RESOLUTION PROPERTIES OF REGULARIZED IMAGE RECONSTRUCTION METHODS

Jeffrey A. Fessler and W. Leslie Rogers

August 1995. Revised December 1996 and 2012-03-19.

Technical Report No. 297

Resolution Properties of
Regularized Image Reconstruction Methods

Jeffrey A. Fessler and W. Leslie Rogers

4240 EECS Bldg., 1301 Beal Ave., University of Michigan, Ann Arbor, MI 48109-2122

email: fessler@umich.edu, phone: 734-763-1434,

Technical Report # 297

Communications and Signal Processing Laboratory

Dept. of Electrical Engineering and Computer Science

The University of Michigan

March 19, 2012

Abstract

This paper examines the spatial resolution properties of penalized-likelihood image reconstruction methods by analyzing the *local impulse response*. The analysis shows that standard regularization penalties induce *space-variant* local impulse response functions, even for space-invariant tomographic systems. Paradoxically, for emission image reconstruction the *local resolution is generally poorest in high-count regions*. We show that the linearized local impulse response induced by quadratic roughness penalties depends on the object only through its projections. This analysis leads naturally to a modified regularization penalty that yields reconstructed images with nearly uniform resolution. The modified penalty also provides a very practical method for *choosing the regularization parameter to obtain a specified resolution* in images reconstructed by penalized-likelihood methods.

Keywords: emission tomography, local impulse response, iterative reconstruction

1 Introduction

Statistical methods for image reconstruction can provide improved spatial resolution and noise properties over conventional filtered backprojection (FBP) methods. However, iterative methods based solely on maximum-likelihood criteria produce images that become unacceptably noisy as the iterations proceed. Methods for reducing this noise include: stopping the iteration before the images become too noisy (long before convergence) [1], iterating until convergence and then post-smoothing the image [2], using smooth basis functions [3] [4], and replacing the maximum-likelihood criterion with a penalized-likelihood (or maximum *a posteriori*) objective function that includes a roughness penalty to encourage image smoothness [5].

Penalized-likelihood approaches for reducing noise have two important advantages over alternatives such as stopping rules and sieves. First, the penalty function improves the conditioning of the problem, so certain iterative algorithms converge very quickly. Second, one can choose penalty functions that control desired properties of the reconstructed images, such as preserving edges [5] or incorporating anatomical side information [6,7]. In contrast, the smoothness that one obtains through stopping rules is limited by the characteristics of the iterative algorithm. A possible disadvantage of penalized-likelihood methods has been the absence of an intuitive method for choosing the value of the regularization parameter, even for simple quadratic penalties. One contribution of this paper is a new object-independent method for specifying the regularization parameter in terms of the desired resolution of the reconstructed image.

This paper describes another possibly undesirable property of penalized-likelihood image reconstruction methods that has not been previously documented (except in [8] to our knowledge), and then proposes a solution to the problem. Through analysis and empirical results we demonstrate that when one uses standard space-invariant roughness penalties, the reconstructed images have *object-dependent nonuniform spatial resolution, even for space-invariant tomographic systems*. For emission imaging the resolution is generally poorest in high-count regions, which is opposite to what one might expect or prefer. In Section 5 we propose a new modified space-variant roughness penalty that yields images with nearly uniform resolution. Based on our analysis, one could extend the method to provide other resolution characteristics, such as "higher resolution in high count regions" etc., in a manner similar to methods for space-varying regularization [9, 10], but in this paper we focus on the goal of providing uniform resolution.

This paper is somewhat in the spirit of previous studies that used the *local impulse response* [11–15] or an effective local Gaussian resolution [16] to quantify the resolution properties of the unregularized maximum-likelihood expectation-

maximization (ML-EM) algorithm for emission tomography. However, there is an important difference in our approach: since the ML-EM algorithm is rarely iterated until convergence, previous studies examined the spatial resolution properties of ML-EM *as a function of iteration*. In contrast, since there are now fast and *globally convergent* algorithms for maximizing both penalized-likelihood [17–20] and penalized weighted least squares [21–23] objective functions, rather than studying the properties of the *algorithms* as a function of iteration, we study directly the properties of the *estimator* as specified by the *objective function* (Sections II and III). This simplifies the practical use and interpretation of our analysis since the specifics of the iterative algorithm are unimportant (provided one uses a globally convergent method). Our main results (14) and (16) should therefore be applicable to a broad range of inverse problems. (Although we focus on image reconstruction, most of the issues also pertain to quantum-limited image restoration.)

In conventional FBP image reconstruction, one controls the tradeoff between resolution and noise by adjusting the cutoff frequency f_c of a filter. Because f_c has units of inverse length, there is an intuitive (and object-independent) relationship between f_c and the spatial resolution of the reconstructed image. For idealized tomographs, one can use the Hankel transform to compute the point spread function (PSF) as a function of f_c [24]. todo: BETTER HANKEL for FBP REFERENCE? But for real systems, one usually determines the (monotonic) relationship between f_c and the full-width half-maximum (FWHM) of the PSF through the following empirical approach. First, acquire a sinogram using a point or line source, possibly at several locations within the scanner. Then pick a filter type (*e.g.*, Hanning) and reconstruct images for several different values of f_c. Finally, compute the FWHM of the PSF for each case, and record a table of (f_c, FWHM) value pairs. In subsequent studies, one typically chooses the desired resolution (FWHM) by experience or by visually observing the resolution-noise tradeoff, and then obtains the appropriate f_c from the table. One needs to perform this tabulation only once for a given scanner, since FBP is linear (and hence its resolution properties are object-independent).

In contrast, in penalized-likelihood image reconstruction, a regularization parameter β controls the tradeoff between resolution and noise, but the units of β are at best opaquely related to spatial resolution. Therefore it is not obvious how to specify the regularization parameter. As a further complication, one finds that for a fixed β, the reconstructed spatial resolution varies between subjects, and even within the same subject (Section IV). One could choose β using statistical criteria such as minimum mean-squared error [25, 26]. However, mean-squared error is composed equally of both bias (resolution) and variance (noise), whereas those two contributions usually have unequal importance in medical imaging, particularly when images are to be interpreted visually. Furthermore, data-driven methods for choosing β can be unstable in imaging problems [27]. Many other alternatives have been proposed, e.g. [28, 29], most of which have again been assessed with respect to mean-squared error. One practical contribution of this paper is that we develop a method for normalizing the penalty function such that the object-dependent component of β is nearly eliminated. This allows one to build an object-independent table relating β to spatial resolution (FWHM) for a given tomographic system, so that one can select β to achieve a consistent specified resolution within planes, between planes, and even between subjects. The task of choosing the "optimal" resolution is left to the user, just as the "optimal" cutoff frequency (and filter) for FBP are determined by different criteria in different contexts.

Nonuniform resolution properties are not unique to penalized-likelihood methods. The ML-EM algorithm for emission tomography also exhibits resolution variation and asymmetry [12] [30]. An advantage of the penalized-likelihood approach is that one can modify the penalty to overcome the resolution nonuniformity (Sections 5, 6, and 7), whereas it is not obvious how to modify ML-EM to achieve uniform resolution.

PET and SPECT systems usually have intrinsically nonuniform spatial resolution [31] (although PET systems are usually nearly space invariant near the center of the scanner [31]). In this paper our simulations focus on an idealized PET system that is essentially space invariant, except perhaps for the effects of discretizing the Radon transform. Thus, the resolution nonuniformities we report are due solely to the interaction between the log-likelihood and the penalty terms of the objective function, and not due to the system response. We hope to study the effects of penalty functions in systems with intrinsically space-variant resolution in future work.

In Section 9 we also analyze a continuous idealization of penalized least-squares image reconstruction. Some readers may prefer to skim that section first.

2 Local Impulse Response

Let $Y = [Y_1, \ldots, Y_N]'$ denote a random measurement vector (*e.g.*, a noisy sinogram) with density function $f(y; \theta)$, where $\theta = [\theta_1, \ldots, \theta_p]'$ is an unknown parameter in a p-dimensional parameter space Θ, and $'$ denotes vector transpose. In imaging problems, θ typically denotes image pixel values in lexicographic ordering and $\Theta = \{\theta : \theta_j \geq 0, \ j = 1, \ldots, p\}$. Given a particular realization $Y = y$, an estimator of the form $\hat{\theta} = \hat{\theta}(y)$ has mean:

$$\mu(\theta) = E_\theta[\hat{\theta}(Y)] = \int \hat{\theta}(y) f(y; \theta) \, dy. \qquad (1)$$

For linear and space-invariant problems, one can characterize the properties of μ either in the spatial domain by specifying the (global) impulse response, or in the spectral domain by specifying the frequency response (Fourier transform of the impulse response), as in Section 9.

Spectral methods are generally inapplicable to nonlinear estimators for which the impulse response is space variant. For nonlinear estimators one can analyze the *local impulse response* (cf [12]). For an estimator with mean $\mu(\theta)$, we define the local

impulse response of the jth parameter (pixel) to be:[2]

$$
\begin{aligned}
l^j(\theta) &= \lim_{\delta \to 0} \frac{\mu(\theta + \delta e^j) - \mu(\theta)}{\delta} \\
&= \frac{\partial}{\partial \theta_j} \mu(\theta), \; j = 1, \ldots, p,
\end{aligned} \tag{2}
$$

where e^j is the jth unit vector of length p. This impulse response is local in two different senses. First, it is a function of the index j, reflecting the space-variant nature of nonlinear estimation. Second, it depends on the location in the parameter space Θ through the argument θ, reflecting the nonlinear object dependence. The local impulse response also depends on the measurement distribution through (1). Thus, the local impulse response characterizes the object, system, and estimator dependent properties. The local impulse response measures the *change in the mean reconstructed image* due to perturbation of a particular pixel in the noiseless object[3].

To confirm that (2) is a natural generalization of the usual definition of impulse response, consider an estimator whose mean is linear in θ: $\mu(\theta) = \boldsymbol{L}\theta$. Then the conventional definition of impulse response is $\mu(e^j)$, which is the jth column of $\boldsymbol{L}$. Evaluating (2), one finds that l^j is also the jth column of $\boldsymbol{L}$. (If in addition $\boldsymbol{L}$ is a circulant matrix, then the impulse response is space-invariant, and $\boldsymbol{L}$ corresponds to a convolution [33].) Also note that $\mu(\theta) = \theta$ for unbiased estimators, in which case $l^j = e^j$. Penalized-likelihood estimators are always biased, so local impulse responses will typically be bump-like functions, rather than the ideal impulse e^j (e.g. Fig. 1).

As a specific example, consider the penalized weighted least-squares estimator [22]:

$$
\hat{\theta} = \hat{\theta}(y) = \arg\min_{\theta} (y - \boldsymbol{A}\theta)'\boldsymbol{W}(y - \boldsymbol{A}\theta) + \beta\theta'\boldsymbol{R}\theta,
$$

where $\boldsymbol{W}$ and $\boldsymbol{R}$ are symmetric nonnegative definite matrices for which the null spaces of $\boldsymbol{R}$ and $\boldsymbol{W}\boldsymbol{A}$ are disjoint. For a fixed $\boldsymbol{W}$, this estimator is linear in y:

$$
\hat{\theta}(y) = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \beta\boldsymbol{R}]^{-1}\boldsymbol{A}'\boldsymbol{W}y,
$$

and assuming $E_\theta[Y] = \boldsymbol{A}\theta$, one can evaluate (2) to show

$$
l^j = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \beta\boldsymbol{R}]^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}e^j. \tag{3}
$$

For such linear estimators, the local impulse response is independent of θ. As we show in Section 3, the local impulse responses of the nonlinear penalized-likelihood estimators for image reconstruction have approximately the same form as (3), except that $\boldsymbol{W}$ and $\boldsymbol{R}$ may depend on θ.

There are at least three reasons to study the local impulse response. The first reason is simply to understand the resolution

---

[2]We restrict our discussion to estimators where the above limit is well defined. The reader is cautioned that non-convex penalties can lead to estimates that are *discontinuous* functions of the data [32]. We focus here on well-behaved convex penalties.

[3]Because of this interpretation, we use the term *point spread function* (PSF) synonymously with local impulse response, even though this stretches the usual meaning of PSF.

properties of penalized-likelihood estimators. The second reason is that the local impulse response allows one to quantify local resolution, which in turn allows one to choose the smoothing parameter β sensibly. The third reason is that comprehension of the resolution properties enables the design of better penalty functions. In particular, we show how to modify the standard regularization penalty to achieve nearly uniform resolution.

2.1 Brute Force Evaluation of Local Impulse Response

Unlike the simple penalized weighted least squares estimator described above, most estimators $\hat{\theta}(y)$ do not have an explicit analytical form. When there is no explicit form for $\hat{\theta}(y)$, there is usually no explicit form for the estimator mean $\mu(\theta)$ either. Thus it would at first appear that to investigate the local impulse response of a nonlinear estimator of interest, one must resort to a numerical approach based on (1) and (2), replacing the expectation in (2) by the sample mean in a computer simulation. The following recipe illustrates this brute-force approach.

- Select an object θ of interest and generate multiple realizations $\{y^{(m)}\}_{m=1}^M$ of noisy measurements according to the density $f(y;\theta)$.

- Apply the estimator of interest to each of the measurement realizations to obtain estimates $\{\hat{\theta}(y^{(m)})\}_{m=1}^M$.

- Estimate the estimator mean using the sample mean:

$$
\widehat{\mu}(\theta) = \frac{1}{M}\sum_{m=1}^M \hat{\theta}(y^{(m)}). \tag{4}
$$

- Choose a pixel j of interest and small value δ, and generate a second set of noisy measurements according to the density $f(y;\theta + \delta e^j)$.

- Apply the estimator to the second set of noisy measurements, and compute the sample mean to obtain an estimate $\widehat{\mu}(\theta + \delta e^j)$.

- Estimate the local impulse response:

$$
l^j(\theta) \approx \frac{\widehat{\mu}(\theta + \delta e^j) - \widehat{\mu}(\theta)}{\delta}. \tag{5}
$$

By taking δ sufficiently small and M sufficiently large, one can obtain arbitrarily accurate estimates of the local impulse response.

2.2 Unbiased Estimator for Local Impulse Response

If one wants to evaluate the local impulse response for pixels $j_1, \ldots, j_L$ of interest, the above procedure requires $(L+1)M$ image reconstructions. The following method [34–36] reduces the computation to only M image reconstructions. Note that from (2),

$$
l^j(\theta) = \frac{\partial}{\partial \theta_j}\mu(\theta) = \frac{\partial}{\partial \theta_j}E_\theta[\hat{\theta}(Y)] = \frac{\partial}{\partial \theta_j}\int \hat{\theta}(y)f(y;\theta)\,dy
$$

$$= E_\theta[\hat{\theta}(Y) \frac{\partial}{\partial \theta_j} \log f(Y; \theta)].$$

Thus one can show [35, 36] that

$$\widehat{l^j(\theta)} = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}(y^{(m)}) - \hat{\mu}(\theta)) \frac{\partial \log f(y^{(m)}; \theta)}{\partial \theta_j} \quad (6)$$

is an unbiased estimator for $l^j(\theta)$, where $\hat{\mu}(\theta)$ was defined in (4). Once one performs the M reconstructions $\{\hat{\theta}(y^{(m)})\}_{m=1}^{M}$, then one can estimate the local impulse response $\widehat{l^j(\theta)}$ for many pixels with little additional effort.

By taking M sufficiently large, one can obtain arbitrarily accurate estimates of the local impulse response. Unfortunately, M may need to be very large for sufficient accuracy. Often we would gladly accept an *approximation* to the local impulse response if we could avoid performing extensive numerical simulations. The remainder of this paper is devoted to approximations suitable for likelihood-based estimators in tomography.

