King's Research Portal

Document Version
Version created as part of publication process; publisher's layout; not normally made publicly available

Link to publication record in King's Research Portal

Download date: 29. Dec. 2024

Synergistic PET and SENSE MR image reconstruction using joint sparsity regularization

Abolfazl Mehranian[†], Martin A. Belzunce, Claudia Prieto, Alexander Hammers and Andrew J. Reader

Abstract–In this work we propose a generalized joint sparsity regularization prior and reconstruction framework for the synergistic reconstruction of PET and undersampled sensitivity encoded (SENSE) MRI data with the aim of improving image quality beyond that obtained through conventional independent reconstructions. The proposed prior improves upon the joint total variation (TV) using a non-convex potential function that assigns a relatively lower penalty for the PET and MR gradients whose magnitudes are jointly large, thus permitting the preservation and formation of common boundaries irrespective of their relative orientation. The alternating direction method of multipliers (ADMM) optimization framework was exploited for the joint PET-MR image reconstruction. In this framework, the joint maximum *a posteriori* objective function was effectively optimized by alternating between well-established regularized PET and MR image reconstructions. Moreover, the dependency of the joint prior on the PET and MR signal intensities was addressed by a novel alternating scaling of the distribution of the gradient vectors. The proposed prior was compared with the separate TV and joint TV regularization methods using extensive simulation and real clinical data. In addition, the proposed joint prior was compared to the recently proposed linear parallel level sets (PLS) method using a benchmark simulation dataset. Our simulation and clinical data results demonstrated the improved quality of the synergistically reconstructed PET-MR images compared to unregularized and conventional separately regularized methods. It was also found that the proposed prior can outperform both joint TV and linear PLS regularization methods in assisting edge preservation and recovery of details which are otherwise impaired by noise and aliasing artifacts. In conclusion, the proposed joint sparsity regularization within the presented ADMM reconstruction framework is a promising technique, nonetheless our clinical results showed that the clinical applicability of joint reconstruction might be limited in current PET-MR scanners, mainly due to the lower resolution of PET images.

Index Terms– Multi-modal imaging, PET-MRI, Synergistic reconstruction, Sensitivity encoding, Sparsity regularization, Total variation.

I. Introduction

Simultaneous PET-MR systems have recently been introduced in clinical practice as a new dual-modality imaging technique offering the capability of combined molecular and morphological assessment of different diseases [1]. In such scanners, the PET and MRI data are simultaneously acquired which opens up new opportunities for more fully realising the benefits of these hybrid imaging modalities [2]. For instance, these dual-modality systems provide elegant solutions to PET motion correction through estimation of the motion field from MRI data [3], PET partial volume correction using high resolution MR images [4], and supplemental functional information for PET pharmacokinetic modelling such as the arterial input function [5]. Moreover, for PET attenuation correction, advanced and clinically feasible MRI data acquisitions are currently under development for providing accurate attenuation maps and thus improved PET quantification [6].

Another promising aspect of simultaneous PET-MR that has been unexplored until recently [7], is to synergistically reconstruct PET and MRI data in order to improve image quality of both PET and MR images beyond what is obtained by the conventional independent image reconstruction techniques. MRI can suffer from long duration data acquisitions. Currently fast MRI data acquisition methods focus on partial Fourier and multi-coil parallel MRI, which are based on acquiring less MRI data (i.e., k-space undersampling) and compensate for the missing data using k-space Hermitian symmetries or the information available in coil sensitivity profiles [8]. However, at high undersampling rates, the reconstructed images exhibit aliasing artifacts that can severely degrade image quality. Emerging trends have therefore focused on the combination of parallel MRI (such as sensitivity encoding –SENSE) and compressed sensing (CS) techniques [9, 10]. In this regard, joint PET and MR image reconstruction can be employed to still further reduce the sampling requirements of CS-MRI [11], by exploiting the common features and similarities between PET and MR images.

Generally, PET and MR images share some common anatomical similarities depending on the choice of PET radiotracer and MRI acquisition protocol. In joint PET-MR image reconstruction, the form of the joint prior function is of paramount importance to accurately exploit the shared anatomical/physiological information and effectively cope with the challenges encountered in multi-modal imaging. The similarity of PET and MR images is assumed to be at the tissue boundaries that separate different regions. However, the signal intensity and the contrast orientation are different between these modalities. Moreover, both PET and MR images can have unique features that do not have a corresponding feature in the other imaging modality. Therefore, the joint prior should exploit and promote structural or physiological similarities between the two images while preserving modality-unique features and avoiding cross-talk artifacts. Finally, it should be amenable to a practical optimization scheme, resulting in numerically feasible joint PET and MR image reconstruction in clinical practice.

The joint processing and reconstruction of multi-channel data have been recently investigated in the context of colour image processing [12, 13], multi-energy CT imaging [14], multi-contrast [15] and multi-coil MRI [16, 17]. In most of these studies, joint prior functions were employed to promote the joint sparsity and alignment of the image gradients of each channel using vectorial joint total variation (TV) or total nuclear variation. Recently, Ehrhardt *et al* [7] reported the first attempt in joint PET-MR image reconstruction based on the structural similarity of PET and MR images measured by the parallelism of their level sets or in practice their gradients at each spatial location. The authors defined a generalized dissimilarity metric between the image gradients

of the two modalities, which achieves its minimal value when the gradients are parallel or anti-parallel. Using simulations, the authors demonstrated that their parallel level sets (PLS) method substantially outperformed single-channel (separate) TV and joint TV regularizations. Despite promising results, the PLS prior has two major limitations: i) its dependency on the magnitude of the image gradients and ii) the suppression and transfer of modality-unique features as can be concluded from the presented results. Inspired by TV Bregman colour image processing [13], Rasch *et al* [18] studied a structural similarity metric, based on the Bregman distance of TV norms of the PET and MR images, that is insensitive to signal intensities. Similar to [13], the infimal convolution of Bregman distances were employed to avoid the dependency on the orientation of edge direction (contrast).

More recently, Knoll *et al* [19] proposed a multi-channel total generalized variation (TGV) regularization for joint PET-MR reconstruction. The multi-channel TGV was defined based on the nuclear norm of the Jacobian matrix of PET and MR images. The nuclear norm was used as a convex surrogate for the rank of a matrix, which counts the number of non-zero singular values of the matrix. At a common edge, the gradient vectors for each modality are parallel or anti-parallel, therefore the Jacobian matrix will be rank one and thus have only one non-zero singular value [14]. In [19], the multi-channel TGV was applied for joint PET-MR image reconstruction and it was demonstrated that the proposed method outperformed conventional methods. However, in contrast to the low-rank prior, the nuclear norm prior depends on the magnitude of the singular values of the Jacobian matrix, which in turn depends on the signal intensity and edge orientation, therefore the multi-channel TGV prior is also potentially adversely dependent on these factors.

In this study, we propose a non-convex joint sparsity prior by the generalization of the joint TV regularization to encourage the formation and preservation of common boundaries between PET and MR images independent of their orientation while preserving modality-unique features. We propose an alternating scaling approach to cope with the dependency of the prior on the magnitudes of PET-MR image gradients. In addition, we present a joint PET and SENSE MR image reconstruction framework based on the augmented Lagrangian method that leads to a numerically efficient distributed optimization. Furthermore, we elaborate the proposed joint prior and its numerical implementation and report the performance of the resulting algorithm using extensive simulated data as well as real clinical data.

II. MATERIAL AND METHODS

A. Problem formulation

In the statistical modelling of PET data, the probability distribution of the measured number of prompt coincidences, p_i, in the ith line of response (LOR) is best modelled by the Poisson distribution, that is:

$$p_i \sim \text{Poisson}\{\bar{y}_i + \bar{r}_i\} \qquad (1)$$

where $\bar{y}_i$ and $\bar{r}_i$ denote the mean number of true and background (random and scatter) coincidences in the ith LOR (or sinogram bin). For multi-coil parallel MR imaging, the MRI measured signal (the electromotive force induced in the receiver coils), s_{li}, at the ith time point in the lth coil is modelled as the true signal $\bar{s}_l(t_i)$ contaminated by measurement errors ϵ_{li}, that are well modelled by additive complex Gaussian noise of variance σ_{li}^2:

$$s_{li} = \bar{s}_l(t_i) + \epsilon_{li}, \qquad \epsilon_{li} \sim \mathcal{N}\{0, \sigma_{li}\} \qquad (2)$$

In model-based PET and MR image reconstruction, the system's response to the object's PET activity (mean radioactivity concentration), $\boldsymbol{u} \in \mathbb{R}^{N_u}$, and MR signal, $\boldsymbol{v} \in \mathbb{C}^{N_v}$, distributions is often modelled as linear and spatially variant, such that discretized PET and MRI acquisition models can be expressed as:

$$\bar{\boldsymbol{y}} = \boldsymbol{Pu}, \qquad \bar{\boldsymbol{s}} = \boldsymbol{Ev} \qquad (3)$$

where $\boldsymbol{P} \in \mathbb{R}^{M_u \times N_u}$ is the PET system matrix and M_u and N_u are the number of sinogram bins and the number of voxels of the PET activity image. The system matrix can be factorized as $\boldsymbol{P} = \boldsymbol{WGB}$, where $\boldsymbol{B}$ is an $N_u \times N_u$ matrix representing the image-space point spread function

(PSF) of the PET scanner, $\boldsymbol{G}$ is an $M_u \times N_u$ geometric transition matrix that maps image space to sinogram space and $\boldsymbol{W}$ is a $M_u \times M_u$ diagonal matrix consisting of the product of the attenuation and normalization factors [20]. $\boldsymbol{E} \in \mathbb{C}^{M_v L \times N_v}$ is the MR Fourier encoding matrix consisting of the product of the discrete Fourier transform matrix $\boldsymbol{F}$ with a sub-sampled k-space and coil sensitivity profiles. M_v, L and N_v are the number of k-space samples, coils and MR image voxels, respectively. The encoding matrix can be factorized as $\boldsymbol{E} = (\boldsymbol{I}_L \otimes \boldsymbol{\Theta F})\boldsymbol{C}$, where $\boldsymbol{C} = [\boldsymbol{C}_1^H, \ldots, \boldsymbol{C}_L^H]^H$ is an $N_v L \times N_v$ matrix composed of $N_v \times N_v$ diagonal spatial sensitivity matrices associated with the L coils, $\boldsymbol{F}$ is a $N_v \times N_v$ Fourier basis matrix, $\boldsymbol{\Theta}$ is an $M_v \times N_v$ k-space undersampling (trajectory) matrix with $M_v \leq N_v$ samples and $\otimes$ is the Kronecker product and $\boldsymbol{I}_L$ an identity matrix of size L [21]. $(\cdot)^H$ denotes the Hermitian transpose, or conjugate transpose.

