform an approximate design using only the central portion. Similarly, even though there may be some structure in $\underline{d}^j$ far away from the origin, these regions are arguably less important for the design. As shown in the top right image in Figure 1, far away from the origin, (small neighborhood) penalties have little influence. Similarly, we expect the approximation in (29) to be less accurate far from position j, which is equivalent to being far from the origin in $\boldsymbol{\Phi}$ and $\underline{d}^j$. Therefore only a small region near the origin need be evaluated. Moreover, one can interpret such a truncation as a specific least-squares weighting $\boldsymbol{V}^j$ that disregards many image positions. This reduction is easily accomplished for columns of $\boldsymbol{\Phi}$. However to decrease computation time for $\underline{d}^j$, one must have fast access to the columns of $\boldsymbol{A}$. Direct access to the columns of $\boldsymbol{A}$ allows one to obtain $\boldsymbol{A}\underline{e}^j$ easily, as is needed for evaluating $\underline{d}^j$, and to compute backprojections to a region. Such is the case when the system matrix has been precomputed and stored. In contrast, system models that are implemented in a procedural form and have only row access will not be able to fully exploit these speed-ups.

Disregarding portions of $\underline{d}^j$ that may be precomputed, one can evaluate (40) using radix-2 FFTs in approximately $3p[\frac{2}{3}\alpha N + \log_2(P)]$ floating point operations, where α represents a factor indicating the sparsity of $\boldsymbol{A}$. For the PET system investigated in Figure 1, where $\alpha = 3.4\%$, $P = 170^2$, and $N = 120 \cdot 240$, each calculation of $\underline{d}^j$ would take approximately 60 Mflops. In contrast, consider the technique where only a subset of pixels are backprojected and blurred. When a subset of $[32 \times 32]$ pixels is used, $P = 32^2$ and calculating $\underline{d}^j$ takes 2 Mflops, a factor of 30 speed-up. The small support method also has the advantage of reducing the dimensions of $\boldsymbol{\Phi}$ and $\underline{d}^j$, so that the constrained design in (32) may be computed more quickly. However, calculations remain dominated by the evaluation of the $\underline{d}^j$ terms.

C. Image Subsampling

Because $\boldsymbol{A'DA}\underline{e}^j$ often varies smoothly with position, the designed penalty weights, $\underline{\hat{w}}^j$ also tend to vary smoothly with position. Therefore another potential speedup can be made by simply evaluating the penalty design of (32) over a subset of image positions and using interpolation to find the weights for the remaining positions. For example, since computation time is dominated by the computation of $\underline{d}^j$, finding $\underline{\hat{w}}^j$ for every third pixel (in both horizontal and vertical directions) and interpolating remaining weights yields a factor of 3^2 speed-up. Selection of the coarseness of this sampling depends on the system characteristics and the desired trade-off between computation time and quality of resolution control.

D. Summary

For fast computation of the penalty, we propose using all of the computational improvements discussed in the previous three subsections. Although each may be applied in varying degrees (*i.e.*: support size and coarseness of subsampling), together they may be used to make shift-variant penalty design practical. The overall procedure for generating the shift-variant penalty (for shift-variant systems) is shown in detail in Table II. These same methods might also be applied to the shift-invariant procedures discussed in [1] to achieve additional speedups; however, in this paper we continue to focus only on systems that are intrinsically shift-variant.

IV. Resolution Results

This section examines the resolution uniformity of our proposed penalized-likelihood technique and of other reconstruction techniques for both a space-variant SPECT system and a space-variant PET system. Both investigations adopt a Poisson noise model and a linear measurement model as in (5). We also describe the practical computational burden for the penalty design.

A. Application to 2D SPECT

We first investigate the resolution properties of a space-variant SPECT system. We adopt a SPECT (reconstruction) system model whose circular orbit contains a field-of-view of $[128 \times 128]$ 2 mm pixels. The detector head rotates at a radius of 12.8 cm, and collects data for 110 projection angles over 360° with 128 evenly spaced 2 mm radial bins. The system response is modeled after a high resolution collimator with a linearly varying depth-dependent Gaussian response that has a 1.75 mm FWHM at face of the collimator and a slope of 0.044, which corresponds to about 7.4 mm FWHM at the center of the field-of-view. We model the true projections (*i.e.*, the $\mathcal{A}$ operator) using a discrete system model that is upsampled by a factor of three. That is, the image-domain support contains $[384 \times 384]$ pixels

TABLE II
OVERALL PROCEDURE FOR GENERATING THE SHIFT-VARIANT PENALTY

Calculate $\phi_q = b_q(m,n) ** l_0(m,n)$ for $q = 1...B$ by convolving the desired response with each basis over a small $[M \times M]$ support centered at (0,0).
Construct columns of Φ in (39) from vectors formed from lexicographically-ordered elements of ϕ_q.
Precompute $\Phi'\Phi$.
For each pixel, j, on a coarse grid:
    Extract the jth column of $\boldsymbol{A}$ to obtain $\boldsymbol{A}\underline{e}^j$.
    Calculate the elements of $\boldsymbol{D}$, using the measurements and (27): $\boldsymbol{D}_{ii} \approx h_i^{11}(Y_i, Y_i)\left[\dot{g}_i\left(g^{-1}(Y_i)\right)\right]^2$
    Multiply $\boldsymbol{A}\underline{e}^j$ element-by-element with $\boldsymbol{D}_{ii}$ to find $\boldsymbol{D}\boldsymbol{A}\underline{e}^j$.
    Perform a local backprojection to obtain a subset of $\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\underline{e}^j$, using only the columns of $\boldsymbol{A}$ that represent pixels within an $[M \times M]$ region of support about the pixel j.
    Convolve this locally backprojected image with the high-pass filter $\delta(m,n) - l_0(m,n)$.
    Reshape this small support filtered backprojection into the vector $\underline{d}^j$.
    Use $\Phi'\Phi$ and $\Phi'\underline{d}^j$ in an algorithm like NNLS [17] to estimate the local penalty weights, $\underline{\hat{w}}^j$.
end
For each pixel, k, not on the coarse grid:
    Use neighboring samples of $\underline{\hat{w}}^j$ from above to estimate $\underline{\hat{w}}^k$ using bilinear interpolation.
end

for the true projector. The projections and reconstruction models are matched in all other respects.

We simulated a 23 cm diameter cold rod phantom with uniform attenuation coefficient of 0.015 mm^{-1} (the approximate attenuation coefficient of water at 140 keV) and rod diameters of 6.4, 9, 10.25, 12.8, 17.9, and 25.6 mm. The emission image for this object is shown in Figure 2A. To represent scatter, the model includes 5% uniformly distributed background events and 500 thousand counts total.

