Joint Estimation of Image and Deformation Parameters in Motion-Corrected PET

M.W. Jacobson and J.A. Fessler

Abstract— We consider an emission tomography reconstruction problem in which projection measurements from several time frames are available. Two strategies for doing motion-corrected image reconstruction are compared. In the first strategy, separate images are reconstructed from the measurements at each time frame. They are then consolidated by post-registration and averaging procedures. In the second strategy, we incorporate parameters to describe the effects of motion into the statistical model of the projections. Joint maximum likelihood estimation of image and motion parameters is then carried out. Each of these strategies involves the minimization of non-convex cost functions. Accordingly, we also propose some relevant optimization algorithm design options.

I. INTRODUCTION

To account for motion effects in Positron Emission Tomography (PET), a common practice is to gate the acquisition period into intervals or collections of intervals (henceforth called *time frames*) over which the object is considered to be motionless. With the time frames enumerated $\kappa = 0, 1, 2, \ldots, K - 1$, one acquires a separate measured projection vector y^κ in each of these time frames so as to get a "snapshot" of the radio-tracer activity image. In this work, we examine different strategies for reconstructing an image of the object based on the measured data $\{y^\kappa\}_{\kappa=0}^{K-1}$.

We model the measurements as Poisson with ensemble means $\{\bar{y}^\kappa\}$ given by,

$$\bar{y}^0(\lambda) = P^0 \lambda + r^0 \tag{1}$$
$$\bar{y}^\kappa(\lambda, \alpha^\kappa) = P^\kappa W^\kappa(\alpha^\kappa)\lambda + r^\kappa, \quad \kappa = 1, \ldots, K - 1. \tag{2}$$

Here λ is the vector of unknown activity image pixel values in time-frame $\kappa = 0$. Also, for each time-frame κ, P^κ is a forward projection matrix, r^κ is a known vector of mean background counts, and $W^\kappa(\alpha^\kappa)$ is an image transformation matrix that depends on an unknown deformation parameter vector α^κ. The transformed image $W^\kappa(\alpha^\kappa)\lambda$ gives the pixel values of the activity image in time frame κ, after undergoing changes due to anatomical motion (and possibly other physical processes). We shall denote $\boldsymbol{\alpha} = (\alpha^1, \alpha^2, \ldots, \alpha^{K-1})$ as the total array of unknown deformation parameters. Typically, each vector α^κ has far fewer components than λ. For example, α^κ could be the coefficients in a B-spline deformation model (see also Section III-B).

We consider two main reconstruction strategies. The first, which we call the Frame-Wise with Post Registration (FWPR)

strategy is to obtain estimates $\hat{\lambda}$ and $\hat{\alpha}^\kappa$, $\kappa = 1, \ldots, K$ according to the following steps:

1) *Frame-Wise Reconstructions.* For each $\kappa \geq 0$, reconstruct the *deformed* image $z^\kappa = W(\alpha^\kappa)\lambda$ using penalized likelihood estimation,

$$\hat{z}^\kappa = \underset{z}{\arg\min} \left\{ L^k(z; y^\kappa) + R^\kappa(z) \right\}. \tag{3}$$

Here

$$L^\kappa(z^\kappa; y^\kappa) = \sum_i (([P^\kappa z^\kappa]_i + r_i^\kappa) - y_i^\kappa \log([P^\kappa z^\kappa]_i + r_i^\kappa))$$

is the negative loglikelihood for a single time frame κ in terms of the deformed image z^κ and $R^\kappa(z^\kappa)$ is a roughness penalty.

2) *Post-Registration.* Fit all $\{z^\kappa\}_{\kappa=1}^{K-1}$ to z^0 by doing regularized least squares post-registration

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\arg\min} \left\{ \sum_\kappa ||W^\kappa(\alpha^\kappa)\hat{z}^0 - \hat{z}^\kappa||^2 + \Psi(\boldsymbol{\alpha}) \right\} \tag{4}$$

where $\Psi(\boldsymbol{\alpha})$ is a regularizing penalty function.

