Variance Images for Penalized-Likelihood Image Reconstruction

Jeffrey A. Fessler
4240 EECS Bldg., University of Michigan, Ann Arbor, MI 48109-2122
Email: fessler@umich.edu, Voice: 734-763-1434, FAX: 734-764-8041

Abstract— **This paper describes new approximations for the variance and covariance of images reconstructed by penalized-likelihood (PL) methods, and presents a simple procedure for computing those approximations. The proposed approximations require less computation than the PL estimates themselves. The method enables the display of variance images, which can provide an indication of uncertainty that may be helpful in medical diagnosis. Using the approximations, we show analytically that penalized-likelihood estimates have lower variance than penalized unweighted least-squares methods at matched spatial resolution, confirming previous empirical comparisons. Simulations of positron emission tomography (PET) scans illustrate the accuracy of the proposed variance approximations in nonzero image pixels.**

I. Introduction

must discuss Qi [1] !!!

When estimating unknown parameters from noisy data, usually one would like to have not only point estimates of the parameters, but also some measure of uncertainty, such as confidence intervals. In conventional tomographic image reconstruction using the filtered backprojection (FBP) method, measures of statistical uncertainty are rarely used or reported (e.g. displayed), except when needed for certain post-processing procedures [2–4]. One possible reason for this absence may be the fact that since FBP is a linear, shift-invariant method, the variances of the reconstructed pixel values are fairly uniform spatially, and hence relatively uninteresting.

Statistical methods for image reconstruction are now being used routinely for certain types of SPECT and PET scans, e.g. [5, 6]. These reconstruction methods are nonlinear and shift-variant, so the resulting image variance can be highly space-variant [7–12]. Thus it may be desirable to be able to augment displays of reconstructed images with displays of variance images to provide an indication of uncertainty to the physician making the diagnostic interpretation. Variance and covariance information should also be useful for purposes such as generalizing Huesman's weighting method [2, 4] from FBP to penalized-likelihood estimators, for searching for statistically significant regions in brain activation studies [13,14], and for statistically-based image post-processing methods [15]. This paper presents a simple method for computing reasonably accurate approximations to the variances of image pixel values in images reconstructed by penalized-likelihood methods.

This work was supported in part by NIH grants CA-60711 and CA-54362 and by the Whitaker Foundation.

The variance calculated here is that due to Poisson noise in the emission sinogram measurements. The analysis could be extended to account for additional variance caused by noisy attenuation correction factors, random coincidences, scatter, etc. Physiological factors such as patient motion or between-subject variability are not considered here, and are probably very difficult to quantify.

In most statistical methods for image reconstruction, one estimates the unknown image parameters $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]'$ from a realization $\boldsymbol{y} = [y_1, \ldots, y_n]'$ of the noisy measurements $\boldsymbol{Y}$ by maximizing an objective function:

$$\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(\boldsymbol{y}) = \arg\max_{\boldsymbol{\lambda} \geq 0} \Phi(\boldsymbol{\lambda}, \boldsymbol{y}). \quad (1)$$

We have previously derived [12] the following general approximation for the covariance of such estimators:

$$\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}\right\} \approx [-\nabla^{20}\Phi(\check{\boldsymbol{\lambda}}, \bar{\boldsymbol{Y}})]^{-1} \, \nabla^{11}\Phi(\check{\boldsymbol{\lambda}}, \bar{\boldsymbol{Y}}) \, \mathrm{Cov}\{\boldsymbol{Y}\} \cdot [\nabla^{11}\Phi(\check{\boldsymbol{\lambda}}, \bar{\boldsymbol{Y}})]' \, [-\nabla^{20}\Phi(\check{\boldsymbol{\lambda}}, \bar{\boldsymbol{Y}})]^{-1}, \quad (2)$$

where "$'$" denotes matrix transpose and

$$\begin{aligned} \bar{\boldsymbol{Y}} &= E_{\boldsymbol{\lambda}^{\mathrm{true}}}[\boldsymbol{Y}] \\ \check{\boldsymbol{\lambda}} &= \hat{\boldsymbol{\lambda}}(\bar{\boldsymbol{Y}}). \end{aligned}$$

Using the methods described in [12], one can use the approximation (2) fairly easily to calculate the variance of selected pixels of interest, e.g.

$$\mathrm{Var}\left\{\hat{\lambda}_j\right\} = [\,\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}\right\}]_{jj} = \boldsymbol{e}_j' \,\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}\right\} \boldsymbol{e}_j$$

for some pixel j, where $\boldsymbol{e}_j$ is the $j$the unit vector of length p. However, for display of a variance image one must compute all p of the diagonal elements of $\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}\right\}$. The methods of [12] require $O(p)$ computation per diagonal element, so the overall computation for a variance image would be $O(p^2)$, which is computationally infeasible for realistic image sizes. In this paper we derive simpler approximations that reduce the required computation to only $O(p)$, provided certain system-dependent factors are precomputed (which only needs to be done once). The resulting computation time is considerably less than that required for computing $\hat{\boldsymbol{\lambda}}$ itself.

We focus on emission tomography, although the methods also apply to transmission scans and to related inverse problems. We assume the usual independent-Poisson statistical model for the emission measurements [16]:

$$Y_i \sim \text{Poisson}\{\bar{Y}_i(\boldsymbol{\lambda}^{\text{true}})\} \quad (3)$$

$$\bar{Y}_i(\boldsymbol{\lambda}) = \sum_j c_i g_{ij} \lambda_j + r_i, \quad (4)$$

where $\boldsymbol{G} = \{g_{ij}\}$ represents the geometric component of the system response [17], the c_i's represent ray-dependent factors such as attenuation and detector efficiency, and the r_i's denote additive background events such as random coincidences. These nonnegative factors are all assumed known.

Under these assumptions, the log-likelihood is

$$L(\boldsymbol{\lambda}, \boldsymbol{y}) = \sum_{i=1}^{n} y_i \log \bar{Y}_i(\boldsymbol{\lambda}) - \bar{Y}_i(\boldsymbol{\lambda}), \quad (5)$$

neglecting constants independent of $\boldsymbol{\lambda}$. We focus on penalized-likelihood image reconstruction methods, for which the objective function has the form:

$$\Phi(\boldsymbol{\lambda}, \boldsymbol{y}) = L(\boldsymbol{\lambda}, \boldsymbol{y}) - \beta R(\boldsymbol{\lambda}), \quad (6)$$

where $R(\boldsymbol{\lambda})$ is a roughness penalty included for regularization. The parameter β controls the resolution/noise tradeoff [17].

