CONSTRAINED MIXTURE MODELING FOR THE ESTIMATION OF KINETIC PARAMETERS IN DYNAMIC PET

†*Yanguang Lin*,**Quanzheng Li*,†*Justin P. Haldar*,†*Richard M. Leahy*

†Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089
*Department of Radiology, Massachusetts General Hospital, Boston, MA 02114

ABSTRACT

The estimation and analysis of kinetic parameters in dynamic PET is frequently confounded by noise and partial volume effects. We propose a new constrained model of dynamic PET to address these limitations. The proposed formulation incorporates an explicit partial volume model in which each image voxel is represented as a mixture of different pure tissue types with distinct temporal dynamics. A two stage algorithm is proposed to solve the resulting problem. In the first stage, a sparse signal processing method is applied to estimate the rate parameters for the different tissue compartments from the noisy PET time series. In the second stage, tissue fractions and the linear parameters of different time activity curves (TACs) are estimated using a combination of sparsity, spatial-regularity, and fractional mixture constraints. A block coordinate descent (BCD) algorithm is combined with a manifold search to robustly estimate these parameters. The method is evaluated with both simulated and experimental dynamic PET data.

***Index Terms*—** Dynamic PET, Mixture Modeling, Sparsity, Kinetic Parameter Estimation

1. INTRODUCTION

Dynamic PET is a powerful technique that provides information about molecular processes in the human body. Pharmacokinetic modeling is commonly used in PET to investigate the physiological processes involved in tracer uptake, and estimated kinetic parameters can play an important role in disease diagnosis and treatment evaluation. It is common to estimate kinetic parameters by fitting PET images of the time-varying tracer density to an appropriate model.

Conventional methods for kinetic parameter estimation [1, 2] make the assumption that tracer kinetics are homogeneous within a voxel or a region of interest. However, tissue heterogeneity and partial volume effects inevitably cause this assumption to be violated. Highly heterogeneous tumors are considered more aggressive with a higher propensity for metastasis or invasion [3]. Therefore quantification of tumor heterogeneity could prove to be a useful metric for treatment assessment or a predictive indicator for treatment failure [4]. In addition, tumor heterogeneity could potentially be exploited for biologically optimized treatment planning in radiation therapy. Because of the significance of tumor heterogeneity in cancer management, the application of conventional homogeneous tissue models is rather limited.

To overcome these limitations, mixture models [5, 6] have been proposed that model the time activity curve (TAC) from each voxel as a linear combination of several TACs representing the tracer dynamics of pure tissues. Since mixture models often use an overcomplete parameterization, solutions are frequently not unique unless additional constraints are incorporated. In [7, 8], basis pursuit and sparse Bayesian learning approaches were proposed to sparsely estimate tissue kinetic parameters, under the assumption that the number of distinct tissue types contributing to any given voxel is relatively small. However, these methods operate voxel-by-voxel, without incorporating the prior information that PET tracer kinetics should be spatially-correlated and smoothly varying. In addition, these methods were only used to derive summary statistics of the ensemble TAC behavior in each voxel, without attempting to accurately resolve the kinetic parameters and volume fractions of each contributing tissue.

Our approach attempts to overcome the drawbacks of these previous approaches by combining accurate kinetic modeling, joint-sparsity modeling [9], and spatial regularity constraints. By using joint sparsity, we can leverage the fact that tissue TACs for the same tissue type are relatively consistent over large regions, which improves the SNR for kinetic parameter estimation. By using an ℓ_0 sparsity constraint, we can robustly recover the tissue volume fractions in the presence of significant image noise. And our method enables estimation of the complete set of kinetic parameters without requiring linearization or use of the Patlak model.

