A Bregman Majorization-Minimization Framework for PET Image Reconstruction

Claire Rossignol, Florent Sureau, Émilie Chouzenoux, Claude Comtat, Jean-Christophe Pesquet

▶ **To cite this version:**

Claire Rossignol, Florent Sureau, Émilie Chouzenoux, Claude Comtat, Jean-Christophe Pesquet. A Bregman Majorization-Minimization Framework for PET Image Reconstruction. ICIP 2022 - 29th IEEE International Conference on Image Processing, Oct 2022, Bordeaux, France. hal-03720547

HAL Id: hal-03720547
https://hal.science/hal-03720547v1

Submitted on 12 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A BREGMAN MAJORIZATION-MINIMIZATION FRAMEWORK FOR PET IMAGE RECONSTRUCTION

Claire Rossignol[⋆]*, Florent Sureau*[†]*, Emilie Chouzenoux*[⋆]*, Claude Comtat*[†]*, Jean-Christophe Pesquet*[⋆]

[⋆] CVN, Inria, CentraleSupélec, University Paris Saclay, France
[†] BioMaps, Université Paris-Saclay, CEA, CNRS, Inserm, SHFJ, France

ABSTRACT

Positron emission tomography (PET) is a quantitative imaging modality widely used in oncology, neurology, and pharmacology. The data acquired by a PET scanner correspond to projections of the concentration activity, assumed to follow a Poisson distribution. The reconstruction of images from tomographic projections corrupted by Poisson noise is a challenging ill-posed large-scale inverse problem. Several available solvers use the majorization-minimization (MM) principle, though relying on various construction strategies with a lack of unifying framework. This work fills the gap by introducing the concept of Bregman majorization. This leads to a unified view of MM-based methods for image reconstruction in the presence of Poisson noise. From this general approach, we exhibit three algorithmic solutions and compare their computational efficiency on a problem of dynamic PET image reconstruction, either using GPU or CPU processing.

***Index Terms*—** PET image reconstruction, Poisson noise, Majorization-Minimisation, Bregman divergence, GPU implementation.

1. INTRODUCTION

Positron emission tomography (PET) is a quantitative medical imaging modality providing information on various physiological processes (e.g. tumor uptake in oncology, concentration of receptors, binding constant in pharmacology, etc). During PET acquisition, a radiotracer emits positrons that annihilate in the body of the patient with an electron, resulting in the emission of two back-to-back photons which can then be detected in coincidence by scintillation crystals located around the patient. The measured count data in PET is modeled as a Poisson random variable with mean related to the projection of the activity concentration [1]. Reconstructing images of this activity concentration (or related physiological parameters) is therefore an ill-posed tomographic inverse problem.

Typical PET image reconstruction relies on the minimization of a penalized cost function, gathering a data fidelity term accounting for the Poisson nature of the noise and a regularization term incorporating priors on the sought image. Probably the most commonly used PET reconstruction algorithm in clinical routine is ML-EM (Maximum-Likelihood Expectation Maximization) [2, 3] and its accelerated version OSEM (Ordered Subset Expectation Maximization) [4], which simply minimize the Poisson fidelity under positivity constraints. Penalized versions are also available [5, 6], typically assuming a smoothness prior on the gradients of the image. The extension of these approaches to the case of dynamic PET image reconstruction is explored in [7, 8, 9]. More recently, deep learning techniques have also been investigated [10, 11, 12] with the advantage of a fast running time when using GPU devices. Remarkably, most aforementioned PET reconstruction algorithms rely on the same optimization methodology, namely they are instances of MM (Majorization-Minimization) approaches. The MM paradigm, introduced in [13] relies upon the concept of majorant approximation. It has become very popular because of its ability to yield simple, efficient, and well sounded algorithms, and it is widely used in several fields of data science such as machine learning, image restoration, and signal processing [14].

In this paper, our contribution is twofold. First, we propose a general framework relying on the concept of Bregman majorization, that encompasses famous image MM-based algorithms for reconstruction from Poisson data. This includes a large variety of existing works [3, 5, 6, 10], and even yields new resolution schemes. Second, we select three possible choices for Bregman majorants and perform an extensive comparative analysis of the three associated algorithms in terms of computational efficiency. Namely, we present time execution comparisons on a problem of dynamic 2D PET image reconstruction. We evaluate both CPU and GPU implementations.