The following general form [18] expresses most of the convex penalty functions that have been proposed for regularization of imaging problems:

$$R(\boldsymbol{\lambda}) = \sum_{k=1}^{K} \psi_k([\boldsymbol{C}\boldsymbol{\lambda} - \boldsymbol{c}]_k), \quad (7)$$

where $\boldsymbol{C}$ is a $K \times p$ matrix and $\boldsymbol{c} \in \mathbb{R}^K$, for some user-defined number K of soft "constraints" of the form $[\boldsymbol{C}\boldsymbol{\lambda}]_k \approx c_k$. The standard roughness penalty, which penalizes differences between neighboring pixel values, is the special case of (7) where K is the number of pairs of neighboring pixels [19, 20]. We consider general convex nonquadratic functions ψ_k that are symmetric, twice-differentiable, and that have bounded, nonzero second derivatives. The Hessian (matrix of second partial derivatives) of the penalty function is given by

$$\boldsymbol{R}(\boldsymbol{\lambda}) = \nabla^2 R(\boldsymbol{\lambda}) = \boldsymbol{C}' \boldsymbol{D}_{\ddot{\psi}} \boldsymbol{C}, \quad (8)$$

where $\boldsymbol{D}_{\ddot{\psi}} = \text{diag}\left\{\ddot{\psi}_k([\boldsymbol{C}\boldsymbol{\lambda} - \boldsymbol{c}]_k)\right\}$ and $\ddot{\psi}_k(t) = \frac{d^2}{dt^2}\psi_k(t)$.

For the penalized-likelihood objective function for emission tomography, we derived in [12] from (2) that

$$\text{Cov}\left\{\hat{\boldsymbol{\lambda}}\right\} \approx [\boldsymbol{F} + \beta \boldsymbol{R}(\check{\boldsymbol{\lambda}})]^{-1} \boldsymbol{F} [\boldsymbol{F} + \beta \boldsymbol{R}(\check{\boldsymbol{\lambda}})]^{-1}, \quad (9)$$

where

$$\boldsymbol{F} = \boldsymbol{G}' \,\text{diag}\{u_i\}\, \boldsymbol{G} \quad (10)$$

$$u_i = c_i^2 / \bar{Y}_i(\boldsymbol{\lambda}^{\text{true}}). \quad (11)$$

The matrix $\boldsymbol{F}$ is the Fisher information for estimating $\boldsymbol{\lambda}$ from $\boldsymbol{Y}$ [21].

This paper develops simpler approximations to the covariance (9), and demonstrates that they are usefully accurate. The expressions give further insight into why weighted statistical methods outperform unweighted methods for image reconstruction, as illustrated qualitatively in [22].

The approximations we present are based on (2), which does not account for any nonnegativity constraint used in the reconstruction. Such constraints obviously will lower the variance of the estimator. In low-count regions where the nonnegativity constraint is very active, our approximations will over-estimate the variance. It is generally preferable to err on the side of over-estimating uncertainty. Nevertheless, developing variance approximations that account for the nonnegativity constraint in penalized-likelihood estimators remains a challenging open problem. Some progress has been made for iterative maximum-likelihood algorithms [9, 10].

II. Regularized Least Squares

Before considering the nonlinear Poisson case in Section III, we first consider in this section the simpler case of the linear regularized least-squares (RLS) estimator for the linear measurement model with white noise. The linear model can be expressed:

$$\boldsymbol{z} = \boldsymbol{G}\boldsymbol{\lambda}^{\text{true}} + \boldsymbol{\varepsilon},$$

where the noise covariance is the identity matrix: $\text{Cov}\{\boldsymbol{\varepsilon}\} = \boldsymbol{I}$. (This assumption is *not* adequate for nuclear imaging and is only considered in this section for didactic purposes.) The regularized least-squares estimator minimizes the following objective function:

$$\frac{1}{2}\|\boldsymbol{z} - \boldsymbol{G}\boldsymbol{\lambda}\| + \beta \frac{1}{2} \boldsymbol{\lambda}' \boldsymbol{R}_0 \boldsymbol{\lambda}$$

and can be written explicitly as

$$\hat{\boldsymbol{\lambda}}^{\text{RLS}} = [\boldsymbol{G}'\boldsymbol{G} + \beta \boldsymbol{R}_0]^{-1} \boldsymbol{G}' \boldsymbol{z}.$$

If $\mathrm{Cov}\{\boldsymbol{\varepsilon}\} = \boldsymbol{I}$ then the covariance of $\hat{\boldsymbol{\lambda}}^{\mathrm{RLS}}$ is exactly

$$\boldsymbol{K}(\beta) = [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{G}'\boldsymbol{G}[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}. \quad (12)$$

The matrix $\boldsymbol{K}(\beta)$ is central to the approximations developed in the next section. In particular, for this hypothetical scenario, the following term

$$\sigma_j(\beta) = \sqrt{[\boldsymbol{K}(\beta)]_{jj}} = \sqrt{\boldsymbol{e}_j'\boldsymbol{K}(\beta)\boldsymbol{e}_j} \quad (13)$$

would be the standard deviation of $\hat{\boldsymbol{\lambda}}_j^{\mathrm{RLS}}$, where $[\boldsymbol{K}(\beta)]_{jj}$ is the jth diagonal entry of $\boldsymbol{K}(\beta)$.

If $\boldsymbol{G}'\boldsymbol{G}$ and $\boldsymbol{R}_0$ are approximately circulant matrices, i.e., corresponding to a shift-invariant imaging system and penalty function, then two simplifications occur. The first is that $\sigma_j(\beta)$ becomes independent of j, since sums, inverses, and products of circulant matrices are also circulant and hence have constant diagonals. The second is that we can easily compute the diagonal value $\sigma(\beta)$ using Fourier methods. Thus we can precompute and store a table of $(\beta, \sigma(\beta))$ pairs for several values of β. This table needs to be computed only once for a given system $\boldsymbol{G}$ and regularizer $\boldsymbol{R}_0$. Then when a subsequent reconstruction is performed for some new value of β, we could determine the standard deviation of $\hat{\boldsymbol{\lambda}}^{\mathrm{RLS}}$ simply by interpolating the values in the table.