2. METHOD

2.1. Mixture Model

We assume an image model comprised of Q distinct tissue types and N voxels. Assuming that we know the plasma input function $C_p(t)$, the time activity curve $TAC^q(t)$ for the q^{th} tissue can be written as [10]

$$TAC^q(t) = (c_q e^{-\lambda_d t} + d_q e^{-\theta_q t}) * C_p(t), \tag{1}$$

where c_q, d_q and θ_q are unknown tissue parameters, λ_d is the known tracer decay constant, and $*$ denotes convolution.

For each voxel, the measured TAC is expressed as the linear combination of these tissue TACs:

$$TAC_j(t) = \sum_{q=1}^{Q-1} TAC^q(t) A_{qj} + C_p(t) A_{Qj}, \tag{2}$$

where j is the voxel index and the A_{qj} are linear mixing coefficients that represent the fractions of different types of tissues

present in the voxel. Substituting (1) into (2), we have

$$\begin{aligned} TAC_j(t) = &(\sum_{q=1}^{Q-1} c_q A_{qj}) e^{-\lambda_d t} * C_p(t) \\ &+ \sum_{q=1}^{Q-1} d_q A_{qj} e^{-\theta_q t} * C_p(t) + A_{Qj} C_p(t). \end{aligned} \tag{3}$$

In equation (3), Q, c_q, d_q, A_{qj}, and θ_q are unknowns to be estimated, under the constraints that

$$\theta_q > \lambda_d \ , \ c_q > 0 \ , \ d_q > 0 \ , \ A_{qj} \geq 0 \ , \ \sum_q A_{qj} \leq 1, \tag{4}$$

that A_{qj} should be spatially smooth for each q, that there is at most one fractional tissue compartment that is not metabolically active (e.g., necrosis), and that non-active tissue compartments are present in only a small number of voxels (i.e., $\sum_q A_{qj} \neq 1$ in only a sparse set of voxels).

The joint estimation of all these parameters is complicated because of nonlinearity, the fact that the model complexity changes significantly for different Q, the bilinearity of A_{qj} with both c_q and d_q, and the constraints that need to be satisfied. As a result, we propose a two stage method to separate the estimation of Q and the θ_q from the remaining variables. In the first stage, we use a sparsity-constrained dictionary approach to simultaneously estimate Q and θ_q. In the second stage we use a block coordinate descent (BCD) algorithm together with a manifold search to robustly estimate the tissue fractions and the linear tissue TAC parameters.

2.2. Stage 1: Dictionary-Based Estimation of Q and θ_q

By a change of variables, (3) can be rewritten as

$$\begin{aligned} TAC_j(t) = &B_{0j} e^{-\lambda_d t} * C_p(t) \\ &+ \sum_{q=1}^{Q-1} B_{qj} e^{-\theta_q t} * C_p(t) + B_{Qj} C_p(t) \end{aligned} \tag{5}$$

where we have used $B_{0j} = \sum_{q=1}^{Q-1} c_q A_{qj}$ and $B_{qj} = d_q A_{qj}$ for $q = 1, \ldots, Q-1$, and $B_{Qj} = A_{Qj}$.

This change of variables complicates the form of the constraints in (4). However, while these constraints will play important roles in estimating c_q, d_q, and A_{qj}, they are not as essential for estimating Q and θ_q. As a result, we neglect these constraints on B for this first stage to ensure that the estimation problem is tractable. Estimation of Q is still nonconvex and nonlinear with respect to the θ_q. To avoid this, we use a discretized dictionary-based approach inspired by [7, 8]. In particular, we express (3) using the overcomplete series representation

$$TAC_j(t) = \sum_{q=0}^{N_b} B_{qj} \psi_q(t), \tag{6}$$

where $N_b \gg Q$, and with

$$\begin{aligned} &\psi_0(t) \stackrel{def}{=} C_p(t), \quad \psi_1(t) \stackrel{def}{=} e^{-\lambda_d t} * C_p(t), \\ &\text{and } \psi_n(t) \stackrel{def}{=} e^{-\theta_n t} * C_p(t), \ 2 < n < N_b. \end{aligned} \tag{7}$$

The set of $N_b - 1$ different values of θ_n are selected from the physiologically meaningful range. From our original model (3), we expect that B_{qj} in (5) will have joint-sparse characteristics, i.e., that the number of q values for which *any* $B_{qj} \neq 0$ will be very small (and equal to $Q+1$). Given the set of non-zero B_{qj}, it is straightforward to retrieve the corresponding rate parameters.

