UNMIXING DYNAMIC PET IMAGES: COMBINING SPATIAL HETEROGENEITY AND NON-GAUSSIAN NOISE

Yanna Cruz Cavalcanti[(1)], *Thomas Oberlin*[(1)], *Nicolas Dobigeon*[(1)], *Cédric Févotte*[(1)], *Simon Stute*[(2)] *and Clovis Tauber*[(3)]

[(1)] IRIT, Université de Toulouse, CNRS, Toulouse, France
[(2)] CEA SHJF, UMR 1023 IMIV, 91400 Orsay, France
[(3)] UMR U1253 iBrain, Inserm, Université de Tours, France

firstname.lastname@irit.fr, simon.sture@cea.fr,clovis.tauber@univ-tours.fr

ABSTRACT

An important task when processing dynamic PET images is to identify the time-activity curves (TACs) of the pure tissues, along with their corresponding spatial proportions. This step, often referred to as unmixing or factor analysis, is based on a loss function which measures the discrepancy between the observed data and the model. This loss function should be chosen according to the statistical properties of the noise, which is in this case hard to characterize. Indeed, while dynamic PET images results from a decay process that can be statistically described by a Poisson distribution, acquisition and post-filtering reconstruction drastically change the nature of the noise. In the literature dedicated to factor analysis of dynamic PET images, a common and underlying assumption consists in assuming that the dynamic PET images are corrupted by an additive Gaussian or by a Poisson noise. These assumptions lead to the choice of the squared Euclidian distance and the Kullback-Leibler divergence. We propose here to consider the β-divergence, which is able to encompass a wide family of divergence measures corresponding to various noise distributions. This loss function is incorporated into three different factor models and evaluated using four sets of synthetic data.

***Index Terms*—** Dynamic PET imaging, factor analysis, β-divergence, nonnegative matrix factorization, blind source separation.

1. INTRODUCTION

Dynamic positron emission tomography (PET) is a functional imaging technique that is able to deliver relevant information on dysfunctions hardly detectable by anatomical imagery. To provide interpretable results, PET images have to be quantified, which often requires a previous estimation of global time-activity curves (TACs) (herein called *factors*) representing the concentration of tracer in a reference tissue or blood over time. The estimation of those factors along with their spatial repartitions is referred to as unmixing, factor analysis or linear blind source separation.

Most of the factor analysis techniques in the PET literature [1, 2] implicitly or explicitly assume that the residual noise corrupting the images follows a Gaussian distribution [3]. Conversely, concurrent methods exploit the intrinsic nature of the count-rates to consider that the noise behavior can be described by a Poisson distribution [4, 5]. However, depending on the acquisition setup, the reconstruction algorithm and its parameter values (e.g., the number of reconstruction iterations), the nature of the noise can be significantly altered [6]. To allow for a more flexible modeling of the noise, hybrid distributions, such as Poisson-Gaussian [7] and Poisson-Gamma [8], have been also used to improve different steps of the PET imaging pipeline. Teymurazyan *et al*. investigated the statistics of images reconstructed with different algorithms [9]. The Gamma distribution is identified as the one that better fits iterative expectation maximization (EM)-reconstructed data. Similarly, Mou et *al.* also discussed the Gamma behavior of PET data [10].

This work proposes to adapt dynamic PET unmixing to the various noise distributions encountered in reconstructed data. To this end, we resort to the β-divergence, a popular and extremely flexible cost function that encompasses a family of standard loss functions, including the Euclidean distance, the Kullback-Leibler (KL) and the Itakura-Saito (IS) divergences, directly related to the Gaussian, Poisson and Gamma noise distributions [11, 12].

To investigate the impact of various β values on the quality of factor analysis of dynamic PET images, the β-divergence is applied to three mixing models from the literature: nonnegative matrix factorization (NMF) [4], linear mixing model (LMM) [13] and specific binding linear mixing model (SLMM) [14]. While NMF simply assumes nonnegative sources, LMM adds a sum-to-one constraint on the mixing coefficients and SLMM provides more flexibility by allowing one source to spatially vary. While β-NMF has been deeply investigated in [12], β-LMM has been used to analyze hyperspectral images in [15]. This paper introduces the β-SLMM algorithm, which considers the mixing model presented in [14]. The influence of the β parameter on the performance of the dynamic PET image factor analysis is then evaluated for the three mixing models presented above.