In the circulant case with white noise, a variance image for $\hat{\boldsymbol{\lambda}}^{\mathrm{RLS}}$ is uniform, which is fairly uninteresting. However, even in the noncirculant case we can precompute $\sigma_j(\beta)$ from (13) for all j for a few values of β, essentially creating p tables—one for each pixel j. Then we again can use table lookup and interpolation to form a variance image for any β that is used subsequently. Using fast-converging iterative methods [19,23], one can precompute $\{\sigma_j(\beta)\}_{j=1}^p$ with $O(p^2)$ computation; subsequent table-lookups require only $O(p)$ computation.

In the above linear case, the covariance $\boldsymbol{K}(\beta)$ is independent of the unknown object parameters $\boldsymbol{\lambda}$, which enables precomputing the table. In the next section we derive approximations that also reduce the nonlinear case to a modified form of table lookup.

III. Variance Approximations

This section presents approximations that simplify the covariance (9). The first approximation is one that we have previously used for analyzing the spatial resolution properties of tomographic image reconstruction in [17]. Roughly speaking, this approximation brings the diagonal measurement covariance matrix outside of the Fisher information matrix (10) as follows:

$$\boldsymbol{F} = \boldsymbol{G}' \operatorname{diag}\{u_i\} \boldsymbol{G} \approx \boldsymbol{D}_\kappa \boldsymbol{G}'\boldsymbol{G}\boldsymbol{D}_\kappa, \quad (14)$$

where $\boldsymbol{D}_\kappa = \operatorname{diag}\{\kappa_j\}$ and

$$\kappa_j = \sqrt{\frac{\sum_{i=1}^n g_{ij}^2 u_i}{\sum_{i=1}^n g_{ij}^2}}, \quad j = 1, \ldots, p. \quad (15)$$

We also apply an approximate exchange analogous to (14) to the regularization term in (8):

$$\beta \boldsymbol{D}_\kappa^{-1}\boldsymbol{C}'\boldsymbol{D}_{\ddot{\psi}}\boldsymbol{C}\boldsymbol{D}_\kappa^{-1} \approx \boldsymbol{D}_\eta \boldsymbol{C}'\boldsymbol{C}\boldsymbol{D}_\eta, \quad (16)$$

where $\boldsymbol{D}_\eta = \operatorname{diag}\{\sqrt{\eta_j}\}$ and

$$\eta_j = \frac{\beta}{\kappa_j^2} \frac{\sum_k c_{kj}^2 \ddot{\psi}_k([\boldsymbol{C}\boldsymbol{\lambda}^{\mathrm{true}} - \boldsymbol{c}]_k)}{\sum_k c_{kj}^2}. \quad (17)$$

Substituting the approximations (14) and (16) into (9) and simplifying yields the following intermediate approximation for the covariance of penalized-likelihood estimates:

$$\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}^{\mathrm{PL}}\right\} \approx \boldsymbol{D}_\kappa^{-1}\boldsymbol{B}^{-1}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{B}^{-1}\boldsymbol{D}_\kappa^{-1}, \quad (18)$$

where

$$\boldsymbol{B} = \boldsymbol{G}'\boldsymbol{G} + \boldsymbol{D}_\eta \boldsymbol{C}'\boldsymbol{C}\boldsymbol{D}_\eta.$$

Thus for the diagonal terms we have

$$\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{PL}}\right\} \approx \frac{1}{\kappa_j^2}\boldsymbol{e}_j'\boldsymbol{B}^{-1}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{B}^{-1}\boldsymbol{e}_j. \quad (19)$$

Our third approximation uses the heuristic that $\boldsymbol{B}^{-1}\boldsymbol{e}_j$ is determined primarily by the value η_j, since typically the η_j's vary slowly spatially. A convenient mathematical expression for this property is the following [19]:

$$\begin{aligned} \boldsymbol{B}^{-1}\boldsymbol{e}_j &= [\boldsymbol{G}'\boldsymbol{G} + \boldsymbol{D}_\eta\boldsymbol{C}'\boldsymbol{C}\boldsymbol{D}_\eta]^{-1}\boldsymbol{e}_j \\ &\approx [\boldsymbol{G}'\boldsymbol{G} + \eta_j\boldsymbol{C}'\boldsymbol{C}]^{-1}\boldsymbol{e}_j. \end{aligned} \quad (20)$$

Thus by identifying $\boldsymbol{R}_0 = \boldsymbol{C}'\boldsymbol{C}$ in (12) we have:

$$\boldsymbol{e}_j'\boldsymbol{B}^{-1}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{B}^{-1}\boldsymbol{e}_j \approx \boldsymbol{e}_j'\boldsymbol{K}(\eta_j)\boldsymbol{e}_j.$$

Substituting this approximation into (19) and simplifying yields:

$$\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{PL}}\right\} \approx \frac{1}{\kappa_j^2}\boldsymbol{e}_j'\boldsymbol{K}(\eta_j)\boldsymbol{e}_j,$$

or from the definition of σ_j in (13):

$$\sqrt{\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{PL}}\right\}} \approx \frac{\sigma_j(\eta_j)}{\kappa_j}. \quad (21)$$

Thus we have the following simple recipe for computing approximate variance images for penalized-likelihood

estimates. First, precompute $\sigma_j(\beta)$ as described in Section II for each pixel j of potential interest and for a range of β values using $\boldsymbol{R}_0 = \boldsymbol{C}'\boldsymbol{C}$. Then for a given reconstruction problem, compute[1] κ_j and η_j (this requires a modified backprojection in (15)). Finally, substitute the κ_j's and η_j's into (21) to compute the variance (or standard deviation) for each pixel.

The essence of what we have done is to make approximations that separate the system component[2] (into the σ_j expressions which can be precomputed) from the object-dependent component—the κ_j's and η_j's. Because $\boldsymbol{G}$ and $\boldsymbol{C}$ are sparse matrices, calculating the κ_j's and η_j's and substituting into (21) requires only $O(p)$ computation (a single modified backprojection and one table lookup per pixel).

A. Quadratic Penalties

The expression simplifies somewhat for quadratic penalties with $\psi_k(t) = t^2/2$. In this case $\ddot{\psi}_k(t) = 1$, so substituting into (17) yields

$$\eta_j = \beta/\kappa_j^2.$$

Substituting into (21) yields:

$$\boxed{\sqrt{\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{PL}}\right\}} \approx \frac{\sigma_j(\beta/\kappa_j^2)}{\kappa_j},} \tag{22}$$

so we only need to compute the κ_j's to form a variance image. We focus on this case hereafter.