The paper is organized as follows. We present the PET image reconstruction inverse problem and the required mathematical background in Section 2. We then introduce in Section 3 our novel Bregman majorization framework and illustrate its applicability through three examples of interest. Section 4 evaluates these three solutions on the regularized reconstruction of a spatio-temporal PET sequence, and Section 5 provides some concluding remarks.

2. NOTATION AND BACKGROUND

2.1. Problem formulation

We consider the reconstruction of an image denoted by $\overline{x} \in [0,+\infty)^N$, from an observed noisy sinogram $y \in [0,+\infty)^M$ given by

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

where $H \in [0,+\infty)^{M\times N}$ is the projection matrix modelling the geometry of the acquisition system [13], $b \in (0,+\infty)^M$ is a strictly positive background term (assumed to be known) [13] and $\mathcal{P}(\cdot)$ models a Poisson noise corruption process. An efficient strategy to tackle the inverse problem of reconstructing $\overline{x}$ given y, b, and H, resorts to performing a maximum a posteriori estimate of $\overline{x}$ [13]. This amounts to solving the constrained optimization problem:

$$\underset{x\in[0,+\infty)^N}{\text{minimize}} \quad (f(x) = L(x) + R(x)). \tag{2}$$

This research work received funding support from the European Research Council Starting Grant MAJORIS ERC-2019-STG-850925.

Herabove, $L : \mathbb{R}^N \to (-\infty, +\infty]$ is the neg-log-likelihood function expressed as

$$(\forall x \in \mathbb{R}^N) \quad L(x) = \mathcal{KL}(Hx + b, y), \tag{3}$$

where $\mathcal{KL}$ is the Kullback-Leibler divergence defined, for every $(u, v) \in \mathbb{R}^2$, as

$$\mathcal{KL}(u, v) = \begin{cases} -v \log u + u & \text{if } u > 0 \text{ and } v > 0, \\ u & \text{if } u \geq 0 \text{ and } v = 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{4}$$

Moreover, $R : \mathbb{R}^N \to (-\infty, +\infty]$ is a regularization term that incorporates prior knowledge on the sought image.

2.2. Majorization-Minimization

Various methods have been proposed in the literature to solve Problem (2) [15, 4, 13, 10]. Most of them belongs to the class of majorization-minimization (MM) approaches [13, 16]. MM approaches are efficient iterative methods that minimize an objective function f through the minimization of a sequence of more tractable surrogates satisfying a majorizing condition. A function q_z is a tangent majorant of f at $z \in \mathbb{R}^N$ if, for every $x \in \mathbb{R}^N$, $f(x) \leq q_z(x) \quad \text{and} \quad f(z) = q_z(z)$. At each iteration $k \in \mathbb{N}$ of an MM algorithm, the new iterate x^{k+1} is the minimizer of the tangent majorant of f at the current iterate x^k. The MM algorithm then reads

$$(\forall k \in \mathbb{N}) \quad x^{k+1} \in \operatorname*{argmin}_{x \in \mathbb{R}^N} q_{x^k}(x), \tag{5}$$

with x^0 an initial vector belonging to the domain of the majorant function. Various algorithms can be built from this generic idea, depending on the structure of the majorant function [17]. This work brings a novel perspective on MM methods for PET reconstruction, recasting them within a Bregmanian framework.

2.3. Bregman divergences

A key mathematical tool in our approach is the Bregman divergence. Let us recall the definition of Bregman divergence, introduced in 1967 in [18]. Let $\phi \colon \mathbb{R}^N \to (-\infty, +\infty]$ be a strictly convex function, which is twice differentiable on the interior $\operatorname{int} \operatorname{dom} \phi$ of its domain. The Bregman divergence associated with ϕ is defined, for every $(x, z) \in (\operatorname{int} \operatorname{dom} \phi)^2$, as

$$D_\phi(x, z) = \phi(x) - \phi(z) - \nabla\phi(z)^\top (x - z). \tag{6}$$

In a nutshell, the quantity $D_\phi(x, z)$ compares the gap between the function ϕ at $x \in \mathbb{R}^N$, and its linear approximation around x evaluated at $z \in \mathbb{R}^N$.