3) *Undeform and Consolidate.* Apply inverse deformations to each frame-wise reconstruction and average to obtain a final activity image estimate

$$\hat{\lambda} = \frac{\hat{z}^0 + \sum_{\kappa=1}^{K-1} W^\kappa([\hat{\alpha}^\kappa]^{-1})\hat{z}^\kappa}{K}. \tag{5}$$

Here $[\hat{\alpha}^\kappa]^{-1}$ denotes the parameter vector that induces the (approximate) inverse transformation to that of $\hat{\alpha}^\kappa$.

The second strategy, which we call Joint Estimation with Deformation Modeling (JEDM), is to do penalized likelihood estimation of *all* the unknown parameters based on the *entire* data set $\{y^\kappa\}_{\kappa=0}^{K-1}$ as follows,

$$(\hat{\lambda}, \hat{\boldsymbol{\alpha}}) = \underset{\lambda, \boldsymbol{\alpha}}{\arg\min} \left\{ L(\lambda, \boldsymbol{\alpha}; y^0, \ldots, y^{K-1}) + R(\lambda) + \Psi(\boldsymbol{\alpha}) \right\}. \tag{6}$$

In FWPR, the deformation parameters are not obtained from the Poisson model, but rather by post-reconstruction fitting using (4). Moreover, since each $\hat{z}^\kappa$ is derived only from the corresponding y^κ, a relatively low count data set, then we can expect considerable errors to propagate into the estimates of $\hat{\boldsymbol{\alpha}}$ and $\hat{\lambda}$ via (4) and (5).

Conversely, in the JEDM approach, all parameters are estimated based on the penalized Poisson loglikelihood in (6), a more statistically principled approach. In addition, both the

deformation parameters and the image parameters are jointly derived from the total data set. We therefore expect the JEDM strategy to better estimate both λ and $\boldsymbol{\alpha}$, provided that suitable algorithms for the minimization (6) can be developed.

II. PRELIMINARY EXPERIMENT

We made a preliminary test of each strategy using simulated acquisitions of a 64×64 torso phantom with a circular hot lesion, 2 pixels in diameter. The projection space was discretized into 180 angles by 64 radial bins. We used 2 time frames. In the second time frame, the phantom was stretched by a factor of $\alpha^1 = 10\%$ in one direction. Tests were based on 50 independent realizations of the acquisition at a count level of 3 average counts per bin per time-frame. In all cases, we used a discretized line integral forward projector and 10% mean background count rates. Presently, all cost function minimizations were unregularized. We discretized the space of α^1 values into 22 values from 5% to 15%. Loglikelihood maximization with respect to λ was done for each fixed α^1 value using ML-EM. Optimization of all cost functions with respect to α^1 was done by exhaustive search. In principal, this approach ensures that all cost functions can be minimized globally.

To test performance, we computed the empirical bias and standard deviation of the estimates of α^1, as well as the percent root mean squared error, given by $\%\mathrm{RMS}(\hat{\theta}) = \sqrt{\mathrm{MSE}(\hat{\theta})}/|\theta_{\mathrm{true}}|$. The same was done for the estimated tracer uptake in the hot lesion, i.e., $\hat{U} = \sum_{j \in J} \hat{\lambda}_j$ where J are the lesion pixels. The true uptake was 24.7. These results are reported in Table I. For comparison, we have also reported the performance of ML-EM as applied to the total acquired data $\sum_{\kappa=1}^{K} y^\kappa$, i.e., when no gating is used.

As predicted in Section I, the post-registration operation (4) in the FWPR reconstructions resulted in less accurate estimates of α^1 than JEDM. Also, as one might expect, both FWPR and JEDM exhibited better accuracy in estimating the lesion uptake than non-gated ML-EM. What is surprising is that FWPR's poorer estimate of α^1 did not translate into a significantly poorer estimate of the uptake than JEDM. At this time, our best conjecture is that errors in the approximation of the inverse transformations in (5) somehow countered the registration errors in (4).

More sophisticated experiments will clearly be necessary to see which of these algorithms has more practical benefit. Whatever we find, it will be necessary to develop iterative algorithms to handle the cost function minimizations in these algorithms. This is discussed next.

