

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

Faster PET reconstruction with a stochastic primal-dual hybrid gradient method

Matthias J. Ehrhardt, Pawel Markiewicz, Antonin Chambolle, Peter Richtárik, Jonathan Schott, et al.

Matthias J. Ehrhardt, Pawel Markiewicz, Antonin Chambolle, Peter Richtárik, Jonathan Schott, Carola-Bibiane Schönlieb, "Faster PET reconstruction with a stochastic primal-dual hybrid gradient method," Proc. SPIE 10394, Wavelets and Sparsity XVII, 103941O (24 August 2017); doi: 10.1117/12.2272946

SPIE.

Event: SPIE Optical Engineering + Applications, 2017, San Diego, California, United States

Faster PET Reconstruction with a Stochastic Primal-Dual Hybrid Gradient Method*

Matthias J. Ehrhardt^a, Pawel Markiewicz^b, Antonin Chambolle^c, Peter Richtárik^{d,e,f},
Jonathan Schott^g, and Carola-Bibiane Schönlieb^a

^aDepartment for Applied Mathematics and Theoretical Physics, University of Cambridge, UK

^bTranslational Imaging Group, CMIC, Department of Medical Physics & Biomedical
Engineering, University College London, UK

^cCMAP, Ecole Polytechnique, CNRS, France

^dVisual Computing Center & Extreme Computing Research Center, KAUST, Thuwal, Saudi
Arabia

^eSchool of Mathematics, University of Edinburgh, UK

^fThe Alan Turing Institute, London, UK

^gDementia Research Centre, Institute of Neurology, University College London, UK

ABSTRACT

Image reconstruction in positron emission tomography (PET) is computationally challenging due to Poisson noise, constraints and potentially non-smooth priors—let alone the sheer size of the problem. An algorithm that can cope well with the first three of the aforementioned challenges is the primal-dual hybrid gradient algorithm (PDHG) studied by Chambolle and Pock in 2011. However, PDHG updates all variables in parallel and is therefore computationally demanding on the large problem sizes encountered with modern PET scanners where the number of dual variables easily exceeds 100 million. In this work, we numerically study the usage of SPDHG—a stochastic extension of PDHG—but is still guaranteed to converge to a solution of the deterministic optimization problem with similar rates as PDHG. Numerical results on a clinical data set show that by introducing randomization into PDHG, similar results as the deterministic algorithm can be achieved using only around 10 % of operator evaluations. Thus, making significant progress towards the feasibility of sophisticated mathematical models in a clinical setting.

Keywords: positron emission tomography, PET, total variation, convex optimization, primal-dual algorithms, saddle point problems, stochastic optimization, inverse problems

1. INTRODUCTION

Positron emission tomography (PET) is an important medical imaging modality in many medical fields such as oncology and neurology as it allows monitoring of body functions.^{1,2} PET image reconstruction can be formulated as a convex optimization problem which is computationally challenging due the optimization set-up that includes Poisson noise, constraints and potentially non-smooth priors. An algorithm that can cope well with the aforementioned challenges is the primal-dual hybrid gradient algorithm (PDHG)³⁻⁶ which can decompose the problem into a sequence of easy operations such as matrix-vector products and proximal operators that have closed-form solutions. However, as the PDHG updates are parallel, every iteration of PDHG is computationally demanding for the large variable sizes (easily exceeding 100 million) encountered with modern PET scanners.

