A FAST FULLY 4D INCREMENTAL GRADIENT RECONSTRUCTION ALGORITHM FOR LIST MODE PET DATA

Quanzheng Li, Evren Asma, Richard M. Leahy

Signal and Image Processing Institute, Univ. of Southern California, Los Angeles, CA 90089

ABSTRACT

We present a fully four-dimensional, globally convergent, incremental gradient algorithm to estimate the continuous-time tracer density from list mode positron emission tomography (PET) data. The rate function in each voxel is modeled as an inhomogeneous Poisson process whose rate function can be reconstructed using a cubic B-spline basis. The rate functions are then estimated by maximizing the objective function formed by the sum of the likelihood of arrival times and spatial and temporal smoothness penalties. We first provide a computable bound for the norms of the optimal temporal basis function coefficients, and based on this bound we construct an incremental gradient algorithm that converges to the solution. Fully four-dimensional simulations demonstrate the convergence of the algorithm for a high count dataset on a 4-ring scanner.

1. INTRODUCTION

Dynamic PET imaging usually involves a sequence of contiguous acquisitions which are then independently reconstructed to form a set of images which can be visualized and used to estimate physiological parameters [1]. This approach involves selection of the set of acquisition times, where one must choose between collecting longer scans with good counting statistics but poor temporal resolution, or shorter scans that are noisy but preserve temporal resolution. Direct reconstruction from list mode data can avoid this trade-off since list mode data extremely high temporal resolution with full spatial resolution by storing event arrival times in addition to sinogram data. One of the difficulties in dynamic image reconstruction directly from list mode data is the large number of image parameters that need to be reconstructed. Furthermore, unlike static PET, reconstruction speed depends on the number of counts in the scan, which can be on the order of hundreds of millions. Therefore there is a need for fast and convergent algorithms that can take full advantage of the high spatial and temporal resolution in list mode data and reconstruct dynamic images with high spatio-temporal resolution. In this paper we present such an algorithm.

Snyder [2] developed a list mode expectation maximization - maximum likelihood (EM-ML) method for estimation of dynamic PET images using inhomogeneous Poisson processes. Our statistical model is similar to Snyder's approach but we work with rate functions formed as a linear combination of known basis functions. This allows a better representation of the dynamic activity seen in experimental data that is not well modeled by more restrictive physiological models. Barrett *et al*. [3] and Reader *et. al.* [4] also describe list mode maximum likelihood methods but for the estimation of a temporally stationary image.

In our previous work on image reconstruction from list mode data [5], we presented a penalized maximum likelihood (ML) approach for the estimation of time-activity curves. However, that method was intrinsically three-dimensional (x, y and t) and the planes along the axial dimension had to be reconstructed separately. Oblique sinogram data could not be used in the reconstruction unless it was rebinned into direct sinogram planes. The algorithm presented in this paper is fully four-dimensional and can use the entire sinogram data without any need for rebinning. Recently, Ahn and Fessler presented a convergent incremental gradient algorithm for penalized ML static PET reconstructions [10]. The incremental gradient methods are similar to the well known OSEM algorithm but differ in the fact that they are provably convergent under reasonable conditions. In this paper we extend this algorithm for dynamic reconstructions and present a globally convergent, penalized ML dynamic PET image reconstruction algorithm.

2. METHODS

2.1. Statistical Modeling using Inhomogeneous Poisson Processes

We model the positron emissions in each voxel in the volume as an inhomogeneous Poisson process [5]. We denote the rate function at voxel j by $\eta_j(t)$ and parameterize it using a set of temporal basis functions so that:

$$\eta_j(t) = \sum_l w_{jl} B_l(t), \eta_j(t) \geq 0 \; \forall t \tag{1}$$

With this formulation, the detection process at detector pair i is also an inhomogeneous Poisson process with rate function:

$$\lambda_i(t) = \sum_j \sum_\ell p_{ij} w_{j\ell} B_\ell(t) \tag{2}$$

where p_{ij} is the probability of an event at voxel j being detected at detector pair i. The continuous time log-likelihood function of event arrival times is then given by:

$$\Phi(\mathbf{w}) = \sum_{i=1}^{n_p} \sum_{k=1}^{x_i} \log \left(\sum_{\ell=1}^{n_b} \sum_{j=1}^{n_v} p_{ij} w_{j\ell} B_\ell(a_{ik}) \right) - \sum_{i=1}^{n_p} \left(\sum_{\ell=1}^{n_b} (\sum_{j=1}^{n_v} p_{ij} w_{j\ell}) \int_0^T B_\ell(t) dt \right) \tag{3}$$