In practice, dynamic PET time series are often reconstructed at a finite number I of time frames, with each frame corresponding to the average tracer behavior over a given time interval. To model this process, we introduce new quantities

$$Y_{ij} = \frac{\int_{t_i^s}^{t_i^e} TAC_j(t)dt}{t_i^e - t_i^s}, \ \ \Psi_{iq} = \frac{\int_{t_i^s}^{t_i^e} \psi_q(t)dt}{t_i^e - t_i^s} \tag{8}$$

where t_i^s and t_i^e are the sequences of start and end frame times, with which the measured data at the ith time frame in the jth voxel is modeled as

$$Y_{ij} \approx \sum_{q=0}^{N_b} \Psi_{iq} B_{qj} \tag{9}$$

for each i, which we can simultaneously express in matrix form for all voxels and all time points in matrix form as

$$\mathbf{Y} = \mathbf{\Psi B}, \text{ with } \mathbf{Y} \in \mathbb{R}^{I \times N}, \ \mathbf{\Psi} \in \mathbb{R}^{I \times (N_b+1)}, \text{ and } \mathbf{B} \in \mathbb{R}^{(N_b+1) \times N}. \tag{10}$$

Although there will not be a unique least-squares solution for $\mathbf{B}$ in the typical case where $N_b + 1 > I$, we can use joint-sparsity information to make the problem well posed and to include the prior information that Q should be relatively small. We define the *row*-ℓ_0 quasi-norm of $\mathbf{B}$ as the number of non-zero rows which is a measure of joint sparsity [9]. Since the first two rows of $\mathbf{B}$ are always present, we impose a joint-sparsity constraint only on a submatrix of $\mathbf{B}$ with the first two rows removed, which we denote by $\mathbf{B}_{2:N_b,:}$. The resulting sparse estimation problem can be posed in regularized form as

$$\min_{\mathbf{B}} = \|\mathbf{Y} - \mathbf{\Psi B}\|_{Fro}^2 + \kappa \|\mathbf{B}_{2:N_b,:}\|_{row\text{-}\ell_0} \tag{11}$$

where κ is a regularization parameter.

Since the *row*-ℓ_0 quasi-norm is nonconvex, it is common to apply convex relaxation methods [9]. Our preliminary experience with this model suggests that convex relaxation approaches are not well suited to this problem, due to the significant temporal coherence between the different ψ_q. However, we have found much better success if we directly optimize (11) using combinatorial methods. Despite being NP-hard, direct optimization of (11), which can ensure global optimality for this subproblem, is not time consuming if we leverage the prior information that the true value of Q should be small, and has proven to be very robust to noise. Our results always use $Q < 5$.

2.3. Stage 2: Estimation of c_q, d_q and A_{qj}

Solving (11) provides us with estimated $\hat{Q}$ and $\hat{\theta}_q$, and with these parameters fixed, we return to solving (3) for the remaining parameters. Defining

$$\mathbf{c} = [c_1, c_2, \ldots, c_{\hat{Q}-1}] \text{ and } \mathbf{D} = diag\{d_1, d_2, \ldots, d_{\hat{Q}-1}\}, \tag{12}$$

our model can be expressed in matrix form as

$$\begin{aligned} \mathbf{Y} &\approx \hat{\mathbf{\Psi}} \begin{pmatrix} \mathbf{c} & 0 \\ \mathbf{0} & 1 \\ \mathbf{D} & \mathbf{0} \end{pmatrix} \mathbf{A} \\ &= \hat{\mathbf{\Psi}} \mathbf{F} \mathbf{A}, \end{aligned} \tag{13}$$