The sequel of this paper is organized as follows. Section 2 provides a general factor analysis formulation and briefly recalls the models to be evaluated in this work. Section 3 presents the β-divergence as a measure of dissimilarity. Section 4 provides the β-SLMM algorithm that was developed for this study. Simulation results obtained with synthetic data are reported in Section 5. Section 6 concludes the paper.

Part of this work has been supported by Coordenação de Aperfeioamento de Ensino Superior (CAPES), Brazil, and the European Research Council (ERC FACTORY-CoG-681839).

Table 1. Summary of NMF, LMM and SLMM under (2)

	$\underline{\boldsymbol{\theta}}$	$\mathbf{X}(\underline{\boldsymbol{\theta}})$	$\mathcal{C}$	$\mathbf{R}(\underline{\boldsymbol{\theta}})$
NMF	$\{\mathbf{M}, \mathbf{A}\}$	$\mathbf{X} = \mathbf{MA}$	$\mathbf{A} \succeq \mathbf{0}_{K,N}$ $\mathbf{M} \succeq \mathbf{0}_{L,K}$	-
LMM	$\{\mathbf{M}, \mathbf{A}\}$	$\mathbf{X} = \mathbf{MA}$	$\mathbf{A} \succeq \mathbf{0}_{K,N}$ $\mathbf{M} \succeq \mathbf{0}_{L,K}$ $\mathbf{A}^{\mathrm{T}}\mathbf{1}_K = \mathbf{1}_N$	-
SLMM	$\{\mathbf{M}, \mathbf{A}, \mathbf{B}\}$	$\mathbf{X} = \mathbf{MA} + \left[\mathbf{E}_1\mathbf{A} \cdot \mathbf{VB}\right]$	$\mathbf{A} \succeq \mathbf{0}_{K,N}$ $\mathbf{M} \succeq \mathbf{0}_{L,K}$ $\mathbf{B} \succeq \mathbf{0}_{N_v,N}$ $\mathbf{A}^{\mathrm{T}}\mathbf{1}_K = \mathbf{1}_N$	$\|\mathbf{B}\|_{2,1} = \sum_{n=1}^{N} \|\mathbf{b}_n\|_2$

2. FACTOR ANALYSIS

In this paper, we apply three factor analysis models from the dynamic PET factor analysis literature. To this end, a general formulation of inverse dynamic PET problems will be provided. Consider an $L \times N$ dynamic PET image $\mathbf{Y}$ comprising N TACs over L time-frames. An approximation model $\mathbf{X}(\underline{\boldsymbol{\theta}})$ of P physically interpretable variables $\underline{\boldsymbol{\theta}} = \{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_P\}$ can be defined to describe matrix $\mathbf{Y}$, yielding

$$\mathbf{Y} \approx \mathbf{X}(\underline{\boldsymbol{\theta}}), \tag{1}$$

where the approximation symbol $\approx$ generalizes the relation between the model and the measured data. Given this general setting, recovering these explanatory variables $\underline{\boldsymbol{\theta}}$ can be formulated as the optimization problem

$$\hat{\underline{\boldsymbol{\theta}}} \in \underset{\underline{\boldsymbol{\theta}} \in \mathcal{C}}{\arg\min} \left\{ \mathcal{D}(\mathbf{Y}|\mathbf{X}(\underline{\boldsymbol{\theta}})) + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{R}(\underline{\boldsymbol{\theta}}), \right\} \tag{2}$$

where $\mathcal{C}$ is a set of constraints and $\mathbf{R}(\underline{\boldsymbol{\theta}}) = [r_1(\underline{\boldsymbol{\theta}}), \ldots, r_T(\underline{\boldsymbol{\theta}})]^{\mathrm{T}}$ gathers T penalizations individually weighted by the parameters in $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_T]^{\mathrm{T}}$. In (2), $\mathcal{D}(\cdot|\cdot)$ stands for the loss function that measures the discrepancy between $\mathbf{Y}$ and $\mathbf{X}(\underline{\boldsymbol{\theta}})$. In this work, we are particularly interested in the choice of this function, that will be further discussed in Section 3.

