Infimal convolution spatiotemporal PET reconstruction using total variation based priors

Maïtine Bergounioux, Evangelos Papoutsellis, Simon Stute, Clovis Tauber

Noname manuscript No.
(will be inserted by the editor)

Infimal convolution spatiotemporal PET reconstruction using total variation based priors.

Maïtine Bergounioux · Evangelos Papoutsellis · Simon Stute · Clovis Tauber

Received: date / Accepted: date

Abstract In this paper, we focus on spatiotemporal regularization of Positron Emission Tomography (PET) reconstruction. Through a minimization problem defined on a dynamic variational framework we consider regularizers based on total variation priors adapted to problems related to Poisson noise degradation. In particular, we consider spatiotemporal total variation and total generalized variation and their corresponding extensions to the infimal convolution regularization. The numerical solutions of the corresponding variational problems are performed using Primal-Dual Hybrid Gradient (PDHG) optimization methods under a diagonal preconditioning. We compare our numerical solutions with the standard Maximum Likelihood Expectation Maximization (MLEM) reconstruction for simulated dynamic brain data for different kind of radiotracers. Our results indicate that the infimal convolution approaches provide better reconstructions compared to the ground truth brain phantom.

Keywords Total Variation · Total generalized variation · Infimal convolution · Spatiotemporal regularization · Dynamic Positron Emission Tomography · Primal-Dual Hybrid Gradient optimization methods.

Mathematics Subject Classification (2000) 94A08 · 68U10 · 65K10 · 49M29

This work is supported by the Région Centre (France) - OTITE4D Project.

Maïtine Bergounioux
Institut Denis Poisson, CNRS, UMR 7013, Université d'Orléans, B.P. 6759 45067 Orléans cedex 2, France E-mail: maitine.bergounioux@gmail.com

Simon Stute
UMRS Inserm U1023 IMIV-CEA SHFJ, 91400 Orsay , France, E-mail: simon.stute@cea.fr

Evangelos Papoutsellis, Clovis Tauber
UMRS Inserm U1253 Imagerie et Cerveau Université de Tours, France, E-mail: evangelos.papoutsellis@univ-orleans.fr, epapoutsellis@gmail.com, clovis.tauber@univ-tours.fr

1 Introduction

Positron emission tomography (PET) is a nuclear imaging method that produce functional images which depict metabolic processes in the human body [37]. After injecting the patient with a radioactive substance, for instance a glucose analog such as [18F]-FDG (fluorodeoxyglucose), it is distributed across the body and retained by organs which contain the target molecule or functionality. Due to the radioactive decay a positron is emitted and annihilated with an electron. At the annihilation point two gamma rays are emitted in opposite directions and are captured through a PET detector ring placed around the patient. These measurements are in form of photon counts or coincidence events and can be presented as projection matrices called *sinograms*. Due to the randomness in the photon counting process, the sinogram images are corrupted by the so called *photon* noise. It is often referred as *Poisson* noise, since the number of events recorded obeys the Poisson probability distribution. PET image reconstruction is a tomographic inverse problem that consists in using the projection data to find the image of the spatial distribution of the radiotracer. Due to the noise, the reconstruction of PET images is an ill-posed problem and its solution is unstable in the sense that a small perturbation of the data can lead to an unpredictable change in the image estimate. It is therefore necessary to use some form of regularization to constrain the solution to physically acceptable values.

In this paper, we focus on dynamic PET imaging. Compared to static PET imaging, the dynamic setting provides a spatiotemporal information of the distribution of a radiotracer. Kinetic variations of the radiopharmaceuticals are described within organ of interest over time. In this case, PET data are a series of frames, a stack of sinogram images that capture the radiotracer distribution over the spatial and temporal domains. In order to describe the abrupt change of the tracer activity, after its administration, it is common to use shorter time intervals in the earlier stage of the scan. Therefore, one encounters dynamic PET images with low number of photons per frame which tends to significantly noisier images compared to static PET imaging. On the other hand, in the later frames, larger time intervals are generally used since there is less radioactivity and lesser radioactivity variations. This reduces the photon noise as more photon counts are recorded. Given the noisy dynamic PET measurements $\boldsymbol{g}$, our goal is to better reconstruct the spatiotemporal radioactive distribution $\boldsymbol{u}$ via a variational problem defined as

$$\operatorname*{argmin}_{\boldsymbol{u}} \mathcal{H}(\boldsymbol{g}, \mathcal{R}\boldsymbol{u}) + \mathcal{N}(\boldsymbol{u}).$$

The first term represents a distance term, determined by the statistics of the noise, i.e., Poisson noise and the second term enforces a prior information on the regularized solution $\boldsymbol{u}$. In the forthcoming analysis we fix the distance term, referred also as *fidelity* term, as the *Kullback-Leibler* divergence and compare different regularizers $\mathcal{N}(\boldsymbol{u})$ on dynamic PET data for different radiotracers. In terms of the regularizers, we choose edge-preserving priors that are extended to a dynamic setting. In the literature, the classical Total Variation (TV) regularization, [29], as well as its recent extension, i.e., the Total Generalized Variation, [4] have been applied to a plethora of imaging applications with respect to emission tomography, PET reconstruction and in general to Poisson related problems. Here, we mention a few of them that are related to both static and dynamic PET reconstruc-

tion. For example, in [5,18], the TV penalization is combined with the standard Expectation-Maximization (EM) algorithm, leading to alternating EM and TV reconstruction steps. Moreover, in [14], a total variation minimization is applied to parametric PET imaging, see also [38] for the dynamic case. Besides, the regularization only on the image space, in [6], the authors propose a combine approach of TV regularization for both the image and the sinogram space. We point out also, the works of [27,34]. In [27], TV is applied to each time-frame and in [34] a post-reconstruction to dynamic PET is proposed using anisotropic diffusion.

Although, TV regularizer can eliminate the noise while preserving the edges in the image, it promotes solutions with piecewise constant structures and the well known *staircasing* artifacts appear in smooth regions. One remedy to this artifact is to extend to high order methods such as the TGV which also allows linear and polynomial smoothness in addition to the preservation of edges. We refer the reader to [13], [19], where TGV regularization is applied to Poisson related problems such as PET deconvolution and a combined MRI-PET reconstruction respectively.

In our framework, we use the aforementioned TV and TGV regularizers that are extended to a spatiotemporal form for dynamic PET reconstruction. In addition, motivated by the recent works of [16,30], we extend the latter priors to the Infimal Convolution (IC) framework, equivalently we have the ICTV and ICTGV regularizers defined as

$$\mathcal{N}(\boldsymbol{u}) = \begin{cases} \mathrm{ICTV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u}) = \min\limits_{\boldsymbol{v}} \beta_1 \,\mathrm{TV}_{\boldsymbol{\kappa}}(\boldsymbol{u}-\boldsymbol{v}) + \beta_0 \,\mathrm{TV}_{1-\boldsymbol{\kappa}}(\boldsymbol{v}) \\ \mathrm{ICTGV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u}) = \min\limits_{\boldsymbol{v}} \beta_1 \,\mathrm{TGV}_{\boldsymbol{\kappa}}(\boldsymbol{u}-\boldsymbol{v}) + \beta_0 \,\mathrm{TGV}_{1-\boldsymbol{\kappa}}(\boldsymbol{v}) \end{cases}$$

The TV and TGV regularizers are weighted by a positive vector $\boldsymbol{\kappa} = (\kappa, 1-\kappa)$, with $\kappa \in (0,1)$ that balances the strength of the penalization between the spatial and temporal domains, for instance the x,y and the t directions in a 2D+t dynamic framework. The ICTV and ICTGV constitute a combination of TV and TGV regularizers respectively, that are weighted symmetrically via $\boldsymbol{\kappa}$. In addition, we have a positive vector $\boldsymbol{\beta} = (\beta_1, \beta_0)$ that balances the two terms in the TV/TGV regularizers in terms of the fidelity term $\mathcal{H}(\boldsymbol{g}, \mathcal{R}\boldsymbol{u})$. As we present in the following, the additional variable $\boldsymbol{v}$ provides a better reconstruction in the spatiotemporal domain and it is able to reduce the artifacts that TV and TGV priors create.

Outline of the paper: In Section 2, we introduce the spatiotemporal framework for dynamic PET reconstruction. We focus on the choice of priors used in our variational problem as well as the corresponding fidelity term. We continue with a brief summary of the proposed algorithm which computes numerically the solutions via a primal-dual approach. In Section 4, we describe the numerical realization of the saddle point form for the proposed regularizers. Finally, in Section 5, we discuss our results for different PET radiotracers applied to simulated brain phantoms and compare them with standard methods such as the MLEM reconstruction.

2 Spatiotemporal Variational Framework for PET reconstruction

In this section, we describe the spatiotemporal variational context suitable for dynamic PET reconstruction. This is achieved by a minimization problem of an energy functional $E : \mathbb{R}^{N\times K} \to \mathbb{R}$ i.e.,

$$\operatorname*{argmin}_{\boldsymbol{u}} E(\boldsymbol{u}) \;\text{ with }\; E(\boldsymbol{u}) = \mathcal{H}(\boldsymbol{g}, \mathrm{R}\boldsymbol{u}) + \mathcal{N}(\boldsymbol{u}). \tag{1}$$

Note that we focus on the finite dimensional case and we refer the reader to Section 4 for the details on the numerical approximation.

The corresponding solution of (1) is an image sequence, a collection of frames that represents the radioactivity of the injected radiotracer denoted as $\boldsymbol{u} = (u_{j,k}) \in \mathbb{R}^{N\times K}$, where N represents the spatial dimension (spatial bins) and K is the number of (temporal) frames. In the following, we describe the form of the fidelity term in the dynamic PET setting as well as the proposed edge preserving priors.

2.1 Fidelity term

During the PET scan, the measurements are associated to a linear operator $\mathcal{R}$ that incorporate the geometry of the acquisition and is described through the Radon transform

$$\mathcal{R}u(\theta, s) = \int_{y\in\theta^{\perp}} u(s\theta + y)\, dy, \tag{2}$$

where the integral is considered along the orthogonal subspace $\theta^{\perp} = \Big\{x \in \mathbb{R}^N : x \cdot \theta = 0\Big\}$, see [24] for more details. The equation (2) is applied for every temporal bin and its discrete form is written as

$$(\mathrm{R}u)_{i,k} := \sum_{j=1}^{N} R_{i,j} u_{j,k}, \quad 1 \le i, k \le M, K \tag{3}$$

where $\mathrm{R} = (\mathrm{R}_{i,j})$ is the *system matrix* corresponding to the discretized operator $\mathcal{R}$. The coefficients of this matrix express in fact probability values. For instance, $\mathrm{R}_{i,j}$ is the probability of an emitted photon from voxel j to be recorded in the ith tube of response (TOR) and when there is no intersection of the ith TOR to a voxel j then $\mathrm{R}_{i,j} = 0$.

In order to capture the physical process of the photon counting process during PET acquisition, we assume that the measurements follow an independent Poisson distribution. In addition to the true coincidence events, one has to take into account the random and scattered events denoted by [RD] and [SC] respectively. The true coincidences occur when both photons from the annihilation point are detected without any interaction or deflection within the coincidence time-window. A scattered event occurs when at least one of the emitted photons is detected at the wrong line of response due to a deflection. A random event happens when two photons from different annihilation points are detected on the same detector within the coincidence time window. Both random and scattered events increase the statistical noise of the data and behave as a background noise. We denote the input PET data e.g., the prompt events with $\boldsymbol{g} = (g_{i,k}) \in \mathbb{R}^{M\times K}$, where M is the dimension of the projection space and for every time frame $k = 1, \ldots, K$ we have that

$$g_{i,k} \sim \mathbf{P}\Big(\gamma_k\big(\sum_{j=1}^{N} D_k \mathrm{R}_{i,j} u_{i,k} + [SC]_{i,k} + [RD]_{i,k}\big)\Big), \tag{4}$$

where $\mathbf{P}(\lambda)$ describes the Poisson distribution with parameter $\lambda = D_k(\mathrm{R}u)_{i,k} + [SC]_{i,k} + [RD]_{i,k}$. Here, $(D_k)_{k=1}^{K}$ is the decay correction factor that depends on the specific radiotracer and the frame duration (see Section 5). Moreover, $\boldsymbol{\gamma} =$

$(\gamma_k)_{k=1}^K$ is a positive parameter that incorporates the temporal duration over two consecutive frames, i.e.,

$$\gamma_k = \Delta t_k = t_{k+1} - t_k, \quad 1 \leq k \leq K-1 \tag{5}$$

In practice, as we observe in Table 1 in Section 5, the time duration over two consecutive frames variates during a PET acquisition where shorter time intervals are considered in the beginning and longer at the end. Hence, we have a non uniform time discretization. In the following we denote $\boldsymbol{\eta} = (\eta_{i,k}) := [SC]_{i,k} + [RD]_{i,k}$.

The fidelity term related to the Poisson noise is known as the *Kullback-Leibler* divergence, see for instance [3]. Recall that the generalized Kullback-Leibler divergence for $\boldsymbol{g} = (g_{i,k}) > 0$ is

$$D_{KL}(\boldsymbol{g}, D\mathrm{R}\boldsymbol{u}+\boldsymbol{\eta}) = \begin{cases} \sum\limits_{i,k=1}^{M,K} \gamma_k \left(g_{i,k} \log \frac{g_{i,k}}{D_k(\mathrm{R}u)_{i,k}+\eta_{i,k}} + D_k(\mathrm{R}u)_{i,k} + \eta_{i,k} - g_{i,k} \right), u_{i,k} \geq 0 \\ +\infty, \text{ otherwise.} \end{cases} \tag{6}$$

Note that $\boldsymbol{\eta} = (\eta_{i,k}) > 0$. The fidelity term with respect to the minimization over $\boldsymbol{u}$ is written as

$$\mathcal{H}(\boldsymbol{g}, D\mathrm{R}\boldsymbol{u}+\boldsymbol{\eta}) = \sum_{i,k=1}^{M,K} \gamma_k \left(D_k(\mathrm{R}u)_{i,k} - g_{i,k} \log \left(D_k(\mathrm{R}u)_{i,k} + \eta_{i,k} \right) \right) + \mathbb{I}_{\{\boldsymbol{u} \geq 0\}}(\boldsymbol{u}). \tag{7}$$

where $\mathbb{I}_C(\boldsymbol{z})$ is the indicator function of C, with $\mathbb{I}_C(\boldsymbol{z}) = 0$ if $\boldsymbol{z} \in C$ and $+\infty$ otherwise. The positivity constraint is necessary in order to guarantee that the fidelity is well defined, since $(\mathrm{R}u)_{i,k} \geq 0$, if $u_{i,k} \geq 0$. The above distance term is a direct consequence of the Bayesian approach and the maximum a-posteriori probability (MAP) estimation of the negative log-likelihood function. Due to (4) the conditional probability of $\boldsymbol{g}$ which obeys a Poisson distribution given an image sequence $\boldsymbol{u}$ is

$$\mathbf{P}(\boldsymbol{g}|\boldsymbol{u}) = \prod_{i,k=1}^{M,K} \frac{\gamma_k (D_k(\mathrm{R}u)_{i,k} + \eta_{i,k})^{g_{i,k}} \, e^{-\gamma_k (D_k(\mathrm{R}u)_{i,k} + \eta_{i,k})}}{g_{i,k}!}. \tag{8}$$

Now, we impose the probability density of $\boldsymbol{u}$, which describes a geometric prior on $\boldsymbol{u}$ and is expressed as a Gibbs type density, [3], i.e.,

$$\mathcal{P}(\boldsymbol{u}) = e^{-\mathcal{N}(\boldsymbol{u})}.$$

Using the Bayes' Theorem,

$$\mathbf{P}(\mathbf{u}|\mathbf{g}) = \frac{\mathbf{P}(\boldsymbol{g}|\boldsymbol{u})\mathbf{P}(\boldsymbol{u})}{\mathbf{P}(\boldsymbol{g})},$$

the MAP estimation of $-\log \mathbf{P}(\boldsymbol{u}|\boldsymbol{g})$ and neglecting the denominator term, we conclude with (1), where the distance term is given in (7).

