

# Fast Gated PET Direct Motion Estimation Using Ordered Subsets

Alexandre Bousse\*, Brian F. Hutton, Kris Thielemans

**Abstract**—Direct motion compensation in gated PET can be achieved with maximum-likelihood (ML) joint activity reconstruction/motion estimation (JRM). Whereas the motion-compensated (MC) activity image can be efficiently updated with ordered subsets (OS) expectation maximisation (EM), the motion optimisation suffers from high computational cost due to the utilisation of line-search algorithms, which require to perform projections/backprojections each time the likelihood and its gradient are evaluated. The idea of this work is to use OS to speed-up the estimation of the motion parameters, by maximising each sub-likelihood sequentially. We show that the utilisation of OS significantly reduces the overall joint-estimation computational cost with negligible effect on the final solution.

**Index Terms**—Motion-compensated PET reconstruction, ordered subsets, maximum-likelihood, direct motion estimation.

## I. INTRODUCTION

Direct (or data-driven) motion compensation methods in gated PET can be achieved by maximum-likelihood (ML) joint activity reconstruction/motion estimation (JRM) methods [1]–[5]. They differ from indirect motion compensation methods which rely on a pre-estimation of the motion fields, either by registration of individually reconstructed PET gates [6] or by using another imaging modality capable of measuring the motion [7], [8]. While they are relatively easy to implement, the former indirect approach suffers from noise due to the limited number of counts at each gate and the absence of accurate attenuation correction, while the latter requires additional costly equipment. In addition, it has been demonstrated that direct approaches achieve better bias/noise trade-off as compared to indirect approaches [3], [5], and only a single attenuation map, deformed alongside the activity image, is needed.

The drawback of direct approaches is their computational cost. Maximisation of the likelihood with respect to the motion needs to be performed in the sinogram domain. When a line-search optimisation approach is used [3], [5], the deformed volumes need to be projected each time the objective function is evaluated. This computational burden can be avoided by using surrogate functions, which allow one to transfer the optimisation problem into the image space without altering the monotonic properties [2], [4]. However, the usual surrogate functions cannot be applied when accounting for deformable attenuation.

This work was supported by GE Healthcare. *Asterisk indicates corresponding author.*

\*A. Bousse is with the Institute of Nuclear Medicine, University College London, London NW1 2BU, UK (email: a.bousse@ucl.ac.uk).

B. F. Hutton and K. Thielemans are with the Institute of Nuclear Medicine, University College London, London NW1 2BU, UK (email: a.bousse@ucl.ac.uk).

B. F. Hutton is also with the Centre for Medical Radiation Physics, University of Wollongong, Wollongong NSW 2522, Australia.

In this paper we propose to utilise the same strategy that was adopted to accelerate ML expectation maximisation (EM) image reconstruction [9], [10] by the utilisation of ordered subsets (OS) [11]. We update the motion with a quasi-Newton line search using one subset of the data only, and sequentially iterate over the subsets. A summary of the model we utilised in [5] and its OS version for motion estimation are presented in section II, and an acceleration evaluation based on the number of projections/backprojections used is presented in Section III.

## II. METHOD

### A. Joint Activity Reconstruction/Motion Estimation

In this section we briefly summarise the model used in [5]. The observed PET data are regrouped into  $n_g$  motion-free PET gates  $\mathbf{g}_1, \dots, \mathbf{g}_{n_g}$ , each  $\mathbf{g}_\ell$  being a realisation of a Poisson random vector with independent entries. The expectation of this Poisson variable is determined by the activity distribution image vector  $\mathbf{f}$  and the motion parameter  $\alpha_\ell$  at gate  $\ell \in \{1, \dots, n_g\}$  as

$$\bar{\mathbf{g}}_\ell(\mathbf{f}, \alpha_\ell, \boldsymbol{\mu}) \triangleq \tau_\ell \mathbf{H}(\alpha_\ell, \boldsymbol{\mu}) \mathbf{f} + \mathbf{r}_\ell,$$