Before discussing the resolution properties of various reconstruction techniques, we first demonstrate the feasibility of the proposed space-variant design in terms of computation time. Table III lists computation times for the space-variant penalty for this SPECT system using a `gcc`-compiled ANSI C implementation of the design algorithm discussed in Section III. For comparison, the time to complete a single projection-backprojection, (*i.e.*, $\boldsymbol{A}'\boldsymbol{A}\underline{\theta}$), is approximately 1.5 seconds. We present results for two different support sizes about the origin and four different spatial subsamplings (*i.e.*, evaluating at every nth pixel and filling in gaps by interpolation). Due to zero padding and the use of radix-2 FFTs, the $[20 \times 20]$ support size uses $[32 \times 32]$ FFTs and the $[12 \times 12]$ support uses $[16 \times 16]$ FFTs. (Technically the zero padding applied these cases is insufficient to completely eliminate wrap-around effects from periodic convolution. However, because the $\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\underline{e}^j$ responses are fairly smooth and the blur operation uses a high pass filter, we accept small amount of wrap-around in the penalty design to reduce computation.) All methods used a second-order penalty, incorporating the eight nearest pixels. The computation times are very reasonable, particularly for the larger subsampling values.

Figure 2 shows images reconstructed from noiseless projections of the cold rod phantom using a variety of techniques. Since the different methods have different resolution properties, we have attempted to match resolution as closely as possible for the *center pixel* in the image. We have chosen the following target impulse response,

$$\underline{l}_0 = [\boldsymbol{A}'\boldsymbol{A} + \boldsymbol{R}_0]^{-1}\boldsymbol{A}'\boldsymbol{A}\underline{e}^{j_0}, \tag{41}$$

where we have selected a conventional space-invariant penalty and j_0 denotes the center pixel in the image. Equation (41) represents the local impulse response for a conventional penalized unweighted least-squares reconstruction. We evaluate the target response (41) using iterative techniques as described in [3]. This response is essentially radially-symmetric since the response lies at the center pixel for a SPECT model that incorporates a circular orbit, and since the object is a centered, uniformly attenuating, disc-shaped object. For $\boldsymbol{R}_0$ in (41), we chose a standard penalty matrix that uses a first-order neighborhood of four equally-weighted pixels with a weighting chosen to yield a FWHM resolution of 10 mm.

Choosing the target response (41) allows one to match exactly the reconstruction resolution for many methods since it represents a form achievable by many penalized-likelihood and filtering methods. Figure 2E shows the true image downsampled to $[128 \times 128]$ and blurred with the target response (41).

TABLE III
CALCULATION TIMES FOR THE PROPOSED PENALTY ON AN 800 MHZ PENTIUM-III PROCESSOR.

Spatial Subsampling	$P = [20 \times 20]$	$P = [12 \times 12]$
1	128 s	60 s
2	33 s	16 s
3	15 s	8 s
5	6 s	4 s

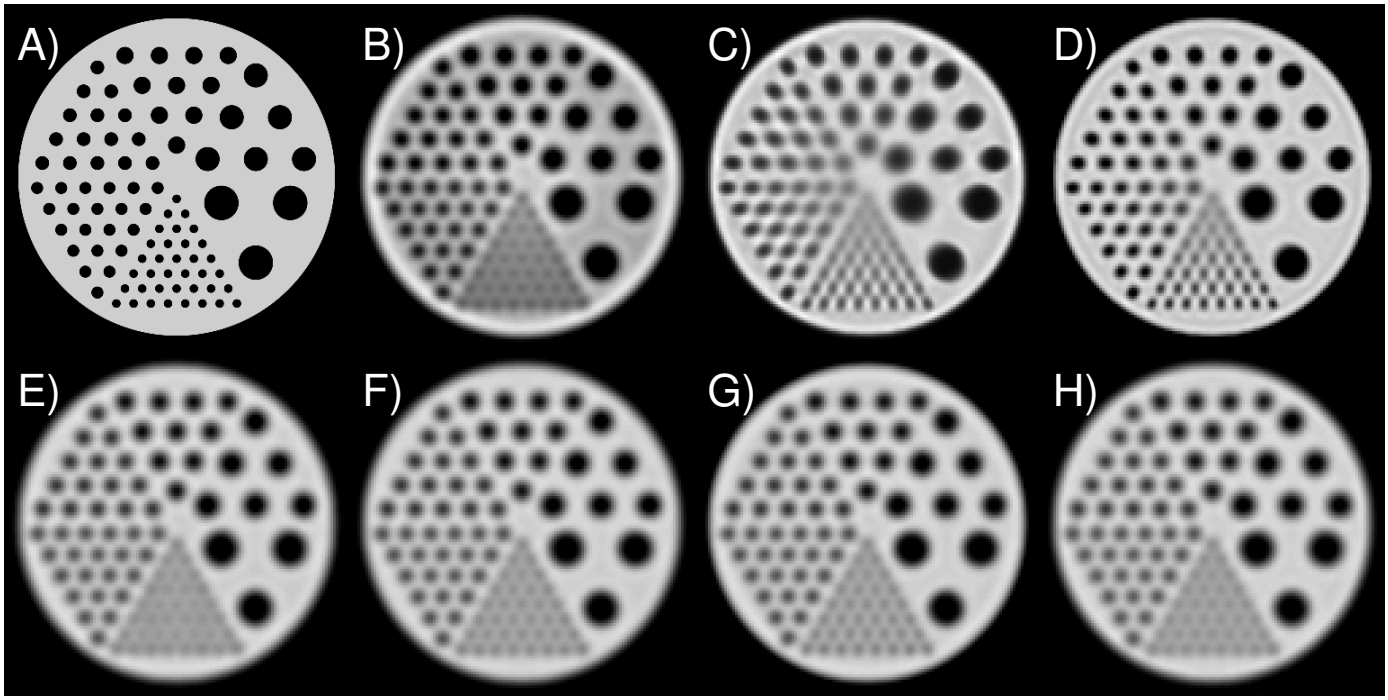

Fig. 2. Noiseless 2D SPECT reconstructions: A) True emission image, B) FBP with uniformity correction using the frequency-distance principle and attenuation correction, C) Truncated OSEM, D) PL with standard space-invariant penalty, E) (Subsampled) true image smoothed with desired blur, F) Post-smoothed ML, G) PL with modified penalty, and H) the hybrid post-smoothed PL approach of Section VI.

Figure 2B shows a filtered-backprojection (FBP) reconstruction using the frequency-distance principle to correct for nonuniform resolution [7] and Chang-type attenuation correction. Because the frequency-distance principle does not completely correct for the nonuniform detector response, we use the following approach to match the resolution properties with the target response in (41).

