Multiresolution constrained least-squares algorithm for direct estimation of time activity curves from dynamic ECT projection data

Jonathan S. Maltz

Center for Functional Imaging,
Lawrence Berkeley National Laboratory, University of California, Berkeley CA 94720

ABSTRACT

We present an algorithm which is able to reconstruct dynamic emission computed tomography (ECT) image series directly from inconsistent projection data that have been obtained using a rotating camera. By finding a reduced dimension time-activity curve (TAC) basis with which all physiologically feasible TAC's in an image may be accurately approximated, we are able to recast this large non-linear problem as one of constrained linear least squares (CLLSQ) and to reduce parameter vector dimension by a factor of 20. Implicit is the assumption that each pixel may be modeled using a single compartment model, as is typical in ^{99m}Tc teboroxime wash-in wash-out studies; and that the blood input function is known. A disadvantage of the change of basis is that TAC non-negativity is no longer ensured. As a consequence, non-negativity constraints must appear in the CLLSQ formulation. A warm-start multiresolution approach is proposed, whereby the problem is initially solved at a resolution below that finally desired. At the next iteration, the number of reconstructed pixels is increased and the solution of the lower resolution problem is then used to warm-start the estimation of the higher resolution kinetic parameters. We demonstrate the algorithm by applying it to dynamic myocardial slice phantom projection data at resolutions of 16×16 and 32×32 pixels. We find that the warm-start method employed leads to computational savings of between 2 and 4 times when compared to cold start execution times. A 20% RMS error in the reconstructed TAC's is achieved for a total number of detected sinogram counts of 1×10^5 for the 16×16 problem and at 1×10^6 counts for the 32×32 grid. These errors are $1.5 - 2$ times greater than those obtained in conventional (consistent projection) SPECT imaging at similar count levels.

Keywords: dynamic ECT, dynamic SPECT, myocardial SPECT, kinetic parameter estimation, compartmental modeling, multiresolution methods, constrained linear least squares, quadratic programming

1. INTRODUCTION

In ECT modalities such as dynamic single photon emission tomography (SPECT), the signal to noise ratio (SNR) of the acquired data is proportional to both the imaging time and the amount of injected radiotracer. Since it is important to keep the patient radiation dose low, lengthening the imaging time is the preferred method of increasing the total number of detected photon counts and hence the SNR. In ECT systems which are unable to measure a full set of angular projections simultaneously, projection data acquired over a long time interval are inconsistent since each angular projection is ostensibly a projection of a different distribution. As a result, standard algorithms which perform tomographic reconstruction via inversion of the Radon transform will introduce artefacts into the resulting image. Since the single-headed SPECT camera is the most commonly available instrument in nuclear medicine, the development of algorithms capable of dynamic reconstruction from inconsistent projections (DRIP) is an important clinical goal.

Methods for the direct estimation of regional time-activity curves (TAC's) from projections may be grouped into two main classes:

1.1. Kinetic parameter estimation algorithms

These methods assume that a complete segmentation of the imaged distribution exists *a priori*. Since the contribution of each dynamic region to each projection bin is thus known, a kinetic model for each region may be fit to the projection measurements. Implementations differ by their segmentation regime and specification of tracer kinetic model.

Figure 1. RMS intensity estimation error versus count rate and number of resolution elements in conventional (static) SPECT imaging (Budinger 1994).

1.1.1. Geometric Model

A conventional rectangular pixel segmentation is appealing for reasons of simplicity and familiarity. Unfortunately, the SNR of clinical dynamic ECT data is low, and consequently only a limited number of resolution elements may be resolved for a given number of total detected events. Guidelines for conventional (non-dynamic) SPECT appear in Figure 1 (Budinger 1994). This problem is aggravated in dynamic ECT, where the objective is the determination of the rates of wash-in and wash-out of radiotracer material from various regions. During the late wash-out phase, for example, count rates can drop to very low levels. Thus, the main drawback of the pixel-based approach for dynamic ECT is that the statistics of the data are unable to support reconstruction of fine pixel grids.

Rather than maintain a fixed resolution throughout the image, regions of interest (ROI's) can be defined within the image space. For such methods to be successful, the entire imaged distribution must be modeled using a reasonably accurate segmentation geometry. This is difficult, since the geometry of the sources need not necessarily correspond to that of anatomical structures whose boundaries might be easily delineated within the ECT reconstruction space using a coregistered magnetic resonance or transmission tomography image. Indeed, it is the determination of the emission source geometry that constitutes the primary goal of the imaging process.

1.1.2. Kinetic model

The kinetics of injected radiotracers are usually described using compartmental models. Since most of the algorithms proposed for the solution of DRIP are applied to dynamic myocardial SPECT studies, we will concentrate on this model. The tracer ^{99m}Tc is commonly used in such studies, owing to the fact that distinct wash-in and wash-out stages are observed in the myocardium minutes after its injection. The wash-in part of the TAC has been shown to correlate with blood perfusion in the myocardium and is consequently a valuable indicator of myocardial viability (Smith, Gullberg, Christian & Datz 1994). During the wash-out phase, the tracer accumulates in the liver, the TAC of which may also be modeled using a single compartment model.

The time-activity within region n, which follows the single compartment model shown in Figure 2, is given by:

$$\phi_n(t) = k_1^n \, i(t) * \mathrm{e}^{-k_2^n t}, \tag{1}$$

where $i(t)$ is the measured blood input function, and the '$*$' operator denotes convolution.

