Learning with fixed point condition for convergent PnP PET reconstruction

M. Savanier, Claude Comtat, Florent Sureau

LEARNING WITH FIXED POINT CONDITION FOR CONVERGENT PNP PET RECONSTRUCTION

Marion Savanier[⋆]*, Claude Comtat*[⋆]*, Florent Sureau*[⋆]

[⋆] BioMaps, Université Paris-Saclay, CEA, CNRS, Inserm, SHFJ, 91401 Orsay, France

ABSTRACT

This work explores plug-and-play algorithms for PET reconstruction, combining deep learning with model-based variational methods. We aim to integrate classical convolutional architectures into algorithms such as ADMM and Forward-Backward (FB) while ensuring convergence and maintaining fixed-point control. We focus on the scenario where only high- and low-count reconstructed PET images are available for training our networks. Experimental results demonstrate that the proposed methods consistently reach fixed points with high likelihood and low mean squared error, thus showcasing the potential of convergent plug-and-play techniques for PET reconstruction.

***Index Terms*—** PET reconstruction, plug-and-play, ADMM, forward-backward, convergence, fixed point

1. INTRODUCTION

Positron Emission Tomography (PET) imaging plays a key role in oncology, neurology, and pharmacology: it allows molecular activity within tissues to be quantified by administering specific radioactive tracers to a patient. Reconstruction of the activity concentration map from PET measurements is an ill-posed inverse problem with Poisson data. It is usually solved using variational methods, minimizing a cost function that combines a data fidelity term and a regularization term that explicitly incorporates image priors whose choice has been a subject of active research [1, 2].

In recent years, deep learning approaches have been introduced for PET reconstruction, primarily to predict high-count quality images from low-count - or low-dose - inputs using convolutional neural networks [3]. However, these approaches have shown instability and limited generalization [4], a concern also observed in other medical imaging modalities [5, 6]. To address these issues, new approaches combine deep learning with robust minimization algorithms derived from classical optimization. These include networks that loosely simulate a fixed number of iterations of a reconstruction algorithm and Plug-and-Play (PnP) algorithms. The latter capitalizes on denoising networks, implicitly serving as image priors to replace the operator related to the regularization (a proximity operator or a gradient[1]) in minimization algorithms. Given that PET reconstruction often requires numerous iterations for convergence, PnP methods emerge as scalable solutions.

Integrating a neural network into an iterative scheme raises questions about convergence and fixed point characteristics. Recent research has partly addressed these issues, providing convergence conditions for PnP variants of well-known optimization algorithms (ADMM, FB, gradient descent) using either results from fixed point theory [7] or from non-convex minimization [8–11]. For instance, PnP-ADMM offers convergence guarantees for a class of firmly nonexpansive (FNE) denoisers. Another approach [9, 11] explores denoisers explicitly defined as gradient steps, relaxing the FNE condition to demonstrate convergence to a stationary point of a cost function with an explicit regularization term derived from the denoiser. The choice of denoiser is crucial in PnP methods. A common approach is to use a Gaussian denoiser as a surrogate for the regularization's proximity operator, interpreted as a maximum-a-posteriori (MAP) estimator. Other works leverage Tweedie identity, using a Gaussian denoiser trained with the mean square error (MSE) loss to approximate the minimum mean square error (MMSE) estimator, giving access to a smoothed version of the log prior. However, one should note that the MAP interpretation is just one possibility [12], and concerns remain about the quality of the MMSE approximation with few training data and the relevance of the smoothed prior. Other works have also shown that using a Gaussian denoiser for reconstruction may not be optimal [13], and there are alternative denoising approaches [14] to remove more general artifacts. Hereinafter, we will follow this line and choose a denoiser mapping low-count to high-count images over a Gaussian denoiser. We will examine two convergent PnP schemes for PET reconstruction and investigate the impact of different denoisers. Apart from our preliminary work [15], we are unaware of any other research on PnP methods for PET reconstruction. Our contributions include (1) comparing two PnP schemes with varying denoiser constraints for our application, (2) training a prox-surrogate network using the reconstruction's fixed point equation, demonstrating lower sensitivity to hyperparameter

[1] When the learned operator is a gradient, the method is often labeled as a RED method.

choices compared to using a low-to-high count mapping, and (3) comparing the PnP approach with classical iterative reconstruction and post-reconstruction processing. The paper is structured as follows: Section 2 introduces PnP-ADMM [7] and PnP-FB [8] and their convergence conditions. Section 3 presents PET datasets and implementation details of the PnP schemes. Section 4 presents experimental results, followed by discussions in Section 5 and conclusions in Section 6.