Table 1 summarizes the three factor analysis techniques, namely NMF, LMM and SLMM, considered in this paper under formulation (2). For both NMF and LMM, $\mathbf{M} = [\mathbf{m}_1, ..., \mathbf{m}_K]$ is a $L \times K$ matrix of factors and $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{a}_N]$ is a $K \times N$ matrix containing the factor coefficients. To produce a low-rank approximation of the matrix $\mathbf{X}$, we choose $K \ll L, N$. For SLMM, the factor matrix is written $\mathbf{M} = [\bar{\mathbf{m}}_1, \ldots, \mathbf{m}_K]$ where $\bar{\mathbf{m}}_1$ is the nominal specific binding factor. Moreover, "$\cdot$" is the Hadamard point-wise product, $\mathbf{E}_1$ is the matrix $[\mathbf{1}_{L,1}\mathbf{0}_{L,K-1}]$, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_{N_v}]$ is the $L \times N_v$ matrix composed of the basis elements used to describe the variability of the specific binding factor (SBF) with $N_v \ll L$, and $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_N]$ is the $N_v \times N$ matrix composed of internal variability proportions. Interested readers are invited to consult [14] for a detailed motivation of this factor model.

Moreover, $\mathbf{0}_{W,H}$ denotes the $W \times H$-matrix made of 0's, $\mathbf{1}_W$ is the W-dimensional vector made of ones and $\succeq$ stands for a component-wise inequality. Therefore the constraints represented by $\boldsymbol{\theta}_i \succeq \mathbf{0}_{W,H}$ express nonnegativity and $\mathbf{A}^{\mathrm{T}}\mathbf{1}_K = \mathbf{1}_N$ is the sum-to-one constraint.

The only model for which a penalization $\mathbf{R}(\underline{\boldsymbol{\theta}})$ with corresponding regularization parameter λ will be applied is SLMM. As in [14], the internal variability $\mathbf{B}$ is penalized to be spatially sparse with the $\ell_{2,1}$-group lasso regularizer defined as

$$\mathbf{R}(\underline{\boldsymbol{\theta}}) \triangleq \|\mathbf{B}\|_{2,1} = \sum_{n=1}^{N} \|\mathbf{b}_n\|_2. \tag{3}$$

3. THE β-DIVERGENCE

As previously discussed, most factor analysis techniques in the dynamic PET literature use the Euclidean distance or the KL divergence as measures of dissimilarity. These loss functions underlie the assumptions of additive Gaussian and Poisson noises, respectively [16]. The noise affecting dynamic PET images can be, however, more complex to model. To introduce more flexibility with respect to the noise distribution of the observed image, we propose to use the β-divergence as the loss function $\mathcal{D}(\cdot|\cdot)$ in (2), yielding

$$\mathcal{D}_\beta(\mathbf{Y}|\mathbf{X}) = \sum_{\ell=1}^{L}\sum_{n=1}^{N} d_\beta(y_{\ell,n}|x_{\ell,n}). \tag{4}$$

First introduced by Basu *et al.* [11] and successfully applied in several domains [16, 17, 18], the β-divergence is defined for $\beta \in \mathbb{R}$ as

$$d_\beta(y|x) = \begin{cases} \frac{1}{\beta(\beta-1)}(y^\beta + (\beta-1)x^\beta - \beta y x^{\beta-1}) & \beta \in \mathbb{R}\backslash\{0,1\} \\ y\log\frac{y}{x} - y + x & \beta = 1, \\ \frac{y}{x} - \log\frac{y}{x} - 1 & \beta = 0. \end{cases} \tag{5}$$

The limit cases $\beta = 1, 0$ correspond to the KL and IS divergences, respectively, while $\beta = 2$ coincides with the squared Euclidean distance. As previously noted, the IS divergence is intrinsically related to the multiplicative Gamma noise distribution and conveniently complements the squared Euclidean distance and KL divergence. Hence, the β-divergence stands out as a relevant tool for dynamic PET factor analysis.

4. BLOCK-COORDINATE DESCENT ALGORITHM

A block-coordinate descent approach is used to solve the optimization problem in (2). With this formulation, the variables $\boldsymbol{\theta}_i$ belonging to one factor analysis model of Table 1 are alternatively updated. Regarding β-SLMM, the updating rules for $\mathbf{A}$ and $\mathbf{B}$ are detailed in the following paragraphs while the update for $\mathbf{M}$ can be computed as in [15]. Algo. 1 presents the resulting β-SLMM unmixing algorithm, where all multiplications (identified by the $\cdot$ symbol), divisions and exponentiations are point-wise operations, $\mathbf{1}_{K,L}$ denotes a $K \times L$-matrix of ones and $\boldsymbol{\Gamma}_{\mathbf{B}} \triangleq \mathrm{diag}[\|\mathbf{b}_1\|_1, \cdots, \|\mathbf{b}_1\|_N]^{-1}$. Regarding the two other factor models in Table 1, the β-NMF updates and algorithm are provided in [12] and the β-LMM unmixing algorithm is a depreciated version of β-SLMM, where $\mathbf{B} = \mathbf{0}$.