When the response of an estimator, such as FBP, is known, and does not match (41) perfectly, one can force a match by applying post-filtering. The overall response is then a combination of the estimator response and the post-filter. Specifically,

$$l_{\text{overall}}(m,n) = l_{\text{est}}(m,n) ** l_{\text{post}}(m,n), \tag{42}$$

where $l_{\text{overall}}(m,n)$, $l_{\text{est}}(m,n)$, and $l_{\text{post}}(m,n)$ represent the overall response, the response due to the estimator, and the post-smoothing filter, respectively. Thus, given an overall desired target response and the estimator response, one can find the appropriate post-smoothing filter by

$$l_{\text{post}}(m,n) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\left\{l_{\text{overall}}(m,n)\right\}}{\mathcal{F}\left\{l_{\text{est}}(m,n)\right\}}\right\}. \tag{43}$$

Depending on the form of $l_{\text{est}}(m,n)$, it may not be possible to obtain any overall desired response because of zeros in the frequency domain. However, one can find approximate post-filters for a wide range of overall desired responses.

Therefore, even though ramp-filtered FBP with the frequency-distance-based uniformity correction yields an imperfect response, we match the overall target response, (41), by using a post-filter calculated from (43). Because the ramp-filtered FBP estimator generally yields space-variant results, we match the target response only at the center pixel. That is, we find $l_{\text{est}}(m,n)$ for the center pixel by propagating an impulse through the ramp-filtered FBP estimator, and find a single shift-invariant post-filter using (43) to match the target response. The resulting reconstruction, shown in Figure 2B, has relatively good resolution uniformity, but suffers from ringing artifacts, most noticeable at the edges of the object.

Figure 2C shows a reconstruction using an ordered subsets expectation maximization (OSEM) algorithm with 10 subsets. We initialize the algorithm with a uniform image and perform nine iterations. Starting with a flat image and using only a few iterations is sometimes used as a noise-control technique, since higher spatial frequencies generally take more iterations to appear in the image estimate. The resolution properties are highly nonuniform, and only roughly matched even at the center due to the poor (object-dependent) resolution control available with this method.

Figure 2D shows a standard penalized-likelihood (PL) reconstruction using a space-invariant penalty. We may write the local impulse response response for this estimator at the center pixel as

$$\underline{l}_{\text{PL}}^{j_0} \quad = \quad [\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} + \beta \boldsymbol{R}_0]^{-1}\boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}\underline{e}^j. \tag{44}$$

However, for the center pixel the diagonal weighting denoted by $\boldsymbol{D}$ is very uniform and can be approximated using a single scalar value d. Therefore, we may rewrite the

response in (44) as

$$\begin{aligned} l_{\mathrm{PL}}^{j_0} &\approx [d\boldsymbol{A}'\boldsymbol{A} + \beta\boldsymbol{R}_0]^{-1} d\boldsymbol{A}'\boldsymbol{A}\underline{e}^j \\ &= [\boldsymbol{A}'\boldsymbol{A} + \frac{\beta}{d}\boldsymbol{R}_0]^{-1}\boldsymbol{A}'\boldsymbol{A}\underline{e}^j. \end{aligned} \quad (45)$$

Thus, using the same penalty as in (41) with an appropriate scaling β, we have matched the center pixel's response nearly exactly. We estimate the solution with 200 iterations of an ordered subsets version of De Pierro's algorithm [18] with 10 subsets, initialized with an FBP reconstruction, followed by 20 iterations with one subset. For typical image reconstruction problems, this represents many more iterations than are generally necessary to form a good image estimate. However, we would like a solution that is well-converged so that we may guarantee that any resolution mismatches (or, noise mismatches in Section V) are due entirely to the objective function, not to insufficient convergence of the algorithm used to find the estimate. While the resolution properties for the PL estimate in Figure 2D are nearly exactly matched at the center, the nonuniform resolution properties away from the center are clearly evident.

Figure 2F is a reconstruction using a post-smoothed ML technique, using 200 OSEM iterations (10 subsets) initialized with an FBP image, followed by 20 EM iterations to ensure a nearly converged solution. Since we have post-smoothed with the desired target response in (41), the resolution properties are essentially exactly matched, as seen by comparing Figure 2E and Figure 2F.

Lastly, we applied our proposed space-variant penalty, using 200 iterations of the ordered-subsets De Pierro's algorithm (10 subsets), initialized with an FBP image and followed by 20 iterations using one subset. Figure 2G shows the reconstruction resulting from our penalty design using the $[20 \times 20]$ support with no spatial subsampling. The resolution properties are virtually exactly matched at the center since the target response is easily achieved using the space-variant design. That is, because a space-invariant penalty achieves this response, the space-variant design easily achieves the same response. The global resolution properties are mostly very uniform, with some mild nonuniformities at the object edges, where approximation (27) is less accurate.

Using the other choices of support size and spatial subsampling shown in Table III yielded nearly identical results in the interior of the object. Significant nonuniformity was noticeable only at the edges of the object when using coarser spatial subsampling. One could use a region-dependent subsampling of positions in (32) to sample more finely at the object edges to provide nearly the same results with fast computation.

Another way to investigate the resolution properties of a technique is to evaluate the local impulse response at a variety of locations and compare them to the target response. For most statistical methods we evaluate (22) using iterative techniques (we choose 100 iterations of a coordinate ascent algorithm initialized with the target response). For ML techniques where the invertibility conditions for (22) may not hold, we use the techniques described in [3] and [19], where the emission image is perturbed with an impulse, and differences in reconstructions with and without the perturbation are obtained. For linear techniques like FBP, we simply propagate an impulse response through the system to find the local response. A sampling of local impulse responses is shown in Figure 3. The local impulse responses are contoured at four levels indicating the 25%, 50%, 75%, and 99% levels of the target response.

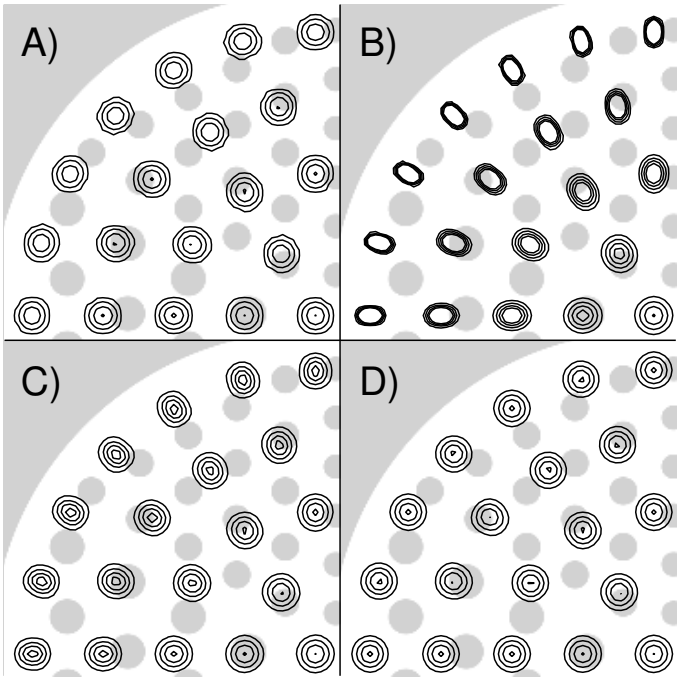

Fig. 3. SPECT local impulse responses for a 10.0 mm FWHM target using A) Uniformity and attenuation-corrected FBP, B) PL with space-invariant penalty, C) PL with proposed penalty, and D) post-smoothed ML. All responses are superimposed on the upper left quadrant of the phantom to illustrate the sample locations for these impulse response.

