JOINT ESTIMATION OF IMAGE AND DEFORMATION PARAMETERS IN TOMOGRAPHIC IMAGE RECONSTRUCTION

M.W. Jacobson and J.A. Fessler

The University of Michigan
Electrical Engineering and Computer Science
1301 Beal Ave.
Ann Arbor, MI 48109-2122

ABSTRACT

We consider an emission tomography reconstruction problem in which projection measurements from several successive 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, parameters to describe the effects of motion are added to the statistical model of the projections. Joint maximum likelihood estimation of image and motion parameters is then carried out.

1. INTRODUCTION

To account for motion effects in Positron Emission Tomography (PET), one may break the scanning period into time frames $= 0, 1, 2, \ldots, K$ in which separate measured projection vectors $\{y^\kappa\}_{\kappa=0}^K$ are acquired by the tomograph. We assume the durations of these time frames are sufficiently short so that the motion of the object over the length of each time frame can be neglected. These measurement vectors then provide "snapshots", in a sense, of the radio-tracer activity image at different points throughout the scan. In this work, we examine different strategies for reconstructing an image of the object based on the measured data $\{y^\kappa\}_{\kappa=0}^K$. We model the measurements as Poisson with ensemble means $\{\bar{y}^\kappa\}$ given by,

$$\bar{y}^0(\lambda) = P^0 \lambda + r^0 \tag{1}$$

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

Here λ is the vector of unknown activity image pixel values in time-frame $\kappa = 0$. Also, for each time-frame k, 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 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 k, after undergoing changes due to anatomical motion and other physical processes. 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.

We consider two main reconstruction strategies. The first, is to obtain estimates $\hat{\lambda}$ and $\hat{\alpha}^\kappa$, $\kappa = 1, \ldots, K$ according to

$$(\hat{\lambda}, \hat{\alpha}^1, \ldots, \hat{\alpha}^K) = F(y^1, \ldots, y^K), \tag{3}$$

where $F(\cdot)$ is a penalized maximum likelihood estimator operating on measurements from all frames. We refer to this as the Joint Estimation with Deformation Modeling (JEDM) strategy. The second, which we call the Frame-Wise Post Registration (FWPR) strategy proceeds as follows,

1. For each frame $\kappa = 0, 1, \ldots, K$ obtain an estimate $\hat{\lambda}^\kappa = f^\kappa(y^\kappa)$ of $W^\kappa(\alpha^\kappa)\lambda$ where the $f^\kappa(\cdot)$ are penalized maximum likelihood estimators.

2. For frames $\kappa = 1, \ldots, K$ obtain a least squares estimate

$$\hat{\alpha}^\kappa = \operatorname*{argmin}_{\alpha} ||W^\kappa(\alpha)\hat{\lambda}^0 - \hat{\lambda}^\kappa||^2 \tag{4}$$

for the deformation parameter vectors.

3. Consolidate the $\hat{\lambda}^\kappa$ according to

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

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

In FWPR, the deformation parameters are not obtained from the Poisson model, but rather by post-reconstruction fitting using (4)[1].

---

[1] As noted by a reviewer, the relationship between these two methods is similar to pre-detection vs. post-detection integration in radar.

2. COST FUNCTION MINIMIZATION

In statistical tomographic imaging, the application of maximum (penalized) likelihood estimation involves minimizing an (often non-convex) cost function $\Phi(\theta)$, one term of which is the negative of the Poisson loglikelihood function, $L(\theta)$. Ignoring, irrelevant constants, the latter has the form

$$L(\theta) = \sum_i [\bar{y}_i(\theta) - y_i \log \bar{y}_i(\theta)], \qquad (6)$$

where $\bar{y}_i(\theta)$ is the Poisson mean of the measurements y_i and θ is a generic vector of unknown parameters. In the model described in (1) and (2), we have $\theta = (\lambda, \alpha)$.