Given an estimate of the PET background coincidences $\bar{\boldsymbol{r}} \in \mathbb{R}^{M_u}$ and the MR sensitivity maps $\boldsymbol{C}$, the task of joint PET and SENSE-MR image reconstruction is to estimate the object's PET activity, $\boldsymbol{u}$, and the MR signal, $\boldsymbol{v}$, from the measured PET sinogram data, $\boldsymbol{y} \in \mathbb{R}^{M_u}$, and MRI k-space data, $\boldsymbol{s} \in \mathbb{C}^{M_v L}$ in such a way that common edges can be shared and propagated between the PET, $\hat{\boldsymbol{u}}$, and MRI, $\hat{\boldsymbol{v}}$, image estimates. The joint reconstruction can be well developed in the Bayesian estimation framework based on the maximum *a posteriori* (MAP) criterion. The MAP image estimates can be found by the following negative log-posterior minimization:

$$(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}) = \underset{\boldsymbol{u}, \boldsymbol{v} \in \Omega}{\text{argmin}}\{\mathcal{D}_u(\boldsymbol{Pu}, \boldsymbol{y}) + \mathcal{D}_v(\boldsymbol{Ev}, \boldsymbol{s}) + R(\boldsymbol{u}, \boldsymbol{v})\} \qquad (4)$$

$$\Omega = \{(\boldsymbol{u}, \boldsymbol{v}): \boldsymbol{u} \in \mathbb{R}_+^{N_u}, \boldsymbol{v} \in \mathbb{C}^{N_v}\}$$

where $\mathcal{D}_u$ and $\mathcal{D}_v$ are, respectively, the PET negative Poisson-log likelihood and the MRI negative Gaussian-log likelihood given by:

$$\mathcal{D}_u(\boldsymbol{Pu}, \boldsymbol{y}) = \sum_{i=1}^{M_u}([\boldsymbol{Pu}]_i + \bar{r}_i - \boldsymbol{y}_i \log([\boldsymbol{Pu}]_i + \bar{r}_i)) \qquad (5)$$

$$\mathcal{D}_v(\boldsymbol{Ev}, \boldsymbol{s}) = \sum_{l=1}^{L} \sum_{i=1}^{M_v} w_{li}([\boldsymbol{Ev}]_{li} - \boldsymbol{s}_{li})^2 \qquad (6)$$

where w_{li} is an element of a $\boldsymbol{W} \in \mathbb{R}^{M_v L \times M_v L}$ weighting matrix obtained from the inversion of the noise covariance matrix [22]. $R(\boldsymbol{u}, \boldsymbol{v})$ is a joint prior function that permits the preservation and propagation of common features between PET and MR images.

B. Generalized joint sparsity regularization

It is well known that the conventional convex TV regularization, defined as the L_2 norm of the spatial image gradients at each voxel, favours sparsity of the image gradients and thus tends to preserve only sharp edges and also limits the effects of noise. In multi-channel image processing and reconstruction, the extension of the TV regularization to *vectorial* TV or joint TV has been studied to exploit the correlated features between image channels [23]. The joint TV thus favours image gradients that are not only sparse but also jointly sparse. In this study, we consider a generalized joint sparsity regularization of the form:

$$R(\boldsymbol{u}, \boldsymbol{v}) = \lambda \sum_{j}^{N} \psi\left(\left\|\begin{pmatrix} \alpha_u \boldsymbol{\Phi}_u \boldsymbol{\nabla u} \\ \alpha_v \boldsymbol{\Phi}_v \boldsymbol{\nabla v} \end{pmatrix}_j\right\|_2\right) \qquad (7)$$

where $\left\|\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}_j\right\|_2 \triangleq \sqrt{\|\boldsymbol{x}_j\|_2^2 + \|\boldsymbol{y}_j\|_2^2}$, $\boldsymbol{\nabla} = [\boldsymbol{\nabla}^{(1)}; \boldsymbol{\nabla}^{(2)}; \boldsymbol{\nabla}^{(3)}] \in$ $\mathbb{R}^{3N \times N}$, with $N \in \{N_u, N_v\}$, is a discrete gradient matrix composed of directional first-order finite difference matrices (horizontal, vertical and axial directions) with periodic boundary conditions. Here, it is assumed that the PET and MR images are of different resolutions with N_u and N_v voxels. The $\boldsymbol{\Phi}_u$ and $\boldsymbol{\Phi}_v$ are two mapping operators that consist of the registration and sampling of the spatial gradient matrices of PET and MR images to a given FOV and image resolution with N voxels. The scalars $\alpha_v, \alpha_u > 0$ weight the gradients, $[\boldsymbol{\nabla u}]_j \in \mathbb{R}^3$, $[\boldsymbol{\nabla v}]_j \in \mathbb{C}^3$, of the PET and MR images, respectively, in order to cope with the differences in signal intensity between these images.

$\psi(t)$ is a potential function that is used to promote the formation of common features between MRI and PET. $\lambda \in \{\lambda_u, \lambda_v\}$ is a regularization parameter controlling the regularity of the solution. In this work, we studied the generalization of the joint total variation with the following quasi-convex potential function [11]:

$$\psi(t) = \frac{1}{\sigma}\left(1 - e^{-\sigma t}\right) \qquad (8)$$

where $\sigma > 0$ is a relaxation parameter controlling the degree of sparsity promotion. As $\sigma \to 0$, the potential function $\psi(t) \to t$, whereby $R(\boldsymbol{u}, \boldsymbol{v})$ becomes convex and approaches the joint TV prior, while for $\sigma > 0$, $R(\boldsymbol{u}, \boldsymbol{v})$ becomes non-convex. Compared to channel-by-channel and joint TV priors, the proposed prior assigns lower penalty on the PET and MR gradients whose magnitudes are jointly large compared to the other priors, leading to further edge-preservation and suppression of smoothing (or diffusion) across those boundaries (see supplementary material Fig. 1). The joint priors can also be characterized by diffusion analysis, which involves the derivative of the functional $R(\boldsymbol{u}, \boldsymbol{v})$. The partial derivative of R with respect to an element of $\boldsymbol{u}$ is given by:

$$\frac{\partial R}{\partial u_j} = \lambda \alpha_u \boldsymbol{\nabla}^T \boldsymbol{\Phi}_u^T \left(\boldsymbol{D}[\boldsymbol{\Phi}_u \boldsymbol{\nabla}\boldsymbol{u}]_j\right)$$

$$\boldsymbol{D} = \mathrm{diag}\{\psi'(\kappa)\kappa^{-1}\}, \qquad \kappa = \left\|\begin{pmatrix} \alpha_u \boldsymbol{\Phi}_u \boldsymbol{\nabla}\boldsymbol{u} \\ \alpha_v \boldsymbol{\Phi}_v \boldsymbol{\nabla}\boldsymbol{v} \end{pmatrix}_j\right\|_2 \qquad (9)$$

where ψ' is the derivative of ψ and $\boldsymbol{D}$ contains what are known as the joint diffusivity coefficients. The partial derivative of R with respect to an element of $\boldsymbol{v}$ is similar to (9) with the same $\boldsymbol{D}$. The single-channel (or separate) diffusion of the PET images is obtained by setting $\alpha_u = 0$, and by setting $\alpha_v = 0$ for the MR images. The diffusivity coefficients tend to suppress diffusion across valid boundaries, while enforcing diffusion within regions between the boundaries thereby preserving edges. The diffusivity coefficients for the joint TV and non-convex (NCX) priors are given by:

$$\boldsymbol{D}_{TV} = \mathrm{diag}\{\kappa^{-1}\}, \quad \boldsymbol{D}_{NCX} = \mathrm{diag}\{\exp(-\sigma\kappa)\,\kappa^{-1}\} \qquad (10)$$

where the $\boldsymbol{D}_{NCX}$ has an additional exponential term that weights the diffusivity coefficients of the joint TV. For large joint PET-MR gradient magnitudes, this exponential term approaches zero, thereby resulting in the reduction of diffusion across the corresponding boundaries. In Fig. 1, the role of the diffusivity coefficients in the identification and preservation of edges is further illustrated. The MRI and PET images show a case where both modalities have common boundaries, but as indicated by the arrows, a common boundary has not been recovered during separate TV PET image regularization compared to the corresponding MR image estimate. The figure compares the diffusivity coefficients of separate TV priors defined on the MR and PET images, as well as those of the joint TV and NCX priors. As MRI provides information about the boundary, its presence has been identified in the joint TV and NCX diffusivity coefficients. As a result, during joint PET-MR reconstruction, the PET image regularization will therefore be supervised by the joint diffusivity coefficients, instead of the separate coefficients, which fail to identify that boundary. As shown, the coefficients of the proposed non-convex prior tend to further suppress diffusion by assigning lower weights or penalty on smoothing across boundaries, especially those of large magnitude.

C. Optimization

To solve the joint optimization problem in (4) with the proposed non-convex joint prior, we employed the alternating direction method of multipliers (ADMM) based on the augmented Lagrangian (AL) method. The AL method, which was originally developed for solving constrained optimization problems, decomposes a large-scale master problem into a set of simpler sub-problems that can be individually solved using more efficient and well-established optimization algorithms [24]. In addition, cost functions with non-continuously differentiable regularizers, such as the TV prior, can be effectively optimized through a variable splitting technique and introduction of an auxiliary variable.
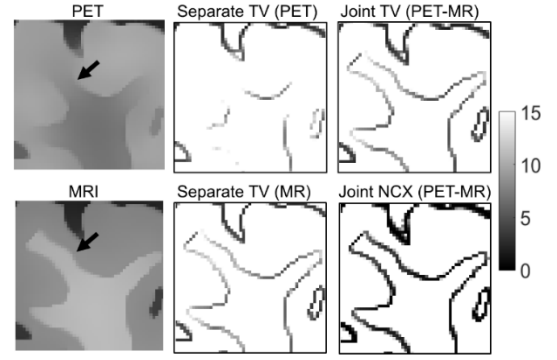

Fig 1. The joint diffusivity coefficients between PET and MRI images.

Let $\boldsymbol{z}_u = \left[\boldsymbol{z}_u^{(1)}; \boldsymbol{z}_u^{(2)}; \boldsymbol{z}_u^{(3)}\right] \in \mathbb{R}^{3N_u}$ and $\boldsymbol{z}_v = \left[\boldsymbol{z}_v^{(1)}; \boldsymbol{z}_v^{(2)}; \boldsymbol{z}_v^{(3)}\right] \in \mathbb{C}^{3N_v}$ be two auxiliary variables, equal to the spatial gradients of the PET, $\boldsymbol{u}$, and MR, $\boldsymbol{v}$, images, respectively. The problem in (4) with the joint prior defined in (7) can then be redefined by the following equality constrained problem that can be optimized in the AL framework:

$$(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}) = \underset{\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{\Omega}}{\mathrm{argmin}} \left\{ \mathcal{D}_u(\boldsymbol{Pu}, \boldsymbol{y}) + \mathcal{D}_v(\boldsymbol{Ev}, \boldsymbol{s}) \right.$$
$$\left. + \lambda \sum_j^N \psi\left(\left\|\begin{pmatrix}\alpha_u \boldsymbol{\Phi}_u \boldsymbol{z}_u \\ \alpha_v \boldsymbol{\Phi}_v \boldsymbol{z}_v\end{pmatrix}_j\right\|_2\right)\right\} \qquad (11)$$

subject to $\boldsymbol{z}_u = \boldsymbol{\nabla}\boldsymbol{u}, \boldsymbol{z}_v = \boldsymbol{\nabla}\boldsymbol{v}$.