III. OPTIMIZATION TRANSFER ALGORITHM DESIGN

In this section, we discuss optimization transfer algorithms for doing the non-convex cost function minimizations[1] in

[1]The cost functions in (3) are convex for typical penalties $R^\kappa(\cdot)$ and their minimization is now standard in statistical image reconstruction.

(4) and (6). It will suffice to discuss the unregularized cost functions

$$\Phi_{\mathrm{FWPR}}(\alpha^\kappa) = ||W^\kappa(\alpha)\hat{z}^0 - \hat{z}^\kappa||^2 \tag{7}$$

$$\Phi_{\mathrm{JEDM}}(\lambda, \boldsymbol{\alpha}) = \sum_{jk\kappa}[s_j^\kappa W_{jk}^\kappa(\alpha^\kappa)\lambda_k - y_i^\kappa \log[P_i W_{jk}^\kappa(\alpha^\kappa)\alpha + r_i^\kappa]] \tag{8}$$

Here, $s_j^\kappa = \sum_i P_{ij}^\kappa$ are the pixel values of the sensitivity images of frame κ. Extension to the regularized case is a routine matter (see, for example [1], [2]).

A. Background

Optimization transfer is a method for addressing cost function minimization problems that has received considerable attention in recent statistical imaging literature (e.g. [1], [2], [3], [4]). Given a cost function $\Phi(\theta)$, the method involves constructing a function $\phi(\cdot; \cdot)$ satisfying $\phi(\theta, \bar{\theta}) \geq \Phi(\theta)$ for all θ, with equality at $\bar{\theta}$. Henceforth, we abbreviate this relationship, by writing

$$\phi(\cdot; \bar{\theta}) \overset{\bar{\theta}}{\succ} \Phi(\cdot). \tag{9}$$

We then call $\phi(\cdot; \cdot)$ a *surrogate generator* for Φ and the function $\phi(\cdot; \bar{\theta})$ a *surrogate* for Φ with expansion point $\bar{\theta}$. Any minimizer θ^* of $\phi(\cdot, \bar{\theta})$ has the property that $\Phi(\theta^*) \leq \Phi(\bar{\theta})$. Applying the iterative algorithm,

$$\theta^{n+1} \in \underset{\theta}{\mathrm{argmin}}\{\phi(\theta; \theta^n)\} \tag{10}$$

one therefore obtains a sequence $\{\theta^n\}$ that monotonically reduces Φ (see Fig. 1) and whose limit points are stationary under fairly weak conditions [5].

When the space Θ of unknowns θ is decomposable into a Cartesian product, one can modify (10) and alternately minimize with respect to a subset of the unknown variables [6], [7], [5]. A natural application of this approach to the minimization of $\Phi_{\mathrm{JEDM}}(\lambda, \boldsymbol{\alpha})$ is to find surrogates $\phi_1(\lambda; \bar{\lambda}, \bar{\boldsymbol{\alpha}})$ and $\phi_2(\alpha; \bar{\lambda}, \bar{\boldsymbol{\alpha}})$ with respect to λ and α respectively, i.e.,

$$\phi_1(\cdot; \bar{\lambda}, \bar{\boldsymbol{\alpha}}) \overset{\bar{\lambda}}{\succ} \Phi_{\mathrm{JEDM}}(\cdot, \bar{\boldsymbol{\alpha}})$$

$$\phi_2(\cdot; \bar{\lambda}, \bar{\boldsymbol{\alpha}}) \overset{\bar{\boldsymbol{\alpha}}}{\succ} \Phi_{\mathrm{JEDM}}(\bar{\lambda}, \cdot),$$

and to iterate as follows

1) Set $n = 0$ and select an initial point $(\lambda^0, \boldsymbol{\alpha}^0)$.
2) Update the image by letting λ^{n+1} be any minimizer of $\phi_1(\cdot; \lambda^n, \boldsymbol{\alpha}^n)$.
3) Update the deformation parameters by letting $\boldsymbol{\alpha}^{n+1}$ be any minimizer of $\phi_2(\cdot; \lambda^{n+1}, \boldsymbol{\alpha}^n)$.
4) Set $n := n + 1$ and go to step 2.