A method for attacking such problems that has received considerable attention in recent years (e.g. [1, 2, 3, 4]) is that of optimization transfer. This technique involves constructing a so-called surrogate function $\phi(\cdot, \bar{\theta})$ satisfying $\phi(\theta, \bar{\theta}) \geq \Phi(\theta)$ for all θ, with equality at $\bar{\theta}$. Henceforth, we abbreviate this relationship, by writing

$$\phi(\cdot; \bar{\theta}) \overset{\bar{\theta}}{\succ} \Phi. \qquad (7)$$

The EM algorithm can be viewed as a particular case of optimization transfer (as explained, for example, [5]). The minimizer θ^* of $\phi(\cdot, \bar{\theta})$ has the property that $\Phi(\theta^*) \leq \Phi(\bar{\theta})$. Applying this iteratively, one obtains a sequence $\{\theta^n\}$ that monotonically reduces Φ and whose limit points are stationary under fairly weak conditions [6]. In this section, we discuss how to devise optimization transfer algorithms for the present problem.

2.1. Algorithm structure

In this subsection, we propose an optimization transfer algorithm that alternately minimizes the (penalized) loglikelihood $\Phi(\lambda, \boldsymbol{\alpha})$ (corresponding to (1) and (2)) with respect to λ and $\boldsymbol{\alpha} = (\alpha^1, \dots, \alpha^K)$. This is the space alternating technique described in [7, 8, 6]. In what follows the functions $\phi_1(\lambda; \bar{\lambda}, \bar{\boldsymbol{\alpha}})$ and $\phi_2(\alpha; \bar{\lambda}, \bar{\boldsymbol{\alpha}})$ are surrogates for Φ with respect to their first arguments, i.e.,

$$\phi_1(\cdot; \bar{\lambda}, \bar{\boldsymbol{\alpha}}) \overset{\bar{\lambda}}{\succ} \Phi(\cdot, \bar{\boldsymbol{\alpha}}) \qquad (8)$$

$$\phi_2(\cdot; \bar{\lambda}, \bar{\boldsymbol{\alpha}}) \overset{\bar{\boldsymbol{\alpha}}}{\succ} \Phi(\bar{\lambda}, \cdot). \qquad (9)$$

The algorithm has the following basic structure

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.

To implement the algorithm, we must be able to find functions $\phi_1(\cdot; \cdot)$ and $\phi_2(\cdot; \cdot)$ satisfying (8) and (9) and which are easy to minimize. Some tools for doing so are presented in the next subsection.

2.2. Construction of surrogates for general Poisson loglikelihoods

A generic penalized Poisson loglikelihood function has the form $\Phi(\theta) = L(\theta) + R(\theta)$, where $L(\theta)$ is as in (6) and $R(\theta)$ corresponds to penalty terms. We wish to find a surrogate function, i.e., a function $\phi(\cdot; \cdot)$ so that (7) holds. Preferably, $\phi(\cdot; \bar{\theta})$ will be convex to facilitate its minimization. To find a surrogate, it is sufficient to find convex surrogates for the separate terms L and R and add these two surrogates together. It is usually a simple matter to find a convex surrogate for common penalties $R(\theta)$ (see, for example,[2]). The problem therefore reduces to one of finding a surrogate for the negative Poisson loglikelihood $L(\theta)$.

However, $L(\theta)$ is often a difficult, non-convex function and finding a convex surrogate for it is a non-routine matter. For certain Poisson loglikelihoods arising in emission and transmission tomography (e.g. [1, 4]), methods for constructing quadratic surrogates have been presented. These methods do not extend in any obvious way to the loglikelihood associated with (1) and (2). We shall therefore pause to present some additional techniques.

The first is encapsulated in the next proposition. Recall that the Kullback-Leibler (KL) distance between two arrays a and b with non-negative components a_{im} and b_{im} is,

$$\text{KL}(a, b) = \sum_{i,m} \left[a_{im} \log \frac{a_{im}}{b_{im}} + b_{im} - a_{im} \right] \qquad (10)$$

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 2.1 (Separation of Additive Poisson Means)
Suppose the means of Poisson measurements has an additive form

$$\bar{y}_i(\theta) = \sum_m^{M_i} \bar{y}_{im}(\theta) \qquad (11)$$

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

$$y_{im}(\theta) = y_i \frac{\bar{y}_{im}(\theta)}{\bar{y}_i(\theta)}$$

$$Q(\theta; \bar{\theta}) = L(\theta) + KL(y_{im}(\bar{\theta}), y_{im}(\theta)) \qquad (12)$$

$$= \sum_{i,m} [\bar{y}_{im}(\theta) - y_{im}(\bar{\theta}) \log \bar{y}_{im}(\theta)] + \text{const.}$$

$$(13)$$

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

Proof. The fact that $Q(\cdot; \bar{\theta}) \overset{\bar{\theta}}{\succ} L(\cdot)$ is an easy consequence of (12) and the properties of the KL distance. The fact that $Q(\theta; \bar{\theta})$ also has the form (13) can be verified by expanding the term $\text{KL}(y_{im}(\bar{\theta}), y_{im}(\theta))$ in (12). □

Remark 2.2 When $\bar{y}_i(\theta) = \sum_{im} p_{im}\theta_m$, then applying Proposition 2.1 with $\bar{y}_{im}(\theta) = p_{im}\theta_m$ one obtains the surrogate used in the analyses of the EMML algorithm in[9, 10].