1.1.3. Implementations of fixed geometry kinetic parameter estimation algorithm

Limber et al. proposed a method of kinetic parameter estimation for DRIP whereby regional TAC's are reconstructed on a pixel grid. This method performs non-linear regression on the projection measurements, and solves for wash-out parameters only. The kinetic model used is equivalent to (1) when $i(t) = \delta(t)$, the Dirac delta distribution. The efficacy of this algorithm was demonstrated through application to a 16×16 resolution grid (Limber, Limber, Celler, Barney & Borwein 1995).

Farncombe et al. increased the flexibility of this method by allowing the reconstructed TAC's to assume any form, so long as all TAC samples remain non-negative, with the added constraint that the TAC's be monotonically decreasing sequences. This is similar to the model used in the algorithm above, in that only the wash-out phase is modeled. The method has the potential to reconstruct the wash-in and wash-out phases as separate problems, but this is not practical when the extrema of several TAC's present in the imaged distribution occur at different abscissae (Farnocombe, Noll, Maeght & Harrop 1999).

Reutter et al. have demonstrated an algorithm capable of fitting single compartment models to both phantom and clinical myocardial studies (Reutter, Gullberg & Huesman 1998, Reutter, Gullberg & Huesman 1999*b*). A stabilized Newton-Raphson optimization algorithm is used to solve the non-linear weighted least squares problem whose solution yields the kinetic parameters directly from the acquired projection data. While this method is effective in providing the desired estimates, the amount of computation required is large for studies involving many dynamic regions.

Two recent algorithms, those of Reutter et al. and Maltz et al. are not reviewed here owing to their similarity to the method we now propose. They will be discussed in later sections (Maltz, Reutter, Huesman & Budinger 1999*b*, Reutter, Gullberg & Huesman 1999*a*).

1.2. Joint geometric and kinetic parameter estimation algorithms

These methods simultaneously estimate the geometry of the source distribution and their associated kinetic parameters.

Chiao et al. describe a maximum-likelihood (ML) method for jointly estimating the parameters of a specified compartmental model and the geometry of the myocardium, directly from projection data (Chiao, Rogers, Clinthorne, Fessler & Hero 1994). The method involves fitting a concentric polygonal myocardial model containing a single tracer compartment to projection data. Endo- and epicardial radii, together with the wash-in, wash-out and spill-over parameters of the compartment, are simultaneously estimated through the maximization of a likelihood function which regards each projection datum as the product of an independent Poisson process. The algorithm produces unbiased parameter estimates having variance close to the Cramér-Rao lower bound at realistic photon count levels.

Maltz et al. extended this method to deal with the case of multiple dynamic regions. Owing to the extremely ill-conditioned nature of the problem, a hybrid global-local optimization algorithm was needed to fit the spatiotemporal model to the single revolution projection data. For proof of concept, the ellipse was used as geometric modeling element. Single compartmental modeling was employed. The method able to fit 11 dynamic regions to a phantom with a useful degree of accuracy at realistic SNR's (Maltz, Polak & Budinger 1999*a*).

The extension of these methods to larger numbers of regions is currently infeasible owing to the vast computational burden required.

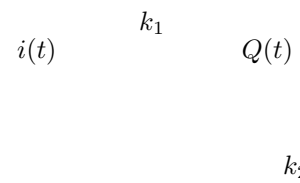

Figure 2. Compartmental model for ^{99m}Tc-teboroxime in the myocardium. Here, $Q(t)$ represents the tracer activity within the myocardium, while other symbols are defined within the text.

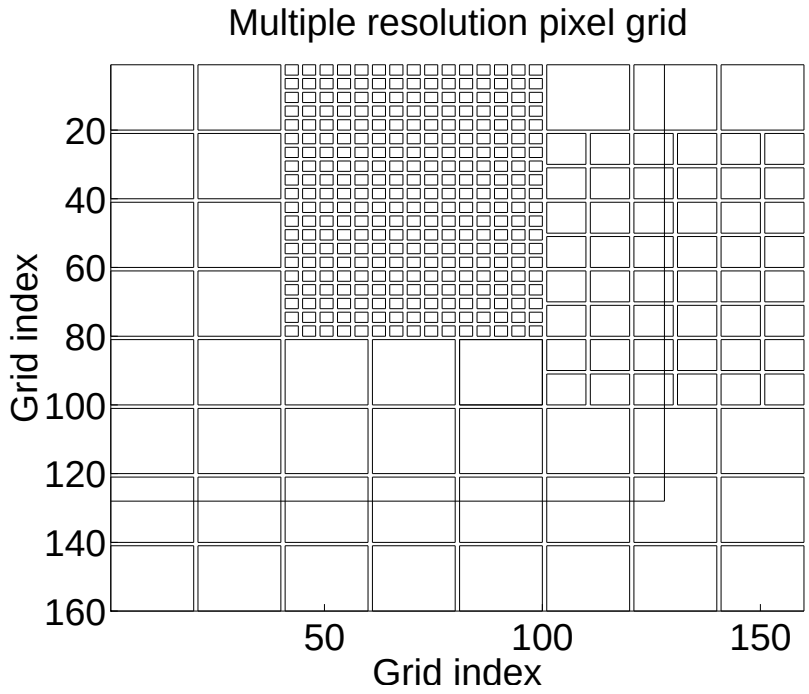

Figure 3. Multiresolution pixel grid. Such a grid is proposed in order to offer increased resolution in areas specifically of clinical interest, while reducing the overall problem dimension. While the pixels of the highest resolution region are not labeled in the figure, TAC's are similarly assigned to thereto.

2. PROBLEM FORMULATION:

At present, none of the methods outlined in Section 1 above is able to reconstruct images containing more than a handful of regions while modeling the tracer kinetics with adequate realism. While explaining the difficulties encountered in solving this problem, we will refer to the most common application of rotating camera ECT imaging, myocardial SPECT.

