Presented at 1997 IEEE Nuc. Sci. Symp. and Med. Im. Conf.

Approximate Variance Images for Penalized-Likelihood Image Reconstruction

Jeffrey A. Fessler
4240 EECS Bldg., University of Michigan, Ann Arbor, MI 48109-2122

Abstract

We present a fairly simple procedure for computing new approximations for the pixel variances in images reconstructed by penalized-likelihood methods. The method enables the display of variance images, which can provide an indication of uncertainty that may be helpful in medical diagnosis and in evaluation of image reconstruction algorithms. Simulations of positron emission tomography (PET) scans illustrate the accuracy of the proposed variance approximations in nonzero image pixels.

I. INTRODUCTION

Statistical methods for image reconstruction are nonlinear, so the image variance can be space-variant [1]. Thus it may be useful to augment displays of reconstructed images with displays of variance images to provide an indication of uncertainty to the physician making the diagnosis. Variance and covariance information should also be useful for purposes such as generalizing Huesman's weighting method [2, 3] from FBP to penalized-likelihood (PL) estimators, for searching for statistically significant regions in brain activation studies [4, 5], or for evaluating image reconstruction algorithm performance. This paper summarizes a new simple method for computing approximations to the variances of image pixel values in images reconstructed by PL methods.

We assume the usual Poisson statistical model for the emission measurements Y_i [6]:

$$\begin{aligned} Y_i &\sim \text{Poisson}\{\bar{Y}_i(\underline{\lambda}^{\text{true}})\} && (1) \\ \bar{Y}_i(\underline{\lambda}) &\triangleq \sum_j c_i g_{ij} \lambda_j + r_i, && (2) \end{aligned}$$

where λ_j is the unknown activity in the jth pixel, $\boldsymbol{G} = \{g_{ij}\}$ represents the geometric system response [7], the c_i's represent ray-dependent factors such as attenuation and detector efficiency, and the r_i's denote random coincidences and scatter. The log-likelihood is:

$$L(\underline{\lambda}, \underline{y}) = \sum_{i=1}^{n} y_i \log \bar{Y}_i(\underline{\lambda}) - \bar{Y}_i(\underline{\lambda}), \tag{3}$$

neglecting constants independent of $\underline{\lambda} = [\lambda_1 \ \ldots \ \lambda_p]'$, where "$'$" denotes matrix transpose. We focus on the case of penalized-likelihood image reconstruction:

$$\hat{\underline{\lambda}} = \hat{\underline{\lambda}}(\underline{y}) = \arg\max_{\underline{\lambda} \geq 0} \Phi(\underline{\lambda}, \underline{y})$$

$$\Phi(\underline{\lambda}, \underline{y}) = L(\underline{\lambda}, \underline{y}) - \beta R(\underline{\lambda}), \tag{4}$$

where $R(\underline{\lambda})$ is a roughness penalty included for regularization. The resolution/noise tradeoff is controlled by β [7].

II. THEORY

For the objective function (4), we derived in [1] that

$$\text{Cov}\{\hat{\underline{\lambda}}\} \approx [\boldsymbol{F} + \beta \boldsymbol{R}(\check{\underline{\lambda}})]^{-1} \boldsymbol{F} [\boldsymbol{F} + \beta \boldsymbol{R}(\check{\underline{\lambda}})]^{-1}, \tag{5}$$

where $\check{\underline{\lambda}} \triangleq \hat{\underline{\lambda}}(\bar{\underline{Y}})$ is the estimate given noiseless data, $\boldsymbol{F} \triangleq \boldsymbol{G}'\boldsymbol{D}(u_i)\boldsymbol{G}$ is the Fisher information matrix, $u_i \triangleq c_i^2 / \bar{Y}_i$ is the inverse variance for each ray, and $\boldsymbol{R}(\underline{\lambda}) = \nabla^2 R(\underline{\lambda})$ is the Hessian matrix of the penalty function.

Using methods described in [1], one can calculate the variance of selected pixels of interest using (5):

$$\text{Var}\{\hat{\underline{\lambda}}_j\} = [\text{Cov}\{\hat{\underline{\lambda}}\}]_{jj} = \underline{e}_j' \text{Cov}\{\hat{\underline{\lambda}}\} \underline{e}_j$$

for some pixel j, where $\underline{e}_j$ is the jth unit vector of length p. However, for display of a variance image one must compute *all* p of the diagonal elements of $\text{Cov}\{\hat{\underline{\lambda}}\}$. The methods of [1] require $O(p)$ computation per diagonal element, so the overall computation for a variance image would be $O(p^2)$, which is impractical. We summarize here a simpler approximation that reduces the computation to only $O(p)$, provided certain system-dependent factors are precomputed (which is only done once). The resulting computation time is less than is required for computing the estimate $\hat{\underline{\lambda}}$ itself.