The relatively narrow responses of conventional PL are evident away from the center of the object in Figure 3B. In contrast, the uniformity-corrected FBP, PL with the space-variant penalty, and post-smoothed ML, shown in Figures 3A, 3C, and 3D, respectively, yield very uniform responses. That is, the responses show a high degree of symmetry and spatial uniformity, and the response peaks and contours are closely matched to the target in (41). The response of the center pixel (shown in the lower right corner of each subfigure) is indistinguishable from the target response for all these methods. (We do not present local impulse responses for OSEM with truncated iterations, however we would expect very nonuniform responses that have mismatch even at the center pixel.) Post-smoothed ML appears to have the best uniformity, whereas our proposed PL method shows very slight asymmetries at the edges of the object.

While post-smoothed ML appears to yield more uniform resolution properties than the proposed PL technique, we

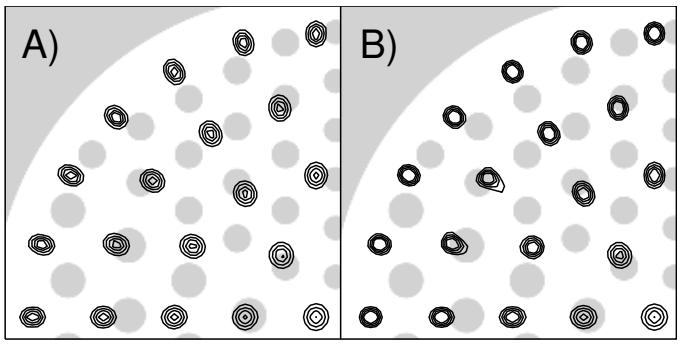

Fig. 4. SPECT local impulse responses for a 7 mm FWHM target using A) PL with proposed penalty and B) post-smoothed ML.

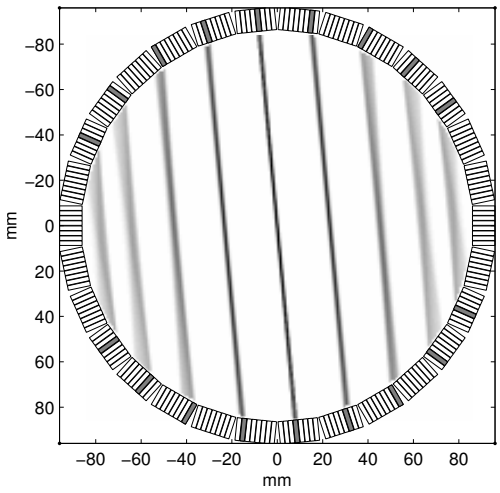

Fig. 5. A space-variant PET system. The above PET system model has physical characteristics chosen to simulate the MicroPET rodent scanner. In the above image, rays connecting every 10th detector are shown. Both nonuniform sampling and varying detector response due to crystal penetration are incorporated in this model.

find that there are still resolution nonuniformities for the post-smoothed ML techniques. When we investigate the resolution properties of conventional ML with no filtering through a systematic evaluation of local impulse responses, we find that the FWHM resolution of the responses varies from about 3 mm at the edges of the phantom to about 6.5 mm at the center. This is an indication that the system matrix, $\boldsymbol{A}$, is rank-deficient, and the ML estimator cannot resolve single pixels. Thus, the post-smoothed estimates must also have nonuniform resolution properties. For relatively large target responses, the post-smoothing blur dominates and these nonuniformities are very small (as we have seen for the 10.0 mm target). However, for smaller desired responses, simple post-smoothing will not yield the desired target. However, we can adopt a post-filter approach that compensates for the intrinsic blur of the ML estimator by applying (43).

We use (43) to find an post-smoothing filter for ML for a target response of the form in (41) with a FWHM resolution of 7 mm. Figure 3 shows local impulse responses for the 7 mm target for the proposed PL estimator and for the post-smoothed ML approach. Despite matching the target response at the center pixel (lower right corners in each subfigure in Figure 4), the ML approach clearly yields nonuniform resolution properties with narrower responses toward the edges. In comparison, the PL approach yields more uniform results.

In summary, the only 2D SPECT reconstruction methods presented here that yield nearly uniform resolution properties are post-smoothed ML (for larger FWHM targets) the proposed PL approach, and FBP with frequency-distance corrections. We compare the noise properties of these methods in Section V.

B. Application to PET

We have also applied our penalty design technique to a space-variant small animal PET system. Specifically, we have modeled a MicroPET rodent scanner. [20] This system has 2 mm (square) by 10 mm crystals in 30 $[8 \times 8]$ blocks. The full field-of-view of $[170 \times 170]$, 1 mm pixels, is modeled using finite integration of over all angles and pixels, and includes crystal penetration effects. Figure 5 shows responses for detectors pairs over regular intervals. Both the nonuniform sampling and the space-variant detector responses are evident in this figure.

Figure 6 shows a sample digital phantom placed in the scanner. This image shows the attenuation map for a digital rat phantom in a slice at the bottom of the pelvis, where the rat takes up a very large portion of the field-of-view. This data was obtained by manually segmenting MRI data obtained from [21]. The attenuation values are 0.0096, 0.013, and 0.010 mm^{-1} for the soft tissue, bone, and the table, respectively. The emission image has a uniform background with emission rate of 1.0, and a single circular lesion in the right half of the phantom with an emission rate of 2.0 (indicated in Figure 6 by the white circle). Projections contain 10 million counts with 5% percent random coincidences.

Figure 7 presents local impulse response contours for four different reconstruction methods. Again, contours are made at the 25%, 50%, 75%, and 99% levels of the target response. The target response was generated by angularly averaging (41) for the PET model to obtain a symmetric response with a FWHM resolution of 4.0 mm. The local impulse response contours are superimposed on the rat slice emission image so that the position of the responses are obvious. We performed FBP reconstruction by radially resampling the cylindrical projections (arc correction). As in the SPECT case, we used the least-squares filter of [19] in an attempt to match resolution properties. Figure 7A shows responses for FBP, which while relatively well-matched at the center, have reduced peaks toward the edges, indicating coarser resolution properties as expected.