In previous work, we developed the convolved-orthogonal basis reconstruction algorithm (COBRA) for a general region model. While this algorithm produces good results with little bias and variance close to the Cramér-Rao bound in phantom experiments with around ten regions present, it is not practical for the reconstruction of images consisting of thousands of regions, as is considered here (Maltz et al. 1999*b*). This is because the previous COBRA algorithm does not enforce non-negativity constraints on the TAC values, and instead relies on the non-negativity of the projection data to keep the reconstructed TAC's non-negative. Preliminary experiments with a large number of regions (16×16 and finer pixel grids) revealed the limitations of this reliance, and the need to introduce constraints. Constraint formulation is therefore discussed in greater detail below. As well as describing the extensions to COBRA, we provide an abridged exposition of the original method here for convenience.

2.1. Geometric model

Without loss of generality, for purposes of illustration we refer to the 2D multiresolution pixel grid shown in Figure 3.

The TAC $\phi_n[\,l\,]$ is assigned to the nth region $\Omega_n(\mathbf{x})$, $n = 1, 2, \ldots, N$. In general, we require that the underlying source distribution $\Omega(\mathbf{x})$ be completely segmented into regions in this way.

2.2. Kinetic model

As kinetic model, we generalize the model given in (1) to the case where the TAC of each region may be composed of linear combinations of the responses of several underlying compartments. This may be useful for modeling tissue

heterogeneity, partial volume and spillover effects (O'Sullivan 1993). We thus have:

$$\phi_n(t) = \sum_{\tilde{m}=1}^{\tilde{M}} k_1^{\tilde{m}n}\, i(t) * \mathrm{e}^{-k_2^{\tilde{m}} t}, \tag{2}$$

where $\tilde{M}$ is the number of exponential modes (compartments) from which the TAC's may derive.

Since the algorithm will be based in discrete time, we regularly sample the imaging interval every T_s minutes giving:

$$\phi_n[\,l\,] = \phi_n(t)\,\delta(t - l\,T_s), \qquad l = 0, 1, \ldots, L-1 \tag{3}$$

where $\delta(t)$ is the Dirac delta distribution.

In (2) we see that each TAC depends non-linearly on the parameter k_2. As the projection measurements constitute sums of TAC values, they are composed of weighted sums of convolved exponential functions. The determination of the values of these parameters constitutes the extremely ill-conditioned problem of resolving the components of exponential sums (Lanczos 1956, Reich 1981). We wish to linearize the problem and improve its condition. To do this, we employ the exponential spectral method of Cunningham et al. and define a set of sampled exponential functions which span the range of physiologically feasible compartmental modes (Cunningham & Jones 1993) expected within the imaged distribution:

$$f_{\tilde{m}}[l] = \mathrm{e}^{-k_{\tilde{m}}\, l\, \Delta t}, \qquad l = 0, 1, \ldots, L-1. \tag{4}$$

We then form the $(L \times \tilde{M})$ matrix $\mathbf{X}$ whose $\tilde{m}$th column is $f_{\tilde{m}}[l]$ as defined in (4), and invoke the singular value decomposition (SVD) to find orthogonal basis vectors for the range of $\mathbf{X}$. These are the left singular (column) vectors $\mathbf{u}_{\tilde{m}}$ of the SVD of $\mathbf{X}$:

$$\mathbf{X} = \mathbf{V}\mathbf{S}\tilde{\mathbf{U}}^T, \qquad \tilde{\mathbf{U}} = \left(\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_{\tilde{M}}\right) \tag{5}$$

where $\mathbf{V}$ is the matrix of right singular vectors, and $\mathbf{S}$ is the diagonal matrix of singular values. We associate the discrete time index l with each row of $\mathbf{U}$. Depending on the degree of accuracy required in the sampled representation of the $\phi_n(t)$, we utilize only the first $M \leq \tilde{M}$ of $\tilde{\mathbf{U}}$ such that:

$$\mathbf{U} = \left(\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_M\right). \tag{6}$$

Typically, $M \approx 3$ is sufficient for the myocardial imaging applications we have studied.

We then form the matrix $\mathbf{C}$ from the columns of $\mathbf{U}$ convolved with the sampled blood input function $i[l]$, which we assume has either been measured or estimated:

$$\mathbf{C}' = \left(\mathbf{c}_1\ \mathbf{c}_2\ \cdots\ \mathbf{c}_M\right) \tag{7}$$

where $\mathbf{c}_m = \mathbf{u}_m * i[\,l\,],\ \ l = 0, 1, \ldots, L-1$. We refer to the $\mathbf{c}_m$ and convolved-orthogonal basis functions (COB's).

With the kinetic model formalized, we wish to estimate the coefficients μ_{mn} of the $\mathbf{c}_m$ for all regions, which form the approximated TAC's as:

$$\hat{\phi}_n[\,l\,] = \sum_{m=1}^{M} \hat{\mu}_{mn}\, \mathbf{c}_m[\,l\,], \qquad l = 0, 1, \ldots, L-1\,; \tag{8}$$

where $L = PR$, the total number of angular projections, given R camera rotations with P angular projections per rotation.