Finding surrogate generators is the art of the optimization transfer method. To facilitate minimizations like (10), one hopes to find $\phi(\cdot; \cdot)$ such that, for every $\bar{\theta}$, $\phi(\cdot; \bar{\theta})$ is convex. Some tools for doing so for the problems at hand are presented in subsequent sections.

TABLE I

PERFORMANCE DATA FOR RECONSTRUCTIONS.

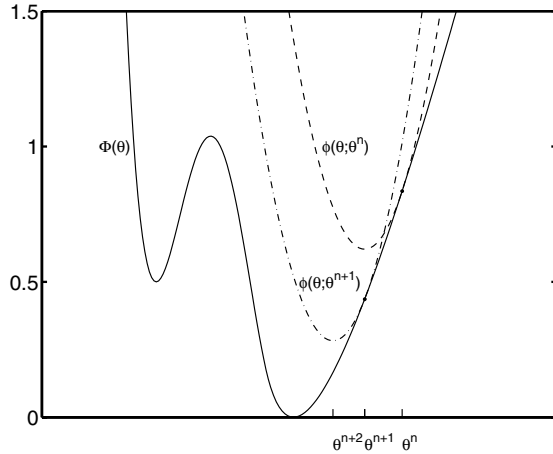| Estimated Quantity	Method	Bias	Std. Dev.	Std. Error (of Mean)	Std. Error (of Std. Dev.)	%RMS Error
Uptake, U	FWPR	-1.35	3.23	0.46	0.32	14.05
	JEDM	-0.72	3.28	0.46	0.33	13.45
	ML-EM	-6.56	3.22	0.46	0.32	29.51
Stretch, α^1	FWPR	0.31	2.45	0.35	0.32	24.46
	JEDM	-0.35	0.85	0.12	0.33	9.08

Fig. 1. A 1D illustration of the optimization transfer principal.

B. Modeling image deformation and related optimization transfer considerations

Here we discuss a typical form for the deformation matrix elements $W_{jk}^{\kappa}(\alpha^{\kappa})$, one which also proves convenient from the point of view of algorithm design. In conventional image registration, a common way of deforming a discrete image λ is by way of the following steps.

1) **Interpolate** the discrete image into continuous space,

$$\boldsymbol{\lambda}(\vec{r}) = \sum_k \lambda_k h_k(\vec{r})$$

where the $h_k(\vec{r})$ are interpolation functions.

2) **Warp** the continuous image using a spatially dependent, additive perturbation

$$\vec{r} \leftarrow \vec{r} + \sum_\ell b_\ell(\vec{r})\alpha_\ell^{\kappa}$$

$$W^{\kappa}\boldsymbol{\lambda}(\vec{r}) = \sum_k \lambda_k h_k\left(\sum_\ell b_\ell(\vec{r})\alpha_\ell^{\kappa} + \vec{r}\right).$$

Here the $b_\ell(\vec{r})$ are deformation basis functions (B-splines are a typical choice) and α_ℓ^{κ} are scalar coefficients.

3) **Resample** the warped image at locations $\vec{r}_j$

$$W^{\kappa}\boldsymbol{\lambda}(\vec{r}_j) = \sum_k \lambda_k \underbrace{h_k(B_j\alpha^{\kappa} + \vec{r}_j)}_{W_{jk}^{\kappa}(\alpha^{\kappa})}$$

where

$$B_j = [b_1(\vec{r}_j) \quad b_2(\vec{r}_j) \quad b_3(\vec{r}_j) \quad \cdots]$$

is a matrix whose columns are the basis functions sampled at $\vec{r}_j$.

In this family of deformations, therefore, we have

$$W_{jk}^{\kappa}(\alpha^{\kappa}) = h_k(B_j\alpha^{\kappa} + \vec{r}_j). \tag{11}$$

That is, the deformation matrix elements are compositions of the interpolation functions $h_k(\vec{r})$ with affine expressions $(B_j\alpha^{\kappa} + \vec{r}_j)$ in α^{κ}.

