A Smoothing Prior with Embedded Positivity Constraint for Tomographic Reconstruction

Ing-Tsung Hsiao[1], Anand Rangarajan[2], Yuxiang Xing[1], Gene Gindi[1]
[1]Departments of Electrical & Computer Engineering and Radiology,
SUNY Stony Brook, Stony Brook, NY 11784, USA
[2]Department of Computer and Information Science and Engineering,
University of Florida, Gainesville, FL 32611, USA

I. Introduction

Positivity in regularized emission computed tomography (ECT) reconstruction is important for quantitative accuracy, especially for low-count data. However, it is often difficult to impose positivity on the reconstruction without suffering some other drawback, such as speed or lack of analyzability of the algorithm. A general framework for positivity-constrained ECT reconstruction has been the formulation and possibly constrained minimization of an objective function comprising a data penalty (usually log likelihood) and penalty term (a.k.a. "prior" in Bayesian terms). Here we propose to embed a positivity constraint via a novel prior for 3D ECT reconstruction.

The prior generalizes a notion of *I*-divergence proposed by Csiszár [1], and also bears a superficial similarity to formulations previously used in ECT [2], [3], [4]. However, unlike previous formulations, our new formulation also includes a notion of object *smoothness* in addition to positivity. In addition, the priors are convex and that makes the reconstruction independent of initial conditions.

In Sec.II, we describe the new priors mathematically and prove convexity. In Sec.III, we present initial results showing 2D and 3D reconstructions using our new priors, and compare these anecdotally to reconstructions with a conventional smoothing prior. In Sec.IV, we compare our work with other relevant work.

II. Theory

A. Regularized Likelihood Reconstruction

Let $\mathbf{g} = \{g_i; i = 1,..,M\}$ and $\mathbf{f} = \{f_n; n = 1,..,N\}$ be the emission data and object, respectively. Here, $\mathbf{f}$ and $\mathbf{g}$ are 1D vectors obtained by lexicographic ordering 3D entities into 1D vector. The projection data $\mathbf{g}$ has an independent Poisson distribution with the (negative) log-likelihood function

$$\Phi_L(\mathbf{g};\mathbf{f}) = -\sum_i \{g_i \log \bar{g}_i - \bar{g}_i\} \tag{1}$$

where $\bar{\mathbf{g}} = \mathcal{H}\mathbf{f} + \bar{\mathbf{r}}$ is the Poisson mean of $\mathbf{g}$ with mean background events $\bar{\mathbf{r}}$. Note that $\mathcal{H}$ is the $M \times N$ system matrix with element $\mathcal{H}_{in}$ indicating the probability of a photon from pixel n detected in sinogram bin i.

From the Bayes theorem, the maximum *a posteriori* (MAP) reconstruction estimate $\hat{\mathbf{f}}$ of the object $\mathbf{f}$ is obtained by optimizing a regularized likelihood objective $\Phi(\mathbf{f};\mathbf{g})$ with prior (or penalty) objective $\Phi_P(\mathbf{f})$. Therefore, the MAP reconstruction $\hat{\mathbf{f}}$ is computed by

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}\geq 0}\{\Phi_L(\mathbf{g};\mathbf{f}) + \lambda\Phi_P(\mathbf{f})\}. \tag{2}$$

where $\lambda > 0$ is the global weight controlling the influence of the prior.

B. Conventional Smoothing Prior

The objective function for a conventional smoothing prior usually takes the following form

$$\Phi_P^s(\mathbf{f}) = \sum_n \sum_{n'\in\mathcal{N}(n)} w_{nn'}\phi(f_n - f_{n'}) \tag{3}$$

where $\mathcal{N}(n)$ indicates the neighborhood system, and $w_{nn'} \geq 0$ the associated weight. The potential function $\phi(f_n - f_{n'})$ penalizes the difference between neighborhood pixels. For example, for a quadratic membrane prior [5], $\phi(f_n - f_{n'}) = 1/2(f_n - f_{n'})^2$. Note that neither Φ_L nor Φ_P^s provides a natural positivity enforcement in Eq.(2). Thus, Eq.(2) is a possibly difficult constrained optimization problem since one needs to enforce the non-negativity constraint in the reconstruction.