3. PROPOSED APPROACH

In this section, we present the main methodological contribution of this paper. We show in Section 3.1 how Bregman divergences offer a both powerful and elegant framework for building tangent majorants for the objective function f involved in Problem (2), and then we derive in Section 3.2 a versatile MM algorithm to tackle this problem, so generalizing existing MM solvers from [10, 6].

3.1. A unifying Bregman majorization framework

Let us start by introducing the key notion of Bregman tangent majorant.

Definition 1. Let g be a differentiable function on an open set $\mathcal{D} \subset \mathbb{R}^N$ and let $z \in \mathcal{D}$. Let E_z be a strictly convex function which is differentiable on $\mathcal{D}$. Let D_{E_z} be the Bregman divergence associated with E_z. $q_z \colon \mathcal{D} \to \mathbb{R}$ is the Bregman tangent majorant of g at z associated with E_z if

$$(\forall x \in \mathcal{D}) \quad \begin{cases} q_z(x) = g(z) + \nabla g(z)^\top (x - z) + D_{E_z}(x, z) \\ g(x) \leq q_z(x). \end{cases} \tag{7}$$

This definition generalizes previous concepts used for instance in [19]. The main novelty here is that we allow the so-called Bregman metric D_{E_z} to vary with the reference point z. Up to the best of our knowledge, this has never been considered in the literature so far.

Starting from the above definition, we aim at building Bregman majorant functions of f in (2). Since f reads as the sum of two terms, we rely on the following useful additivity property.

Lemma 1. *Let g_1 and g_2 be differentiable functions on an open set $\mathcal{D} \subset \mathbb{R}^N$ and let $z \in \mathcal{D}$. Let $E_{1,z}$ (resp. $E_{2,z}$) be strictly convex functions, which are differentiable on $\mathcal{D}$, from which a Bregman tangent majorant of g_1 (resp. g_2) can be built. Let $(\alpha_1, \alpha_2) \in [0, +\infty)^2$, let $\theta \in \mathbb{R}^N$, and let $\eta \in \mathbb{R}$. Then a Bregman tangent majorant of $x \mapsto \alpha_1 g_1(x) + \alpha_2 g_2(x) + \theta^\top x + \eta$ is associated with $\alpha_1 E_{1,z} + \alpha_2 E_{2,z}$.*

It is therefore possible to construct a Bregman tangent majorant of our cost function f, by simple affine combination of a Bregman majorant of the data fidelity term L and the regularization term R. In the following, we present three examples of strategies for building a Bregman majorant function for L.

3.1.1. Bregman majorant functions for Poisson data fidelity

Let us first focus on the majorization of function L defined in (3). For every $m \in \{1, \ldots, M\}$, and $x \in \mathbb{R}^N$, let us define $\ell_m(x) = h_m^\top x + b_m > 0$, with $h_m \in \mathbb{R}^N$ the m-th row of H. Then, on the positive orthant, Function (3) takes a finite value, namely it can be rewritten as

$$(\forall x \in [0, +\infty)^N) \quad L(x) = \sum_{m=1}^{M} -y_m \log(\ell_m(x)) + \ell_m(x). \tag{8}$$

We first show how judicious choices for the underlying Bregman divergence allows us to derive suitable majorants leading to simple MM updates.

Example 1. Let us define, for every $z \in (0, +\infty)^N$, the function

$$(\forall x \in (0, +\infty)^N) \quad E_{1,z}(x) = -\sum_{n=1}^{N} a_{1,n}(z) \log x_n, \tag{9}$$

$$\text{where } (\forall n \in \{1, \ldots, N\}) \quad a_{1,n}(z) = \sum_{m=1}^{M} y_m \frac{h_{m,n} z_n}{\ell_m(z)}. \tag{10}$$