2. METHOD

2.1. PET reconstruction problem

We aim to reconstruct of an image $\mathbf{x} \in [0, +\infty[^N$ from noisy measurements or projections $\mathbf{y} \in \mathbb{N}^M$ given by

$$\mathbf{y} = \mathcal{P}(\mathbf{H}\overline{\mathbf{x}} + \mathbf{b}), \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{M\times N}$ is the projection matrix, $\mathbf{b} \in]0, +\infty[^M$ is a known background term and $\mathcal{P}(\cdot)$ a Poisson process. Variational methods estimate $\overline{\mathbf{x}}$ by

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}} \quad \Big(C(\mathbf{x}) = LL(\mathbf{x}) + R(\mathbf{x}) + \iota_{[0,+\infty[^N}(\mathbf{x})\Big), \tag{2}$$

where $LL(\mathbf{x}) = -\sum_{m=1}^{M}[\mathbf{y}]_m \log([\mathbf{H}\mathbf{x} + \mathbf{b}]_m) + [\mathbf{H}\mathbf{x} + \mathbf{b}]_m + \text{cte}$ is the negative log-likelihood corresponding to the generalized Kullback Leibler divergence on the projection space, ι_S, the indicator function over the convex set S and R a convex non necessarily smooth regularization term typically based on the intensity difference between neighboring pixels. Equivalently, $\hat{\mathbf{x}}$ reads as the zero of the sum of three operators

$$0 \in \nabla LL(\hat{\mathbf{x}}) + \partial\iota_{[0,+\infty[^N}(\hat{\mathbf{x}}) + \partial R(\hat{\mathbf{x}}), \tag{3}$$

and can be computed using various algorithms.

2.2. PET PnP algorithms

In the PnP framework, a denoising neural network ($\mathbf{D}_\theta$,$\mathbf{D}_{\theta,\text{GS}}$) with a set of learnable parameters θ replaces an operator related to R in the reconstruction algorithm. We now focus on two algorithms.

Algorithm 1 PnP ADMM

Require: K, $\mathbf{z}_\rho^0$, $\mathbf{u}_\rho^0$, $\rho > 0$, $\mathbf{D}_\theta : \mathbb{R}^N \mapsto [0, +\infty[^N$
  for $k = 0$ to $K-1$ **do**
    $\mathbf{x}_\rho^{k+1} = \text{prox}_{1/\rho LL}(\mathbf{z}_\rho^k - \mathbf{u}_\rho^k)$
    $\mathbf{z}_\rho^{k+1} = \mathbf{D}_\theta(\mathbf{x}_\rho^{k+1} + \mathbf{u}_\rho^k)$
    $\mathbf{u}_\rho^{k+1} = \mathbf{u}_\rho^k + \mathbf{x}_\rho^{k+1} - \mathbf{z}_\rho^{k+1}$
  end for

PnP ADMM (Algo 1) When $\mathbf{D}_\theta$ is FNE ($2\mathbf{D}_\theta - \mathbf{Id}$ is 1-Lipschitz), $\mathbf{x}_\rho^k$, $\mathbf{z}_\rho^k$, $\mathbf{u}_\rho^k$ defined in Algo 1 converge to $\mathbf{x}_\rho^*$, $\mathbf{z}_\rho^*$, $\mathbf{u}_\rho^*$ such that

$$\mathbf{x}_\rho^* = \mathbf{z}_\rho^* \quad \mathbf{u}_\rho^* = \nabla LL(\mathbf{x}_\rho^*)/\rho \quad \mathbf{x}_\rho^* = \mathbf{D}_\theta(\mathbf{x}_\rho^* - \nabla LL(\mathbf{x}_\rho^*)/\rho) \tag{4}$$