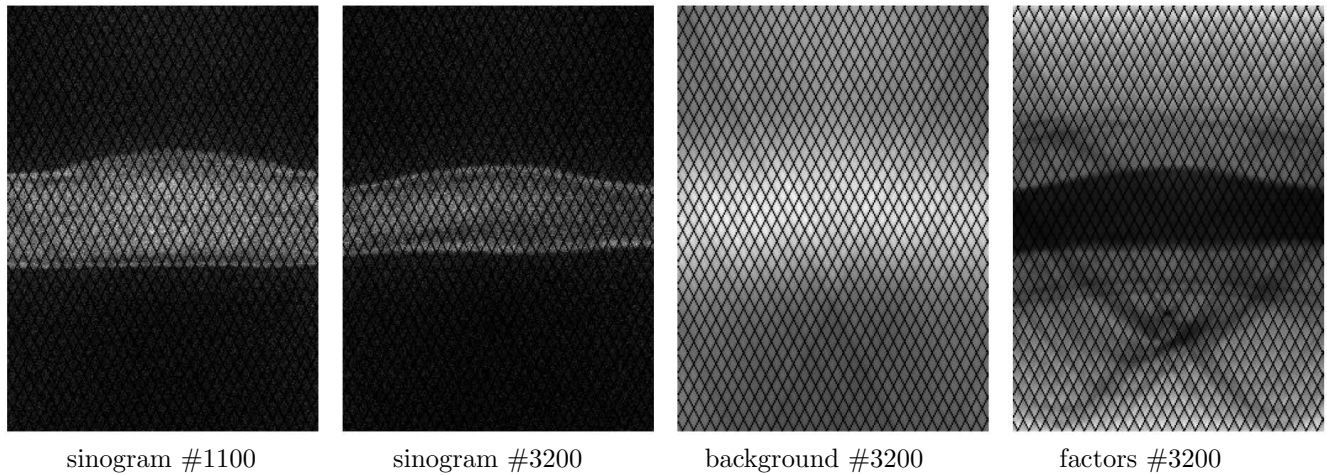


Figure 1. PET data from the Siemens Biograph mMR in span-1 consists of 4084 sinograms (two shown) each of size 252×344 . Next to it are the background and the factors needed to reconstruct from the data for the 3200th sinogram.

In this work, we numerically study the usage of SPDHG⁷—a stochastic extension of the PDHG that makes use of the dual separability of the problem (cf. e.g.^{8–11}). Instead of updating all dual variables in parallel, in each iteration only a randomly selected subset of the dual variables are updated. Despite its randomness, it is still guaranteed to converge to a solution of the deterministic optimization problem with similar rates as PDHG.⁷ Conceptually, the randomness in the dual update is similar to the Kaczmarz algorithm^{12–15} (also called ART¹⁶), stochastic gradient descent and surrogate approaches with subsets.^{17,18} Unlike other methods,^{19–21} SPDHG does not rely on any consistency conditions on the system of linear equations nor a “good” subset selection for convergence. This robustness comes from the fact that—as in incremental optimization algorithms^{22–26}—we progressively build up information from previous iterations which reduces variance. An interesting feature of SPDHG is that it can be run with any iteration-independent sampling.

Over the course of the last years, also other random variants of PDHG have been proposed,^{8–10,27–32} some being special cases of SPDHG.^{8–10} A few comparisons have been made on small simulated data.⁷ An extensive comparison of these methods is out of the scope of this contribution.

2. MATHEMATICAL MODEL FOR PET RECONSTRUCTION

In this section we briefly describe the optimization model used for PET reconstruction.

Before the actual PET scan is performed, a radioactive tracer is injected into the patient.^{1,2} The radioactive tracer emits positrons which in turn annihilate with free electrons in the human tissue resulting in two photons traveling in (almost) opposite directions and measured by rings of detectors around the patient. Assuming the directions to be exactly opposite, the detected photons yield information about the line integral of the tracer along the line that connects the detectors that recorded these photons. The reconstruction task in PET is to estimate the distribution of the tracer given more than a billion of these events (example taken from a data set acquired on the Siemens Biograph mMR).

Mathematically, the reconstruction task is usually posed as the maximum a-posteriori estimation^{1,2} of an image given the measured data which is equivalent to the minimization problem

$$\min_{x \in \mathbb{X}} \left\{ \sum_{i=1}^N \varphi_i(a_i^T x) + \psi(x) \right\} \quad (1)$$

where for each $i \in \{1, \dots, N\}$, the function $\varphi_i : \mathbb{R} \rightarrow \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}$ (also called Kullback–Leibler divergence)

$$\varphi_i(y) = \begin{cases} y + r_i - b_i + b_i \log\left(\frac{b_i}{y+r_i}\right) & \text{if } y + r_i > 0 \\ \infty & \text{else,} \end{cases} \quad (2)$$

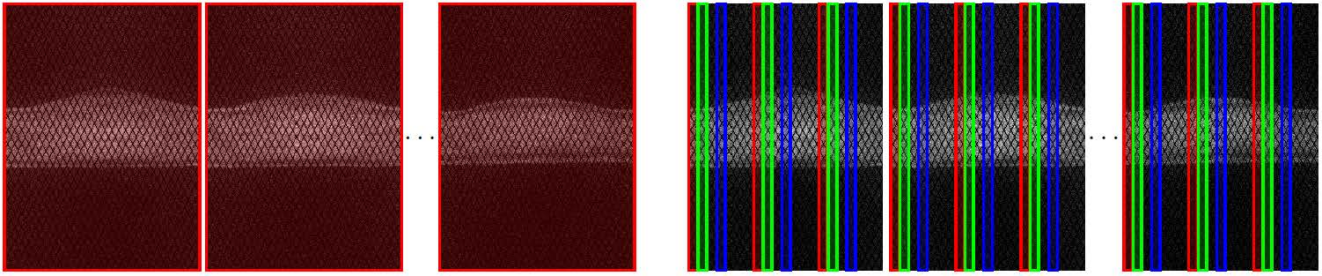


Figure 2. The selection of subsets is a free parameter in the reconstruction. On the left one subset (marked in red) has been chosen, such that all parts of all sinograms will be used in every iteration. Displayed on the right is the situation with seven subsets (three of which are highlighted in red, green and blue), where for each sinogram the angles (columns in this picture) are uniformly divided.