where  $\mathbf{r}_\ell$  is the expected background event vector (random and scatter) and  $\tau_\ell$  is the duration of gate  $\ell$ . Respiratory gating can be achieved using external hardware [12] or data-driven methods [13], [14]. The motion-compensated attenuation-corrected PET system matrix  $\mathbf{H}(\alpha_\ell, \boldsymbol{\mu})$  is given by the deformation field defined by the parameter  $\alpha_\ell$  and the attenuation map  $\boldsymbol{\mu}$  as

$$\mathbf{H}(\alpha_\ell, \boldsymbol{\mu}) \triangleq \text{diag} \left\{ e^{-R\mathbf{W}(\alpha_\ell)} \right\} \mathbf{P}\mathbf{W}(\alpha_\ell), \quad (1)$$

where  $\mathbf{P}$  is the “motion-free” PET system matrix,  $\mathbf{R}$  is a line-integral operator that computes the attenuation coefficients for each bin pair, and  $\mathbf{W}(\alpha_\ell)$  is the image warping matrix associated to the motion field defined by  $\alpha_\ell$  [5]. The choice of the motion model parameterisation depends on the application. For example in [5] we used a B-spline non-rigid model for respiratory motion whereas in [15] we used a rigid model for head motion. Note that in the projection model (1) the attenuation  $\boldsymbol{\mu}$  is affected by the same motion  $\mathbf{W}(\alpha_\ell)$  as  $\mathbf{f}$ .

Denoting  $\boldsymbol{\vartheta} = \{\alpha_\ell\}_{\ell=1}^{n_g}$ , the log-likelihood takes the form of a sum over the gates  $\ell \in \{1, \dots, n_g\}$  and the detector bins  $i \in \mathcal{I}$ :

$$\Phi(\mathbf{f}, \boldsymbol{\vartheta}, \boldsymbol{\mu}) \triangleq \sum_{\ell=1}^{n_g} \sum_{i \in \mathcal{I}} \Lambda(g_{i,\ell} | \bar{g}_{i,\ell}(\mathbf{f}, \alpha_\ell, \boldsymbol{\mu})), \quad (2)$$

$$\Lambda(x | y) \triangleq x \log y - y$$

with  $g_{i,\ell} \triangleq [g_\ell]_i$  and  $\bar{g}_{i,\ell} \triangleq [\bar{g}_\ell]_i$ . Joint activity reconstruction/motion estimation (JRM) by maximum-likelihood is achieved by maximising  $\Phi$  with respect to  $\mathbf{f}$  and  $\vartheta$ :

$$(\hat{\mathbf{f}}, \hat{\vartheta}) = \arg \max_{\mathbf{f} \geq 0, \vartheta} \Phi(\mathbf{f}, \vartheta, \boldsymbol{\mu}). \quad (3)$$

In [5], (3) was achieved by alternating maximisation in  $\mathbf{f}$  and  $\vartheta$ . While we used motion-compensated (MC) ML-EM steps for  $\mathbf{f}$ , we used a limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) quasi-Newton line-search [16] for  $\vartheta$ .

The reconstructed activity images at each gate are

$$\hat{\mathbf{f}}_\ell = \mathbf{W}(\hat{\boldsymbol{\alpha}}_\ell) \hat{\mathbf{f}}, \quad \ell = 1, \dots, n_g \quad (4)$$

and are considered as the final output of the method.