2.3 Linearized Local Impulse Response

In the context of emission tomography, several investigators have observed [14, 15, 37, 38] that the ensemble mean of a likelihood-based estimator is approximately equal to the value that one obtains by applying the estimator to noiseless data:

$$\mu(\theta) = E_\theta[\hat{\theta}(Y)] \approx \hat{\theta}(\bar{Y}(\theta)) \triangleq \check{\theta}. \quad (7)$$

Here

$$\bar{Y}(\theta) = E_\theta[Y] = \int y f(y; \theta) \, dy \quad (8)$$

denotes the mean of the measurement vector, and $\check{\theta}$ denotes the value of the estimator when given noiseless data $\bar{Y}(\theta)$. This approximation is equivalent to assuming that the estimator is *locally linear*. Let $\nabla_y = [\frac{\partial}{\partial y_1} \ldots \frac{\partial}{\partial y_N}]$ and consider the first-order Taylor expansion of $\hat{\theta}(Y)$ about $\bar{Y}(\theta)$:

$$\hat{\theta}(Y) \approx \hat{\theta}(\bar{Y}(\theta)) + \nabla_y \hat{\theta}(\bar{Y}(\theta)) \cdot (Y - \bar{Y}(\theta));$$

taking the expectation of both sides yields (7). The remainder of this paper uses this local linearity approximation.

Substituting (7) into (2) yields the following definition of the *linearized local impulse response*:

$$\begin{aligned} l^j(\theta) &= \lim_{\delta \to 0} \frac{\hat{\theta}(\bar{Y}(\theta + \delta e^j)) - \hat{\theta}(\bar{Y}(\theta))}{\delta} \\ &= \frac{\partial}{\partial \theta_j} \hat{\theta}(\bar{Y}(\theta)). \end{aligned} \quad (9)$$

Since we focus on this form in the remainder of this paper, for brevity we usually omit the adjective "linearized."

The form of (9) leads to a much simpler recipe for numerically evaluating the local impulse response.

- Select an object θ of interest, a pixel j of interest, and a small value δ. Generate two noiseless measurements vectors: $\bar{Y}(\theta)$ and $\bar{Y}(\theta + \delta e^j)$.

- Apply the estimator of interest to each of the two noiseless measurements, obtaining estimates $\hat{\theta}(\bar{Y}(\theta))$ and $\hat{\theta}(\bar{Y}(\theta + \delta e^j))$.

- Estimate the local impulse response:

$$l^j(\theta) \approx \frac{\hat{\theta}(\bar{Y}(\theta + \delta e^j)) - \hat{\theta}(\bar{Y}(\theta))}{\delta}. \quad (10)$$

By taking δ sufficiently small, one can obtain very accurate estimates of the linearized local impulse response. If $\hat{\theta}$ is linear in y, then (10) is exact of course.

To illustrate this method, Fig. 1 shows a profile through several local impulse response functions of FBP and of the emission ML-EM algorithm [39] (stopped at 30 iterations, well before convergence). The object θ was a uniform ellipse of activity within a uniform elliptical attenuator[4]. Despite the fact that the elliptical object has uniform activity, the resolution of the non-linear ML-EM estimator is clearly nonuniform, whereas the FBP resolution is uniform since the smoothing provided by the Hanning window is space-invariant. Using a similar perturbation approach applied to both the noiseless mean of the data $\bar{Y}(\theta)$ and to a single noisy realization Y, Stamos *et al.* [11] reported strongly object-dependent point response functions for the ART and ML-EM algorithms.

Several investigators have used this easily implemented empirical approach to study the properties of maximum-likelihood estimators in emission tomography. However, being empirical, it fails to reveal general estimator properties. An *analytical* expression for the linearized local impulse response would facilitate understanding general properties of image reconstruction methods. The next section derives an analytical expression for the local impulse response of implicitly defined estimators.

3 Implicitly Defined Estimators

Many estimators in tomography are defined implicitly as the maximizer of some objective function:

$$\hat{\theta} = \hat{\theta}(y) = \arg \max_{\theta \in \Theta} \Phi(\theta, y). \quad (11)$$

We assume Φ has a unique global maximum, so that $\hat{\theta}(y)$ is well defined. There is often no analytical form for such estimators; hence the ubiquitous use of iterative algorithms for performing the required maximization. Fortunately, the linearized local impulse response (9) depends *only on the partial derivatives* of the implicitly defined estimator $\hat{\theta}(y)$. As discussed in [38], even though $\hat{\theta}(y)$ itself is unknown, one can determine its partial derivatives using the implicit function theorem and the chain rule. Disregarding the nonnegativity constraint[5], the maximizer

---

[4]Image size 128×64 at 3mm pixels, 128 radial bins, 110 angles, 3mm ray spacing, 6mm strip width, ellipse radii 58,26 pixels, attenuator radii 180,84 mm, with $\mu = 0.0095$/mm.

[5]Although it appears we are assuming that (12) holds for any y, from (9) one sees we really only need (12) to hold near the case $y = \bar{Y}(\theta)$, i.e. the noiseless case. The nonnegativity constraint is often largely inactive for noiseless data, so (12) is a reasonable assumption.

FBP (Hanning Window)
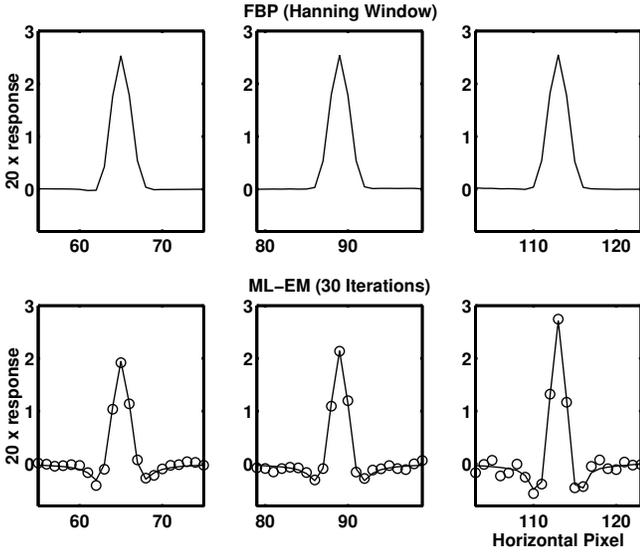
ML–EM (30 Iterations)

Figure 1: Horizontal profiles through the local impulse response functions of FBP with a Hanning window (top) and of the ML-EM algorithm at 30 iterations (bottom), for three pixels located along the horizontal midline of an elliptical object. Solid line: computed using the linearized approximation (10); Circles: computed using the unbiased estimator (6) from $M = 2000$ realizations.

of Φ satisfies:

$$\left.\frac{\partial}{\partial \theta_j}\Phi(\theta, y)\right|_{\theta=\hat{\theta}(y)} = 0, \ j = 1, \ldots, p, \tag{12}$$

for any y. In vector notation:

$$\nabla^{10}\Phi(\hat{\theta}(y), y) = 0 \ \forall y,$$

where $\nabla^{10} = [\frac{\partial}{\partial\theta_1} \ \ldots \ \frac{\partial}{\partial\theta_p}]$ is the row gradient operator (with respect to the first argument of Φ). Now differentiate again with respect to y using the chain rule:

$$\nabla^{20}\Phi(\hat{\theta}(y), y)\nabla_y\hat{\theta}(y) + \nabla^{11}\Phi(\hat{\theta}(y), y) = 0, \tag{13}$$

where the (j,k)th element of ∇^{20} is $\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}$ and the (j,i)th element of ∇^{11} is $\frac{\partial^2}{\partial\theta_j\,\partial y_i}$. For simplicity we drop the dependence of $\bar{Y}$ on θ except where explicitly needed. Assuming that $-\nabla^{20}\Phi(\check{\theta}, \bar{Y})$ is positive definite, substitute $y = \bar{Y}$ into (13) and solve for the partial derivatives of $\hat{\theta}(\bar{Y}(\theta))$:

$$\nabla_y\hat{\theta}(\bar{Y}(\theta)) = [-\nabla^{20}\Phi(\check{\theta}, \bar{Y})]^{-1}\nabla^{11}\Phi(\check{\theta}, \bar{Y}).$$

Combining with the chain rule applied to (9):

$$\begin{aligned} l^j(\theta) &= \frac{\partial}{\partial\theta_j}\hat{\theta}(\bar{Y}(\theta)) = \nabla_y\hat{\theta}(\bar{Y}(\theta))\frac{\partial}{\partial\theta_j}\bar{Y}(\theta) \\ &= [-\nabla^{20}\Phi(\check{\theta}, \bar{Y})]^{-1}\nabla^{11}\Phi(\check{\theta}, \bar{Y})\frac{\partial}{\partial\theta_j}\bar{Y}(\theta). \end{aligned} \tag{14}$$

This equality expresses the local impulse response solely in terms of the partial derivatives of the objective function and the measurement mean, i.e., we have eliminated the dependence on the implicitly defined estimator $\hat{\theta}(y)$.

3.1 Penalized-Likelihood Estimators

In the remainder of this paper, we focus on penalized-likelihood objective functions Φ of the form:

$$\Phi(\theta, y) = L(\theta, y) - \beta R(\theta), \tag{15}$$

where $L(\theta, y) = \log f(y; \theta)$ denotes the log-likelihood, $R(\theta)$ is a roughness penalty function, and β is a nonnegative regularization parameter that controls the influence of the penalty, and hence the tradeoff between resolution and noise.

Define $\boldsymbol{R}(\theta) = \nabla^2 R(\theta)$ to be the Hessian of the penalty, and note that $\nabla^{11} R = 0$. For penalized-likelihood estimators of the form (15) we have from (14) the following expression for the local impulse response[6]:

$$l^j(\theta) = [-\nabla^{20}L(\check{\theta}, \bar{Y}) + \beta\boldsymbol{R}(\check{\theta})]^{-1}\nabla^{11}L(\check{\theta}, \bar{Y})\frac{\partial}{\partial\theta_j}\bar{Y}(\theta). \tag{16}$$

This expression should be useful for investigating estimators in a variety of imaging problems. Next we evaluate expression (16) for Poisson distributed measurements, which will be the focus of the remainder of this paper.

3.2 Poisson Statistics

Both emission and transmission tomographic systems yield independent measurements with Poisson statistics; the primary difference is in the form of their assumed measurement means $\bar{Y}(\theta)$. In both cases the assumed log-likelihood has the form:

$$L(\theta, y) = \sum_i y_i \log \bar{Y}_i(\theta) - \bar{Y}_i(\theta),$$

neglecting constants independent of θ. Thus

$$\begin{aligned} \frac{\partial}{\partial\theta_j}L(\theta, y) &= \sum_i \left(\frac{y_i}{\bar{Y}_i(\theta)} - 1\right)\frac{\partial}{\partial\theta_j}\bar{Y}_i(\theta) \\ \frac{\partial^2}{\partial\theta_j\,\partial y_i}L(\theta, y) &= \frac{1}{\bar{Y}_i(\theta)}\frac{\partial}{\partial\theta_j}\bar{Y}_i(\theta) \\ -\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}L(\theta, y) &= \sum_i \left(\frac{y_i}{\bar{Y}_i^2(\theta)}\right)\frac{\partial}{\partial\theta_j}\bar{Y}_i(\theta)\frac{\partial}{\partial\theta_k}\bar{Y}_i(\theta) \\ &\quad - \sum_i \left(\frac{y_i}{\bar{Y}_i(\theta)} - 1\right)\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\bar{Y}_i(\theta) \end{aligned} \tag{17}\tag{18}$$

3.3 Emission Tomography

For emission tomography [39], θ_j denotes the radioisotope concentration in the jth voxel, and the measurement mean is linear in θ:

$$\bar{Y}_i(\theta) = \sum_{j=1}^p a_{ij}\theta_j + r_i. \tag{19}$$

---

[6]We consider the class of objectives Φ for which the Hessian $-\nabla^{20}L(\check{\theta}, \bar{Y}) + \beta\boldsymbol{R}(\check{\theta})$ is positive definite; i.e., $\Phi(\theta, y)$ is at least *locally* strictly concave near the noiseless case $(\check{\theta}, \bar{Y}(\theta))$.

5

The $\{a_{ij}\}$ are nonnegative constants that characterize the tomographic system, and the $\{r_i\}$ are nonnegative constants that represent the mean contribution of background events (random coincidences, scatter, etc.). Substituting (19) into (17) and (18):

$$-\nabla^{20}L(\theta, y) = \mathbf{A}'D\left[\frac{y_i}{\bar{Y}_i^2(\theta)}\right]\mathbf{A}$$

$$\nabla^{11}L(\theta, y) = \mathbf{A}'D\left[\frac{1}{\bar{Y}_i(\theta)}\right],$$

where $\mathbf{A} = \{a_{ij}\}$ is an $N \times p$ sparse matrix and $D[u_i]$ denotes a $N \times N$ diagonal matrix with diagonal entries $u_1, \ldots, u_N$. Noting that $\frac{\partial}{\partial\theta_j}\bar{Y}(\theta) = \mathbf{A}e^j$ and substituting into (16) yields the local impulse response:

$$l^j(\theta) = [\mathbf{A}'D\left[\frac{\bar{Y}_i(\theta)}{\bar{Y}_i^2(\check\theta)}\right]\mathbf{A} + \beta\mathbf{R}(\check\theta)]^{-1}\mathbf{A}'D\left[\frac{1}{\bar{Y}_i(\check\theta)}\right]\mathbf{A}e^j.$$

For moderate or small values of β, $\check\theta$ is a slightly blurred version of θ (see (7)). Since the projection operation $\mathbf{A}\theta$ is a smoothing operator, the projections $\bar{Y}(\theta)$ and $\bar{Y}(\check\theta)$ are approximately equal. Therefore[7], we simplify the above expression to

$$l^j(\theta) \approx [\mathbf{A}'D\left[u_i^{\text{emis}}(\theta)\right]\mathbf{A} + \beta\mathbf{R}(\check\theta)]^{-1}\mathbf{A}'D\left[u_i^{\text{emis}}(\theta)\right]\mathbf{A}e^j, \tag{20}$$

where

$$u_i^{\text{emis}}(\theta) = \frac{1}{\bar{Y}_i(\theta)} \tag{21}$$

is the reciprocal of the variance of Y_i under the assumed Poisson model. For penalized-likelihood estimators in emission tomography, (20) is our final approximation to the local impulse response.

3.4 Transmission Tomography

For transmission tomography [39], θ_j denotes the linear attenuation coefficient of the jth pixel. The measurement means are nonlinear functions of the attenuation coefficients:

$$\bar{Y}_i(\theta) = b_i \exp\left(-\sum_{j=1}^{p} a_{ij}\theta_j\right) + r_i. \tag{22}$$

The $\{a_{ij}\}$, $\{b_i\}$, and $\{r_i\}$ are nonnegative constants that characterize the system, transmission source strength, and background events respectively. From (17) and (18) one can show [38] that

$$-\nabla^{20}L(\theta, y) = \mathbf{A}'D\left[(\bar{Y}_i(\theta) - r_i)(1 - \frac{r_i y_i}{\bar{Y}_i^2(\theta)})\right]\mathbf{A}$$

$$\nabla^{11}L(\theta, y) = -\mathbf{A}'D\left[1 - \frac{r_i}{\bar{Y}_i(\theta)}\right].$$

Under the same assumptions about the similarity of $\bar{Y}(\check\theta)$ and $\bar{Y}(\theta)$ used in the emission case, one can substitute the above formulae into (16) to derive an approximate local impulse response for transmission tomography:

$$l^j(\theta) \approx [\mathbf{A}'D\left[u_i^{\text{tran}}(\theta)\right]\mathbf{A} + \beta\mathbf{R}(\check\theta)]^{-1}\mathbf{A}'D\left[u_i^{\text{tran}}(\theta)\right]\mathbf{A}e^j, \tag{23}$$

which is of the same form as (20). However, in this case

$$u_i^{\text{tran}}(\theta) = \frac{(\bar{Y}_i(\theta) - r_i)^2}{\bar{Y}_i(\theta)} \tag{24}$$

is approximately the reciprocal of the variance (cf [38]) of $\log(b_i/(Y_i - r_i))$.

To summarize, we have derived a general local impulse response expression (14) for penalized-likelihood estimators, and specific expressions (20) and (23) for emission and transmission tomography.