This work was supported by National Institute of Biomedical Imaging and Bioengineering under Grant No. R01 EEB00363.

where a_{ik} is the arrival time of the k^{th} photon pair, $k = 1, \ldots, x_i$ at the i^{th} detector pair. We use a spatial smoothing function equivalent to the pair-wise quadratic penalty used previously in Bayesian estimation of static PET images [6]. Our temporal roughness penalty is in the form of integrated squared curvature. When cubic B-splines are used as temporal basis functions, both of these penalties become quadratic in the control vertices, resulting in the following objective function:

$$L(\mathbf{w}) = -\sum_{i=1}^{n_p}\sum_{k=1}^{x_i}\log\lambda_i(a_{ik}) + \sum_{i=1}^{n_p}\int_0^T \lambda_i(t)dt + \beta\frac{1}{2}\mathbf{w}^T\mathbf{S}\mathbf{w} + \gamma\frac{1}{2}\mathbf{w}^T\mathbf{T}\mathbf{w} \quad (4)$$

where $\mathbf{S}$ and $\mathbf{T}$ are the second derivative matrices of the quadratic spatial and temporal penalties and $\mathbf{w}$ is the large vector containing all control vertices at all voxels. Our dynamic PET penalized ML control vertex estimates are then given as the solution of the following optimization problem:

$$\mathbf{w}^* = arg\min_{\mathbf{w}\in\mathcal{W}} L(\mathbf{w}) \quad (5)$$

where $\mathcal{W} = \{\mathbf{w} | \sum_{\ell=1}^{n_b} w_{j\ell}B_\ell(t) \geq 0 \quad \forall j, t\}$.

2.2. The Incremental Gradient Algorithm

Our dynamic incremental gradient algorithm has the following update equation:

$$\mathbf{w}^{n,m} = P_{\mathcal{T}}[\mathbf{w}^{n,m-1} + \alpha_n\mathbf{D}(\mathbf{w}^{n,m-1})\nabla f_m(\mathbf{w}^{n,m-1})] \quad (6)$$

where n denotes the iteration number, m denotes the subiteration number, f_m denotes the m^{th} subobjective function, α_n denotes the stepsize and $\mathcal{T} \equiv \{\mathbf{w} | \sum_{\ell=1}^{n_b} w_{j\ell}B_\ell(t) \geq \epsilon \; for \; all \; j, \ell, t\}$ for some small $\epsilon > 0$. The elements of the diagonal matrix $[\mathbf{D}(\mathbf{w})]_{jl,jl} \equiv d_{jl}(\mathbf{w})$ are given according to the following rule:

$$d_{jl}(\mathbf{w}) = \begin{cases} \frac{w_{jl}}{p_{jl}} & -\frac{1}{3}U \leq w_{jl} < \frac{1}{3}U \\ \frac{U - w_{jl}}{2p_{jl}} & \frac{1}{3}U \leq w_{jl} \leq U \\ \frac{U + w_{jl}}{2p_{jl}} & -U \leq w_{jl} < -\frac{1}{3}U \end{cases}$$

We chose $p_{jl} \equiv \sum_{i=1}^{n_d} p_{ij}\int_0^T B_\ell(t)dt$ motivated by the EM algorithm in static PET and by observing that the sensitivity of the i^{th} detector pair to activity due to the ℓ^{th} basis function over the entire scan duration is given by $p_{ij}\int_0^T B_\ell(t)dt$ in our dynamic PET problem.

It can be shown that this update equation guarantees that all iterates lie inside the set $\mathcal{U}$ (de ned below) if the initial estimate is inside $\mathcal{U}$ but it does not necessarily guarantee that all iterates lie inside $\mathcal{W}$. The projection of the resulting estimate into the interior of the constraint set (i.e. $\mathcal{T}$) ensures that all iterates lie inside $\mathcal{W}$ and does not affect the convergence proof because the stepsize falls below ϵ when the iteration number is large enough, making the projection operation unnecessary.