Since evaluating (5) would require an impractical amount of computation, we would like an approximation. In [7], we proposed the following approximation to the Fisher information:

$$\boldsymbol{F} \approx \boldsymbol{D}(\kappa_j) \boldsymbol{G}' \boldsymbol{G} \boldsymbol{D}(\kappa_j), \tag{6}$$

where $\boldsymbol{D}(\kappa_j)$ denotes the diagonal matrix with elements κ_j, and where

$$\kappa_j \triangleq \sqrt{\frac{\sum_{i=1}^n g_{ij}^2 u_i}{\sum_{i=1}^n g_{ij}^2}}, \quad j = 1, \ldots, p \tag{7}$$

reflects the effective "certainty" of the jth pixel [7].

Elsewhere we consider general convex penalty functions [8]. Here, we restrict attention to the modified penalty function proposed in [7], defined by

$$R(\underline{\lambda}) \triangleq \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} \kappa_j \kappa_k \frac{1}{2} (\lambda_j - \lambda_k)^2, \tag{8}$$

where κ_j is defined by (7) and where $\mathcal{N}_j$ is the four nearest neighbors of pixel j. As shown in [7], the Hessian of $R(\underline{\lambda})$ is approximately the following:

$$\boldsymbol{R} \triangleq \nabla^2 R(\underline{\lambda}) \approx \boldsymbol{D}(\kappa_j) \boldsymbol{R}_0 \boldsymbol{D}(\kappa_j) \tag{9}$$

where $R_0(\underline{\lambda})$ denotes the standard penalty function:

$$R_0(\underline{\lambda}) \triangleq \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} \frac{1}{2} (\lambda_j - \lambda_k)^2$$

with corresponding Hessian $\boldsymbol{R}_0 \triangleq \nabla^2 R_0(\underline{\lambda})$. (The Hessian $\boldsymbol{R}_0$ is block-Toeplitz with 4's along the diagonal and -1's in the off-diagonal positions corresponding to each pixel's neighbors.)

Substituting the approximations (6) and (9) into (5) and simplifying yields the following approximation for the image covariance:

$$\boxed{\mathrm{Cov}\{\hat{\underline{\lambda}}\} \approx \boldsymbol{D}(\kappa_j^{-1}) \boldsymbol{K}(\beta) \boldsymbol{D}(\kappa_j^{-1})} \tag{10}$$

where

$$\boldsymbol{K}(\beta) \triangleq [\boldsymbol{G}'\boldsymbol{G} + \beta \boldsymbol{R}_0]^{-1} \boldsymbol{G}'\boldsymbol{G} [\boldsymbol{G}'\boldsymbol{G} + \beta \boldsymbol{R}_0]^{-1}. \tag{11}$$

The matrix $\boldsymbol{K}(\beta)$ is the image covariance matrix that one would obtain if we had assumed the white Gaussian noise model

$$\tilde{\underline{y}} = \boldsymbol{G}\underline{\lambda} + \mathrm{noise},$$

and had used the standard quadratically-penalized unweighted least-squares (QPULS) estimator:

$$\hat{\underline{\lambda}}_{\mathrm{QPULS}} \triangleq [\boldsymbol{G}'\boldsymbol{G} + \beta \boldsymbol{R}_0]^{-1} \boldsymbol{G}' \tilde{\underline{y}}.$$

The additional κ_j terms in the proposed covariance approximation (10) account for the nonuniform variance of the Poisson measurements.

In PET, the matrix $\boldsymbol{K}(\beta)$ is independent of all object-dependent factors, so portions of this matrix can be precomputed. In particular, we can precompute and store its diagonal elements $[\boldsymbol{K}(\beta)]_{jj}$:

$$\sigma_j^2(\beta) \triangleq [\boldsymbol{K}(\beta)]_{jj} = \underline{e}_j' \boldsymbol{K}(\beta) \underline{e}_j. \tag{12}$$

We can precompute and store a table of $(\beta, \sigma_j^2(\beta))$ pairs for several values of β. This table needs to be computed only once for a given system $\boldsymbol{G}$ and regularizer $\boldsymbol{R}$. From (11), one can see that to compute $\sigma_j^2(\beta)$, we can first solve the system of equations

$$[\boldsymbol{G}'\boldsymbol{G} + \beta \boldsymbol{R}_0] \underline{x} = \underline{e}_j$$

for $\underline{x}$, and then compute the following weighted sum of the forward projection of $\underline{x}$:

$$\sigma_j^2(\beta) = \sum_i u_i [\boldsymbol{G}\underline{x}]_i^2.$$

For the case of shift-invariant PET systems, the computation is trivial using DFT's and Parseval's theorem.