Then, it can be shown that the Bregman tangent majorant of L for $z \in (0,+\infty)^N$, which is associated to $E_{1,z}$ is

$$(\forall x = (x_n)_{1\leq n\leq N} \in (0,+\infty)^N)$$
$$q_{1,z}(x) = \sum_{m=1}^{M} \frac{b_m}{\ell_m(z)} \left(\ell_m(z) - y_m \log(\ell_m(z))\right)$$
$$+ \sum_{n=1}^{N}\sum_{m=1}^{M} \frac{h_{m,n} z_n}{\ell_m(z)} \left(\frac{\ell_m(z) x_n}{z_n} - y_m \log\left(\frac{\ell_m(z) x_n}{z_n}\right)\right). \quad (11)$$

We recover the separable majorant for L which was used to derive the recent FBSEM (Forward-Backward Splitting Expectation Maximisation) algorithm [10]. This algorithm is a regularised version of the MLEM algorithm from [15]. FBSEM is an MM algorithm whose majorant is built following a strategy initially introduced in [3].

Example 2. For every $z \in (0,+\infty)^N$, let

$$(\forall x \in [0,+\infty)^N) \quad E_{2,z}(x) = \sum_{n=1}^{N} a_{2,n}(z) \frac{x_n^2}{2}, \quad (12)$$

where, for every $n \in \{1,\dots,N\}$,

$$a_{2,n}(z) = \sum_{m=1}^{M} y_m \frac{h_{m,n}(z_n + \zeta_m b_m)}{\ell_m(z)} c(z_n,\rho). \quad (13)$$

Hereabove, $h_{m,n}$ stands for the (m,n)-th entry of matrix H, and we set

$$c(z_n,\rho) = \begin{cases} \dfrac{1}{\rho^2} & \text{if } z_n = 0 \\ -\dfrac{2}{z_n}\left(\dfrac{1}{z_n}\log\left(\dfrac{\rho}{z_n+\rho}\right) + \dfrac{1}{z_n+\rho}\right) & \text{otherwise,} \end{cases} \quad (14)$$

$$\text{with} \quad \begin{cases} \rho = \min_{1\leq m\leq M} \zeta_m b_m, \\ (\forall m = \{1,\dots,M\}) \quad \zeta_m = 1/\sum_{n=1}^{N} h_{m,n}. \end{cases} \quad (15)$$

The quantity $\sum_{n=1}^{N} h_{m,n}$ is assumed to be positive without loss of generality on H.

After some calculations, it can be proved that the Bregman tangent majorant of L at z associated to $E_{2,z}$ is

$$(\forall x \in (0,+\infty)^N)$$
$$q_{2,z}(x) = L(z) + \nabla L(z)^\top (x-z) + \sum_{n=1}^{N} \frac{\theta_n(z)}{2}(x_n - z_n)^2, \quad (16)$$

where, for every $n \in \{1,\dots,N\}$,

$$(\forall z \in (0,+\infty)^N) \quad \theta_n(z) = \sum_{m=1}^{M} c(z_n,\rho) \frac{h_{m,n}(z_n+\zeta_m b_m)}{\ell_m(z)}, \quad (17)$$

This quadratic separable majorant function can be obtained in a more classical manner by a clever combination of the majorant construction from [6] and Jensen's inequality.

The two previous examples have illustrated that majorants used in popular MM algorithms for image reconstruction under Poisson noise are instances of the proposed Bregman framework. It is also possible to construct new majorants, as shown hereafter.

Example 3. Let us construct a variant of (11). For every $z \in (0,+\infty)^N$, set

$$(\forall x \in (0,+\infty)^N) \quad E_{3,z}(x) = -\sum_{n=1}^{N} a_{3,n}(z) \log(x_n + \rho) \quad (18)$$

where, for every $n \in \{1,\dots,N\}$,

$$a_{3,n}(z) = \sum_{m=1}^{M} y_m \frac{h_{m,n}(z_n + \zeta_m b_m)}{\ell_m(z)}, \quad (19)$$

The same constants ρ and $(\zeta_m)_{1\leq m\leq M}$ as in the previous examples are used.

The Bregman tangent majorant of L at z then reads

$$(\forall x \in (0,+\infty)^N)$$
$$q_{3,z}(x) = L(z) + \nabla L(z)^\top (x-z) + D_{E_{3,z}}(x,z). \quad (20)$$

Remark 1. The majorant in the first example involves log terms and is thus valid only on $(0,+\infty)^N$. In turn, in the second and third examples, it is possible to extend the majorizing property to every $(x,z) \in [0,+\infty)^N$ by continuity arguments.