The AL functional for the above problem is defined as:

$$\mathcal{L}(\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{z}_u, \boldsymbol{z}_v, \boldsymbol{\gamma}_u, \boldsymbol{\gamma}_v) =$$
$$\Gamma(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{z}_u, \boldsymbol{z}_v) + \boldsymbol{\gamma}_u^T(\boldsymbol{\nabla}\boldsymbol{u} - \boldsymbol{z}_u) + \boldsymbol{\gamma}_v^T(\boldsymbol{\nabla}\boldsymbol{v} - \boldsymbol{z}_v)$$
$$+ \frac{\rho_u}{2}\|\boldsymbol{\nabla}\boldsymbol{u} - \boldsymbol{z}_u\|_2^2 \qquad (12)$$
$$+ \frac{\rho_v}{2}\|\boldsymbol{\nabla}\boldsymbol{v} - \boldsymbol{z}_v\|_2^2$$

where $\Gamma(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{z}_u, \boldsymbol{z}_v)$ is the objective function defined in (11), $\boldsymbol{\gamma}_u \in \mathbb{R}^{3N_u}$, $\boldsymbol{\gamma}_v \in \mathbb{C}^{3N_v}$ and $\rho_u, \rho_v > 0$ are respectively the Lagrange multipliers and the penalty parameter associated with the two equality constraints in (11). The ADMM algorithm seeks to minimize (12) by alternating through the following sub-problems, in which $\mathcal{L}$ is minimized with respect to a given variable while the other variables are held constant. Following completing the squares in (12), the alternating minimizations are given by:

$$\boldsymbol{u}^{k+1} = \underset{\boldsymbol{u}}{\mathrm{argmin}}\left\{\mathcal{D}_u(\boldsymbol{Pu}, \boldsymbol{y}) + \frac{\rho_u}{2}\|\boldsymbol{\nabla}\boldsymbol{u} - \boldsymbol{z}_u^k + \boldsymbol{\gamma}_u^k/\rho_u\|_2^2\right\} \qquad (13)$$

$$\boldsymbol{v}^{k+1} = \underset{\boldsymbol{v}}{\mathrm{argmin}}\left\{\mathcal{D}_v(\boldsymbol{Ev}, \boldsymbol{s}) + \frac{\rho_v}{2}\|\boldsymbol{\nabla}\boldsymbol{v} - \boldsymbol{z}_v^k + \boldsymbol{\gamma}_v^k/\rho_v\|_2^2\right\} \qquad (14)$$

$$\boldsymbol{z}_u^{k+1} = \underset{\boldsymbol{z}_u}{\mathrm{argmin}}\left\{\lambda_u \sum_j^{N_u} \psi\left(\left\|\begin{pmatrix}\boldsymbol{z}_u \\ \alpha_v^k \boldsymbol{\Phi}_v \boldsymbol{z}_v^k\end{pmatrix}_j\right\|_2\right)\right.$$
$$\left. + \frac{\rho_u}{2}\|\boldsymbol{\nabla}\boldsymbol{u}^{k+1} - \boldsymbol{z}_u + \boldsymbol{\gamma}_u^k/\rho_u\|_2^2\right\} \qquad (15)$$

$$\boldsymbol{z}_v^{k+1} = \underset{\boldsymbol{z}_v}{\mathrm{argmin}}\left\{\lambda_v \sum_j^{N_v} \psi\left(\left\|\begin{pmatrix}\alpha_u^k \boldsymbol{\Phi}_u \boldsymbol{z}_u^k \\ \boldsymbol{z}_v\end{pmatrix}_j\right\|_2\right)\right.$$
$$\left. + \frac{\rho_v}{2}\|\boldsymbol{\nabla}\boldsymbol{v}^{k+1} - \boldsymbol{z}_v + \boldsymbol{\gamma}_v^k/\rho_v\|_2^2\right\} \qquad (16)$$

$$\boldsymbol{\gamma}_u^{k+1} = \boldsymbol{\gamma}_u^k + \rho_u(\boldsymbol{\nabla}\boldsymbol{u}^{k+1} - \boldsymbol{z}_u^{k+1}) \qquad (17)$$
$$\boldsymbol{\gamma}_v^{k+1} = \boldsymbol{\gamma}_v^k + \rho_v(\boldsymbol{\nabla}\boldsymbol{v}^{k+1} - \boldsymbol{z}_v^{k+1}) \qquad (18)$$

In sub-problem (15), the MR image gradients $\boldsymbol{z}_v^k$, evaluated at the kth iteration, are spatially registered and sampled to the resolution and FOV of the PET images by $\boldsymbol{\Phi}_v: \mathbb{R}^{3N_v \times N_v} \to \mathbb{R}^{3N_u \times N_u}$ and scaled by α_v^k to the magnitude of the PET gradients in order to cope with the resolution and signal intensity differences between the PET and MR images. Similarly, in sub-problem (16), the PET image gradients $\boldsymbol{z}_u^k$, evaluated at the kth iteration, are mapped and scaled by $\boldsymbol{\Phi}_u: \mathbb{R}^{3N_u \times N_u} \to \mathbb{R}^{3N_v \times N_v}$ and α_u^k so as to correspond to the MR images' resolution/FOV and magnitude. As a result, in (15-16) PET(MR) image gradients are estimated at a corresponding magnitude to the

matched MR(PET) gradients. In this study, we employed an alternating scaling of the distributions of the image gradients using the parameters α_v^k, α_u^k defined by:

$$\alpha_u^k = \frac{\|\mathbf{z}_v^k\|_F}{\|\mathbf{\Phi}_u \mathbf{z}_u^k\|_F}, \qquad \alpha_v^k = \frac{\|\mathbf{z}_u^k\|_F}{\|\mathbf{\Phi}_v \mathbf{z}_v^k\|_F} \qquad (19)$$

where $\|\mathbf{x}\|_F = \left(\sum_j \|\mathbf{x}_j\|_2^2\right)^{1/2}$ is the Frobenius norm or L$_{2,2}$ norm. By this definition, these scaling factors are global and spatially invariant (see the *Discussion* section for alternative approaches to finding these scaling factors). Note that in (16) the MR image gradients were set to be updated using the PET image gradients estimated at the previous iteration ($\mathbf{z}_u^k$) instead of the current iteration ($\mathbf{z}_u^{k+1}$). Therefore, any prioritization of the variable updates is avoided, providing a more balanced regularization. Moreover, as the PET and MR images have different intensities, we consider two different regularization parameters λ_u and λ_v in Eqs. (15) and (16).

1) Minimization w.r.t. $\mathbf{u}$

The minimization in (13) corresponds to a MAP reconstruction of a PET image at the kth iteration using a quadratic regularization. The solution to this problem can be iteratively obtained using a MAP expectation maximization (EM) approach, such as Green's one-step-late (OSL) approximation, as follows:

$$\mathbf{u}^{n+1} = \frac{\mathbf{u}^n}{P^T \mathbf{1} + \rho_u \nabla^T (\nabla \mathbf{u}^n - \mathbf{z}_u^k + \gamma_u^k / \rho_u)} P^T \left(\frac{\mathbf{y}}{P \mathbf{u}^n + \bar{r}_i}\right) \qquad (20)$$

where $\mathbf{1} \in \mathbb{R}^{M_u}$ and $\mathbf{u}^n$ is initialized by the PET image estimate at the kth global iteration. After n iterations of the MAP-OSL EM algorithm, the $(k+1)$th global iteration of the PET image, $\mathbf{u}^{k+1}$, is obtained from $\mathbf{u}^{n+1}$. It should be noted that the OSL approach does not have guaranteed convergence, especially when using a large regularization parameter [25]. An alternative approach to OSL is to use De Pierro's convexity lemma [26] to define a separable surrogate for the quadratic regularizer in (19), which can guarantee the convergence to the minimizer of the problem. As will be discussed later, for practical implementation we perform only a few iterations of the MAP-OSL EM method at each global iteration k. The Siddon algorithm was utilized for the calculation of the geometric components of the system matrix P, while an image-space PSF was used to account for the blurring components of the model.

2) Minimization w.r.t. $\mathbf{v}$

The minimization of the problem (14) is relatively straightforward as it corresponds to a regularized weighted least squares problem with a quadratic regularization. The solution is achieved by taking the derivative of the objective of the problem with respect to the vector of parameters to estimate, and equating it to zero. This approach results in the following normal equations:

$$(E^H W E + \rho_v \nabla^T \nabla) \mathbf{v}^{k+1} = E^H W \mathbf{s} + \nabla^T (\rho_v \mathbf{z}_v^k - \gamma_v^k) \qquad (21)$$

In general, for arbitrary k-space sampling patterns, the solution of the above SENSE MR image reconstruction problem can be iteratively obtained using the conjugate-gradient (CG) algorithm that has been extensively used for iterative MR image reconstruction [22].

3) Minimization w.r.t. $\mathbf{z}_u$ *and* $\mathbf{z}_v$

The proposed joint prior in (7) is non-convex for the potential function in (8) when $\sigma > 0$. Therefore, the problems (15-17) will possibly have multiple local minima, and their solution will highly depend on the choice of the optimization algorithm and initialization. These problems can be optimized using the optimization transfer technique in which a convex surrogate is defined at each step such that its minimization will guarantee progression towards the minimization of the original problem [27]. In this study, we considered the linear local approximation (linearization) of the potential function $\psi_\lambda(t)$ in (8) near the point t^k as follows:

$$L(t, t^k) = \psi(t^k) + \psi'(t^k)(t - t^k) = \psi'(t^k)t + C \qquad (22)$$

where $\psi'(t^k) = e^{-\sigma t^k}$ is the first-derivative of the $\psi(t^k)$ and C is a constant term independent of t. By substituting ψ with its linear surrogate L in (15), this problem is redefined as:

$$\mathbf{z}_u^{k+1} = \underset{\mathbf{z}_u}{\operatorname{argmin}} \left\{ \lambda_u \sum_j^{N_u} \psi' \left(\left\| \left(\frac{\mathbf{z}_u^k}{\alpha_v^k \mathbf{\Phi}_v \mathbf{z}_v^k}\right)_j \right\|_2 \right) \left\| \left(\frac{\mathbf{z}_u}{\alpha_u^k \mathbf{\Phi}_v \mathbf{z}_v^k}\right)_j \right\|_2 \right.$$
$$\left. + \frac{\rho_u}{2} \left\| \nabla \mathbf{u}^{k+1} - \mathbf{z}_u + \frac{\gamma_u^k}{\rho_u} \right\|_2^2 \right\} \qquad (23)$$

The above problem is solvable with respect to each jth component of $\mathbf{z}_u$, therefore by equating the derivative of its objective function to zero, it can be shown that the solution is given by the following vectorial weighted soft-thresholding estimator [28]:

$$\mathbf{z}_{uj}^{k+1,(d)} = \max\left(0, \left\| \left(\frac{\tilde{\mathbf{z}}_u}{\alpha_v^k \mathbf{\Phi}_v \mathbf{z}_v^k}\right)_j \right\|_2 - \frac{\lambda_u}{\rho_u} \omega_j^k \right) \frac{\tilde{\mathbf{z}}_{uj}^{(d)}}{\left\| \left(\frac{\tilde{\mathbf{z}}_u}{\alpha_v^k \mathbf{\Phi}_v \mathbf{z}_v^k}\right)_j \right\|_2},$$

$$\tilde{\mathbf{z}}_u = \nabla \mathbf{u}^{k+1} + \frac{\gamma_u^k}{\rho_u}, \qquad (24)$$

$$\omega_j^k = exp\left(-\sigma \left\| \left(\frac{\mathbf{z}_u^k}{\alpha_v^k \mathbf{\Phi}_v \mathbf{z}_v^k}\right)_j \right\|_2\right)$$

where $d = \{1,2,3\}$ and $\boldsymbol{\omega}^k \in \mathbb{R}^N$ is the vector of joint weights that are calculated from the gradients of both the PET and MR image. A similar solution can be obtained for problem (16). The σ parameter defined in ω_j^k in (24) should be properly selected with respect to the magnitude of the gradient vectors. Hence, this parameter was iteratively normalized by the Frobenius norm of the PET-MR gradients defined in ω_j^k. For brevity, we denote the above soft-thresholding rule for both problems (15) and (16) as follows:

$$\mathbf{z}_u^{k+1} = ST_{\rho_u}(\tilde{\mathbf{z}}_u, \mathbf{z}_u^k, \alpha_v^k \mathbf{\Phi}_v \mathbf{z}_v^k,) \qquad (25)$$
$$\mathbf{z}_v^{k+1} = ST_{\rho_v}(\tilde{\mathbf{z}}_v, \mathbf{z}_v^k, \alpha_u^k \mathbf{\Phi}_u \mathbf{z}_u^k,) \qquad (26)$$

The proposed joint reconstruction algorithm is summarized in Algorithm 1.