4 Resolution Properties

The local impulse response approximations for penalized-likelihood image reconstruction in emission tomography (20) and transmission tomography (23) differ only by the definitions of the u_i terms in the diagonal matrix. Thus, the local impulse response has the following generic form:

$$l^j(\theta) \approx [\mathbf{A}'\mathbf{D}_\theta\mathbf{A} + \beta\mathbf{R}(\check\theta)]^{-1}\mathbf{A}'\mathbf{D}_\theta\mathbf{A}e^j, \tag{25}$$

where $\mathbf{D}_\theta = D[u_i(\theta)]$ is an object-dependent diagonal matrix with $u_i(\theta)$ defined by (21) for emission tomography and (24) for transmission tomography.

Many penalty functions used in tomography can be written in the following form[8]:

$$R(\theta) = \sum_{j=1}^{p}\frac{1}{2}\sum_{k\in\mathcal{N}_j}w_{jk}\psi(\theta_j - \theta_k), \tag{26}$$

where $\mathcal{N}_j$ is a neighborhood of pixels near pixel j, ψ is a symmetric convex function, and $w_{jk} = w_{kj}$. For a "first-order" neighborhood one chooses w_{jk} to equal 1 for horizontal and vertical neighboring pixels, and 0 otherwise; for a "second-order" neighborhood one also includes $w_{jk} = 1/\sqrt{2}$ for diagonal neighbors. With either of these standard choices for the w_{jk}'s, we refer to $R(\theta)$ as a *uniform penalty*, since it is shift-invariant; i.e., translating the image yields an identical value of $R(\theta)$.

One of the simplest uniform penalties is the *uniform quadratic penalty*, which refers to the case where $\psi(x) = x^2/2$. In this case the penalty has a quadratic form:

$$R(\theta) = \frac{1}{2}\theta'\mathbf{R}\theta,$$

---

[7]The diagonal terms in (20) and the preceding equation are sandwiched between the backprojection and projection operators $\mathbf{A}'$ and $\mathbf{A}$, which smooth out most differences between $\bar{Y}(\theta)$ and $\bar{Y}(\check\theta)$. In a sense, the heavy-tailed $1/r$ kernel that makes tomography ill-posed works to our advantage when making the above approximations.

[8]If $\ddot\psi(x) > 0$ for all x, then it is easily shown that the only vectors in the null space of the matrix $\nabla^2 R(\theta)$ are of the form $v = 1_p v_1$, where 1_p is the length-p vector of ones. For any tomographic system that satisfies $\mathbf{D}_\theta\mathbf{A}1_p \neq 0$ (i.e. the projection of a uniform image is nonzero), we can then conclude that $\mathbf{A}'\mathbf{D}_\theta\mathbf{A} + \beta\mathbf{R}(\check\theta)$ is positive definite and therefore invertible, as required by (16).

where $\boldsymbol{R}$ is a θ-independent $p \times p$ matrix defined by:

$$\boldsymbol{R}_{jk} = \begin{cases} \sum_{l \in \mathcal{N}_j} w_{jl}, & k = j \\ -w_{jk}, & k \neq j \end{cases}.$$

In the quadratic case the local impulse response simplifies to:

$$l^j(\theta) \approx [\boldsymbol{A}' \boldsymbol{D}_\theta \boldsymbol{A} + \beta \boldsymbol{R}]^{-1} \boldsymbol{A}' \boldsymbol{D}_\theta \boldsymbol{A} e^j. \tag{27}$$

4.1 Projection Dependence

When $R(\theta)$ is a quadratic form so that $\boldsymbol{R}$ is independent of θ, then remarkably the local impulse response approximation $l^j(\theta)$ given by (27) *depends on the object θ only through its projections* $\bar{Y}(\theta)$ (see (21) and (24)). Even if the object is unknown, its projections are approximately known through the noisy measurements y. Thus, even for real noisy measurements, we can predict the local impulse response simply by replacing $\bar{Y}(\theta)$ with y in (20) or (23). This simple approach is effective primarily because the diagonal terms in (20) and (23) are sandwiched between the backprojection and projection operators $\boldsymbol{A}'$ and $\boldsymbol{A}$, which greatly smooth out the noise in y, i.e.

$$\boldsymbol{A}' D [\bar{Y}_i(\theta)]^{-1} \boldsymbol{A} \approx \boldsymbol{A}' D[y_i]^{-1} \boldsymbol{A}. \tag{28}$$

4.2 Nonuniformity

One might expect that a uniform penalty such as (26) would induce uniform spatial resolution, just as space-invariant sieves do [2]. Using the preconditioned conjugate gradient [23, 40] or Gauss-Siedel [21, 22] algorithms, one can evaluate (25) or (27) and then display the local impulse response for several locations within the object. Upon doing this, one immediately finds that the local impulse response is very *nonuniform*, even for standard uniform quadratic penalties. (See Section 6.)

The next section elaborates on this property, but one can partially understand the source of the resolution nonuniformity by considering (27). If the measurement noise was homoscedastic with variance ν, then $\boldsymbol{D}$ would be simply a scaled identity matrix: $\boldsymbol{D} = \nu^{-1} \boldsymbol{I}$, and from (27) the local impulse response would be

$$\begin{aligned} l^j(\theta) &= [\nu^{-1} \boldsymbol{A}' \boldsymbol{A} + \beta \boldsymbol{R}]^{-1} \nu^{-1} \boldsymbol{A}' \boldsymbol{A} e^j \\ &= [\boldsymbol{A}' \boldsymbol{A} + \nu \beta \boldsymbol{R}]^{-1} \boldsymbol{A}' \boldsymbol{A} e^j. \end{aligned} \tag{29}$$

In other words, noise with variance ν leads to an impulse response that corresponds to an "effective" smoothing parameter $\nu\beta$. Thus, the influence of the smoothing penalty is *not invariant to changes in the noise variance*, which perhaps explains in part why choosing β is considered by many investigators to be a difficult process. The Poisson case is more complicated since the values of $\boldsymbol{D}_\theta$ vary along the diagonal. Since a given pixel is primarily affected by the detectors whose rays intersect it, each pixel sees a different "effective variance" and hence a different effective smoothing parameter.

This resolution nonuniformity can also be explained from a Bayesian perspective. The Fisher information $\boldsymbol{A}' \boldsymbol{D}_\theta \boldsymbol{A}$ is a measure of the *certainty* in the data. For pixels where this data certainty is smaller (due to higher noise variance in the rays that

intersect that pixel), the posterior estimate will give more weight to the prior, which (being a smoothness prior) will cause more smoothing. In emission tomography, pixels with higher activity yield rays with higher counts and hence higher *absolute* variance or *lower* certainty. Paradoxically, penalized-likelihood methods using the standard uniform penalty thus have *lower spatial resolution in high-count regions*. This property is certainly undesirable, and may explain in part why many authors have characterized the uniform quadratic penalty as causing "oversmoothing," since the most prominent image features are generally smoothed the most!

4.3 Choosing β for one pixel

Since (27) allows one to predict the local impulse response (and hence the spatial resolution) at any pixel j as a function of β, one could use (27) to choose a value for β that induces a desired resolution at some pixel j of interest in the image. However, the induced resolutions at other points in the image would still be different, which motivates the modified penalty developed in the next section.

5 Resolution Uniformity

This section analyzes the problem of resolution nonuniformity more closely. This analysis leads to a natural modified penalty function that induces more uniform resolution. For simplicity we focus on emission tomography; parallel arguments apply to transmission tomography.

5.1 Emission Tomography

In emission tomography, the Fisher information matrix $\boldsymbol{A}' \boldsymbol{D}_\theta \boldsymbol{A}$ is an operator that, due to the lexicographic ordering of pixels, one can treat as a mapping from image space to image space. The operator $\boldsymbol{A}' \boldsymbol{D}_\theta \boldsymbol{A}$ is shift-variant for emission tomography, which is the crux of the problem of resolution nonuniformity. The previous section noted that the nonuniform diagonal of the $\boldsymbol{D}_\theta$ term is partially responsible for the nonuniform local impulse response. But even without that term, the spatial resolution would still be nonuniform because typically even $\boldsymbol{A}' \boldsymbol{A}$ is a shift-variant operator in PET and SPECT. However, one often models the system matrix $\boldsymbol{A}$ as a product of three factors: $a_{ij} = c_i g_{ij} s_j$, such that $\boldsymbol{G}' \boldsymbol{G}$ is approximately shift-invariant, where $\boldsymbol{G} = \{g_{ij}\}$ represents the object-independent[9] geometric portion of the tomographic system response. The c_i's represent ray-dependent factors that change between studies, including detector efficiency factors, dead time, radioisotope decay, and (in PET) attenuation factors. The s_j's represent pixel-dependent factors such as spatial variations in sensitivity, and (in SPECT) "first-order" attenuation correction factors (cf the image-space Chang method [41] for SPECT attenuation correction). For our PET work thus far, we have simply used $s_j = 1$. In matrix

---

[9]In SPECT $\boldsymbol{G}$ will only be approximately object-independent due to attenuation.

notation:

$$\boldsymbol{A} = D[c_i]\,\boldsymbol{G}D[s_j]\,. \tag{30}$$

This factorization is not unique. If one desires resolution uniformity, then the analysis that follows suggests that one should strive to choose $\{c_i\}$ and $\{s_j\}$ so that $\boldsymbol{G}'\boldsymbol{G}$ is "as shift-invariant as possible" (cf (43) below). (See [42] for additional analyses of shift-invariant and shift-variant imaging systems.)

Substituting (30) into (20) and simplifying:

$$l^j(\theta) \approx [D[s_j]\,\boldsymbol{G}'D[q_i(\theta)]\,\boldsymbol{G}D[s_j] + \beta\boldsymbol{R}(\check{\theta})]^{-1}$$
$$\cdot\, D[s_j]\,\boldsymbol{G}'D[q_i(\theta)]\,\boldsymbol{G}D[s_j]\,e^j, \tag{31}$$

where

$$q_i(\theta) = c_i^2/\bar{Y}_i(\theta). \tag{32}$$

In PET, these q_i's are very nonuniform due to attenuation correction factors that range from 1 to 100, detector efficiencies that vary over an order of magnitude in block crystal systems, and the intrinsic count variations of Poisson measurements.
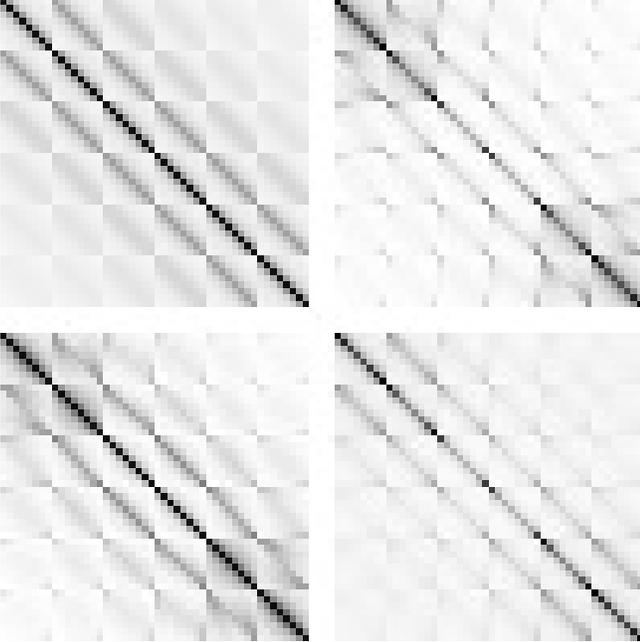
Figure 2: Illustration of the approximation (36). Upper left: the matrix $\boldsymbol{G}'\boldsymbol{G}$ which is approximately Toeplitz-block-Toeplitz. Upper right: the Fisher information $\boldsymbol{F} = \boldsymbol{G}'D[q_i(\theta)]\,\boldsymbol{G}$ including Poisson noise covariance. The nonuniform diagonal is caused by the nonuniform Poisson noise variance. Lower right: the approximation $\Lambda\boldsymbol{G}'\boldsymbol{G}\Lambda$; note the agreement with the upper right matrix, i.e. $\boldsymbol{F} \approx \Lambda\boldsymbol{G}'\boldsymbol{G}\Lambda$. Lower left: $\Lambda^{-1}\boldsymbol{G}'D[q_i(\theta)]\,\boldsymbol{G}\Lambda^{-1}$; note that this matrix is a reasonable approximation to $\boldsymbol{G}'\boldsymbol{G}$.

The Fisher information matrix for estimating θ is:

$$\boldsymbol{F}(\theta) = \boldsymbol{A}'D[u_i(\theta)]\,\boldsymbol{A} = D[s_j]\,\boldsymbol{G}'D[q_i(\theta)]\,\boldsymbol{G}D[s_j]\,. \tag{33}$$

As a consequence of the nonuniformity of the q_i's, the diagonal of $\boldsymbol{F}(\theta)$ is also nonuniform, which contributes greatly to the shift-variance of the $\boldsymbol{F}(\theta)$ operator in PET.

Understanding the structure of $\boldsymbol{F}(\theta)$ is the key to correcting the resolution nonuniformity. From (33) the diagonal elements of $\boldsymbol{F}(\theta)$ can be written:

$$\boldsymbol{F}_{jj}(\theta) = s_j^2\sum_i g_{ij}^2 q_i(\theta) = \kappa_j^2(\theta)\sum_i g_{ij}^2,\ j = 1,\dots,p, \tag{34}$$

where we define

$$\kappa_j(\theta) = s_j\sqrt{\sum_i g_{ij}^2 q_i(\theta)/\sum_i g_{ij}^2}. \tag{35}$$

Due to the $1/r$ response of tomographs, $\boldsymbol{F}(\theta)$ is fairly concentrated about its diagonal, so (34) suggests the approximation:

$$\boldsymbol{F}(\theta) = D[s_j]\,\boldsymbol{G}'D[q_i(\theta)]\,\boldsymbol{G}D[s_j] \approx \Lambda_\theta\boldsymbol{G}'\boldsymbol{G}\Lambda_\theta, \tag{36}$$

where

$$\Lambda_\theta = D[\kappa_j(\theta)] \tag{37}$$

is a $p \times p$ diagonal matrix. From (34) one sees that approximation (36) is exact along the diagonal of $\boldsymbol{F}(\theta)$, and would be exact on the off-diagonal elements if the q_i's were equal. The approximation (36) turns out to be reasonably accurate even for very nonuniform q_i's because the κ_j's vary slowly as a function of j, due to the smoothing implicit in (35). This approximation also reflects the fact that the local impulse response of pixel j depends primarily on the q_i's that correspond to rays that intersect pixel j.

To visualize (36), Fig. 2 shows the various matrices for a toy PET problem[10] (with $s_j = 1$). The nearly Toeplitz-block-Toeplitz structure of $\boldsymbol{G}'\boldsymbol{G}$ is apparent.

Substituting (36) into (31) and rearranging yields the following approximation to the local impulse response:

$$\begin{aligned} l^j(\theta) &\approx [\Lambda_\theta\boldsymbol{G}'\boldsymbol{G}\Lambda_\theta + \beta\boldsymbol{R}(\check{\theta})]^{-1}\Lambda_\theta\boldsymbol{G}'\boldsymbol{G}\Lambda_\theta e^j \\ &= \Lambda_\theta^{-1}[\boldsymbol{G}'\boldsymbol{G} + \beta\Lambda_\theta^{-1}\boldsymbol{R}(\check{\theta})\Lambda_\theta^{-1}]^{-1}\boldsymbol{G}'\boldsymbol{G}\Lambda_\theta e^j \\ &= \kappa_j(\theta)\Lambda_\theta^{-1}[\boldsymbol{G}'\boldsymbol{G} + \beta\Lambda_\theta^{-1}\boldsymbol{R}(\check{\theta})\Lambda_\theta^{-1}]^{-1}\boldsymbol{G}'\boldsymbol{G}e^j, \end{aligned} \tag{38}$$

because $\Lambda_\theta e^j = \kappa_j(\theta)e^j$.