3.1.2. Bregman majorant of the complete loss function

We derived in the previous section three constructions of Bregman majorant functions for L. In general, we can now assume that L possesses a Bregman tangent majorant associated with some E_z and a domain of validity $\mathcal{D}$. There remains to majorize the regularization term R. Let us assume that R is differentiable on $\mathbb{R}^N$ with a β-Lipschitzian gradient, that is there exists $\beta > 0$ such that for every $(x,z) \in (\mathbb{R}^N)^2$, $\|\nabla R(x) - \nabla R(z)\| \leq \beta \|x - z\|$. It follows from the descent lemma [20] that a Bregman tangent majorant of R at every $z \in \mathbb{R}^N$ is associated to $\beta\|\cdot\|^2/2$. Therefore, by Lemma 1, a Bregman tangent majorant of the loss function f in Problem (2) at $z \in \mathcal{D}$ is associated with $\widetilde{E}_z = E_z + \beta\|\cdot\|^2/2$. In the following, we will denote $\widetilde{q}_z$ such a majorant.

3.2. One MM algorithm to rule them all

We are now ready to state our general Bregman MM algorithm to solve Problem (2). First, let us point out that all the three constructions presented in Sec. 3.1 can be recast in a single framework. More precisely, for every $(x,z) \in \mathcal{D}^2$, $f(x) \leq \widetilde{q}_z(x)$. The k-th iteration of the MM algorithm reads as the minimizer of $\widetilde{q}_z$ on $\mathcal{D}$. In all our constructions, $\widetilde{q}_z$ is a strictly convex separable function. Moreover, the components $(\widehat{x}_n(z))_{1\leq n\leq N}$ of its unique minimizer $\widehat{x}(z)$ satisfy, for each $n \in \{1,\dots,N\}$,

$$-\frac{a_n^{(0)}(z)}{\widehat{x}_n(z) + \widetilde{\rho}} + (a_n^{(1)}(z) + \beta)\widehat{x}_n(z) + d_n(z) = 0, \quad (21)$$

$$\text{with} \quad d_n(z) = \frac{a_n^{(0)}(z)}{z_n + \widetilde{\rho}} - (a_n^{(1)}(z) + \beta) z_n + [\nabla R(z)]_n + \sum_{m=1}^{M} h_{m,n}\Big(1 - \frac{y_m}{h_m^\top z + b_m}\Big), \quad (22)$$

where $(a_n^{(0)}(z), a_n^{(1)}(z), \tilde{\rho})$ are some non-negative values depending on the selected majorant. The N problems in (21) have closed form

solutions, each of them being the root of second-order polynomial equation. The final algorithm then reads

$$x^{k+1} = \left(\max\{\widehat{x}_n(x^k), 0\}\right)_{1\leq n\leq N}, \tag{23}$$

initialized with $x^0 \in (0, +\infty)^N$. It is remarkable that, when $R = 0$ and the majorant from Example 1 is chosen, we retrieve the famous MLEM algorithm. If $R \neq 0$ and the same majorant is employed, the FBSEM method can be retrieved for a specific choice of its step-size.

4. EXPERIMENTAL RESULTS

In this section, we compare the three MM strategies derived previously, in a 2D+t simulation scenario of dynamic PET brain imaging.

4.1. Settings and dataset

We simulated a 2D 1 hour ^{18}F-FDG dynamical PET exam on a Biograph 6 TruePoint TrueV PET system (Siemens Healthcare, Erlangen, Germany), using a slice of the Zubal phantom [21]. We followed the set up of [22] to generate dynamic realistic projection data with an analytical simulator [23]. Generic projector and backprojector based on Joseph's method [24] were defined as operators for PyTorch and Tensorflow, using C++ and Cuda for the GPU version / C++ with OpenMP parallelization for multithreaded CPU. Each projection was only defined by the position of the two crystals involved, without assumption on a specific geometry (e.g. cylindrical system). For reconstruction, we used a mixed regularization term consisting of smoothed spatial TV and temporal Tikhonov regularisation, as in [25], which has indeed a Lipschitzian gradient. The dynamic exam was subdivided into 24 time frames reconstructed simultaneously with a voxel size of 2.2 mm $\times$ 2.2 mm, and took into account all corrections as in [22].