ALGORITHM 1: SYNERGISTIC PET-MR RECONSTRUCTION USING THE ADMM ALGORITHM

- Choose the parameters $\rho_u, \rho_v, \lambda_u, \lambda_v, \sigma$, and the PET and MR sub-problem iteration numbers n and m
- Initialize the variables $\mathbf{u}^0, \mathbf{v}^0, \gamma_u^0, \gamma_v^0$, set $k = 0$

While *a criterion is not met* **do**
1. Set $\mathbf{z}_u^k = \nabla \mathbf{u}^k$ and $\mathbf{z}_v^k = \nabla \mathbf{v}^k$
2. Estimate $\mathbf{u}^{k+1}$ using MAP-EM-OSL in (20) with n image updates.
3. Estimate $\mathbf{v}^{k+1}$ using MAP-CG-SENSE in (21) with m image updates.
4. Set the temporary variables $\tilde{\mathbf{z}}_u = \nabla \mathbf{u}^{k+1} + \gamma_u^k/\rho_u$, $\tilde{\mathbf{z}}_v = \nabla \mathbf{v}^{k+1} + \gamma_v^k/\rho_v$.
5. Set the temporary variables $\boldsymbol{\theta}_u = \mathbf{\Phi}_u \mathbf{z}_u^k$ and $\boldsymbol{\theta}_v = \mathbf{\Phi}_v \mathbf{z}_v^k$ by the mapping of PET(MR) gradients to the resolution and FOV of MR(PET) gradients.
6. Calculate α_u^k and α_v^k based on (19).
7. Perform the shrinkage rule $\mathbf{z}_u^{k+1} = ST_{\rho_u}(\tilde{\mathbf{z}}_u, \mathbf{z}_u^k, \alpha_v^k \boldsymbol{\theta}_v)$ in (25).
8. Perform the shrinkage rule $\mathbf{z}_v^{k+1} = ST_{\rho_v}(\tilde{\mathbf{z}}_v, \mathbf{z}_v^k, \alpha_u^k \boldsymbol{\theta}_u)$ in (26).
9. Update the Lagrange multipliers in (17) and (18).
Output: $\mathbf{u}^{k+1}, \mathbf{v}^{k+1}$

The convergence of the algorithm is assumed when the relative difference between successive iterates falls below a tolerance (τ), or the global number of iterations reaches a predefined maximum number N_{max}. The relative difference was defined as the normalized L$_2$ distance of the $\mathbf{u}, \mathbf{v}$ image pairs at follows:

$$\eta(k) = \frac{\left\| \binom{\mathbf{u}^{k+1}}{\mathbf{v}^{k+1}} - \binom{\mathbf{u}^k}{\mathbf{v}^k} \right\|_2}{\left\| \binom{\mathbf{u}^k}{\mathbf{v}^k} \right\|_2} \qquad (28)$$

In this study, we set $\tau = 1 \times 10^{-4}$ and $N_{max} = 400$ in all our experiments, elaborated in the following section. The number of iterations for MAP-EM PET and CG-SENSE MR image reconstructions were set to $n = 2$ and $m = 2$ respectively for our simulations as well as for
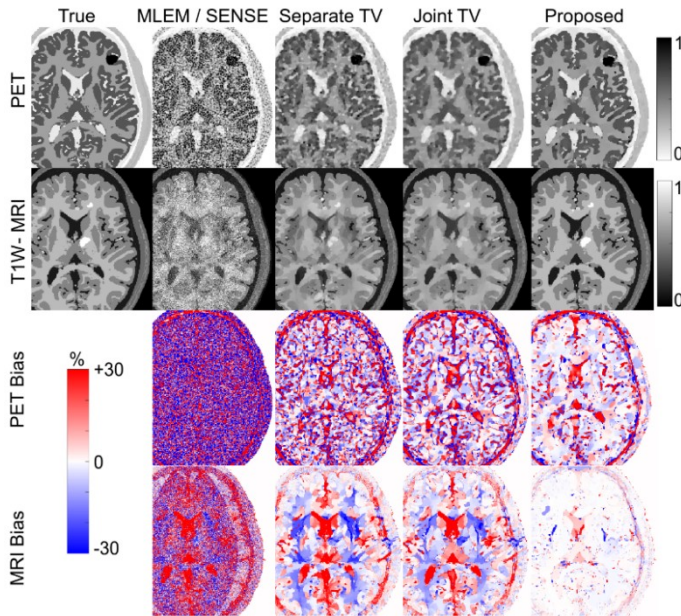
Fig 2. The reconstruction results for the Cartesian data acquisition of the T1-weighted MR image of the 2D brain phantom with acceleration factor of 8 and FDG PET image. The top two rows show the PET and MR images reconstructed without a prior (MLEM and SENSE respectively), with separate TV, joint TV and the proposed joint priors. The bottom two rows show the error maps of the reconstructed images.
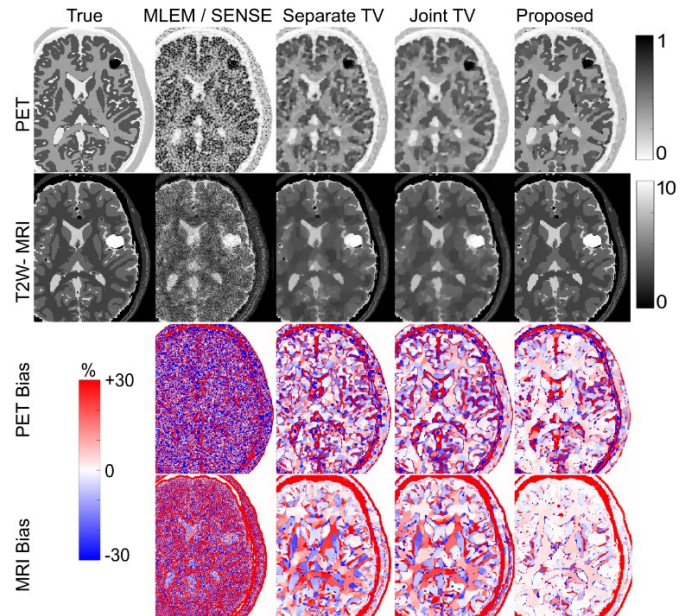

Fig 3. Similar to Fig. 2, except that the reconstruction results are for the multi-shot spiral data acquisition of the T2-weighted MR image of the brain phantom with 10 spiral interleaves.

the real data evaluations. The performance of the proposed algorithm was evaluated using simulation and clinical datasets. In our simulations, the joint reconstructed images were compared with the ground truth images and those reconstructed by the conventional unregularized PET and SENSE MR image reconstructions and those reconstructed with separated TV, joint TV and the linear PLS method proposed by Ehrhardt *et al* [7]. The PLS joint reconstructions were performed using the L-BFGS-B optimization algorithm, which jointly reconstructs the PET and MR images, while the other algorithms were reconstructed using the ADMM algorithm, which, as mentioned above, alternatingly reconstructs the PET and MR images.

D. Numerical simulations and real data

To objectively evaluate the algorithms in terms of image quality, bias-variance and convergence performance, three different simulations were performed for: i) a 2D brain phantom, ii) a 2D resolution phantom and iii) a 3D brain phantom.

1) 2D brain phantom

A 2D high-resolution brain phantom was derived from the BrainWeb database [29] with a matrix size of 512×512 and a pixel size of 0.53×0.53 mm^2. The phantom was used to simulate T1- and T2-weighted MR images and an [^{18}F]FDG PET image (see Figs. 2 and 3). In these phantoms, the ground truth PET and MR images share common edges with different magnitudes and orientations and contain several unique lesions. The PET acquisition was simulated for a 2D scanner with 180 azimuthal views, 729 radial bins and with a spatial resolution of 4 pixels full width at half maximum (FWHM). Twenty Poisson noise realizations with total counts of $\sim10\times10^6$ were generated. Scatter and randoms coincidences were ignored in these simulations. The MRI data were simulated for multi-coil, undersampled MRI data acquisition. An 8-channel circular head coil array was considered and the coil sensitivity maps simulated based on Biot–Savart's law [30]. A Cartesian trajectory with an undersampling factor (R) of 8 and multi-shot spiral k-space trajectories with 10 interleaves were considered for the T1 and T2 MR images. For all simulations, 27 dB noise was added to the k-space complex values, thereby simulating datasets with

a typical level of noise [30]. In the PET, T1 and T2 ground truth images, we also simulated modality unique lesions.

2) 2D resolution phantom

To evaluate our proposed joint reconstruction method against the PLS method, we used the PET-MR resolution phantom developed by Ehrhardt *et al* [7]. In this study, we considered the *Lines2* and *Radial20* benchmarks corresponding to the Cartesian sampling of MR k-space with every other phase encoding line ($R = 2$) and 20 radial lines (spokes). For the PET simulation, a total number of 1×10^6 counts was simulated for a 2D scanner with 300 azimuthal views and 128 radial bins. We only considered the linear PLS method as Ehrhardt *et al* had already concluded that this regularizer results in the lowest quantification errors compared to the quadratic PLS prior.

3) 3D brain phantom

3D realistic simulations were also performed for the native geometry of the Siemens Biograph mMR scanner (Siemens Healthcare, Erlangen), including all physical and data degradation processes during a PET scan (i.e. attenuation, normalization factors, randoms and scatter coincidences and PSF). The BrainWeb phantom was used to simulate an FDG activity distribution and a T1-weighted MR phantom (Fig. 6) both with the matrix sizes of $344\times344\times127$ and voxels of size $2.086\times2.086\times2.03$ mm^3. The normalization factors were obtained from a 5-hour ^{68}Ge phantom scan. The scatter sinograms were simulated by Gaussian smoothing of the mean prompts sinograms and rescaling them to simulate a scatter fraction of 50%. A randoms faction of 30% was also simulated using uniformly distributed Poisson noise (constant mean). The PET system's PSF was simulated using a spatially-invariant 4-mm FWHM Gaussian. For the MRI system, a 5-channel coil array was considered. A 100×10^6-count FDG scan was simulated for the PET scan and a Cartesian undersampling factor of R=6 was used for the undersampled MRI scan in the transverse phase encoding direction.

4) Real data

A brain PET-MR scan was acquired on the mMR scanner. The patient was injected with 214.7 MBq of [^{18}F]FDG and underwent a 30-minute PET scan about 60 minutes post-injection. The MR acquisition was performed on the 3T MRI subsystem of the scanner using the 5-channel head and neck coil array. For PET attenuation correction, a standard Dixon sequence and a UTE sequence was performed, which was

TABLE 1. NORMALIZED ROOT MEAN SQUARE DIFFERENCES (%) FOR THE RECONSTRUCTED PET AND MR IMAGES FOR THE 2D BRAIN, 2D RESOLUTION AND 3D BRAIN PHANTOMS WITH DIFFERENT k-SPACE SAMPLING PATTERNS. THE RESULTS FOR THE 2D BRAIN PHANTOM WITH CARTESIAN AND SPIRAL RECONSTRUCTION SETUPS CORRESPOND TO FIGS. 2 AND 3, RESPECTIVELY. THE RESULTS FOR THE 2D RESOLUTION PHANTOM WITH LINES2 AND RADIAL20 RECONSTRUCTION SETUPS CORRESPOND TO FIG. 6 AND SUPPLEMENTARY MATERIAL FIG. 2, RESPECTIVELY. THE RESULTS FOR THE 3D BRAIN PHANTOM WITH CARTESIAN SETUP CORRESPOND TO FIG. 7.

| METHODS | 2D BRAIN | | | | 2D RESOLUTION | | | | 3D BRAIN | |
| | CARTESIAN | | SPIRAL | | LINES2 | | RADIAL20 | | CARTESIAN | |
	PET	MRI	PET	MRI	PET	MRI	PET	MRI	PET	MRI
MLEM/SENSE	52.3	50.1	40.8	57.5	31.7	51.8	62.1	39.9	52.3	43.1
SEPARATE TV	22.2	12.7	23.4	12.6	---	---	---	---	38.8	18.3
JOINT TV	21.3	11.8	22.8	17.8	---	---	---	---	37.7	17.2
LINEAR PLS	---	---	---	---	11.0	19.9	11.8	5.7	---	---
PROPOSED	17.1	5.4	20.2	9.1	9.7	4.7	11.6	2.5	36.4	16.8

followed by a Cartesian 3D fully sampled T1-MPRAGE for anatomical and diagnostic imaging with the following parameters: repetition time (TR): 1700 ms, echo time (TE): 2.63 ms, inversion time (TI): 900 ms, number of averages (NEX): 1, flip angle: 9 degrees and acquisition time (TA) of 382 seconds. Data was retrospectively undersampled with acceleration factors of R=4 and R=3 in the transverse phase encoding direction to give 2 different datasets with different undersampling levels. In this study, we aimed to particularly preserve the native resolution of the MR images, therefore both the PET and the MR images were reconstructed with their default resolutions specified in the reconstruction console of the scanner. The MR images were reconstructed with a matrix size of 404×244×244, corresponding to the number of frequency-encoding steps and the transverse and axial phase-encoding steps. The voxel size was 1.05×1.05×1.1 mm³. The PET images were reconstructed with the same matrix sizes used in our 3D simulation. To match the spatial sampling of the gradient vectors of the PET and MR images during their soft-thresholding in Eqs.(25-26), the operator Φ was used to resample the gradients of PET to that of the MR images using B-spline interpolation and vice versa. For PET image reconstruction, all correction sinograms were generated using the VB e7 tools. For the real data, the PET images were reconstructed without PSF modelling. As the MRI data were fully sampled, the MRI coil sensitivity maps were estimated from each coil's MR image using the method described in [31].