In contrast, the responses arising from conventional PL (first-order penalty) shown in Figure 7B are narrower at the edges. There are competing effects in PL reconstruction. While the system model suggests decreased resolu-

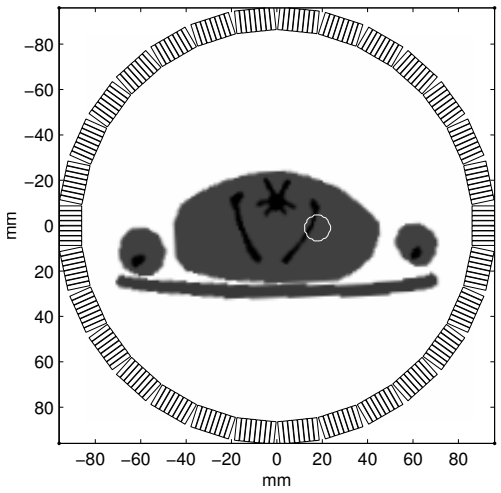

Fig. 6. The above image shows the PET system with a simulated rat phantom inside. This figure shows the attenuation map for a transverse slice of a rat's lower pelvis and thighs, where the animal takes up a very wide field-of-view. The emission image is uniform, except for a hot lesion in the location indicated by the white circle.

tion at the edges due to the detector responses, there is actually finer sampling at the edges (in effect better conditioning the reconstruction than if uniformly sampled data were acquired). However, for emission tomography the FWHM resolution of conventional PL varies inversely with ray certainty. Thus at the edges, where ones obtains lower count measurements and thus increased certainty (under the Poisson model), one expects decreased (finer) resolution. While these competing effects actually appear to yield more uniform resolution than if the system model were idealized to have uniformly sampled projections, the effects of attenuation are clear in the responses, resulting in greater vertical smoothing.

Figure 7C shows contours for PL with the proposed space-variant (second-order) penalty. The responses are very uniform in the interior of the object, but degrade near the edges and outside the object. In general, the proposed technique yields more uniform results than conventional PL. If more uniform results are desired, a larger order penalty neighborhood may be required, or relaxed design constraints (as in [16]) may need to be applied.

Lastly, we present the contours for the case of post-smoothed ML in Figure 7D. These responses are very uniform throughout the image and are very well matched to the target response. We find the greatest uniformity and the ability to match a target response with the post-smoothed ML and proposed PL techniques.

V. Estimator Performance

Given different methods that provide uniform resolution properties, we would like to be able to choose the "best" method for reconstruction. Of course, "best" is a highly subjective term until a particular figure of merit is selected. Computation time, degree of resolution uniformity, and the variance and autocorrelation functions for

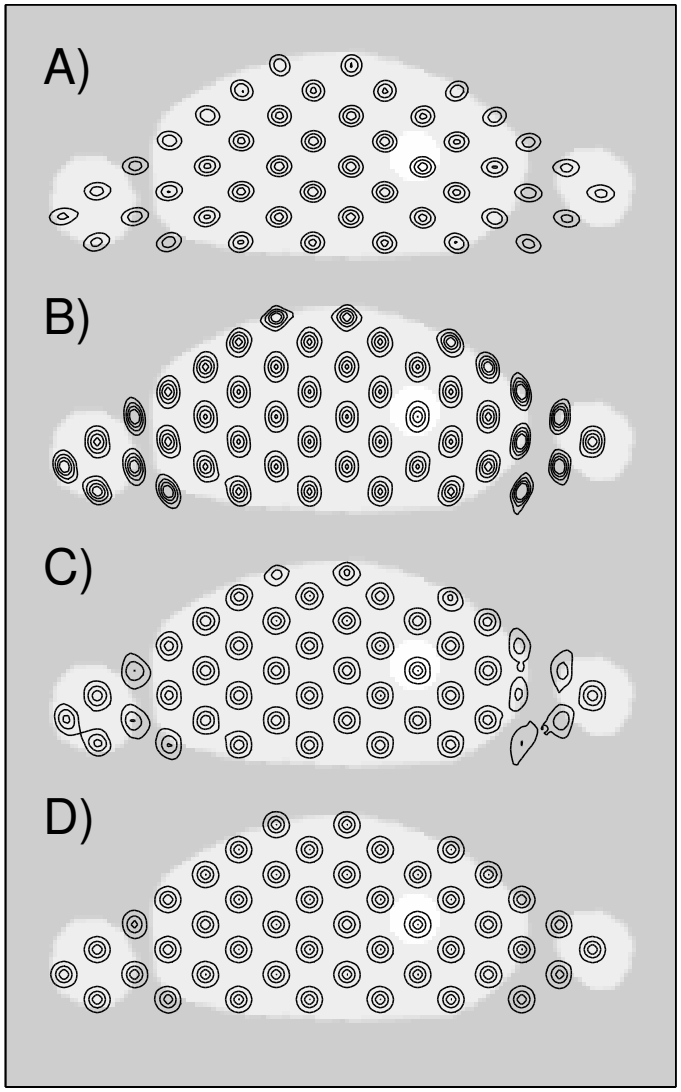

Fig. 7. Contours of the local impulse responses for A) filtered-backprojection, B) PL with space-invariant penalty, C) PL with proposed penalty, and D) post-smoothed ML. Contours are superimposed on the emission image to show position.

the reconstructed pixels are a few categories that might be important. While pixel variances are often maligned as being only obliquely related to performance for certain tasks, they are a simple place to start, and we would like to ask the question, "Among estimators with exactly matched resolutions, which one yields the lowest variance in the reconstruction?"

We first must identify estimators with nearly exactly matched resolution properties. We have found that simply matching FWHM resolution is insufficient for comparison, as the sidelobe behavior and overall shape of the response can greatly affect the noise performance. Recalling the investigations in the previous sections, the only uniform resolution methods we have investigated with well matched responses are uniformity-corrected FBP, post-smoothed ML, and our proposed PL technique. Additionally, we know that these methods are not globally exactly matched. In

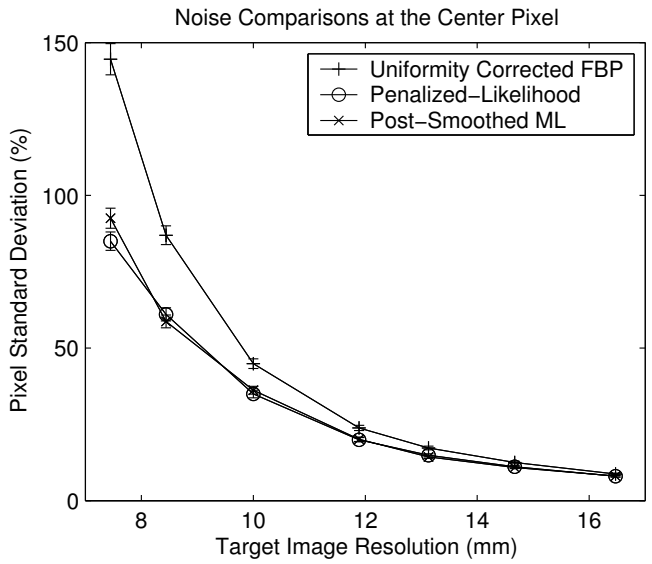

Fig. 8. Standard deviations for the uniformity-corrected FBP (+), the penalized-likelihood (∘), and the post-smoothed maximum-likelihood (×) techniques for the center pixel where the local impulse responses are exactly matched over a range of target FWHM resolutions.

practice we can match these methods at (at least) one pixel by choosing to post-smooth the ML and FBP approaches using (43) and a target response equal to the estimated PL response. Other pixel positions will generally be only approximately matched.