2.3. Convergence Proof

2.3.1. Existence, Uniqueness and Differentiability

Since $L(\mathbf{w})$ is continuous and coercive, the level set $\mathcal{A} \equiv \{\mathbf{w} \in \mathcal{W} | L(\mathbf{w}) \leq L(\mathbf{1})\}$ is closed and bounded (i.e. compact). The compactness of A and the fact that it is non-empty (has at least one element which is the vector of all ones) guarantees the existence of a vector $\mathbf{w}$ such that $L(\mathbf{w}) = \inf_{\mathbf{w}\in\mathcal{W}} L(\mathbf{w})$ due to the Theorem of Weierstrass [11, Prop. A.8 (3)]. $L(\mathbf{w})$ can be shown to be strictly convex if either $\beta > 0$ or $\gamma > 0$ holds and $\mathbf{y}^T\mathbf{P}\mathbf{1} \neq 0$ which means that the solution is unique. We also note that $L(\mathbf{w})$ is not differentiable over the set $\mathcal{D} = \{\mathbf{w} | \; l_{ik}(\mathbf{w}) = 0\}$; however as in [10], it can be replaced with another objective function that is differentiable over the entire feasible set $\mathcal{W}$ and has the same global minimum as $L(\mathbf{w})$.

The convergence proof relies heavily on the determination of a bounded region in which the solution is guaranteed to lie. We determine this region in the next subsection.

2.3.2. Computable Bound on Optimal Control Vertices

It can be shown that $\mathcal{W}$ is a convex cone and that the objective function $L(\mathbf{w})$ is convex and differentiable. Therefore the second part of the lemma [8] (p. 227), which states that $\delta L(\mathbf{w}^*; \mathbf{w}^*) = 0$ where δL denotes the Fréchet derivative and $\mathbf{w}^*$ denotes the optimal solution, can be applied. Applying the lemma we get:

$$\sum_{i=1}^{N}\sum_{j=1}^{n_v}\sum_{\ell=1}^{n_b} p_{ij}w^*_{j\ell}\int_0^T B_\ell(t)dt - \sum_{i=1}^{N} y_i + \mathbf{w}^{*T}\mathbf{S}\mathbf{w}^* + \mathbf{w}^{*T}\mathbf{T}\mathbf{w}^* = 0 \quad (7)$$

Since $\mathbf{w}^{*T}\mathbf{S}\mathbf{w}^* \geq 0$, $\mathbf{w}^{*T}\mathbf{T}\mathbf{w}^* \geq 0$ due to the positive de - niteness of $\mathbf{S}$ and $\mathbf{T}$, we have:

$$\sum_{i=1}^{N}\int_0^T r_i^*(t)dt \leq \sum_{i=1}^{N} y_i \quad (8)$$

where $r_i^*(t) \equiv \sum_j\sum_\ell p_{ij}w^*_{j\ell}B_\ell(t)$ denote the reconstructed rate functions at the detectors (i.e. forward projections of reconstructed rate functions). Since the sum of N positive terms on the left hand side of (8) is less than or equal to the sum of the N positive terms on the right hand side, we have:

$$\min_i \int_0^T r_i^*(t)dt \leq \max_i y_i \quad (9)$$

which can be simpli ed and expressed in terms of the integrals of reconstructed rate functions $\lambda_j^*(t)$:

$$\int_0^T \lambda_j^*(t)dt \leq \frac{\max_i y_i}{\min_{ij,\; p_{ij}\neq 0} p_{ij}} \quad \forall j \quad (10)$$

Using the lemma in [9] (p. 155), which provides a bound to the norms of the control vertices of a B-spline in terms of its maximum, and noting that the integrals of reconstructed rate functions are the maxima of the higher order B-splines corresponding to their running integrals, we can convert (10) to a bound on the norms of

the control vertices of the integrals of the reconstructed rate functions:

$$|\beta_{j\ell}^*| \leq D_{k+1,\infty} \frac{\max_i y_i}{\min_{ij,\ p_{ij}\neq 0} p_{ij}} \tag{11}$$

where $D_{k+1,\infty}$ is a constant that is independent of knot locations (i.e. only a function of the order of the B-splines). The control vertices of the integral of a B-spline basis $\{\beta_{j\ell}\}$ are related to the control vertices of this B-spline basis of order k, $w_{j\ell}$, as follows [9]:

$$\beta_{j\ell} = \beta_{j\ell-1} + w_{jl}(t_{\ell+k} - t_\ell)/k \qquad \forall \ell \tag{12}$$

which implies:

$$|w_{j\ell}^*| \leq \frac{2 \cdot k \max_\ell |\beta_{j\ell}^*|}{\min_\ell (t_{\ell+k} - t_\ell)} \qquad \forall j, \ell \tag{13}$$