What does Λ_θ represent statistically? From (35), we see that $\kappa_j(\theta)$ is a normalized backprojection of $\{q_i\}$, where q_i is the inverse of the variance of the ith corrected measurement y_i/c_i. Thus $\kappa_j(\theta)$ represents an aggregate *certainty* of the measurement rays that intersect the jth pixel. Since the local impulse response l^j is typically concentrated about pixel j, a somewhat cruder but nevertheless very useful approximation that follows from (38) is

$$l^j(\theta) \approx [\boldsymbol{G}'\boldsymbol{G} + \beta/\kappa_j^2(\theta)\,\boldsymbol{R}(\check{\theta})]^{-1}\boldsymbol{G}'\boldsymbol{G}e^j, \tag{39}$$

(cf (29)). The accuracy of this approximation improves as β decreases (and hence l^j approaches the impulse e^j). This expression again illustrates the property that the effective amount of smoothing $\beta/\kappa_j^2(\theta)$ increases with decreasing measurement certainty $\kappa_j(\theta)$.

---

[10]The object was a 6×2 uniform rectangle in a 8×6 image. We used $c_i = 1$, so the only nonuniformity in the q_i's was due to the $1/\bar{Y}_i(\theta)$ contribution of Poisson noise.

Approximation (39) illuminates the paradoxical oversmoothing of high-count regions with the uniform penalty. If pixel j is transected by rays with high counts, then from (32) and (35) we see that q_i and hence $\kappa_j(\theta)$ will be small, so the effective smoothing parameter $\beta/\kappa_j(\theta)^2$ above will be large, causing lower resolution. As θ_j increases, the rays that intersect it will also increase, so the local resolution decreases[11].

5.2 A Modified Penalty

The form of (38) suggests several possible methods for modifying the penalty function to improve resolution uniformity. We focus on one approach that is easily implemented. Let $R^\star(\theta)$ denote a "target" penalty function of the form (26) (presumably shift-invariant) whose properties would be suitable if we had $\boldsymbol{D}_\theta = \boldsymbol{I}$. Suppose we have estimates $\{\hat{\kappa}_j\}$ of $\{\kappa_j(\theta)\}$, and consider the *modified penalty*:

$$R(\theta) = \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \hat{\kappa}_j \hat{\kappa}_k \psi(\theta_j - \theta_k). \qquad (40)$$

If $\boldsymbol{R}(\theta) = \nabla^2 R(\theta)$ denotes the Hessian of this modified penalty[12], then one can show that

$$\boldsymbol{R}_{jk}(\theta) = \begin{cases} \sum_{l \neq j} w_{jl} \hat{\kappa}_j \hat{\kappa}_l \ddot{\psi}(\theta_j - \theta_k), & j = k \\ -w_{jk} \hat{\kappa}_j \hat{\kappa}_k \ddot{\psi}(\theta_j - \theta_k), & j \neq k \end{cases},$$

so that if $D[\hat{\kappa}_j] \approx \Lambda_\theta$ and we let $\boldsymbol{R}^\star(\theta) = \nabla^2 R^\star(\theta)$, then

$$\boldsymbol{R}(\theta) \approx \Lambda_\theta \boldsymbol{R}^\star(\theta) \Lambda_\theta. \qquad (41)$$

This approximation relies on the fact that neighboring pixels have very similar certainties, i.e. $\kappa_k(\theta) \approx \kappa_j(\theta)$ for $k \in \mathcal{N}_j$, which again follows from the smoothing effect of (35). Substituting (41) into the expression for the local impulse response (38) yields the new approximation

$$l^j(\theta) \approx \kappa_j(\theta)\Lambda_\theta^{-1}[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star(\check{\theta})]^{-1}\boldsymbol{G}'\boldsymbol{G}e^j. \qquad (42)$$

If the geometric response $\boldsymbol{G}$ is nearly space invariant, then to within our approximation accuracy, (42) corresponds to nearly uniform resolution except for the following features.

- Unlike the uniform quadratic target penalty, for which $\boldsymbol{R}^\star$ is constant along its diagonal, nonquadratic penalties lead to object-dependent Hessians $\boldsymbol{R}^\star(\check{\theta})$. However, users of nonquadratic penalties presumably desire certain nonuniformities, i.e. more smoothing in flat regions and less smoothing near edges. Our modified penalty (40) preserves this important characteristic of nonquadratic penalties. Our modification only corrects for the resolution nonuniformities that are induced by the interaction between the nonuniform statistics and the penalty function. Essentially we are correcting for the $\Lambda_\theta^{-1}\boldsymbol{R}\Lambda_\theta^{-1}$ term in (38).

- Since $\kappa_j(\theta)/\kappa_k(\theta) \approx 1$ for $k \in \mathcal{N}_j$, the term $\kappa_j(\theta)\Lambda_\theta^{-1}$ in (42) effectively acts as an identity matrix for pixels near j, so for local impulse responses that are fairly narrow we can disregard the $\kappa_j(\theta)\Lambda_\theta^{-1}$ term, leading to the approximation

$$l^j(\theta) \approx [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star(\check{\theta})]^{-1}\boldsymbol{G}'\boldsymbol{G}e^j. \qquad (43)$$

By "narrow" we mean relative to the scale of the spatial fluctuations in $\kappa_j(\theta)$. However, in regions where the certainty $\kappa_j(\theta)$ is more rapidly varying as a function of spatial position (such as near the edge of an object), the presence of the term $\kappa_j(\theta)\Lambda_\theta^{-1}$ indicates that there will be some asymmetry in the local impulse response. As illustrated in Section 6, such asymmetry can occur with or without our modifications to the penalty. Further work is needed to correct these asymmetries.

5.3 Practical Implementation

In practice, the term $\kappa_j(\theta)$ is unknown, since it depends on the noiseless measurement mean $\bar{Y}(\theta)$. Fortunately, we can manipulate the noisy data to provide a reasonable estimate $\hat{\kappa}_j$ of $\kappa_j(\theta)$.

We first compute from the measurements an estimate $\hat{q}_i$ of the term $q_i(\theta)$ defined by (32):

$$\hat{q}_i = \frac{c_i^2}{\max\{y_i, 10\}}. \qquad (44)$$

The "10" factor ensures that the denominator is not too close to zero, and hopefully provides a little robustness to model mismatch by giving no rays an inordinate weighting. We then replace the $q_i(\theta)$ term in (35) with $\hat{q}_i$ to precompute $\hat{\kappa}_j$, which we then use in (40). Thus, implementing the modified penalty (40) simply requires one extra backprojection. (To save a little computation, one could probably replace (35) with an approximate backprojector.) The cost of multiplying by $\hat{\kappa}_j\hat{\kappa}_k$ in (40) is negligible compared to the forward projections required by iterative reconstruction algorithms.

Since the $\hat{\kappa}_j$'s depend on the data, our modified penalty (40) is data-dependent! Bayesian-minded readers may find the idea of a data-dependent "prior" to be somewhat disconcerting. We make absolutely no pretense that this approach has any Bayesian interpretation. The purpose of the penalty is solely to control noise, and the purpose of our modification to the penalty is solely to control the resolution properties. As an alternative to (44), one could periodically update the $\hat{\kappa}_j$'s by substituting one's current estimate of $\hat{\theta}$ into (35) within an iterative algorithm. But the extra effort is unlikely to change the final estimate very much, since, as noted earlier, small changes in the q_i's have minor effects on the estimate due to the "sandwich" effect described in footnote 7 and by (28).

Since (40) and (44) define the modified penalty $R(\theta)$ to be a function that depends on y, the matrix $\nabla^{11}R$ is no longer exactly 0, so strictly speaking the steps between (14) and (16) need modification. However, because of the effective smoothing in the definition (35), the partial derivatives with respect to y of the modified penalty are very small, so we ignore this second-order effect.

---

[11]However, note that even uniform objects (e.g. $\theta = [1 \ \dots \ 1]$) lead to nonuniform resolution (i.e. to shift-variant local impulse response), since $\bar{Y}(\theta)$ will be a nonuniform vector due to the different lengths of the line integrals through the object.

[12]One can easily verify that this Hessian is nonnegative definite if $\ddot{\psi} > 0$.

5.4 Transmission Tomography

For transmission tomography, we use a modified penalty similar to (40). The only difference is that rather than using the $\{q_i\}$ defined in (32), we use the $\{u_i\}$ defined by (24) for computing $\{\hat{\kappa}_j\}$ in (35). Without our modification to the penalty, it follows from (39) and (24) (and from empirical experience) that attenuation maps reconstructed with standard uniform penalties have the ==poorest== resolution in high density regions (through which the fewest photons pass unabsorbed, so the corresponding rays have fewer counts and thus the κ_j values are small). Again this property is undesirable, particularly since mismatch between the spatial resolution of attenuation maps and emission measurements can cause image artifacts in PET [43]. With our modified penalty, the resolution is nearly uniform. Furthermore, if one uses the same geometric model G for both emission and transmission reconstruction[13], then one can use the same parameter β for both cases, encouraging consistency between emission and transmission spatial resolution.

5.5 Choosing β

For a quadratic target penalty $R^\star(\theta)$, the local impulse response (43) induced by our modified penalty (40) is independent of the object θ. Thus the process of choosing the smoothing parameter β is significantly simplified by the following approach. Let j be a pixel in the center of the image, for example. For a given system geometric response G, precompute the local impulse response (43) for a range of values of β. For each β, tabulate some measure of resolution, such as the FWHM of l^j. Then, when presented with a new data set to be reconstructed at some *user-specified resolution*, simply interpolate the table to determine the appropriate value for β. Finally, reconstruct the object using the modified quadratic penalty. Section 6 presents results that demonstrate the effectiveness of this approach. Analytical results in Section 9 further simplify the process of building this table for certain tomographs.

Many (but not all) nonquadratic penalties are locally quadratic near 0, and it is this quadratic portion of the penalty that is active within relatively flat regions in the image. For such penalties, one could use the table approach described above to specify the desired "resolution" in the flat parts of the image, and then adjust any remaining penalty parameters to control the influence of edges etc. For penalties that are not even locally quadratic, such as the generalized Gaussian Markov random field prior [32], further investigation is needed.

6 Examples

This section demonstrates the improved resolution uniformity induced by the modified penalty (40) within a penalized-likelihood image reconstruction method for PET emission measurements. For θ, we used the 128×64 emission image shown in Fig. 3, which has relative emission intensities of 1, 2, and 3

in the cold disk (left), background ellipse, and hot disk (right) respectively. We included the effects of nonuniform attenuation in the c_i's by using an attenuation map qualitatively similar to Fig. 3, but with attenuation coefficients 0.003/mm, 0.0096/mm, and 0.013/mm for the cold disk, background ellipse, and hot disk respectively. The pixel size was 3mm. Rather than being anthropomorphic, this phantom was designed to demonstrate that the modified penalty induces nearly uniform spatial resolution even for problems where the standard penalty yields highly nonuniform spatial resolution.

We simulated a PET emission scan with 128 radial bins and 110 angles uniformly spaced over $180°$. The g_{ij} factors corresponded to 6mm wide strip integrals with 3mm center-to-center spacing. We set $r_i = 0.1 \frac{1}{N} \sum_{i'} \sum_j a_{i'j}\theta_j$, which corresponds to 10% random coincidences.

6.1 Resolution Uniformity

We computed local impulse response functions $l^j(\theta)$ for three pixels j, corresponding to the center of the cold disk, the center of the image, and the center of the hot disk. We used the recipe following (9) with $\delta = 0.01$ to evaluate $l^j(\theta)$, for both the standard penalty (26) and the modified penalty (40) with $\psi(x) = x^2/2$. For both penalties we used a first-order neighborhood. We used this recipe rather than any of the approximations that followed it (such as (20)) to provide a more convincing demonstration; for routine work we usually just use (31). (The results of (31) are not shown in Fig. 4 since they turn out to be indistinguishable from the curves shown, which supports the accuracy of the approximations leading to (31).) We maximized the objective function (15) to compute $\hat{\theta}$ in (5) using 20 iterations of the PML-SAGE-3 algorithm [19].

Fig. 4 displays horizontal and vertical profiles through the local impulse responses for the estimators corresponding to the two penalty functions. The circles in Fig. 4 are for the unbiased estimator (6) for $M = 2000$ realizations. The standard penalty has highly nonuniform spatial resolution, whereas the modified penalty yields nearly uniform spatial resolution. These results are typical.

6.2 Asymmetry

In part because of the large eccentricity of the ellipse in Fig. 3, the local impulse responses of both penalties are asymmetric. Fig. 5 displays contours at levels 25, 50, 75, and 99% of the peak value for each PSF, computed using the `contour` function of Matlab. We hope to extend the analysis in this paper to develop suitable modifications to the penalty that will reduce this asymmetry. (The corresponding contours for FBP were virtually circular.)

6.3 Choosing β

We now describe how we selected β for this simulation, which illustrates the effectiveness of the table-based approach described in Section 5.5. First, we decided for illustration purposes to use a FWHM of 4 pixels. Since the strip width is twice the

---

[13]For transmission imaging, the detector efficiences etc. are included in $\{b_i\}$, so $A = G$, i.e., there are no c_i's.

pixel size in this example, the detector response (as discussed in Section 9) is a rectangular function with frequency response $S_{\text{true}}(u) = \text{sinc}(2u)$ From the analytical results shown in Fig. 18 for that detector blur, we need $\log_2 \beta_0 \approx 9.3$. From (52) in Section 9

$$\log_2 \beta = \log_2 \frac{\beta_0 \Delta_\theta \Delta_b}{4\pi^4} = 9.3 + \log_2 \frac{\frac{\pi}{110}\frac{3}{3}}{4\pi^4} \approx -4.44,$$

so we used $\beta = 2^{-4.44}$ for the modified penalty[14]. Did this choice of β actually give the desired 4 pixels FWHM resolution? Since Fig. 5 shows that the local impulse response is asymmetric, clearly the resolution is not exactly 4 pixels FWHM isotropically. In particular, for the same 3 pixels considered above, the horizontal resolutions were 3.10, 3.38, and 3.34 pixels FWHM, whereas the vertical resolutions were 5.28, 4.83, and 4.76 pixels FWHM. However, the averages of the horizontal and vertical resolutions were 4.19, 4.10, and 4.05 pixels FWHM, all of which are within 5% of the target resolution of 4 pixels FWHM. Thus, although further work is needed to correct the asymmetry in such eccentric objects, the proposed method for selecting β appears to yield local impulse responses whose *average* resolution is very close to the desired resolution. These results are typical in our experience.
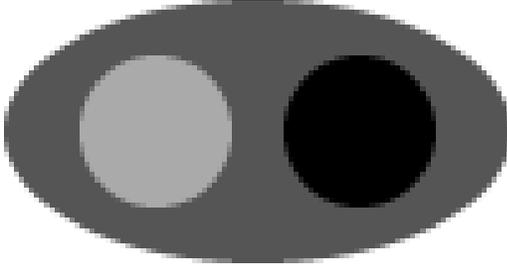
Figure 3: Digital phantom used to examine spatial resolution properties.

7 What Happens to the Variance?

It is well known that the global smoothing parameter β controls an overall tradeoff between resolution and noise: larger β's lead to coarser resolution but less noise, and vice-versa. The analysis in preceding sections shows that for the modified penalty to induce uniform spatial resolution, the "local" smoothing parameter must effectively be larger in some areas, and smaller in others. Thus, it is natural to expect that these changes in the local resolution will also influence the noise—but is the influence global or local? I.e., if the modified penalty increases the resolution (and hence the noise) at a given pixel, will that noise

---

[14]For the standard penalty, we used the above value of β scaled down by κ_j^2 for the single j corresponding to the pixel at the center of the image, as suggested by (39) and described in Section 4.3. This choice matched the resolution at the image center for the two penalties, as illustrated in the center plots of Figs. 4 and 5.
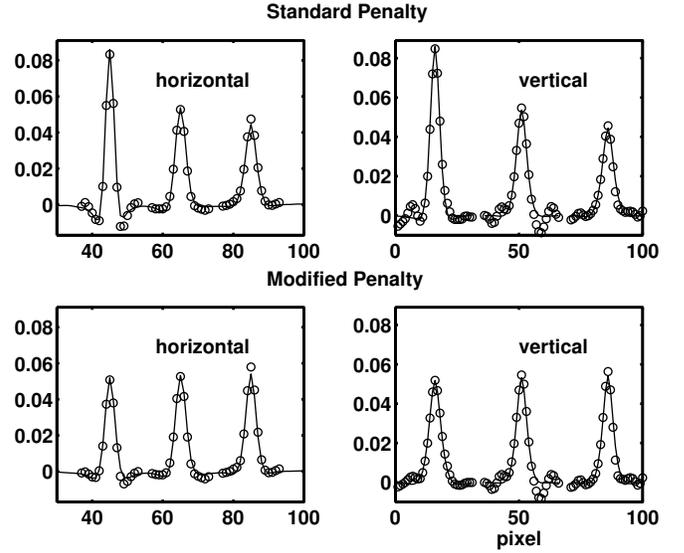
Figure 4: Horizontal and vertical profiles (concatenated left to right) through three local impulse response functions for a penalized-likelihood estimate of the image shown in Fig. 3. The standard quadratic penalty yields highly nonuniform resolution (upper profiles), whereas the proposed modified penalty leads to nearly uniform spatial resolution (lower profiles). Note that for the standard penalty the resolution is poorest in the high-count disk.

somehow propagate to distant pixels and lead to an overall worse resolution/noise tradeoff?