3. PROBLEM SOLUTION

We begin by lexicographically stacking the projections of the $RP \times Q$ measured sinogram $\tilde{\mathbf{Y}}$ into the vector $\tilde{\mathbf{y}}$, where each angular projection is made up of Q bins. Similarly the $P \times Q$ sinogram $\mathbf{Y}'_n$ for each of the N segmented regions $f_n(\mathbf{x})$ are stacked into the vectors $\mathbf{y}'_n$. We then define the $RPQ \times M$ matrices $\mathbf{G}_n$ which consist of $R \times M$ replicates of $\mathbf{y}'_\mathbf{n}$. The geometric weighting matrix for the activity contributions of each region is then given by:

$$\mathbf{G} = \left[\ \mathbf{G}_1 \quad \mathbf{G}_2 \quad \cdots \quad \mathbf{G}_N\ \right]. \tag{9}$$

The second matrix we will describe consists of blocks containing the convolved basis functions $\mathbf{c}_m$. For each time sample l, we form the $PQ \times M$ matrices

$$\mathbf{C}_l = \begin{bmatrix} u_1[l] & u_2[l] & \cdots & u_M[l] \\ u_1[l] & u_2[l] & \cdots & u_M[l] \\ \vdots & \vdots & \vdots & \vdots \\ u_1[l] & u_2[l] & \cdots & u_M[l] \end{bmatrix} \tag{10}$$

from which the $RPQ \times NM$ basis weighting matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_0 & \mathbf{C}_0 & \cdots & \mathbf{C}_0 \\ \mathbf{C}_1 & \mathbf{C}_1 & \cdots & \mathbf{C}_1 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{C}_{L-1} & \mathbf{C}_{L-1} & \cdots & \mathbf{C}_{L-1} \end{bmatrix} \tag{11}$$

is composed.

Finally, we define the matrix F as:

$$\tilde{\mathbf{y}} = \mathbf{F}\,\hat{\boldsymbol{\mu}} = (\mathbf{G} \cdot \mathbf{C})\,\hat{\boldsymbol{\mu}} \tag{12}$$

where the operator '$\cdot$' denotes element-by-element multiplication.

3.1. TAC non-negativity constraints

While the employment of the COB provides dimensionality reduction and linearizes the problem, it has the disadvantage of allowing the reconstruction of negative values of the regional TAC's (Maltz et al. 1999*b*). Although this does not appear to constitute a serious practical problem when the number of dynamic regions is of the order of 100, when solving large pixel grids of 32×32 elements, for example, constraint enforcement proves essential.

A vector of all the TAC elements may be produced via the operation

$$\boldsymbol{\phi} = \mathbf{A}\boldsymbol{\mu} \tag{13}$$

where $\mathbf{A} \in \Re^{\Xi N \times N}$ is a block diagonal matrix which via (13) allows us to implement the constraints:

$$\hat{\phi}_n[\,l\,] = \sum_{m=1}^{M} u_m[l]\,\mu_{mn} \geq 0, \qquad l = 0, 1, \ldots, L-1. \tag{14}$$

The LHS of this expression corresponds to the TAC value for the nth region at discrete time index l. The value $\Xi \leq L$ corresponds to the number of time indices as which the constraints are imposed. We define the total number of constraints $\Upsilon \triangleq \Xi N$.

We desire that $\hat{\phi}_n[\,l\,]$ be non-negative for all n and l. However, this leads to a very large matrix $\mathbf{A}$, which is undesirable. We define the set R

$$R \triangleq \Big\{ t \;\; : \;\; \frac{dc_m(t)}{dt} = 0, \quad (L-1)\,T_s > t > 0, \quad m = 1, 2, \ldots, M; \tag{15}$$

$$t \;\; : \;\; t = 0, \quad t = (L-1)\,T_s \Big\} \tag{16}$$

where T_s is the time between successive angular projections. The $c_m(t)$ are the continuous convolved orthogonal basis functions. The set R includes the start and end times of the data acquisition interval, as well as the extrema of all the COB's. To remain assured of the non-negativity of all TAC values, it is sufficient to impose the constraints (14) for $t \in R$ only, since the COB are monotonic between successive points in R. We consequently achieve a significant reduction in the row dimension of $\mathbf{A}$.

As we have access to sampled versions of the $c_m(t)$ only, the approximate extrema are found by accurately fitting a polynomial of between 8th and 12th order to each of the $\mathbf{c}_m$, and then finding the extrema of these functions. The entire set of extrema and interval values is then mapped to discrete indices, whereupon duplicate indices are removed. These values of l are the Ξ elements of the set we define as S. Figure 4 illustrates the COB and their extrema.

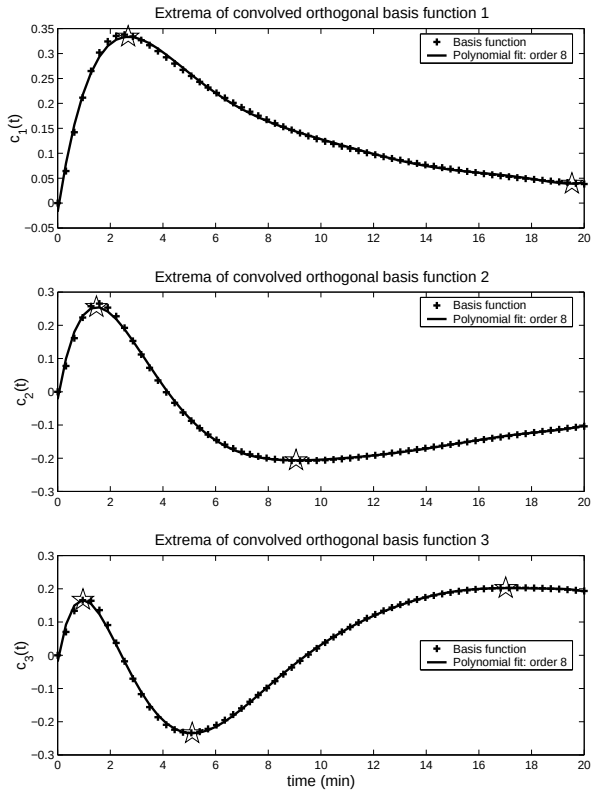

Figure 4. Convolved orthogonal basis functions and their extrema, indicated by pentagrams. TAC non-negativity constraints are enforced at these points and at the interval bounds only.