A. Noise Study

Returning to the SPECT model, we performed 400 noisy reconstructions for the uniformity-corrected FBP techniques, our PL approach, and the post-smoothed ML method. This was performed over a range of target resolutions with FWHM from 7.5 mm to 16.5 mm, using the target response of (41). No targets below 7 mm were calculated because even unpenalized ML yields a response of about 6.5 mm at the center pixel. This minimum resolution represents a barrier for both methods since the PL method approaches the ML estimate for small target resolutions. We chose the post-filters for the FBP and ML techniques using (43) over the entire range of targets. Thus the resolution properties are essentially exactly matched for all methods at the center even for the smaller target responses.

Figure 8 shows standard deviations for the center pixel for these methods. One standard deviation error bars are shown for each estimate. The plots for the proposed PL approach and the post-smoothed ML estimates are nearly identical with small differences well within the error bars. (The one exception is the mismatch at the finest resolution point. This is most likely due to a residual resolution mismatch which stems from a difficulty in matching resolutions exactly for small responses.) Thus, in terms of variance the methods appear to have the same noise performance when the spatial resolutions are carefully matched. In contrast, the FBP approach suffers from increased noise in the reconstructions.

One can also study the covariances in the reconstructions. Covariance functions are arguably a more important feature than variances for evaluating different methods with specific tasks in mind. We calculated the sample covariance function for the center pixel for the uniform resolutions methods using the 400 reconstructions. The covariance for the post-smoothed ML and our proposed PL approaches were indistinguishable. Thus, for this system and target, the post-smoothed ML and PL approaches have essentially the same noise performance. Neither method appears to have an advantage over a wide range of practical reconstruction resolutions. This result is not entirely unexpected. In the Appendix we present an analysis for a linear measurement model and a Gaussian noise model, and argue that the post-smoothed ML and exactly matched PL methods should yield identical covariance properties. Thus for cases where the Poisson statistics are modeled well by a Gaussian approximation, it is not surprising that the same conclusions hold. In contrast, the uniformity-corrected FBP yields different covariance functions. However, whether or not FBP's covariance is desirable will depend on the task for which the images are made and if the associated reconstruction artifacts are tolerable.

We also explored the noise performance at other pixel locations found similar equivalence of post-smoothed ML and PL for the target in (41). However, for other pixel positions and other targets at finer matched resolutions sometimes one or the other algorithm would have lower standard deviation depending on the particular location and target response. Rather than attempting to draw general conclusions about the relative merits of the two approaches it seems advisable for algorithm designers to compare the two for the given system model and target resolutions of interest.

B. Convergence Rates

Since the noise performance for PL and post-smoothed ML are indistinguishable for the investigations in the previous section, other considerations such as computation time may be more important. It is popularly held that unregularized methods converge more slowly than regularized methods due to the conditioning of the problems. However, unregularized algorithms converge to different limits than the regularized algorithms making analytical comparisons difficult.

We performed a simple investigation of the convergence rates of matched post-smoothed ML and PL approaches. We compared the normalized mean squared difference between the image estimate at the nth iteration, $\underline{\hat{\theta}}_n$, and the fully-converged solution, $\underline{\hat{\theta}}_\infty$. For the PL approach, $\underline{\hat{\theta}}_n$ is simply the estimate at the nth iteration. For the post-smoothed ML technique, $\underline{\hat{\theta}}_n$ is the ML estimate at the nth iteration, with a post-smoothing filter applied. Thus, $\underline{\hat{\theta}}_n$ represents the *post-smoothed* ML estimate at the nth iteration. In this way, $\underline{\hat{\theta}}_\infty$ for each method is nearly the same; however, there will be marginal differences from the residual resolution mismatch of the PL approach.

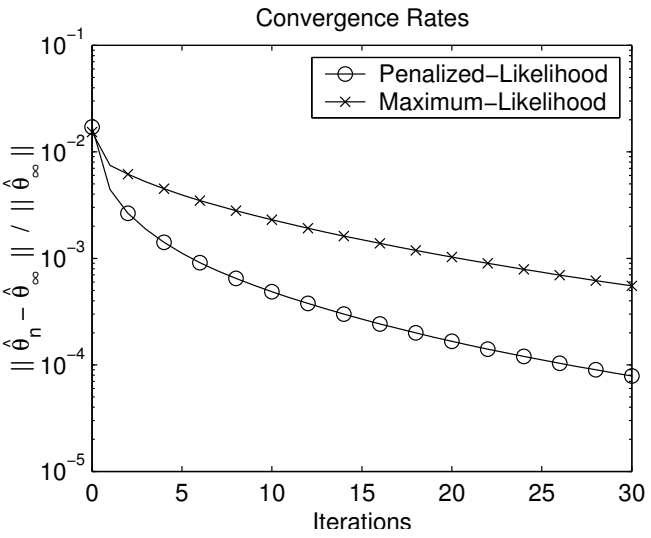

Fig. 9. Convergence rates of PL (∘) and post-smoothed ML (×) approach for a 10 mm target FWHM resolution.

We initialized both methods using the same FBP image and used the same ordered-subsets techniques and the same 10.0 mm target response mentioned in Section IV. Estimates of $\underline{\hat{\theta}}_\infty$, were calculated using 500 ordered-subsets iterations, followed by 100 single subset iterations.

Figure 9 shows that the PL approach converges more quickly than the ML approach. For a similar level of convergence, it appears that the ML technique takes roughly three times the number of iterations. Such speed-ups depend on the target resolution, since increased regularization leads to better conditioning. For example, smaller target resolutions require more iterations, and as the target is made arbitrarily small we will approach the unregularized ML problem both in terms of solution and in terms of convergence rate. However, we expect similar rank performance for the two methods for different target resolutions, as long as the same iterative algorithms are applied to each method.