From (13), one can see that $Q(\theta; \bar{\theta})$ has the form of a Poisson loglikelihood function (ignoring irrelevant constants) with measured data $y_{im}(\bar{\theta})$ and means $\bar{y}_{im}(\theta)$. However, the terms $\bar{y}_{im}(\theta) - y_{im}(\bar{\theta})\log\bar{y}_{im}(\theta)$ are often simpler than the original terms $\bar{y}_i(\theta) - y_i\log\bar{y}_i(\theta)$. For example, when $\bar{y}_{im}(\theta)$ has an exponential form $b_{im}\exp(\langle a_im, \theta\rangle)$, as in [4], then the terms $\bar{y}_{im}(\theta) - y_{im}(\bar{\theta})\log\bar{y}_{im}(\theta)$ are convex, whereas the original terms in (6) are not.

Even when the $\bar{y}_{im}(\theta)$ do not have such an exponential form, it may still be the case that $\log\bar{y}_{im}(\theta)$ is concave. In this case, a convex surrogate can be obtained as indicated in the following Proposition.

Proposition 2.3 *Suppose the assumptions of Proposition 2.1 are satisfied and, in addition, all of the functions $\log\bar{y}_{im}(\theta)$ are concave. Then, apart from irrelevant constants,*

$$\phi(\theta; \bar{\theta}) = \sum_{i,m}[\bar{y}_{im}(\bar{\theta})\exp\left(\left\langle \frac{\nabla\bar{y}_{im}(\bar{\theta})}{\bar{y}_{im}(\bar{\theta})}, (\theta - \bar{\theta})\right\rangle\right)$$
$$- y_{im}(\bar{\theta})\log\bar{y}_{im}(\theta)] \qquad (14)$$

is a convex surrogate for $L(\theta)$.

Proof. As established in Proposition 2.1, the expression on the right hand side of (13) is, ignoring irrelevant constants, a surrogate for $L(\theta)$. It is therefore sufficient to show that

$$\bar{y}_{im}(\bar{\theta})\exp\left(\left\langle \frac{\nabla\bar{y}_{im}(\bar{\theta})}{\bar{y}_{im}(\bar{\theta})}, (\theta - \bar{\theta})\right\rangle\right) \overset{\bar{\theta}}{\succ} \bar{y}_{im}(\theta). \quad (15)$$

Using the assumed concavity of $\log\bar{y}_{im}(\theta)$ and the monotonicity of $\exp(\cdot)$,

$$\bar{y}_{im}(\theta) = \exp(\log\bar{y}_{im}(\theta))$$
$$\leq \exp(\log\bar{y}_{im}(\bar{\theta}) + \langle\nabla\log\bar{y}_{im}(\bar{\theta}), (\theta - \bar{\theta})\rangle)$$
$$= \bar{y}_{im}(\bar{\theta})\exp\left(\left\langle \frac{\nabla\bar{y}_{im}(\bar{\theta})}{\bar{y}_{im}(\bar{\theta})}, (\theta - \bar{\theta})\right\rangle\right)$$

and equality holds at $\bar{\theta}$, verifying (15). □

In the next subsection, we apply these results to the loglikelihoods associated with (1) and (2).

2.3. A surrogate for the deformed emission loglikelihood

We use the results of the previous section to develop surrogates to the loglikelihood corresponding to (1) and (2). These may be used in the algorithm of Subsection 2.1 once a penalty surrogate is added.