4.1. Update of the factor proportions A

The optimization problem for $\mathbf{A}$ writes

Algorithm 1: β-SLMM unmixing

Data: $\mathbf{Y}$

Input: $\mathbf{A}$, $\mathbf{M}$, $\mathbf{B}$, λ

$\tilde{\mathbf{X}} \leftarrow \mathbf{MA} + \left[\mathbf{E}_1\mathbf{A} \cdot \mathbf{VB}\right]$

while *stopping criterion not satisfied* **do**

  % Update variability matrix

  $\mathbf{B} \leftarrow \mathbf{B} \cdot \left[\frac{\mathbf{1}_{N_v}^{\mathrm{T}} \mathbf{A}_{1,:} \cdot (\mathbf{V}^{\mathrm{T}}(\mathbf{Y} \cdot \tilde{\mathbf{X}}^{\beta-2}))}{\mathbf{1}_{N_v}^{\mathrm{T}} \mathbf{A}_{1,:} \cdot (\mathbf{V}^{\mathrm{T}} \tilde{\mathbf{X}}^{\beta-1}) + \lambda \mathbf{B}^k \boldsymbol{\Gamma}_{\mathbf{B}}}\right]^{\frac{1}{3-\beta}}$

  $\tilde{\mathbf{X}} \leftarrow \mathbf{MA} + \left[\mathbf{E}_1\mathbf{A} \cdot \mathbf{VB}\right]$

  % Update factor TACs

  $\mathbf{M}_{2:K} \leftarrow \mathbf{M}_{2:K} \cdot \left[\frac{(\mathbf{Y} \cdot \tilde{\mathbf{X}}^{\beta-2})\mathbf{A}_{2:K}^T}{\tilde{\mathbf{X}}^{\beta-1}\mathbf{A}_{2:K}^T}\right]$

  $\tilde{\mathbf{X}} \leftarrow \mathbf{MA} + \left[\mathbf{E}_1\mathbf{A} \cdot \mathbf{VB}\right]$

  % Update SBF factor proportion

  $\mathbf{A}_1 \leftarrow \mathbf{A}_1 \cdot \left[\frac{\mathbf{1}_L^{\mathrm{T}}((\mathbf{M}_1\mathbf{1}_N^{\mathrm{T}} + \mathbf{VB}) \cdot (\mathbf{Y} \cdot \tilde{\mathbf{X}}^{\beta-2}) + \tilde{\mathbf{x}}^{\beta})}{\mathbf{1}_L^{\mathrm{T}}((\mathbf{M}_1\mathbf{1}_N^{\mathrm{T}} + \mathbf{VB}) \cdot \tilde{\mathbf{X}}^{\beta-1} + \mathbf{Y} \cdot \tilde{\mathbf{X}}^{\beta-1})}\right]$

  % Update other factor proportions

  $\mathbf{A}_{2:K} \leftarrow \mathbf{A}_{2:K} \cdot \left[\frac{\mathbf{M}_{2:K}^T(\mathbf{Y} \cdot \tilde{\mathbf{X}}^{\beta-2}) + \mathbf{1}_{K-1,L}\tilde{\mathbf{X}}^{\beta}}{\mathbf{M}_{2:K}^T\tilde{\mathbf{X}}^{\beta-1} + \mathbf{1}_{K-1,L}(\mathbf{Y} \cdot \tilde{\mathbf{X}}^{\beta-1})}\right]$

  $\tilde{\mathbf{X}} \leftarrow \mathbf{MA} + \left[\mathbf{E}_1\mathbf{A} \cdot \mathbf{VB}\right]$

Result: $\mathbf{A}$, $\mathbf{M}$, $\mathbf{B}$

$$\min_{\mathbf{A}} \mathcal{J}(\mathbf{A}) = \mathcal{D}(\mathbf{Y}|\mathbf{MA} + \left[\mathbf{E}_1\mathbf{A} \cdot \mathbf{W}\right]) \quad \text{s.t. } \mathbf{A} \succeq \mathbf{0}_{K,N}, \mathbf{A}^{\mathrm{T}}\mathbf{1}_K = \mathbf{1}_N, \tag{6}$$

with $\mathbf{W} = \mathbf{VB}$. The sum-to-one constraint is incorporated by normalizing the factor proportions at each corresponding update with the use of a change of variable [19, 15]. More precisely, an auxiliary matrix $\mathbf{U}$ is updated and the factor proportions matrix $\mathbf{A}$ is then defined as