VI. A Hybrid Technique

An interesting alternative to choosing between post-smoothed ML and the space-variant PL approaches is to use both! One can use a hybrid method that includes a degree of regularization that keeps responses fairly uniform and increases convergence rates for iterative algorithms, and then apply a post-smoothing filter to set the overall target resolution. This approach is attractive for a number of reasons. Using the PL approach keeps the responses uniform even for fairly small target responses. Post-smoothing will generally reduce any of the remaining resolution nonuniformities, and can be applied quickly for a number of desired FWHM resolutions or responses without additional iterative reconstructions. And, convergence rates are increased over the unregularized ML approach, reducing computation.

This hybrid post-smoothed PL approach can implemented easily, using (43) to find the appropriate post-smoothing filter for a desired overall response. In this case, the l_{est} term represents the "first-pass" resolution induced by the penalized-likelihood objective. For our proposed space-variant penalty, l_{est} is equal to the "first-pass" target response. Figure 2H shows a sample reconstruction using this technique for the SPECT problem. For this hybrid estimator, we apply our PL approach with a target of the form in (41) with a 7.5 mm FWHM, followed post-filtering via (43) using same overall target as the other methods shown in that figure. One can see the increased uniformity as compared with the non-hybrid PL approach in Figure 2G.

We expect this method to yield similar noise performance as the individual PL or post-smoothed ML approaches. However, other trade-offs remain. Specifically, the optimal choice of the "first-pass" response must be determined by balancing overall resolution uniformity against computation time (both for the obtaining the space-variant penalty and for the convergence rates of the iterative algorithm used to solve the reconstruction objective).

VII. Discussion

We have presented a general penalized-likelihood framework for reconstructing images with uniform resolution for space-variant imaging systems. Additionally, we have demonstrated computationally feasible techniques for calculating the space-variant penalty when the system matrix has been precomputed. In an investigation of emission tomographic systems, we found uniformity-corrected FBP, post-smoothed ML, and PL with our space-variant penalty to yield nearly uniform resolution properties. For small target responses, the post-smoothed ML approach can have significant nonuniformities, if the system model is rank deficient. However, the PL approach still possesses residual nonuniformities most notably at the edges of the object. One solution that can eliminate these nonuniformities is to adopt the hybrid approach discussed in Section VI. This hybrid approach is also attractive since different resolutions can be quickly applied without further applications of the iterative estimation procedure.

Because this penalty design approach neglects nonnegativity constraints on the object, it may be difficult to exert fine control of the resolution properties of the reconstructed object in regions where the reconstructed object values are nearly zero. For example, nonnegativity constraints would obviously affect the negative sidelobes of a target response at edges near a zero-valued region. However, coarse resolution control can still be made using the techniques outlined in this paper. We should note that while some target responses (such as those with negative sidelobes) may incompatible with certain objects and nonnegativity constraints, there are many other targets (e.g.: nonnegative responses) that might benefit from a future penalty design that has been derived with these constraints in mind.

In a noise investigation of these uniform resolution approaches we found that the post-smoothed ML and PL methods have essentially identical noise performance over a wide range of target responses when spatial resolution

properties are very carefully matched. These empirical results are not unexpected, since under certain system and noise models, one can show theoretically the equivalence of the methods in terms of noise (see Appendix). Because these methods yield the same noise performance, selection of one technique over another may depend on other trade-offs like computation time and overall resolution uniformity. These trade-offs may be explored further using the hybrid post-smoothed PL technique. In contrast, the FBP approach has different noise properties with a significant increase in reconstructed pixel variance.

Because our investigations really only apply to the systems we have investigated, other imaging systems, noise models, and target resolutions may yield different trade-offs or conclusions about which techniques are preferred. We have presented a class of uniform resolution methods from which one may select a specific estimator according to the desired features of the estimator. Future work should include investigations of other systems and noise models, to determine whether the conclusions presented here apply more generally.

Appendix

Following [22], this appendix describes conditions under which a post-filtered weighted least-squares reconstruction is identical to a penalized weighted least-squares (PWLS) reconstruction. This mathematical equivalence corroborates our empirical findings for post-filtered ML and penalized-likelihood reconstructions in Section V. For a linear measurement model, a PWLS estimate maximizes an objective function of the following form:

$$\boldsymbol{\Phi}(\underline{\theta}, \underline{Y}) = -\left(\underline{Y} - \boldsymbol{A}\underline{\theta}\right)' \boldsymbol{K}^{-1} \left(\underline{Y} - \boldsymbol{A}\underline{\theta}\right) - \underline{\theta}' \boldsymbol{R} \underline{\theta}, \qquad (46)$$

where $\boldsymbol{K} = \mathrm{cov}\{\underline{Y}\}$. Assuming appropriate invertibility conditions hold, the minimizer has the following closed-form solution

$$\hat{\underline{\theta}} = [\boldsymbol{A}'\boldsymbol{K}^{-1}\boldsymbol{A} + \boldsymbol{R}_{\mathrm{sym}}]^{-1} \boldsymbol{A}'\boldsymbol{K}^{-1}\underline{Y}, \qquad (47)$$

where $\boldsymbol{R}_{\mathrm{sym}}$ was defined in (18). We will also assume that the actual system model is exactly matched to the reconstruction model with measurements related to the object through the system matrix, $\boldsymbol{A}$. Since the estimator is a linear function of the measurements, the local impulse responses can be expressed as [3]

$$\underline{l}^j = \boldsymbol{P}\underline{e}^j = [\boldsymbol{F} + \boldsymbol{R}_{\mathrm{sym}}]^{-1} \boldsymbol{F} \underline{e}^j, \qquad (48)$$

where $\underline{l}^j$ represents the response centered at the jth pixel, and $\boldsymbol{F} = \boldsymbol{A}'\boldsymbol{K}^{-1}\boldsymbol{A}$ is the Fisher information matrix. The matrix, $\boldsymbol{P}$, represents the collection of local impulse responses for all image positions. Thus, if the estimator yields space-invariant resolution properties, then $\boldsymbol{P}$ is a Toeplitz matrix. For a given desired response matrix, $\boldsymbol{P}_0$, in theory one can solve (48) for $\boldsymbol{R}$. Specifically, if $\boldsymbol{P}_0$ is invertible then the required regularization matrix is

$$\boldsymbol{R}^{\star}_{\mathrm{sym}} = \boldsymbol{F}[\boldsymbol{P}_0^{-1} - \boldsymbol{I}], \qquad (49)$$

where $\boldsymbol{I}$ is the identity matrix. For the solution in (49) to exist, the right-hand side of the equation must yield a symmetric matrix. Rearranging terms, one finds that if $\boldsymbol{P}_0'\boldsymbol{F} = \boldsymbol{F}\boldsymbol{P}_0$, then a solution exists. In general, one may not be able to exactly match arbitrary responses for specific systems. Moreover, equation (49) may not be a practical design for finding the penalty matrix; however, this expression provides a convenient analytical form relating the desired responses $\boldsymbol{P}_0$ to the penalty matrix.