where $\mathbf{A} \in \mathbb{R}^{Q \times N}$ is the matrix of A_{qj}, $\mathbf{F} \in \mathbb{R}^{(\hat{Q}+1) \times \hat{Q}}$ is given by

$$\mathbf{F} = \begin{pmatrix} \mathbf{c} & 0 \\ \mathbf{0} & 1 \\ \mathbf{D} & \mathbf{0} \end{pmatrix}, \tag{14}$$

and we form the matrix $\hat{\mathbf{\Psi}}$ as the submatrix of the columns from $\mathbf{\Psi}$ corresponding the non-zero rows of $\hat{\mathbf{B}}$.

There are $2(\hat{Q}-1)$ unknown parameters (c_q, d_q) in $\mathbf{F}$ and $N\hat{Q}$ unknown parameters in $\mathbf{A}$. In addition, without use of the constraints in

(4), the least squares fit of these parameters would not be unique because of bilinearity . We need additional prior information to solve for $\hat{\mathbf{F}}$ and $\hat{\mathbf{A}}$. In addition to (4), we also apply:

1. Sparsity-enforcing ℓ_1 regularization on $\left(1-\sum_q A_{qj}\right)$ to impose the prior information that most voxels are completely filled with metabolically active tissues. This constraint removes the bilinear scale ambiguity between $\mathbf{F}$ and $\mathbf{A}$.
2. Quadratic smoothness regularization of the tissue fractions A_{qj}, to impose the prior information that tissue contributions will be vary smoothly standard PET images.

Combining these two constraints, we define our cost function as:

$$cost(\mathbf{F},\mathbf{A}) = ||\mathbf{Y}-\hat{\Psi}\mathbf{F}\mathbf{A}||^2_{Fro} + \beta\sum_q^{\hat{Q}}\sum_j^N \sum_{l\in Nbr(j)} \phi(A_{qj}-A_{ql}) + \gamma\sum_j^N |1-\sum_{q=1}^{\hat{Q}} A_{qj}| \quad (15)$$

where $Nbr(j)$ denotes the set of neighboring voxels for voxel j, $\phi(x)=x^2$, and γ and β are regularization parameters.

The optimal solution to (15) is:

$$[\hat{\mathbf{F}},\hat{\mathbf{A}}] = \underset{\mathbf{F}\in\Omega_F,\mathbf{A}\in\Omega_A}{\operatorname{argmin}} cost(\mathbf{F},\mathbf{A}) \quad (16)$$

where $\Omega_F = \{\mathbf{F} : \mathbf{F} = \begin{pmatrix} \mathbf{c} & 0 \\ \mathbf{0} & 1 \\ \mathbf{D} & \mathbf{0} \end{pmatrix}, \mathbf{D} = diag(\mathbf{d}), c_q > 0, d_q > 0\}$, and $\Omega_A = \{\mathbf{A} : A_{qj} \geq 0, \sum_q A_{qj} \leq 1\}$.

To find the optimal solution, we alternate between estimating $\mathbf{F}$ and $\mathbf{A}$ using a BCD approach. Although our regularization framework eliminates potential bilinear scale ambiguity between $\mathbf{F}$ and $\mathbf{A}$, the bilinear nature of the problem prevents convergence of our alternating algorithm when the current estimates of $\mathbf{F}$ and $\mathbf{A}$ lie in a solution valley where $\mathbf{Y} \approx \hat{\Psi}\mathbf{F}\mathbf{A}$ but $\left(1-\sum_q A_{qj}\right)$ is not optimally sparse. This is because the alternating BCD method does not allow for simultaneous update of $\mathbf{A}$ and $\mathbf{F}$. To overcome this problem, we introduce an additional manifold search into our alternating descent algorithm, to resolve problems of scale ambiguity in this region of the cost function. A sketch of the full algorithm is given below:

1. At the mth iteration, fix $\mathbf{F}$ to be equal to the estimate from the previous iteration $\hat{\mathbf{F}}_{m-1}$ and solve

$$\begin{aligned}\hat{\mathbf{A}}_m &= \underset{\mathbf{A}\in\Omega_A}{\operatorname{argmin}} cost(\hat{\mathbf{F}}_{m-1},\mathbf{A}) \\ &= ||(\mathbf{Y}-\hat{\Psi}\hat{\mathbf{F}}_{m-1}\mathbf{A})||^2_{Fro} \\ &+ \beta\sum_{q=1}^{\hat{Q}}\sum_{j=1}^N \sum_{l\in N(j)} \phi(A_{qj}-A_{ql}) + \gamma\sum_j^N |1-\sum_{q=1}^{\hat{Q}} A_{qj}|\end{aligned} \quad (17)$$

   using an alternating direction method of multipliers (ADMM) algorithm [11].
2. For fixed $\hat{\mathbf{A}}_m$ fixed,we solve

$$\hat{\mathbf{F}}_m = \underset{\mathbf{F}\in\Omega_F}{\operatorname{argmin}} cost(\mathbf{F},\hat{\mathbf{A}}_m) = ||\mathbf{Y}-\hat{\Psi}\mathbf{F}\hat{\mathbf{A}}_m||^2_{Fro} \quad (18)$$

   using a nonnegative least squares method.
3. We introduce an auxiliary vector $\rho \in \mathbb{R}^{\hat{Q}-1}$ and perform a manifold search to resolve potential scale ambiguity between $\hat{\mathbf{F}}_m$ and $\hat{\mathbf{A}}_m$. Defining $\Lambda_\rho = diag\{1,\rho\}$ we note that $\hat{\mathbf{F}}_m\hat{\mathbf{A}}_m = \left(\hat{\mathbf{F}}_m\Lambda_\rho^{-1}\right)\left(\Lambda_\rho\hat{\mathbf{A}}_m\right)$ for any strictly positive choices for the elements of ρ. This implies that we can shift the scaling between $\hat{\mathbf{F}}_m$ and $\hat{\mathbf{A}}_m$ without modifying the data consistency term of (15), while avoiding being trapped in a suboptimal stationary point of the BCD algorithm. Thus, we calculate

$$\rho_m = \underset{\rho\in\Omega_\rho}{\operatorname{argmin}} cost(\hat{\mathbf{F}}_m\Lambda_\rho^{-1},\Lambda_\rho\hat{\mathbf{A}}_m) \quad (19)$$

   with $\Omega_\rho = \{\rho : \rho_q \geq 0; \rho_q\hat{A}_{m,qj} \leq 1, \forall j; 1-\sum_q^{\hat{Q}-1}\rho_q\hat{A}_{m,qj} - \hat{A}_{m,\hat{Q}j} \geq 0, \forall j\}$
   This is a small scale quadratic programming problem and is solved using a standard interior point method.
4. Set $\hat{\mathbf{F}}_m = \hat{\mathbf{F}}_m\Lambda_\rho^{-1}$ and $\hat{\mathbf{A}}_m = \Lambda_\rho\hat{\mathbf{F}}_m$.
5. Iterate steps 1-4 until convergence.