### B. Ordered Subsets for Image and Motion Update

A MC-OS-EM update of  $\mathbf{f}$  consists of updating each activity values  $f_j \triangleq [\mathbf{f}]_j$  simultaneously using data from successive disjoint bin subsets  $\mathcal{S}_n$ , with  $\cup_n \mathcal{S}_n = \mathcal{I}$ ,

$$\hat{f}_j^n = \frac{\hat{f}_j^{n-1}}{p_{n,j}(\vartheta, \boldsymbol{\mu})} \sum_{\ell=1}^{n_g} \sum_{i \in \mathcal{S}_n} \frac{g_{i,\ell}}{\bar{g}_{i,\ell}(\hat{\mathbf{f}}^{n-1}, \boldsymbol{\alpha}_\ell, \boldsymbol{\mu})} \quad (5)$$

where  $p_{n,j}(\vartheta, \boldsymbol{\mu}) = \sum_{i \in \mathcal{S}_n} \sum_{\ell=1}^{n_g} \tau_\ell [\mathbf{H}(\boldsymbol{\alpha}_\ell, \boldsymbol{\mu})]_{i,j}$ , which is less computationally expensive than a standard ML-EM update. It is worth noting that performing (5) is a maximisation step of the partial log-likelihood

$$\phi_n(\mathbf{f}, \vartheta, \boldsymbol{\mu}) \triangleq \sum_{i \in \mathcal{S}_n} \sum_{\ell=1}^{n_g} \Lambda(g_{i,\ell} | \bar{g}_{i,\ell}(\mathbf{f}, \boldsymbol{\alpha}_\ell, \boldsymbol{\mu})),$$

and a ‘‘pseudo-maximiser’’ in  $\mathbf{f}$  of the complete log-likelihood  $\Phi(\mathbf{f}, \vartheta, \boldsymbol{\mu}) = \sum_n \phi_n(\mathbf{f}, \vartheta, \boldsymbol{\mu})$  is obtained by circularly iterating over  $n$ .

The main idea of this paper is to use a similar approach for the estimation of the motion parameters  $\boldsymbol{\alpha}$ , *i.e.* by performing successive maximisations of the partial log-likelihoods  $\phi_n$

$$\hat{\vartheta}^n = \arg \max_{\vartheta} \phi_n(\mathbf{f}, \vartheta, \boldsymbol{\mu}) \quad (6)$$

with an L-BFGS quasi-Newton line-search initialised from  $\hat{\vartheta}^{n-1}$ . Thus, (6) is the analog of the OS-EM step (5) for  $\vartheta$ . Note that as for OS-EM, this approach theoretically does not converge to a maximiser of  $\Phi$  because of the inconsistencies between subsets due to the noise. For this reason, it would be desirable not to use too many L-BFGS iterations per subset as a maximiser of  $\phi_n$  does not necessarily maximise  $\phi_{n+1}$ . However, this might have limited impact in practice, especially during the early iterations.

The overall methodology is summarised in Algorithm 1.  $\vartheta$  and  $\mathbf{f}$  are initialised with  $\mathbf{0}$  and  $\mathbf{1}$  respectively. For each sub-routine, the variables on the right of the semicolon correspond to the initialisations. For example,  $\vartheta^n \leftarrow \text{L-BFGS}(\phi_n(\mathbf{f}, \cdot, \boldsymbol{\mu}); \vartheta^{n-1})$  is an L-BFGS maximisation of  $\vartheta \mapsto \phi_n(\mathbf{f}, \vartheta, \boldsymbol{\mu})$  initialised from  $\vartheta^n$ . Prior to the main loop, a first reconstruction of  $\mathbf{f}$  is achieved with an ML-EM reconstruction from a single gate  $g_{\ell_0}$ , which is usually chosen so that it is roughly aligned with the input attenuation map  $\boldsymbol{\mu}$ . Note that we demonstrated in [5], [17] that  $\boldsymbol{\mu}$  does not need to be aligned with any gates  $\ell$ , but convergence will be faster with proper initialisation.