4. PROBLEM SOLUTION:

We wish to find the parameter estimate $\hat{\boldsymbol{\mu}}$ which solves the constrained least squares problem:

$$\theta(\hat{\boldsymbol{\mu}}) = \min_{\boldsymbol{\mu} \in Q} \left\| \tilde{\mathbf{y}} - \mathbf{F}\,\boldsymbol{\mu} \right\|^2 \tag{17}$$

where the feasible set Q is given by

$$Q \triangleq \left\{ \boldsymbol{\mu} \in \Re^{MN} : \ \mathbf{h} \geq \mathbf{A}\boldsymbol{\mu} \geq \mathbf{l};\ b^h_{mn} \geq \mu_{mn} \geq b^l_{mn} \right\} \tag{18}$$

with $\mathbf{h}$ and $\mathbf{l}$ the respective upper and lower bounds on the constraints. The b^h_{mn} and b^l_{mn} similarly bound the optimization parameters.

As (17) subject to (18) constitutes an inequality and bound constrained least squares problem (LSI), this is problem is convex, and hence the minimum value of the objective function $\theta(\hat{\boldsymbol{\mu}})$ is a global minimum. A unique solution exists, however, if and only if

$$\operatorname{rank}\big(\mathbf{A}\big) = \Upsilon \quad \text{and} \quad \operatorname{rank}\begin{bmatrix} \mathbf{F} \\ \mathbf{A} \end{bmatrix} = D = NM, \tag{19}$$

where Υ is the number of constraints and D is the parameter vector dimension. If $\operatorname{rank}\big(\mathbf{A}\big) < \Xi$, the constraints are inconsistent, and no solution exists (Björck 1996).

We discuss two major approaches to the solution of the LSI problem. The first is to recast the problem as a least distance problem (LDP), and then as one of non-negative least squares (NNLS). This is the approach taken by Lawson and Hanson (Lawson & Hanson 1974).

Alternatively, it can be reexpressed as a problem in quadratic programming (QP) (Björck 1996)

$$\boldsymbol{\theta}(\hat{\boldsymbol{\mu}}) = \min_{\boldsymbol{\mu} \in Q'} \; \boldsymbol{\mu}^T \boldsymbol{\Lambda}\, \boldsymbol{\mu} + \mathbf{a}^T \boldsymbol{\mu}, \tag{20}$$

with

$$Q' \triangleq \left\{ \boldsymbol{\mu} \in \Re^{MN} : \; \mathbf{h} \geq \boldsymbol{\Lambda}\boldsymbol{\mu} \geq \mathbf{l} \right\} \tag{21}$$

where

$$\boldsymbol{\Lambda} = \mathbf{F}^T \mathbf{F} \quad \text{and} \quad \mathbf{a} = -2\mathbf{F}^T \boldsymbol{\mu}. \tag{22}$$

Since problem QP is a subproblem solved at each iteration of many constrained non-linear optimization algorithms; efficient, well-tested and numerically robust code exists for its solution. The formation of the matrix $\mathbf{F}^T\mathbf{F}$ is undesirable as it introduces numerical errors, so conventional QP methods must be modified to obviate the need to calculate this matrix. Gill et al. describe such an algorithm, which is available as part of the commercial LSSOL and NAG libraries (Gill, Murray, Saunders & Wright 1984, Björck 1996).

Preliminary experiments indicated that the NAG QP code proves more robust for the solution of this particular problem than the LDP code available as part of the free SLATEC numerical library and the similar code given by Lawson and Hanson (Lawson & Hanson 1974). Consequently, we utilize the former for all experiments (NAG subroutine E04NCF mode LS1; Numerical Algorithms Group, Oxford, UK).

We refer to the algorithm described above as the constrained COBRA algorithm (CCOBRA).

4.1. Multiresolution acceleration

By first solving problem (17-18) at a lower resolution than that finally desired, we may accelerate convergence when the same problem is subsequently solved for a greater number of regions. This is easily achieved when using QP algorithms, which begin their search for the solution based on an initial parameter estimate $\hat{\boldsymbol{\mu}}_0$. This estimate is likely to fall within the feasible set of solutions for the higher dimension problem, and this will speed the execution of the latter through elimination of the feasibility search phase of the QP algorithm. Once feasibility is achieved, the optimality stage of the algorithm proceeds, and the optimization parameter vector remains within the feasible region thereafter (Gill et al. 1984).

Let us consider a subimage of uniform resolution, such as the 6×6 pixel block in upper left interior of Figure 3. Suppose we have just completed iteration i of the CCOBRA algorithm, which has yielded the solution vector $\hat{\boldsymbol{\mu}}^i$. The current resolution of this grid is $\rho_i \times \rho_i$, $\rho_i = 16$.

Suppose, on the next iteration that we wish to resolve this region as a 12×12 pixel $\rho_{i+1} \times \rho_{i+1}$, $\rho_{i+1} = 32$ grid. The $(i+1)$ th iteration may then be initialized with

$$\hat{\boldsymbol{\mu}}_{0n'}^{i+1} = \left(\frac{\rho_i}{\rho_{i+1}} \right)^2 \hat{\boldsymbol{\mu}}_n^i = \varrho_i^{i+1}\, \hat{\boldsymbol{\mu}}_n^i, \qquad n' = 1, 2, \ldots, (\rho_{i+1})^2; \tag{23}$$

where the $(\rho_{i+1}/\rho_i)^2 \rightarrow 1$ mapping between n' and n is based on geometrical overlap between the two pixel grids.