E. Evaluation metrics

For the objective evaluation of the algorithms in our simulations, the reconstructed PET and MR images of a single noise realization were compared to their ground truth images based on the error maps (E) and normalized root mean square difference (NRMSD) defined here by:

$$E_j = 100 \times \frac{x_j - x_j^{True}}{x_j^{True}}$$
$$NRMSD = 100 \times \frac{\|x - x^{True}\|_2}{\|x^{True}\|_2} \quad (29)$$

where x is either a reconstructed PET or MR image, and x^{True} is its corresponding ground truth image. The error map was calculated for each voxel while the NRMSD was obtained over the entire image. The bias-variance performance of the studied algorithms was evaluated for twenty noise realizations ($N_r = 20$) using the 2D T1-weighted MR and FDG PET phantom, shown in Fig. 2. A region-of-interest (ROI) based approach was followed to assess the bias vs. variance trade-off for the studied algorithms. Three ROIs were considered: white matter (WM), grey matter (GM) and tumours for both the PET and the MR images. The regionally-averaged absolute voxel-level bias in each ROI was calculated as:

$$Bias = 100 \times \frac{1}{N_{ROI}} \sum_{j \in ROI} \frac{|\bar{x}_j - x_j^{True}|}{x_j^{True}} \quad (30)$$

where $\bar{x}_j = \frac{1}{N_r} \sum_{r=1}^{N_r} x_j^r$ represents the ensemble mean value of each voxel calculated for all N_r noise realizations. N_{ROI} is the total number

of voxels in the ROI. The variance was calculated using the average of the pixel-level coefficient of variation (COV) for each ROI:

$$COV = 100 \times \frac{1}{N_{ROI}} \sum_{j \in ROI} \frac{\sqrt{\frac{1}{N_r - 1} \sum_{r=1}^{N_r} (x_j^r - \bar{x}_j)^2}}{\bar{x}_j} \quad (31)$$

For the clinical dataset, for which the ground truth is unknown, the performance of the proposed algorithm was subjectively compared to TV regularization and also for the case of the image reconstructed without a prior. In our evaluations, the reconstructed MR images were compared with those reconstructed from fully sampled *k*-spaces, whereas the reconstructed PET images were compared with the MLEM reconstructed images followed by post-reconstruction convolution with a 3-mm FWHM Gaussian.

III. RESULTS

A. 2D Simulation studies

Fig. 2 shows the PET and T1-weighted MR images of the 2D brain phantom simulated for the Cartesian MR data acquisition. The images were reconstructed using MLEM, SENSE-CG, separate TV, joint TV and the proposed prior. In this simulation, the ground truth PET and MR images were assumed to have the same range of intensities in [0,1]. The results reveal that the MLEM and SENSE-CG images substantially suffer from noise and/or artifacts leading to high quantification errors shown via the error maps. Such noisy and undesirable solutions can be attributed to the ill-conditioning of the image reconstruction problem in PET and the ill-posed nature of undersampled MRI. The reconstructions with the TV and joint TV priors clearly reduce the artifacts. There are the familiar stair-casing and residual wraparound artifacts in the PET and MR images respectively, which are notably reduced by the joint reconstruction of the images. As can be seen, the proposed joint prior outperforms its counterparts in terms of reducing the noise and artifacts and the recovery of the common boundaries and unique features without transferring artifacts or unique features between the two modalities. The algorithms were further evaluated based on their NRMSD performance. Table 1 summarizes the results for all simulations and phantoms used in this study. For the 2D brain phantom with *Cartesian k*-space undersampling, the proposed algorithm achieves the lowest NRMSDs for both the PET and the MR images. Note that in this study, we compared the proposed algorithm with the linear PLS algorithm only in the reconstruction of the resolution phantom (a published benchmark).

Fig. 3 shows the reconstruction results for the simulations in the brain phantom using the T2-weighted MR images and multi-shot spiral MR data acquisition. In this simulation, we considered the T2W MR image with an intensity range of 0 to 10, while the PET image intensity was in [0, 1]. As in Fig. 2, the reconstruction results are shown for the studied algorithms together with their error maps. Note that for the joint TV regularization we used the same scaling scheme as was used for the proposed prior based on Eq. (19). The results show that the PET and MR images reconstructed without regularization suffer from noise and/or artifacts. The separate TV and joint TV regularizations

TABLE 2. THE HYPER-PARAMETERS USED FOR THE RECONSTRUCTION OF ALL SIMULATION AND CLINICAL DATASETS OF THIS STUDY.

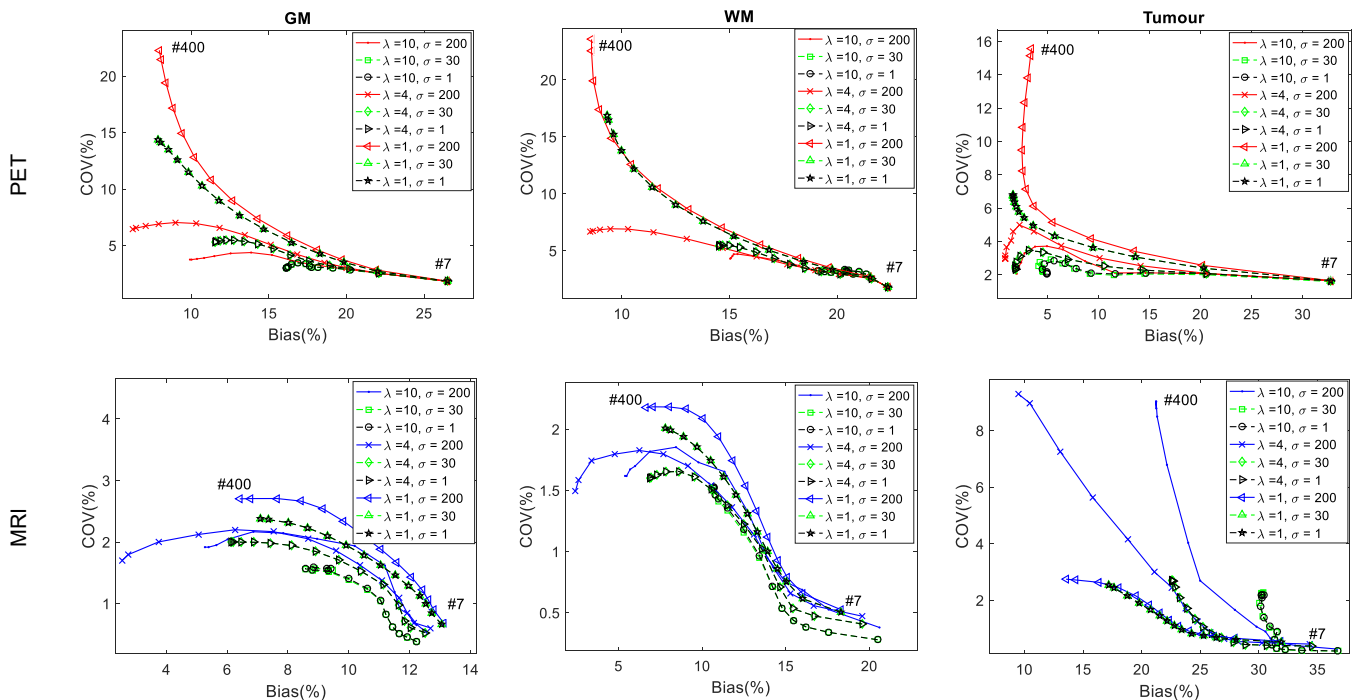| METHODS	PARAMETERS	2D BRAIN: CARTESIAN	2D BRAIN: SPIRAL	2D RESOLUTION: LINE 2	2D RESOLUTION: RADIAL20	3D BRAIN:	T1/FDG R=4	T1/FDG R=3
PROPOSED	σ	200	600	200	300	300	200	500
	λ_u	4	3	5	5	120	15	15
	λ_v	4	3	5	5	10	180	300
	ρ_u	0.07	0.035	0.05	0.01	0.01	0.5	0.5
	ρ_v	0.008	0.05	1.1	1	0.01	0.1	0.5
JOINT TV	λ_v	4	3	----	----	120	15	15
	λ_u	4	3	----	----	15	180	300
	ρ_u	0.05	0.035	----	----	0.01	0.5	0.5
	ρ_v	0.008	0.03	----	----	0.01	0.1	0.5
SEPARATE TV-MRI	λ_v	4	3	----	----	90	180	300
	ρ_v	0.01	0.035	----	----	0.01	0.1	0.5
SEPARATE TV-PET	λ_u	4	3	----	----	20	15	15
	ρ_u	0.05	0.04	----	----	0.1	0.5	0.5

Fig 4. Bias-variance analysis of the proposed algorithm for the 2D Cartesian MR-PET dataset for different combinations of λ and σ values in the grey matter (GM), white matter (WM) and the tumours. Note that for these results λ_u and λ_v were set equal to λ. The numbers on the graphs indicate the iteration number. The parameters $\lambda = 4$ and $\sigma = 200$ resulted in best-case performance of the algorithm for this dataset.

substantially reduced the noise and artifacts and partially recovered the sharpness and visibility of the common boundaries. Nonetheless, the joint TV method slightly outperformed the TV method in the PET image reconstruction, yet performed almost equivalently to the TV method in the reconstruction of the MR images, as can be observed in the error map. The visual inspection of the results shows that the artifacts and noise are efficiently reduced by the proposed method and that both the PET and MR images retain the details on the common boundaries. The error maps and NMRSD results in Table 1 also show that this method achieves the lowest error. Close inspection of the images reconstructed using the proposed prior shows that in the lesion area where the PET and MRI data do not share anatomical similarities, the temporal gyri have not been well recovered in the PET image compared to other regions for which the two modalities have common features. Moreover, it is noticeable that the MR-specific lesion has not been propagated into the PET image as a cross-talk artifact. In our experiments, we heuristically optimized the required regularization parameters, i.e. $\rho_u, \rho_v, \lambda_u, \lambda_v$ and σ. The regularization parameters were

initially optimized for the separate TV regularizations, in order to find the range of parameters that can subjectively result in acceptable reconstructions, then for joint TV and NCX regularizations, the same range of parameters with the proper adjustment was used. Table 2 summarizes the selected parameters for all the results presented in this work.