---

**Algorithm 1:** Ordered subsets joint PET maximum-likelihood motion estimation/image reconstruction

---

**Input:** Gated PET data  $\{g_\ell\}_{\ell=1}^{n_g}$ , attenuation map  $\boldsymbol{\mu}$ , number of subsets  $N_{\text{sub}}$

**Output:** PET image  $\mathbf{f}$ , motion parameter  $\vartheta$

*initialisation* ;

$\vartheta \leftarrow \mathbf{0}$  ;

$\mathbf{f} \leftarrow \mathbf{1}$  ;

$\mathbf{f} \leftarrow \text{ML-EM}(g_{\ell_0}, \boldsymbol{\mu}; \mathbf{f})$  ;

**for**  $k = 1, \dots, \#$ outer iterations **do**

$\vartheta^0 \leftarrow \vartheta$  ;

**for**  $n = 1, \dots, N_{\text{sub}}$  **do**

$\vartheta^n \leftarrow \text{L-BFGS}(\phi_n(\mathbf{f}, \cdot, \boldsymbol{\mu}); \vartheta^{n-1})$ ;

**end**

$\vartheta \leftarrow \vartheta^{N_{\text{sub}}}$  ;

$\mathbf{f} \leftarrow \text{MC-OS-EM}(\{g_\ell\}_{\ell=1}^{n_g}, \vartheta, \boldsymbol{\mu}; \mathbf{f})$  ;

**end**

---

## III. VALIDATION

### A. Data Generation

We generated  $n_g = 3$  XCAT activity and attenuation volumes ( $133 \times 133 \times 42$ , 3.125 mm edge cubic voxels)  $\{\mathbf{f}_\ell\}_{\ell=1}^3$  and  $\{\boldsymbol{\mu}_\ell\}_{\ell=1}^3$  (Figures 1(a) and 2(a)), at 3 different stages of the respiratory cycle. Gated PET data were simulated as

$$g_\ell \sim \text{Poisson}(\tau_\ell \text{diag}\{e^{-\mathbf{R}\boldsymbol{\mu}_\ell}\} \mathbf{P} \mathbf{f}_\ell + \mathbf{r}_\ell).$$

The same duration  $\tau_\ell$  was used for each gate. The total number of counts was  $3 \times 10^7$ , including background  $\mathbf{r}_\ell$  (30% of the total counts). The same projector was used for  $\mathbf{P}$  and  $\mathbf{R}$  and included a 5 mm FWHM resolution model.

### B. Projections/Backprojections vs Log-Likelihood

1) *Reconstructions:* We proceeded with 2 JRM reconstructions: standard JRM (noOS-JRM) as described in [5], and JRM with OS for motion estimation (OS-JRM), using  $N_{\text{sub}} = 15$  subsets. We used  $\boldsymbol{\mu} = \boldsymbol{\mu}_1$  (the XCAT attenuation volume corresponding to the first gate) for attenuation correction, and gate  $\ell_0 = 1$  was used to obtain a first ML estimate of  $\mathbf{f}$ . The B-splines motion model used was the same as in [5], and  $\vartheta = \{\boldsymbol{\alpha}_\ell\}_{\ell=1}^{n_g}$  corresponds to the B-spline coefficients.

The estimated activity images and motion parameters are denoted

$$\hat{\mathbf{f}}^{\text{noOS}}, \hat{\vartheta}^{\text{noOS}} = \{\hat{\boldsymbol{\alpha}}_\ell^{\text{noOS}}\}_{\ell=1}^{n_g}, \quad \hat{\mathbf{f}}^{\text{OS}}, \hat{\vartheta}^{\text{OS}} = \{\hat{\boldsymbol{\alpha}}_\ell^{\text{OS}}\}_{\ell=1}^{n_g}.$$