measures the deviance from the estimated line integral $a_i^T x + r_i$ to the measured data $b_i \in \mathbb{N}$ in a Poisson sense where it is convention to define $0 \log 0 := 0$. The functions φ_i are a distance measure, in the sense that $\varphi_i(y) \geq 0$ and $\varphi_i(y) = 0$ if and only if $a_i^T y + r_i = b_i$. Note that the φ_i are convex but not strictly or strongly convex and not differentiable (on the whole domain \mathbb{R}).

The vectors $a_i = \eta_i w_i$ do not only contain the vector of coefficients for the line integral between the i th detector pair w_i , but also contain weights $\eta_i > 0$ to compensate for attenuation and detector efficiency. The background $r_i > 0$ is estimated from the data prior to reconstruction to compensate for non-linear effects such as scatter and randoms. Figure 1 shows parts of the data y_i , together with the factors η_i and the background r_i .

The space of all possible (discrete) tracer distributions is denoted by $\mathbb{X} = \mathbb{R}^d := \mathbb{R}^{d_1 \times d_2 \times d_3}$. The function $\psi : \mathbb{X} \rightarrow \mathbb{R}_\infty$ models the a-priori knowledge about the desired solution such as constraints or regularity. It is standard to assume the tracer is non-negative. Moreover, we model the tracer's regularity with the total variation.^{33,34} Thus, the prior function ψ becomes

$$\psi(x) = \lambda \|\nabla x\|_{1,2} + \iota_{\geq 0}(x) \quad (3)$$

where $\lambda > 0$ denotes the regularization parameter, and the gradient operator $\nabla x \in \mathbb{X}^3$ discretized by forward differences, cf. e.g.⁵ for details. The 1, 2-norm of these gradients is defined as $\|y\|_{1,2} := \sum_\alpha \sqrt{\sum_{i=1}^3 (\nabla_i y_\alpha)^2}$ and the constraint function as

$$\iota_{\geq 0}(x) := \begin{cases} 0 & \text{if } x \geq 0 \\ \infty & \text{else} \end{cases}. \quad (4)$$

For computational efficiency, it is usually desirable to reformulate problem (1) and to group measurements together into so-called “subsets” $B_i \subset \{1, \dots, N\}$, $i = 1, \dots, n-1$ that are mutually disjoint $B_i \cap B_j = \emptyset$ if $i \neq j$ and cover all measurements $\cup_{i=1}^{n-1} B_i = \{1, \dots, N\}$. Two examples of subset selection is shown in Fig. 2. While “subsets” is a medical imaging terminology, this is often being referred to as “mini-batches” in machine learning. We then define the measurement operator for a subset i as $\mathbf{A}_i := [a_j]_{j \in B_i}$ where the vectors a_j form the rows of \mathbf{A}_i and the data fit as

$$f_i(\mathbf{A}_i x) := \sum_{j \in B_i} \varphi_i(a_j^T x). \quad (5)$$

With $f_n(y) := \lambda \|y\|_{1,2}$, $\mathbf{A}_n := \nabla$, $g(x) := \iota_{\geq 0}(x)$, we can rewrite (1) into the standard form

$$\min_{x \in \mathbb{X}} \left\{ \sum_{i=1}^n f_i(\mathbf{A}_i x) + g(x) \right\}. \quad (6)$$

3. STOCHASTIC PRIMAL-DUAL HYBRID GRADIENT METHOD

The following exposition holds true in general, and not only for the special case of PET image reconstruction. We only assume that the $f_i, i = 1, \dots, n$ and g are proper, convex, and lower semi-continuous. Under those assumptions,

$$f_i(z_i) = f_i^{**}(z_i) = \sup_{y_i} \langle z_i, y_i \rangle - f_i^*(y_i) \quad (7)$$

where f_i^* is the Fenchel conjugate of f_i (and f_i^{**} its biconjugate defined as the conjugate of the conjugate).³⁵ Thus, instead of solving the optimization problem (6) directly, it can be reformulated as a saddle point problem

$$\min_x \sup_{y_1, \dots, y_n} \left\{ \sum_{i=1}^n \langle \mathbf{A}_i x, y_i \rangle - f_i^*(y_i) + g(x) \right\}. \quad (8)$$