5. EXPERIMENTAL EVALUATION

The main objective of the experiments performed here is to establish whether dynamic reconstruction of source distributions segmented into large numbers of regions is feasible using the proposed method. These experiments are preliminary, in that they do not constitute thorough testing. Comprehensive experiments would need to evaluate algorithm performance versus:

1. Total number of reconstructed regions N.
2. Total detected events.
3. Average number of counts per pixel.
4. Extent and relative activity variation of sources within the imaged distribution.

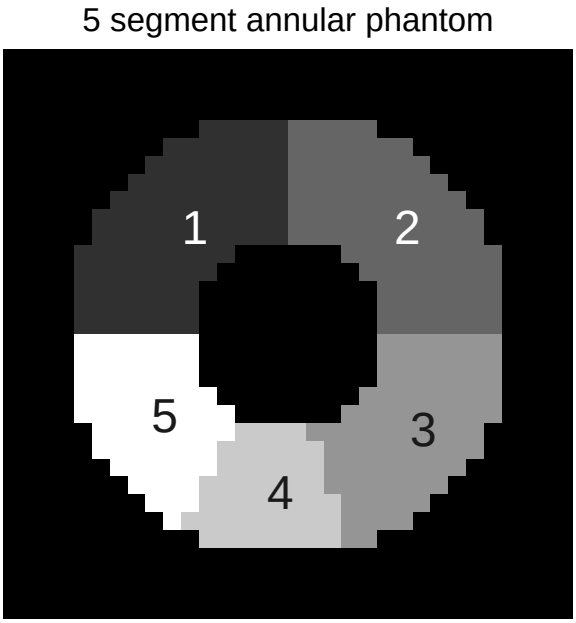

Figure 5. Dynamic phantom used in experiments

5. The effect of physiological departures from the assumed modified single compartment model (2).
6. Inaccuracies in the system matrix $\mathbf{G}$.

Here, we seek to determine how well we are able to recover individual TAC's using a single phantom, and uniform pixel grids of 16×16 ($N = 256$) and 32×32 ($N = 1024$) pixels. To our knowledge, dynamic ECT algorithms which allow a kinetic model as flexible as (2) have not been tested with so great a number of regions. Although it may be argued that the tests presented here lack sufficient realism, we believe that in the context of what has proved a very difficult, ill-conditioned problem, merely finding the limits of feasibility under the best possible circumstances, is important.

As well as testing algorithm performance, we also determine the efficacy of the multiresolution acceleration method described in Section 4.1.

5.1. Phantom simulation description

5.2. Phantom geometry

The phantoms used are 8×8, 16×16 and 32×32 versions of the segmented annulus which appears in Figure 5.

5.3. Kinetic model specification

The kinetic parameters for each of the regions are given in Table 3. The blood input function,

$$I(t) = t\,\mathrm{e}^{\,-t/\tau}$$

with $\tau = 0.7$ minutes, is used in all experiments. It is similar in form to input functions commonly encountered in clinical myocardial studies. The entire data acquisition interval is chosen to be 20 minutes.

The orthogonal basis functions were calculated through the application of the SVD to a matrix of 100 sampled exponential functions parameterized by rate constants logarithmically spaced in the interval $[\,k_2^l,\, k_2^h\,] = [\,0.025,\, 1.6\,]$. This interval includes the true range of $k_2 \in [0.1,\, 0.4]$ (see Table 3), from which the TAC's are derived. In practice, of course, the true range is unknown, so the choice of interval for k_2 should ensure that all physiologically feasible modes are accommodated. The number M of left singular vectors $\mathbf{u}_m$ retained after application of the SVD is selected as the minimum number needed to approximate all of the exponential functions $f_m[l]$ to within 1% peak deviation, using the reduced-dimension basis.

Quantity	Value
R	1
$\tilde{M}$	100
M	3
Ξ	8
QP feasibility tolerance (ϵ)	1×10^{-14}
Constraint safety factor (λ)	2
Event distribution	i.i.d. Poisson
System response	ideal
Execution platform	Pentium II 400MHz
Operating system	Linux 2.2.10
Algorithm implementation	GNU C (gcc-2.91)

Table 1. General parameter specifications, experimental assumptions, and implementation details

5.4. Simulation parameters

Algorithm parameters used for each of the three resolution tests appear in Table 5.6. Those constant throughout all tests and universal modeling assumptions appear in Table 1. The significance of several of the quantities listed are explained below.

5.5. Specification of constraint and interval bounds

In order to select the interval bounds on the μ_{mn}, the range of values of μ necessary to represent any of the original $\tilde{M}$ exponentials was determined. By convolving these exponentials with the input function, the peak value of all TAC's ϕ_h was then found. The μ were then scaled by a factor κ so that all TAC's could be represented using linear combinations of the COB, and these bounds modified to allow for TAC's with no activity as follows:

$$b_m^h = \lambda \max\left\{ \max_n (\kappa\, \mu_{mn}),\, 0 \right\} \tag{24}$$

$$b_m^l = \lambda \min\left\{ \min_n (\kappa\, \mu_{mn}),\, 0 \right\}. \tag{25}$$

The safety factor $\lambda = 2$ was introduced so as to weaken these constraints somewhat, in order to account for possible inaccuracies in the estimated or measured input function. At the $(i+1)$th resolution iteration, these bounds are scaled by ϱ_i^{i+1} relative to bounds used at the preceding iteration.