Fig. 4 presents the bias-variance performance of the proposed algorithm for reconstructing the Cartesian MR-PET dataset for 20 noise realizations for different λ and σ parameters. The results are presented for ROIs defined on whole regions of grey matter, white matter and tumours in the PET and MR images. For this dataset, the combination of $\lambda_u = 4$ and $\lambda_v = 4$ and $\sigma = 200$ results in the best bias-variance trade-off (as reported also in Table 2). As with other MAP reconstructions, for a given σ, decreasing the regularization parameters, λ_u and λ_v, will generally increase both noise (COV) and bias, and the performance of the regularized reconstruction will approach conventional unregularized methods.
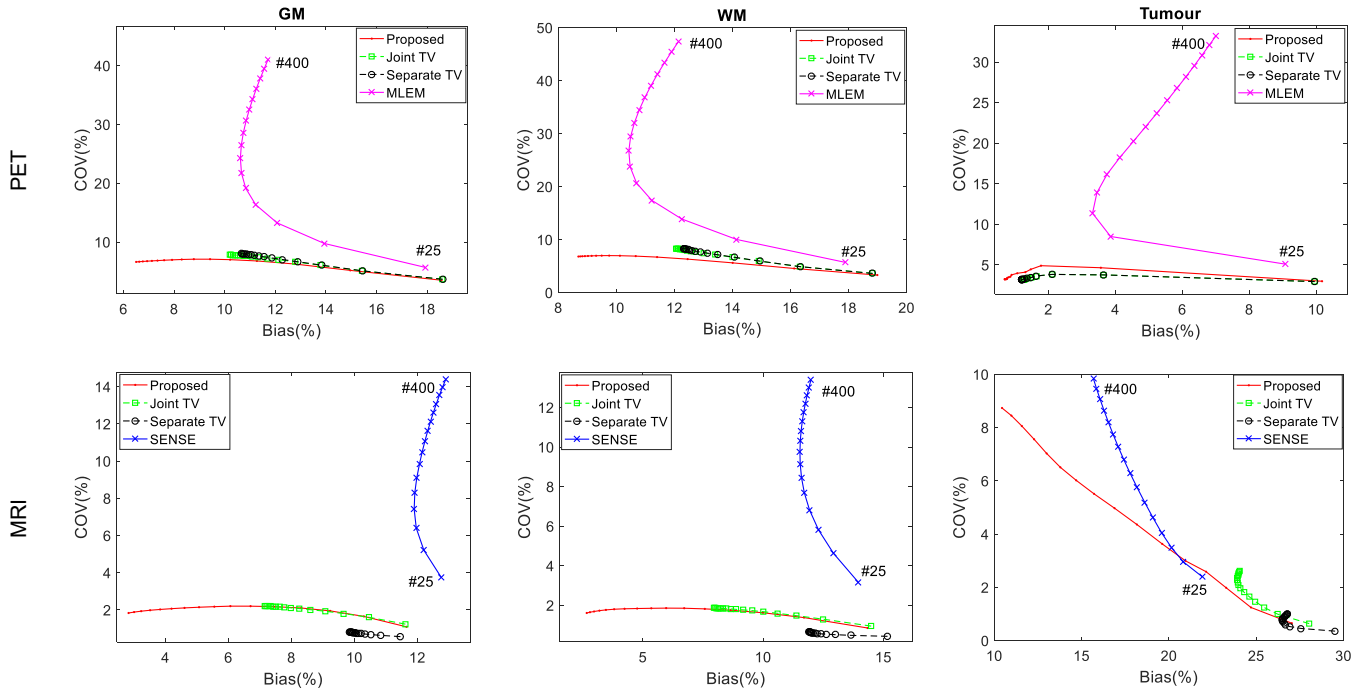
Fig 5. Bias-variance analysis of the algorithms for the Cartesian MR-PET dataset of the 2D brain phantom with T1-weighted MRI. The evaluations of PET and MRI data are shown for the grey matter (GM), white matter (WM) and the tumours. The same regularization parameters as used in Fig. 2 and reported in Table 2 were used. The different points on the graphs are obtained from different iteration numbers (low iterations corresponding to higher bias).

In Fig. 5, the bias-variance performance of all algorithms in the image reconstruction of the Cartesian MR-PET dataset is shown for the same 20 noise realizations. The results of the proposed prior is the same as those shown in Fig. 4 with $\lambda = 4$ and $\sigma = 200$. As could be expected, the MLEM and SENSE reconstructions present the worst bias-variance trade-off, while the regularized reconstructions show substantial improvement. It is noticeable that the TV and joint TV priors performed almost similarly in the reconstruction of PET images of the dataset. Whereas in the reconstruction of MR images, the joint TV outperformed the separate TV reconstructions. The results also show that the proposed prior considerably outperforms its counterparts by achieving lower bias and variance in grey and white matter. For the ROIs defined on the PET tumours, the proposed prior performs comparably to the TV and joint TV reconstruction while it outperforms the methods in terms of bias reduction in the regions of the MRI tumours. For all noise realizations, the same regularization parameters were used. Comparison of the results shows that for the reconstructions with $\lambda = 4$, as σ is reduced, the performance of the proposed prior approaches that of the joint TV prior (c.f the joint TV's graph in Fig. 5 and the graph of $\lambda = 4, \sigma = 1$ in Fig. 4).

In this study, the 2D brain simulations were designed to demonstrate the best case of joint PET-MR reconstruction using an ideal high-resolution PET-MR scanner. In the supplementary materials Fig. 2, the reconstruction results of the Cartesian MR-PET data for different PSF kernel widths are given, while using the same reconstruction and regularization parameters. As shown, as the PET scanner's resolution is degraded the quality of the jointly reconstructed images is similarly reduced, especially for the PET images. In the next sections, we therefore evaluate the joint reconstructions using more realistic 3D simulations and the clinical datasets to show their performance in practical situations with current PET-MR scanners.

The proposed joint prior was also compared with the linear PLS prior. Fig. 6 compares the PET and MR images of the *Lines2* setup, reconstructed by the conventional MLEM and zero-filling followed by inverse Fourier transform, and those jointly reconstructed by the linear PLS and the proposed algorithms.
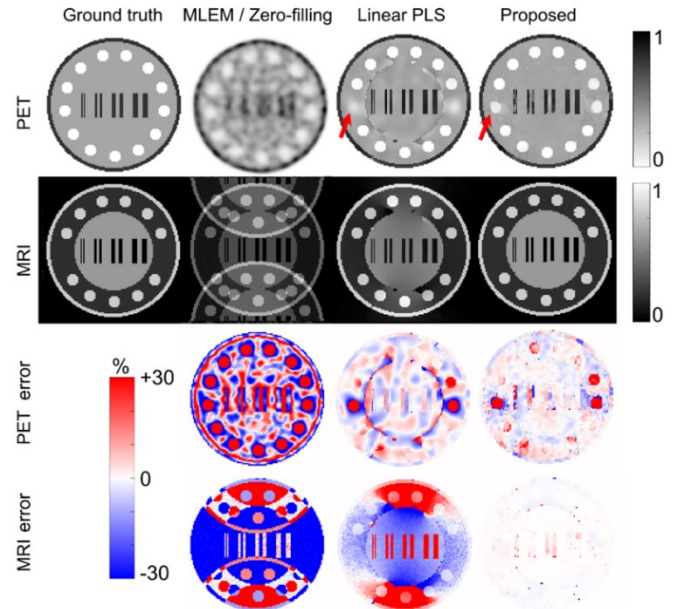

Fig 6. Comparison of the reconstruction results of the linear PLS and the proposed algorithms for the *Lines2* dataset of the resolution phantom [7]. Top: the reconstructed PET and MR images compared with their ground truth images. Bottom: the error maps calculated for each reconstruction in comparison with the ground truth images.

The quantitative performance of the algorithms has been shown using error maps and NRMSD errors in Table 1. Both joint prior algorithms considerably reduce the noise and aliasing artifacts, which leads to their improved quantitative performance. The PLS algorithm however introduces new artifacts in the PET images by transferring the unique features of the MR images (the central disk) into the PET ones. Despite this, the algorithm perfectly reconstructed the shared features such as the vertical bars and spheres, but it suppressed the spheres that are not
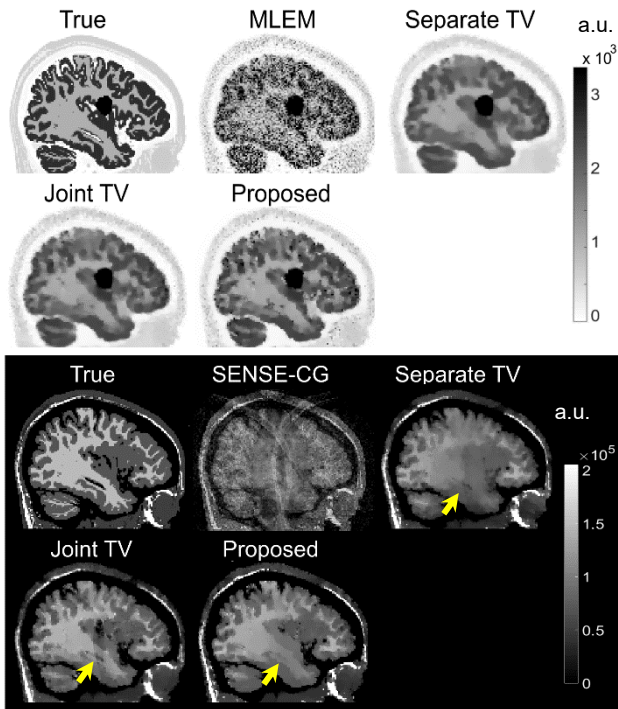
Fig 7. The results of separate and synergistic reconstructions for the 3D brain phantom simulated for the mMR scanner. Top: the reconstructed PET image compared to the ground truth, bottom: the reconstructed MR image compared to the ground truth.

in common between the PET and MR images. In comparison, the proposed algorithm reconstructed the images with almost no cross-talk artifacts. It is noticeable that this prior restores the two PET unique spheres more accurately, since in such regions it reduces to a single-modality prior, relying only on PET or only on MRI data. The results of the *Radial20* experiments, shown in the supplementary material Fig. 3 and Table 1, also demonstrates the outperformance of this prior. However, the performance of the proposed prior depends on the shape parameter σ. In Fig. 4 of the supplementary material, this dependency has been demonstrated for the *Line2* setup of the resolution phantom. For this experiment other parameters were kept fixed to their values presented in Table 2. The figure shows the PET and MR images and the MR weighting coefficients, ω_j^k, defined in (24). With increasing σ, the MR residual artifacts are reduced and at the same time the PET image details are recovered.

B. 3D Simulation studies

Fig. 7 compares the results in the 3D brain phantom for the PET and MR images reconstructed using the separate TV, joint TV and the proposed non-convex priors. As shown, in the ground truth images, the PET and MR images share several common boundaries with different contrasts and amplitudes. In this phantom, we simulated a large and active tumour in the PET data and a mismatched anatomical region in the MR image in the vicinity of the tumour. As shown, the regularized reconstruction notably reduces the noise in PET-MR images reconstructed by MLEM and SENSE-CG. The separate TV regularization seems to be very effective in reduction of noise and artifacts, particularly for the MR image reconstruction. However, as shown by the arrows, this algorithm cannot recover the indicated missing white matter in the MR image due to aliasing artifacts. On the other hand, the joint reconstruction algorithms attempt to utilize the corresponding PET boundary information to recover those missing parts of the white matter in the MR image. As also indicated by the arrows, the MR images reconstructed by the joint priors improve the overall quality of the separate TV MR image, showing the potential of synergistic PET-MR image reconstruction for highly undersampled MRI data acquisition.
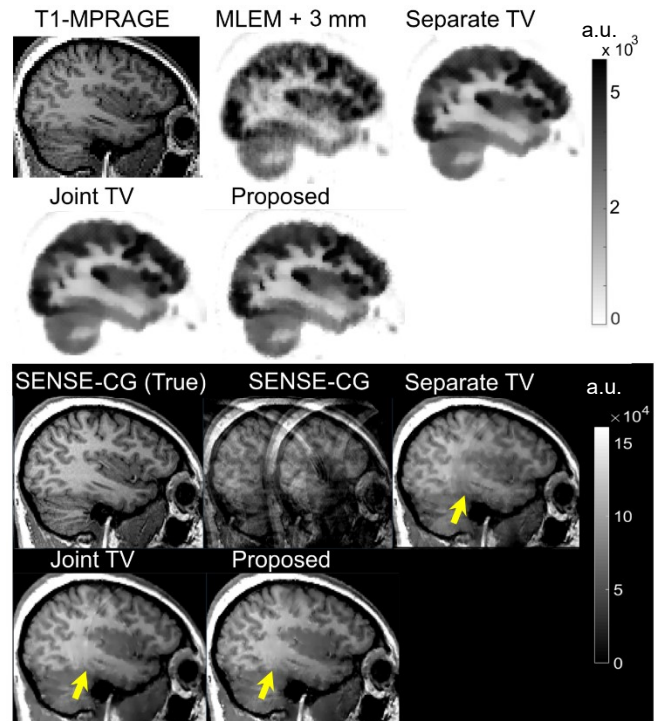