B. A Modified Penalty

In [17] we proposed a modified quadratic penalty function that leads to more uniform spatial resolution than conventional penalty functions. This modification is essentially equivalent to choosing

$$\boldsymbol{C} = \boldsymbol{C}_0\boldsymbol{D}_\kappa, \tag{23}$$

where $\boldsymbol{C}_0$ corresponds to the usual shift-invariant penalty neighborhood structure for which $\boldsymbol{C}_0'\boldsymbol{C}_0$ is nearly circulant. This modified penalty also leads to an even simpler approximation for the estimator covariance. Substituting in (23) and (14) into (9) and simplifying yields:

$$\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}^{\mathrm{PL}}\right\} \approx \boldsymbol{D}_\kappa^{-1}\boldsymbol{K}(\beta)\boldsymbol{D}_\kappa^{-1}, \tag{24}$$

[1]The expressions for both κ_j and η_j depend on the unknown parameter $\boldsymbol{\lambda}^{\mathrm{true}}$. We have found previously that simply "plugging in" the estimate $\hat{\boldsymbol{\lambda}}$ for $\boldsymbol{\lambda}^{\mathrm{true}}$ works quite well in such expressions [12], so we adopt that approach here as well.

[2]In SPECT, the system matrix $\boldsymbol{G}$ also depends on the attenuation properties of the object. So the method as described is more readily applicable to PET than SPECT.

provided we identify $\boldsymbol{R}_0 = \boldsymbol{C}_0'\boldsymbol{C}_0$ in the definition of $\boldsymbol{K}(\beta)$ in (12). We thus have the following approximation to the standard deviation of the penalized-likelihood estimate $\hat{\boldsymbol{\lambda}}_j^{\mathrm{PL}}$ for the modified penalty:

$$\boxed{\sqrt{\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{PL}}\right\}} \approx \frac{\sigma_j(\beta)}{\kappa_j}.} \tag{25}$$

If the system is shift-invariant so that $\boldsymbol{G}'\boldsymbol{G}$ is also approximately circulant, then the numerator $\sigma_j(\beta)$ is a constant independent of j, so the estimator standard deviation is simply inversely proportional to κ_j. (We have previously referred to κ_j as a measure of "certainty" [17].) Using this approximation, we can easily compute a "standard deviation map," where the jth pixel intensity is given by (25).

C. Quadratically Penalized Unweighted Least Squares

FBP is an example of an *unweighted* estimator: all measurements are treated equally [24]. FBP with a ramp filter is closely related to an unweighted least-squares estimate [25], at least in the special case when the system matrix $\boldsymbol{G}$ is the Radon transform (i.e. no system blur). Furthermore, FBP with certain apodizing windows is essentially equivalent to using a quadratically penalized unweighted least-squares (QPULS) estimator [26] (without the nonnegativity constraint). We now derive an approximate expression for the covariance of the QPULS estimator; which in turn approximates the covariance of FBP images reconstructed with the appropriate window [26]. The QPULS estimator is defined by

$$\begin{aligned}\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}} &= \arg\min_{\boldsymbol{\lambda}\geq 0} \|\tilde{\boldsymbol{y}} - \boldsymbol{G}\boldsymbol{\lambda}\|^2 + \beta\boldsymbol{\lambda}'\boldsymbol{R}_0\boldsymbol{\lambda} \\ &\approx [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{G}'\tilde{\boldsymbol{y}},\end{aligned}$$

where

$$\tilde{\boldsymbol{y}} = \mathrm{diag}\left\{c_i^{-1}\right\}\boldsymbol{y} - \boldsymbol{r}$$

is a sinogram precorrected for attenuation, randoms, etc. Besides nonnegativity, an additional difference between QPULS and the RLS method described in Section II is that here we account for the Poisson noise of (3) in $\boldsymbol{y}$, whereas in Section II the noise was assumed white. Note that there is no need to use the modified penalty of [17] for this unweighted estimator, since the local impulse response [17] is easily shown to be

$$l^j = [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{e}_j,$$

which is independent of $\boldsymbol{\lambda}$. Furthermore, this local impulse response is the same as that of the penalized-likelihood estimator with the modified penalty (23) [17],

so for the same β, $\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}}$ and $\hat{\boldsymbol{\lambda}}^{\mathrm{PL}}$ have nearly equal spatial resolution. Except for the nonnegativity constraint, the estimator $\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}}$ is linear, so an approximation for its covariance is:

$$\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}}\right\} \approx$$

$$[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{G}'\,\mathrm{diag}\{v_i\}\,\boldsymbol{G}[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1},$$

where $v_i = u_i^{-1}$ is the variance of $\tilde{y}_i$, and u_i was defined in (11). An approximation analogous to (14) is:

$$\boldsymbol{G}'\,\mathrm{diag}\{v_i\}\,\boldsymbol{G} \approx \boldsymbol{D}_\chi \boldsymbol{G}'\boldsymbol{G}\boldsymbol{D}_\chi \tag{26}$$

where $\boldsymbol{D}_\chi = \mathrm{diag}\{\chi_j\}$ is diagonal with entries

$$\chi_j = \sqrt{\frac{\sum_{i=1}^n g_{ij}^2 v_i}{\sum_{i=1}^n g_{ij}^2}}, \quad j = 1, \ldots, p. \tag{27}$$

The expression for χ_j^2 is closely related to the definition of "mean effective number of counts" given in equation (41) of [27]. The above approximation leads to the following covariance approximation:

$$\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}}\right\} \approx$$

$$[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{D}_\chi\boldsymbol{G}'\boldsymbol{G}\boldsymbol{D}_\chi[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}. \tag{28}$$

If $\boldsymbol{D}_\chi$ were a scaled identity matrix, it would exactly commute with $[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}$. Since the diagonal elements of $\boldsymbol{D}_\chi$ vary slowly spatially, the matrices $\boldsymbol{D}_\chi$ and $[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}$ approximately commute. Figure 1 illustrates that $[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{e}_j$ is well localized near pixel j, so the following approximation is reasonable:

$$\boldsymbol{D}_\chi[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{e}_j \approx [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{e}_j\chi_j. \tag{29}$$

It follows from (29) that

$$\boldsymbol{D}_\chi[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1} \approx [\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{D}_\chi, \tag{30}$$

which, substituted into (28), leads to the following approximation for the covariance of the QPULS estimator:

$$\mathrm{Cov}\left\{\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}}\right\} \approx \boldsymbol{D}_\chi\boldsymbol{K}(\beta)\boldsymbol{D}_\chi, \tag{31}$$

where $\boldsymbol{K}(\beta)$ was defined by (12). Thus[3] by (13)

$$\boxed{\sqrt{\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{QPULS}}\right\}} \approx \chi_j\sigma_j(\beta).} \tag{32}$$

The results in Section IV demonstrate the accuracy of the approximation (32).