A very popular algorithm to solve saddle point problems of the form (8) is the primal-dual hybrid gradient (PDHG)³⁻⁶ which reads (with extrapolation on the dual variable y_i)

$$x^{k+1} = \text{prox}_{\tau g} \left(x^k - \tau \sum_{i=1}^n \mathbf{A}_i^* \bar{y}_i^k \right) \quad (9)$$

$$y_i^{k+1} = \text{prox}_{\sigma f_i^*} (y_i^k + \sigma \mathbf{A}_i x^{k+1}) \quad i = 1, \dots, n \quad (10)$$

$$\bar{y}_i^{k+1} = y_i^{k+1} + \theta (y_i^{k+1} - y_i^k) \quad i = 1, \dots, n. \quad (11)$$

It consists of simple operations like matrix-vector product and seemingly complicated operations like evaluations of *proximal operators* which are defined as

$$\text{prox}_{\tau f}(y) := \arg \min_x \left\{ \frac{1}{2} \|x - y\|^2 + \tau f(x) \right\}. \quad (12)$$

However, for many problems—including PET reconstruction as established in the previous section—these proximal operators can be solved in closed-form and are separable,⁵⁻⁷ thus, allowing for efficient parallel implementation. The PDHG is guaranteed to converge if the step size parameters σ, τ are positive and satisfy $\sigma \tau \|\mathbf{A}\|^2 < 1$ and $\theta = 1$.^{5,6}

While the PDHG is computationally attractive as it only needs simple operations, in every iteration many of those have to be executed. For instance, every iteration needs the evaluation of all operators \mathbf{A}_i and their adjoints \mathbf{A}_i^* for $i = 1, \dots, n$. To overcome this issue, a randomized variant of PDHG has recently been introduced⁷ depicted in Algorithm 1. In each iteration, instead of all n variables y_i , only one of those is updated. The sampling of those variables is allowed to be arbitrary as long as the probability p_i —that i is selected during each iteration—is positive and constant during the iterations. Such a sampling is called “serial” and “proper”.^{36,37} Other samplings are possible, too, and convergence for a much wider class can be proven,⁷ but for simplicity we stick to this special case. Note that SPDHG not only allows for a randomized update of the dual variables, but also allows the dual step size σ_i to be different for each variable y_i . Moreover, if $n = 1$ and the dual variable gets updated in each iteration, i.e. $p_1 = 1$, then SPDHG coincides with the deterministic PDHG. The following theorem establishes convergence of SPDHG in terms of the generalized Bregman distance

$$D_g^{\mathbf{g}^\#}(x, x^\#) := g(x) - g(x^\#) - \langle \mathbf{g}^\#, x - x^\# \rangle \quad (13)$$

where $\mathbf{g}^\# \in \partial g(x^\#)$ is an element of the subgradient of g at $x^\#$.

Algorithm 1 Stochastic Primal-Dual Hybrid Gradient algorithm (**SPDHG**). **Input:** primal and dual variable x^0, y^0 , step length parameters σ_i, τ, θ , probabilities p_i . **Initialize:** $z^0 = \mathbf{A}^* y^0, \bar{z}^0 = z^0$

for $k = 0, 1, \dots$ **do**
 $x^{k+1} = \text{prox}_{\tau g}(x^k - \tau \bar{z}^k)$
 Select $j \in \{1, \dots, n\}$ with probability $p_j > 0$.
 $y_i^{k+1} = \begin{cases} \text{prox}_{\sigma_i f_i^*}(y_i^k + \sigma_i \mathbf{A}_i x^{k+1}) & \text{if } i = j \\ y_i^k & \text{else} \end{cases}$
 $\Delta z^{k+1} = \mathbf{A}_j^*(y_j^{k+1} - y_j^k), \quad z^{k+1} = z^k + \Delta z^{k+1}, \quad \bar{z}^{k+1} = z^{k+1} + \frac{\theta}{p_j} \Delta z^{k+1}$
end for

Theorem 1 (Convergence of SPDHG⁷). Let $\gamma < 1$, the extrapolation parameter $\theta = 1$ and the step sizes τ, σ_i be chosen such that

$$\frac{\sigma_i \tau}{p_i} \|\mathbf{A}_i\|^2 \leq \gamma^2 \quad i = 1, \dots, n. \quad (14)$$