$$a_{kn} = \frac{u_{kn}}{\sum_k u_{kn}}. \tag{7}$$

This formulation yields a new minimization problem

$$\min_{\mathbf{U}} \mathcal{J}(\mathbf{U}) \text{ s.t. } \mathbf{U} \succeq \mathbf{0}_{K,N}, \tag{8}$$

with

$$\mathcal{J}(\mathbf{U}) = \sum_{ln} d\left(y_{ln} | \sum_k m_{lk}\left[\frac{u_{kn}}{\|\mathbf{u}_n\|_1}\right] + \left[\frac{u_{1n}}{\|\mathbf{u}_n\|_1}\right] w_{ln}\right). \tag{9}$$

The gradient of $\mathcal{J}(\mathbf{U})$ can be expressed as

$$\nabla_{u_{kn}} \mathcal{J}(\mathbf{U}) = \nabla^+_{u_{kn}} \mathcal{J}(\mathbf{U}) - \nabla^-_{u_{kn}} \mathcal{J}(\mathbf{U}) \tag{10}$$

To solve this problem, we resort to the heuristic formulation of [16], yielding the following nonnegativity-preserving update rule for the auxiliary matrix $\mathbf{U}$

$$u_{kn} = \tilde{u}_{kn}\left(\frac{\nabla^-_{u_{kn}} \mathcal{J}(\mathbf{U})}{\nabla^+_{u_{kn}} \mathcal{J}(\mathbf{U})}\right), \tag{11}$$

where $\tilde{u}_{kn}$ is the current state of u_{kn}. By adopting $\tilde{x}_{ln} = \sum_{k \neq 1} m_{lk}\tilde{a}_{kn} + \tilde{a}_{1n}w_{ln}$, it yields

$$u_{kn} = \tilde{u}_{kn} \upsilon_{kn}^{\gamma(\beta)}$$

where $\gamma(\beta)$ is $\frac{1}{2-\beta}$ for $\beta < 1$, 1 for $\beta \in [1, 2]$ and $\frac{1}{\beta-1}$ for $\beta > 2$ and

$$\upsilon_{kn} = \begin{cases} \frac{\sum_l \left(\tilde{x}_{ln}^{\beta} + (m_{l1} + w_{ln})\tilde{x}_{ln}^{\beta-2} y_{ln}\right)}{\sum_l \left((m_{l1} + w_{ln})\tilde{x}_{ln}^{\beta-1} + y_{ln}\tilde{x}_{ln}^{\beta-1}\right)}, & \text{if } k = 1; \\ \frac{\sum_l \left(\tilde{x}_{ln}^{\beta} + m_{lk} y_{ln} \tilde{x}_{ln}^{\beta-2}\right)}{\sum_l \left(m_{lk}\tilde{x}_{ln}^{\beta-1} + y_{ln}\tilde{x}_{ln}^{\beta-1}\right)}, & \text{otherwise.} \end{cases}$$

4.2. Update of the internal variability $\mathbf{B}$

The optimization problem for $\mathbf{B}$ writes

$$\min_{\mathbf{B}} \mathcal{J}(\mathbf{B}) = \mathcal{D}(\mathbf{Y}|\mathbf{MA} + \left[\mathbf{E}_1\mathbf{A} \cdot \mathbf{VB}\right]) + \lambda\|\mathbf{B}\|_{2,1} \quad \text{s.t. } \mathbf{B} \succeq \mathbf{0}_{N_v,N}, \tag{12}$$

where the parameter λ controls the trade-off between the data-fitting term and the spatial sparsity-inducing group lasso regularizer. Denoting by $\tilde{\mathbf{B}}$ the current state of $\mathbf{B}$, the model-based reconstructed data using the current estimates is defined by $\tilde{x}_{ln} = s_{ln} + \sum_i a_{1n} v_{li} \tilde{b}_{in}$ with $s_{ln} = \sum_k m_{lk} a_{kn}$. Following [15, 20], the update of $\mathbf{B}$ can be handled with majorization-minimization (MM). The data-fitting term is majorized using Jensen's inequality and the penalty function $\|\mathbf{B}\|_{2,1}$ is majorized by its tangent thanks to concavity of the square-root function. The data-fitting term is further majorized as in [20] to match the quadratic upper bound of the regularization term. Minimizing the final auxiliary function gives the following update

$$b_{in} = \tilde{b}_{in}\left(\frac{a_{1n}\sum_l v_{li} y_{ln} \tilde{x}_{ln}^{\beta-2}}{a_{1n}\sum_l v_{li}\tilde{x}_{ln}^{\beta-1} + \lambda \frac{\tilde{b}_{in}}{\|\tilde{\mathbf{b}}_n\|_2}}\right)^{\frac{1}{3-\beta}}. \tag{13}$$