To address this question, we generated 100 realizations of Poisson distributed simulated PET measurements for the object shown in Fig. 3, and for the system properties described in Section 6. For each realization $y^{(1)}, \ldots, y^{(100)}$, we used 20 iterations of PML-SAGE-3 [19] to compute penalized-likelihood estimates $\{\hat{\theta}(y^{(m)})\}_{m=1}^{100}$ for several values of β for both the standard and the modified quadratic penalties. For each value of β, we computed the empirical standard deviation of $\hat{\theta}_j$ for the pixels at the centers of the two disks in Fig. 3. (The results were similar for the pixel at the image center, so are not shown.)

7.1 Just What You Expected

Fig. 6 shows the tradeoff between resolution (measured by the average FWHM of the local impulse response) and noise (measured by the empirical standard deviation) as β is varied. Fig. 6 also shows predicted standard deviations computing using the variance approximations described in [38]. (The good agreement between empirical and predicted results in Fig. 6 is further confirmation of the utility of the approximations in [38].)

In Fig. 6, the resolution/noise data points follow an essentially identical tradeoff curve *for both the standard and the modified penalty*. This is true both for the analytically predicted tradeoff (the solid line and the dashed line overlap almost perfectly) as well as for the empirical results (the circle and the plus symbols lie on the same curve). These results suggest that the effects of the modified penalty are essentially local: a given pixel moves
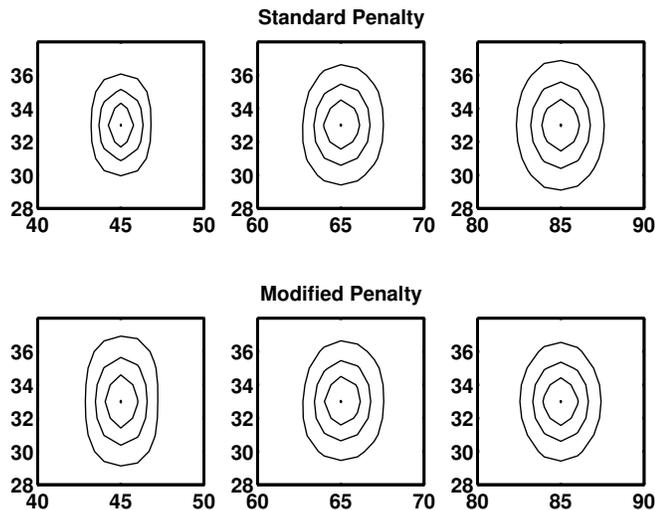
Standard Penalty

Modified Penalty

Figure 5: Contours of the local impulse response functions at 25, 50, 75, and 99% of each peak. Left: center of cold disk, middle: center of image, right: center of hot disk.

up or down its own resolution/noise tradeoff curve to the specified resolution, and then has a variance that is the same value as would be obtained if one were to use the standard penalty but globally adjust β to enforce that specified resolution at the given pixel. This property probably hinges on the fact that the κ_j factors are spatially smooth. If one were to artificially create an κ_j map having discontinuities and then apply the modification (40), then it is plausible that the results would be less regular than indicated in Fig. 6. Readers who apply variations of (40) to induce some type of data-based non-uniform resolution will need to consider the resolution/noise tradeoff in more detail.

Fig. 25 shows central horizontal profiles through empirical standard deviation maps of the penalized likelihood estimates for both the modified and the standard quadratic penalties. Also shown is a calculated prediction of the variance, an approximation developed in [44]. As noted in footnote 14, the penalties were chosen to have matched resolution at the image center, and in Fig. 25 the estimator variance is also matched at the image center. Note however that whereas the variance for the standard penalty is fairly uniform (at least for this object), the variance for the modified penalty is nonuniform. (Of course as we have shown it is the other way around for the spatial resolution.) This nonuniformity is consistent with the results of Fig. 6.

7.2 Quadratic Penalties <u>Are</u> Useful

Fig. 7 compares the resolution/noise tradeoff of penalized likelihood with that of images reconstructed by FBP with a Hamming window and with a constrained least-squares (CLS) window corresponding to (50) of Section 9:

$$\frac{\operatorname{sinc}(2u)\,/\operatorname{sinc}(u)}{\operatorname{sinc}^2(2u)+\beta u^3}, \quad u \in [0, \tfrac{1}{2}], \tag{45}$$

(where u denotes spatial frequency: cycles per radial sample). (Dividing by $\operatorname{sinc}(u)$ compensates for the linear interpolation

step of backprojection in the FBP algorithm. We found this correction improved the match between the PSF of FBP and the PSF of penalized likelihood.) This window induces a PSF indistinguishable from that of penalized-likelihood estimates with the first-order quadratic penalty . As shown by Fig. 7, at any given resolution the empirical standard deviations for the FBP images are higher than for the penalized-likelihood estimates. This demonstrates that even using the oft-maligned quadratic penalty, penalized-likelihood image reconstruction can outperform FBP in terms of the tradeoff between resolution and noise. Of course nonquadratic prior models may give even better results for objects that are consistent with those models, but results such as Fig. 7 show that quadratic penalties provide a useful reduction in image noise over a large range of spatial resolutions.
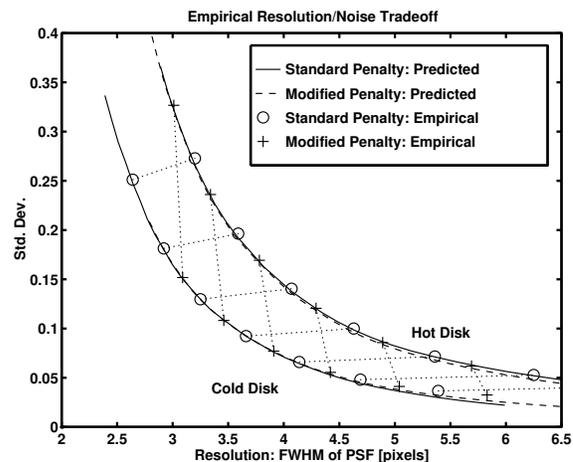

Figure 6: Resolution/noise tradeoff for penalized-likelihood emission image reconstruction with standard and modified quadratic penalties. The two penalties induce virtually identical tradeoff curves. (The dotted lines connect points that correspond to the same β value.)

8 Discussion

We have analyzed the local impulse response of implicitly defined estimators (14) and of penalized-likelihood estimators for emission tomography (20) and transmission tomography (23). The analysis and empirical results show that the local impulse response is asymmetric and has nonuniform resolution for Poisson distributed measurements. We proposed a modified regularization penalty (40) that improved the spatial resolution uniformity but not the asymmetry.

For the space-invariant tomographs considered here, the resolution nonuniformity arises from the nonuniform diagonal of the Fisher information matrix, which in turn is a consequence of the nonuniform variance of Poisson noise. In principle one could "avoid" this problem altogether by using an *unweighted* least-squares estimator. We have shown qualitatively in [22] that nonuniform weighting is essential to achieve the desirable noise properties of statistical methods. In Section 10, we provide additional analyses and quantitative results that demonstrate the

12

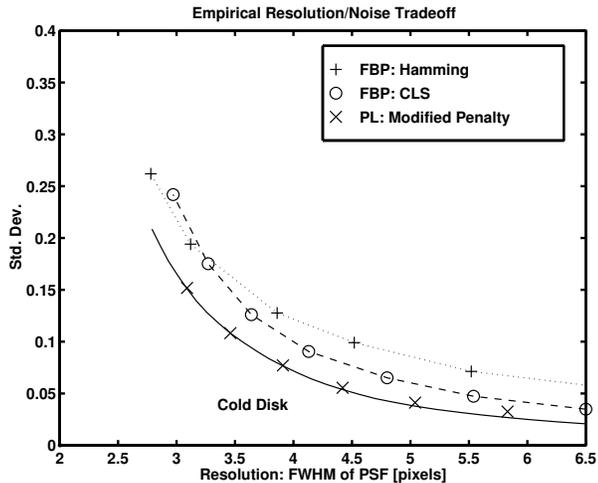**Empirical Resolution/Noise Tradeoff**

Figure 7: Resolution/noise tradeoff of FBP with Hamming window and the constrained least-squares (CLS) window (45). At any given resolution, the variances of the penalized-likelihood estimates are smaller than those of FBP.

importance of weighting. Therefore we advocate retaining the nonuniform weighting that is natural for Poisson statistics, but modifying the penalty to compensate for the undesirable spatial resolution properties. Fortunately this modification does not destroy the benefits of the weighting, as shown in Section 10 and in Fig. 7, apparently because the nonuniform weighting is applied in sinogram space, whereas the penalty acts on the image space. It is an open question as to whether the modified penalty would be effective for problems such as restoration of quantum-limited image measurements, where both the unknown parameters and the data are images.

Some colleagues have argued that nonuniform resolution is desirable and expected. This opinion is presumably based on the idea that statistical methods can make better use of the measurement information and thus provide higher resolution in high-count regions. Ironically, our analysis shows that the effect of uniform penalties is just the opposite: more smoothing occurs in high-count regions. Although we have emphasized methods for achieving resolution uniformity, one could apply our analysis to develop alternative modified penalties that yield higher resolution in high-count regions according to some user-specified criterion. Since we now see that the statistics of the data themselves do not automatically provide a natural resolution-noise tradeoff in penalized-likelihood estimators (contrary to what may have been a widely held misconception), any such user-specified criteria will probably be considered somewhat arbitrary.

We have shown the somewhat remarkable result that the local impulse response induced by quadratic penalties depends on the object only through its projections. Thus, one does not need to know the object to predict the reconstructed resolution, since the noisy measurements serve as an adequate approximation to the object's projections. In contrast, the local impulse response for nonquadratic penalties depends explicitly on the (unknown) object (cf (25)) through the Hessian of the penalty. Being able to predict and control the resolution properties induced by such

penalty functions remains an important challenge.

For nonquadratic edge-preserving potential functions ψ, the nonuniform diagonal in (25) may induce additional types of nonuniformities beyond the resolution effects reported here. Specifically, we conjecture that the "propensity to retain edges" (as opposed to smoothing them out) will be space-variant, again due to coupling between the Hessian of the log-likelihood and the Hessian of the penalty in (25). If so, then modified penalties such as (40) may be useful for restoring the (presumably desirable) space invariance of the effects of edge-preserving penalties. The importance of such modifications is more likely to appear in rigorous studies of the *ensemble* characteristics of edge-preserving methods, rather than in anecdotal examples.

This paper has emphasized space-invariant tomographs. Further investigation is needed for space-variant systems such as SPECT emission measurements and truncated data such as fan-beam transmission SPECT and 3D cylindrical PET.

Part II: Resolution and Noise Properties of Penalized Least-Squares Image Reconstruction

9   PSF of Idealized Tomograph

Although the discrete formulation used in Section 3 is appropriate for computer implementation, it lacks the insight one can obtain from analytical methods. This section derives the PSF of one form of penalized-likelihood tomography by analyzing an idealized tomograph having a continuum of detectors and angles. The model is partially realistic, however, since we account for a radial detector blur. The analytical form we derive for this continuous tomograph shows remarkable agreement with simulations using discrete sampled systems.

Let $f(x_1, x_2)$ denote an object defined on $\mathbb{R}^2$. Let $\mathcal{P}$ denote the continuous Radon transform operator. If $p = \mathcal{P}f$, then

$$p_\phi(r) = \int f(l\cos\phi + r\sin\phi, l\sin\phi - r\cos\phi)\, dl.$$

Let $\mathcal{S}_{\text{true}}$ denote a space-invariant radial sinogram blurring operator with symmetric kernel $s_{\text{true}}(r)$. If $y = \mathcal{S}_{\text{true}}p$, then

$$y_\phi(r) = s_{\text{true}}(r) \star p_\phi(r)$$

where $\star$ denotes 1D convolution.

Define $\mathcal{G}_{\text{true}} = \mathcal{S}_{\text{true}}\mathcal{P}$ to be the blurred Radon transform operator, which is analogous to the matrix $\boldsymbol{G}$ in Section 5. Given measurements with additive zero-mean noise:

$$y = \mathcal{G}_{\text{true}}f + \text{noise},$$

we would like to recover f from y. We may not know the blur function $\mathcal{S}_{\text{true}}$ exactly, but rather may only have an approximation $\mathcal{S}_{\text{model}}$ with kernel $s_{\text{model}}(r)$. A penalized least squares (PLS) approach to this problem is:

$$\hat{f} = \arg\min_f \|y - \mathcal{G}_{\text{model}}f\|^2 + \alpha\langle f, \mathcal{R}f\rangle, \qquad (46)$$

where $\mathcal{G}_{\text{model}} = \mathcal{S}_{\text{model}}\mathcal{P}$ and where the norms and inner products are the standard L_2 functionals on $\mathbb{R}^2$.

If we desire smooth solutions $\hat{f}$, then we would like $\langle f, \mathcal{R}f\rangle$ to be a measure of roughness. Therefore we define $\mathcal{D}_j$ to be the differentiation operator with respect to the jth spatial coordinate:

$$\mathcal{D}_j f = \frac{\partial}{\partial x_j} f(x_1, x_2), \quad j = 1, 2.$$

Thus

$$\|\mathcal{D}_j f\|^2 = \iint \left(\frac{\partial}{\partial x_j} f(x_1, x_2)\right)^2\, dx_1\, dx_2.$$

By defining

$$\mathcal{R} = \sum_{k=0}^{m} \binom{m}{k} (\mathcal{D}_1'\mathcal{D}_1)^k (\mathcal{D}_2'\mathcal{D}_2)^{m-k},$$

we have specified $\langle f, \mathcal{R}f\rangle$ to be an isotropic measure of the roughness of f. (Note that operators $\mathcal{D}_1$ and $\mathcal{D}_2$ commute.) In particular, for the usual choice $m = 1$, we have

$$\mathcal{R} = \mathcal{D}_1'\mathcal{D}_1 + \mathcal{D}_2'\mathcal{D}_2,$$

which is analogous to the matrix $\boldsymbol{R}$ in Section 3.

Assuming $s_{\text{model}}(r)$ is a low-pass filter whose transfer function $S_{\text{model}}(u)$ is nonzero at $u = 0$, one can easily show that the null spaces of $\mathcal{G}_{\text{model}}$ and $\mathcal{R}$ are disjoint, so the solution to (46) is

$$\hat{f} = [\mathcal{G}_{\text{model}}'\mathcal{G}_{\text{model}} + \alpha\mathcal{R}]^{-1}\mathcal{G}_{\text{model}}'y,$$

and the estimator mean is:

$$\mu(f) = E\{\hat{f}\} = [\mathcal{G}_{\text{model}}'\mathcal{G}_{\text{model}} + \alpha\mathcal{R}]^{-1}\mathcal{G}_{\text{model}}'\mathcal{G}_{\text{true}}f. \quad (47)$$

This is a space-invariant mean response, i.e. $\mu(f)$ corresponds to a filtered version of f. We now use Fourier methods to derive the frequency response of that filter.