The GPU solution was run on a 32 GB NVIDIA Tesla V100-SXM2, with block size optimized to reduce computation time. A comparison was made with a multi-threaded CPU implementation using OpenMP and 52 threads run on 52 distinct 2.10GHz Intel(R) Xeon(R) Gold 6230R cores.

4.2. Numerical results

Algorithms 1-3 correspond to MM algorithm (23) based on the majorants (11), (16), (20), respectively. We show, in Fig. 1, the convergence speed of the algorithms in terms of the relative distance between the outer iterate (corresponding to x^k) and the limit point x_∞, approximated for each method by performing 10,000 iterations. One can observe that Algorithm 1 outperforms the two others in terms of convergence speed. Fig. 2 provides a visual translation of these differences in convergence speeds. These results can be explained by considering the number of projections or backprojections needed to minimize each surrogate. It turns out that the computational cost of each iteration is dominated by the time needed to perform a backprojection (0.98s for a GPU-based implementation versus only 0.0016s for a projection). One iteration of Algorithm 1 requires only one backprojection against two for Algorithm 3 and three for Algorithm 2. This is also reflected by Tab. 1 where Algorithm 2 appears to be more than 2.5 times slower than Algorithm 1. This table also shows that our GPU implementation leads to a reduction of computation time by a factor larger than 10 compared to the use of the multithreaded CPU with 52 cores. We have focused on the PyTorch implementation, but similar results have also been obtained with Tensorflow.

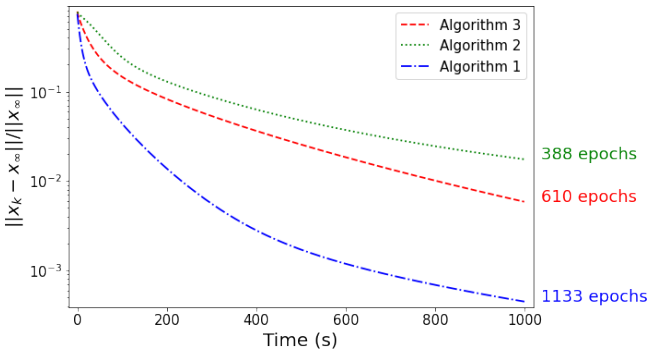

Fig. 1. Relative distance between the outer iterate and the limit point as a function of computational time, for GPU-based implementations.

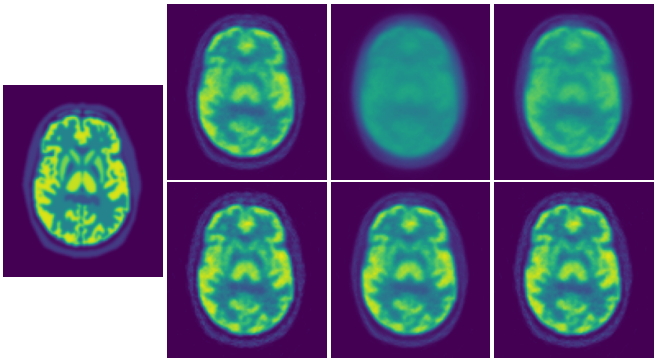

Fig. 2. Reference image (first column) ; Reconstructed images after 100 s (first row) and 1000 s (second row) of GPU computation time, for Algorithm 1 (second column), Algorithm 2 (third column) and Algorithm 3 (fourth column).

5. CONCLUSION

In this paper, we have proposed a new Bregman majorization framework which unifies MM approaches used for PET image reconstruction. In our PET simulations, three MM algorithms derived from specific choices of Bregman majorants have been compared in terms of convergence speed. We have also evaluated the gain in computation time resulting from a GPU implementation instead of a multithreaded CPU implementation. It is worthy to note that the proposed framework is quite general and also applies to many other inverse problems involving Poisson data. This work also opens perspectives in terms of optimization of the algorithm parameters through differential programming techniques on GPUs [10].