The reconstructed activities at each gate are defined following (4):

$$\hat{\mathbf{f}}_\ell^{\text{noOS}} = \mathbf{W}(\hat{\boldsymbol{\alpha}}_\ell^{\text{noOS}}) \hat{\mathbf{f}}^{\text{noOS}}, \quad \hat{\mathbf{f}}_\ell^{\text{OS}} = \mathbf{W}(\hat{\boldsymbol{\alpha}}_\ell^{\text{OS}}) \hat{\mathbf{f}}^{\text{OS}},$$

and the corresponding warped attenuation maps are:

$$\hat{\boldsymbol{\mu}}_\ell^{\text{noOS}} = \mathbf{W}(\hat{\boldsymbol{\alpha}}_\ell^{\text{noOS}}) \boldsymbol{\mu}, \quad \hat{\boldsymbol{\mu}}_\ell^{\text{OS}} = \mathbf{W}(\hat{\boldsymbol{\alpha}}_\ell^{\text{OS}}) \boldsymbol{\mu}.$$

Note that for both approaches we used OS-EM to update  $\mathbf{f}$  (15 subsets as well, 1 iteration per subset), but OS motion updates were used for OS-JRM only. 10 L-BFGS motion inner

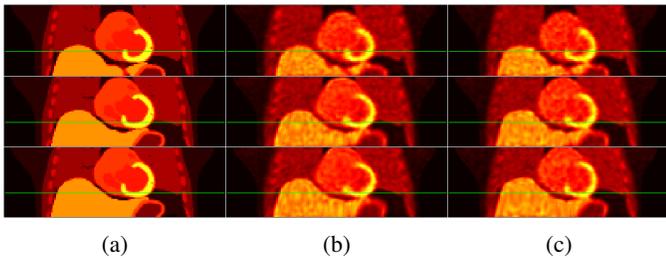


Fig. 1: Activity images at each gates: (a)  $f_\ell$  (XCAT phantoms); (b)  $\hat{f}_\ell^{\text{noOS}}$ ; (c)  $\hat{f}_\ell^{\text{OS}}$ .

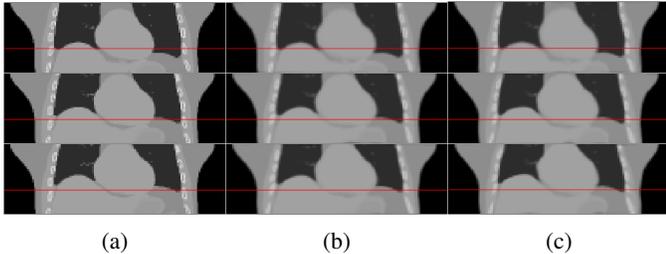


Fig. 2: Attenuation images at each gates: (a)  $\mu_\ell$  (XCAT phantoms); (b)  $\hat{\mu}_\ell^{\text{noOS}}$ ; (c)  $\hat{\mu}_\ell^{\text{OS}}$ .

iterations were used for OS-JRM and noOS-JRM, and 3 outer iterations were used for both. We also added a small quadratic penalty on  $\vartheta$ , as we previously did in [5].

Figures 1(b) and 1(c) show the reconstructed gates  $\{\hat{f}_\ell^{\text{noOS}}\}_{\ell=1}^{n_g}$  and  $\{\hat{f}_\ell^{\text{OS}}\}_{\ell=1}^{n_g}$ . Both look similar to the XCAT phantoms  $\{f_\ell^{\text{noOS}}\}_{\ell=1}^{n_g}$  and  $\{f_\ell\}_{\ell=1}^{n_g}$  used to generate data (Figure 1(a)). This result suggests that using OS to estimate motion in JRM does not affect the joint estimation of  $(f, \vartheta)$ . Similarly, the warped attenuation maps  $\{\hat{\mu}_\ell^{\text{noOS}}\}_{\ell=1}^{n_g}$  and  $\{\hat{\mu}_\ell^{\text{OS}}\}_{\ell=1}^{n_g}$  appear similar to  $\{\mu_\ell\}_{\ell=1}^{n_g}$  (Figure 2(a)).