Then, the iterates (x^k, y^k) of SPDHG (Alg. 1) almost surely converge to a saddle point $(x^\#, y^\#)$ in a Bregman sense:

$$D_g^\#(x^k, x^\#) \rightarrow 0, \quad D_{f^*}^\#(y^k, y^\#) \rightarrow 0 \quad \text{a.s.} \quad (15)$$

Moreover, the ergodic sequence $(x_K, y_K) := \frac{1}{K} \sum_{k=1}^K (x^k, y^k)$ converges to a saddle point with rate $1/K$ in an expected Bregman sense:

$$\mathbb{E} \left\{ D_g^\#(x_K, x^\#) + D_{f^*}^\#(y_K, y^\#) \right\} \leq \frac{c}{K}. \quad (16)$$

For an explicit formula of the constant c , see.⁷

Remark. Stronger convergence rates—in line with the deterministic results^{5,6}—can be proven⁷ if more regularity is assumed on g and/or f_i .

4. NUMERICAL RESULTS

Software The numerical examples are implemented in python using numpy and the operator discretization library (ODL).³⁸ The ray tracing for the PET forward operator is performed using CUDA-based NiftyPET.³⁹

Data The clinical data are from a 70 year old cognitive normal subject, scanned as part of the “Insight 46”—a neuroscience sub-study of the MRC National Survey of Health and Development.⁴⁰ The PET data is accompanied by the T1-, T2-weighted MRI, all acquired simultaneously on a Siemens Biograph mMR PET-MRI scanner. The PET data was acquired dynamically for 60 mins using the amyloid tracer 18F-florbetapir tracer with activity of 409 MBq at the time of injection. The subject’s -map was synthesized from the T1- and T2-weighted images for better accuracy.⁴¹ The MRI was co-registered to the PET image using NiftyReg software.⁴² Randoms and scatter have been computed prior to reconstruction.^{39,43,44}

Sampling The sampling is chosen to be non-uniform to allow for the difference between the $n - 1$ subsets that correspond to the data and the subset that corresponds to the gradient. In each iteration we uniformly choose either data or gradient. In case the data is selected, then we again choose uniformly among the data subsets. Thus, overall the probability that subset i is selected is given by

$$p_i = \begin{cases} \frac{1}{2(n-1)} & \text{if } i < n \\ \frac{1}{2} & \text{if } i = n \end{cases}. \quad (17)$$

After $2(n - 1)$ iterations the SPDHG is expected to have evaluated all operators \mathbf{A}_i once, thus, having computed the same number of rays as the deterministic PDHG after only one iteration.

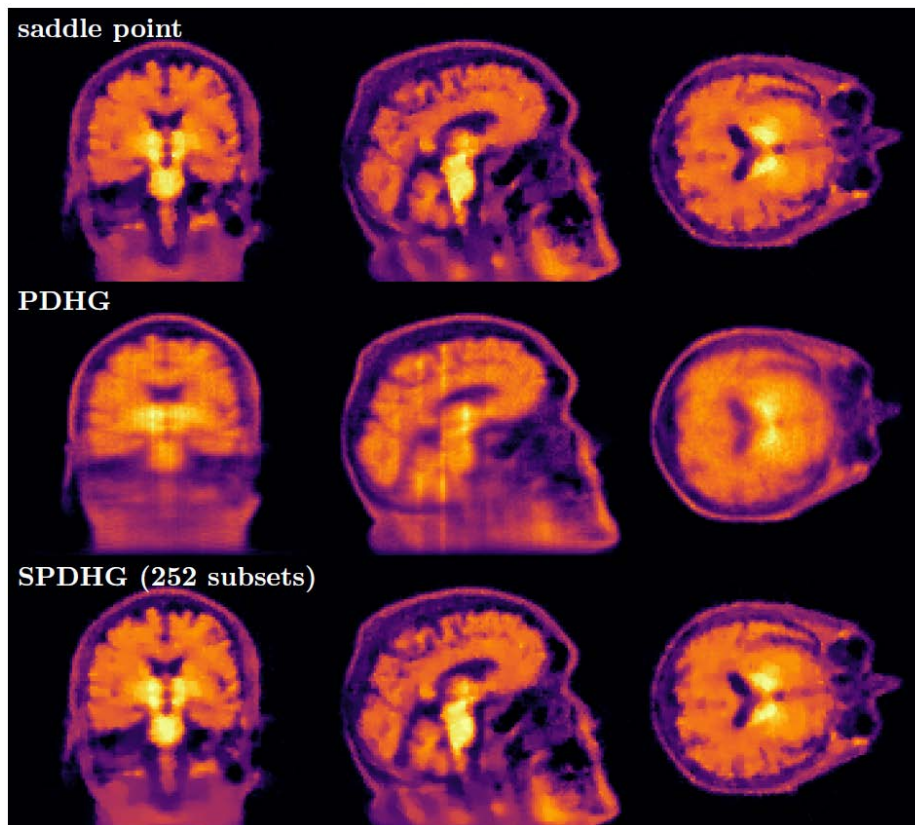


Figure 3. PET reconstruction with TV prior; results after 100 expected operator evaluations. While PDHG still has artifacts in the reconstructed image, SPDHG is visually already very close to the saddle point.