Fig 8. The results of the separate and synergistic reconstructions of the clinical [^{18}F]FDG PET and T1-MPRAGE MRI brain scan with MR acceleration of 4 in the phase encoding direction. Top: the T1-MR image resampled into the coordinate of PET images, the MLEM reconstruction followed by a 3-mm Gaussian post-reconstruction filtering, and the PET images reconstructed by separate and joint priors. Middle: the MR image reconstructed using full k-space sampling (as a ground truth), those reconstructed from undersampled data using the studied algorithm.
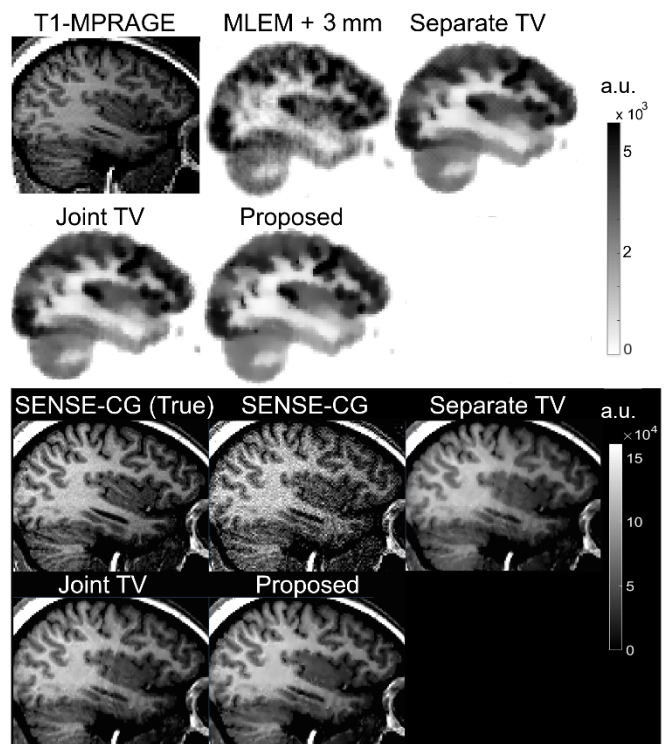

Fig 9. The same as in Fig. 8 but with an MR acceleration of 3 in the phase encoding direction.
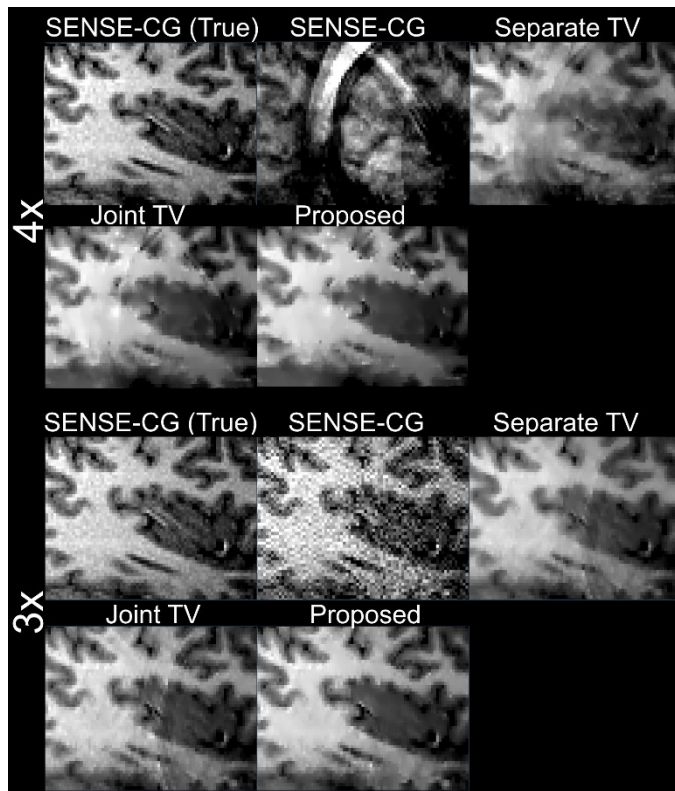
Fig 10. The zoomed-in views of the MR images shown in Fig. 8 (top) and Fig. 9 (bottom).

The joint priors are robust in preserving unique edges (because they become separate regularizations for those edges), however, they might induce artificial edges in the other modality by preserving noise and aliasing artifacts residing on the corresponding boundaries. Close inspection of the MR images reconstructed using the joint TV and proposed priors for the simulated PET tumour show that both priors have induced some faint pseudo-edges formed by noise. However, these artifacts can be acceptably reduced by the proper selection of the regularization parameters. For the proposed joint non-convex prior, we used the same regularization parameters optimized for joint TV reconstruction. This prior has an additional σ parameter that controls the degree of non-convexity. The comparison of the regularized PET images show that the proposed prior leads to increased edge preservation, however, at the expense of a slight increase in noise.

Table 1 also provides the NRMSD performance of the reconstructions, where the proposed prior achieves the lowest errors compared to the other methods. The supplementary material Fig. 5 compares the reconstructed image in the coronal plane with some representative activity profiles.

C. Evaluations with real data

Fig. 8 shows the results of the PET-MR image reconstruction of the clinical dataset with MR acceleration of R = 4. The PET and MR images were reconstructed in their native resolution and field of view, mainly to preserve the quality of the MR images. Therefore, in our alternating reconstruction approach, during PET image reconstruction-the MR image gradients are registered to the native coordinates of the PET and vice versa. On top of the figure, the T1-MPRAGE MR image registered to the PET coordinate system is shown together with the PET images reconstructed by the MLEM algorithm, followed by a post-reconstruction convolution with a 3-mm FWHM Gaussian, and the regularization algorithms. As can be seen, the PET image reconstructed by the proposed joint prior trends to slightly better preserve

functional boundaries while suppressing noise. In this dataset, the separate TV and joint TV methods perform almost equally, however, their impacts on the MR image reconstruction are notably different. We observed the same trend in the bias-variance analysis of the simulation results in Fig. 5. For the comparison of MR reconstructions, we reconstructed the T1-MPRAGE image using the full *k*-space sampling. To demonstrate the spatial appearance of aliasing artifacts in the conventional unregularized SENSE-CG reconstruction, we reconstructed the data with early termination of iterations (~5 iterations). With increased number of iterations, the aliasing artifacts are gradually reduced, however as the reconstruction of highly undersampled data corresponds to an ill-posed inverse problem, the noise is amplified and obscures the residual artifacts as seen in our simulation results. The aliasing artifacts are considerably reduced by the separate TV regularization, however, as indicated by the arrows the joint reconstructions resulted in further reduction of the artifacts and recovery of valid anatomical boundaries that are otherwise incompletely restored by the separate TV method. The results also show that the proposed method can slightly improve upon the joint TV results by removing the residual artifacts. The supplementary material Fig. 6 also compares the performance of the methods in a coronal plane.

Fig. 9 shows the same results with an MR acceleration factor of R=3. In this experiment, the result of the unregularized SENSE-CG reconstruction has been shown after 400 iterations, therefore the aliasing artifacts are reduced but noise has been amplified. As seen, the reconstruction algorithms give rise to MR images of the same overall quality as the fully sampled MR (true) image. In the PET reconstructions, the proposed prior results in improved enhancement of anato-functional boundaries.

Fig. 10 compares the zoomed-in MR images reconstructed by the studied algorithms with MR acceleration factors of 4 and 3. It is noticeable that in both experiments the proposed synergistic reconstruction method removes the residual artifacts to a greater extent than the other separate and synergistic reconstruction methods. However, it should be noted that in both experiments some detail cannot be recovered even by synergistic reconstruction, since the PET image is much lower resolution than the anatomical details.

IV. DISCUSSION

In this study, we aimed at developing a joint PET-MR image reconstruction framework for the practical and numerically feasible synergistic reconstruction of PET and MR images. For the joint priors considered here, the balanced formation of the joint norm of the PET and MR image gradients depends on the proper scaling of these gradients, otherwise the one with higher magnitude will dominate the resulting joint norm. The alternating scaling scheme proposed in Eq. (19) was defined as the ratio of the Frobenius norm of the magnitude of the PET image gradients to that for the MR image. Here, it is assumed that the magnitudes of the image gradients, especially over prominent boundaries, are approximately of the same order. As a result, image gradients at all voxels are scaled by a global spatially-invariant scaling factor. Fig. 7 in the supplementary material shows the scaling factors for the MR image gradients, α_v, as a function of iteration for the simulations presented in Figs. 2–3. In each case, the estimated scaling factors asymptotically approach the true scaling factors obtained from the ground truth images, which are in the range of 1 and 0.1 for the Cartesian and spiral MR simulation set-ups, respectively. However, in more realistic situations, the ratio of the magnitudes of the PET and MR image gradients is not necessarily the same for all image voxels and application of a single global factor to scale the gradients might not be effective for all voxels. On the other hand, a voxel-specific local scaling of the gradients would have a major disadvantage. Suppose that a given boundary in the PET image cannot be properly recovered during PET reconstruction, therefore the magnitude of the PET image gradient along the boundary remains small and close to zero. As a result of local scaling, the corresponding MR gradients will thus be scaled to those small values, making inefficient use of the MR image gradients

to recover those missing boundaries in PET. Another approach could be the calculation of local scaling factors from the PET and MR images reconstructed separately. The challenge of the proper scaling of the PET-MR gradients however leaves room for the application of information theoretic priors such as joint entropy and mutual information used in multi-modal image registration and anatomically-guided PET image reconstruction [32], which are in principle independent of the image intensities of the PET and MR images. Although, the selection of bandwidths for the Parzen windows for the proper estimation of the joint probability distribution of PET and MR image values is potentially dependent on the distribution and range of PET and MR image intensities.

The improved performance of proposed prior over the joint TV one can be attributed to the non-convex potential function used in the proposed prior in Eq. (8). As described in the section II, we made the resulting joint prior convex by iterative linearization of the potential function, which in fact resulted in a weighted soft-thresholding rule. As defined in Eq. (24), the weighting coefficients, ω_j^k, were jointly defined from the gradients of the PET and MR images. These coefficients were iteratively derived and aimed to recognize the valid features in PET and MR images. Fig. 8 in the supplementary material shows the joint weighting coefficients used in soft-thresholding of the PET gradients in the spiral simulation set-up shown in Fig. 3, as a function of iterations. With increasing iterations, the coefficients distinguish true edges from those arising from aliasing and streaking artifacts, since the norms of the gradients of noise and noise-like artifacts are usually of low-amplitude compared to valid edges.

In these comparisons with the PLS prior, we used the same dataset, optimization and regularization parameters provided in [7]. The proposed prior encourages the joint sparsity of the gradients and similarly to the PLS relies on the magnitude of the PET and MR image gradients. However, the PLS method depends on the PET and MR signal intensities, which is the main limitation of the method, therefore we only compared our proposed prior with the PLS prior using the resolution phantom in which the PET and MR images have the same range of intensities. Moreover, PLS is a non-convex prior and the authors used a Newton-type optimization algorithm which renders the performance of the algorithm highly dependent on the PET-MR initialization and selection of regularization parameters. In comparison, we defined a convex surrogate for the employed non-convex potential function. Furthermore, we used the ADMM algorithm that makes the optimization of the joint PET-MRI objective function tractable and almost independent of the initialization. It is worth mentioning that the ADMM algorithm has also been applied for the optimization of other joint objective functions such as joint PET activity and attenuation map estimation [33] as well as joint MR image reconstruction and coil sensitivity estimation [34].