C. Priors Based on Cross-Entropy

As a stepping stone towards our new priors, we first consider another, older form based on *I*-divergence. The *I*-divergence between two positive vectors $\mathbf{a}$ and $\mathbf{b}$ is $S(\mathbf{a},\mathbf{b}) = \sum_n (a_n \log \frac{a_n}{b_n} - a_n + b_n)$. $S(\mathbf{a},\mathbf{b})$ is also termed "cross-entropy" in [3], [4] though definitions of cross-entropy vary. In [3], and [4], priors of the form $\Phi_P^{ce}(\mathbf{f};\mathbf{m}) \equiv S(\mathbf{f},\mathbf{m})$ or $S(\mathbf{m},\mathbf{f})$ have been proposed. Here, $\mathbf{m}$ is an external "reference" vector chosen empirically. Each f_n is attracted towards its corresponding m_n, and positivity of $\mathbf{f}$ is maintained by the form of S. The "reference vector" $\mathbf{m}$ must be determined empirically! (For all m_n's equal, the cross-entropy prior becomes a max entropy prior.) In addition to the problem of determining $\mathbf{m}$, there is

no explicit smoothing of $\mathbf{f}$ implicit in $\Phi_P^{ce}(\mathbf{f};\mathbf{m})$. Thus $\Phi_P^{ce}(\mathbf{f};\mathbf{m})$ suffers from two problems.

D. A New Prior: Smoothed I-Divergence

To solve these problems while retaining the positivity, convexity and differentiability desiderata of $S(\mathbf{f},\mathbf{m})$ priors, we define a new prior, termed a smoothed *I*-divergence prior. In this definition, $\mathbf{m}$ is a variable to be estimated, and the m_n's are defined on a lattice coincident with the f_n. The definition of our first form of the prior is:

$$\begin{aligned}\Phi_P^{fm}(\mathbf{f},\mathbf{m}) &= \sum_n \sum_{n'\in\mathcal{N}(n)} w_{nn'}\phi(f_n, m_{n'}) \\ &= \sum_n \sum_{n'\in\mathcal{N}(n)} w_{nn'}\{f_n \log\frac{f_n}{m_{n'}} - f_n + m_{n'}\} \qquad (4)\end{aligned}$$

where $\mathcal{N}(n)$ again defines a neighborhood of n, and $w_{nn'} \geq 0$ are positive weights. The form of Eq.(4) embeds positivity in $\mathbf{f}$, and the prior is also differentiable.

To explore Φ_P^{fm} further, consider its role in MAP reconstruction. Since $\mathbf{m}$ is now a variable, the optimization in Eq.(2) (now *un*constrained) becomes a joint estimation:

$$\hat{\mathbf{f}},\hat{\mathbf{m}} = \arg\min_{\mathbf{f},\mathbf{m}}\{\Phi_L(\mathbf{g};\mathbf{f}) + \lambda\Phi_P^{fm}(\mathbf{f},\mathbf{m})\}. \qquad (5)$$

We propose to implement this joint estimation by a form of alternating (on $\mathbf{f}$ and $\mathbf{m}$) descent, which for iteration k becomes

$$\begin{aligned}\hat{\mathbf{f}}^{k+1} &= \arg\min_{\mathbf{f}}\{\Phi_L(\mathbf{g};\mathbf{f}) + \lambda\Phi_P^{fm}(\mathbf{f},\hat{\mathbf{m}}^k)\} \qquad (6)\\ \hat{\mathbf{m}}^{k+1} &= \arg\min_{\mathbf{m}}\{\Phi_P^{fm}(\hat{\mathbf{f}}^{k+1},\mathbf{m})\} \qquad (7)\end{aligned}$$

Equation (6), with $\hat{\mathbf{m}}^k$ fixed can be carried out by a suitable unconstrained gradient method. Equation (7) can be solved in closed form and has the interesting solution

$$\hat{m}_{n'}^k = \frac{\sum_{n\in\mathcal{N}(n')} w_{nn'}\hat{f}_n^k}{\sum_{n\in\mathcal{N}(n')} w_{nn'}}, \qquad (8)$$

that is, a weighted arithmetic mean of the $\hat{f}_n^k$'s in a neighborhood surrounding n'. Plugging Eq.(8) into Eq.(6), one thus sees the smoothing nature of the new prior, thus solving the first problem. The second problem has also been solved: vector $\mathbf{m}$ need not be determined empirically, and its formulation yields an appealing interpretation.