2.2 Regularizers

The next step is the choice of the regularizer $\mathcal{N}(\boldsymbol{u})$. We present the edge preserving priors that are used in this paper and are adapted for the spatiotemporal PET reconstruction.

2.2.1 TV/TGV regularizers

For the spatiotemporal TV regularization, an additional temporal direction is considered and through two positive regularizing parameters it defines a trade-off between spatial and temporal penalization. We let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, where $\alpha_1 > 0$, $\alpha_2 \geq 0$ for the 2D+time framework and define

$$\begin{aligned} \mathrm{TV}_{\boldsymbol{\alpha}}(\boldsymbol{u}) &= \|\nabla_{\boldsymbol{\alpha}}\boldsymbol{u}\|_1 = \sum \boldsymbol{\gamma}|\nabla_{\boldsymbol{\alpha}}\boldsymbol{u}|, \text{ with} \\ |\nabla_{\boldsymbol{\alpha}}\boldsymbol{u}| &= \sqrt{(\alpha_1\partial_x\boldsymbol{u})^2 + (\alpha_1\partial_y\boldsymbol{u})^2 + (\alpha_2\partial_t\boldsymbol{u})^2}. \end{aligned} \tag{9}$$

Similarly to the fidelity term, we allow a non uniform time discretization for the regularizer using the parameter $\boldsymbol{\gamma}$, see (5). The parameters α_1, α_2 are not predefined and require a specific tuning in order to find the optimal spatial and temporal TV smoothing. If $\alpha_2 = 0$, then the TV regularizer acts only spatially and no temporal information is considered. This first-order regularizer allows jump discontinuities but introduces the staircasing effect. In the static case, this is a result of piecewise constant approximations of smooth regions. Consequently, in the dynamic setting except of the spatial staircasing, one witness staircasing along the temporal direction where two consecutive frames for example are overlapping with each other and certain structures appear falsy to some regions. A suitable path towards an edge preserving and at the same time smooth reconstruction is to combine first and second order derivatives, as in[2,15,25].

We examine the second order total generalized variation (TGV), presented in [4], where an additional variable $\boldsymbol{v}$ finds an optimal balance between the first and second derivative of a function. For the spatiotemporal TGV, we write the following minimization problem

$$\mathrm{TGV}_{\boldsymbol{\alpha}}(\boldsymbol{u}) = \min_{\boldsymbol{v}} \|\nabla_{\boldsymbol{\alpha}}\boldsymbol{u} - \boldsymbol{v}\|_1 + \sqrt{2}\,\|\mathcal{E}_{\boldsymbol{\alpha}}\boldsymbol{v}\|_1\,. \tag{10}$$

where $\mathcal{E}_{\boldsymbol{\alpha}}\boldsymbol{v} = \frac{1}{2}(\nabla_{\boldsymbol{\alpha}}\boldsymbol{v} + \nabla_{\boldsymbol{\alpha}}\boldsymbol{v}^T)$ denote the weighted $\boldsymbol{\alpha}$-symmetrized gradient. In (10), the minimization is considered over vector fields $\boldsymbol{v}$ and the weights have been fixed in front of the two norms of (10). In previous imaging applications related to TGV, these parameters have been proven to give desirable results, see for instance [21,16,20]. The intuition behind TGV is that in smooth regions, both on the spatial as well as in the temporal domains, $\nabla^2_{\boldsymbol{\alpha}}\boldsymbol{u}$ is relatively small which forces that $\nabla_{\boldsymbol{\alpha}}\boldsymbol{u} = \boldsymbol{v}$. Moreover, in piecewise constant regions, $\nabla^2_{\boldsymbol{\alpha}}\boldsymbol{u} = \boldsymbol{v}$ is significantly large and therefore is reasonable to have $\boldsymbol{v} = 0$. In this way, the additional variable $\boldsymbol{v}$ enables the regularized solution $\boldsymbol{u}$ to find an *optimal* structure in the piecewise constant and smooth regions.

2.2.2 An infimal convolution approach

In the dynamic PET measurements one has to choose between longer scans with good counting statistics and shorter scans that are noisy but preserve temporal resolution. Therefore, an usual and fair choice is to select shorter scans in the beginning of the acquisition where there is an abrupt change of activity of the radioactive tracer and longer scans at the end. In this setting, different parts of the PET data may require either strong spatial regularization or strong temporal regularization. One way to model this behaviour is using an infimal convolution operation, see [8]. In the static case, the concept of the infimal convolution operation is to combine two different regularizers J_1, J_2 with different geometric priors via a minimization problem, i.e.,

$$\mathcal{N}(u) = \inf_v J_1(u - v) + J_2(v).$$

Using this approach, we can decompose our image into two components that promotes different geometrical structures. For instance using first and second derivatives, which capture piecewise constant and smooth structures respectively, see [31].

In a recent work, the authors in [16] introduce the infimal convolution total variation regularization extended to the video image processing framework. Compared to the static case, they use the same TV geometrical prior but weighted differently and either promote strong or weak spatiotemporal regularization. Let $\boldsymbol{\kappa} = (\kappa, 1-\kappa)$, where $\kappa \in (0,1)$, then the infimal convolution total variation regularization, denoted as ICTV, is

$$\begin{aligned} \mathrm{ICTV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u}) &= \min_{\boldsymbol{v}} \beta_1 \left\|\nabla_{\boldsymbol{\kappa}}(\boldsymbol{u}-\boldsymbol{v})\right\|_1 + \beta_0 \left\|\nabla_{\mathbf{1}-\boldsymbol{\kappa}}\boldsymbol{v}\right\|_1 \\ &= \min_{\boldsymbol{v}} \beta_1 \,\mathrm{TV}_{\boldsymbol{\kappa}}(\boldsymbol{u}-\boldsymbol{v}) + \beta_0 \,\mathrm{TV}_{1-\boldsymbol{\kappa}}(\boldsymbol{v}) \end{aligned} \tag{11}$$

where the positive parameter $\boldsymbol{\beta} = (\beta_1,\, \beta_0)$ is responsible for a proper balance between the terms of the ICTV functional and the corresponding fidelity term in (7). Under the infimal convolution method, we can decompose the regularized solution $\boldsymbol{u}$ into two components $\boldsymbol{u} - \boldsymbol{v}$ and $\boldsymbol{v}$ which are penalized by a total variation prior and capture different spatiotemporal regularity depending on the triple $(\beta_1, \beta_0, \kappa)$. As we observe in Section 5, finding an optimal balancing of these parameters, one can achieve a higher quality reconstruction compared to the TV one, since a more tailored treatment is forced via the additional variable $\boldsymbol{v}$. In fact, we show that different degrees of temporal activity are captured through this robust decomposition of $\boldsymbol{v}$ and $\boldsymbol{u} - \boldsymbol{v}$.

The reason behind the choice of $\boldsymbol{\kappa}$ is that we would like to enforce locally either a strong or weak spatial-temporal regularization. Under the extreme cases, where $\kappa \to 0, 1$, one applies only spatial regularization on the $\boldsymbol{u} - \boldsymbol{v}$ component and only strong temporal regularization on the $\boldsymbol{v}$ component and vice versa. Hence, we do not treat the temporal noise on the $\boldsymbol{u} - \boldsymbol{v}$ component and the spatial noise in the $\boldsymbol{v}$ component. The restriction on the $(0, 1)$ interval will allow an interchange between these two components in terms of spatiotemporal smoothing.

In a recent paper, see [30], a natural extension of the ICTV regularizer is proposed using a TGV regularizer instead and applied to dynamic MRI recon-

struction. It is defined below as

$$\mathrm{ICTGV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u}) = \min_{\boldsymbol{v}} \beta_1 \,\mathrm{TGV}_{\boldsymbol{\kappa}}(\boldsymbol{u}-\boldsymbol{v}) + \beta_0 \,\mathrm{TGV}_{1-\boldsymbol{\kappa}}(\boldsymbol{v}). \tag{12}$$

For the ICTGV regularizer, not only we can eliminate the spatiotemporal staircasing but also benefit using the infimal convolution approach. In this situation the decomposition on the $\boldsymbol{v}$, $\boldsymbol{u}-\boldsymbol{v}$ components relies on the TGV regularity instead of TV. Hence, spatiotemporal smooth structures are promoted which is a more reasonable assumption in the dynamic PET reconstruction.

Our goal is to examine the effect on these regularizers defined above, i.e.,

$$\mathcal{N}(\boldsymbol{u}) \in \left\{\mathrm{TV}_{\boldsymbol{\alpha}}(\boldsymbol{u}), \mathrm{ICTV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u}), \mathrm{TGV}_{\boldsymbol{\alpha}}(\boldsymbol{u}), \mathrm{ICTGV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u})\right\}. \tag{13}$$

for different types of radiotracers in the dynamic PET reconstruction. In the following section, we present a brief summary of the Primal-Dual numerical realization of (1), see [9] under a diagonal preconditioning proposed in [26]. Finally, let us mention that existence and uniqueness of the minimizer $\boldsymbol{u}$ for the problem (1) using the fidelity term (7) and the regularizers in (13) is proved using the direct method of calculus of variations. Since we deal with the discrete setting the proof for existence is straightforward and we omit it. In terms of uniqueness, note that the fidelity term is strictly convex since we assume that $\boldsymbol{g} = (g_{i,k}) > 0$.

3 An overview of the PDHG algorithm

In order to obtain a numerical solution of (1), we use the modified primal-dual hybrid gradient (PDHGM) algorithm as proposed in [9] ; it can be identified as a generalization of the PDHG algorithm found in [11]. In this section, we review the basic concepts on the primal-dual framework and focus on the diagonal preconditioning version presented in [26].

Let X and Y with $\dim X = n$, $\dim Y = m$, two finite dimensional vector spaces endowed with standard inner products $\langle\cdot,\cdot\rangle$. Let $\mathcal{K} \in \mathcal{L}(X,Y)$ a continuous linear operator with operator norm

$$\|\mathcal{K}\| = \max\left\{\|\mathcal{K}x\|_Y : \|x\|_X \leq 1\right\}.$$

We define $\mathcal{G} : X \to \mathbb{R}\cup\{\infty\}$, $\mathcal{F} : Y \to \mathbb{R}\cup\{\infty\}$ two linear convex functionals with a *simple* structure in the sense that

$$x = (I + \tau\partial\mathcal{G})^{-1}(\hat{x}) = \operatorname*{argmin}_x \mathcal{G}(x) + \frac{1}{2\tau}\|x-\hat{x}\|_X^2 \tag{14}$$

i.e., the *resolvent* operator is easy to solve and has a closed-form solution. Our goal is to write the general minimization problem

$$\min_{x\in X} \mathcal{G}(x) + \mathcal{F}(\mathcal{K}x) \tag{15}$$

that describes (1) into a general saddle-point problem,

$$\min_{x\in X}\max_{y\in Y} \langle \mathcal{K}x, y\rangle_Y + \mathcal{G}(x) - \mathcal{F}^*(y), \tag{16}$$

using the Fenchel duality (see [10] for example). Here $\mathcal{F}^*$ denotes the convex conjugate of $\mathcal{F}$:

$$\forall y_0 \in Y \qquad \mathcal{F}^*(y_0) = \sup_{y \in Y} \langle y, y_0 \rangle - \mathcal{F}(y) \ .$$

We recall that the dual maximization problem of (15) is

$$\max_{y \in Y} -\mathcal{G}^*(-\mathcal{K}^T y) - \mathcal{F}^*(y). \tag{17}$$

The generic PDHG method is Algorithm 1, described below.

Algorithm 1 Generic PDHG

Require: Choose τ, σ two positive real numbers, $\theta \in (0,1)$ and $(x_0, y_0) \in X \times Y$

Update: x_ω, y_ω as

$$\begin{cases} x_{\omega+1} = (I + \tau \partial \mathcal{G})^{-1} (x_\omega - \tau \mathcal{K}^T y_\omega) \\ \overline{x}_{\omega+1} = x_{\omega+1} + \theta(x_{\omega+1} - x_\omega) \\ y_{\omega+1} = (I + \sigma \partial \mathcal{F}^*)^{-1} \left(y_\omega + \sigma \mathcal{K}(\overline{x}_{\omega+1}) \right) \end{cases}$$

In general the efficiency of the PDHG methods relies on the splitting of the initial problem (16) into subproblems that are easier to solve and have the form described in Algorithm 1. Note that the solution of these subproblems can be computed efficiently with high numerical precision using (14) for example. Here, $\partial \mathcal{F}^*$, $\partial \mathcal{G}$ denote the subdifferential of $\mathcal{F}^*$, $\mathcal{G}$ and the expressions $(I + \tau \partial \mathcal{G})^{-1}$ and $(I + \sigma \partial \mathcal{F}^*)^{-1}$ are the resolvent operators of $\mathcal{G}$ and $\mathcal{F}^*$ respectively. Under the conditions that $\theta = 1$ and $\tau\sigma \|\mathcal{K}\| < 1$, one can show that Algorithm 1 converges, see in [9, Theorem 1].

However, one of the main difficulties of these methods is the appropriate choice of the primal-dual step sizes τ and σ. Indeed, this choice has an important impact on the speed of the algorithm and in some cases, the computation of $\|\mathcal{K}\|$ is quite difficult to obtain, due to its complicated structure. In order to handle these issues, different methods have been proposed for a suitable tuning of these parameters, especially via an adaptive scheme which improves the convergence speed. For instance, assuming additional conditions on $\mathcal{F}$ or $\mathcal{G}$, the above algorithm can be accelerated with an iterative update of θ, σ and τ, see [9,35]. In addition, the authors in [12], introduce a *linesearch* approach where the primal-dual step sizes are tuned adaptively so the primal and dual residuals of (15),(17) have roughly the same value. In the same direction, we refer also to [16] and [22].

Algorithm 2 Preconditioned PDHG

Require: Choose T, Σ two symmetric positive definite matrices, $\theta \in [0,1]$ and $(x_0, y_0) \in X \times Y$

Update: x_ω, y_ω as

$$\begin{cases} x_{\omega+1} = (I + T \partial \mathcal{G})^{-1} (x_\omega - T \mathcal{K}^T y_\omega) \\ \overline{x}_{\omega+1} = x_{\omega+1} + \theta(x_{\omega+1} - x_\omega) \\ y_{\omega+1} = (I + \Sigma \partial \mathcal{F}^*)^{-1} \left(y_\omega + \Sigma \mathcal{K}(\overline{x}_{\omega+1}) \right) \end{cases}$$

Here, we choose to follow the preconditioning technique for the PDHG algorithm, see Algorithm 2. The convergence is guaranteed provided that $\theta = 1$ and $\left\|\Sigma^{1/2}\mathcal{K}T^{1/2}\right\| < 1$, see [26, Theorem 1]. For the preconditioning of the symmetric, positive definite matrices T and Σ, we use Lemma 2 in [26] where the corresponding matrices have a diagonal structure defined as $T = \mathrm{diag}(\boldsymbol{\tau})$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$ and $\Sigma = \mathrm{diag}(\boldsymbol{\sigma})$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_m)$ with

$$\tau_j = \frac{1}{\sum_{i=1}^{m} |\mathcal{K}_{i,j}|^{2-\alpha}} \quad \text{and} \quad \sigma_i = \frac{1}{\sum_{j=1}^{n} |\mathcal{K}_{i,j}|^{\alpha}}, \tag{18}$$

for any $\alpha \in [0, 2]$. The resolvent operators are now defined as

$$x = (I + T\partial\mathcal{G})^{-1}(\hat{x}) = \underset{x\in X}{\mathrm{argmin}}\ \mathcal{G}(x) + \frac{1}{2}\|x - \hat{x}\|_X^2 \quad \text{with} \quad \|x\|_X^2 = \left\langle T^{-1}x, x\right\rangle \tag{19}$$

$$y = \left(I + \Sigma\partial\mathcal{F}^*\right)^{-1}(\hat{y}) = \underset{y\in Y}{\mathrm{argmin}}\ \mathcal{F}^*(y) + \frac{1}{2}\|y - \hat{y}\|_Y^2 \quad \text{with} \quad \|y\|_Y^2 = \left\langle \Sigma^{-1}y, y\right\rangle. \tag{20}$$

Compared to other adaptive methods, it provides a significant improvement on acceleration of these problems, see for instance [32]. Additionally, we avoid to estimate the operator norm of $\mathcal{K}$ which in the case of PET reconstruction is quite complicated, see Section 4.2. Moreover, it is quite simple to implement since only one auxiliary parameter needs to be tuned, see (18). Finally, let us mention that in the case of static PET reconstruction, this approach has been studied in [7] and [32] with a total variation penalization. In the next section, we detail the numerical implementation of our problem and discuss the saddle point forms for each regularizer.