Normally, one would choose interpolators $h_k(\vec{r})$ that are continuous and whose supports are of small extent. Therefore, from (11), any function of the matrix entries $W_{jk}^{\kappa}(\alpha^{\kappa})$ having the form

$$f(W^{\kappa}(\alpha^{\kappa})) = \sum_j f_j(W_j^{\kappa}(\alpha^{\kappa})), \tag{12}$$

where $W_j^{\kappa}(\alpha^{\kappa})$ is the j-th row of $W^{\kappa}(\alpha^{\kappa})$ and $f_j(\cdot)$ is a polynomial function of the elements of this row, can be expressed

$$f(W^{\kappa}(\alpha^{\kappa})) = \sum_j \sum_n c_{nj}\tilde{h}_n(\vec{r})\big|_{\vec{r}=B_j\alpha^{\kappa}+\vec{r}_j}, \tag{13}$$

where the $\tilde{h}_n(\vec{r})$ are again continuous and have small supports.

It is useful to find convex quadratic surrogate generators $q_n^+(\vec{r}; \vec{r}_0)$ to $\tilde{h}_n(\vec{r})$ and $q_n^-(\vec{r}; \vec{r}_0)$ to $-\tilde{h}_n(\vec{r})$, i.e.,

$$q_n^-(\cdot; \vec{r}_0) \stackrel{\vec{r}_0}{\prec} \tilde{h}_n(\cdot) \stackrel{\vec{r}_0}{\prec} q_n^+(\cdot; \vec{r}_0). \tag{14}$$

This is illustrated for a hypothetical 1D $\tilde{h}(\vec{r})$ in Fig. 2. Upon doing so, one can then construct convex quadratic surrogate generators $Q^+(\alpha^{\kappa}; \bar{\alpha}^{\kappa})$ to $f(W^{\kappa}(\alpha^{\kappa}))$ and $Q^-(\alpha^{\kappa}; \bar{\alpha}^{\kappa})$ to $-f(W^{\kappa}(\alpha^{\kappa}))$, i.e.,

$$Q^-(\cdot; \bar{\alpha}^{\kappa}) \stackrel{\bar{\alpha}^{\kappa}}{\prec} f(W^{\kappa}(\cdot)) \stackrel{\bar{\alpha}^{\kappa}}{\prec} Q^+(\cdot; \bar{\alpha}^{\kappa}), \tag{15}$$

by considering (13) and defining

$$Q^+(\alpha^{\kappa}; \bar{\alpha}^{\kappa}) = \sum_{c_{nj}\geq 0} c_{nj}q_n^+(B_j\alpha^{\kappa} + \vec{r}_j; B_j\bar{\alpha}^{\kappa} + \vec{r}_j)$$
$$+ \sum_{c_{nj}<0} |c_{nj}|q_n^-(B_j\alpha^{\kappa} + \vec{r}_j; B_j\bar{\alpha}^{\kappa} + \vec{r}_j)$$

and similarly for $Q^-(\alpha^{\kappa}; \bar{\alpha}^{\kappa})$.

Since $\tilde{h}_n(\vec{r})$ is a function of only spatial coordinates – a small number of variables – finding $q_n^+(\vec{r}; \vec{r}_0)$ and $q_n^-(\vec{r}; \vec{r}_0)$ satisfying (14) would not be an overly difficult task. For example, one could exhaustively test different curvature parameters and pre-tabulate them for a suitable range of $\vec{r}_0$.
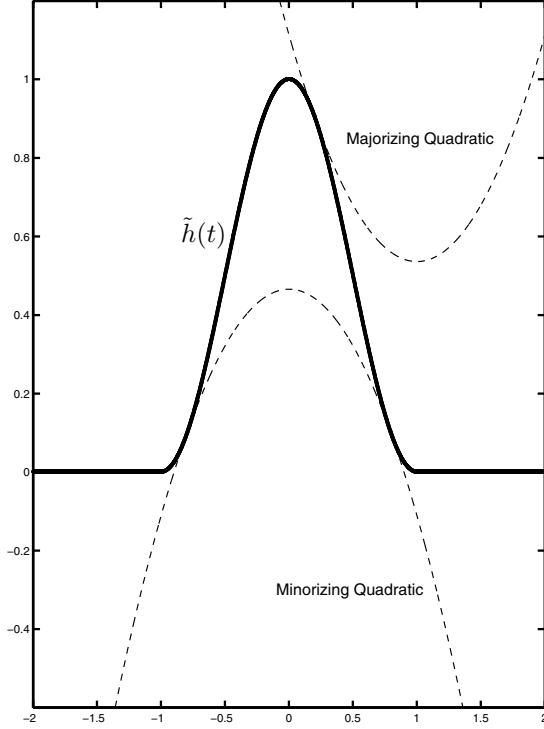

Fig. 2. Quadratic majorizers and minorizers for a 1D finite support function.