A second form of the new prior may be obtained by swapping $\mathbf{f}$, $\mathbf{m}$ in Eq.(4) to obtain

$$\begin{aligned}\Phi_P^{mf}(\mathbf{f},\mathbf{m}) &= \sum_n \sum_{n'\in\mathcal{N}(n)} w_{nn'}\phi(m_{n'}, f_n) \\ &= \sum_n \sum_{n'\in\mathcal{N}(n)} w_{nn'}\{m_{n'}\log\frac{m_{n'}}{f_n} - m_{n'} + f_n\} \qquad (9)\end{aligned}$$

Again, Φ_P^{mf} maintains positivity and is differentiable. When plugged into the alternation Eqs.(6)(7), however, the update Eq.(8) becomes something different:

$$\hat{m}_{n'}^k = \exp\left[\frac{\sum_{n\in\mathcal{N}(n')} w_{nn'}\log\hat{f}_n^k}{\sum_{n\in\mathcal{N}(n')} w_{nn'}}\right], \qquad (10)$$

which is a weighted *geometric* mean of the neighborhood pixels $\mathcal{N}(n')$. Again, we get a prescription for $\mathbf{m}$, and a novel form of smoothing for $\mathbf{f}$. For convenience, we refer to Eq.(4) as the FM prior, and Eq.(9) as the MF prior, with the quadratic version of Eq.(3) the MM prior.

E. Convexity

We would like to show the global convexity of the regularized likelihood objective with the proposed smoothing priors. Since the Poisson likelihood is convex, it is sufficient to show that the prior is convex (w.r.t. both $\mathbf{f}$ and $\mathbf{m}$).

The second derivatives of the MF prior objective are

$$\begin{aligned}\frac{\partial^2\Phi_P^{mf}}{\partial f_n^2} &= \sum_{n'} w_{nn'}\frac{m_{n'}}{f_n^2} \qquad (11)\\ \frac{\partial^2\Phi_P^{mf}}{\partial m_{n'}^2} &= \sum_{n} w_{nn'}\frac{1}{m_{n'}} \qquad (12)\\ \frac{\partial^2\Phi_P^{mf}}{\partial m_{n'}\partial f_n} &= -w_{nn'}\frac{1}{f_n} \qquad (13)\end{aligned}$$

It follows, for all y_n, $z_{n'}$ that

$$\begin{aligned}&\sum_n y_n^2\frac{\partial^2\Phi_P^{mf}}{\partial f_n^2} + 2\sum_{n,n'} y_n z_{n'}\frac{\partial^2\Phi_P^{mf}}{\partial m_{n'}\partial f_n} + \sum_n z_{n'}^2\frac{\partial^2\Phi_P^{mf}}{\partial m_{n'}^2} \\ &\quad= \sum_{n,n'}\frac{w_{nn'}}{f_n^2 m_{n'}}(y_n m_{n'} - z_{n'} f_n)^2 \geq 0. \qquad (14)\end{aligned}$$

The Hessian matrix is positive semi-definite (w.r.t. $\mathbf{f}$,$\mathbf{m}$), and thus the MF prior is convex. For the convexity of the FM prior, the proof follows by symmetry. Convexity of the prior or likelihood alone does not guarantee a unique solution, but when the likelihood and prior are combined as in Eq.(2), the solution is unique.

III. Results

In this section, we anecdotally explore 2D and 3D reconstructions using Φ_P^{mf} and Φ_P^{fm} and show that the reconstructions are qualitatively similar to those obtained with a (positivity constrained) conventional quadratic smoothing prior.