Combined with (11) we arrive at the computable bound on the norms of the optimal control vertices:

$$|w_{jl}^*| \leq \frac{2kD_{k+1,\infty}(\max_i y_i)}{[\min_{ij,\ p_{ij}\neq 0} p_{ij}][\min_\ell (t_{\ell+4} - t_\ell)]} \equiv V \tag{14}$$

For cubic B-splines, $k = 4$ and $D_{5,\infty} = \frac{35}{3}$ [9]. We also de ne $U = V + \zeta$ where $\zeta > 0$ is a small positive number (e.g. ζ = .001*V). This gives the strict inequality $|w_{j\ell}{}^*| < U$ and is helpful in the proof of Lemma 4 in the next subsection. We also de ne two useful sets:

$$\mathcal{U} \equiv \{\mathbf{w} | -U \leq w_{j\ell} \leq U,\ \forall j, \ell\} \tag{15}$$

$$\mathcal{B} \equiv \{\mathbf{w} | \sum_{\ell=1}^{n_b} w_{j\ell} B_\ell(t) \geq 0, -U \leq w_{j\ell} \leq U,\ \forall j, \ell, t\} \tag{16}$$

where $\mathcal{U}$ is the bounded set which will include the solution in its interior and $\mathcal{B} = \mathcal{W} \cap \mathcal{U}$ is the bounded, convex set where the solution is guaranteed to lie.

2.3.3. Convergence

In this subsection we outline the global convergence proof of our dynamic incremental gradient algorithm. The following lemma, for which we omit the proof, shows that if the initial estimate is inside the feasible set, then the remaining points generated by the algorithm are also guaranteed to remain inside the feasible set:

Lemma 1 *Suppose that* $\mathbf{w}^{n,m}$ *is a sequence generated by the update equation with* $\mathbf{w}^0 \in Int\ \mathcal{B}$. *Then there exists* $\alpha_0 > 0$ *such that if* $0 < \alpha_n < \alpha_0, \forall n$, *then* $\mathbf{w}^{n,m} \in Int\ \mathcal{B}\ \forall n, m$.

The next lemma, whose proof in [10] applies directly to our dynamic PET problem, shows that under certain conditions imposed on the stepsize, the images generated by the algorithm converge to a limit point in $\mathcal{B}$ which satis es $D(\mathbf{w}^*)\nabla L(\mathbf{w}^*) = \mathbf{0}$. In particular, we have the following:

Lemma 2 *Suppose that* $\{\mathbf{w}^n\}$ *is a sequence generated by (24) with* $\alpha_n > 0$ *such that* $\sum_{n=0}^{\infty} \alpha_n = \infty$ *and* $\sum_{n=0}^{\infty} \alpha_n^2 < \infty$. *If* $\mathbf{w}^{n,m} \in \mathcal{B}\ \ \forall n, m$ *then* $\{L(\mathbf{w}^n)\}$ *converges in* $\mathbf{R}$ *and there exists a limit point* $\mathbf{w}^* \in \mathcal{B}\ \ of\ \ \{\mathbf{w}^n\}$ *such that* $D(\mathbf{w}^*)\nabla L(\mathbf{w}^*) = \mathbf{0}$.