Combining (12) with (10), we have the following simple variance approximation:

$$\boxed{\mathrm{Var}\{\hat{\lambda}_j\} \approx \frac{\sigma_j^2(\beta)}{\kappa_j^2}.} \tag{13}$$

Thus we have the following simple recipe for computing approximate variance images for PL estimates. First, precompute $\sigma_j^2(\beta)$ in (12) for each pixel j of potential interest and for a range of β values. Then, for a given reconstruction problem, compute κ_j, which requires a modified backprojection in (7), and interpolate the σ_j^2 for the β of interest. Finally, substitute the κ_j's and σ_j's into (13) to compute the approximate variances for each pixel.

Because $\boldsymbol{G}$ is a sparse matrix, calculating the κ_j's and substituting into (13) requires only $O(p)$ computation (one backprojection and one table lookup per pixel).

III. Results

To evaluate the proposed variance approximations we simulated 2000 realizations of PET emission scans using the emission phantom shown in Fig. 1. (The simulation parameters are identical to those discussed in [7].) From each sinogram realization we computed $\hat{\underline{\lambda}}$ using 10 iterations of the PML-SAGE-3 algorithm of [9] and the modified quadratic penalty of [7]. We also computed the PL estimates for the standard quadratic penalty function (which has nonuniform spatial resolution [7]). Both algorithms enforced the nonnegativity constraint.

For each of the reconstructed pixel values, we computed the sample standard deviations across the 2000 realizations, which are displayed as images in Fig. 2, along with the *predicted* standard deviations computed from (13).

Figures 3 and 4 display profiles through the standard deviation maps shown in Fig. 2. These (representative) predicted and empirical results agree quite well in the object interior. The largest disparities are 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 there is good agreement between predicted and empirical results.

We also examined autocorrelation functions to investigate the covariance approximation. The empirical autocorrelation functions were asymmetric, but this asymmetry was absent in the predictions computed from (10). However, the radial average of the predicted and empirical autocorrelation functions agreed fairly well.

Computing each realization of $\underline{\hat{\lambda}}$ required approximately 3.0 CPU seconds on a DEC AlphaStation 600-5/333, whereas calculating (13) for all pixels required less than 1.0 CPU second. Thus the computation time is clinically feasible.

IV. Discussion

We have presented a new approximation (13) to the variances of pixels in images reconstructed by penalized-likelihood methods. We have presented results for a particular quadratic penalty function, but the method generalizes easily to other penalty functions [8]. We also show in [8] that the analytical results give further insight into why weighted statistical methods outperform unweighted methods for image reconstruction, as illustrated qualitatively in [10] and quantitatively in empirical results in [8].

The analytical approximations overestimate the variance in regions where the nonnegativity constraint plays a significant role. The nonnegativity constraint is a severe nonlinearity, whereas the approximations are based on a local linearization of the estimator [1]. Further work on this difficult problem is needed. Our empirical histograms of $\hat{\lambda}_j$ show that the log-normal model proposed in [11,12] for ML-EM does not appear applicable to the penalized-likelihood case. Further work is also needed to develop analytical autocorrelation function approximations that reflect the asymmetric properties of the empirical autocorrelation functions. We have also observed such asymmetries in the autocorrelation functions for FBP reconstructions, so this "problem" is not unique to statistical image reconstruction methods.

V. References

[1] 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.

[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. 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] J. A. Fessler, "Variance images for penalized-likelihood image reconstruction," *IEEE Tr. Med. Im.*, 1997. Submitted.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

See `http://www.eecs.umich.edu/~fessler` for current preprints.

VI. Acknowledgement

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

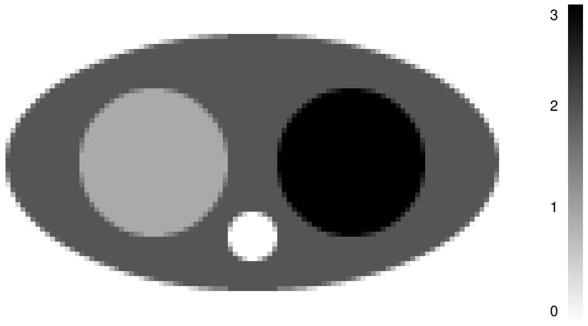

Fig. 1 Emission phantom used for simulations.