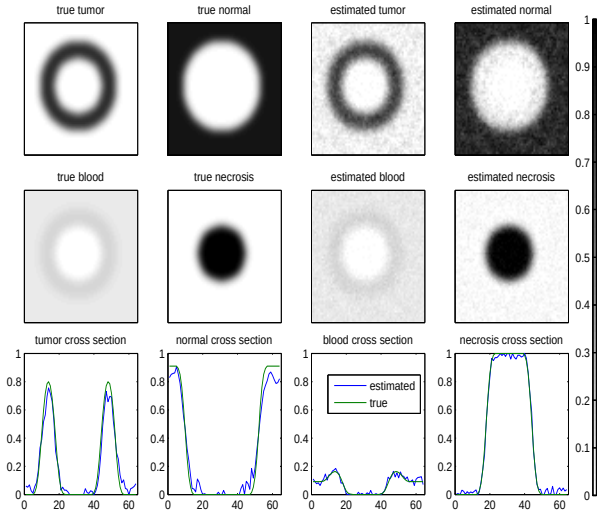

Fig. 1. True and estimated tissue fractions

3. RESULTS

3.1. Dynamic simulation

We simulated a dynamic image sequence representing a ROI containing a donut shaped tumor with a necrotic center embedded in a square 64x64 voxel background of normal tissue. Images of the true fractions $\mathbf{A}$ are shown in the top left portion of Fig. 1. We simulated noisy TACs with SNR = 50, and applied our proposed method with $\kappa = 1e^{-2}$, $\beta = 1.2e^{-5}$, and $\gamma = 1.3e^{-7}$. The estimated tissue fraction images are shown in the top right portion of Fig. 1, and the estimated tissue TACs are shown in Fig. 2. The estimated images match the original tissue fraction images quite closely.

3.2. Real mouse data

A one-hour dynamic FDG PET mouse scan was performed using a Siemens Inveon PET scanner. Dynamic list mode data was binned into 30 inhomogeneous time frames, and images were generated using MAP reconstruction [12]. Our two stage mixture model was applied to a torso ROI (see Fig. 3). Our proposed method was applied with $\kappa = 1e^{-2}$, $\beta = 3.3e^{-5}$, and $\gamma = 9.3e^{-7}$. Results are shown in Figs. 3 4.

The proposed method extracted three types of tissues. "Tissue 1" indicates tissues with accumulated FDG uptake (myocardium and kidneys); "Tissue 2" correspond to normal abdominal or-

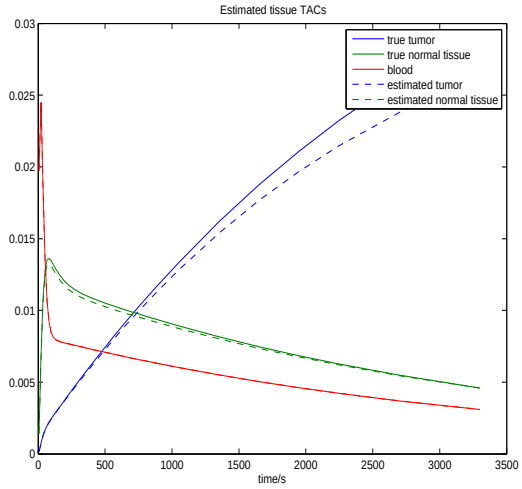

Fig. 2. Tissue TACs estimated from the simulation data.

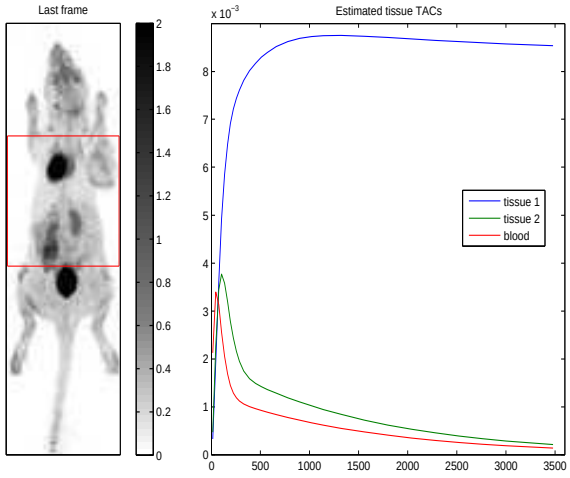

Fig. 3. (Left) The last frame of the dynamic PET time series. The red square indicates the torso ROI. (Right) Estimated tissue TACs.

gans;"blood" seems to correspond to organs with rich blood vessels or major blood vessels themselves. Although we don't have ground truth validation, the estimated tissue distribution is physiologically reasonable.