Fixed point $\mathbf{x}_\rho^*$ also writes as the zero of a sum of operators:

$$0 \in (\mathbf{D}_\theta^{-1} - \mathbf{Id})\mathbf{x}_\rho^* + \nabla LL(\mathbf{x}_\rho^*)/\rho. \tag{5}$$

Note that $\mathbf{x}_\rho^*$ is no longer a minimizer of C and ρ does not only influence convergence speed but also fixed point selection, contrary to ADMM.

It is well known that the FNE property is difficult to enforce numerically. As [7], we will enforce it only locally when training the denoiser. We will also set $\mathbf{z}_\rho^0$ to an ordered subset expectation maximization (OSEM) solution and $\mathbf{u}_\rho^0$ to zero.

Algorithm 2 PnP FB

Require: K, $\mathbf{x}_\lambda^0$, $\lambda > 0$, $\tau > 0$, $\mathbf{D}_{\theta,\text{GS}} : \mathbb{R}^N \mapsto \mathbb{R}^N$
  for $k = 0$ to $K-1$ **do**
    $\mathbf{x}_\lambda^{k+1} = \text{prox}_{\tau LL + \iota_{[0,+\infty[^N}}(\lambda\tau\mathbf{D}_{\theta,\text{GS}}(\mathbf{x}_\lambda^k) + (1-\lambda\tau)\mathbf{x}_\lambda^k)$
  end for

PnP FB (Algo 2) If the residual operator $\mathbf{Id} - \mathbf{D}_{\theta,\text{GS}}$ is L-Lipschitz and the gradient of a lower semi-continuous but non-necessarily convex function R_θ, the sequence $(\mathbf{x}_\lambda^k)_{k\in\mathbb{N}}$ converges to a fixed point $\mathbf{x}_\lambda^*$ provided the step size τ satisfies $\tau < \frac{1}{\lambda L}$. The fixed point is a critical point of the cost function C with regularization λR_θ. Additionally, the value of $C(\mathbf{x}_\lambda^k)$ is non-increasing and $\mathbf{x}_\lambda^*$ satisfies

$$0 \in \nabla LL(\mathbf{x}_\lambda^*) + \partial\iota_{[0,+\infty[^N}(\mathbf{x}_\lambda^*) + \lambda(\mathbf{x}_\lambda^* - \mathbf{D}_{\theta,\text{GS}}(\mathbf{x}_\lambda^*)). \tag{6}$$

When comparing PnP-ADMM and PnP-FB, we see that the constraint on the Lipschitz constant of $2\mathbf{D}_\theta - \mathbf{Id}$ in PnP-ADMM for convergence is traded for a constraint on the architecture of the network in PnP-FB. In [8], the authors suggest using $R_\theta(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{N}_\theta(\mathbf{x})\|^2$ where $\mathbf{N}_\theta$ is a neural network, leading to

$$\mathbf{D}_{\theta,\text{GS}}(\mathbf{x}) = \mathbf{N}_\theta(\mathbf{x}) + J_{\mathbf{N}_\theta(x)}^\top(\mathbf{x} - \mathbf{N}_\theta(\mathbf{x})). \tag{7}$$

When comparing (3), (5), and (6), we observe that the denoisers play a distinct role in the fixed point equation. In PnP-FB, $\mathbf{D}_{\theta,\text{GS}}$ is a surrogate to a gradient residual, i.e., $\mathbf{Id} - \nabla R_\theta$, whereas in PnP-ADMM, $\mathbf{D}_\theta$ appears implicitly through its inverse as a surrogate to the resolvent [16] of $\partial(R + \iota_{[0,+\infty[^N})$, i.e., $(\mathbf{Id} + \partial(R + \iota_{[0,+\infty[^N}))^{-1}$.