C. Construction of surrogates for the FWPR fitting function

By inspection of (7), one can see that $\Phi_{\text{FWPR}}(\alpha^\kappa)$ has the form (12) with

$$f_j(W_j^\kappa(\alpha^\kappa)) = \sum_\kappa (W_j^\kappa(\alpha^\kappa)\hat{z}^0 - \hat{z}_j^\kappa)^2.$$

Therefore, one can apply the methodology of Section III-B directly to obtain a convex quadratic surrogate generator $Q^+(\cdot; \bar{\alpha}^\kappa) \overset{\bar{\alpha}^\kappa}{\succ} \Phi_{\text{FWPR}}(\cdot)$.

D. Construction of surrogates for the JEDM Loglikelihood

In the JEDM strategy, the relevant cost function Φ_{JEDM} is a Poisson negative loglikelihood function. In statistical tomographic imaging, a now classical way to derive surrogate generators for negative loglikelihoods (e.g. [8], [9]) is via the Expectation Maximization (EM) methodology [10]. The next proposition describes EM-based surrogates for general Poisson negative loglikelihoods,

$$L(\theta; y) = \sum_i [\bar{y}_i(\theta) - y_i \log(\bar{y}_i(\theta))]. \qquad (16)$$

First, recall that the Kullback-Leibler (KL) distance between two arrays a and b with non-negative components a_{im} and b_{im} is,

$$\text{KL}(a, b) = \sum_{i,m} \left[a_{im} \log \frac{a_{im}}{b_{im}} + b_{im} - a_{im} \right] \qquad (17)$$

with the conventions $0 \log 0 = 0$, $0/0 = 0$, and $\log(1/0) = \infty$. The KL distance has the property that $\text{KL}(a, b) \geq 0$ with equality iff $a = b$.

Proposition 3.1 (Poisson Model EM-based Surrogates):
Suppose we are given Poisson measurements y_i with ensemble means

$$\bar{y}_i(\theta) = \sum_m^{M_i} \bar{y}_{im}(\theta), \qquad (18)$$

where all $\bar{y}_{im}(\theta) \geq 0$, and define

$$y_{im}(\theta) = y_i \frac{\bar{y}_{im}(\theta)}{\bar{y}_i(\theta)}$$
$$Q(\theta; \bar{\theta}) = L(\theta) + \text{KL}(y(\bar{\theta}), y(\theta)) \qquad (19)$$
$$= \sum_{im} [\bar{y}_{im}(\theta) - y_{im}(\bar{\theta}) \log \bar{y}_{im}(\theta)] + \text{const.} \qquad (20)$$

Then $Q(\cdot; \bar{\theta}) \overset{\bar{\theta}}{\succ} L(\cdot; y)$, i.e., Q is a surrogate generator for $L(\cdot; y)$.

Proof. The fact that $Q(\cdot; \bar{\theta}) \overset{\bar{\theta}}{\succ} L(\cdot)$ is an easy consequence of (19) and the properties of the KL distance. The fact that $Q(\theta; \bar{\theta})$ also has the form (20) can be verified by expanding the term $\text{KL}(y(\bar{\theta}), y(\theta))$ in (19). $\square$

Remark 3.2: The surrogate generator given by Proposition 3.1 can also be derived using the classical EM methodology. To do so, one expresses each y_i as the sum of complete data $\{Y_{im}\}$, which are chosen to be i.i.d. Poisson random variables with means $\{\bar{y}_{im}(\theta)\}$.