[3]In the unregularized case where $\beta = 0$, one can relate (32) to the variance bound given by equations (39) and (45) of [27].

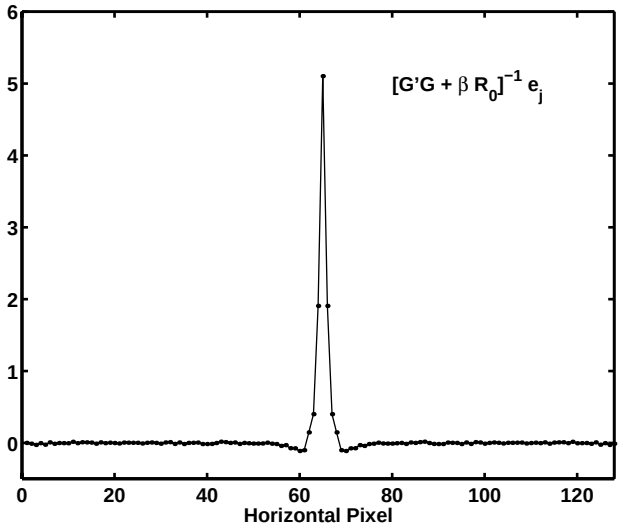

Fig. 1. Typical horizontal profile through the image represented by $[\boldsymbol{G}'\boldsymbol{G} + \beta\boldsymbol{R}_0]^{-1}\boldsymbol{e}_j$, in this case for $\beta = 2^{-4}$, for the pixel j at the image center.

D. Comparison of PL with QPULS

We have previously reported several empirical comparisons that demonstrate that statistical image reconstruction methods lead to lower variance images than FBP when spatial resolution is matched [12, 17, 22, 28, 29]. Using the variance approximations developed above, we now present an *analytical* comparison of penalized-likelihood with penalized unweighted least-squares (which is very closely related to FBP as discussed above). This is the first such comparison that we are aware of. Considering any value of β and any pixel j, then from (25), (14), (32), and (27), we have:

$$\begin{aligned}
\frac{\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{PL}}\right\}}{\sigma_j(\beta)} &\approx \frac{\sum_{i=1}^n g_{ij}^2}{\sum_{i=1}^n g_{ij}^2 u_i} = \frac{1}{\sum_{i=1}^n h_i u_i} \\
\frac{\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{QPULS}}\right\}}{\sigma_j(\beta)} &\approx \frac{\sum_{i=1}^n g_{ij}^2 v_i}{\sum_{i=1}^n g_{ij}^2} = \sum_{i=1}^n h_i v_i,
\end{aligned}$$

where $h_i = g_{ij}^2 / \sum_{i'} g_{i'j}^2$. Since $h_i \geq 0$ and $\sum_{i=1}^n h_i = 1$, it follows from Jensen's inequality that

$$\mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{PL}}\right\} < \mathrm{Var}\left\{\hat{\boldsymbol{\lambda}}_j^{\mathrm{QPULS}}\right\}, \tag{33}$$

with equality if and only if all the v_i's are equal to the same value, i.e., if and only if the measurements are homoscedastic. Homoscedasticity never occurs for Poisson measurements in tomography, so the PL estimator will always have smaller variance than the QPULS estimator at equivalent spatial resolutions. The results in Section IV

confirm the advantage of penalized-likelihood estimation predicted by (33).

The analytical result (33), though approximate, adds further evidence of the importance of using weighting for heteroscedastic measurements, either explicitly as in penalized weighted least-squares estimators [22, 24], 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 or attenuation has the effect of making likelihood-based estimators more like an unweighted estimator, thereby destroying some of the benefits of using statistical methods. Furthermore, although we have focussed on quadratic penalties in deriving (33), similar analysis for the case of nonquadratic (convex) penalty functions leads to the same conclusions.

IV. Simulation Results

To evaluate the variance approximations (22), (25), and (32), we simulated 2000 realizations of PET emission scans using the emission phantom shown in Fig. 2. The simulation parameters are identical to those discussed in [17], so we refer the reader to Section VII of that paper for details, except that for better realism we also simulated nonuniform detector efficiencies according to a log-normal distribution with standard deviation 0.3. From each of the 2000 sinogram measurement realizations we computed $\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}}$ for the standard quadratic penalty and $\hat{\boldsymbol{\lambda}}^{\mathrm{PL}}$ for both the standard and the modified quadratic penalty of [17]. We computed $\hat{\boldsymbol{\lambda}}^{\mathrm{PL}}$ using 10 iterations of the space-alternating generalized expectation-maximization (PML-SAGE-3) algorithm of [30] and $\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}}$ using 10 iterations of the successive overrelaxation algorithm of [22]. Both algorithms enforced the nonnegativity constraint, so some discrepancy between predicted and empirical results is expected.

From each set of 2000 realizations we computed the sample standard deviations pixel-by-pixel. These empirical standard deviations are displayed as images in Fig. 3 and Fig. 6. (The standard deviation images for $\hat{\boldsymbol{\lambda}}^{\mathrm{PL}}$ with the standard penalty were very uniform in this case, so are not shown.) Fig. 3 and Fig. 6 also show the *predicted* standard deviations computed from (25) and (32). Qualitatively the agreement is very good within the object's interior.

Figures 4, 5, 7 and 8 display profiles through the standard deviation maps shown in Fig. 3 and Fig. 6, and through the map (not shown) for PL with the standard penalty. Despite our use of several approximations that might seem fairly crude, these (representative) predicted and empirical results agree remarkably well. The largest disparity is in the background region outside of the object and in the cold disk. In these regions the true pixel values are zero, so the empirical standard deviations are very small due to the nonnegativity constraint, whereas the predicted standard deviations are somewhat higher. But over the nonzero interior of the object the agreement is quite good.