2) *Results:* To assess the estimation speed-up, we looked at the log-likelihood values  $\Phi$  (2) as a function of the number of projections/backprojections (P&B) at each inner iteration of Algorithm 1 (L-BFGS and MC-OS-EM). When a subset is used, the number of P&B was incremented by  $1/N_{\text{sub}}$ . Since the computational cost is mostly explained by the number of P&B, it provides an accurate indicator of the computational time. The image  $f$  was initialised

Figure 3 shows the log-likelihood values versus the number of P&B for both noOS-JRM and OS-JRM. The first plot corresponds to the  $\Phi$ -value following the first L-BFGS iteration (after the initial ML-EM initialisation using gate  $\ell = 1$ ). These curves show that JRM speed is significantly increased when OS are used for motion estimation, in a similar fashion as OS-EM accelerates EM. The sudden increase of the  $\Phi$ -values at 90 P&B for OS-JRM and 170 P&B for noOS-JRM corresponds to the beginning of the first MC-OS-EM update of  $f$ , following the first estimation of  $\vartheta$ . Note that the OS-JRM plot oscillates slightly. This is due to the fact that the motion estimation switches between subsets that are not consistent due to the noise, which does not guarantee an increase of  $\Phi$  at each iteration.

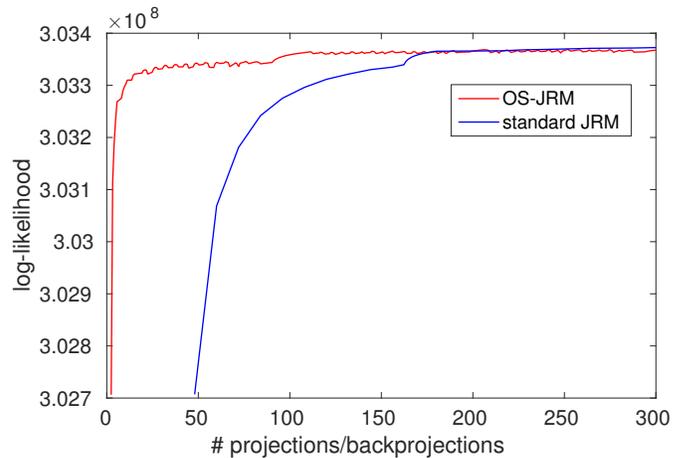


Fig. 3: Number of P&B vs the log-likelihood for noOS-JRM and OS-JRM.

### C. Projections/Backprojections vs “Image Registration Error”

The estimation of  $\vartheta$  can be seen as a registration task at each gate, except that the target image is observed in the sinogram space only. Therefore investigating the difference image  $W(\hat{\alpha})f^{\text{source}} - f^{\text{target}}$  at different stages of the estimation of  $\alpha$  is useful to illustrate the acceleration.

Here we used  $f^{\text{target}} = f_3$  (gate  $\ell = 3$ ) as the target image, from which the projection  $g^{\text{target}} = g_3$  was generated. The input image  $f^{\text{source}}$  was set to  $f_1$  (gate  $\ell = 1$ , corresponding to the input attenuation map  $\mu = \mu_1$ ), and we proceeded with the maximisation of the log-likelihood function

$$\alpha \mapsto \sum_{i \in \mathcal{I}} \Lambda(\bar{g}_i^{\text{target}} | \bar{g}_{i,3}(f^{\text{source}}, \alpha, \mu))$$

with both noOS-L-BFGS (60 iterations) and OS-L-BFGS (15 iterations per subset). We also used  $N_{\text{sub}} = 15$ . Note that since the activity image  $f^{\text{source}}$  is known, there is no need for a reconstruction step and therefore there is no outer loop.

The relative difference images are shown in Figure 4, and the normalised squared error, defined as

$$\text{NSE}(W(\hat{\alpha})f^{\text{source}}, f^{\text{target}}) = \frac{\|W(\hat{\alpha})f^{\text{source}} - f^{\text{target}}\|_2^2}{\|f^{\text{target}}\|_2^2}, \quad (7)$$

is displayed in Figure 5. While it takes about 200 P&B to achieve a good match with noOS-L-BFGS, a similar result can be achieved with less than 30 P&B when using OS-L-BFGS. This observation is consistent with the results displayed in Figure 3.