9.1   Frequency Response

It is well known [45] that

$$\mathcal{P}'\mathcal{P}f = f ** \frac{1}{r},$$

where $**$ denotes 2D convolution. If we let u_1 and u_2 denote the 2D spatial frequency coordinates, and define $\rho = \sqrt{u_1^2 + u_2^2}$, then since the 2D Fourier transform of $1/r$ is $1/\rho$, we have

$$\mathcal{P}'\mathcal{P} = \mathcal{F}_2' \cdot \frac{1}{\rho} \cdot \mathcal{F}_2,$$

where $\mathcal{F}_2$ denotes the 2D Fourier operator.

Since the blur operator $\mathcal{S}_{\text{model}}$ acts radially, by the Fourier-slice theorem:

$$\mathcal{G}_{\text{model}}'\mathcal{G}_{\text{model}} = \mathcal{F}_2' \cdot \frac{|S_{\text{model}}(u)|^2}{\rho} \cdot \mathcal{F}_2. \qquad (48)$$

Similarly

$$\mathcal{G}_{\text{model}}'\mathcal{G}_{\text{true}} = \mathcal{F}_2' \cdot \frac{S_{\text{model}}^\star(u)S_{\text{true}}(u)}{\rho} \cdot \mathcal{F}_2.$$

From the differentiation property of Fourier transforms:

$$\mathcal{D}_j'\mathcal{D}_j = \mathcal{F}_2' \cdot (2\pi u_j)^2 \cdot \mathcal{F}_2,$$

so

$$\begin{aligned}
\mathcal{R} &= \mathcal{F}_2' \cdot \left[\sum_{k=0}^{m} \binom{m}{k} (2\pi u_1)^{2k}(2\pi u_2)^{2(m-k)}\right] \cdot \mathcal{F}_2 \\
&= \mathcal{F}_2' \cdot [(2\pi u_1)^2 + (2\pi u_2)^2]^m \cdot \mathcal{F}_2 \\
&= \mathcal{F}_2' \cdot (2\pi\rho)^{2m} \cdot \mathcal{F}_2. \qquad (49)
\end{aligned}$$

Combining (47), (48), and (49) shows that

$$\mu(f) = \mathcal{F}_2' \cdot L_0(\rho) \cdot \mathcal{F}_2 f,$$

14

where the radially-symmetric frequency response $L_0(\rho)$ of the PLS estimator is therefore:

$$L_0(\rho) = \frac{\frac{S_{\mathrm{model}}^\star(u) S_{\mathrm{true}}(u)}{\rho}}{\frac{|S_{\mathrm{model}}(u)|^2}{\rho} + \alpha(2\pi\rho)^{2m}} = \frac{S_{\mathrm{model}}^\star(u) S_{\mathrm{true}}(u)}{|S_{\mathrm{model}}(u)|^2 + \beta_0 \rho^{2m+1}}, \quad (50)$$

where $\beta_0 = \alpha(2\pi)^{2m}$. Not surprisingly, the response is unity near $\rho = 0$, and then typically decreases with increasing spatial frequency. The form of (50) is very similar to a constrained least-squares restoration filter [46], or for an ideal system with $S_{\mathrm{model}}(u) = S_{\mathrm{true}}(u) = 1$, to a Butterworth low-pass filter. Typical plots of $L_0(\rho)$ are shown in Figs. 10-13.

9.2 Point Spread Function

Since the frequency response is radially symmetric, one can compute the corresponding PSF $l_0(r)$ using the Hankel transform [47]:

$$l_0(r) = 2\pi \int_0^\infty L_0(\rho) \, J_0(2\pi\rho r) \, \rho \, d\rho,$$

where J_0 is the 0th order Bessel function. We do not know of an analytical form for the Hankel transform of (50), even when $S_{\mathrm{true}}(u) = S_{\mathrm{model}}(u) = 1$, but it is easy to evaluate the integral numerically. As $\beta_0 \to 0$, the FWHM of $l_0(r)$ approaches 0, whereas for a discrete system, the smallest possible FWHM is 1 pixel. Therefore, we also define the blurred response:

$$l_1(r) = l_0(r) \star 1_{\{|r| \le 1/2\}}, \quad (51)$$

where $1_{\{|r| \le 1/2\}}$ denotes the standard rectangular function. As $\beta_0 \to 0$, the FWHM of $l_1(r)$ approaches 1, which better agrees with the discrete results.

Typical plots of the normalized PSF $l_0(\rho)/l_0(0)$ are shown in Figs. 14-17, for several value of β_0 and for different radial blurs $S_{\mathrm{true}}(u) = S_{\mathrm{model}}(u)$ with no model mismatch. Note that when β_0 is small, the point response functions exhibit ringing.

Using analytically computed $l_1(\rho)$, we can tabulate the relationship between β_0 and the FWHM of the PSF. Typical curves are shown in Fig. 18, for different $S_{\mathrm{true}}(u) = S_{\mathrm{model}}(u)$ cases. In principle, one can choose a desired resolution, and then read off the appropriate β_0 from Fig. 18. This value of β_0 will be proportional to the value of β that should be used with the modified penalty of Section 5. The proportionality constant is object independent, so only needs to be determined once for a given geometric system matrix $\boldsymbol{G}$ as defined in Section 5. The constant depends on the units one uses when defining $\boldsymbol{G}$ and $R^\star$. For a strip-integral tomographic system $\boldsymbol{G}$, we normalize the elements of $\boldsymbol{G}$ so that $\sum_i g_{ij} = 1$, which is "count preserving." In this case, careful bookkeeping showed that

$$\beta_0 = \beta \frac{4\pi^4}{\Delta_\theta \Delta_b}, \quad (52)$$

where Δ_θ is the angular spacing and Δ_b is the radial center-to-center spacing of the strip integrals.

Using this relationship between β and β_0, we computed the *discrete* local impulse response (43) and the corresponding *analytical* impulse response (51) and plotted the FWHM of the

two in Fig. 19, for the case of a rectangular radial blur with FWHM=2 pixels, i.e. $S_{\mathrm{true}}(u) = S_{\mathrm{model}}(u) = \mathrm{sinc}(2u)$. For resolutions greater than about 2 pixels FWHM, the resolution of the discrete and analytical impulse responses agree quite well. (For FWHM below 2 pixels, the effects of the discrete pixels apparently yield a slightly greater FWHM than predicted by (51).) Figure 20 displays the analytical PSF (51) and the discrete local impulse response (43) for the case $\beta_0 = 175$, for a tomograph with rectangular strip integrals with two pixel width. The agreement shown in Figs. 19 and 20 confirms that one can use Fig. 18 in conjunction with relationship like (52) to determine a value for β that will provide reconstructed images having the user's desired spatial resolution.

We remind the reader that the fact that this PLS estimator gives a response similar to a Butterworth filter does not imply that in general penalized-likelihood estimation is equivalent to Butterworth filtering! The above analysis does not include nonnegativity, Poisson statistics, realistic system modeling, nonquadratic penalties, and the other well-known advantages of statistical methods. However, the above analysis is useful for understanding and quantifying basic resolution properties. It is also useful for designing preconditioners for fast gradient-based iterative methods [23, 40].

9.3 A FWHM Rule of Thumb

It is well known that the FWHM of bump-shaped point spread functions is approximately equal to $1/(2\rho_h)$, where ρ_h is the half-amplitude frequency, i.e. $L_0(\rho_h) = \frac{1}{2} L_0(0)$. From (50), it is clear that the half-amplitude frequency for the PLS estimator is $\beta_0^{-1/(2m+1)}$ for the ideal case when $S_{\mathrm{true}}(u) = 1$. Thus for the usual $m = 1$ case, the FWHM of $l_0(r)$ is approximately $\beta_0^{1/3}$. Since we are more interested in $l_1(r)$, Fig. 9 displays the FWHM of $l_1(r)$ versus $\beta_0^{1/3}$ for various detector blurs. There is a nearly affine relationship between the FWHM of $l_1(r)$ and $\beta_0^{1/3}$ for $\beta_0 > 5$, which may be used to simplify further the table lookup method for relating β to FWHM described in Section 5.5.

10 Variance Approximations

COMPARE WITH (45) of [42].

The approximation (36) provides an opportunity to derive simple approximate expressions for the variance of penalized-likelihood and unweighted penalized least-squares estimators. This section derives such approximations and demonstrates that they are usefully accurate. The expressions give further insight into why weighted statistical methods outperform unweighted statistical methods for image reconstruction, as illustrated qualitatively in [22].

In this section we focus on quadratic penalties, although one can extend the analysis to nonquadratic penalties. The following matrix will be central to all the approximations that follow:

$$\boldsymbol{M}_\beta = [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1} \boldsymbol{G}'\boldsymbol{G} [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}, \quad (53)$$

where $\boldsymbol{R}^\star$ is the Hessian of the standard quadratic penalty. The matrix $\boldsymbol{M}_\beta$ has the following interpretation. Suppose we could

15

observe hypothetical noisy measurements y with mean $G\theta$ and with the identity covariance matrix. If we applied the penalized least-squares estimator $\hat{\theta} = [G'G + \beta R^\star]^{-1}G'y$ to such measurements, then M_β would be the covariance of $\hat{\theta}$. In this hypothetical scenario, the following term

$$\sigma(\beta) = \sqrt{(e^j)'M_\beta e^j} = \sqrt{[M_\beta]_{jj}} \qquad (54)$$

would be the standard deviation of $\hat{\theta}_j$, where $(e^j)'M_\beta e^j = [M_\beta]_{jj}$ is the jth diagonal entry of M_β. Note that for quadratic penalties, the matrix M_β is independent of the object θ, so one could precompute M_β or key portions of M_β (such as the central few diagonals) using algorithms similar to those discussed in [38]. For nearly space-invariant tomographs, the diagonal of M_β is approximately constant, so one can easily compute diagonal elements from (53) using Fourier methods.

The approximations that follow below stem from the following approximation for the covariance of penalized-likelihood estimates $\hat{\theta}^{\mathrm{PL}}$ in emission tomography [38]:

$$\mathsf{Cov}\left\{\hat{\theta}^{\mathrm{PL}}\right\} \approx [F + \beta R]^{-1}F[F + \beta R]^{-1}, \qquad (55)$$

where F was defined by (33).

10.1 The Standard Penalty

For the standard penalty one uses $R = R^\star$. Substituting in the approximation (36) into (55) and simplifying yields

$$\mathsf{Cov}\left\{\hat{\theta}^{\mathrm{PL}}\right\} \approx$$

$$\Lambda_\theta^{-1}[G'G + \beta\Lambda_\theta^{-1}R^\star\Lambda_\theta^{-1}]^{-1}G'G[G'G + \beta\Lambda_\theta^{-1}R^\star\Lambda_\theta^{-1}]^{-1}\Lambda_\theta^{-1}.$$

That is not much simplification, but if one restricts attention to examining the variance of particular pixels, then using approximations in the same spirit as (39) we have

$$\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{PL}}\right\} = (e^j)'\mathsf{Cov}\left\{\hat{\theta}^{\mathrm{PL}}\right\}e^j \approx$$

$$\frac{(e^j)'[G'G + \beta\Lambda_\theta^{-1}R^\star\Lambda_\theta^{-1}]^{-1}G'G[G'G + \beta\Lambda_\theta^{-1}R^\star\Lambda_\theta^{-1}]^{-1}e^j}{\kappa_j^2(\theta)}$$

$$\approx \frac{(e^j)'[G'G + \beta/\kappa_j^2(\theta)R^\star]^{-1}G'G[G'G + \beta/\kappa_j^2(\theta)R^\star]^{-1}e^j}{\kappa_j^2(\theta)}.$$

Thus for the standard quadratic penalty:

$$\sqrt{\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{PL}}\right\}} \approx \frac{\sigma(\beta/\kappa_j^2(\theta))}{\kappa_j(\theta)}. \qquad (56)$$

One can easily tabulate $\sigma(\beta)$ for a range of values of β, and then apply linear interpolation to evaluate (56) for many pixels j. (See Fig. 25.)

10.2 The Modified Penalty

The modified penalty leads to a simpler approximation for the estimator covariance. Substituting in the approximations (36) and (41) into (55) and simplifying yields:

$$\mathsf{Cov}\left\{\hat{\theta}^{\mathrm{PL}}\right\} \approx \Lambda_\theta^{-1}M_\beta\Lambda_\theta^{-1}. \qquad (57)$$

If one only wants the variances, then further simplification is possible. Note that from (37), Λ_θ is a diagonal matrix, so using (57):

$$\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{PL}}\right\} = (e^j)'\mathsf{Cov}\left\{\hat{\theta}^{\mathrm{PL}}\right\}e^j \approx \frac{(e^j)'M_\beta e^j}{\kappa_j^2(\theta)}.$$

We thus have the following approximation to the standard deviation of $\hat{\theta}_j^{\mathrm{PL}}$ for the modified penalty:

$$\sqrt{\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{PL}}\right\}} \approx \frac{\sigma(\beta)}{\kappa_j(\theta)}. \qquad (58)$$

Using this approximation, we can easily compute a "standard deviation map," where the jth pixel intensity is $\sqrt{\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{PL}}\right\}}$. (Note that the numerator is a constant independent of j.) Fig. 24 displays such a map, along with the empirical standard deviation map from the $M = 2000$ realizations described in Section 6, in this case for $\beta = 2^{-4.44}$. Fig. 25 displays horizontal profiles through the two maps shown in Fig. 24. Despite our use of several approximations that might as first seem fairly crude, the predicted and empirical results agree remarkably well. The largest disparity is at the edge where the object is zero and the empirical standard deviation is near zero small due to the nonnegativity constraint. (We would expect similar discrepancies in interior cold regions.) But for usual regions that are not too close[15] to zero, the above approximation should be useful for purposes such as generalizing Huesman's weighting method [48, 49] from FBP to penalized-likelihood estimators, or for searching for statistically significant regions in brain activation studies [50, 51].