The hyperparameters ρ and λ have distinct interpretations as well: in (6), λ defines $\mathbf{x}_\lambda^*$ as the equilibrium point between two sets zer $(\partial(LL + \iota_{[0,+\infty[^N}))$ and Fix $(\mathbf{D}_{\theta,\text{GS}})$, while, in (4), ρ appears in the input of the denoiser: (4) is akin to a denoising operation where $\nabla LL(\mathbf{x}_\rho^*)/\rho$ is an additive image-dependent degradation. In the following, we will demonstrate the advantage of learning a resolvent and thus using PnP-ADMM because one can train the denoiser using the fixed point equation.

3. NUMERICAL EXPERIMENTS

3.1. Data

The database used for learning and evaluation consists of 14 brain [^{18}F]-FDG PET scans from healthy subjects, along with their corresponding T1-weighted MRI images. The T1 images were segmented into 100 regions and the corresponding PET values were then measured in a frame between 30 and 60 minutes after injection to generate 14 3D phantoms with anatomical and functional variability. Noisy sinograms were simulated by considering normalization, attenuation but also scatter, and random effects for a Biograph 6 TruePoint TrueV PET system. For these, we derived paired high-count and low-count sinograms. The high-count sinograms were simulated by sampling for each phantom the total number of detected events in the range observed in the 14 original scans, and the low-count sinograms by further simulating a dose reduction factor of 5 (with appropriate scaling of the randoms). The resulting sinograms were reconstructed through CASToR [17] (OSEM with 8 iterations of 14 subsets) to yield paired high-count and low-count reconstructions ($\mathbf{x}_j^{i,\mathrm{HD}}$ and $\mathbf{x}_j^{i,\mathrm{LD}}$). In total, we simulated 10 doses for each phantom and considered 11 phantoms for training and 3 for validation, leading to 110 pairs ($\mathbf{x}_j^{i,\mathrm{HD}}$, $\mathbf{x}_j^{i,\mathrm{LD}}$) for training and 30 for validation.

3.2. Architecture

As advocated in [7–9], $\mathbf{D}_\theta$ is built as a DRUnet [18]. It comprises three levels, a single residual block, 32 channels, ELU activations, 3D strided and transposed convolutions, and a global skip connection for learning functions close to the identity operator, which is already FNE. For $\mathbf{D}_\theta$, we added an outer ReLU activation enforcing positivity. For PnP-FB, $\mathbf{D}_{\theta,\mathrm{GS}}$ writes as (7) with $\mathbf{N}_\theta$ a DRUnet.

3.3. Training strategy

For each high-count target ($\overline{\mathbf{x}}^q = \mathbf{x}_j^{i,\mathrm{HD}}$), we considered two types of inputs for the networks: $\mathbf{x}^{q,1} = \mathbf{x}_j^{i,\mathrm{LD}}$ and $\mathbf{x}^{q,2} = \mathbf{x}^{i,\mathrm{HD}} + \nabla LL(\mathbf{x}^{i,\mathrm{HD}})/\rho$.
$\mathbf{D}_\theta$ is either trained using $\mathbf{x}^{q,1}$, yielding $\mathbf{D}_\theta^{\mathrm{LH}}$, or using $\{\mathbf{x}^{q,1}, \mathbf{x}^{q,2}\}$ for several values of $\rho \in \{10^{-4}, 10^{-5}, 10^{-6}\}$, yielding $\mathbf{D}_\theta^{\mathrm{FP}}$. The rationale for keeping $\mathbf{x}^{q,1}$ is that in the first iteration of ADMM, the network is applied on an image close to an OSEM solution. To locally enforce the FNE property and handle images with varying doses and thus value ranges, we use the following regularized loss inspired by [7]:

$$\frac{\|\overline{\mathbf{x}}^q - \mathbf{D}_\theta(\mathbf{x}^{q,s})\|_2^2}{\|\overline{\mathbf{x}}^q\|_2} + \beta \max\{|||J_{2\mathbf{D}_\theta - \mathbf{Id}}(\tilde{\mathbf{x}}^{q,s})||| + \epsilon - 1, 0\}^{1+\alpha}, \tag{8}$$