From (47), it is straightforward to write the covariance matrix for $\hat{\underline{\theta}}$ as

$$[\boldsymbol{F} + \boldsymbol{R}_{\mathrm{sym}}]^{-1} \boldsymbol{F} [\boldsymbol{F} + \boldsymbol{R}_{\mathrm{sym}}]^{-1}. \qquad (50)$$

Thus, for a specific set of desired responses, we may plug (49) into (50) to obtain the covariance matrix for penalized-likelihood reconstruction:

$$\begin{aligned} \mathrm{Cov}\{\hat{\underline{\theta}}_{\mathrm{PL}}\} &= [\boldsymbol{F}\boldsymbol{P}_0^{-1}]^{-1} \boldsymbol{F} [\boldsymbol{F}\boldsymbol{P}_0^{-1}]^{-1} \\ &= \boldsymbol{P}_0 \boldsymbol{F}^{-1} \boldsymbol{P}_0. \end{aligned} \qquad (51)$$

Similarly, because post-smoothing is a linear operation: $\hat{\underline{\theta}}_{\mathrm{PSML}} = \boldsymbol{P}_0 \hat{\underline{\theta}}_{\mathrm{ML}}$, we may find the covariance for post-smoothed ML reconstruction by first finding the covariance for the ML approach by setting $\boldsymbol{R}_{\mathrm{sym}} = \boldsymbol{0}$ in (50). That is,

$$\begin{aligned} \mathrm{Cov}\{\hat{\underline{\theta}}_{\mathrm{PSML}}\} &= \boldsymbol{P}_0 \mathrm{Cov}\{\hat{\underline{\theta}}_{\mathrm{ML}}\} \boldsymbol{P}_0 \\ &= \boldsymbol{P}_0 \boldsymbol{F}^{-1} \boldsymbol{P}_0. \end{aligned} \qquad (52)$$

Thus, (51) and (52) are identical. Therefore, when resolution properties are exactly matched under this system model, the penalized weighted least-squares and post-smoothed weighted least-squares approaches will yield the exact same noise performance. Section V describes similar results under a Poisson model.

It is plausible that under other noise models, where the noise cannot be well approximated by a Gaussian model, or when (49) cannot be solved, the noise performance will be significantly difference in the post-smoothed ML and PL cases.

References

[1] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Tr. Med. Im.*, vol. 19, pp. 601–15, June 2000.

[2] B. Karuta and R. Lecomte, "Effect of detector weighting functions on the point spread function of high-resolution PET tomographs," *IEEE Tr. Med. Im.*, vol. 11, pp. 379–85, Sept. 1992.

[3] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Tr. Im. Proc.*, vol. 5, pp. 1346–58, Sept. 1996.

[4] L. A. Kunyansky, "A new SPECT reconstruction algorithm based on the Novikov explicit inversion formula," *Inverse Prob.*, vol. 17, pp. 293–306, Apr. 2001.

[5] C. E. Metz and X. Pan, "A unified analysis of exact methods of inverting the 2-D exponential Radon transform with implications for noise control in SPECT," *IEEE Tr. Med. Im.*, vol. 14, pp. 643–58, Dec. 1995.

[6] L. T. Chang, "A method for attenuation correction in radionuclide computed tomography," *IEEE Tr. Nuc. Sci.*, vol. 25, pp. 638–643, Feb. 1978.

[7] W. Xia, R. M. Lewitt, and P. R. Edholm, "Fourier correction for spatially variant collimator blurring in SPECT," *IEEE Tr. Med. Im.*, vol. 14, pp. 100–15, Mar. 1995.

[8] C. R. Appledorn, "An analytical solution to the nonstationary reconstruction problem in single photon emission computed tomography," in *Info. Proc. in Med. Im.* (D. A. Ortendahl and J. Llacer, eds.), pp. 69–79, NY: Liss, 1991.

[9] E. J. Soares, C. L. Byrne, S. J. Glick, C. R. Appledorn, and M. A. King, "Implementation and evaluation of an analytical solution to the photon attenuation and nonstationary resolution reconstruction problem in SPECT," *IEEE Tr. Nuc. Sci.*, vol. 40, pp. 1231–1237, Aug. 1993.

[10] P. R. G. Virador, W. W. Moses, and R. H. Huesman, "Reconstruction in PET cameras with irregular sampling and depth of interaction capability," *IEEE Tr. Nuc. Sci.*, pp. 1225–1230, June 1998.

[11] J. S. Liow and S. C. Strother, "The convergence of object dependent resolution in maximum likelihood based tomographic image reconstruction," *Phys. Med. Biol.*, vol. 38, pp. 55–70, Jan. 1993.

[12] R. Yao, J. Siedel, C. A. Johnson, M. E. Daube-Witherspoon, M. V. Green, and R. E. Carson, "Performance characteristics of the 3-D OSEM algorithm in the reconstruction of small animal PET images," *IEEE Tr. Med. Im.*, vol. 19, pp. 798–804, Aug. 2000.

[13] D. L. Snyder, M. I. Miller, L. J. Thomas, and D. G. Politte, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography," *IEEE Tr. Med. Im.*, vol. 6, pp. 228–38, Sept. 1987.

[14] J. W. Stayman and J. A. Fessler, "Penalty design for uniform spatial resolution in 3D penalized-likelihood image reconstruction," in *Proc. of the 1999 Intl. Mtg. on Fully 3D Im. Recon. in Rad. Nuc. Med.*, 1999.

[15] J. Qi and R. M. Leahy, "Resolution and noise properties MAP reconstruction for fully 3D PET," *IEEE Tr. Med. Im.*, vol. 19, pp. 493–506, May 2000.

[16] J. W. Stayman and J. A. Fessler, "Nonnegative definite quadratic penalty design for penalized-likelihood reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2001.

[17] C. L. Lawson and R. J. Hanson, *Solving least squares problems.* New Jersey: Prentice-Hall, 1974.

[18] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Tr. Med. Im.*, vol. 14, pp. 132–137, Mar. 1995.

[19] J. A. Fessler, "Resolution properties of regularized image reconstruction methods," Tech. Rep. 297, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Aug. 1995.

[20] "MicroPET scanner." Crump Institute for Biological Imaging, UCLA. [Online]. Available: `http://www.crump.ucla.edu/user-files/resprojects/microPET/`.

[21] "The Whole Frog Project." Lawrence Berkeley National Laboratory, Office of Science, U.S. Department of Energy. [Online]. Available: `http://www.itg.lbl.gov/Frog`.

[22] J. A. Fessler and A. O. Hero, "Comparing estimator covariances at matched spatial resolutions for imaging system design," in *Proc. 1999 IEEE Information Theory Workshop on Detection, Estimation, Classification and Imaging (DECI)*, p. 15, 1999.