10.3 Unweighted Penalized Least Squares

FBP is an example of an *unweighted* estimator: all measurements are treated equally. As noted in [52], FBP (with a ramp filter) is closely related to an unweighted least-squares estimate, in the special case when the system matrix is the Radon transform (i.e. no blur). Using an apodizing window with FBP is essentially equivalent to using a quadratic penalty with an unweighted penalized least-squares estimator (cf Section 9). We now derive an approximate expression for the covariance of the following unweighted penalized least-squares estimator:

$$\begin{aligned} \hat{\theta}^{\mathrm{QPULS}} &= \arg\min_{\theta \geq 0} \|\tilde{y} - G\theta\|^2 + \beta\theta'R^\star\theta \\ &\approx [G'G + \beta R^\star]^{-1}G'\tilde{y}, \end{aligned}$$

where

$$\tilde{y} = D\left[c_i^{-1}\right]y - r$$

---

[15]For the above approximation to be accurate, the estimate should be at least one or maybe two standard deviations above zero.

is a sinogram precorrected for attenuation, randoms, etc. Note that there is no need to use the modified penalty for this unweighted estimator, since (ignoring the nonnegativity constraint) the local impulse response is easily shown to be

$$l^j = [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}\boldsymbol{G}'\boldsymbol{G}e^j,$$

which is independent of θ. Except for the nonnegativity constraint, the estimator $\hat{\theta}^{\mathrm{QPULS}}$ is linear, so an nearly exact expression for its covariance is:

$$\mathsf{Cov}\left\{\hat{\theta}^{\mathrm{QPULS}}\right\} \approx [\boldsymbol{G}'\boldsymbol{G}+\beta\boldsymbol{R}^\star]^{-1}\boldsymbol{G}'D\left[q_i^{-1}\right]\boldsymbol{G}[\boldsymbol{G}'\boldsymbol{G}+\beta\boldsymbol{R}^\star]^{-1},$$

where q_i^{-1} is the variance of $\tilde{y}_i$, and $\{q_i\}$ were defined in (21). An approximation analogous to (36) is:

$$\boldsymbol{G}'D\left[q_i^{-1}\right]\boldsymbol{G} \approx \Omega\boldsymbol{G}'\boldsymbol{G}\Omega$$

where $\Omega = D[\omega_j]$ is diagonal with entries

$$\omega_j = \sqrt{\sum_i g_{ij}^2 q_i^{-1} / \sum_i g_{ij}^2},$$

which leads to the covariance approximation:

$$\mathsf{Cov}\left\{\hat{\theta}^{\mathrm{QPULS}}\right\} \approx [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}\Omega\boldsymbol{G}'\boldsymbol{G}\Omega[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}.$$

If Ω were a scaled identity matrix, it would commute with $[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}$. Since the diagonal elements of Ω vary slowly spatially, it is reasonable to suppose that Ω and $[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}$ approximately commute. More specifically, let e^j be any unit vector, then we maintain that

$$\Omega[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}e^j \approx [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}\omega_j e^j, \qquad (59)$$

since $[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}e^j$ is fairly well localized (see Fig. 8). It follows from (59) that

$$\Omega[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1} \approx [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}\Omega, \qquad (60)$$

which leads to the following approximation for the covariance of the unweighted penalized least squares estimator:

$$\mathsf{Cov}\left\{\hat{\theta}^{\mathrm{QPULS}}\right\} \approx \Omega\boldsymbol{M}_\beta\Omega,$$

where $\boldsymbol{M}_\beta$ was defined by (53). Thus

$$\sqrt{\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{QPULS}}\right\}} \approx \sigma(\beta)\omega_j. \qquad (61)$$

For the 100 realizations described in Section 6, we computed $\hat{\theta}^{\mathrm{QPULS}}$ using the iterative algorithm described in [22], *including the nonnegativity constraint*. Fig. 26 shows the empirical sample standard deviations map of those estimates, as well as the approximation (61). Fig. 27 shows profiles through those maps, and again demonstrates the accuracy of the variance approximation (61). Again, the greatest disagreement is outside of the object where the nonnegativity constraint is active.
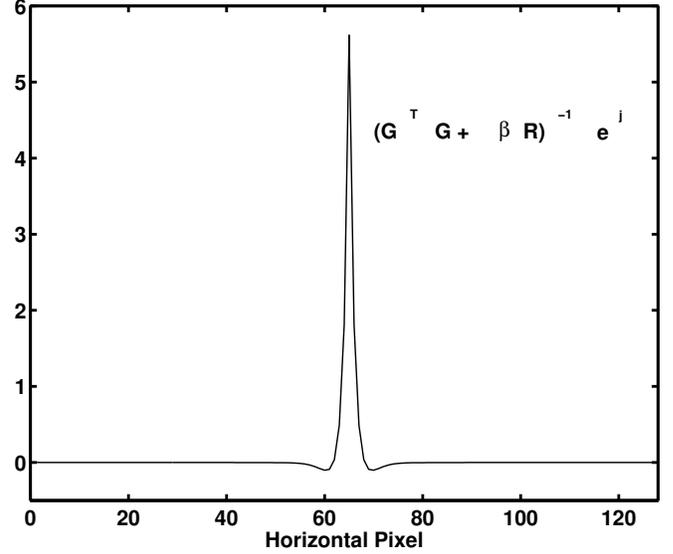
Figure 8: Horizontal profile through typical case of $[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}^\star]^{-1}e^j$, in this case for $\beta = -4$.

10.4 Comparison

From (58), (61), and (35), we have:

$$\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{PL}}\right\} \approx c^2\sum_i g_{ij}^2 / \sum_i g_{ij}^2 q_i = c^2 / \sum_i h_{ij}q_i$$

$$\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{QPULS}}\right\} \approx c^2\sum_i g_{ij}^2 q_i^{-1} / \sum_i g_{ij}^2 = c^2\sum_i h_{ij}q_i^{-1},$$

where $h_{ij} = g_{ij}^2 / \sum_{i'} g_{i'j}^2$ and $c = \sigma(\beta)$. Since $h_{ij} > 0$ and $\sum_i h_{ij} = 1$, it follows from Jensen's inequality that

$$\mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{PL}}\right\} < \mathsf{Var}\left\{\hat{\theta}_j^{\mathrm{QPULS}}\right\},$$

with equality if and only if all of the q_i's are equal to the same value, i.e., if and only if the measurements are homoscedastic (or if $\boldsymbol{G} = \boldsymbol{I}$). Homoscedasticity never happens for Poisson measurements in tomography, so the PL estimator will always have smaller variance than the UPLS estimator at equivalent spatial resolutions.

Fig. 27 compares profiles through the standard deviation maps for the penalized-likelihood and unweighted penalized least-squares estimates. The variances of the PL estimates are significantly lower than those of the UPLS estimates.

This result, though approximate, adds further evidence to the importance of using weighting for heteroscedastic measurements, either explicitly as in penalized weighted least-squares estimators [22,53], or implicitly by using penalized-likelihood estimators. We note without proof that using analyses similar to the above, one can show that precorrecting PET data for multiplicative effects such as detector efficiency and attenuation has the effect of making likelihood-based estimators more like an unweighted estimator, thereby destroying some of the benefits of using statistical methods.

11   Acknowledgment

A   FFT Simplifications

If the geometric system response G is approximately space invariant, and if the target penalty $R^\star$ is the standard quadratic penalty, then if one uses the modified quadratic penalty of [54], then the local impulse response is approximately given by (43) . In this important case, one can use FFT's to compute (43) approximately, thereby eliminating the need for an iterative method to compute (43) .

Let j denote the index of a pixel near the center of the image, and compute the kernel $k_{G'G} = G'Ge^j$. Let k_R be the kernel of R: $k_R = Re^j$. Let $F_{G'G}(u_1, u_2)$, $F_R(u_1, u_2)$, and $F_{\text{UPLS}}(u_1, u_2)$ denote the 2D FFT's of $k_{G'G}$, k_R, and $[G'G + \beta R^\star]^{-1}G'Ge^j$ respectively.

For the standard second-order quadratic penalty, the kernel of the regularization matrix $R^\star$ is

$$
\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix}. \tag{62}
$$

Then from standard properties of circulant matrices [33]:

$$
F_{\text{UPLS}}(u_1, u_2) = \frac{F_{G'G}(u_1, u_2)}{F_{G'G}(u_1, u_2) + \alpha F_R(u_1, u_2)}.
$$

Thus one can compute (43) using two 2D FFT's and one 2D inverse FFT.

For systems where the geometric response can be factored into product of the discretized Radon transform with a space-invariant blur, one can further simplify the calculation above.

One useful approximation to the kernel of $G'G$ is

$$
f(r) = \begin{cases} \pi - 2r, & r \in [0,1] \\ 2(\arcsin(1/r) - (r - \sqrt{r^2 - 1})) & r > 1, \end{cases}
$$

which is shown in Figure 2 of [22] (cf [21, Fig. 11] and [40, Fig. 1]). This function has the expected $1/r$ asymptotic form, but is well behaved near zero—as it must be for a discrete system.

References

[1] E. Veklerov and J. Llacer. Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Trans. Med. Imag.*, 6(4):313–9, December 1987.

[2] 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 Trans. Med. Imag.*, 6(3):228–38, September 1987.

[3] K. M. Hanson and G. W. Wecksung. Local basis-function approach to computed tomography. *Appl. Optics*, 24(23):4028–39, December 1985.

[4] R. M. Lewitt. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. *J. Opt. Soc. Am. A*, 7(10):1834–46, October 1990.

[5] K. Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Trans. Med. Imag.*, 9(4):439–46, December 1990. Corrections, T-MI, 10:2(288), June 1991.

[6] R. Leahy and X. H. Yan. Statistical models and methods for PET image reconstruction. In *Proc. of Stat. Comp. Sect. of Amer. Stat. Assoc.*, pages 1–10, 1991.

[7] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers. Regularized emission image reconstruction using imperfect side information. *IEEE Trans. Nuc. Sci.*, 39(5):1464–71, October 1992.

[8] J. A. Fessler and W. L. Rogers. Uniform quadratic penalties cause nonuniform image resolution (and sometimes vice versa). In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 4, pages 1915–9, 1994.

[9] S. J. Reeves. Optimal space-varying regularization in iterative image restoration. *IEEE Trans. Im. Proc.*, 3(3):319–23, May 1994.

[10] J. Nunez and J. Llacer. Variable resolution Bayesian image reconstruction. In *Proc. IEEE Workshop on Nonlinear Signal and Image Processing*, 1995.

[11] J. A. Stamos, W. L. Rogers, N. H. Clinthorne, and K. F. Koral. Object-dependent performance comparison of two iterative reconstruction algorithms. *IEEE Trans. Nuc. Sci.*, 35(1):611–4, February 1988.

[12] D. W. Wilson and B. M. W. Tsui. Spatial resolution properties of FB and ML-EM reconstruction methods. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1189–93, 1993.

[13] D. W. Wilson and B. M. W. Tsui. Noise properties of filtered-backprojection and ML-EM reconstructed emission tomographic images. *IEEE Trans. Nuc. Sci.*, 40(4):1198–1203, August 1993.

[14] H. H. Barrett, D. W. Wilson, and B. M. W. Tsui. Noise properties of the EM algorithm: I. Theory. *Phys. Med. Biol.*, 39(5):833–46, May 1994.

[15] D. W. Wilson, B. M. W. Tsui, and H. H. Barrett. Noise properties of the EM algorithm: II. Monte Carlo simulations. *Phys. Med. Biol.*, 39(5):847–72, May 1994.

[16] J. S. Liow and S. C. Strother. The convergence of object dependent resolution in maximum likelihood based tomographic image reconstruction. *Phys. Med. Biol.*, 38(1):55–70, January 1993.

[17] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans. Im. Proc.*, 5(3):480–92, March 1996.

[18] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. *IEEE Trans. Med. Imag.*, 13(3):687–701, December 1994.

[19] J. A. Fessler and A. O. Hero. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *IEEE Trans. Im. Proc.*, 4(10):1417–29, October 1995.

[20] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange. Fast parallelizable algorithms for transmission image reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1346–50, 1995.

[21] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. Sig. Proc.*, 41(2):534–48, February 1993.

[22] J. A. Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans. Med. Imag.*, 13(2):290–300, June 1994.

[23] S. D. Booth and J. A. Fessler. Combined diagonal/Fourier preconditioning methods for image reconstruction in emission tomography. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pages 441–4, 1995.

[24] G. T. Gullberg and T. F. Budinger. The use of filtering methods to compensate for constant attenuation in single-photon emission computed tomography. *IEEE Trans. Biomed. Engin.*, 28(2):142–57, February 1981.

[25] Y. Pawitan and F. O'Sullivan. Data-dependent bandwidth selection for emission computed tomography reconstruction. *IEEE Trans. Med. Imag.*, 12(2):167–172, June 1993.

[26] N. P. Galatsanos and A. K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Trans. Im. Proc.*, 1(3):322–336, July 1992.

[27] J. W. Hilgers and W. R. Reynolds. Instabilities in the optimal regularization parameter relating to image recovery problems. *J. Opt. Soc. Am. A*, 9(8):1273–9, August 1992.

[28] L. Kaufman and A. Neuman. PET regularization by envelope guided conjugate gradients. *IEEE Trans. Med. Imag.*, 15(3):385–6, June 1996.

[29] T. J. Hebert and R. Leahy. Statistic-based MAP image reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Sig. Proc.*, 40(9):2290–303, September 1992.

[30] T. R. Miller and J. W. Wallis. Clinically important characteristics of maximum-likelihood reconstruction. *J. Nuc. Med.*, 33(9):1678–84, September 1992.

[31] E. J. Hoffman, S. C. Huang, D. Plummer, and M. E. Phelps. Quantitation in positron emission tomography: 6 effect of nonuniform resolution. *J. Comp. Assisted Tomo.*, 6(5):987–99, October 1982.

[32] C. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Im. Proc.*, 2(3):296–310, July 1993.

[33] A. K. Jain. *Fundamentals of digital image processing*. Prentice-Hall, New Jersey, 1989.

[34] A. O. Hero, J. A. Fessler, and M. Usman. Exploring estimator bias-variance tradeoffs using the uniform CR bound. *IEEE Trans. Sig. Proc.*, 44(8):2026–41, August 1996.

[35] M. Usman, A. O. Hero, and J. A. Fessler. Uniform CR bound: implementation issues and applications. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1443–7, 1994.

[36] M. Usman, A. O. Hero, and J. A. Fessler. Bias-variance tradeoffs analysis using uniform CR bound for image reconstruction. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pages 835–9, 1994.

[37] R. E. Carson, Y. Yan, B. Chodkowski, T. K. Yap, and M. E. Daube-Witherspoon. Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm. *IEEE Trans. Med. Imag.*, 13(3):526–37, September 1994.

[38] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Trans. Im. Proc.*, 5(3):493–506, March 1996.

[39] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *J. Comp. Assisted Tomo.*, 8(2):306–16, April 1984.

[40] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Trans. Med. Imag.*, 12(1):78–83, March 1993.

[41] L. T. Chang. A method for attenuation correction in radionuclide computed tomography. *IEEE Trans. Nuc. Sci.*, 25(1):638–643, February 1978.

[42] H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers. Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance. *J. Opt. Soc. Am. A*, 12(5):834–52, May 1995.

[43] S. R. Meikle, M. Dahlbom, and S. R. Cherry. Attenuation correction using count-limited transmission data in positron emission tomography. *J. Nuc. Med.*, 34(1):143–50, January 1993.

[44] J. A. Fessler. Resolution properties of regularized image reconstruction methods. Technical Report 297, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, August 1995.

[45] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. IEEE Press, New York, 1988.

[46] B. C. Penney, M. A. King, and R. B. Schwinger. Constrained least-squares restoration of nuclear medicine images: selecting the coarseness function. *Med. Phys.*, 14(5):849–58, September 1987.

[47] R. Bracewell. *The Fourier transform and its applications*. McGraw-Hill, New York, 1978.

[48] R. H. Huesman. A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography. *Phys. Med. Biol.*, 29(5):543–52, May 1984.

[49] R. H. Huesman and B. M. Mazoyer. Kinetic data analysis with a noisy input function. *Phys. Med. Biol.*, 32(12):1569–79, December 1987.

[50] K. J. Worsley, A. C. Evans, S. Marrett, and P. Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cerebral Blood Flow and Metabolism*, 12(6):900–18, 1992.

[51] J. B. Poline and B. M. Mazoyer. Analysis of individual brain activation maps using hierarchical description and multiscale detection. *IEEE Trans. Med. Imag.*, 13(4):702–10, December 1994.

[52] F. O'Sullivan. A study of least squares and maximum likelihood for image reconstruction in positron emission tomography. Technical Report 258, Dept. of Stat., Univ. of Washington, July 1993.

[53] J. A. Fessler. Tomographic reconstruction using information weighted smoothing splines. In H H Barrett and A F Gmitro, editors, *Information Processing in Medical Im.*, volume 687 of *Lecture Notes in Computer Science*, pages 372–86. Springer-Verlag, Berlin, 1993.

[54] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996.