Parameters In this experiment we choose $\gamma = 0.99$, $\theta = 1$. The number of data subsets varied between 1 (deterministic case), 21 and 252. The other step size parameters are chosen for PDHG as

$$\sigma_1 = \frac{\gamma}{\|\mathbf{A}\|}, \quad \tau = \frac{\gamma}{\|\mathbf{A}\|} \quad (18)$$

and for SPDHG as

$$\sigma_i = \frac{\gamma}{\|\mathbf{A}_i\|} \quad i = 1, \dots, n, \quad \tau = \min_i \frac{\gamma p_i}{\|\mathbf{A}_i\|}. \quad (19)$$

The norms of the operators are estimated with 100 power iterations and allowing for an error of 5 %.

Results Some example reconstructed images are found in Figs. 3 and 4 quantitative results in Fig. 5. The saddle point for the visual and quantitative comparisons has been approximated with 2000 iterations of the deterministic PDHG. It can be seen from the reconstructed images after 100 expected operator evaluations in 3 that SPDHG is much faster than the deterministic PDHG. While the reconstruction with PDHG still has artifacts, the images that are reconstructed with SPDHG is visually very similar to the saddle point already after 100 iterations. A comparison of SPDHG with 21 and 252 subsets with PDHG after only 20 expected operator evaluations. The result of PDHG is clinically useless, as major anatomical structures are not yet visible. On the other hand, with more and more subsets, even with this little effort, a reasonable (although not perfect) image can be reconstructed. It is also easy to see, that the artifacts get reduced with more subsets indicating that with more subsets the algorithm converges faster. This is confirmed by the quantitative results in 5, where the same trend is observable for all four figures of merit.

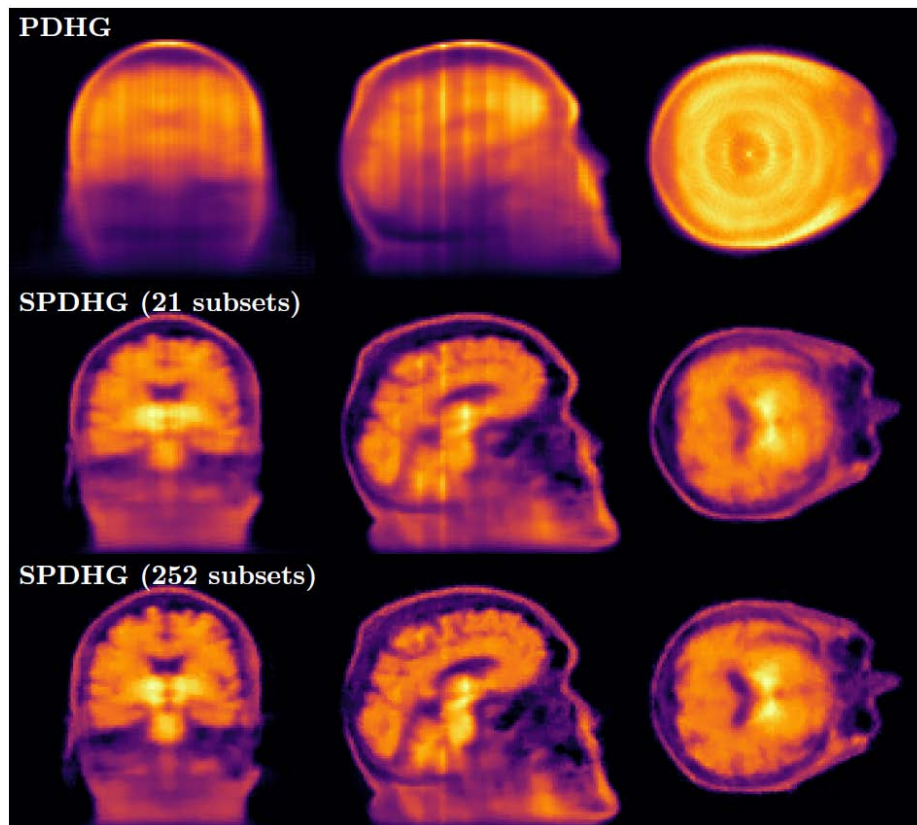


Figure 4. PET reconstruction with TV prior; results after 20 expected operator evaluations. With the same number of expected operator evaluations both variants of SPDHG make much more progress towards the saddle point than the deterministic PDHG.