Compared to the results presented in Knoll *et al* [19], where a generalized joint TV prior is used based on the nuclear norm of the gradients of the PET and MR images, the results presented here highlight the potential advantage of synergistic PET-MR image reconstruction, specifically for highly undersampled MR image reconstruction, where separate MR regularization fails to completely remove residual aliasing artifacts. In current clinical MRI scanners, the combination of parallel imaging and *k*-space undersampling is usually employed to reduce the scan time. Generally, for a given *k*-space undersampling factor, as the number of coils is increased, the aliasing artifacts are reduced. In [19], the author used $R = 4$ (as used in this study) however, a 12-element coil array was employed for the SENSE MR reconstruction, which substantially reduced the aliasing artifacts compared to our results, leaving less scope for the proper evaluation of the proposed joint prior. However, our 3D simulations and real data evaluations show that there are residual artifacts particularly in the MR images of the synergistic reconstructions. The performance of these algorithms is affected by several factors: i) the choice of regularization parameters (up to 5 parameters are required), ii) the optimization algorithm and the number of iterations (given that the PET and MR images converge

at different rates), iii) the reconstruction of the real PET-MR data in their native FOV and resolution, and application of registration and sampling to spatially match the common PET-MR features, and iv) the limitations arising from use of a global scaling factor to match the magnitudes of the PET-MR gradient vectors. Consequently, the resulting image quality benefits for the jointly reconstructed MR images can be limited, especially for MR acceleration factors higher than 4.

Another different aspect of our reconstruction framework compared to [19] is that we reconstructed the clinical PET and MRI datasets in their native resolution and FOV rather than in an intermediate resolution, or only that of MRI. In our alternating reconstruction, the gradients of the PET and MR images are alternatingly registered to each other using a one-off predefined transformation field obtained from the co-registration of separately reconstructed PET and MR images. The main advantage of this approach is that the possible movement and misregistration between PET and MR images can be addressed. It is even possible to include the estimation of transformation fields between PET and MR images within the synergistic reconstruction, obviating the need for separate reconstruction. In this study, we implemented the PET and MR image reconstructions in C++ and MATLAB R2015b, respectively, running on a 12-core Intel Xeon 3.5GHz workstation with 128 GB RAM. The total reconstruction time of the clinical dataset, which includes separate PET and MR image reconstructions, registration of the MR image gradients to those of PET and vice versa, and the soft-thresholding of those gradients, was close to 20 hours for about 400 SENSE-CG and 400 MLEM iterations. The most computationally intensive part of this pipeline is the PET image reconstruction, which can be substantially accelerated by GPU implementation of the forward and transpose PET operators.

In this proof-of-concept study, we heuristically selected the required hyper-parameters. As mentioned earlier, given the high computational expense of the synergistic reconstruction, in the evaluation of the proposed joint prior we considered using the same range of regularization parameters optimized for the joint TV regularization (i.e. $\rho_u, \rho_v, \lambda_u, \lambda_v$), in addition to one extra parameter to be optimized (i.e. σ). In our experience, the impact of the λ and σ parameters on the results were found to be more important than that of ρ. Therefore, the proper selection of these parameters is of paramount priority. As pointed out in [21], the ρ parameters mainly control the convergence rate of the ADMM algorithm and in principle do not affect the final solution. The introduction of two scaling factors, α, in (7) and application of two regularization parameters, λ, results in the joint prior having two energy states, one corresponding to the PET image when the MR image gradients are mapped and scaled to those of the PET and one corresponding to the MR image. Therefore, the MAP problem in (4) has two energy or likelihood states which will be minimized by the alternating minimization. However, it should be noted that the convergence properties of the resulting optimization algorithm, which uses both a convex surrogate and a rescaling step at every iteration, can potentially deviate from the known properties of the ADMM algorithm and therefore the algorithm's convergence to the joint global maximizer of the MAP problem might not be guaranteed.

More work will be required to further evaluate the proposed synergistic reconstruction for more clinical datasets and assessment of the impact on clinical imaging of these new methods for simultaneous PET-MR, especially in the case of mismatched functional and anatomical features. This is especially expected to be more of a concern for tracers other than FDG. In addition, the synergistic reconstruction of PET and multi-contrast MR images (i.e. T1-weighted and FLAIR) remains to be investigated. Finally, it should be noted that joint reconstruction in the context of PET-MR is an emerging yet challenging concept compared to, for example, multi-spectral and multi-contrast MR images. This is due to the substantial differences in the resolution and signal intensities of the two modalities. In particular, given the lower resolution of current PET scanners, one should not expect substantial improvement of the MR image quality using PET data, and so

the clinical benefit of these emerging reconstruction methods may be limited.

V. CONCLUSION

In this work, we proposed a novel non-convex joint sparsity regularization for synergistic PET and MR image reconstruction in order to preserve and encourage the formation of common boundaries between PET and MR images and therefore to improve the image quality beyond what can be obtained from separate regularization techniques. The synergistic reconstruction was achieved using the ADMM algorithm, resulting in the reconstruction of PET and MR images via well-established optimization algorithms. Our extensive simulations and evaluation on clinical data showed the proposed joint sparsity prior can considerably improve upon the currently existing joint and separate total variation regularizations and the PLS regularization. However, our clinical results showed that the clinical applicability of joint reconstruction might be limited in current PET-MR scanners, mainly due to the lower resolution of PET images.

ACKNOWLEDGEMENTS

REFERENCES

[1] C. Catana, A. Drzezga, W.-D. Heiss, and B. R. Rosen, "PET/MRI for Neurologic Applications," *Journal of Nuclear Medicine,* vol. 53, pp. 1916-1925, December 1, 2012 2012.

[2] T. E. Yankeelov, T. E. Peterson, R. G. Abramson, D. Garcia-Izquierdo, L. R. Arlinghaus, X. Li, *et al.*, "Simultaneous PET–MRI in oncology: a solution looking for a problem?," *Magnetic Resonance Imaging,* vol. 30, pp. 1342-1356, 11// 2012.

[3] J. Ouyang, Q. Li, and G. El Fakhri, "Magnetic resonance-based motion correction for positron emission tomography imaging," *Semin Nucl Med,* vol. 43, pp. 60-7, Jan 2013.

[4] B. Bai, Q. Li, and R. M. Leahy, "Magnetic resonance-guided positron emission tomography image reconstruction," *Semin Nucl Med,* vol. 43, pp. 30-44, Jan 2013.

[5] H. Sari, K. Erlandsson, K. Thielemans, D. Atkinson, S. Ourselin, S. Arridge, *et al.*, "Incorporation of MRI-AIF Information For Improved Kinetic Modelling of Dynamic PET Data," *IEEE Trans Nucl Sci,* vol. 62, pp. 612-618, 2015.

[6] G. Delso, F. Wiesinger, L. I. Sacolick, S. S. Kaushik, D. D. Shanbhag, M. Hullner, *et al.*, "Clinical evaluation of zero-echo-time MR imaging for the segmentation of the skull," *J Nucl Med,* vol. 56, pp. 417-22, Mar 2015.

[7] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, *et al.*, "Joint reconstruction of PET-MRI by exploiting structural similarity," *Inverse Problems,* vol. 31, Jan 2015.

[8] D. J. Larkman and R. G. Nunes, "Parallel magnetic resonance imaging," *Phys Med Biol,* vol. 52, pp. R15-55, Apr 7 2007.

[9] K. G. Hollingsworth, "Reducing acquisition time in clinical MRI by data undersampling and compressed sensing reconstruction," *Phys Med Biol,* vol. 60, pp. R297-322, Nov 7 2015.

[10] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magn Reson Med,* vol. 57, pp. 1086-98, Jun 2007.

[11] J. Trzasko and A. Manduca, "Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic L0-Minimization," *IEEE Tran Med Imaging,* vol. 28, pp. 106-121, 2009.

[12] K. M. Holt, "Total Nuclear Variation and Jacobian Extensions of Total Variation for Vector Fields," *IEEE Trans. Image Process.,* vol. 23, pp. 3975-3989, 2014.

[13] M. Moller, E. Brinkmann, M. Burger, and T. Seybold, "Color Bregman TV," *SIAM J. Sci. Comput.,* 2014.

[14] D. S. Rigie and P. J. La Riviere, "Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization," *Phys Med Biol,* vol. 60, pp. 1741-62, Mar 7 2015.

[15] A. Martin, I. Chatnuntawech, B. Bilgic, K. Setsompop, E. Adalsteinsson, and E. Schiavi, "Total generalized variation based joint multi-contrast parallel imaging

[16] S. Lefkimmiatis, A. Saucedo, S. Osher, and S. Kyunghyun, "Vectorial non-local total variation regularization for calibration-free parallel MRI reconstruction," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on,* 2015, pp. 1256-1259.

[17] I. Y. Chun, B. Adcock, and T. M. Talavage, "Efficient Compressed Sensing SENSE pMRI Reconstruction With Joint Sparsity Promotion," *IEEE Trans Med Imaging,* vol. 35, pp. 354-68, Jan 2016.

[18] J. Rasch, E.-M. Brinkmann, F. Knoll, T. Koesters, F. Wuebbeling, and M. Burger, "PET-MRI Joint Reconstruction via Multichannel Bregman-TV," in *IEEE NSS/MIC conference record,* 2015.

[19] F. Knoll, M. Holler, T. Koesters, R. Otazo, K. Bredies, and D. Sodickson, "Joint MR-PET reconstruction using a multi-channel image regularizer," *IEEE Tran Med Imaging,* vol. PP, pp. 1-1, 2016.

[20] A. J. Reader, S. Ally, F. Bakatselos , R. Manavaki, R. J. Walledge, A. P. Jeavons, *et al.*, "One-pass list-mode EM algorithm for high-resolution 3-D PET image reconstruction into large arrays " *IEEE Trans. Nucl. Sci.,* vol. 49, pp. 693-9, 2002.

[21] S. Ramani and J. A. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Trans Med Imaging,* vol. 30, pp. 694-706, Mar 2011.

[22] K. P. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magn Reson Med,* vol. 46, pp. 638-51, Oct 2001.

[23] P. Blomgren and T. F. Chan, "Color TV: Total variation methods for restoration of vector-valued images," *IEEE Trans. Image Process.,* vol. 7, pp. 304-309, 1998.

[24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning,* vol. 3, pp. 1-122, 2011.

[25] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans Med Imaging,* vol. 9, pp. 439-46, 1990.

[26] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans Med Imaging,* vol. 14, pp. 132-7, 1995.

[27] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *J. Comput. Graph. Stat.,* vol. 9, pp. 1–20, 2000.

[28] C. Li, "An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing," Master's Thesis, Rice University, 2010, http://hdl.handle.net/1911/62229.

[29] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, *et al.*, "Design and construction of a realistic digital brain phantom," *IEEE Trans Med Imaging,* vol. 17, pp. 463-8, Jun 1998.

[30] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser, "Realistic analytical phantoms for parallel magnetic resonance imaging," *IEEE Trans Med Imaging,* vol. 31, pp. 626-36, Mar 2012.

[31] D. O. Walsh, A. F. Gmitro, and M. W. Marcellin, "Adaptive reconstruction of phased array MR imagery," *Magn Reson Med,* vol. 43, pp. 682-90, May 2000.

[32] S. Somayajula, C. Panagiotou, A. Rangarajan, Q. Li, S. R. Arridge, and R. M. Leahy, "PET image reconstruction using information theoretic anatomical priors," *IEEE Trans Med Imaging,* vol. 30, pp. 537-49, Mar 2011.

[33] Y. Hsu and C. Y. Chou, "Alternating-direction method of multipliers estimation of attenuation and activity distributions in time-of-flight flat-panel positron emission tomography," *EJNMMI Phys,* vol. 2, p. A40, Dec 2015.

[34] M. Benning, F. Knoll, C. Schönlieb, and T. Valkonen, "Preconditioned ADMM with nonlinear operator constraint," presented at the IFIP Systems Modelling and Optimization, 2016.