A. Optimization Details

Since the new priors embed positivity constraints, any unconstrained method can be used to optimize Eq.(6). Here, we use the Polak-Ribiere form of the preconditioned conjugate-gradient (PCG) method [6] with a simple diagonal preconditioner. For the line search, we use the

method of cubic interpolation [6]. We use, for 2D, a 4 nearest-neighborhood (4NN) system with weights=1.0 for each neighborhood pixel, and 4.0 for the center pixel. For 3D, we use a 6NN system with weights=1.0 for each neighborhood voxel, and 6.0 for the center voxel.

For comparison, we also performed a series of 2D reconstructions using the conventional quadratic smoothing Φ_P^s. We again use a 4NN neighborhood with weights as above. To maintain positivity, we used the slow EM-MAP-ICM algorithm presented in [5].

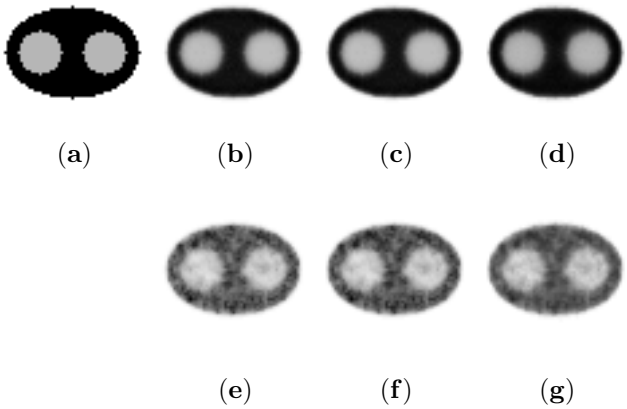

Fig. 1. (a) The 64×64 phantom used in the 2-D simulations. The noiseless anecdotal reconstructions with priors of (b) MF, (c) FM, and (d) MM. The anecdotal reconstructions of 100K counts data using regularized likelihood reconstruction with priors of (e) MF, (f) FM, and (g) MM.

B. Anecdotal Reconstructions

To test the proposed new priors, we first generated noiseless and noisy sinograms using the 2D 64×64 phantom shown in Fig.1(a). The 2D phantom has an intensity ratio of (3.3:1:0) for (ellipse:circle:background). The projection data had dimensions of 65 angles by 96 detector bins. Only Poisson noise is simulated and no other physical or geometrical blurring effects or background events are modeled here. The noisy sinogram is simulated with 100K counts.

The sinograms are then reconstructed using MF, FM and MM priors as described above, with the same smoothing parameter value $\lambda = 1.2$ for the noiseless case, and $\lambda = 2.0$ for the noisy case. The noiseless reconstructions are displayed in Fig.1 for (b) MF, (c) FM, and (d) MM, respectively, while Fig.1(e)(f)(g) are for noisy reconstructions of MF, FM, and MM. A profile plot along the center row of the noisy reconstructions in Fig.1(e)(f)(g) is illustrated in Fig.2 along with the phantom. The profile result demonstrates the similar behavior of the new prior to that of a positivity-constrained quadratic smoothing prior.

A few performance measures are shown in Fig.3. Plots of log posterior vs. iteration, and *rmse* vs. iteration appear in Fig.3(a) and (b), respectively, for the FM results of Fig.1(f), but we draw no conclusions from these since the simulations are not yet sufficiently realistic. We also display the bias and STD (standard deviation) images

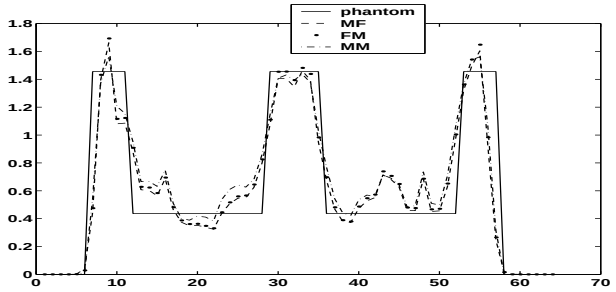

Fig. 2. Profile plots along the center row of phantom, and regularized likelihood reconstructions in Figs.1 (e) MF, (f) FM, and (g) MM priors for the noisy data of 100K counts.