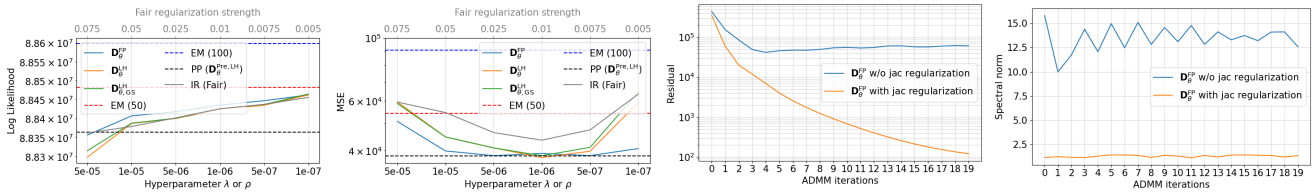
where $\beta = 0.01$, $\epsilon = 0.05$, $\alpha = 0.1$ and $\tilde{\mathbf{x}}^{q,s} = \kappa\overline{\mathbf{x}}^q + (1-\kappa)\mathbf{x}^{q,s}$, $\kappa \sim \mathcal{U}[0,1]$. The spectral norm $|||\cdot|||$ of the jacobian was computed using 10 iterations of the power method and autodifferentiation. Since the computation of the spectral norm is intensive, we pre-trained the network without the jacobian regularization for 100 epochs, then we trained using (8) for 20 epochs.
Contrary to PnP-ADMM, PnP-FB does not allow for training from the fixed point equation for a range of λ. $\mathbf{D}_{\theta,\mathrm{GS}}$ was trained on $\mathbf{x}^{q,1}$ for 100 epochs using the scaled MSE loss, yielding $\mathbf{D}_{\theta,\mathrm{GS}}^{\mathrm{LH}}$. $\mathbf{N}_\theta$ is initialized to our previous DRUnet pretrained on low-count images ($\mathbf{D}_\theta^{\mathrm{Pre,LH}}$).