The lower constraint bound is set as $\mathbf{l} = \mathbf{0}$ to enforce TAC non-negativity. A very small QP algorithm feasibility tolerance $\epsilon = 1 \times 10^{-14}$ is required for this problem, and $\mathbf{l} = -\eta\mathbf{1}$, becomes the true lower bound on the TAC values, where $\mathbf{1} \in \Re^{MN}$ is a vector of ones. The upper bound is specified as:

$$\mathbf{h} = \lambda\, \phi_h\, \mathbf{1}. \tag{26}$$

5.6. Error metric

The most important test of the efficacy of a dynamic reconstruction algorithm is how well the individual regional TAC's are recovered. When recovered TAC's are accurate, a 1D fit to a kinetic model of choice may be easily effected. To this end we define the metric:

$$M_{\text{rms}} = \sqrt{\frac{\sum_{n=1}^{N} \sum_{m=1}^{M} \sum_{l=0}^{L-1} \left(\phi_n[\,l\,] - \hat{\phi}_n[\,l\,] \right)^2}{\sum_{n=1}^{N} \sum_{m=1}^{M} \sum_{l=0}^{L-1} \phi_n[\,l\,]^2}} \times 100 \tag{27}$$

which gives as a percentage, the square root of the sum of power of the errors in all recovered TAC's as a fraction of the total power in all of the original TAC's.

Resolution	P	Q	Ξ
8×8	32	12	8
16×16	64	32	8
32×32	64	32	8

Table 2. Specific setting used with each phantom.

	Region number				
Mode rate constant:	1	2	3	4	5
$k_2^1 = 0.1$	1	0	0	1	0
$k_2^2 = 0.2$	0	1	0	0	1
$k_2^3 = 0.4$	0	0	1	0	0

Table 3. Coefficients (k_1^{mn}) of exponential modes having rate constants k_2^m which define the TAC's for each region of the phantoms depicted in Figure 5

6. RESULTS

Test results for the 16×16 region test, at various noise levels, appear in Table 4, while those for the 32×32 region experiments are listed in Table 5. Experimental results detailing algorithm performance for the 8×8 tests which provided the warm-start conditions for the 16×16 studies are not shown.

7. DISCUSSION

The proposed CCOBRA algorithm exhibits the following improvements over the earlier COBRA algorithm:

1. Under noise-free conditions, perfect reconstruction of 1024 dynamic regions is possible. COBRA produced TAC's with significant negative excursions when as few as 256 regions were reconstructed.
2. The multiresolution warm-start approach leads to reduction in total algorithm execution time by a factor of 3-4 for 16×16 region reconstructions, and by about 2 for 32×32 reconstructions.

Compared with some other fixed geometry DRIP algorithms:

1. CCOBRA offers considerable speed and flexibility advantages over the method of Limber et al., which requires approximately 10 times the computation time to produce estimates of the wash-out time activities only, for a 16×16 phantom (Limber et al. 1995).

Total events	Warm start	M_{rms} (%)	Iterations of QP	Objective fn. value	Exec. time (sec)
Infinite	Yes	0	1246	0	400
1×10^6	Yes	14.48	1636	5.64E-5	400
1×10^5	Yes	21.50	1629	6.43E-4	400
1×10^4	Yes	35.82	1559	6.47E-3	400
1×10^4	No	33.32	2398	5.00E-3	1400
5×10^3	Yes	67.34	1104	1.22E-1	400
1×10^3	Yes	88.71	934	5.48E-1	400

Table 4. Results for tests using the 16×16 phantom. Execution times are rounded to nearest hundred seconds. We note that warm-starting reduces execution time to about one third. Warm-starts were initialized using the results of the 8×8 experiments. The value of the objective function (17) at the solution is given in the penultimate column.

Total events	Warm start	M_{rms} (%)	Iterations of QP	Objective fn. value	Exec. time (hrs)
Infinite	No	0.00	7643	0	12
1×10^6	No	22.61	8331	0.011	12
1×10^6	Yes	22.26	3668	0.011	6
5×10^5	Yes	40.33	3775	0.015	6
1×10^5	Yes	55.50	4667	0.095	6
5×10^4	Yes	64.60	4798	0.204	6
1×10^4	Yes	84.48	4816	1.080	6

Table 5. Results for tests using the 32×32 phantom. Warm-starts were initialized using the results of the 16×16 experiments. We note that warm-starting reduces execution time to about half. Execution times are rounded to nearest hour. The value of the objective function (17) at the solution is given in the penultimate column.

2. Although the linear least squares method of Farncombe et al. is likely to incur similar computational cost, CCOBRA is free of the constraint that all TAC values must decrease monotonically (Farnocombe et al. 1999).

3. The linear least squares method of Reutter et al., which is ostensibly identical to COBRA except that the basis functions are chosen as a set of 16 B-splines rather the 3 or 4 COB's, may rival CCOBRA in speed and estimation accuracy for large problems. This is because the employment of a non-negative basis set allows positivity constraints to be implemented by imposing simple interval constraints on the μ_{mn}. This obviates the need to form the cumbersome matrix $\mathbf{A}$. As yet this algorithm has not yet been evaluated for the solution of problems of large dimension (Reutter et al. 1999*a*). As is equally true of the COB, splines may be used to represent many functions other than the exponentials we ideally seek to be the exclusive realizations of all possible linear combinations of a reduced dimension basis set. It is thus likely that certain constraints will need to be imposed on the coefficients of a spline basis to restrict the nature of the functions they represent, when large problems are considered.

Let us consider as acceptable for practical purposes a TAC recovery error of $M_{\text{rms}} \approx 20\%$. For the 256 region phantom, this can be achieved at count levels of 1×10^5 counts per slice at a resolution of 16×16. For 32×32 region reconstructions, the number of detected events required is 1×10^6, an increase of an order of magnitude. This is a sobering finding in that it suggests that it may not be possible to reconstruct dynamic series of the 64×64 and 128×128 SPECT images to which clinicians are accustomed.