4 Numerical Implementation

Here, we focus on the numerical solution of the minimization problem (1) using the regularizers and the fidelity terms defined in Section 2. We use the diagonal preconditioning, described in the previous section. We first give an insight of the spatiotemporal 2D+time discrete framework and continue with the saddle point forms for each choice of the regularizer. A similar analysis can be extended to 3D+time setting.

4.1 Discrete Setting

We begin with the discretization of the 2D+time spatiotemporal domains of $\Omega \times \mathcal{T}$ and $\Sigma \times \mathcal{T}$ which denotes the image and sinogram domains respectively. The discrete domains Ω and Σ have the following grid forms

$$\begin{aligned} \Omega &= \{(i, j) : i, j \in \mathbb{N},\ 1 \le i \le N_1, 1 \le j \le M_1\}, \\ \Sigma &= \left\{(\hat{\imath}, \hat{\jmath}) : \hat{\imath}, \hat{\jmath} \in \mathbb{N},\ 1 \le \hat{\imath} \le \hat{N}_1, 1 \le \hat{\jmath} \le \hat{M}_1\right\}, \\ \mathcal{T} &= \{k : k \in \mathbb{N}, 1 \le k \le K\}. \end{aligned} \tag{21}$$

Here K is the number of frames, $\hat{N}_1$ the numbers of rays and $\hat{M}_1$ represents the number of angles in the projection space. The solution $\boldsymbol{u}$ and the input dynamic

PET data $\boldsymbol{g}$ of (1) belongs to $U = \mathbb{R}^{N_1 \times M_1 \times K}$ and $\hat{U} = \mathbb{R}^{\hat{N}_1 \times \hat{M}_1 \times K}$ respectively. We also define the corresponding sets needed for the definition of the discrete gradient and symmetrized gradient operators associated to the TV and TGV regularizers as

$$V = U \times U \times U,\ W = \underbrace{U \times \cdots \times U}_{6}.$$

Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ be a vector of parameters with $\alpha_1 > 0$ and $\alpha_2 \geq 0$, penalizing the spatial and temporal domains. The gradient operator is

$$\nabla_{\boldsymbol{\alpha}} : U \to V \text{ with } \nabla_{\boldsymbol{\alpha}} \boldsymbol{u} = (\alpha_1 \partial_x^+ u, \alpha_1 \partial_y^+ u, \alpha_2 \partial_t^+ u)^T,$$

where $\partial_x^+, \partial_y^+, \partial_t^+$ are the forward finite difference operators with Neumann boundary conditions. For the spatial and temporal domains, i.e., with respect to x, y and t directions, we write

$$\partial_x^+ u = \begin{cases} \frac{u_{i+1,j,k} - u_{i,j,k}}{\Delta x_i}, & \text{if } 1 \leq i < N_1 \\ 0, & \text{if } i = N_1 \end{cases}, \ \partial_y^+ u = \begin{cases} \frac{u_{i,j+1,k} - u_{i,j,k}}{\Delta y_j}, & \text{if } 1 \leq j < M_1 \\ 0, & \text{if } j = M_1. \end{cases}$$

$$\partial_t^+ u = \begin{cases} \frac{u_{i,j,k+1} - u_{i,j,k}}{\Delta t_k}, & \text{if } 1 \leq k < K \\ 0, & \text{if } k = K. \end{cases}$$

Note that in the spatial domain the discrete step size along the x and y directions is constant, i.e., $\Delta x_i = \Delta y_j = x_{i+1} - x_i = y_{j+1} - y_j$. In terms of temporal direction, we have $\Delta t_k = t_{k+1} - t_k$ which is not necessarily constant. In fact, in our simulations the time step size variates during the acquisition see Table 1 in Section 5. We also define the divergence operator $\mathrm{div}_{1,\boldsymbol{\alpha}} : V \to U$ such that $\mathrm{div}_{1,\boldsymbol{\alpha}} = -\nabla_{\boldsymbol{\alpha}}^*$: for $\boldsymbol{v} = (v^1, v^2, v^3) \in V$, we write

$$\mathrm{div}_{1,\boldsymbol{\alpha}}(\boldsymbol{v}) = \alpha_1(\partial_x^- v^1 + \partial_y^- v^2) + \alpha_2 \partial_t^- v^3.$$

Here, the $\partial_x^-, \partial_y^-, \partial_t^-$ are the corresponding backward finite differences operators, i.e.,

$$\partial_x^- u = \begin{cases} \frac{u_{i,j,k} - u_{i-1,j,k}}{\Delta x_i}, & \text{if } 1 < i < N_1 - 1, \\ \frac{u_{1,j,k}}{\Delta x_1}, & \text{if } i = 1, \\ \frac{-u_{N_1-1,j,k}}{\Delta x_{N_1-1}} & \text{if } i = N_1. \end{cases}, \ \partial_y^- u = \begin{cases} \frac{u_{i,j,k} - u_{i,j-1,k}}{\Delta y_j}, & \text{if } 1 < j < M_1 - 1, \\ \frac{u_{i,1,k}}{\Delta y_1}, & \text{if } j = 1, \\ \frac{-u_{i,M_1-1,k}}{\Delta y_{M_1-1}} & \text{if } j = M_1. \end{cases}$$

$$\partial_t^- u = \begin{cases} \frac{u_{i,j,k}}{\Delta t_k} - \frac{u_{i,j,k-1}}{\Delta t_{k-1}}, & \text{if } 1 < k < K - 1, \\ \frac{u_{i,j,1}}{\Delta t_1}, & \text{if } k = 1, \\ \frac{-u_{i,j,K-1}}{\Delta t_{K-1}} & \text{if } k = K. \end{cases}$$

The symmetrized gradient, $\mathcal{E}_{\boldsymbol{\alpha}} : V \to W$ is defined as $\mathcal{E}_{\boldsymbol{\alpha}}(\boldsymbol{v}) = \frac{1}{2}(\nabla_{\boldsymbol{\alpha}} \boldsymbol{v} + \nabla_{\boldsymbol{\alpha}} \boldsymbol{v}^T)$ where

$$\mathcal{E}_{\boldsymbol{\alpha}}(\boldsymbol{v}) = \begin{bmatrix} \alpha_1 \partial_x^- v^1 & \frac{\alpha_1 \partial_y^- v_1 + \alpha_1 \partial_x^- v^2}{2} & \frac{\alpha_2 \partial_t^- v^1 + \alpha_1 \partial_x^- v^3}{2} \\ \frac{\alpha_1 \partial_y^- v^1 + \alpha_1 \partial_x^- v^2}{2} & \alpha_1 \partial_y^- v^2 & \frac{\alpha_2 \partial_t^- v^2 + \alpha_1 \partial_y^- v^3}{2} \\ \frac{\alpha_2 \partial_t^- v^1 + \alpha_1 \partial_x^- v^3}{2} & \frac{\alpha_2 \partial_t^- v^2 + \alpha_1 \partial_y^- v^3}{2} & \alpha_2 \partial_t^- v^3 \end{bmatrix}$$

and the divergence operator is defined as $\mathrm{div}_{2,\boldsymbol{\alpha}} = -\mathcal{E}^*_{\boldsymbol{\alpha}}$. For $\boldsymbol{w} = (w^{i,j})_{1\le i,j\le 3} \in W$ it writes

$$\mathrm{div}_{2,\boldsymbol{\alpha}} : W \to V, \qquad \mathrm{div}_{2,\boldsymbol{\alpha}}(w) = \begin{bmatrix} \alpha_1\partial_x^+ w^{11} + \alpha_1\partial_y^+ w^{12} + \alpha_2\partial_t^+ w^{13} \\ \alpha_1\partial_x^+ w^{21} + \alpha_1\partial_y^+ w^{22} + \alpha_2\partial_t^+ w^{23} \\ \alpha_1\partial_x^+ w^{31} + \alpha_1\partial_y^+ w^{32} + \alpha_2\partial_t^+ w^{33} \end{bmatrix},$$

with $w^{21} = w^{12}$, $w^{31} = w^{13}$ and $w^{32} = w^{23}$. For any $\boldsymbol{v} = (v^1, v^2, v^3) \in V$, we set

$$\|\boldsymbol{v}\|_1 = \sum_{i,j,k} |\boldsymbol{v}|_{i,j,k}, \quad \|\boldsymbol{v}\|_\infty = \max_{i,j,k} |\boldsymbol{v}|_{i,j,k} \text{ with } |\boldsymbol{v}| = \left(\sqrt{(v^1)^2+(v^2)^2+(v^3)^2}\right) \quad (22)$$

and for any $\boldsymbol{w} \in W$, we write

$$\begin{aligned} \|\boldsymbol{w}\|_1 = \sum_{i,j,k} |\boldsymbol{w}|_{i,j,k}, \quad \|\boldsymbol{w}\|_\infty = \max_{i,j,k} |\boldsymbol{w}|_{i,j,k} \quad \text{with} \\ |\boldsymbol{w}| = \left(\sqrt{w_{11}^2 + w_{22}^2 + w_{33}^2 + 2w_{12}^2 + 2w_{13}^2 + 2w_{23}^2}\right). \end{aligned} \quad (23)$$

4.2 Saddle point forms

Our goal is to write the minimization problem

$$\operatorname*{argmin}_{\boldsymbol{u}\in U} \mathcal{H}(\boldsymbol{g}, \mathrm{R}\boldsymbol{u}) + \mathcal{N}(\boldsymbol{u}) \quad (24)$$

as a saddle point problem of the form

$$\min_{x\in X}\max_{y\in Y} \langle \mathcal{K}x, y\rangle - \mathcal{F}^*(y) + \mathcal{G}(x) \quad (25)$$

in order to apply the PDHG algorithm.
We start with the fidelity term defined in (7):

$$\mathcal{H}(\boldsymbol{g}, D\mathrm{R}\boldsymbol{u} + \boldsymbol{\eta}) = \left\langle \boldsymbol{\gamma}(D\mathrm{R}\boldsymbol{u} - \boldsymbol{g}\log(D\mathrm{R}\boldsymbol{u} + \boldsymbol{\eta})), \mathbb{1}_{\hat{U}} \right\rangle + \mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}), \quad (26)$$

using the scalar product on the $\Sigma \times \mathcal{T}$ domain, i.e.,

$$\left\langle \boldsymbol{g}^1, \boldsymbol{g}^2 \right\rangle := \sum_{\hat{\imath},\hat{\jmath},k=1}^{\hat{N}_1,\hat{M}_1,K} \gamma_k (\boldsymbol{g}^1 \boldsymbol{g}^2)_{\hat{\imath},\hat{\jmath},k}.$$

Here, $\mathbb{1}_{\hat{U}}$ is a 3D array of ones. We recall that the temporal non uniform discretization Δt_k is incorporated in the positive parameter $\boldsymbol{\gamma}$. Next, we need the dual form of (26), which is obtained by computing of the convex conjugate of the following function

$$h(x) := \left\langle \gamma(\lambda_1 x - g\log(\lambda_1 x + \lambda_2)), \mathbb{1}_{\hat{U}} \right\rangle + \mathbb{I}_{\{x\ge 0\}}(x).$$

where λ_1, λ_2 are positive weights, using that $f(x) = \sup_{x^*} \langle x, x^*\rangle - f^*(x^*)$. In practice, λ_1, λ_2 represent the decay factor $\boldsymbol{D} = (D_k)_{k=1}^K$ and the noisy component $\boldsymbol{\eta}$ respectively. We conclude that the convex conjugate of h is

$$h^*(x^*) = \sum \boldsymbol{\gamma}\boldsymbol{g}\left(\log\left(\frac{\boldsymbol{\gamma}\lambda_1\boldsymbol{g}}{\boldsymbol{\gamma}\lambda_1 - \boldsymbol{x}^*}\right) - 1\right) - \lambda_2\left(\frac{\boldsymbol{x}^*}{\lambda_1} - \boldsymbol{\gamma}\right) + \mathbb{I}_{\{\boldsymbol{x}^* \ge \boldsymbol{\gamma}\lambda_1 - \boldsymbol{g}\frac{\boldsymbol{\gamma}\lambda_1}{\lambda_2}\}}(x^*), \quad (27)$$

where the multiplications are performed componentwise. For the total variation regularization, $\mathcal{N}(u) = \mathrm{TV}_{\boldsymbol{\alpha}}(\boldsymbol{u})$ using the duality of the ℓ^1 norm and the above computation, we derive the following saddle point form

$$\min_{\boldsymbol{u}} \max_{\boldsymbol{\phi_1},\boldsymbol{\phi_2}} \langle \nabla_{\boldsymbol{\alpha}} \boldsymbol{u}, \boldsymbol{\phi_1}\rangle - \mathbb{I}_{\{\|\cdot\|_\infty \le 1\}}(\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}}) + \langle \mathrm{R}\boldsymbol{u}, \boldsymbol{\phi_2}\rangle - \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_2}) + \mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}), \quad (28)$$

where $\mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_2}) := h^*(\boldsymbol{\phi_2})$. Note that the division $\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}}$ is understood componentwise with respect to the time direction i.e., $\frac{\phi_{i,j,k}}{\gamma_k}$. Overall, we obtain the following setting related to (25), i.e.,

$$\begin{cases} x = (\boldsymbol{u}, 0)^T \in U \times U = X,\ y = (\boldsymbol{\phi_1}, \boldsymbol{\phi_2}) \in V \times \hat{U} = Y, \\ \mathcal{F}^*(y) = \mathcal{F}^*(\boldsymbol{\phi_1}, \boldsymbol{\phi_2}) = \mathbb{I}_{\{\|\cdot\|_\infty \le 1\}}(\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}}) + \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_2}) \\ \mathcal{G}(x) = \mathcal{G}(\boldsymbol{u}, 0) = \mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}), \\ \mathcal{K} = \begin{bmatrix} \nabla_{\boldsymbol{\alpha}} & 0 \\ \mathrm{R} & 0 \end{bmatrix} \quad \mathcal{K}^T = \begin{bmatrix} -\mathrm{div}_{1,\boldsymbol{\alpha}} & \mathrm{R}^* \\ 0 & 0 \end{bmatrix}. \end{cases} \quad (29)$$