3.4. Reconstruction details and reference methods

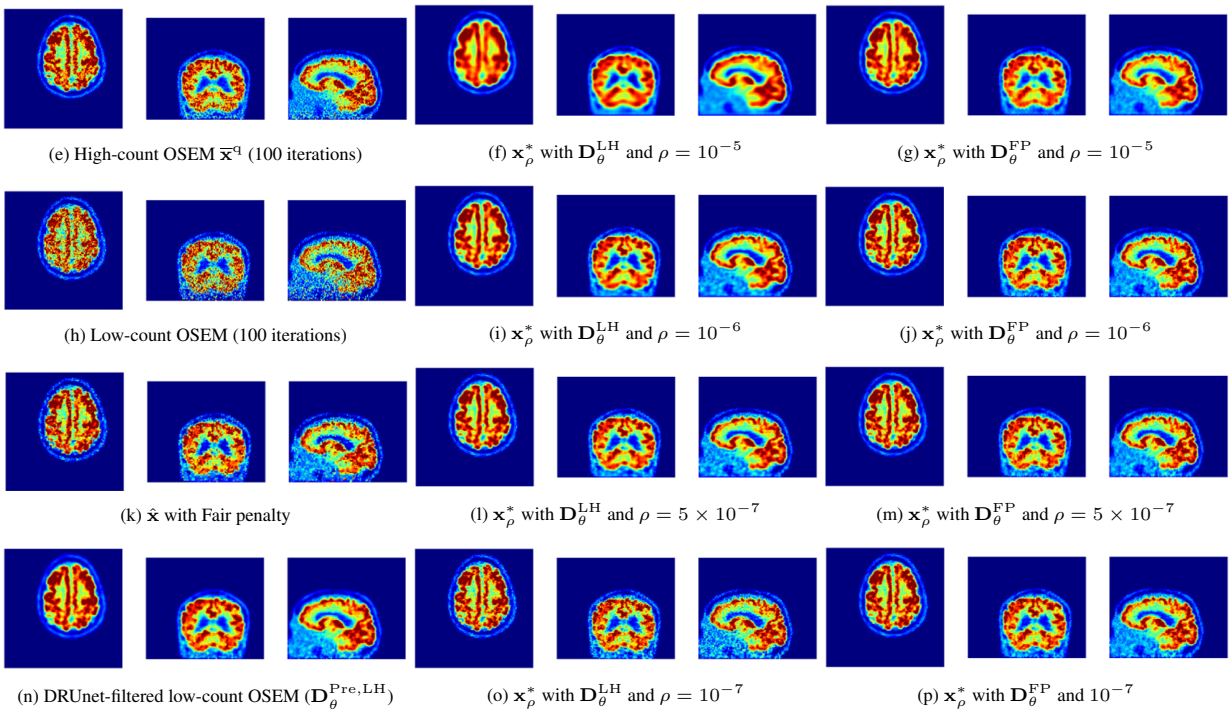
Computation of the proximity operators in PnP-ADMM and PnP-FB are penalized PET reconstruction problems for which efficient algorithms exist. Following [8], a backtracking loop is used in PnP-FB to set the step size. The PnP schemes are compared to an OSEM reconstruction processed by $\mathbf{D}_\theta^{\mathrm{Pre,LH}}$, a variational method solving (2) with the *Fair* regularization function on the difference between the first-order neighboring pixels [2]. PnP-ADMM was run until $\|\mathbf{x}_\rho^{k+1} - \mathbf{v}_\rho^k\|_2/\|\mathbf{v}_\rho^k\|_2$ and $\|\mathbf{v}_\rho^{k+1} - \mathbf{v}_\rho^k\|_2/\|\mathbf{v}_\rho^k\|_2$ are below 5×10^{-4}. PnP-FB was run until $\|\mathbf{x}_\lambda^{k+1} - \mathbf{x}_\lambda^k\|_2/\|\mathbf{x}_\lambda^k\|_2 < 5 \times 10^{-4}$.

4. RESULTS

Figures 1a and 1b display the log-likelihood and MSE for different denoisers and varying values of ρ in PnP-ADMM and λ in PnP-FB. We see the performance of PnP-ADMM and PnP-FP with $\mathbf{D}_\theta^{\mathrm{LH}}$ and $\mathbf{D}_{\theta,\mathrm{GS}}^{\mathrm{LH}}$ is highly sensitive on the choice of these hyperparameters and that the lowest MSE/high likelihood is achieved for $\lambda = 10^{-6}$ and $\rho = 10^{-6}$. For these values, the MSE and likelihood are comparable for PnP-ADMM and PnP-FP ($\approx 3.8 \times 10^4$ for the MSE and 8.842×10^7 for the likelihood), which is corroborated by the images (not shown). In addition, we observe that both schemes exhibited rapid convergence (not shown) in about 20 iterations.
As depicted on figures 1a and 1b, using $\mathbf{D}_\theta^{\mathrm{FP}}$ results in constant MSE ($\approx 3.9 \times 10^4$) and higher log-likelihood for a wide range of ρ values, close to the optimal value obtained at $\lambda = \rho = 10^{-6}$ with $\mathbf{D}_\theta^{\mathrm{LH}}$ and $\mathbf{D}_{\theta,\mathrm{GS}}^{\mathrm{LH}}$. This improvement also translates visually (fig. 1g-1p). We note that when $\mathbf{D}_\theta^{\mathrm{FP}}$ is trained without regularizing the spectral norm, $|||J_{2\mathbf{D}_\theta^{\mathrm{FP}} - \mathbf{Id}}|||$ explodes rapidly (fig. 1d) and PnP-ADMM does not converge as observed in our previous work [15] (fig. 1c). This observation highlights the importance of satisfying convergence conditions when implementing PnP schemes.
Finally, PnP-ADMM with $\mathbf{D}_\theta^{\mathrm{FP}}$ outperforms post-filtered OSEM using $\mathbf{D}_\theta^{\mathrm{Pre,LH}}$ (fig. 1n, MSE: 3.8 $\times 10^4$, likelihood: 8.831×10^7) and the IR solution (fig. 1k, MSE: 4.3 $\times 10^4$, likelihood: 8.842×10^7) with the lowest MSE on figure 1b.