However, multiresolution grids such as the example depicted in Figure 3 offer one way of reducing dimension. Decoupling of the problem into smaller more tractable subproblems is another. In all dimensions, the exterior problem of the Radon transform may be solved, and for distributions $\Omega(\mathbf{x})$, $\mathbf{x} \in \Re^n$, n odd, the interior problem is likewise soluble (Natterer 1986, p.158, 166-169). It is possible that exploitation of these properties might allow decoupling of the large estimation problem into smaller, more tractable ones.

We note also that the convolved orthogonal basis used can probably be improved upon. We are currently investigating methods to convert the COB to a basis set of the same dimension which possesses no negative excursions. This will drastically decrease the memory requirements of the algorithm, and is likely to improve low SNR performance at higher resolutions. Our justification for this statement is based on the observation that the rank of the matrix

$$\begin{bmatrix} \mathbf{F} \\ \mathbf{A} \end{bmatrix} \tag{28}$$

decreases as the problem dimension increases (see 19). When this occurs, the global minimum is no longer unique, and many sets of TAC's, when projected, are able to produce sinograms which are very similar to the acquired sinogram. Through further implicit constraints on the form of the basis functions, the matrix $\mathbf{A}$ will become smaller, and the probability of the existence of a unique minimum norm solution will increase. Referring to Figure 1, it is interesting to note that the RMS % errors given for static SPECT are about $1.5 - 2$ times those obtained in these dynamic studies. This is encouraging, and it is hoped that through further refinement of the CCOBRA algorithm, TAC's of comparable quality to the activity estimates obtained in the static ECT might be achieved.

8. ACKNOWLEDGMENTS

This work was supported in part by the National Heart, Lung, and Blood Institute of the U.S. Department of Health and Human Services under grant HL-07367; in part by the Director, Office of Science, Office of Biological and Environmental Research, Medical Sciences Division of the U.S. Department of Energy under contract DE-AC03-76SF00098; and in part by the South African National Research Foundation.

REFERENCES

Björck, A. (1996), *Numerical methods for least squares problems*, SIAM.

Budinger, T. F. (1994), Dynamic SPECT opens new horizons for technetium radiopharmaceutical applications and new challenges for ligand design, *in* 'Proceedings of the 4th International Symposium on Technetium in Chemistry and Nuclear Medicine', Vol. 4, pp. 475–489.

Chiao, P. C., Rogers, W. L., Clinthorne, N. H., Fessler, J. A. & Hero, A. O. (1994), 'Model-based estimation dynamic cardiac studies using ECT', *IEEE Transactions on Medical Imaging* **13**(2), 217–226.

Cunningham, V. J. & Jones, T. (1993), 'Spectral analysis of dynamic PET studies', *Journal of Cerebral Blood Flow and Metabolism* **13**, 15–23.

Farnocombe, T., Noll, D., Maeght, J. & Harrop, R. (1999), An evaluation of dynamic SPECT imaging methods, *in* 'IEEE Nuclear Science Symposium and Medical Imaging Conference Record', pp. 1959–1963.

Gill, P. E., Murray, W., Saunders, M. A. & Wright, M. H. (1984), 'Procedures for optimization problems with a mixture of bounds and general linear constraints', *ACM Transactions on Mathematical Software* **10**, 282–298.

Lanczos, C. (1956), *Applied analysis*, Prentice-Hall.

Lawson, C. L. & Hanson, R. J. (1974), *Solving lease squares problems*, Prentice-Hall.

Limber, M. A., Limber, M. N., Celler, A., Barney, J. S. & Borwein, J. M. (1995), 'Direct reconstruction of functional parameters for dynamic SPECT', *IEEE Transactions on Nuclear Science* **42**(4), 1249–1256.

Maltz, J. S., Polak, E. & Budinger, T. F. (1999*a*), 'Multistart optimization algorithm for joint spatial and kinetic parameter estimation from dynamic ECT projection data', *IEEE Transactions on Medical Imaging, in review.*

Maltz, J. S., Reutter, B. W., Huesman, R. H. & Budinger, T. F. (1999*b*), An evaluation of dynamic SPECT imaging methods, *in* 'IEEE Nuclear Science Symposium and Medical Imaging Conference Record', pp. M7–92.

Natterer, F. (1986), *The Mathematics of Computerized Tomography*, John Wiley & Sons.

O'Sullivan, F. (1993), 'Imaging radiotracer model parameters in PET: A mixture analysis approach', *IEEE Tranactions on Medical Imaging* **12**(3), 399–411.

Reich, J. G. (1981), On parameter redundancy in curve fitting of kinetic data, *in* L. Endrenyi, ed., 'Kinetic data analysis : design and analysis of enzyme and pharmacokinetic experiments', Plenum Press, chapter IV, pp. 39–60.

Reutter, B. W., Gullberg, G. T. & Huesman, R. H. (1998), 'Kinetic parameter estimation from attenuated SPECT projection measurements', *IEEE Transactions on Nuclear Science* **45**(6), 3007–3013.

Reutter, B. W., Gullberg, G. T. & Huesman, R. H. (1999*a*), Direct least squares estimation of spatiotemporal distributions from dynamic cardiac SPECT projections, *in* 'IEEE Nuclear Science Symposium and Medical Imaging Conference Record', pp. M6–3.

Reutter, B. W., Gullberg, G. T. & Huesman, R. H. (1999*b*), Kinetic parameter estimation from dynamic cardiac patient SPECT projection measurements, *in* '1998 IEEE Nuclear Science Symposium and Medical Imaging Conference Record', pp. 1953–1958.

Smith, A. M., Gullberg, G. T., Christian, P. E. & Datz, F. L. (1994), 'Kinetic modeling of teboroxime using dynamic SPECT imaging of a canine model', *Journal of Nuclear Imaging* **35**(3), 484–495.