The approach is similar for the other two regularizers. We summarize below the saddle point forms for $\mathrm{ICTV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u})$, $\mathrm{TGV}_{\boldsymbol{\alpha}}(u)$ and $\mathrm{ICTGV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u})$ regularizers. Starting with $\mathcal{N}(u) = \mathrm{ICTV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u})$ the minimization problem (24) can be written in the (25) form as

$$\begin{aligned} \min_{\boldsymbol{u},\boldsymbol{v}} \max_{\boldsymbol{\phi_1},\boldsymbol{\phi_2},\boldsymbol{\phi_3}} & \langle \nabla_{\boldsymbol{\kappa}}(\boldsymbol{u} - \boldsymbol{v}), \boldsymbol{\phi_1}\rangle - \mathbb{I}_{\{\|\cdot\|_\infty \le \beta_1\}}(\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}}) + \langle \nabla_{\boldsymbol{1-\kappa}}, \boldsymbol{\phi_2}\rangle \\ & - \mathbb{I}_{\{\|\cdot\|_\infty \le \beta_0\}}(\frac{\boldsymbol{\phi_2}}{\boldsymbol{\gamma}}) + \langle \mathrm{R}u, \boldsymbol{\phi_3}\rangle - \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_3}) + \mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}), \end{aligned} \quad (30)$$

$$\begin{cases} x = (\boldsymbol{u}, \boldsymbol{v})^T \in U \times U = X,\ y = (\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \boldsymbol{\phi_3})^T \in V \times V \times \hat{U} = Y, \\ \mathcal{F}^*(y) = \mathcal{F}^*(\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \boldsymbol{\phi_3}) = \mathbb{I}_{\{\|\cdot\|_\infty \le \beta_1\}}(\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}}) + \mathbb{I}_{\{\|\cdot\|_\infty \le \beta_0\}}(\frac{\boldsymbol{\phi_2}}{\boldsymbol{\gamma}}) + \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_3}) \\ \mathcal{G}(x) = \mathcal{G}(\boldsymbol{u}, \boldsymbol{v}) = \mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}), \\ \mathcal{K} = \begin{bmatrix} \nabla_{\boldsymbol{\kappa}} & -\nabla_{\boldsymbol{\kappa}} \\ 0 & \nabla_{\boldsymbol{1-\kappa}} \\ \mathrm{R} & 0 \end{bmatrix}, \mathcal{K}^T = \begin{bmatrix} -\mathrm{div}_{1,\boldsymbol{\kappa}} & 0 & \mathrm{R}^* \\ \mathrm{div}_{1,\boldsymbol{\kappa}} & \mathrm{div}_{1,\boldsymbol{1-\kappa}} & 0 \end{bmatrix}. \end{cases} \quad (31)$$

For $\mathcal{N}(u) = \mathrm{TGV}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{u})$ we write that

$$\begin{aligned} \min_{\boldsymbol{u},\boldsymbol{w}} \max_{\boldsymbol{\phi_1},\boldsymbol{\phi_2},\boldsymbol{\phi_3}} & \langle \nabla_{\boldsymbol{\alpha}} \boldsymbol{u} - \boldsymbol{w}, \boldsymbol{\phi_1}\rangle - \mathbb{I}_{\{\|\cdot\|_\infty \le \beta_1\}}(\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}}) + \langle \mathcal{E}_{\boldsymbol{\alpha}} \boldsymbol{w}, \boldsymbol{\phi_2}\rangle \\ & - \mathbb{I}_{\{\|\cdot\|_\infty \le \beta_0\}}(\frac{\boldsymbol{\phi_2}}{\boldsymbol{\gamma}}) + \langle \mathrm{R}\boldsymbol{u}, \boldsymbol{\phi_3}\rangle - \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_3}) + \mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}), \end{aligned} \quad (32)$$

$$\begin{cases} x = (\boldsymbol{u}, \boldsymbol{w})^T \in U \times V = X,\ y = (\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \boldsymbol{\phi_3})^T \in V \times W \times \hat{U} = Y, \\ \mathcal{F}^*(y) = \mathcal{F}^*(\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \boldsymbol{\phi_3}) = \mathbb{I}_{\{\|\cdot\|_\infty \le \beta_1\}}(\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}}) + \mathbb{I}_{\{\|\cdot\|_\infty \le \beta_0\}}(\frac{\boldsymbol{\phi_2}}{\boldsymbol{\gamma}}) + \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_3}) \\ \mathcal{G}(x) = \mathcal{G}(\boldsymbol{u}, \boldsymbol{w}) = \mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}), \\ \mathcal{K} = \begin{bmatrix} \nabla_{\boldsymbol{\alpha}} & -I \\ 0 & \mathcal{E}_{\boldsymbol{\alpha}} \\ \mathrm{R} & 0 \end{bmatrix}, \mathcal{K}^T = \begin{bmatrix} -\mathrm{div}_{2,\boldsymbol{\alpha}} & 0 & \mathrm{R}^* \\ -I & -\mathrm{div}_{2,\boldsymbol{\alpha}} & 0 \end{bmatrix}. \end{cases} \quad (33)$$

and for ICTGV$_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u})$ we conclude with

$$\begin{aligned}\min_{\substack{\boldsymbol{u},\boldsymbol{v}\\ \boldsymbol{w_1},\boldsymbol{w_2}}}\max_{\substack{\boldsymbol{\phi_1},\boldsymbol{\phi_2},\boldsymbol{\phi_3}\\ \boldsymbol{\phi_4},\boldsymbol{\phi_5}}} &\langle \nabla_{\boldsymbol{\kappa}}(\boldsymbol{u}-\boldsymbol{v})-\boldsymbol{w_1},\boldsymbol{\phi_1}\rangle - \mathbb{I}_{\{\|\cdot\|_\infty\le\beta_1\}}(\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}})\\ &+\langle \mathcal{E}_{\boldsymbol{\kappa}}\boldsymbol{w_1},\boldsymbol{\phi_2}\rangle - \mathbb{I}_{\{\|\cdot\|_\infty\le\beta_1\sqrt{2}\}}(\frac{\boldsymbol{\phi_2}}{\boldsymbol{\gamma}})\\ &+\left\langle \nabla_{\boldsymbol{1-\kappa}}(\boldsymbol{v})-\boldsymbol{w_2},\boldsymbol{\phi_3}\right\rangle - \mathbb{I}_{\{\|\cdot\|_\infty\le\beta_0\}}(\frac{\boldsymbol{\phi_3}}{\boldsymbol{\gamma}})\\ &+\left\langle \mathcal{E}_{\boldsymbol{1-\kappa}}\boldsymbol{w_2},\boldsymbol{\phi_4}\right\rangle - \mathbb{I}_{\{\|\cdot\|_\infty\le\beta_0\sqrt{2}\}}(\frac{\boldsymbol{\phi_4}}{\boldsymbol{\gamma}})\\ &+\langle \mathrm{R}\boldsymbol{u},\boldsymbol{\phi_5}\rangle - \mathcal{H}^*(\boldsymbol{g},\boldsymbol{\phi_5}) + \mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}),\end{aligned} \tag{34}$$

$$\begin{cases} x=(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w_1},\boldsymbol{w_2})^T\in U\times U\times V\times W=X,\\ y=(\boldsymbol{\phi_1},\boldsymbol{\phi_2},\boldsymbol{\phi_3},\boldsymbol{\phi_4},\boldsymbol{\phi_5})^T\in (V\times W)^2\times\hat{U}=Y,\\ \mathcal{F}^*(y)=\mathcal{F}^*(\boldsymbol{\phi_1},\boldsymbol{\phi_2},\boldsymbol{\phi_3},\boldsymbol{\phi_4},\boldsymbol{\phi_5})=\mathbb{I}_{\{\|\cdot\|_\infty\le\beta_1\}}(\frac{\boldsymbol{\phi_1}}{\boldsymbol{\gamma}})+\mathbb{I}_{\{\|\cdot\|_\infty\le\beta_1\sqrt{2}\}}(\frac{\boldsymbol{\phi_2}}{\boldsymbol{\gamma}})\\ \qquad\qquad +\mathbb{I}_{\{\|\cdot\|_\infty\le\beta_0\}}(\frac{\boldsymbol{\phi_3}}{\boldsymbol{\gamma}})+\mathbb{I}_{\{\|\cdot\|_\infty\le\beta_0\sqrt{2}\}}(\frac{\boldsymbol{\phi_4}}{\boldsymbol{\gamma}})+\mathcal{H}^*(\boldsymbol{g},\boldsymbol{\phi_5})\\ \mathcal{G}(x)=\mathcal{G}(\boldsymbol{u},\boldsymbol{w})=\mathbb{I}_{\{\boldsymbol{u}\ge 0\}}(\boldsymbol{u}) \text{ and } \mathcal{K}=\begin{bmatrix}\nabla_{\boldsymbol{\kappa}} & -I & -\nabla_{\boldsymbol{\kappa}} & 0\\ 0 & \mathcal{E}_{\boldsymbol{\kappa}} & 0 & 0\\ 0 & 0 & \nabla_{\boldsymbol{1-\kappa}} & -I\\ 0 & 0 & 0 & \mathcal{E}_{\boldsymbol{1-\kappa}}\\ \mathrm{R} & 0 & 0 & 0\end{bmatrix}. \end{cases} \tag{35}$$

4.3 Numerical Solution

Based on the algorithm presented in Section 3, we can precompute the diagonal preconditioners T and Σ for each case of (29), (31), (33) and (35) and solve the corresponding proximal maps (19) and (20). In particular, the proximal map $x=(I+T\partial\mathcal{G})^{-1}(\hat{x})$ is a projection on the positive cone $\{x\ge 0\}$ and common for all the regularizers. Indeed, for the (29) case, we write using Algorithm 2

$$\begin{aligned}&\boldsymbol{u}=(I+T\partial\mathcal{G})^{-1}(\hat{\boldsymbol{u}})\Leftrightarrow \boldsymbol{u}=\max\{\hat{\boldsymbol{u}},0\},\quad \text{where}\\ &\hat{\boldsymbol{u}}=\boldsymbol{u}-T(-\mathrm{div}_{1,\boldsymbol{\alpha}}\boldsymbol{\phi_1}+\mathrm{R}^*\boldsymbol{\phi_2}).\end{aligned}$$

In terms of $y=(I+\Sigma\partial\mathcal{F}^*)^{-1}(\hat{y})$ we need to compute it separately based on the definition of $\mathcal{F}^*$. For instance in the (29) case, the functional $\mathcal{F}^*(y)$ can be decoupled and obtain two closed formed solutions namely

$$\begin{bmatrix}\boldsymbol{\phi_1}\\ \boldsymbol{\phi_2}\end{bmatrix}=(I+\Sigma\partial\mathcal{F}^*)^{-1}\begin{bmatrix}\hat{\boldsymbol{\phi_1}}\\ \hat{\boldsymbol{\phi_2}}\end{bmatrix}\Leftrightarrow\begin{cases}\boldsymbol{\phi_1}=\underset{\|\boldsymbol{\phi_1}\|_\infty\le 1}{\mathrm{argmin}}\ \frac{1}{2}\left\|\boldsymbol{\phi_1}-\hat{\boldsymbol{\phi_1}}\right\|_V^2\\ \boldsymbol{\phi_2}=\underset{\boldsymbol{\phi_2}}{\mathrm{argmin}}\ \frac{1}{2}\left\|\boldsymbol{\phi_2}-\hat{\boldsymbol{\phi_2}}\right\|_{\hat{U}}^2+\mathcal{H}^*(\boldsymbol{g},\boldsymbol{\phi_2})\end{cases},$$

where using Algorithm 2 we write

$$\begin{bmatrix}\hat{\boldsymbol{\phi_1}}\\ \hat{\boldsymbol{\phi_2}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\phi_1}\\ \boldsymbol{\phi_2}\end{bmatrix}+\begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\begin{bmatrix}\nabla_{\boldsymbol{\alpha}} & 0\\ \mathrm{R} & 0\end{bmatrix}\begin{bmatrix}\overline{\boldsymbol{u}}\\ 0\end{bmatrix}.$$

The first minimization problem corresponds to a projection on the ℓ^∞ ball, i.e.,

$$\boldsymbol{\phi_1} = \underset{\|\boldsymbol{\phi_1}\|_\infty \leq 1}{\operatorname{argmin}} \frac{1}{2} \left\|\boldsymbol{\phi_1} - \hat{\boldsymbol{\phi_1}}\right\|_V^2 \Leftrightarrow \boldsymbol{\phi_1} = \frac{\hat{\boldsymbol{\phi_1}}}{\max\left\{1, \frac{|\hat{\boldsymbol{\phi_1}}|}{\boldsymbol{\gamma}}\right\}}. \tag{36}$$

For the second minimization problem, we use [23, Lemma 3] and conclude to

$$\begin{aligned} &\boldsymbol{\phi_2} = \underset{\boldsymbol{\phi_2}}{\operatorname{argmin}} \frac{1}{2} \left\|\boldsymbol{\phi_2} - \hat{\boldsymbol{\phi_2}}\right\|_{\hat{U}}^2 + \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_2}) \Leftrightarrow \\ &\boldsymbol{\phi_2} = 0.5\left(\hat{\boldsymbol{\phi_2}} + \frac{\lambda_2}{\lambda_1}\Sigma_2 + \lambda_1\boldsymbol{\gamma} - \sqrt{(\hat{\boldsymbol{\phi_2}} + \frac{\lambda_2}{\lambda_1}\Sigma_2 - \lambda_1\boldsymbol{\gamma})^2 + 4\Sigma_2 g\boldsymbol{\gamma}}\right), \end{aligned} \tag{37}$$

where $\lambda_2 = \boldsymbol{\eta}$, $\lambda_1 = (D_k)_{k=1}^K$, $\boldsymbol{\gamma} = (\gamma_k)_{k=1}^K$ and the multiplication/division is understood componentwise. For ICTV$_{\boldsymbol{\beta},\boldsymbol{\kappa}}$, TGV$_{\boldsymbol{\alpha}}$ and ICTGV$_{\boldsymbol{\beta},\boldsymbol{\kappa}}$ regularizers the solution with respect to $\boldsymbol{\phi_3}$ and $\boldsymbol{\phi_5}$ variables respectively is the same as (37).

For the TGV$_{\boldsymbol{\alpha}}$ case the solution with respect to $\boldsymbol{\phi_1}$, $\boldsymbol{\phi_2}$ the solution is similar to (36) taking into account the additional weights β_1, β_0 i.e.,

$$\boldsymbol{\phi_1} = \frac{\hat{\boldsymbol{\phi_1}}}{\max\left\{1, \frac{|\hat{\boldsymbol{\phi_1}}|}{\boldsymbol{\gamma}\beta_1}\right\}} \text{ and } \frac{\hat{\boldsymbol{\phi_2}}}{\max\left\{1, \frac{|\hat{\boldsymbol{\phi_2}}|}{\boldsymbol{\gamma}\beta_0}\right\}}$$

Here, we have that $\phi_2 \in W$. Similarly, we compute the solutions for the ICTGV$_{\boldsymbol{\beta},\boldsymbol{\kappa}}$ regularizer.