(a) Log-likelihood as a function of ρ or λ (b) MSE as a function of ρ or λ (c) $\|\mathbf{v}_\rho^{k+1} - \mathbf{v}_\rho^k\|$ along ADMM iterations (d) Spectral norm along ADMM iterations



(e) High-count OSEM $\overline{\mathbf{x}}^{\mathrm{q}}$ (100 iterations) (f) $\mathbf{x}_\rho^*$ with $\mathbf{D}_\theta^{\mathrm{LH}}$ and $\rho = 10^{-5}$ (g) $\mathbf{x}_\rho^*$ with $\mathbf{D}_\theta^{\mathrm{FP}}$ and $\rho = 10^{-5}$

(h) Low-count OSEM (100 iterations) (i) $\mathbf{x}_\rho^*$ with $\mathbf{D}_\theta^{\mathrm{LH}}$ and $\rho = 10^{-6}$ (j) $\mathbf{x}_\rho^*$ with $\mathbf{D}_\theta^{\mathrm{FP}}$ and $\rho = 10^{-6}$

(k) $\hat{\mathbf{x}}$ with Fair penalty (l) $\mathbf{x}_\rho^*$ with $\mathbf{D}_\theta^{\mathrm{LH}}$ and $\rho = 5 \times 10^{-7}$ (m) $\mathbf{x}_\rho^*$ with $\mathbf{D}_\theta^{\mathrm{FP}}$ and $\rho = 5 \times 10^{-7}$

(n) DRUnet-filtered low-count OSEM ($\mathbf{D}_\theta^{\mathrm{Pre,LH}}$) (o) $\mathbf{x}_\rho^*$ with $\mathbf{D}_\theta^{\mathrm{LH}}$ and $\rho = 10^{-7}$ (p) $\mathbf{x}_\rho^*$ with $\mathbf{D}_\theta^{\mathrm{FP}}$ and 10^{-7}

Fig. 1: Comparison of the different PnP reconstructions, post-reconstruction filtering and IR reconstructions

5. DISCUSSION

Our experiments demonstrated that PnP-ADMM and PnP-FB with denoisers mapping low- to high-count images produce similar images despite having distinct denoiser constraints. Both PnP schemes display high sensitivity to hyperparameters governing fixed point selection and convergence speed.

PnP-FB stands out for its use of a backtracking line search, eliminating the need for knowledge of L. Consequently, the scheme is guaranteed to converge, albeit potentially slowly, especially when L is high (close to 1 in our experiments). In contrast, because we enforced a local FNE constraint, PnP-ADMM is not guaranteed to converge per se. However, our interpolation rule was found sufficient to satisfy the FNE condition on the PnP-ADMM iterates.

Only PnP-ADMM allows enforcing proximity of the fixed point to high-count images across varying ρ values, which is crucial as ρ varies per patient per exam. While this approach does not offer complete control over the optimization path, it consistently produced interesting fixed points in a range of ρ values, close to the best MSE/log-likelihood trade-off. We hypothesize that this is due to the input of the denoiser being a perturbation of a proximity operator, keeping the estimate close to the training images, especially if ρ is not excessively large - the sole regime of interest. This effect would not be the same in algorithms such as primal dual schemes where reconstruction and regularization are performed jointly. Complete fixed point control would require solving a costly bilevel optimization problem via implicit differentiation.

Our reconstructed images remain noisy due to noisy target images. The Artifact2Artifact approach of [14] may be worth investigating to address this issue.