We now show that as the number of iterations tend towards in nity, the difference between the images in consecutive iterations go to zero:

Lemma 3 *Suppose that* $\{\mathbf{w}^n\}$ *is a sequence generated by (24) with* $\alpha_n > 0$ *such that* $\lim_{n\to\infty} \alpha_n = 0$. *If* $\mathbf{w}^{n,m} \in \mathcal{B}$ *for all n, m, then* $\lim_{n\to\infty}(\mathbf{w}^{n,m} - \mathbf{w}^n) = 0$ *for all m.*

The proof of this lemma is based on the boundedness of $D(\mathbf{w})\nabla f_m(\mathbf{w})$ on $\mathcal{B}$ for all m and α_n approaching 0. An immediate corollary obtained by setting $m = M$ is:

$$\lim_{n\to\infty} (\mathbf{w}^{n+1} - \mathbf{w}^n) = 0 \tag{17}$$

Our nal lemma shows that the limit point $\mathbf{w}^*$ in Lemma 2 is a maximizer of $L(\mathbf{w})$ over $\mathcal{B}$ if all iterates lie inside $\mathcal{B}$. In other words, it shows that $D(\mathbf{w}^*)\nabla L(\mathbf{w}^*) = \mathbf{0}$ implies $\nabla L(\mathbf{w}^*) = \mathbf{0}$. Although, this property holds by inspection for points at the interior of the constraint set, it can be shown that it also holds at the boundary.

Lemma 4 *The limit point* $\mathbf{w}^*$ *in Lemma 2 such that* $D(\mathbf{w}^*)\nabla L(\mathbf{w}^*) = \mathbf{0}$ *is a maximizer of* $L(\mathbf{w})$ *over* $\mathcal{B}$ *if* $\mathbf{w}^{n,m} \in \mathcal{B}\ \forall n, m$.

Combining these lemmas, we have:

$$\lim_{n\to\infty} \mathbf{w}^{n,m} = arg \min_{\mathbf{w}\in\mathcal{W}} L(\mathbf{w})\ \forall m \tag{18}$$

which proves the global convergence of the algorithm.

3. SIMULATIONS

We simulated a 4-ring scanner with the same geometry as the microPET R4 small animal scanner. The reconstructed image volume was $128 \times 128 \times 7$ and the sinogram dimensions were $84 \times 96 \times 13$ after rebinning the data with a span of 3. The time activity curve (TAC) at each voxel was simulated as a scaled version of the TAC shown in Figure 2. All 7 image planes had identical activity distributions. We used 11 B-splines with equally spaced knots as our temporal basis functions. We used repeated knots (4 knots for cubic B-splines) at both end-points so that the reconstructed rate functions were not restricted to be continuous or differentiable at the end points. Our current implementation of the algorithm uses a non-negativity penalty in place of the projection operation which is computationally intensive due to the objective function calculations in the line searches. In the future, we will replace the non-negativity penalty with an ef cient projection operation.

We reconstructed a 120M count dataset using 12 subsets and a stepsize function of $\alpha_n = 1.5/(n/2 + 1.5)$. Only the sinogram data was divided into subsets and the arrival times corresponding to those subsets were used in computing the gradients. The objective function values as a function of subiteration number are shown in Figure 1.

Figure 2 shows the true and reconstructed TACs at the center of the left cylinder in the central plane (i.e. voxel (64,43) of the fourth plane). The reconstructed TAC is in very good agreement with the true TAC. Figure 3 shows the true and reconstructed central plane images for the fourth control vertex together with the corresponding central transaxial pro les after 15 iterations. The algorithm shows good contrast recovery in both cylinders.

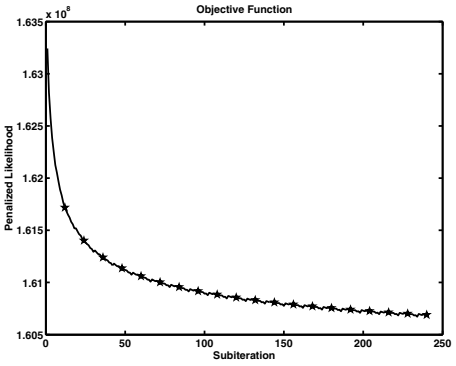

Fig. 1. Cost function values versus subiterations. The stars denote the cost function values at the end of each full iteration (12 subiterations).

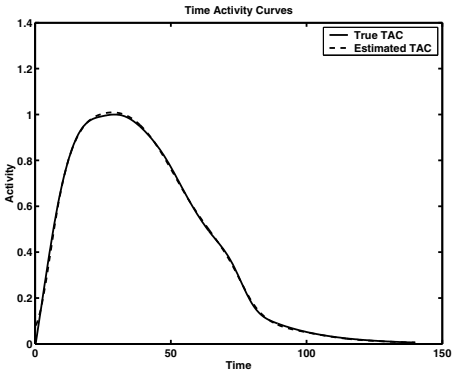

Fig. 2. The true TAC at voxel (64,43) in the fourth plane compared against the TAC reconstructed with our fully 4D dynamic incremental gradient algorithm.