4.4 Stopping Criteria

In order to compute an accurate saddle point solution and validate the convergence of our algorithm, we use a primal-dual gap of (25) defined as

$$PD_{gap} = \mathcal{F}(\mathcal{K}x) + \mathcal{G}(x) + \mathcal{F}^*(-y) + \mathcal{G}^*(-\mathcal{K}^T y). \tag{38}$$

As the primal-dual gap approaches to zero, the solution gets asymptotically close to the desired saddle point. To be more precise, we do not use the exact form of (38) since we do not take into account some of the indicator functions that are present, for instance $\mathbb{I}_{\{\|\cdot\|_\infty \leq 1\}}(\boldsymbol{\phi_1})$ or the positivity constraints related to (26), (27). In fact these constraints are satisfied at each iteration and can be neglected, see for instance [32]. However, we need to consider the convex conjugate of $\mathcal{G}$, i.e. the convex conjugate of the positive cone $\{x \geq 0\}$. Hence, we have that

$$\mathcal{G}(x) = \mathbb{I}_{\{x \geq 0\}}(x) \Rightarrow \mathcal{G}^*(x^*) = \max_{x \geq 0} \langle x, x^* \rangle = \begin{cases} 0, & \text{if } \langle x^*, x \rangle \leq 0 \\ \infty, & \text{otherwise,} \end{cases}$$

see [28]. For all the regularizers the modified primal-dual gaps are explicitly written below

$$
\begin{aligned}
PD_{TV-gap} &= \mathrm{TV}_{\boldsymbol{\alpha}} + \mathcal{H}(\boldsymbol{g}, \mathrm{R}\boldsymbol{u}) + \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_2}) + \sum \max(\mathrm{div}_{1,\boldsymbol{\alpha}}\, \boldsymbol{\phi_1} - \mathrm{R}^T \boldsymbol{\phi_2}, 0) \\
PD_{TGV-gap} &= \mathrm{TGV}_{\boldsymbol{\alpha}}(\boldsymbol{u}) + \mathcal{H}(\boldsymbol{g}, \mathrm{R}\boldsymbol{u}) + \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_3}) + \sum \max(\mathrm{div}_{1,\boldsymbol{\alpha}}\, \boldsymbol{\phi_1} - \mathrm{R}^T \boldsymbol{\phi_3}, 0) \\
&\quad + \sum \max(\boldsymbol{\phi_1} + \mathrm{div}_{2,\boldsymbol{\alpha}}\, \boldsymbol{\phi_2}, 0) \\
PD_{ICTV-gap} &= \mathrm{ICTV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u}) + \mathcal{H}(\boldsymbol{g}, \mathrm{R}\boldsymbol{u}) + \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_3}) + \sum \max(\mathrm{div}_{1,\boldsymbol{\kappa}}\, \boldsymbol{\phi_1} - \mathrm{R}^T \boldsymbol{\phi_3}), 0) \\
&\quad + \sum \max(-\mathrm{div}_{1,\boldsymbol{\kappa}}\, \boldsymbol{\phi_1} + \mathrm{div}_{1,\boldsymbol{1-\kappa}}\, \boldsymbol{\phi_2}, 0) \\
PD_{ICTGV-gap} &= \mathrm{ICTGV}_{\boldsymbol{\beta},\boldsymbol{\kappa}}(\boldsymbol{u}) + \mathcal{H}(\boldsymbol{g}, \mathrm{R}\boldsymbol{u}) + \mathcal{H}^*(\boldsymbol{g}, \boldsymbol{\phi_5}) + \sum \max \mathrm{div}_{1,\boldsymbol{\kappa}}\, \boldsymbol{\phi_1} - \mathrm{R}^T \boldsymbol{\phi_5}), 0) \\
&\quad + \sum \max(\boldsymbol{\phi_1} + \mathrm{div}_{2,\boldsymbol{\kappa}}\, \boldsymbol{\phi_2}, 0) + \sum \max(-\mathrm{div}_{1,\boldsymbol{\kappa}}\, \boldsymbol{\phi_1} + \mathrm{div}_{1,\boldsymbol{1-\kappa}}\, \boldsymbol{\phi_3}, 0) \\
&\quad + \sum \max(\boldsymbol{\phi_3} + \mathrm{div}_{2,\boldsymbol{1-\kappa}}\, \boldsymbol{\phi_4}, 0)
\end{aligned}
$$

In order to be independent of the size of the input data, we rescale the primal-dual gap by the dimension of the problem. We use the primal-dual gap in order to verify the convergence of the algorithm. However, evaluating the primal-dual gap increases the computational cost since it requires two additional computations of the system matrix per iteration. Hence, we rather use a fixed number of iterations, i.e., 2000 iterations as a stopping criterion, or alternatively a certain threshold ε such that the relative error between two consecutive iterates satisfies

$$
\frac{\left\| u^{k+1} - u^k \right\|_2}{\left\| u^{k+1} \right\|_2} < \varepsilon.
$$

In practice we noticed that after 1000 iterations the primal-dual gap is in the range of $[10^{-5}, 10^{-2}]$ and there is no significant difference among the iterations.

4.5 Quantitative criteria

The proposed regularizers involve parameters that need to be optimized to produce the best reconstructions compared to the ground truth. In particular for the TV and TGV regularizers one has to optimize α_1 and α_2 only, see (9), (10). In terms of their infimal convolution extensions, three parameters need to be tuned, i.e., β_1, β_0 and κ, see (11), (12). For this purpose, we follow a trial-error approach on large grid sizes where the parameters are optimized based on the highest average of structural similarity index (SSIM) over all frames, see [36]. In addition, we report the mean squared error (MSE) over all the brain regions, as well as on some *regions of interest* (ROI) such as the striatum, cortex, thalamus and white matter regions, see Figure 1. Let us mention that both MSE and SSIM criteria have been tested in order to acquire the best reconstructions. However, we find that the best reconstructions with respect to the lowest MSE allow more noisy structures compared to SSIM. Here, we would like to mention that although 3 parameters on the infimal convolutions case are difficult to optimize, we can benefit from the symmetry of ICTV and ICTGV and reduce the total number of combination of

$(\beta_1, \beta_0, \kappa)$ For instance when $\beta_1 = \beta_0$ and we replace κ by $1 - \kappa$. Similarly the solution are the same for $(\beta_1, \beta_0, \kappa)$ and $(\beta_0, \beta_1, 1 - \kappa)$.

Finally, we evaluate the relative bias defined as

$$\text{bias} = \frac{1}{\#\text{pixels}} \sum_{i,j,k}^{\#\text{pixels}} \left(\frac{|\hat{u} - u|}{|\hat{u}|} \right)_{i,j,k} \tag{39}$$

where the number of pixels depends on the specific ROI, $\hat{u}$ denote the ground truth and u the reconstructed image sequence. Apart from the aforementioned priors, we present the standard Maximum Likehood Expectation Maximization (MLEM) reconstruction and a post processed MLEM reconstruction with a Gaussian filtering. A full width at half maximum (FWHM) of 12mm is used for all the radiotracers. As a stopping criterion of the MLEM and the post filtered MLEM reconstructions, we use a fixed number of iterations (200) and recover the one with minimum MSE. Although this is out of the scope of this paper, we would like to mention that besides the trial-error method for optimizing these parameters, one can employ a bilevel optimization strategy such as in [1].

5 Numerical Experiments

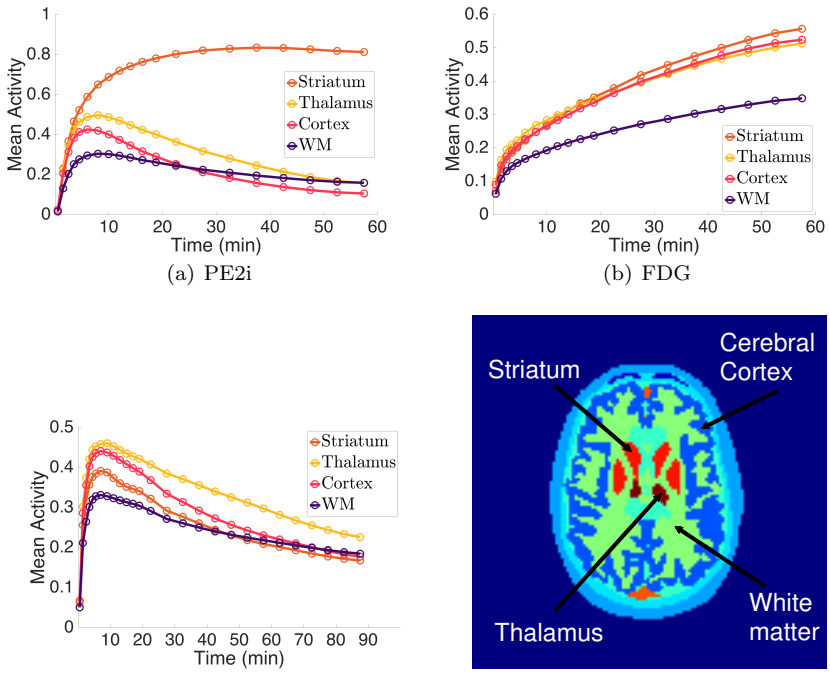
In this section, the proposed regularizers are examined on simulated two dimensional dynamic brain phantoms.

To simulate the acquisitions, we measured Time Activity Curves (TACs) in three different acquisitions with the Siemens HRRT camera of healthy volunteers injected with 11C-PE2I, 18F-FDG and 18F-DPA714. The TACs were measured on ROIs delineated on a T1-weighted MRI, from PET images that were reconstructed with PSF modeling. The HRRT camera was used for its relatively high resolution in order to extract input TACs with minimal partial volume effect, see [33]. In Figure 1, we present the mean activity on different regions of interest (ROI) with respect of the acquisition time, used as an input to the simulation. Outside of the brain was set to 0 and Artery and Vein were set from blood samples measurements. Scatter and random fractions were also extracted from these three scans in order to produce realistic simulations.

All numerical examples are implemented using MATLAB under GPU on a Macbook pro with 16GB RAM and 2.2 GHz Intel Core i7 CPU.

5.1 Noisy Data Generation

In order to create two dimensional dynamic data, we fix a specific slice from the 3D+time volume as in Figure 2 and create a 2D+time array where the third coordinate corresponds to the temporal dimension. The original 3D+time brain phantom has been rescaled to a [0,1] pixel intensity interval. We first create the noiseless dynamic sinograms via the forward model. This is performed using the ground truth brain phantom and the system matrix R which describes the geometric properties of the PET scanner. Then, the noise-free sinograms are corrupted with noise based on the prompt counts that follow Poisson distribution. In addition, we consider a uniform background noise that simulates the scattered and



(a) PE2i (b) FDG

(c) DPA (d) Regions of interest (ROI

Fig. 1 Mean Time Activity Curves for the radiotracers PE2i, FDG and DPA for the ROI: Striatum, Cortex, White matter, Thalamus

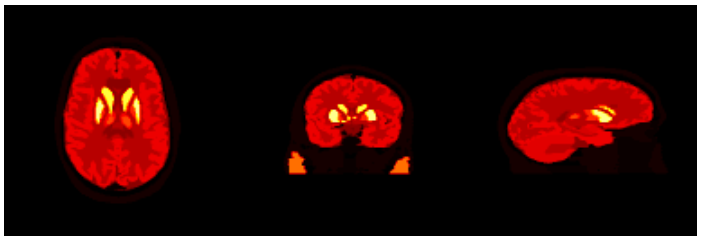


Fig. 2 Axial, Coronal, Sagittal views of PE2i radiotracer, t = 11min.

random events. Finally, based on the radiotracer, we use a decay correction on the activity of each radiotracer, namely

$$D_t = \frac{1 - e^{\lambda\, \Delta t}}{\lambda\, \Delta t\, e^{\lambda t_1}} \tag{40}$$

where $\lambda = \frac{ln2}{\tau_{1/2}}$, $\Delta t = t_2 - t_1$ with t_1, t_2 are the starting and ending frame times. For this numerical simulation no attenuation correction is considered. In Table 1, we present all technical details for each radiotracer.

	Isotope	Prompt Counts	$\tau_{1/2}$	Duration (Frames)
PE2i	^{11}C	31 000 000	1223 sec	$5 \times 60, 5 \times 120, 2 \times 150, 8 \times 300$ sec (20)
FDG	^{18}F	47 767 000	6586.2 sec	$5 \times 60, 5 \times 120, 2 \times 150, 8 \times 300$ sec (20)
DPA	^{18}F	22 349 700	6586.2 sec	$6 \times 60, 7 \times 120, 14 \times 300$ sec (27)

Table 1 Isotope, Counts, Half-life and time-frame configurations for the PE2i, FDG and DPA tracers.

For the simulation of the system matrix R, we use the open-source Matlab toolbox *Image Reconstruction*, see [17], designed in a C-compiled mex framework. The Poisson noise is simulated using *poisson* m-file from this toolbox. The resolution of the brain phantom is 128×128 of 2.2mm pixels. The sinogram is discretized into 150 radial bins and 150 angles uniformly spaced over 180^o with 2mm ray spacing.

5.2 PE2i/FDG/DPA radiotracers

Here, we briefly describe the functionality of each radiotracers used in our numerical experiments. [11C]PE2i is a radiopharmaceutical for in vivo exploration of the dopamine transporter (DAT). The DAT is involved in many dopamine-related disorders including attention deficit hyperactivity disorder, Parkinson's disease, clinical depression or alcoholism. In vivo investigations in healthy humans have demonstrated that PE2i has high striatal uptake. In our experiments, we considered this radiotracer as it has a very well identified ROI with specific uptake in the striatum, and almost only non-specific fixation everywhere else in the brain. [18F]FDG was also considered as it is the most widely used radiopharmaceutical for clinical PET imaging. It is based on a glucose analog and can be used to assess glucose metabolism. Many neurodegenerative diseases produce significant brain functions alterations that are detectable in PET with [18F]FDG. The uptake of [18F]FDG in the brain is higher in the grey matter compared to the white matter. Finally, we considered a third radiotracer, the [18F]DPA-714, which is specific to the translocator protein (TSPO), a biochemical marker of neuroinflammation. Full understanding of the in vivo TSPO function in central and peripheral inflammatory processes and other pathologies is yet a challenge. In this case, the thalamus is the region that generally presents the highest specific binding in healthy controls.

In the following section, we present our results with respect to the PE2i, FDG and DPA radiotracers.

5.3 Results: PE2i

For the sake of readability, results obtained on the PE2i radiotracer are separated into two figures. The first one (Fig. 3) contains the results obtained with methods that do not penalize the temporal domain, i.e. spatial TV, spatial TGV and classic MLEM reconstructions. While the second one (Fig. 4) contains the results obtained with the proposed spatiotemporal regularizers, i.e. spatiotemporal TV and TGV, ICTV and ICTGV. In Figure 3, it can be observed a varying image quality across the frames for all the spatial-only reconstruction methods. This is predictable as a global regularization parameter is used, the noise is signal dependent and the number of counts variates over the time in dynamic acquisitions. For instance, in

the case of TV reconstruction, the solution clearly suffers from staircasing and loss of contrast artifacts. Most of the thin structures in the cortex and thalamus regions have disappeared due to loss of contrast. In the case of TGV reconstructions, the method recovers sharper structures especially around the cortex and the striatum regions. However, we observe that many of the contours are blurred. This is due to the fact that the optimized regularizing parameter with respect to SSIM forces the TGV prior to act mostly as a TV^2 prior, see [21]. Finally, we would like to highlight that although there are still some strong noise artifacts on the filtered version of the MLEM reconstruction, it behaves better on the striatum and thalamus regions in the late frames ($25min$). These visual results are confirmed by the quantitative criteria that can be found in the last 4 lines of Tables 2 and 3. An analysis of the striatum TAC confirms that the four methods based on spatial regularizers have a close behaviour early in the acquisition, and that all of them have a tendency to underestimate the true activity in these regions, see Figure 5.

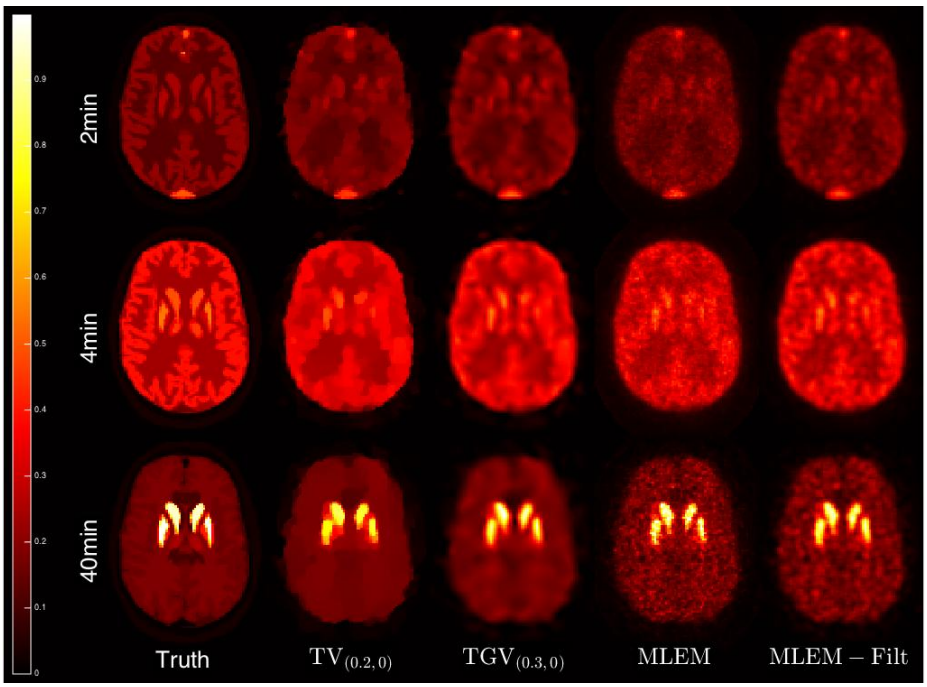


Fig. 3 Dynamic PET simulations with PE2i radiotracer: only the spatial regularization and the corresponding MLEM reconstructions.