With $\alpha = \bar{\alpha}$ held fixed and applying Proposition 3.1 to the additive form $\bar{y}_i^\kappa(\lambda) = \sum_{jk} P_{ij}^\kappa W_{jk}^\kappa(\bar{\alpha}^\kappa)\lambda_k$, one obtains the surrogate

$$\phi_1(\lambda; \bar{\lambda}, \bar{\alpha}) =$$
$$\sum_{i\kappa} \bar{y}_i^\kappa(\lambda) - \sum_k \left(\sum_{ij\kappa} \frac{P_{ij}^\kappa W_{jk}^\kappa(\bar{\alpha}^\kappa)}{\bar{y}_i^\kappa(\bar{\lambda}, \bar{\alpha}^\kappa)} \right) \bar{\lambda}_k \log(\lambda_k) \qquad (21)$$

This is the standard ML-EM surrogate of [8]. Similarly, with $\lambda = \bar{\lambda}$ held fixed and applying Proposition 3.1 to the additive form $\bar{y}_i^\kappa(\alpha) = \sum_{jk} P_{ij}^\kappa W_{jk}^\kappa(\alpha^\kappa)\bar{\lambda}_k$ one obtains,

$$\phi_2(\alpha; \bar{\lambda}, \bar{\alpha}) = \sum_{jk\kappa} s_j^\kappa \bar{\lambda}_k W_{jk}^\kappa(\alpha^\kappa)$$
$$+ \sum_{jk\kappa} \bar{\lambda}_k \left(\sum_i \frac{y_i P_{ij}^\kappa}{\bar{y}_i^\kappa(\bar{\lambda}, \bar{\alpha}^\kappa)} \right) W_{jk}^\kappa(\bar{\alpha}^\kappa)(-\log W_{jk}^\kappa(\alpha^\kappa)) \qquad (22)$$

The surrogate generators $\phi_1(\cdot; \cdot)$ and $\phi_2(\cdot; \cdot)$ are candidates for the block alternating procedure described in Section III-A.

A drawback of $\phi_2(\boldsymbol{\alpha}; \bar{\lambda}, \bar{\boldsymbol{\alpha}})$ is that it may not be convex. However, the first sum on the RHS of (22) has the form (12) with

$$f_j(W_j^\kappa(\alpha^\kappa)) = s_j^\kappa \sum_k \bar{\lambda}_k W_{jk}^\kappa(\alpha^\kappa). \tag{23}$$

One can therefore find a convex quadratic surrogate for it as in (15). Furthermore, if one chooses $h_k(\vec{r}) = h(\vec{r} - \vec{r}_k)$ where

$$h(x_1, x_2, x_3) = \prod_{\ell=1}^3 0.5(1 + \cos(\pi x_\ell))\mathrm{rect}(x_\ell/2). \tag{24}$$

is the separable product of windowed raised cosines then, noting (11), the expressions $-\log W_{jk}^\kappa(\alpha^\kappa)$, and hence also the second sum on the RHS of (22), will be convex. It follows that a convex surrogate $\phi_2'(\boldsymbol{\alpha}; \bar{\lambda}, \bar{\boldsymbol{\alpha}}) \overset{\bar{\boldsymbol{\alpha}}}{\succ} \phi_2(\boldsymbol{\alpha}; \bar{\lambda}, \bar{\boldsymbol{\alpha}})$ can be obtained, which is also a surrogate for $\Phi_{\mathrm{JEDM}}(\bar{\lambda}, \cdot)$.

Another potential difficulty is that, when surrogate generators $\phi(\theta; \bar{\theta})$ are used that have singularities as functions of $(\theta, \bar{\theta})$, the iterates $\{\theta^n\}$ may have non-stationary limit points. It is for this reason that most optimization transfer literature (e.g. [11], [7], [12]) forbids such discontinuity. In the case of (21) and (22), both surrogate generators have singularities of this kind. The limiting behavior of the corresponding optimization transfer algorithm is consequently unpredictable.

One remedy in these kinds of situations is to generate further surrogates $\phi'(\cdot, \bar{\theta}) \overset{\bar{\theta}}{\succ} \phi(\cdot; \bar{\theta})$ which do not suffer from these discontinuities. Another option is given by the following proposition.