5. CONCLUSIONS AND FUTURE WORK

A bottleneck in the translation of mathematical ideas into clinical applications in PET imaging has been the computational effort that modern algorithms need in order to solve the optimization problems. In this work we have numerically evaluated SPDHG—a stochastic variant of PDHG—to reconstruct images from clinical data with a total variation prior. As the results clearly indicated, by introducing randomness into the algorithm a similar performance as the deterministic variant may be achieved with only around 10 % of operator evaluations and thereby cutting the computational effort significantly. The algorithm is general enough to solve the optimization problem with other priors such as the directional total variation / parallel level sets,^{45,46} total generalized variation (TGV),^{47–49} directional TGV⁵⁰ or structure tensor-based total variation⁵¹ to name a few. In addition to PET reconstruction, SPDHG may help easing the computational effort in multi-modal medical imaging as well, where the PDHG has been used to jointly reconstruct from simultaneously acquired PET-MRI data with a nuclear norm variant of TGV⁵² or coupled Bregman iterations.⁵³ Similarly, a comparable speed up is expected in other imaging modalities that involve line integrals like computerized tomography (CT) where PDHG also found its applications.⁵⁴

Future work will encompass further numerical investigations of PET image reconstruction with different priors, other sampling strategies and step-size choices. While the randomness already allowed a significant speed up, we will investigate the use of smoothness and strong-convexity for algorithmic acceleration to increase the speed-up even further.

ACKNOWLEDGMENTS

M. J. E. and C.-B. S. acknowledge support from Leverhulme Trust project “Breaking the non-convexity barrier”, EPSRC grant “EP/M00483X/1”, EPSRC centre “EP/N014588/1”, the Cantab Capital Institute for the Math-