Regarding the reconstructions for the proposed spatiotemporal regularizers (Fig. 4), we deal with a simultaneous spatial-temporal regularization in order to take into account the temporal variation. A visual comparison indicates that globally, all spatiotemporal methods produced better results than their spatial counterparts. It also indicates that compared with the spatiotemporal TV and TGV reconstructions, their corresponding infimal convolution versions behave better and produce sharper reconstructions. In particular, we can observe that compared

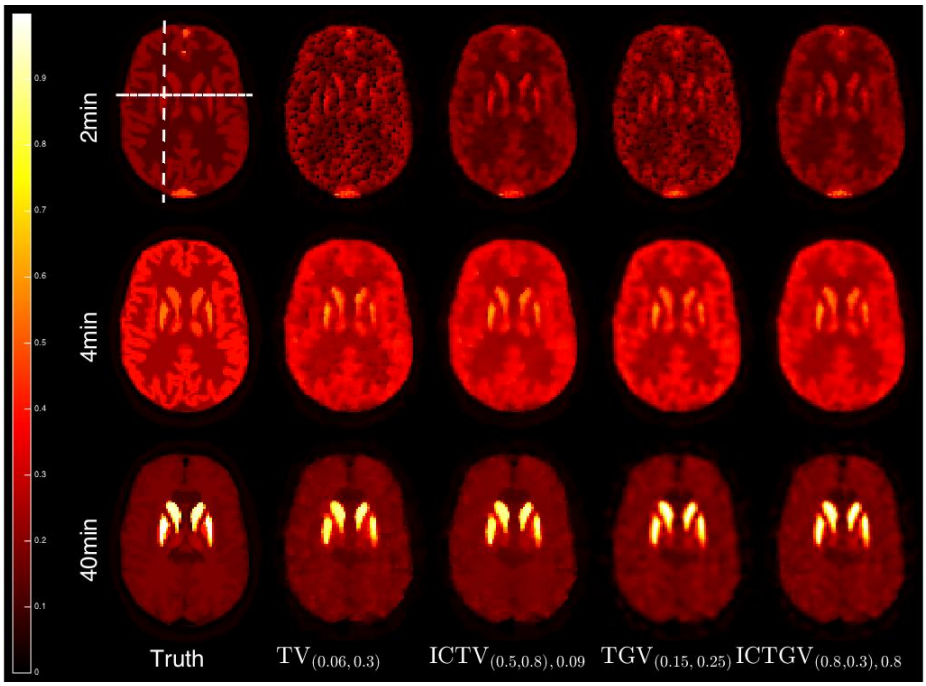


Fig. 4 Dynamic PET simulations with PE2i radiotracer: reconstruction methods with spatiotemporal regularization.

to the ground truth, the results obtained with ICTV and ICTGV are more robust and accurate to the activity across the acquisition time, see the TAC with respect to the striatum in Figure 5.

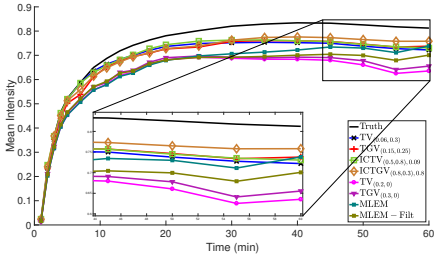
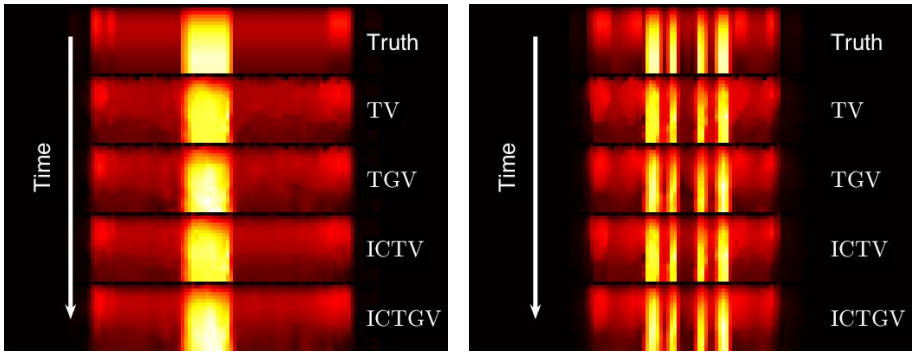


Fig. 5 Results obtained on dynamic PET simulations with PE2i radiotracer: Time Activity Curves for the region of striatum.

This is also verified using the time-lines presented in Figure 6. This plots concerns the spatiotemporal reconstructions for two reference lines presented in the upper left corner in Figure 4. We notice that a spatiotemporal staircasing and blurring appears for the TV and TGV reconstructions respectively. Furthermore, using the infimal convolution regularizers we manage to reduce these artifacts.



(a) Reference line: - - (b) Reference line: - .

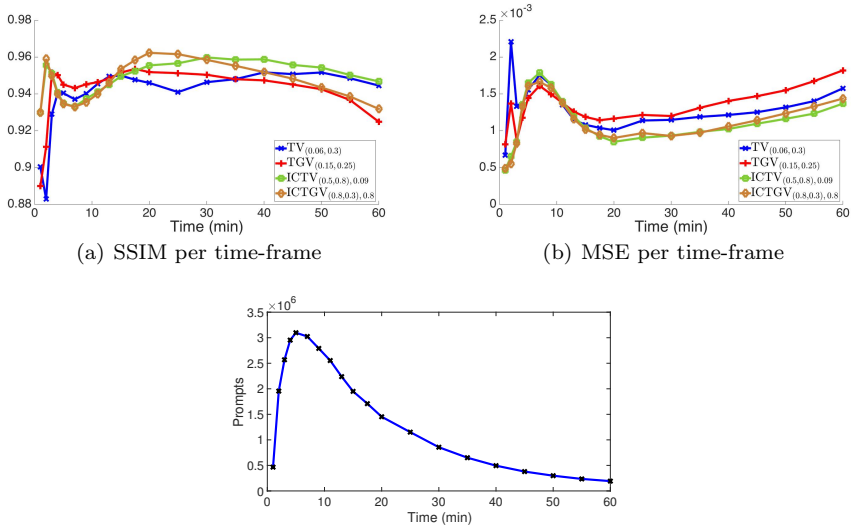
Fig. 6 PE2i: Time-lines of the spatiotemporal reconstructions. The reference lines are shown in Figure 4.

For the case of PE2i, ICTV is the one that produces reconstructions with less noise and better contrast. This is justified from the comparison of the quality measures in the Tables 2, 3. In terms of MSE, SSIM and bias, we obtain the best scores and especially compared to ICTGV. We refer also to the MSE and SSIM values per frame in Figures 7(a), 7(b). In addition, we have the lowest scores in the striatum region for MSE and bias. Comparing the timelines of ICTV and ICTGV in Figure 6 for both the striatum and white matter regions, we observe that for the ICTV reconstruction the white matter region is closer to the ground truth whereas we obtain a better contrast in the striatum region with ICTGV. Similar behaviour can be observed for the TAC plot in Figure 5.

We continue with the interpretation of the infimal convolution components. In the first time-frames of the TV and TGV reconstructions, we witness a significant amount of noise that is not eliminated compared to the other two cases. This is due to the fact that at the beginning of the acquisition the recorded photon counts increase abruptly over a short period of time, specially in the first 5 minutes, as it can be observed in the Figure 7(c). Therefore, the temporal variation for these time-frames is significantly larger. In this situation, one can benefit using the infimal convolution approaches. In Figure 8, we present the corresponding decomposition for the $\mathrm{ICTV}_{(0.5,0.8),0.09}$ case for the first five consecutive frames. In addition, we present the mean activity curves for the corresponding ICTV reconstruction for two regions of interest, e.g., striatum and cortex in Figure 9. This indicates a decomposition into the $\boldsymbol{u} - \boldsymbol{v}$ and $\boldsymbol{v}$ components.

For visualization purposes, we consider a rescaled version of the $\boldsymbol{v}$ component. One can observe that the $\boldsymbol{u} - \boldsymbol{v}$ component depicts nearly the activity of the specified region and stores morphological structures. On the other hand for the $\boldsymbol{v}$ component, we have a weak temporal weight which enforces noisy variations along the temporal direction to be captured. Equivalently, the $\boldsymbol{v}$ component depicts more rapid intensity changes that are strong in the beginning (Fig. 8) of the acquisition and decreases at the end, Figure 9.

Regarding the quantification of the radiotracer, which measures the impact of the reconstruction method on the bias of image values compared to ground truth,



(a) SSIM per time-frame

(b) MSE per time-frame

(c) Counts per time-frame

Fig. 7 PE2i: SSIM, MSE values and number of counts with respect to time.

we notice that the mean time activity curves especially in the striatum region are better for the infimal convolution cases (Fig. 5). This is particularly important for this radiotracer as striatum is the region that presents the highest specific uptake and it is the region used in clinical practice to evaluate pathologies associated to the dopamine transporter.

	SSIM	MSE (10^{-3})	BIAS
$TV_{(0.06,0.3)}$	0.9396	1.3181	0.3422
$TGV_{(0.15,0.25)}$	0.9417	1.3281	0.3580
$ICTV_{(0.5,0.8),0.09}$	**0.9484**	**1.1251**	**0.3105**
$ICTGV_{(0.8,0.3),0.8}$	0.9466	1.1266	0.3201
$TV_{(0.2,0)}$	0.9155	2.4955	0.3926
$TGV_{(0.3,0)}$	0.9099	2.5429	0.4485
MLEM	0.7899	3.4633	0.5402
MLEM-Filt	0.8586	2.6265	0.4767

Table 2 PE2i: Comparison of SSIM, MSE and bias quality criteria for the whole brain.

5.4 Results: FDG

Results obtained with the FDG were comparable with the ones obtained in the PE2i case. While the same time-frame protocol as in the PE2i case was used, see Table 1, the number of counts increases with a different rate since the FDG is an almost irreversible radiotracer. In addition, it does not decrease as its half-life is much higher. The change of the prompt counts across time between the

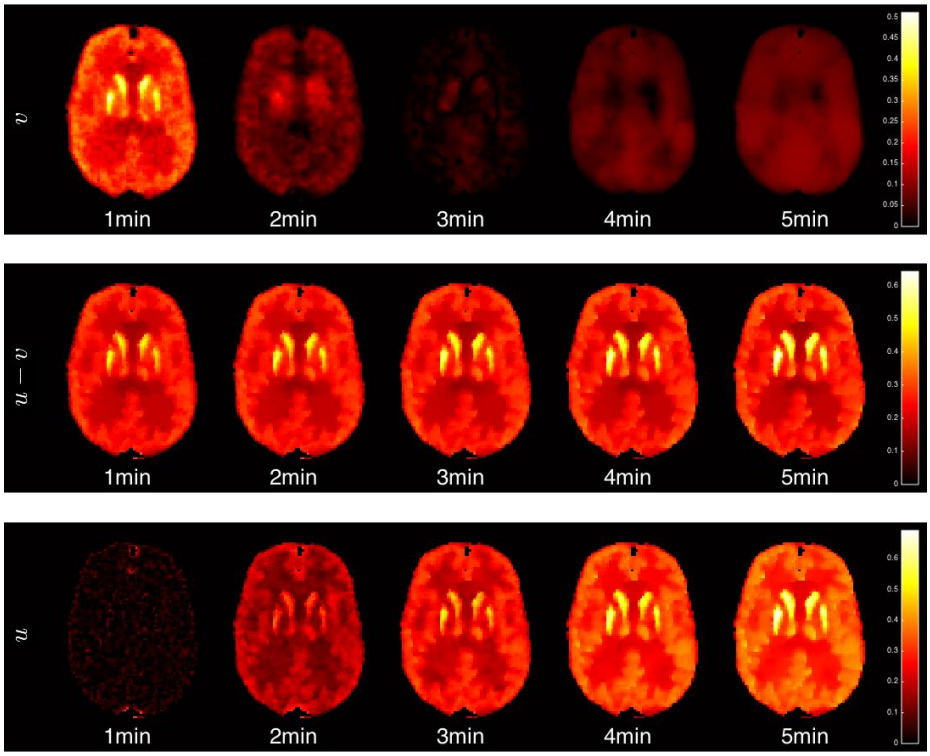
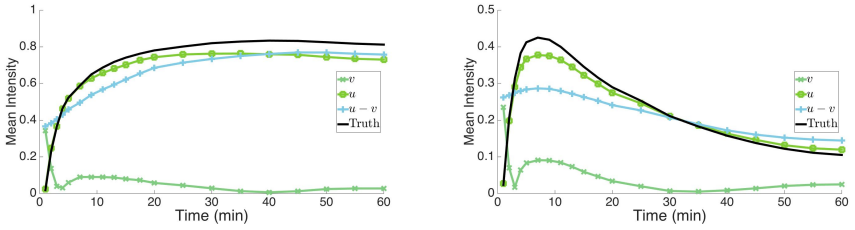


Fig. 8 Dynamic PET simulations with PE2i radiotracer: Decomposition of $\mathrm{ICTV}_{(0.5,0.8),0.09}$ for the first 5 minutes. For visualization purposes we consider the rescaled version for $\boldsymbol{v}$.



(a) Striatum (b) Cortex

Fig. 9 PE2i: The $\mathrm{ICTV}_{(0.5,0.8),0.09}$ decomposition into the $\boldsymbol{u} - \boldsymbol{v}$ and $\boldsymbol{v}$ components for the striatum and cortex regions. The mean intensity for the $\boldsymbol{v}$ component is rescaled for visualization purposes.

FDG radiotracer can be compared in Figure 12(c) and in Figure 7(c) for the PE2i radiotracer.

In the reconstructions with no temporal regularization, one observes noise artifacts in the early frames and low contrast between the white and gray matter, see Figure 10. In the late frames, the contrast was better recovered but with relatively

	MSE (10^{-3})/BIAS ROI			
$TV_{(0.06,0.3)}$	8.3354/0.1621	1.7234/0.1543	1.3244/0.1684	1.2615/0.1646
$TGV_{(0.15,0.25)}$	10.505/0.1747	**1.2094**/0.1438	1.0885/0.1575	1.1067/0.1584
$ICTV_{(0.5,0.8),0.09}$	**7.0436/0.1537**	1.7702/0.1388	1.2056/0.1498	0.9940/**0.1261**
$ICTGV_{(0.8,0.3),0.8}$	8.0809/0.1573	1.2176/**0.1195**	**1.0645**/0.1420	**0.9314**/0.1277
$TV_{(0.2,0)}$	25.642/0.1987	3.9084/0.1682	1.2714/**0.1285**	2.4995/0.1493
$TGV_{(0.3,0)}$	24.128/0.1907	2.8140/0.1542	1.5272/0.1389	2.4122/0.1586
MLEM	25.029/0.1861	4.3788/0.1803	3.7225/0.2128	2.8891/0.1908
MLEM-Filt	20.730/0.1698	2.8498/0.1536	2.2530/0.1716	2.4044/0.1709
	Striatum	Thalamus	Cortex	WM

Table 3 PE2i: Comparison of MSE and bias quality criteria for regions of interest.