Implementation	Algorithm 1	Algorithm 2	Algorithm 3
PyTorch CPU	1351	3536	2246
PyTorch GPU	87	255	162

Table 1. Comparisons of execution time for CPU-based and GPU-based implementations, for 100 iterations, in seconds.

6. REFERENCES

[1] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.*, vol. 80, no. 389, pp. 8–20, Mar. 1985.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 39, no. 1, pp. 1–22, Sept. 1977.

[3] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 328–333, June 1993.

[4] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, 1994.

[5] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 1, pp. 132–137, Mar. 1995.

[6] H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 801–814, 1999.

[7] G. Wang and J. Qi, "Acceleration of the direct reconstruction of linear parametric images using nested algorithms," *Phys. Med. Biol.*, vol. 55, no. 5, pp. 1505–1517, Feb. 2010.

[8] J. C. Matthews, G. I. Angelis, F. A. Kotasidis, P. J. Markiewicz, and A. J. Reader, "Direct reconstruction of parametric images using any spatiotemporal 4D image based model and maximum likelihood expectation maximisation," in *Proc. IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, Knoxville, USA, Oct. 2010, pp. 2435–2441.

[9] N. Pustelnik, C. Chaux, J.-C. Pesquet, and C. Comtat, "Parallel algorithm and hybrid regularization for dynamic PET reconstruction," in *IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, Knoxville, USA, Oct. 2010, pp. 2423–2427.

[10] A. Mehranian and A. J. Reader, "Model-based deep learning PET image reconstruction using Forward-Backward Splitting Expectation-Maximization," *IEEE Trans. Radiat. and Plasma Med. Sci.*, vol. 5, no. 1, pp. 54–64, Jan. 2021.

[11] K. Gong, J. Guan, K. Kim, X. Zhang, J. Yang, Y. Seo, G. E. Fakhri, J. Qi, and Q. Li, "Iterative PET image reconstruction using Convolutional Neural Network representation," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 675–685, Mar. 2019.

[12] K. Gong, C. Catana, J. Qi, and Q. Li, "PET image reconstruction using Deep Image Prior," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1655–1665, July 2019.

[13] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics in Applied Mathematics, 1970.

[14] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Proc.*, vol. 65, no. 3, pp. 794–816, 2016.

[15] K. L. Lange and R. E. Carson, "EM reconstruction algorithms for emission and transmission tomography.," *J. Comput. Assist. Tomo.*, vol. 8 2, pp. 306–16, 1984.

[16] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[17] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2411–2422, 2007.

[18] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. & Math. Phys.*, vol. 7, no. 3, pp. 200–217, Jan. 1967.

[19] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, "Bregman monotone optimization algorithms," *SIAM J. Control Optim.*, vol. 42, no. 2, pp. 596–636, 2003.

[20] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.

[21] I. G. Zubal, C. R. Harrell, E. O. Smith, Z. Rattner, G. Gindi, and P. B. Hoffer, "Computerized three-dimensional segmented human anatomy," *Med. Phys.*, vol. 21, no. 2, pp. 299–302, Feb. 1994.

[22] Z. Chalampalakis, S. Stute, M. Filipović, F. Sureau, and C. Comtat, "Use of dynamic reconstruction for parametric Patlak imaging in dynamic whole body PET," *Phys. Med. Biol.*, vol. 66, no. 18, pp. 185017, Sept. 2021.

[23] S. Stute, C. Tauber, C. Leroy, M. Bottlaender, V. Brulon, and C. Comtat, "Analytical simulations of dynamic PET scans with realistic count rates properties," in *Proc. IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, San Diego, USA, Oct. 2015, pp. 1–3.

[24] P. M. Joseph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Trans. Med. Imag.*, vol. 1, no. 3, pp. 192–196, Nov. 1982.

[25] S. Cadoni, E. Chouzenoux, J.-C. Pesquet, and C. Chaux, "A block parallel majorize-minimize memory gradient algorithm," in *Proc. IEEE Internat. Conf. Image Proc.*, Phoenix, USA, Sep. 2016, pp. 3194–3198.