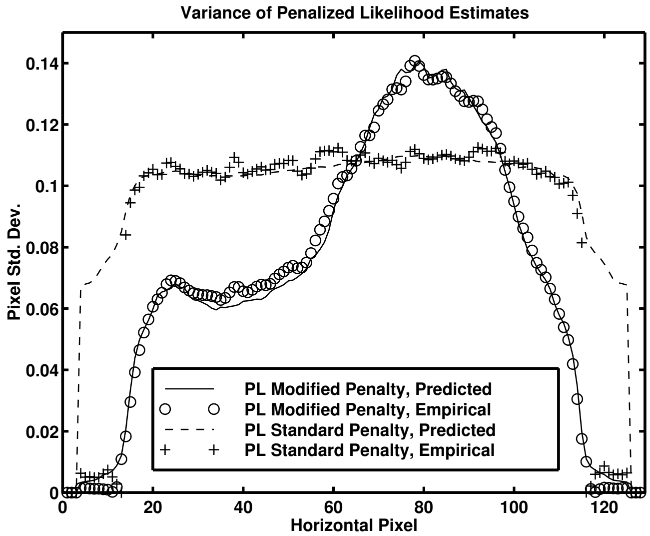

Fig. 3 Central horizontal profiles through Fig. 2, and through corresponding variance image (not shown) for PL with the standard penalty.

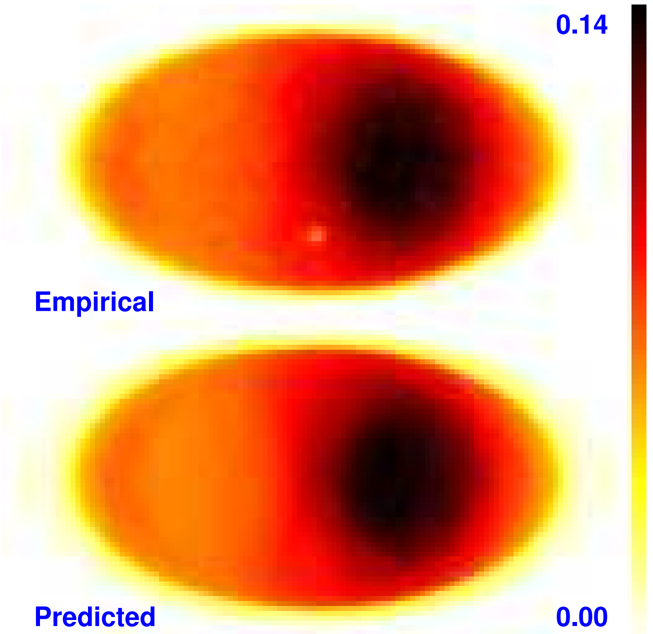

Fig. 2 Empirical and predicted standard deviation maps (eqn. (13)) for $\hat{\underline{\lambda}}$: penalized-likelihood emission image reconstruction using the modified quadratic penalty.

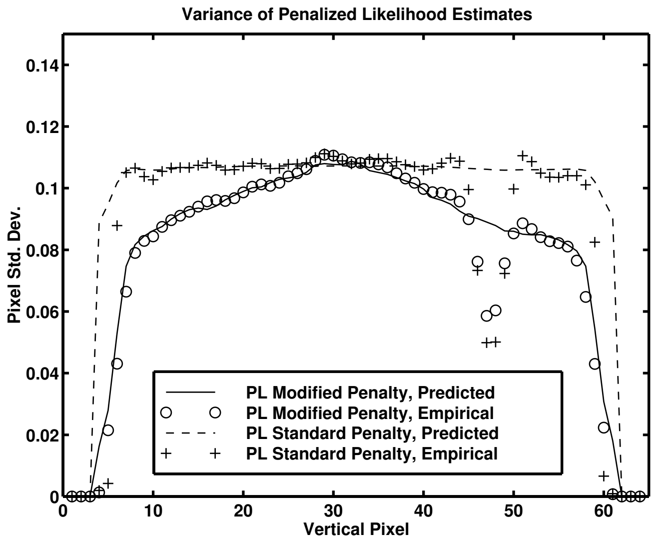

Fig. 4 Central vertical profiles through Fig. 2.