5. EXPERIMENTS WITH SYNTHETIC DATA

5.1. Synthetic data generation

To evaluate the impact of varying β, the three considered factor analysis algorithms were applied to two sets of $128 \times 128 \times 64$-voxel synthetic PET images with $L = 20$ time-frames. The first set is generated from a phantom derived of a clinical PET image that was acquired using the ^{11}C-PE2I radioligand, as in [21], herein referred to as Phantom I. Phantom I is introduced on a chain of generation from the SLMM model, detailed in [14], to produce Phantom II, for which the ground-truth of the SLMM variables are known. The generation process described in [21] is applied to both phantoms to yield synthetic images with realistic count-rates properties. In particular, the process in [21] includes a reconstruction step that, in our case, is done with the standard ordered-subset expectation-maximization (OSEM) algorithm. To account for the impact of the reconstruction parameters on noise properties, the OSEM algorithm is applied with 3 and 30 reconstruction iterations for each image, yielding two datasets referred to as *3it* and *30it*, respectively. For each image and each value of reconstruction iterations, 16 samples were generated to assess statistical performance.

5.2. Compared methods

Phantom I is used to evaluate the factor modeling of *3it* and *30it* images through the reconstruction error in terms of $\mathrm{PSNR}(\hat{\mathbf{X}}) = 10\log_{10}\frac{\max(\mathbf{X}^*)^2}{\|\hat{\mathbf{X}} - \mathbf{X}^*\|_F^2}$, where $\max(\mathbf{X}^*)$ is the maximum value of the ground-truth image $\mathbf{X}^*$ and $\hat{\mathbf{X}} \triangleq \mathbf{X}(\hat{\boldsymbol{\theta}})$ is the estimated image. In this setting, β-LMM and β-NMF are compared with β in the range

Table 2. Variability penalization parameters

	λ		
	β=0	β=1	β=2
3it	2.10^{-4}	1.10^{-3}	1.10^{-3}
30it	1.10^{-4}	1.10^{-3}	1.10^{-3}

$(0, 2.4)$ with a stepsize of 0.2. Factors and factor proportions are initialized by vertex component analysis (VCA) [22] and SUnSAL [23], respectively.

Phantom II is used to compare the β-SLMM and β-LMM algorithms for $\beta \in \{0, 1, 2\}$. Factors and factor proportions are initialized with K-means and SUnSAL, respectively. The variability matrix $\mathbf{B}$ is initialized randomly. Additionally to the PSNR, for each variable we compute the normalized mean square error $\text{NMSE}(\hat{\theta}_i) = \frac{\|\hat{\theta}_i - \theta_i^*\|_F^2}{\|\theta_i^*\|_F^2}$, where θ_i^* and $\hat{\theta}_i$ are the actual and estimated variables, respectively. The stopping criterion ε is set to 10^{-4} in both settings and the parameter λ is empirically tuned with the values provided in Table 2.

5.3. Results on Phantom I

Fig. 1 shows the PSNR mean and standard deviation of the *3it* and *30it* results estimated by β-NMF. For *3it*, the higher PSNRs are obtained around $\beta \in [0.2, 0.8]$, which indicates a noise distribution that is between Gamma and Poisson. For *30it*, the best PSNRs are obtained around $\beta \in [1, 1.6]$, suggesting a more Poisson-Gaussian noise distribution.

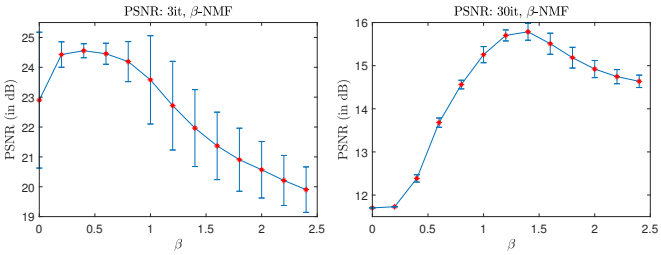

Fig. 1. PSNR mean and standard deviation obtained on the *3it* (left) and *30it* (right) images after factorization with β-NMF.