for the FM prior using 200 noisy trials of 100K counts from the same 2D phantom. The bias and STD images are computed by $bias = \frac{1}{200}\sum_l(\hat{\mathbf{f}}^l - \mathbf{f})$ and $STD = \sqrt{\frac{1}{199}\sum_l(\hat{\mathbf{f}}^l - \bar{\hat{\mathbf{f}}})^2}$ where $\bar{\hat{\mathbf{f}}}$ is the mean reconstruction over 200 trials, $\mathbf{f}$ the phantom, and $\hat{\mathbf{f}}^l$ is the lth reconstruction. The results for MF and MM priors are similar. As λ increases, the variance as seen in Fig.3(d) tends to become more uniform.

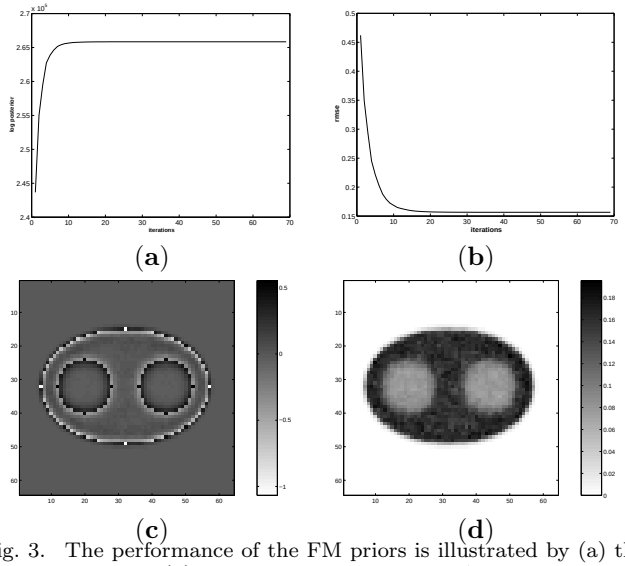

Fig. 3. The performance of the FM priors is illustrated by (a) the log posterior and (b) the *rmse* vs. iterations. Also, the bias and standard deviation images over 200 noise trials for the FM prior regularized reconstructions are shown in (c) and (d), respectively.

We also generated a 3D phantom with 10 slices of 64×64 each. Each slice has two small hot and cold circles with an ellipse background. The intensity ratio is (hot:ellipse:cold)=(1.5:1:0.6). The hot and cold circles in the top slice have diameters of 10 and 12 voxels, respectively, while each slice below contains hot and cold circles of reducing diameters at 1 voxel per slice. A noisy sinogram is generated with 1000K total counts and reconstructed with a 3D FM prior of 6NN. Here, we show only 4 slices of the 3D phantom in the top row of Fig.4, and the FM reconstructions with smoothing parameter $\lambda = 5.0$ in the bottom row of Fig.4. The 3D results for the MF and MM priors are qualitatively similar.

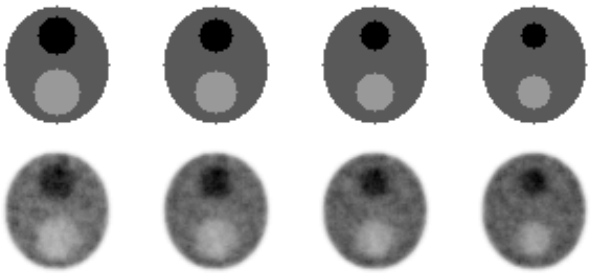

Fig. 4. This figure shows the anecdotal regularized likelihood reconstruction of a 3D phantom (10 slices and each with the size of 64×64) using the FM smoothing prior. The top row shows 4 consecutive slices of the phantom, and the bottom row shows their corresponding FM reconstructions. The projection data is simulated with a noisy level of 1 million total counts. The smoothing parameter for this 3D case is λ=5.

IV. Discussion

To locate our new work within the galaxy of ECT reconstruction methods, we take a distant (and *very* incomplete!) overview of objective-function based methods in the context of the following desiderata: (a) speed, (b) need to specify additional parameters beyond a global weight λ, (c) whether a notion of object smoothness is contained in the objective. We consider 3 categories of methods, (where "method" = objective + algorithm.) We do not consider any cases for which positivity is not imposed or the objective is non-convex.