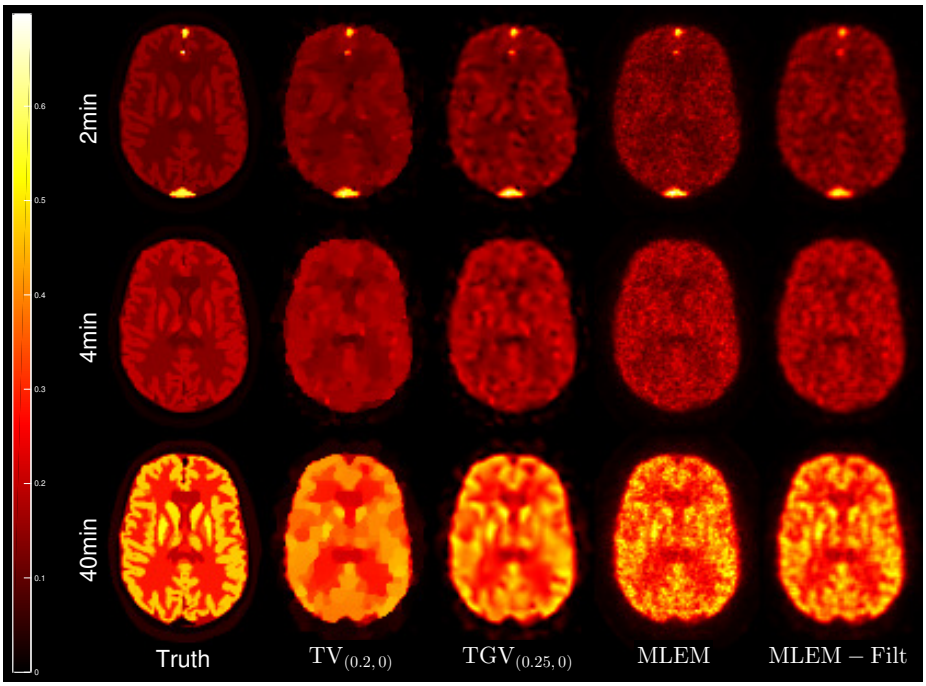


Fig. 10 Dynamic PET simulations with FDG radiotracer: only the spatial regularization and the corresponding MLEM reconstructions.

poor spatial resolution. Again, for both the TV and TGV reconstructions, we encounter either a significant amount of staircasing and loss of contrast or blurring effect in the whole brain area. We refer to the time activity curves in Figure 13, for the regions of interest with the highest uptake, which in this case are the striatum and the cortex. These kind of artifacts reflect to the values of the corresponding quality measures reported in the Tables 4 and 5.

A visual comparison indicates that overall, the quality of images reconstructed with spatiotemporal regularization (Fig. 11) is much higher than their spatial counterparts. Noise is better controlled and the small contrast in the early frames is nicely recovered. In the early frames, we can see noise artifacts in the TV and TGV reconstructions, which are reduced in both cases of the infimal convolution,

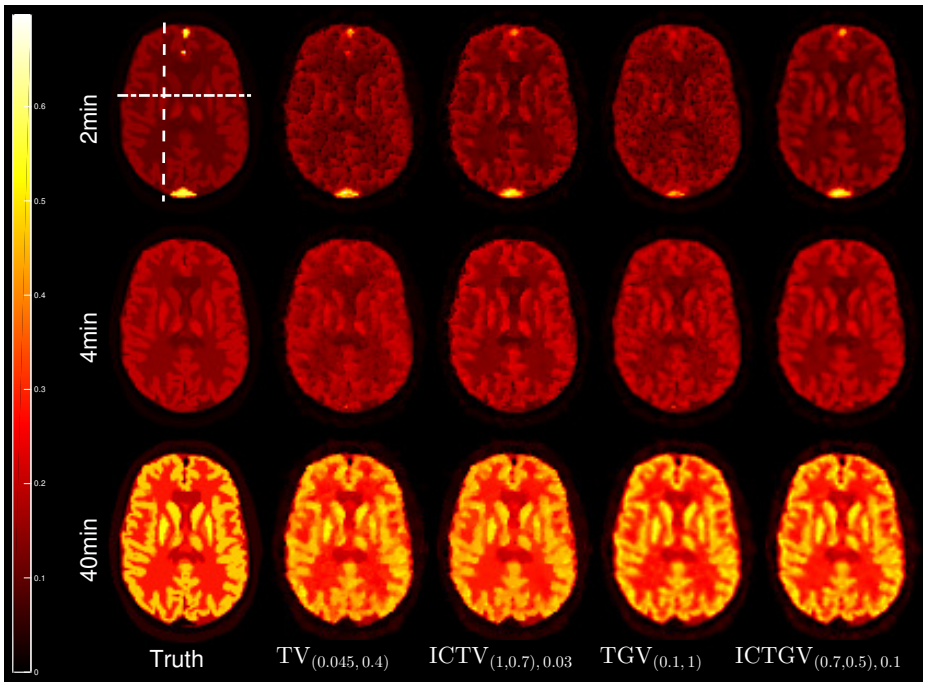
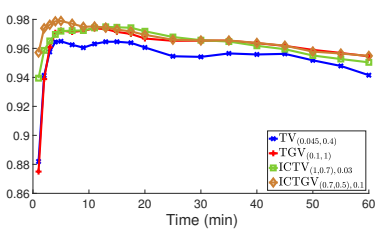
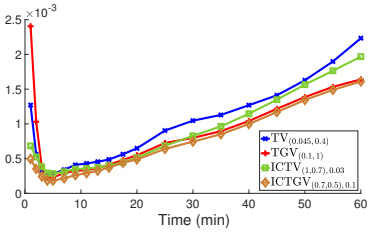


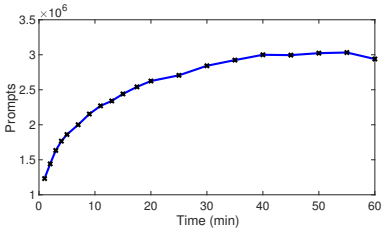
Fig. 11 Dynamic PET simulations with FDG radiotracer: reconstruction methods with spatiotemporal regularization.



(a) SSIM per time-frame



(b) MSE per time-frame



(c) Counts per time-frame

Fig. 12 FDG: SSIM, MSE values and number of counts with respect to time.

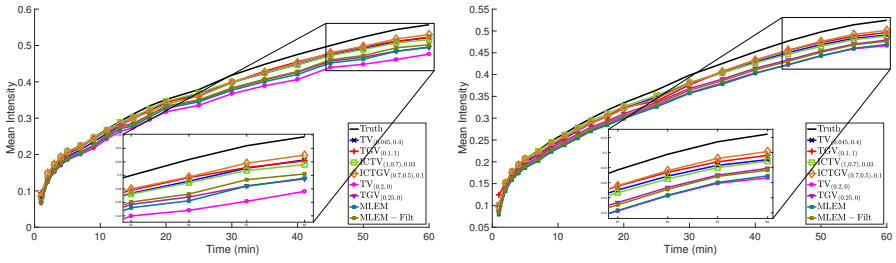
see for instance the SSIM and MSE values per time-frame in Figure 12. Note that in the FDG case, for the whole acquisition time we have overall a smooth variation of the ground truth activity. Therefore, high order regularizers such as TGV and ICTGV are in favour in this case. In the late frames, we observe similar behaviour among the TV/ICTV and TGV/ICTGV reconstructions since after 40th minute the noise level does not variates significantly with respect to time, see Figure 12(c). The quantitative results for all the reconstructions are presented in Tables 4, 5.

	SSIM	MSE (10^{-3})	BIAS
$TV_{(0.045,0.4)}$	0.9534	0.8787	0.2332
$TGV_{(0.1,1)}$	0.9604	0.8132	0.2243
$ICTV_{(1,0.7),0.03}$	0.9648	0.7593	0.2165
$ICTGV_{(0.7,0.5),0.1}$	**0.9686**	**0.6348**	**0.2062**
$TV_{(0.2,0)}$	0.9138	1.6369	0.3380
$TGV_{(0.25,0)}$	0.9185	1.4988	0.3656
MLEM	0.8379	2.1210	0.4347
MLEM-Filt	0.9000	1.6007	0.4084

Table 4 FDG: Comparison of MSE and bias quality criteria for the whole brain.

	MSE (10^{-3})/BIAS ROI			
$TV_{(0.045,0.4)}$	1.4186/0.1059	0.646/0.0756	1.1718/0.0941	0.8739/0.1111
$TGV_{(0.1,1)}$	1.2196/0.1030	0.6105/0.0742	1.1514/0.0975	0.6730/0.1034
$ICTV_{(1,0.7),0.03}$	1.2833/0.0981	0.5519/0.0675	1.0761/0.0845	0.6417/0.0794
$ICTGV_{(0.7,0.5),0.1}$	**1.0316/0.0836**	**0.4763/0.0593**	**0.9300/0.0724**	**0.5414/0.0740**
$TV_{(0.2,0)}$	3.6531/0.1317	2.8986/0.1302	2.0604/0.1085	1.5952/0.1231
$TGV_{(0.25,0)}$	2.7182/0.1158	1.4961/0.0928	2.1229/0.1106	1.4406/0.1270
MLEM	3.5984/0.1383	2.7860/0.1297	3.8144/0.1518	1.6384/0.1377
MLEM-Filt	2.5685/0.1143	1.5662/0.0954	2.4927/0.1193	1.3407/0.1234
	Striatum	Thalamus	Cortex	WM

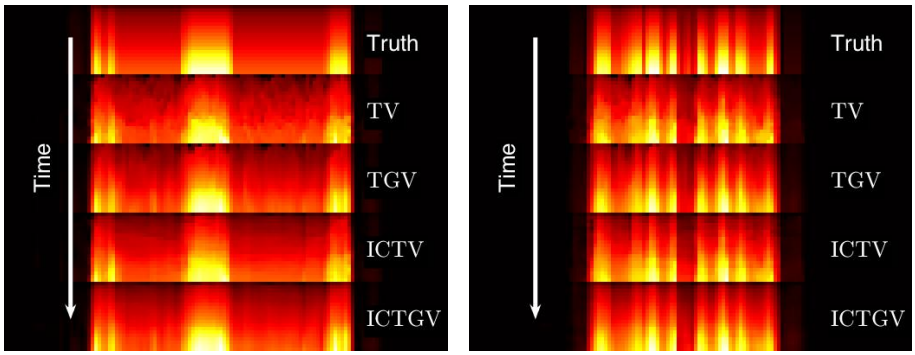
Table 5 FDG: Comparison of MSE and bias quality criteria for regions of interest.



(a) Striatum (b) Cortex

Fig. 13 Results obtained on dynamic PET simulations with FDG radiotracer: Time Activity Curves for regions of interest.

In the FDG dynamic PET simulations, all the best scores were obtained with the ICTGV regularization, see Tables 4 and 5. Similar to the PE2i case, we present the time-line plots in Figure 14. The corresponding reference lines are reported in Figure 11 and are similar to the PE2i radiotracer since we focus again to the striatum region. Note that there are staircasing issues in the ICTV reconstruction and especially in the white matter regions. In terms of the TGV reconstruction, we observe blurring effects that are reduced with ICTGV. Finally, in Figure 15, we present the ICTGV$_{(0.7,0.5),0.1}$ decomposition into the $\boldsymbol{v}$ and $\boldsymbol{u}-\boldsymbol{v}$ components for the first 5 frames as well as the decomposition for the whole sequence for the mean activity of the striatum and cortex regions in Figure 16. Notice that noisy structures are captured at the very first frames, but are not as apparent as in the PE2i case, since in this case the change of prompt counts between time-frames is less compared to the PE2i radiotracer.



(a) Reference line: - - (b) Reference line: - .

Fig. 14 FDG: Time-lines of the spatiotemporal reconstructions. The reference lines are shown in Figure 11.

5.5 Results: DPA

The last set of experiments was conducted on simulations of dynamic PET images with [18F]DPA-714. Compared to the two other types of simulations, the time-frame configuration was set differently than in the two previous cases, see Table 1. The evolution of the number of counts across time was relatively similar to the PE2i case, and one can observe a sudden change in the number of counts for the first 5 minutes, see Figures 18(c) and 7(c). A visual and quantitative analysis of the results lead to the conclusion that the infimal convolution reconstructions and in particular the ICTGV are noticeably better again. The reconstruction with only spatial regularizers were again worse than the other approaches and therefore we only display the spatiotemporal reconstructions along with filtered MLEM, see Figure 17. One can notice that the noise cannot be eliminated using the spatiotemporal TV and TGV regularizers, see also the SSIM and MSE values for the first time-frames in Figure 18. For the ICTV reconstruction, there is still a staircasing

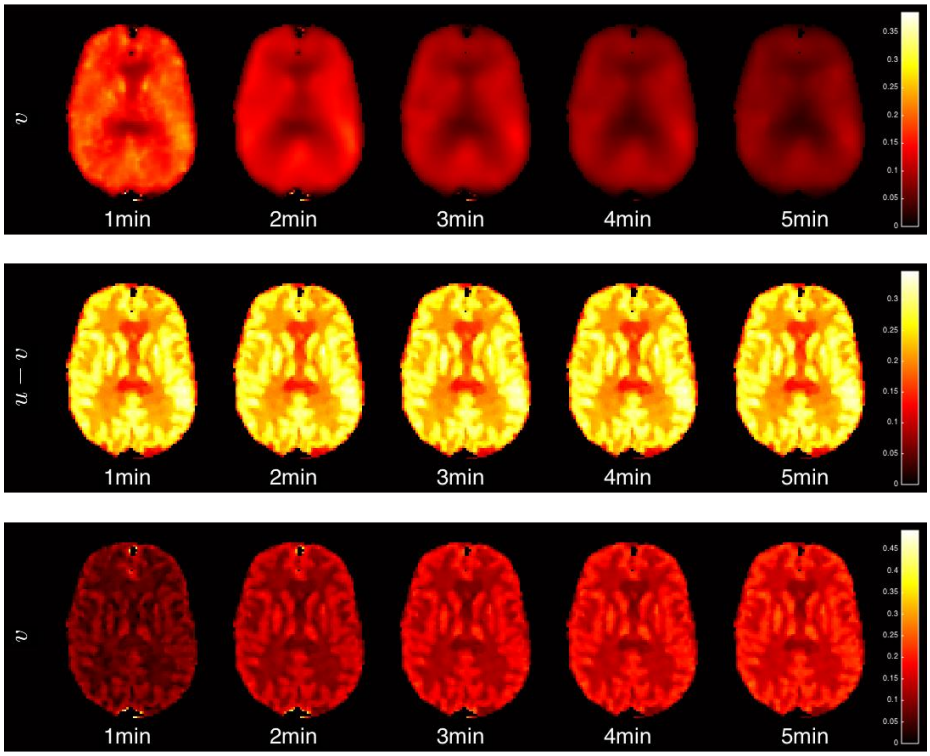
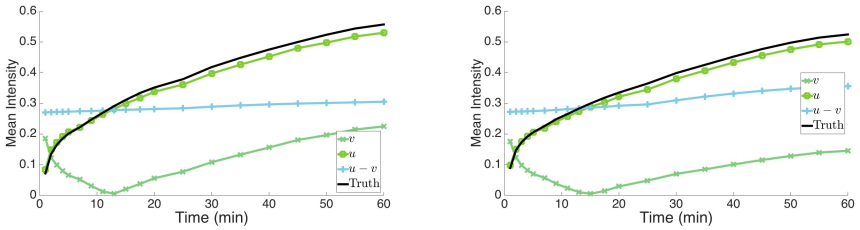


Fig. 15 Dynamic PET simulations with FDG radiotracer: Decomposition of $\mathrm{ICTGV}_{(0.7,0.5),0.1}$ for the first 5 minutes. For visualization purposes we consider the absolute values for $\boldsymbol{u} - \boldsymbol{v}$ and $\boldsymbol{v}$.



(a) Striatum (b) Cortex

Fig. 16 FDG: The $\mathrm{ICTGV}_{(0.7,0.5),0.1}$ decomposition into the $\boldsymbol{u} - \boldsymbol{v}$ and $\boldsymbol{v}$ components for the striatum and cortex regions. The mean intensity for the $\boldsymbol{v}$ component is rescaled for visualization purposes.

artifact especially in the cortex region, see also the time-lines plots in Figure 19. The reference lines are focus on the thalamus region and appear in Figure 17. In healthy subjects, thalamus and cortex are the regions that should present the highest uptake. The ICTGV reconstructions leads to a result where this contrast

is preserved and the thalamus and cortex are well identified. This is justified with the TAC for the thalamus and cortex regions in Figure 20. Finally, we report all the quantitative results in Tables 6 and 7, where for the ICTGV the best values were obtained.