The author's preprints are available on the WWW from URL `http://www.eecs.umich.edu/˜fessler`.
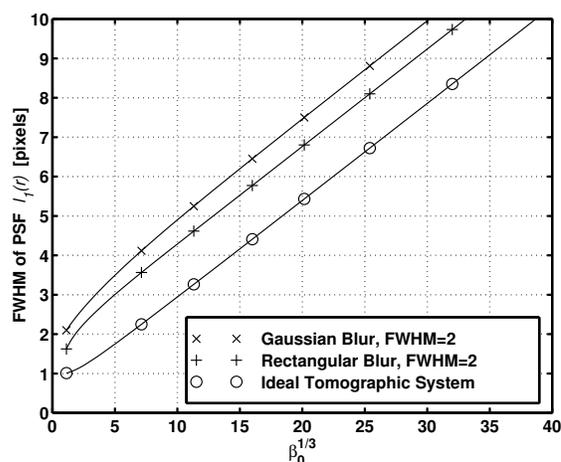
Figure 9: The $\beta_0^{1/3}$ rule of thumb.
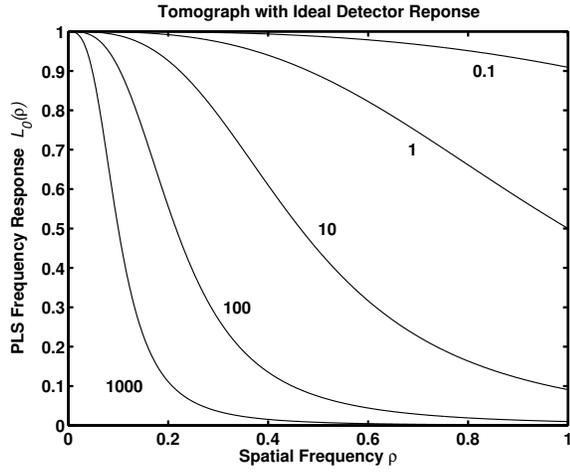
Figure 10: Plots of $L_0(\rho)$, the spatial frequency response of penalized least squares, under the continuous model described in Section 9, for several values (shown) of the regularization parameter β_0, for the ideal tomograph with $S_{\text{true}}(u) = 1$.
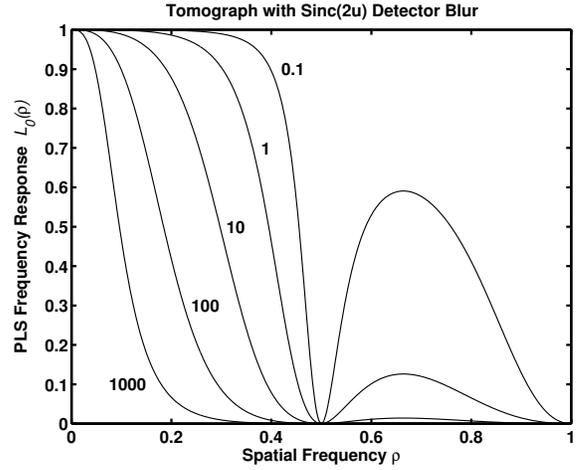
Figure 12: As in the previous figure, except $S_{\text{true}}(u) = \text{sinc}(2u)$.
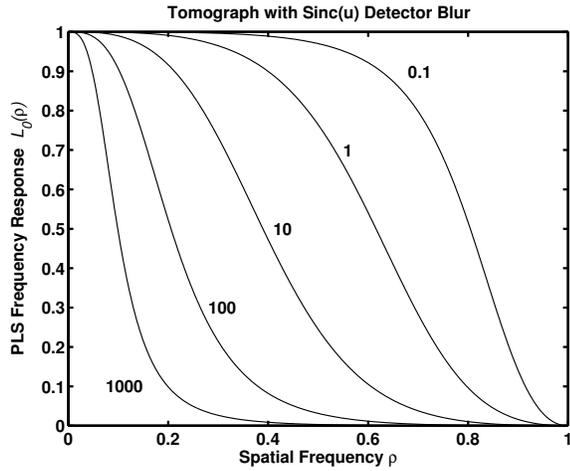
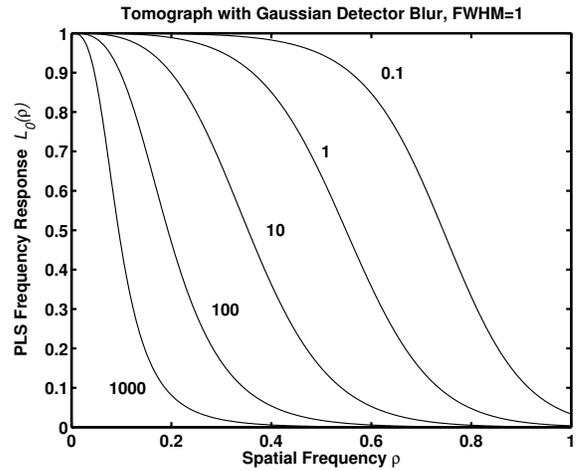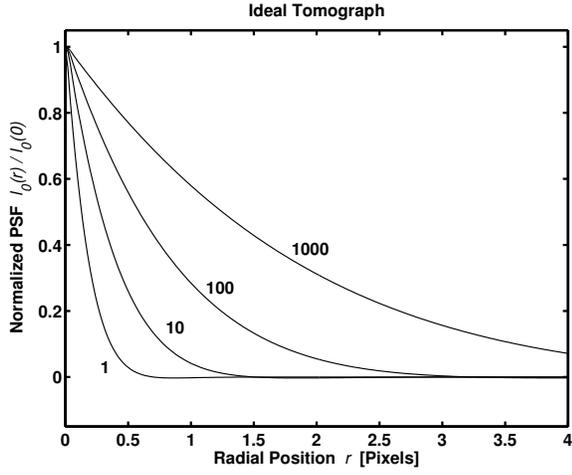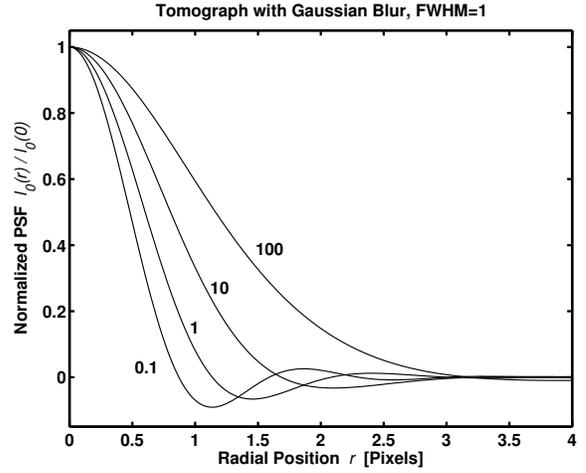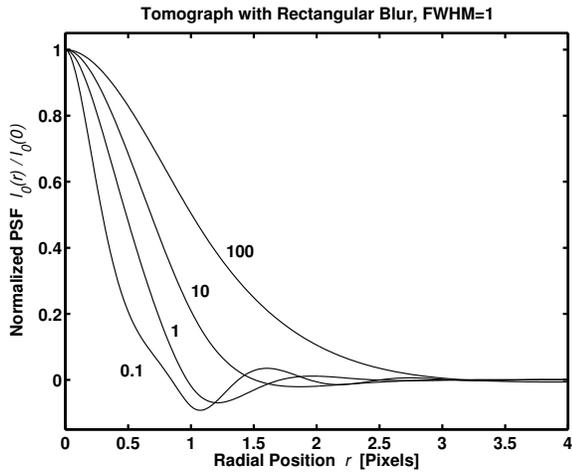Figure 11: As in Fig. 10, but for a tomograph with a rectangular detector blur: $S_{\text{true}}(u) = \text{sinc}(u)$.

Figure 13: As in Fig. 10, but for a tomograph with a Gaussian detector blur: $S_{\text{true}}(u) = \exp(-2\pi^2 u^2/\sigma^2)$, where $\sigma = (8\log 2)^{-1/2}$ so that the FWHM of $s_{\text{true}}(r)$ is approximately 1.

Figure 14: Plots of $l_0(r)/l_0(0)$, the impulse response or PSF of penalized least squares, under the continuous model described in Section 9, for several values (shown) of the regularization parameter β_0, for the ideal case $S_{\text{true}}(u) = 1$.

Figure 15: As in Fig. 14, but for a tomograph with a rectangular detector blur: $S_{\text{true}}(u) = \text{sinc}(u)$.
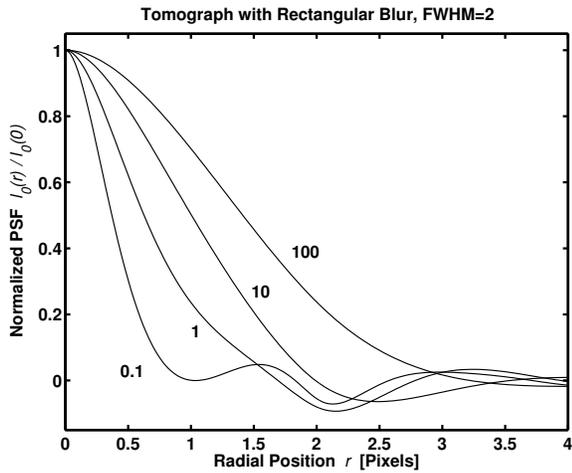
Figure 16: As in the previous figure, except $S_{\text{true}}(u) = \text{sinc}(2u)$.

Figure 17: As in Fig. 14, but for a tomograph with a Gaussian detector blur: $S_{\text{true}}(u) = \exp\left(-2\pi^2 u^2/\sigma^2\right)$.
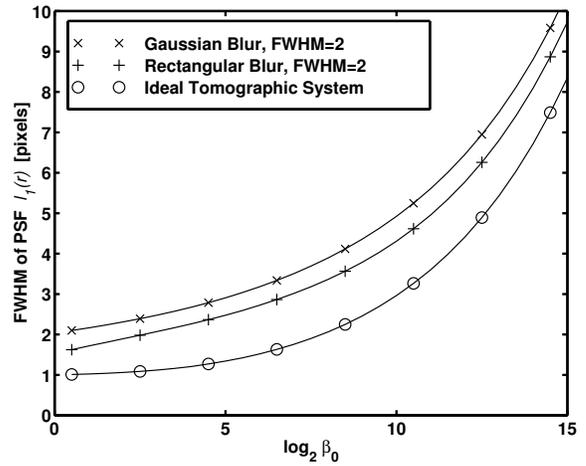
Figure 18: FWHM of the point spread functions $l_1(r)$ corresponding to the $l_0(r)$ shown in Figs. 14-17, as a function of the regularization parameter β_0.

22
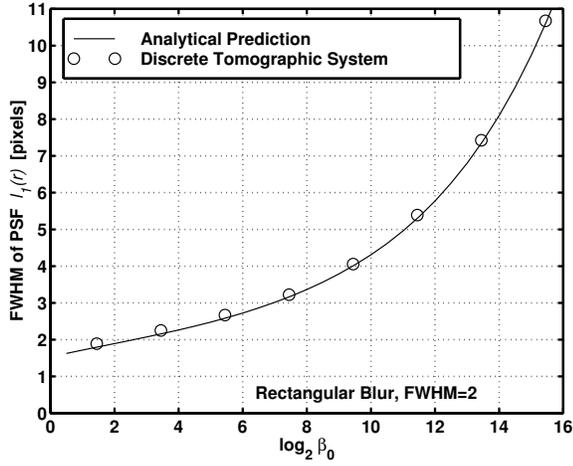
Figure 19: Comparison of the resolution of the analytically computed PSF (51) with the resolution of the discrete PSF ((43)), for a tomograph with rectangular blur: $S_{\text{true}}(u) = \text{sinc}(2u)$.

Figure 20: Comparison of analytically predicted PSF using (51) and discrete PSF from (43) , for a tomograph with rectangular blur: $S_{\text{true}}(u) = \text{sinc}(2u)$.
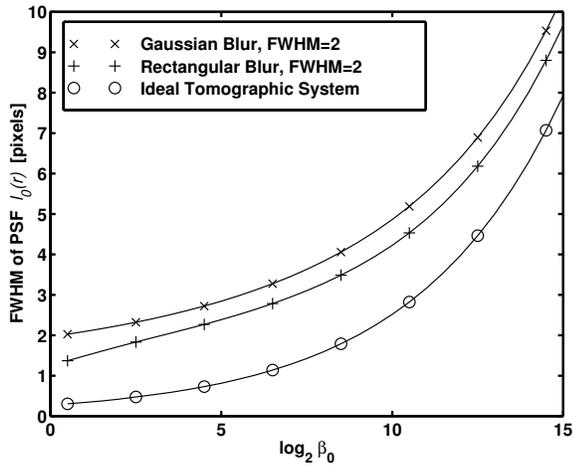

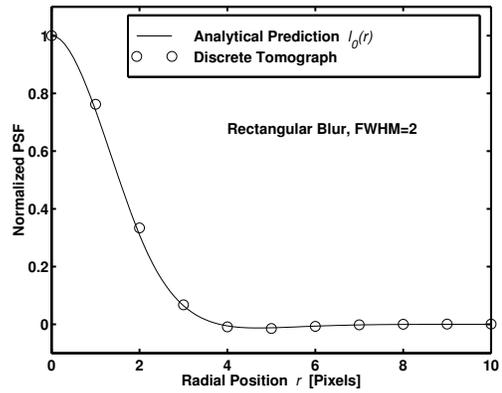Figure 21: FWHM of $l_0(r)$ versus $\log_2(\beta_0)$, for various detector responses.

Figure 22: As in Fig. 19 but for $l_0(r)$.

Figure 23: As in Fig. 20 but for $l_0(r)$

23
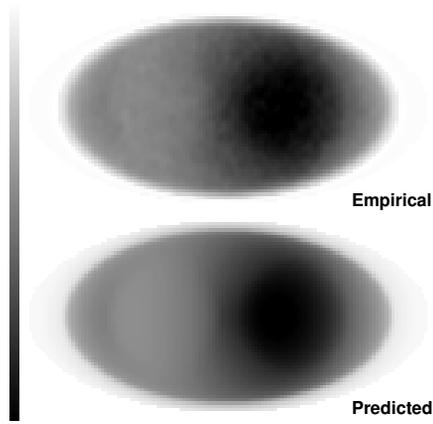
Figure 24: Empirical and predicted standard deviation maps (eqn. (58)) for $\hat{\theta}^{\mathrm{PL}}$: penalized-likelihood emission image reconstruction using the modified quadratic penalty.

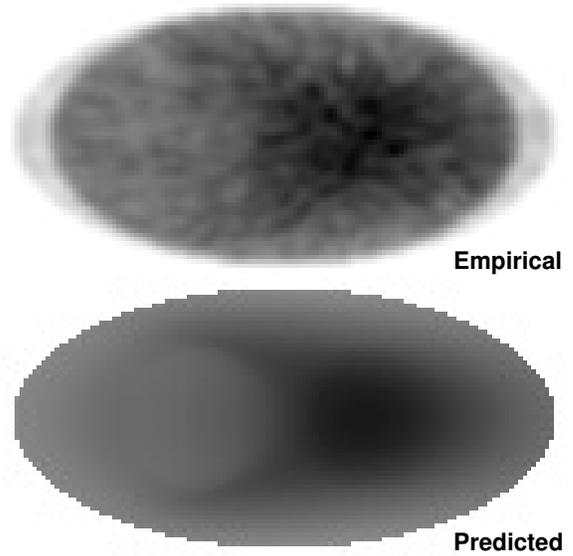Figure 25: Central horizontal profile through Fig. 24.

Figure 26: Empirical and predicted standard deviation maps (eqn. (61)) for $\hat{\theta}^{\mathrm{QPULS}}$: unweighted penalized least-squares estimator.
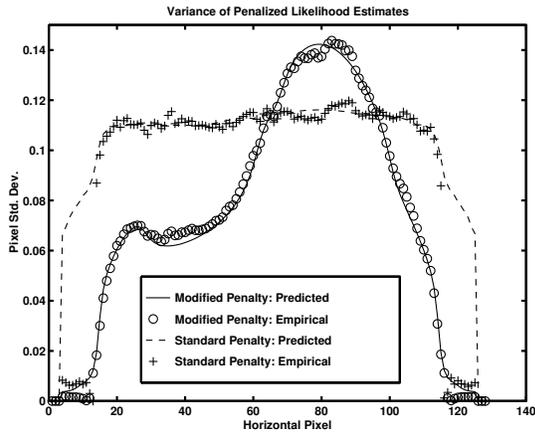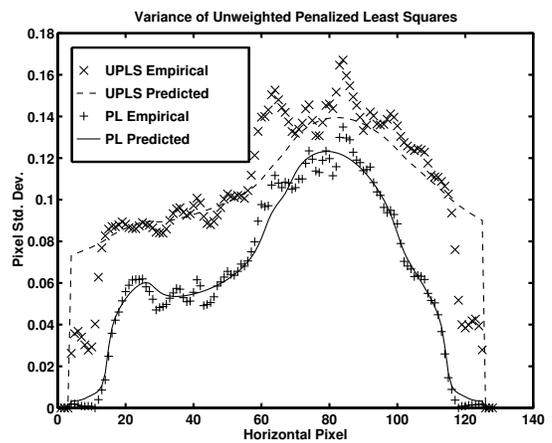

Figure 27: Central horizontal profile through Fig. 26