Figures 7 and 8 compare profiles through the standard deviation maps for the penalized-likelihood images reconstructed with the modified quadratic penalty and for unweighted penalized least-squares estimates. The spatial resolution of the two methods was matched using the approach described in [17]. Figures 7 and 8 also display a profile through the standard deviation map for FBP reconstructions of the 2000 realizations using the filter described in [17, 26] to match spatial resolution. (We also used these FBP images to initialize the iterative algorithms.) Note that FBP and QPULS have essentially the same variance, except outside of the object where the nonnegativity constraint is active for QPULS. More importantly, the variances of the PL estimates *are significantly lower* than those of the QPULS (and FBP) estimates, as predicted by our analysis (33). The standard deviations of QPULS average about 30-40% higher than for PL, which in terms of scan time or injected dose is quite significant.

The above comparisons focused on the pixel variance, i.e. the diagonal elements of the image estimate covariance matrix. To evaluate the predicted covariance given by (24), we computed the autocorrelation function for the pixel at the image center for both the PL and QPULS estimators, i.e., the jth column of the covariance matrix, where j is the index of the center pixel. The 2D autocorrelation functions turned out to be asymmetric, considerably more so than predicted by (24). This is unsurprising given the analysis of nonuniform and asymmetric spatial resolution described in [17]. As a more modest comparison, we *averaged* the autocorrelation functions radially, which reduced them to 1D curves (and averaged out the asymmetries). Figure 9 compares the empirical autocorrelation functions computed from the 2000 measurement realizations with the autocorrelation function predicted from (24). In terms of these radial averages, the agreement is again quite good given the number of approximations involved. However, further work is needed to accurately represent the asymmetries that are present in the true covariance (and in the initial covariance approximation (9)) but that are diminished in the approximations leading to (24).

Computing each realization of $\hat{\boldsymbol{\lambda}}^{\mathrm{PL}}$ required approximately 3.0 CPU seconds on a DEC AlphaStation 600-5/333, whereas calculating (25) for all pixels required less than 1.0 CPU second. Thus the computation time is clinically feasible.

V. Discussion

The principle contribution of this paper is the new approximations for variance given by (22), (25), and (32), and the covariance approximations (24) and (31). The empirical results from simulated PET scans demonstrate that these approximations are usefully accurate in nonzero pixels. (We would expect that $E[\hat{\lambda}_j]$ should be one or two standard deviations higher than zero for the approximations to be accurate). These approximations are easily calculated so provide a practical means to display variance images for indicating uncertainty.

In the nuclear imaging community, there sometimes a tendency to lump all "iterative" algorithms together, as if it were the iteration itself that leads to improved image quality. Our analytical and empirical comparison of penalized-likelihood with penalized unweighted least-squares estimators, both of which are maximized by iterative algorithms, clearly demonstrates that it is the objective function that determines image quality, and the iterative algorithm is only a method for maximizing that objective function. Objective functions that are based on accurate statistical models lead to lower variance estimates, as we have shown both analytically (33) and with empirical results.

The proposed approximations over-estimate the variance in low-count regions where the nonnegativity constraint is active in a significant fraction of the realizations. We have confirmed numerically (results not shown) that this inaccuracy is due to the *initial* covariance approximation (2), as opposed to the subsequent approximations presented in this paper. Interestingly, in our results the effects of the nonnegativity constraint appear to be predominately local, i.e., even though there were many zero-valued pixels surrounding the object and in the cold disk within the object, the variance approximations were very accurate in all pixels that are a few pixels away from all such zero-valued pixels.

The foundation for this work is the covariance formula (2), which is applicable to estimators formed by maximizing an objective function. Thus this work is inapplicable to iterative algorithms such as ordered-subsets expectation-maximization (OS-EM) [31], since unregularized OS-EM is never run to convergence. The author considers this to be less a limitation of the analysis method than an inherent shortcoming of the use of unregularized iterations with ad-hoc stopping rules. It is inherently more difficult to analyze properties such as convergence, spatial resolution, and noise for unregularized methods, so the behavior of such methods is harder to predict. Caution would seem advisable in the clinical use of unpredictable methods.

Future work includes possibly improving the approximations to account for asymmetries in the 2D autocorrelation functions (perhaps in conjunction with different penalty functions that reduce those asymmetries), and to apply the methods in generalizing Huesman's weighting method [2, 4] from FBP to penalized-likelihood estimators.

REFERENCES

[1] J. Qi and R. M. Leahy, "Fast computation of the covariance of MAP reconstsructions of PET images," in *Proc. SPIE ?, Med. Im. 1999: Im. Proc. ??*, 1999.

[2] R. H. Huesman, "A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography," *Phys. Med. Biol.*, vol. 29, no. 5, pp. 543–552, 1984.

[3] R. Maitra and F. O'Sullivan, "Estimating the variability of reconstructed PET data: a technique based on approximation the reconstruction filter by a sum of Gaussian kernels," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1995.

[4] R. H. Huesman and B. M. Mazoyer, "Kinetic data analysis with a noisy input function," *Phys. Med. Biol.*, vol. 32, no. 12, pp. 1569–1579, 1987.

[5] E. P. Ficaro, J. A. Fessler, P. D. Shreve, J. N. Kritzman, P. A. Rose, and J. R. Corbett, "Simultaneous transmission/emission myocardial perfusion tomography: Diagnostic accuracy of attenuation-corrected 99m-Tc-Sestamibi SPECT," *Circulation*, vol. 93, no. 3, pp. 463–73, February 1996.

[6] S. Meikle et al. *Information at World Wide Web address* `http://www.cs.nsw.gov.au/rpa/pet/`, 1997.

[7] D. W. Wilson and B. M. W. Tsui, "Noise properties of filtered-backprojection and ML-EM reconstructed emission tomographic images," *IEEE Tr. Nuc. Sci.*, vol. 40, no. 4, pp. 1198–1203, August 1993.

[8] Z. Zhou, R. M. Leahy, and E. U. Mumcuoglu, "A comparative study of the effect of using anatomical priors in PET reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pp. 1749–53, 1993.

[9] H. H. Barrett, D. W. Wilson, and B. M. W. Tsui, "Noise properties of the EM algorithm: I. Theory," *Phys. Med. Biol.*, vol. 39, pp. 833–846, 1994.