Fig. 2 shows the PSNR mean and standard deviation of the *3it* and *30it* results estimated by β-LMM. For the same quantity of factors, LMM produces more constrained results than NMF due to the sum-to-one constraints of the factor proportions. The resulting PSNRs are therefore of smallest standard deviations. The optimal values of β are around the same locations. For*3it*, the higher PSNRs are obtained around $\beta \in [0, 1]$, still corresponding to the same noise distribution as before. For *30it*, the best PSNRs are in the range of $\beta \in [0.8, 1.4]$. The distribution can still be considered more Poisson-Gaussian, though the differences between the PSNR for $\beta = 0.8$ and the previous ones ($\beta \in [0, 0.6]$) are significant.

5.4. Results on Phantom II

Table 3 presents the mean NMSE for $\mathbf{A}_1$, $\mathbf{A}_{2:K}$ and $\mathbf{A}_1 \cdot \mathbf{B}$ as well as the PSNR for the *3it* and *30it* images estimated with β-SLMM and β-LMM algorithms. As the partial volume effect may appear in $\mathbf{A}_1$ or in $\mathbf{B}$, the performance of $\mathbf{A}_1 \cdot \mathbf{B}$ is evaluated. For *3it* images, the

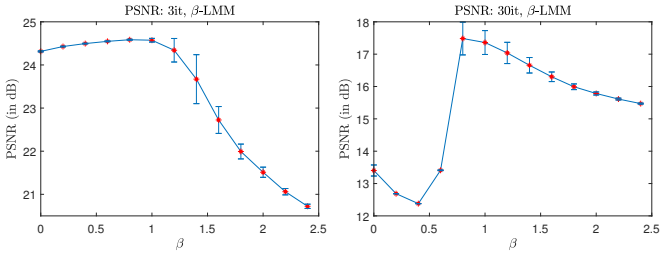

Fig. 2. PSNR mean and standard deviation obtained on the *3it* (left) and *30it* (right) images after factorization with β-LMM.

Table 3. Mean NMSE of $\mathbf{A}_1$, $\mathbf{A}_{2:K}$, $\tilde{\mathbf{M}}^1$, $\mathbf{M}^{2:K}$ and $\mathbf{A}_1 \cdot \mathbf{B}$ and PSNR of estimated image estimated by β-LMM and β-SLMM for different values of β.

		β-LMM			β-SLMM		
	β	0	1	2	0	1	2
3it	$\mathbf{A}_1$	0.312	0.286	0.301	0.305	0.303	0.306
	$\mathbf{A}_{2:K}$	0.522	0.513	**0.509**	0.531	0.527	**0.518**
	$\tilde{\mathbf{M}}^1$	**0.096**	0.187	0.273	**0.007**	**0.007**	0.008
	$\mathbf{M}^{2:K}$	0.388	0.367	**0.311**	0.385	0.373	**0.321**
	$\mathbf{A}_1 \cdot \mathbf{B}$	-	-	-	0.487	**0.476**	0.532
	PSNR	22.16	25.27	**28.26**	30.0	**30.22**	27.93
30it	$\mathbf{A}_1$	0.583	0.629	**0.558**	**0.680**	0.717	0.698
	$\mathbf{A}_{2:K}$	0.577	**0.563**	0.591	**0.580**	0.588	0.590
	$\tilde{\mathbf{M}}^1$	0.758	0.507	**0.409**	0.012	**0.010**	0.011
	$\mathbf{M}^{2:K}$	0.269	0.282	**0.239**	0.260	0.248	**0.237**
	$\mathbf{A}_1 \cdot \mathbf{B}$	-	-	-	0.872	**0.723**	0.799
	PSNR	20.36	25.51	**26.09**	25.82	**27.38**	27.02

highest PSNR is achieved by β-SLMM with $\beta = 1$, closely followed by $\beta = 0$, once again suggesting a Gamma-Poisson distribution as a good fit for the noise. Due to the strong non-convexity of β-SLMM, the minimum NMSE for the variables do not follow the same line, appearing either for all values of β. This may be not surprising, given the undetermined nature of PET images noise. For *30it*, the best PSNR is again obtained by β-SLMM, this time with $\beta = 1$, suggesting a Poisson behavior, that was also perceived in the previous *30it* setting. Once again, the minimum NMSEs are obtained with other values of β. This simulation was conducted only for illustration purposes. The β-SLMM should not be used to choose the optimal value of β, since results also depend on the parameter λ. Thus, β should be adjusted based on the standard β-NMF or β-LMM and then used into β-SLMM.