## IV. DISCUSSION

Utilising OS for motion estimation has an effect that is comparable to utilising OS for EM image reconstruction, as limiting the number of bins for a motion parameter update reduces the computational burden. Similarly to OS-EM, this approach does not converge to a maximum, as the update obtained from the maximisation of a sub-likelihood  $\phi_n$  does not maximise the other sub-likelihoods. In principle, each subset should be used to maximise the information and reduce the variance and the accuracy of motion estimation, but less subsets may be used for more acceleration. Further work to

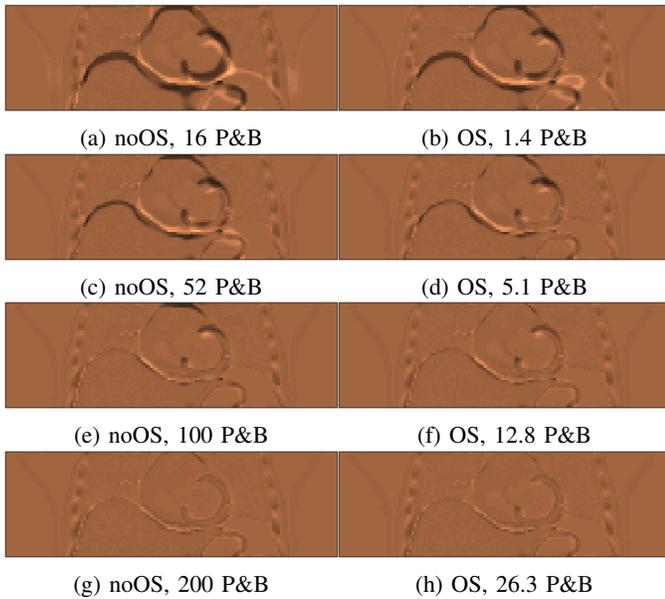


Fig. 4: Relative difference  $W(\hat{\alpha})f^{\text{source}} - f^{\text{target}}$  with nonOS-L-BFGS and OS-L-BFGS at different stages of the optimisation.

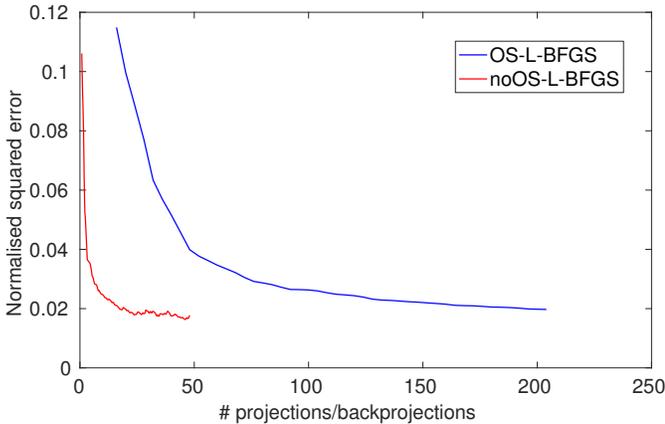


Fig. 5: Number of P&B vs the normalised squared error (NSE) for noOS-L-BFGS and OS-L-BFGS, as defined in (7).

assess the number of subsets for optimal performance should be conducted.

Despite the absence of convergence, the speed boost provided by OS is significant and can be very useful, especially in situations where the computational cost is high such as in time-of-flight (TOF) PET JRM [17], or where the input attenuation is misaligned with all the PET gates, which requires a large number of iterations [5], [17].

## V. CONCLUSION

We introduced an OS approach for direct motion estimation in gated PET, inspired from OS-EM. This new approach provides significant acceleration as compared to the conventional approach proposed in previous work [5]. Further work includes: (i) evaluation of the variance and computational costs for different subset sizes and L-BFGS iterations; (ii) application of OS-JRM on TOF-PET/CT patient data; (iii) evaluation of the acceleration with misaligned attenuation map/PET gates.