[10] D. W. Wilson, B. M. W. Tsui, and H. H. Barrett, "Noise properties of the EM algorithm: II. Monte Carlo simulations," *Phys. Med. Biol.*, vol. 39, pp. 847–872, 1994.

[11] S.-J. Lee, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Tr. Med. Im.*, vol. 14, no. 4, pp. 669–80, December 1995.

[12] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Tr. Im. Proc.*, vol. 5, no. 3, pp. 493–506, March 1996.

[13] 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*, vol. 12, no. 6, pp. 900–918, 1992.

[14] J. B. Poline and B. M. Mazoyer, "Analysis of individual brain activation maps using hierarchical description and multiscale detection," *IEEE Tr. Med. Im.*, vol. 13, no. 4, pp. 702–710, December 1994.

[15] R. D. Nowak and R. G. Baraniuk, "Wavelet-domain filtering for photon imaging systems," *IEEE Tr. Im. Proc.*, vol. 8, no. 5, pp. 666–78, May 1999.

[16] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, no. 2, pp. 306–316, April 1984.

[17] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Tr. Im. Proc.*, vol. 5, no. 9, pp. 1346–58, September 1996.

[18] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Tr. Med. Im.*, vol. 14, no. 1, pp. 132–137, March 1995.

[19] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Tr. Im. Proc.*, vol. 8, no. 5, pp. 688–99, May 1999.

[20] J. A. Fessler, "Conjugate-gradient preconditioning methods: numerical results," Technical Report 303, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, January 1997. Available from `http://www.eecs.umich.edu/~fessler`.

[21] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Tr. Sig. Proc.*, vol. 44, no. 8, pp. 2026–41, August 1996.

[22] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Tr. Med. Im.*, vol. 13, no. 2, pp. 290–300, June 1994.

[23] A. O. Hero and J. A. Fessler, "A recursive algorithm for computing Cramer-Rao-type bounds on estimator covariance," *IEEE Tr. Info. Theory*, vol. 40, no. 4, pp. 1205–10, July 1994.

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

[25] 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.

[26] 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.

[27] 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*, vol. 12, no. 5, pp. 834–52, May 1995.

[28] J. A. Fessler, "Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1439–50, October 1995.

[29] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Tr. Med. Im.*, vol. 16, no. 2, pp. 166–75, April 1997.

[30] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1417–29, October 1995.

[31] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Tr. Med. Im.*, vol. 13, no. 4, pp. 601–609, December 1994.

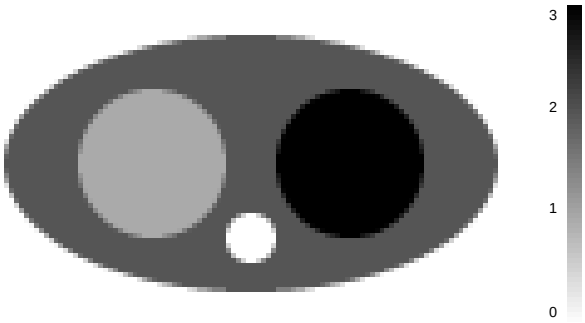

Fig. 2. Emission phantom used for simulations.

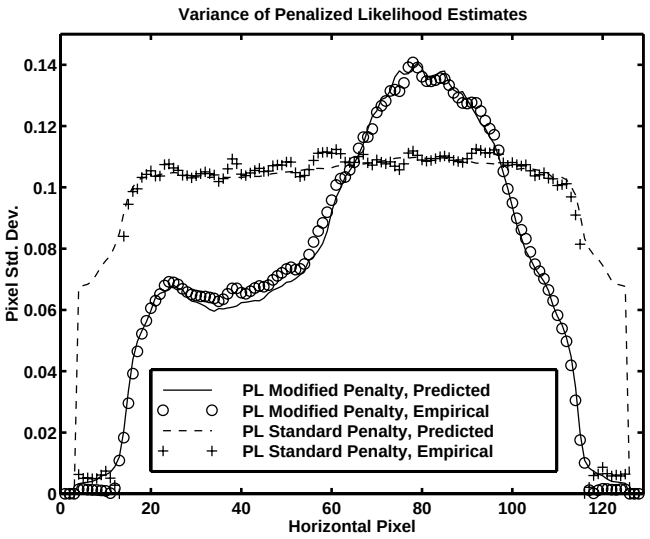

Fig. 4. Central horizontal profiles through Fig. 3, and through corresponding variance image (not shown) for PL with the standard penalty.

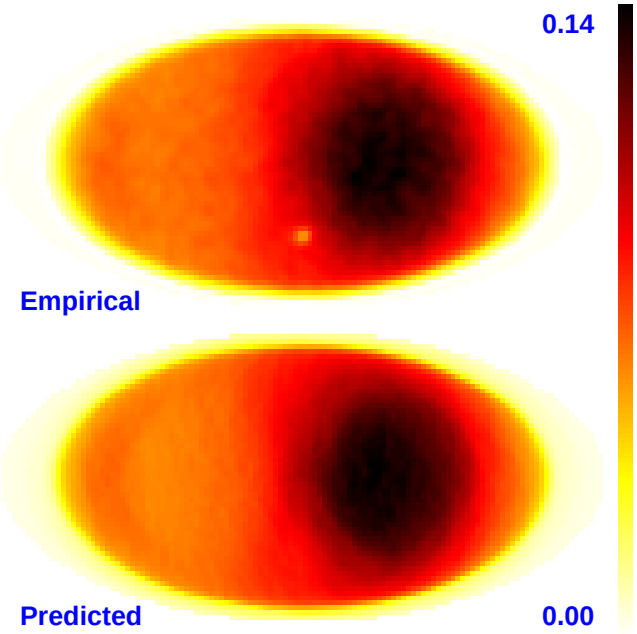

Fig. 3. Empirical and predicted standard deviation maps (eqn. (25)) for $\hat{\boldsymbol{\lambda}}^{\mathrm{PL}}$: penalized-likelihood emission image reconstruction using the modified quadratic penalty.