6. CONCLUSION

This paper introduced the β-divergence as a divergence measure for the factor analysis of dynamic PET images. A new algorithm for the SLMM model, using the β-divergence was also derived with multiplicative updates. Simulations illustrated the interest of considering different values of β to fit the noise distribution in dynamic PET data. This work opens the discussion on applying the β-divergence into different steps of the PET imaging pipeline, such as denoising and reconstruction.

7. REFERENCES

[1] D. C. Barber, "The use of principal components in the quantitative analysis of gamma camera dynamic studies," *Physics Med. Biol.*, vol. 25, no. 2, pp. 283–292, 1980.

[2] H.-M. Wu *et al.*, "Factor analysis for extraction of blood time-activity curves in dynamic FDG-PET studies," *J. Nuclear Med.*, vol. 36, no. 9, pp. 1714–1722, Sept. 1995.

[3] A. Sitek, E. V. R. Di Bella, and G. T. Gullberg, "Factor analysis with a priori knowledge-application in dynamic cardiac SPECT," *Physics Med. Biol.*, vol. 45, pp. 2619–2638, 2000.

[4] J. S. Lee *et al.*, "Non-negative matrix factorization of dynamic images in nuclear medicine," in *Proc. IEEE Nuclear Sci. Symp (NSS)*, 2001.

[5] D. Schulz, "Non-negative matrix factorization based input function extraction for mouse imaging in small animal PET - comparison with arterial blood sampling and factor analysis," *J. Molecular Imag. Dynamics*, vol. 02, 2013.

[6] P. Razifar *et al.*, "Noise correlation in PET, CT, SPECT and PET/CT data evaluated using autocorrelation function: a phantom study on data, reconstructed using FBP and OSEM," *BMC Med. Imag.*, vol. 5, no. 1, Aug. 2005.

[7] M. Slifstein, O. R. Mawlawi, and M. Laruelle, "Partial volume effect correction: Methodological considerations," in *Physiological Imaging of the Brain with PET*, A. Gjedde *et al.*, Eds. San Diego: Academic, 2000, ch. 11, p. 413.

[8] Z. Irace *et al.*, "Bayesian segmentation of chest tumors in PET scans using a Poisson-Gamma mixture model," in *Proc. IEEE-SP Workshop Stat. and Signal Process. (SSP)*, June 2011, pp. 809–812.

[9] A. Teymurazyan *et al.*, "Properties of noise in positron emission tomography images reconstructed with filtered-backprojection and row-action maximum likelihood algorithm," *J. Digit. Imag.*, vol. 26, no. 3, pp. 447–456, Aug. 2012.

[10] T. Mou, J. Huang, and F. O'Sullivan, "The Gamma characteristic of reconstructed PET images: Implications for ROI analysis," *IEEE Trans. Med. Imag.*, vol. PP, no. 99, pp. 1–1, 2017.

[11] A. Basu *et al.*, "Robust and efficient estimation by minimising a density power divergence," *Biometrika*, vol. 85, no. 3, pp. 549–559, 1998.

[12] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[13] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 5, no. 2, pp. 354–379, April 2012.

[14] Y. C. Cavalcanti *et al.*, "Unmixing dynamic pet images with variable specific binding kinetics," *Medical Image Analysis*, vol. 49, pp. 117 – 127, 2018.

[15] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.

[16] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[17] D. FitzGerald, M. Cranitch, and E. Coyle, "On the use of the beta divergence for musical source separation," in *Proc. IET Irish Signals Systems Conf (ISSC)*, June 2009, pp. 1–6.

[18] A. Cichocki and S. Amari, "Families of alpha- beta- and Gamma- divergences: Flexible and robust measures of similarities," *Entropy*, vol. 12, no. 6, pp. 1532–1568, 2010.

[19] J. Eggert and E. Korner, "Sparse coding and NMF," in *Proc. IEEE Int. Joint Conf. Neural Net. (IJCNN)*, vol. 4, July 2004, pp. 2529–2533 vol.4.

[20] V. Y. F. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization with the β-divergence," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1592–1605, July 2013.

[21] S. Stute *et al.*, "Analytical simulations of dynamic PET scans with realistic count rates properties," in *Proc. IEEE Nuclear Sci. Symp. Med. Imag. Conf. (NSS-MIC)*, Nov. 2015.

[22] J. Nascimento and J. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, April 2005.

[23] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, 2010.