The original ECT EM-ML algorithm and its variants led to a natural positivity imposition, but, in terms of (b) required a stopping criterion and initial condition specification. True EM-based methods are also notoriously slow, thus failing on point (a). Regularization helps, and EM-MAP approaches [7], [5] incorporating smoothing, satisfied (c), helped (b), but still suffered in terms of speed (a). Our own method, if using gradient based methods as in Sec. III, is likely to be faster than an EM version.

A second (non-EM) category of methods includes constrained gradient and coordinate-descent methods for optimization. Imposing positivity for a gradient method is often complex and difficult, and requires judicious specification of algorithm-specific parameters [8], [9] to attain a good result, thus failing on point (b), though excelling on point (a) and satisfying (c). To a great extent, the failing on (b) can be overcome [10] with improved methods. Sequential Gauss-Seidel coordinate descent methods [11], [12] incorporate positivity easily, and can be fast.

A third category of approaches embeds positivity directly into the prior. Our approach, and ones based on cross-entropy [2], [4], [3] and max-entropy [13] apply here. A very similar prior is the independent gamma prior proposed in [14]. The cross-entropy and gamma formulations fail badly on point (b), requiring empirical specification of a pointwise "reference image" as outlined in Sec. II-C, and no explicit notion of object smoothness is incorporated in these approaches. (In [2], [4], smoothing is indeed introduced in an empirical fashion.) Our own priors thus surmount the difficulties listed above.

A final, nice feature of our formulation is that the performance of the method (i.e. mean and covariance of $\hat{\mathbf{f}}$) may be analyzable using the theoretical methods in [15]. To qualify for [15], the objective should be (i) smooth, (ii) nearly quadratic, and (iii) be used in an unconstrained optimization. We are currently investigating this issue.

V. Acknowledgments

This work was supported by NIH NS32879.

References

[1] I. Csiszár, "MAXENT, mathematics, and information theory," in *Proc. 15th Inter. Work. Maximum Entropy and Bayesian Methods*, K. Hanson and R. Silver, Eds., 1995, pp. 35–50.

[2] J. Liang, R. Jaszczak, and K. Greer, "On Bayesian image reconstruction from projections: Uniform and nonuniform *a priori* source information," *IEEE Trans. Med. Imag.*, vol. 8, no. 3, pp. 227–235, 1989.

[3] C. L. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Tran. Imag. Proc.*, vol. 2, no. 1, pp. 96–103, Jan. 1993.

[4] B. A. Ardekani, M. Braun, B. F. Hutton, I. Kanno, and H. Iida, "Minimum cross-entropy reconstruction of PET images using prior anatomical information," *Phys. Med. Biol.*, vol. 41, no. 11, pp. 2497–2517, Nov. 1996.

[5] S. J. Lee, A. Rangarajan, and G. R. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Trans. Med. Imag.*, vol. vol. 14, pp. 669–680, 1995.

[6] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, second edition, 1984.

[7] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction for Poisson data using Gibbs priors," *IEEE Trans. on Medical Imaging*, vol. 8, no. 2, pp. 194–202, June 1989.

[8] 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.*, vol. 13, no. 4, pp. 687–701, Dec. 1994.

[9] C. Johnson, J. Seidel, and A. Sofer, "Interior-point methodology for 3-D PET reconstruction," *IEEE Trans. Med. Imag.*, vol. 19, no. 4, pp. 271–285, Apr. 2000.

[10] E. U. Mumcuoglu, R. M Leahy, and S. R. Cherry, "Bayesian reconstruction of PET images: Methodology and performance analysis," *Phys. Med. Bio.*, vol. 41, pp. 1777–1807, 1996.

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

[12] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Imag. Proc.*, vol. 5, no. 3, pp. 480–492, Mar. 1996.

[13] J. Nunez and J. Llacer, "A fast Bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images," *IEEE Trans. Med. Imag.*, vol. 9, no. 2, pp. 159–171, 1990.

[14] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imag.*, vol. 6, no. 2, pp. 106–114, June 1987.

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