6. CONCLUSION

In this study, we introduced two novel PET reconstruction methods combining deep learning and convergence guarantees. The first employs the plug-and-play forward-backward algorithm with minimal denoiser requirements. The second relies on the plug-and-play ADMM framework, providing enhanced fixed-point control. Our experiments demonstrate that these methods outperform CNN post-filtering and a classical variational approach, thus highlighting their potential to enhance PET image quality for improved clinical applications.

Compliance with Ethical Standards

This is a numerical simulation study for which no ethical approval was required.

Acknowledgment

We acknowledge financial support from the French National Research Agency (ANR) under grant ANR-20-CE45-0020 (ANR MULTIRECON). This work was partly funded by the France Life Imaging (ANR-11-INBS-0006 grant from the French "Investissements d'Avenir" program).

7. REFERENCES

[1] J. Nuyts, D. Beque, P. Dupont, and L. Mortelmans, "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography," *IEEE Transactions on Nuclear Science*, vol. 49, no. 1, pp. 56–60, Feb. 2002.

[2] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Transactions on Medical Imaging*, vol. 9, no. 4, pp. 439–446, Dec. 1990.

[3] S. Kaplan and Y.-M. Zhu, "Full-Dose PET Image Estimation from Low-Dose PET Image Using Deep Learning: a Pilot Study," *Journal of Digital Imaging*, vol. 32, no. 5, pp. 773–778, Oct. 2019.

[4] K. Gong, J. Guan, C.-C. Liu, and J. Qi, "PET Image Denoising Using a Deep Neural Network Through Fine Tuning," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 153–161, Mar. 2019.

[5] G. Nataraj and R. Otazo, "Investigating robustness to unseen pathologies in model-free deep multicoil reconstruction," in *Proc. ISMRM Workshop on Data Sampling & Image Reconstruction*, 2020.

[6] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30088–30095, Dec. 2020.

[7] J.-C. Pesquet, A. Repetti, M. Terris, and Y. Wiaux, "Learning Maximally Monotone Operators for Image Recovery," *SIAM Journal on Imaging Sciences*, vol. 14, no. 3, pp. 1206–1237, Jan. 2021.

[8] S. Hurault, A. Leclaire, and N. Papadakis, "Gradient Step Denoiser for convergent Plug-and-Play," Feb. 2022.

[9] S. Hurault, A. Leclaire, and N. Papadakis, "Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization," in *International Conference on Machine Learning*. 2022, pp. 9483–9505, PMLR.

[10] S. Hurault, A. Chambolle, A. Leclaire, and N. Papadakis, "A Relaxed Proximal Gradient Descent Algorithm for Convergent Plug-and-Play with Proximal Denoiser," in *Scale Space and Variational Methods in Computer Vision*, 2023, pp. 379–392.

[11] R. Cohen, Y. Blau, D. Freedman, and E. Rivlin, "It Has Potential: Gradient-Driven Denoisers for Convergent Solutions to Inverse Problems," in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 18152–18164.

[12] R. Gribonval, "Should Penalized Least Squares Regression be Interpreted as Maximum A Posteriori Estimation?," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2405–2410, May 2011.

[13] X. Xu, Y. Sun, J. Liu, B. Wohlberg, and U. S. Kamilov, "Provable Convergence of Plug-and-Play Priors with MMSE denoisers," *IEEE Signal Processing Letters*, vol. 27, pp. 1280–1284, 2020.

[14] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "RARE: Image Reconstruction using Deep Priors Learned without Ground Truth," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1088–1099, Oct. 2020.

[15] F. Sureau, M. Latreche, M. Savanier, and C. Comtat, "Convergent ADMM Plug and Play PET Image Reconstruction," in *Proceedings of the 17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2023.

[16] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer International Publishing, 2017.

[17] T. Merlin, S. Stute, D. Benoit, J. Bert, T. Carlier, C. Comtat, M. Filipovic, F. Lamare, and D. Visvikis, "CASToR: a generic data organization and processing code framework for multi-modal and multi-dimensional tomographic reconstruction," *Physics in Medicine & Biology*, vol. 63, no. 18, pp. 185005, Sept. 2018.

[18] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-Play Image Restoration With Deep Denoiser Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.