With $\boldsymbol{\alpha} = \bar{\boldsymbol{\alpha}}$ held fixed and applying Proposition 2.1 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

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

As noted in Remark 2.2, using this in step 2 in the algorithm of Subsection 2.1 is the EMML surrogate with respect to λ. Similarly, with $\lambda = \bar{\lambda}$ held fixed and applying Proposition 2.1 to the additive form $\bar{y}_i^\kappa(\boldsymbol{\alpha}) = \sum_{jk} P_{ij}^\kappa W_{jk}^\kappa(\alpha^\kappa)\bar{\lambda}_k$ one obtains,

$$\phi_2(\boldsymbol{\alpha}; \bar{\lambda}, \bar{\boldsymbol{\alpha}}) = \sum_{i\kappa}\bar{y}_i^\kappa(\lambda)$$
$$- \sum_{jk}\left(\sum_{i\kappa}\frac{P_{ij}^\kappa}{\bar{y}_i^\kappa(\bar{\lambda}, \bar{\alpha}^\kappa)}\right)\bar{\lambda}_k W_{jk}^\kappa(\bar{\alpha}^\kappa)\log W_{jk}^\kappa(\alpha^\kappa)$$

Since the matrices $W^\kappa(\alpha^\kappa)$ are typically sparse, this expression for ϕ_2 is of moderate complexity.

The surrogate $\phi_1(\lambda; \bar{\lambda}, \bar{\boldsymbol{\alpha}})$ obtained above is convex, however $\phi_2(\boldsymbol{\alpha}; \bar{\lambda}, \bar{\boldsymbol{\alpha}})$ may not be. One can apply Lemma 2.3 to obtain a convex surrogate if $\log W_{jk}^\kappa(\alpha^\kappa)$ is convex. This is a situation that is encountered when the deformations are modeled using log concave interpolators, e.g., raised cosine interpolators.

3. PRELIMINARY EXPERIMENTS

We tested 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 $K = 2$ time frames. In time frame $\kappa = 2$, 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 count levels of 0.3 and 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, maximum likelihood estimation was carried out without a roughness penalty. To better serve the purpose of these preliminary experiments, likelihood maximizations were not attempted using the optimization transfer technique approach proposed in Section 2.1 the previous section. Instead, we discretized the space of α^1 values into 22 values from 5% to 15%. Maximization over λ was done for each fixed α^1 value using EMML and subsequently over α^1 by exhaustive search. Since the negative loglikelihood is convex as a

Table 1. Performance data for reconstructions.

Average Counts Per Bin Per Time-Frame	Estimated Quantity	Method	Bias	Std. Dev.	%RMS Error	Std. Error (of Mean)	Std. Error (of Std. Dev.)
0.3	Uptake, U	FWPR	-3.88	6.46	30.27	0.91	0.65
		JEDM	-6.50	8.32	42.42	1.18	0.83
		EMML	-7.83	7.85	44.60	1.11	0.78
	Stretch, α^1	FWPR	1.33	3.52	37.34	0.50	0.65
		JEDM	-6.46	5.42	83.98	0.77	0.83
3	Uptake, U	FWPR	-1.35	3.23	14.05	0.46	0.32
		JEDM	-0.72	3.28	13.45	0.46	0.33
		EMML	-6.56	3.22	29.51	0.46	0.32
	Stretch, α^1	FWPR	0.31	2.45	24.46	0.35	0.32
		JEDM	-0.35	0.85	9.08	0.12	0.33

function of λ, this approach ensured that the iterates could not be trapped at a sub-optimal local minima.

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 1. For comparison, we have also reported the performance of EMML as applied to the total acquired data $\sum_{\kappa=1}^{K} y^{\kappa}$, i.e., when object motion is not accounted for. We see that FWPR is the most robust to very low count levels. As the count level increases, however, JEDM seems to overtake FWPR, while plain EMML falls distinctly behind.

In future work, we shall explore these trends for more elaborate deformation models and, additionally, evaluate the practical performance of the optimization techniques discussed in this article.

4. 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] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *J. Computational and Graphical Stat.*, vol. 9, no. 1, pp. 1–20, Mar. 2000.

[6] M. W. Jacobson and J. A. Fessler, "Properties of optimization transfer algorithms on convex feasible sets," In preparation.

[7] 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.

[8] 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.

[9] C. Byrne, "Likelihood maximization for list-mode emission tomographic image reconstruction," *IEEE Tr. Med. Im.*, vol. 20, no. 10, pp. 1084–92, Oct. 2001.

[10] A R De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Tr. Med. Im.*, vol. 12, no. 2, pp. 328–33, June 1993.