4. CONCLUSION

We investigated partial volume model kinetic parameter estimation for dynamic PET. Our proposed algorithm successfully estimates tissue kinetic parameters and tissue fractions for real and realistic data with relatively high noise. The estimated tissue fractions could be used as a measure of tumor heterogeneity to reflect heterogeneity of ongoing pathophysiological processes that can in turn be used for tumor staging, treatment assessment and optimization.

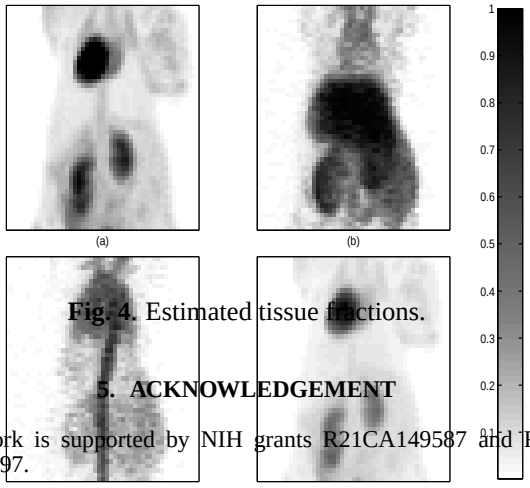

Fig. 4. Estimated tissue fractions.

5. ACKNOWLEDGEMENT

This work is supported by NIH grants R21CA149587 and R01-EB010197.

6. REFERENCES

[1] M. Kamasak, C. Bouman, E. Morris, and K. Sauer, "Direct reconstruction of kinetic parameter images from dynamic pet data," *Medical Imaging, IEEE Transactions on*, vol. 24, pp. 636 –650, may 2005.

[2] T. Nichols, J. Qi, E. Asma, and R. Leahy, "Spatiotemporal reconstruction of list-mode pet data," *Medical Imaging, IEEE Transactions on*, vol. 21, pp. 396 –404, april 2002.

[3] J. Yokota, "Tumor progression and metastasis," *Carcinogenesis*, vol. 21, no. 3, p. 497, 2000.

[4] J. Eary, F. O'Sullivan, J. O'Sullivan, and E. Conrad, "Spatial heterogeneity in sarcoma 18f-fdg uptake as a predictor of patient outcome," *Journal of Nuclear Medicine*, vol. 49, no. 12, pp. 1973–1979, 2008.

[5] F. O'Sullivan, "Imaging radiotracer model parameters in pet: a mixture analysis approach," *Medical Imaging, IEEE Transactions on*, vol. 12, pp. 399 –412, sep 1993.

[6] F. O'Sullivan, "Locally constrained mixture representation of dynamic imaging data from pet and mr studies," *Biostatistics*, vol. 7, no. 2, pp. 318–338, April 2006.

[7] R. N. Gunn, S. R. Gunn, F. E. Turkheimer, J. A. D. Aston, and V. J. Cunningham, "Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling," *J Cereb Blood Flow Metab*, vol. 22, no. 12, pp. 1425–1439, 2002.

[8] J. Peng, J. Aston, R. Gunn, C. Liou, and J. Ashburner, "Dynamic positron emission tomography data-driven analysis using sparse bayesian learning," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 9, pp. 1356–1369, 2008.

[9] J. A. and Tropp, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589 – 602, 2006.

[10] Y. Lin, Q. Li, and R. Leahy, "Kinetic parameters estimation for heterogeneous tumor model," in *Nuclear Science Symposium Conference Record (NSS/MIC)*, pp. 3069 – 3073, nov 2009.

[11] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pp. 1 –4, june 2010.

[12] 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, p. 1001, 1998.