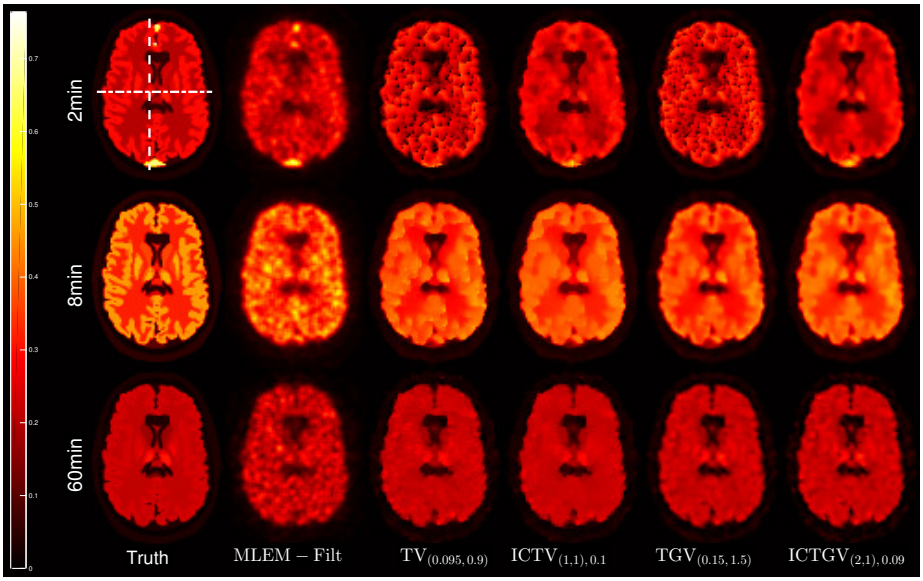


Fig. 17 Dynamic PET simulations with DPA radiotracer: The spatiotemporal and the MLEM-filtered reconstructions.

	SSIM	MSE (10^{-3})	BIAS
$TV_{(0.1,1)}$	0.9333	1.2995	0.3085
$TGV_{(0.15,1.5)}$	0.9369	1.3658	0.3168
$ICTV_{(14,1.5),0.003}$	0.9394	1.1260	0.2907
$ICTGV_{(2,1.5),0.1}$	**0.9397**	**1.0495**	**0.2906**
$TV_{(0.4,0)}$	0.8983	1.9324	0.4069
$TGV_{(0.4,0)}$	0.8874	2.1019	0.4574
MLEM	0.7580	3.5197	0.5661
MLEM-Filt	0.8314	2.5058	0.4814

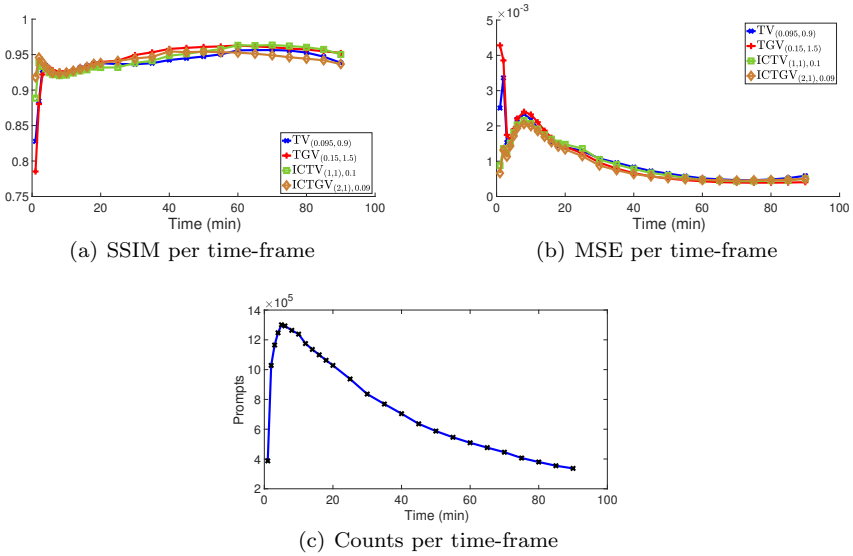
Table 6 DPA: Comparison of MSE and bias quality criteria for the whole brain.

6 Conclusion

In this paper, we deal with spatiotemporal variational regularization methods for dynamic PET reconstruction. We extend the standard total variation and total generalized variation regularizers in the dynamic framework and in addition, we

	MSE (10^{-3})/BIAS ROI			
$TV_{(0.1,1)}$	1.3170/0.1225	1.7956/0.1249	1.9099/0.1294	0.9007/0.1091
$TGV_{(0.15,1.5)}$	1.5339/0.1300	1.4561/0.1159	2.2425/0.1390	0.9052/0.1169
$ICTV_{(14,1.5),0.003}$	1.0533/0.0993	1.6182/0.1045	1.6093/0.1087	**0.7151**/0.0831
$ICTGV_{(2,1.5),0.1}$	**1.0058/0.0966**	**1.0720/0.0849**	**1.5308/0.1024**	0.7410/**0.0856**
$TV_{(0.4,0)}$	1.6892/0.1186	6.1348/0.1995	2.4684/0.1216	1.0207/0.0866
$TGV_{(0.4,0)}$	1.7457/0.1253	3.1974/0.1333	3.0400/0.1374	1.3987/0.1138
MLEM	2.9163/0.1634	6.1558/0.1853	6.3402/0.2093	2.2793/0.1532
MLEM-Filt	2.2980/0.1463	3.8636/0.1433	4.1258/0.1650	1.7533/0.1329
	Striatum	Thalamus	Cortex	WM

Table 7 DPA: Comparison of MSE and bias quality criteria for regions of interest.



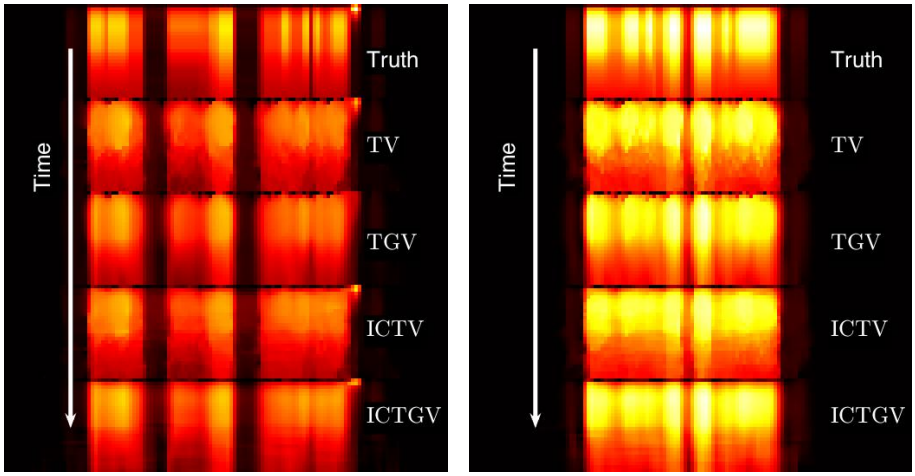
(a) SSIM per time-frame

(b) MSE per time-frame

(c) Counts per time-frame

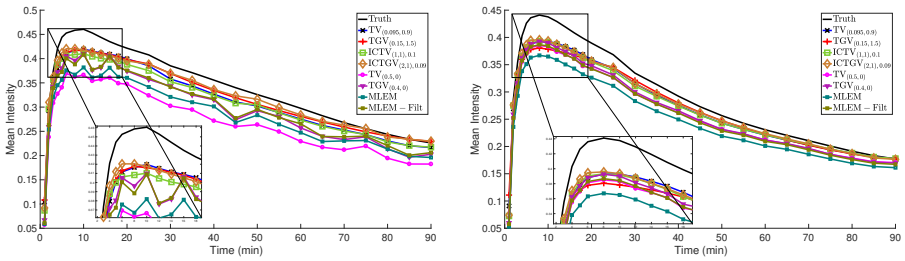
Fig. 18 DPA: SSIM, MSE values and number of counts with respect to time.

consider their infimal convolution variants. In the fidelity term, we take into account the random and scattered events as well as the decay factor for specific radiotracers and the non-uniform time discretization along the temporal direction. The corresponding minimization problem is solved numerically, with a preconditioned primal-dual gradient technique proposed in [26]. For our numerical experiments, we use simulated brain phantoms and test our proposed regularizers for three different radiotracers. We observe that spatiotemporal reconstructions are superior to MLEM and also to the non-temporal reconstructions, where most of the brain regions are either noisy or with staircasing and smoothing issues. Based on the quantitative and qualitative comparative results, we emphasize on the significant potential benefit of the infimal convolution for dynamic PET image reconstruction, especially when the radioactivity changes abruptly over a short period of time. As a future work, we aim to test the proposed regularizers on realistic data obtained from a microPET for dynamic PET reconstruction on mice.



(a) Reference line: - - (b) Reference line: - .

Fig. 19 DPA: Time-lines of the spatiotemporal reconstructions. The reference lines are shown in Figure 17.



(a) Thalamus (b) Cortex

Fig. 20 Results obtained on dynamic PET simulations with DPA radiotracer: Time Activity Curves for regions of interest (thalamus and cortex).

7 Acknowledgments

E. P. would like to thank Martin Holler and Kristian Bredies for discussions on the model and the research visit in Graz.

References

1. M. Benning, C.-B. Schönlieb, T. Valkonen, and V. Vlačić. *preprint arXiv*.
2. M. Bergounioux and L. Piffet. A second-order model for image denoising. *Set-Valued and Variational Analysis*, 18(3-4):277–306, 2010. http://dx.doi.org/10.1007/s11228-010-0156-6.
3. M. Bertero, H. Lanteri, and L. Zanni. Iterative image reconstruction: a point of view. *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Ther-*

apy (IMRT), (CRM Series vol 7) ed Y Censor, M Jiang and A K Louis, pages 37–63, 2008.
4. K. Bredies, K. Kunisch, and T. Pock. Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010. http://dx.doi.org/10.1137/090769521.
5. C. Brune, A. Sawatzky, F. Wübbeling, T. Kösters, and M. Burger. Forward backward em-tv methods for inverse problems with poisson noise. *Preprint*, 2010.
6. M. Burger, J. Müller, E. Papoutsellis, and C.-B. Schönlieb. Total variation regularization in measurement and image space for PET reconstruction. *Inverse Problems*, 30(10):105003, 2014. http://stacks.iop.org/0266-5611/30/i=10/a=105003.
7. M. Burger, A. Sawatzky, and G. Steidl. First order algorithms in variational image processing. pages 345–407, 2016. "http://dx.doi.org/10.1007/978-3-319-41589-5_10".
8. A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997. "http://dx.doi.org/10.1007/s002110050258".
9. A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. http://dx.doi.org/10.1007/s10851-010-0251-1.
10. I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1999. http://epubs.siam.org/doi/abs/10.1137/1.9781611971088.
11. E. Esse, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010. http://dx.doi.org/10.1137/09076934X.
12. T. Goldstein, M. Li, X. Yuan, E. Esser, and R. Baraniuk. Adaptive primal-dual hybrid gradient methods for saddle-point problems. *arXiv preprint arXiv:1305.0546*, 2013.
13. S. Guérit, L. Jacques, B. Macq, and J. A. Lee. Post-reconstruction deconvolution of pet images by total generalized variation regularization. pages 629–633, Aug 2015. 10.1109/EUSIPCO.2015.7362459.
14. H. Guo, R. A. Renaut, K.Chen, and E. Reiman. Fdg-pet parametric imaging by total variation minimization. *Computerized Medical Imaging and Graphics*, 33(4):295 – 303, 2009. https://doi.org/10.1016/j.compmedimag.2009.01.005.
15. W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing*, 76(1-2):109–133, 2006. http://dx.doi.org/10.1007/s00607-005-0119-1.
16. M. Holler and K. Kunisch. On Infimal Convolution of TV-Type Functionals and Applications to Video and Image Reconstruction. *SIAM Journal on Imaging Sciences*, 7(4):2258–2300, 2014. http://dx.doi.org/10.1137/130948793.
17. J.Fessler. Michigan image reconstruction toolbox.
18. E. Jonsson, S-C.Huang, and T. Chan. Total-variation regularization in positron emission tomography. 1998.
19. F. Knoll, M. Holler, T. Koesters, R. Otazo, K. Bredies, and D. K. Sodickson. Joint mr-pet reconstruction using a multi-channel image regularizer. *IEEE Transactions on Medical Imaging*, 36(1):1–16, Jan 2017.
20. S. Lefkimmiatis, A. Roussos, P. Maragos, and M. Unser. Structure tensor total variation. *SIAM Journal on Imaging Sciences*, 8(2):1090–1122, 2015. https://doi.org/10.1137/14098154X.
21. J. C. De los Reyes, C.-B.Schönlieb, and T. Valkonen. Bilevel parameter learning for higher-order total variation regularisation models. *Journal of Mathematical Imaging and Vision*, 57(1):1–25, 2017. "http://dx.doi.org/10.1007/s10851-016-0662-8".
22. Y. Malitsky and T. Pock. A first-order primal-dual algorithm with linesearch. https://arxiv.org/pdf/1608.08883.pdf.
23. C. Melo, Y. Boursier, J.-F. Aujol, and S. Anthoine. Some proximal methods for Poisson intensity CBCT and PET. *Inverse Problems and Imaging*, 6(4):565–598, 2012.
24. F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. Society for Industrial and Applied Mathematics, 2001. http://dx.doi.org/10.1137/1.9780898718324.
25. K. Papafitsoros and C.-B Schönlieb. A combined first and second order variational approach for image reconstruction. *Journal of Mathematical Imaging and Vision*, 48(2):308–338, Feb 2014. https://doi.org/10.1007/s10851-013-0445-4.
26. T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *2011 International Conference on Computer Vision*, pages 1762–1769, 2011. http://dx.doi.org/10.1109/ICCV.2011.6126441.

27. N. Pustelnik, C. Chaux, J.-C. Pesquet, and C. Comtat. Parallel algorithm and hybrid regularization for dynamic pet reconstruction. In *IEEE Medical Imaging Conference*, pages 2423–2427, Knoxville, Tennessee, United States, October 2010. https://hal-upec-upem.archives-ouvertes.fr/hal-00733493.
28. R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
29. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
30. M. Schloegl, M. Holler, A. Schwarzl, K.Bredies, and R.Stollberger. Infimal convolution of total generalized variation functionals for dynamic mri. *Magnetic Resonance in Medicine*, 2016. = http://dx.doi.org/10.1002/mrm.26352.
31. S. Setzer, G. Steidl, and T.Teuber. Infimal convolution regularizations with discrete l1-type functionals. *Comm. Math. Sci*, 9(3):797–872, 2011.
32. E. Y. Sidky, J. H .Jørgensen, and X.Pan. Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle-pock algorithm. *Physics in Medicine and Biology*, 57(10):3065, 2012. http://stacks.iop.org/0031-9155/57/i=10/a=3065.
33. S. Stute, C. Tauber, C. Leroy, M. Bottlaender, V. Brulon, and C. Comtat. Analytical simulations of dynamic pet scans with realistic count rates properties. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015 IEEE*, pages 1–3. IEEE, 2015.
34. C. Tauber, S. Stute, M. Chau, P. Spiteri, S. Chalon, D. Guilloteau, and I. Buvat. Spatio-temporal diffusion of dynamic pet images. *Physics in Medicine and Biology*, 56(20):6583, 2011. http://stacks.iop.org/0031-9155/56/i=20/a=004.
35. T. Valkonen and T. Pock. Acceleration of the pdhgm on partially strongly convex functions. *Journal of Mathematical Imaging and Vision*, pages 1–21, 2016. http://dx.doi.org/10.1007/s10851-016-0692-2.
36. Z. Wang, A. C. Bovik, H. R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
37. M. N. Wernick and J. N. Aarsvold. *Emission tomography : the fundamentals of PET and SPECT*. Elsevier Academic Press, 2004.
38. Haiqing Yu, Shuhang Chen, Yunmei Chen, and Huafeng Liu. Joint reconstruction of dynamic pet activity and kinetic parametric images using total variation constrained dictionary sparse coding. *Inverse Problems*, 33(5):055011, 2017. http://stacks.iop.org/0266-5611/33/i=5/a=055011.