4. CONCLUSIONS

We have proved the convergence of a fast fully four-dimensional dynamic image reconstruction algorithm. We also demonstrated its convergence by reconstructing a dynamic dataset from a simulated 4-ring scanner. In addition to its convergence properties, the algorithm is straightforward to implement because it only requires the computation of the gradient from a subset of the data and a projection onto the interior of the constraint set. Since the algorithm operates directly on list mode data, it can take advantage of the full spatio-temporal content of list mode data and reconstruct dynamic images with high spatio-temporal resolution. Since both gradient computation and projection (via line search) operations can be parallelized, the algorithm lends itself to multithreaded and distributed implementations. Such implementations can reduce the reconstruction times for real datasets to clinically acceptable levels.

5. REFERENCES

[1] S. Huang and M. Phelps, *Principles of tracer kinetic modeling in positron emission tomography and autoradiography, in Principles and Applications for the Brain and Heart* , New York: Raven, 1986, Positron Emission Tomography and Autoradiography, ch. 7.

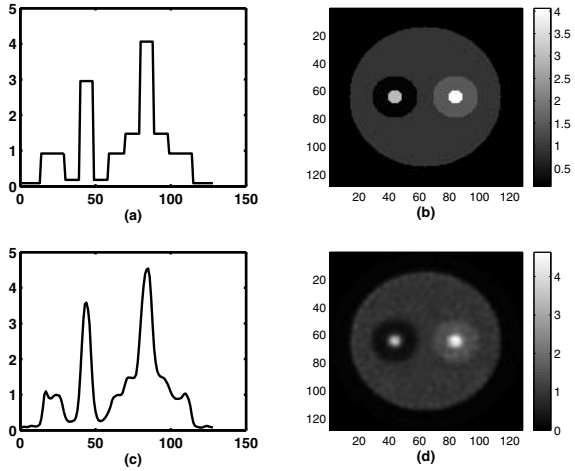

Fig. 3. True (b) and reconstructed (d) control vertex images and their corresponding central pro les (a and c) for the fourth control vertex.

[2] D. Snyder, *Parameter estimation for dynamic studies in emission-tomography systems having list-mode data*, IEEE Trans. Nucl. Sci., vol.31, pp. 925931, Apr. 1984.

[3] L. Parra and H. Barrett, *List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET*, IEEE Trans. Med. Imag.,vol. 17, pp. 228-235, Apr. 1998.

[4] A. J. Reader, S. Ally, F. Bakatselos, R. Manavaki, R.J. Walledge, A.P. Jeavons, P.J. Julyan, Zhao Sha, D.L. Hastings, J. Zweit: *One-pass list-mode EM algorithm for high-resolution 3-D PET image reconstruction into large arrays* IEEE Trans. Nuc. Sci., vol. 49, no. 3, June 2002, pp. 693-699

[5] T.E. Nichols, J. Qi, E. Asma, R.M. Leahy, *Spatiotemporal reconstruction of list mode PET data* , IEEE Transactions on Medical Imaging, vol.21, No.4, April 2002

[6] J. Qi, R. Leahy, S. Cherry, A. Chatziioannou, and T. Farquhar, *High resolution 3D Bayesian image reconstruction using the microPET small animal scanner* , Physics in Medicine and Biology, vol. 43, no. 4, pp. 1001 1013, 1998.

[7] A. Nedic, D. P. Bertsekas: *Incremental subgradient methods for nondifferentiable optimization*. SIAM J. OPTM. vol. 12, pp109-138, 2001

[8] D. Luenberger: *Optimization by Vector Space Methods*, Wiley Interscience, 1968

[9] C. de Boor: *A Practical Guide to Splines*, Springer-Verlag, 1978

[10] S. Ahn and J.A. Fessler, *Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms*, IEEE Transactions on Medical Imaging, Vol. 22, No.5 pp. 613-626, 2003.

[11] D. Bertsekas, *Nonlinear Programming*, Athena Scienti c, 1999