Proposition 3.3 (Majorization/Minorization Technique):
Given Poisson measurements y_i with ensemble means $\bar{y}_i(\theta)$, suppose one can find functions $Q_i^+(\theta; \bar{\theta})$ and $Q_i^-(\theta; \bar{\theta})$ such that

$$Q_i^-(\cdot; \bar{\theta}) \overset{\bar{\theta}}{\prec} \bar{y}_i(\cdot) \overset{\bar{\theta}}{\prec} Q_i^+(\cdot; \bar{\theta}). \tag{25}$$

Suppose further that for every $\bar{\theta}$, the functions $Q_i^+(\cdot; \bar{\theta})$ are convex and the functions $Q_i^-(\cdot; \bar{\theta})$ are concave. Then

$$\phi(\cdot; \bar{\theta}) = \sum_i [Q_i^+(\cdot; \bar{\theta}) - y_i \log Q_i^-(\cdot; \bar{\theta})] \tag{26}$$

is a convex surrogate for $L(\cdot; y)$ for every $\bar{\theta}$.

Proof. The fact that $\phi(\cdot; \bar{\theta}) \overset{\bar{\theta}}{\succ} L(\cdot; y)$ is immediate from (25). It remains to establish convexity. It is easily shown that substituting a concave function into a convex, nonincreasing function such as $-\log(\cdot)$ produces a convex function. Hence, the RHS of (26) is a sum of convex terms. □

An advantage of Proposition 3.3 is that if $Q_i^+(\theta; \bar{\theta})$ and $Q_i^-(\theta; \bar{\theta})$ are continuous in $(\theta, \bar{\theta})$, then $\phi(\theta; \bar{\theta})$, as given by (26), is always continuous at all pairs $(\theta, \bar{\theta})$ in some level set of $L(\theta; y)$. So the aforementioned difficulty with EM-based surrogates is avoided.

Furthermore, it is often simple to find the majorizing and minorizing functions required in (25). In particular, holding $\lambda = \bar{\lambda}$ fixed in (2) yields

$$\bar{y}_i^\kappa(\alpha^\kappa) = \sum_{jk} P_{ij}^\kappa \bar{\lambda}_k W_{jk}^\kappa(\alpha^\kappa).$$

which has the form (12) with

$$f_j(W_j^\kappa(\alpha^\kappa)) = P_{ij}^\kappa \sum_k \bar{\lambda}_k W_{jk}^\kappa(\alpha^\kappa). \tag{27}$$

Suitable quadratic majorizing and minorizing functions for $\bar{y}_i^\kappa(\alpha^\kappa)$ are then obtainable as in (15).

IV. CONCLUSIONS

We have considered a statistical model for PET which includes deformation parameters to describe the motion of the activity image throughout the scan. Preliminary experimental results suggested that JEDM performs better than FWPR for estimating motion, but were inconclusive as to which is better for quantifying lesion uptake. Future experiments with more elaborate deformation models may show more conclusive trends. These more elaborate models will entail more complicated cost function minimizations. To this end, several options for the design of optimization transfer algorithms have been discussed and await testing.

REFERENCES

[1] H Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Tr. Med. Im.*, vol. 18, no. 9, pp. 801–14, Sept. 1999.
[2] Hakan Erdoğan, *Statistical image reconstruction algorithms using paraboloidal surrogates for PET transmission scans*, Ph.D. thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI., July 1999.
[3] J. A. Fessler and H Erdoğan, "A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1998, vol. 2, pp. 1132–5.
[4] D. F. Yu, J. A. Fessler, and E. P. Ficaro, "Maximum likelihood transmission image reconstruction for overlapping transmission beams," *IEEE Tr. Med. Im.*, vol. 19, no. 11, pp. 1094–1105, Nov. 2000.
[5] M. W. Jacobson and J. A. Fessler, "Properties of optimization transfer algorithms on convex feasible sets," In preparation.
[6] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Tr. Sig. Proc.*, vol. 42, no. 10, pp. 2664–77, Oct. 1994.
[7] 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, Oct. 1995.
[8] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Tr. Med. Im.*, vol. 1, no. 2, pp. 113–22, Oct. 1982.
[9] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, no. 2, pp. 306–16, Apr. 1984.
[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
[11] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, 1983.
[12] D. Nettleton, "Convergence properties of the EM algorithm in constrained parameter spaces," *The Canadian Journal of Statistics*, vol. 27, no. 3, pp. 639–48, 1999.