## REFERENCES

- [1] M. W. Jacobson and J. A. Fessler, "Joint estimation of image and deformation parameters in motion-corrected PET," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, 2003, pp. 3290–3294.
- [2] M. W. Jacobson, "Approaches to motion-corrected PET image reconstruction from respiratory gated projection data," PhD thesis, The University of Michigan, Ann Arbor, USA, 2006.
- [3] M. Blume, A. Martinez-Müller, A. Keil, N. Navab, and M. Rafecas, "Joint reconstruction of image and motion in gated positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1892–1906, 2010.
- [4] G. Wang and J. Qi, "A monotonic image-space algorithm for joint PET image reconstruction and motion estimation," in *IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 2015.
- [5] A. Bousse, O. Bertolli, D. Atkinson, S. Arridge, S. Ourselin, B. F. Hutton, and K. Thielemans, "Maximum-likelihood joint image reconstruction/motion estimation in attenuation-corrected respiratory gated PET/CT using a single attenuation map," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 217–228, 2016.
- [6] M. Dawood, T. Kösters, M. Fieseler, F. Büther, X. Jiang, F. Wübbeling, and K. P. Schäfers, "Motion correction in respiratory gated cardiac PET/CT using multi-scale optical flow," *Med. Image Comput. Comput. Assist. Interv.*, vol. 11, no. 2, pp. 155–162, 2008.
- [7] F. Qiao, T. Pan, J. W. Clark, and O. R. Mawlawi, "A motion-incorporated reconstruction method for gated PET studies," *Phys. Med. Biol.*, vol. 51, pp. 3769–3783, 2006.
- [8] R. Manber, K. Thielemans, B. F. Hutton, A. Barnes, S. Ourselin, S. Arridge, C. O'Meara, S. Wan, and D. Atkinson, "Practical PET respiratory motion correction in clinical PET/MR," *J. Nucl. Med.*, vol. 56, no. 6, pp. 1–6, 2015.
- [9] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, 1982.
- [10] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–316, 1984.
- [11] H. M. Hudson and R. S. Larkin, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, 1994.
- [12] Y. Otani, I. Fukuda, N. Tsukamoto, Y. Kumazaki, H. Sekine, E. Imabayashi, O. Kawaguchi, T. Nose, T. Teshima, and T. Dokiya, "A comparison of the respiratory signals acquired by different respiratory monitoring systems used in respiratory gated radiotherapy," *Med. Phys.*, vol. 37, no. 2, pp. 6178–6186, 2010.
- [13] K. Thielemans, S. Rathore, F. Engbrant, and P. Razifar, "Device-less gating for PET/CT using PCA," in *IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 2011, pp. 3904–3910.
- [14] J. Xu and B. M. W. Tsui, "Improved intrinsic motion detection using time-of-flight PET," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 2131–2145, 2015.
- [15] J. Jiao, A. Bousse, K. Thielemans, N. Burgos, P. Weston, P. Markiewicz, J. Schott, D. Atkinson, S. Arridge, B. F. Hutton, and S. Ourselin, "Direct parametric reconstruction with joint motion estimation/correction for dynamic brain PET data," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 203–213, 2017.
- [16] R. H. Byrd, P. Lu, and J. Nocedal, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Stat. Comp.*, vol. 16, no. 5, pp. 1190–1208, 1995.
- [17] A. Bousse, O. Bertolli, D. Atkinson, S. Arridge, S. Ourselin, B. F. Hutton, and K. Thielemans, "Maximum-likelihood joint image reconstruction and motion estimation with misaligned attenuation in TOF-PET/CT," *Phys. Med. Biol.*, vol. 61, no. 3, pp. L11–19, 2016.