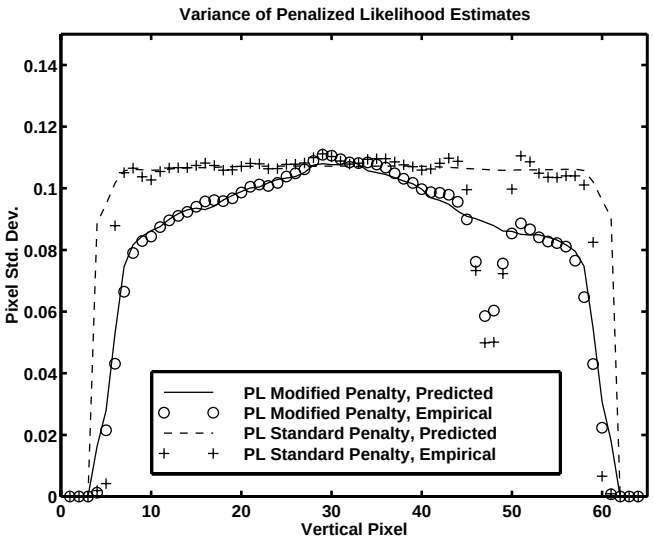

Fig. 5. Central vertical profiles through Fig. 3.

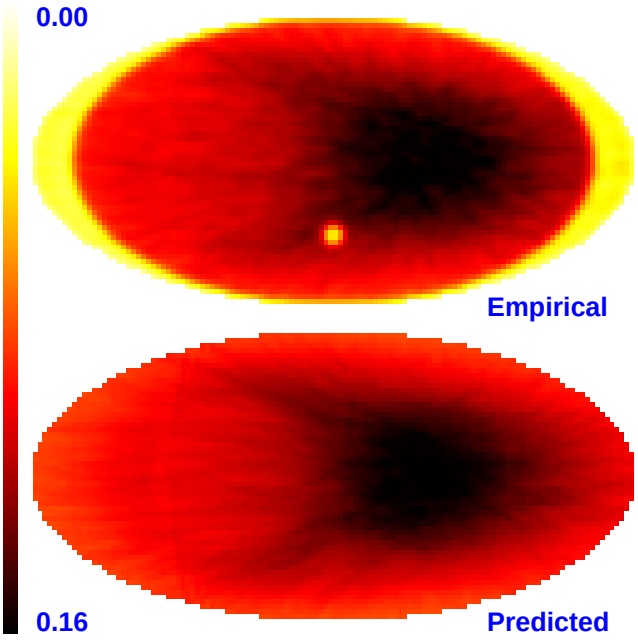

Fig. 6. Empirical and predicted standard deviation maps (eqn. (32)) for $\hat{\boldsymbol{\lambda}}^{\mathrm{QPULS}}$, a penalized unweighted least-squares estimator.

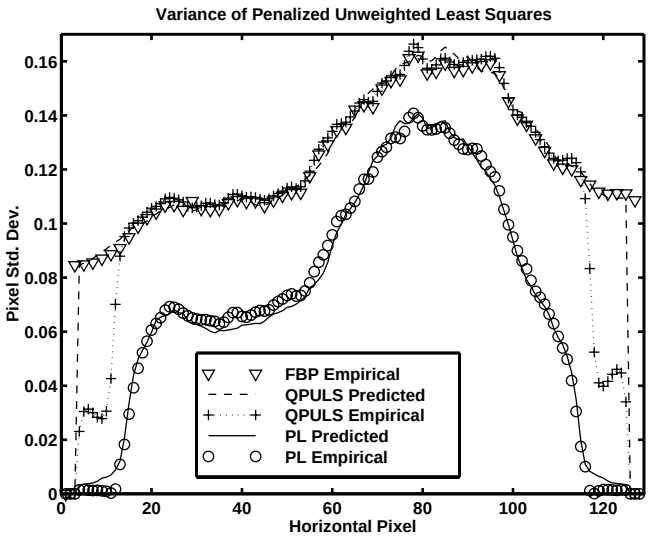

Fig. 7. Central horizontal profiles through Figs. 6 and 3.

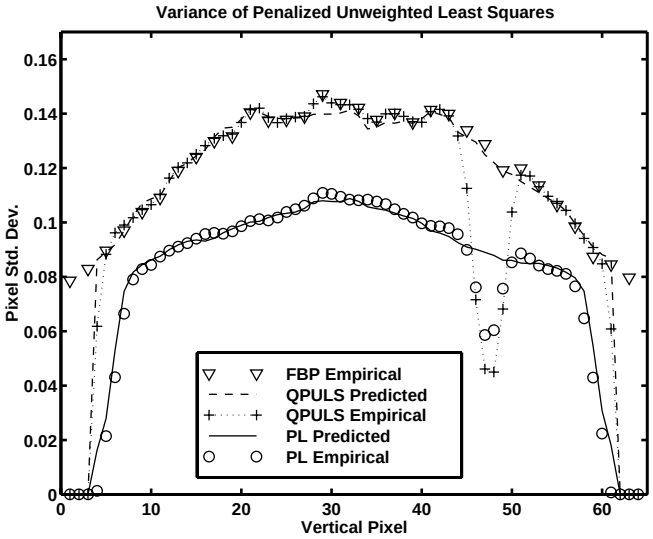

Fig. 8. Central vertical profiles through Figs. 6 and 3.

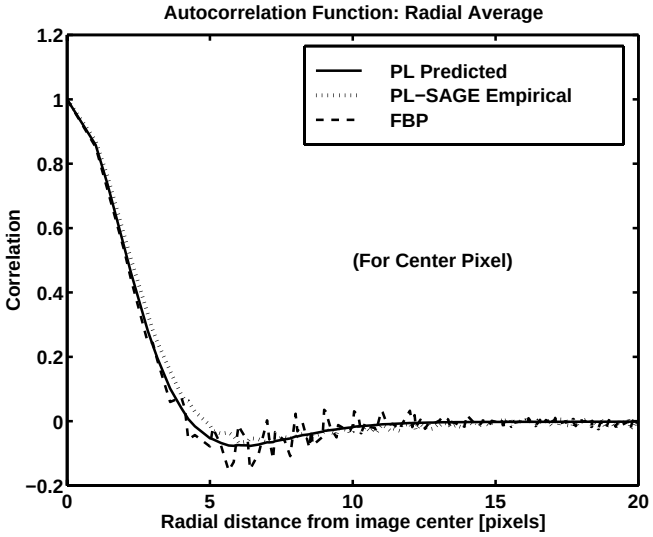

Fig. 9. Radial average of autocorrelation function of penalized-likelihood (PL) and penalized unweighted least-squares (QPULS) estimators for pixel at image center, compared with predicted autocorrelation function (24).