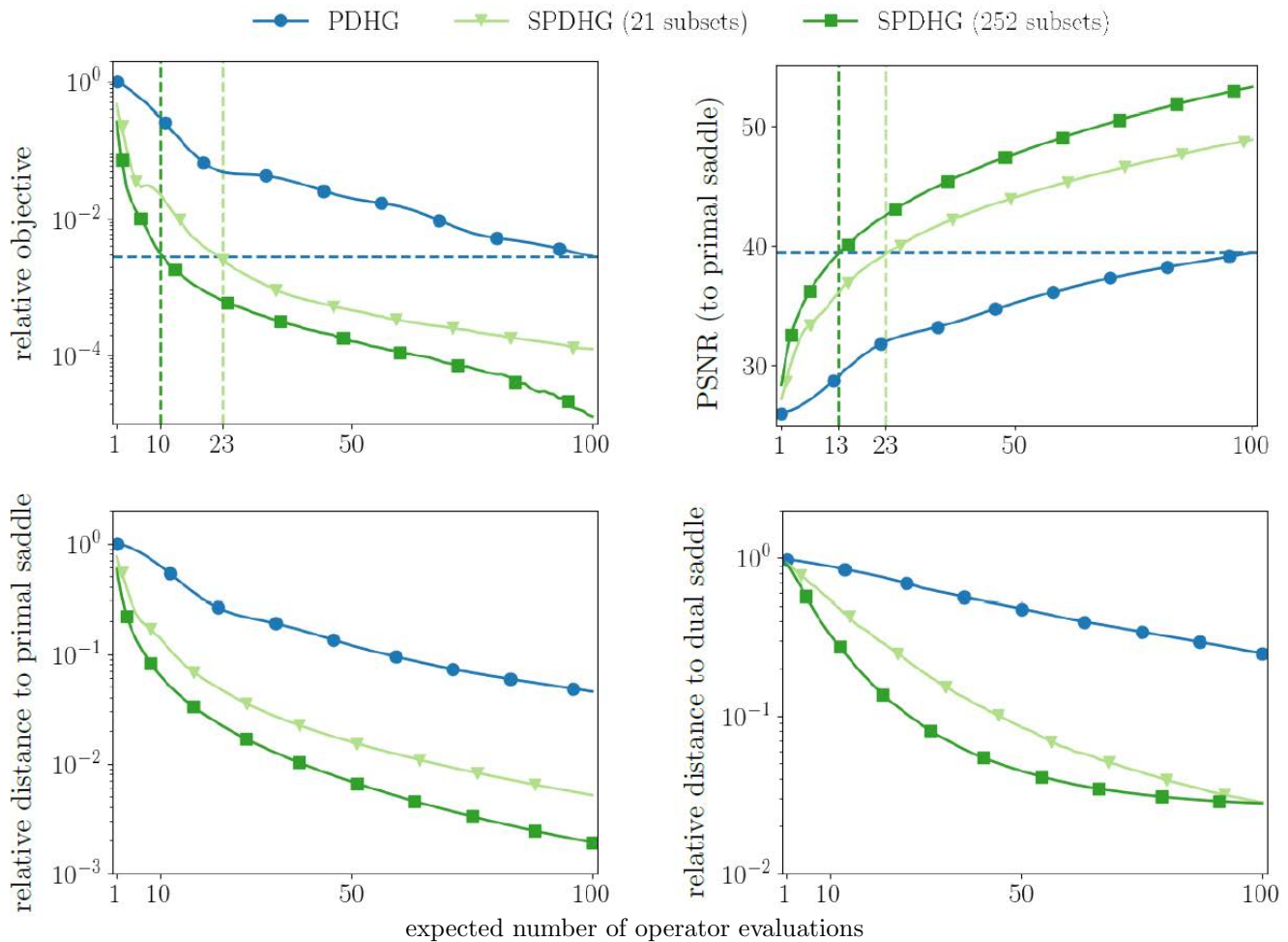


Figure 5. Quantitative evaluation of reconstruction algorithms. In all four measures, both variants of SPDHG are much faster than the deterministic PDHG. Moreover, with more subsets the algorithm becomes even faster.

ematics of Information, and from CHiPS (Horizon 2020 RISE project grant). Moreover, C.-B. S. is thankful for support by the Alan Turing Institute. P. M. was supported by the Medical Research Council (MR/N025792/1) and AMYPAD (European Commission project ID: ID115952, H2020-EU.3.1.7. - Innovative Medicines Initiative 2). A. C. benefited from a support of the ANR, “EANOI” Project I1148 / ANR-12-IS01-0003 (joint with FWF). Part of this work was done while he was hosted in Churchill College and DAMTP, Centre for Mathematical Sciences, University of Cambridge, thanks to a support of the French Embassy in the UK and the Cantab Capital Institute for Mathematics of Information. P. R. acknowledges the support of EPSRC Fellowship in Mathematical Sciences “EP/N005538/1” entitled “Randomized algorithms for extreme convex optimization”. The computations have been made on a GPU that was kindly made available by the NVIDIA GPU Grant Program. The Florbetapir PET tracer was provided by AVID Radiopharmaceuticals (a wholly owned subsidiary of Eli Lilly & Co).

REFERENCES

- [1] Ollinger, J. M. and Fessler, J. A., “Positron Emission Tomography,” *IEEE Signal Process Mag* **14**, 43–55 (1997).
- [2] Bailey, D. L., Townsend, D. W., Valk, P. E., and Maisey, M. N., [*Positron Emission Tomography - Basic Sciences*], Springer, London (2005).
- [3] Pock, T., Cremers, D., Bischof, H., and Chambolle, A., “An Algorithm for Minimizing the Mumford-Shah Functional,” in [*Proc IEEE ICCV*], (813396), 1133–1140 (2009).

- [4] Esser, E., Zhang, X., and Chan, T. F., “A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science,” *SIAM J Imaging Sci* **3**(4), 1015–1046 (2010).
- [5] Chambolle, A. and Pock, T., “A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging,” *J Math Imaging Vision* **40**(1), 120–145 (2011).
- [6] Chambolle, A. and Pock, T., “An Introduction to Continuous Optimization for Imaging,” *Acta Numerica* **25**, 161–319 (2016).
- [7] Chambolle, A., Ehrhardt, M. J., Richtárik, P., and Schönlieb, C.-B., “Stochastic Primal-Dual Hybrid Gradient Algorithm with Arbitrary Sampling and Imaging Application,” (2017).
- [8] Dang, C. D. and Lan, G., “Randomized Methods for Saddle Point Computation,” (2014).
- [9] Zhang, Y. and Xiao, L., “Stochastic Primal-Dual Coordinate Method for Regularized Empirical Risk Minimization,” in [*Proc ICML*], 1–34 (2015).
- [10] Zhu, Z. and Storkey, A. J., “Adaptive Stochastic Primal-Dual Coordinate Descent for Separable Saddle Point Problems,” in [*Proc ECML PKDD*], Appice, A., Rodrigues, P. P., Costa, V. S., Soares, C., Gama, J., and Jorge, A., eds., 643–657, Springer, Porto (2015).
- [11] Peng, Z., Wu, T., Xu, Y., Yan, M., and Yin, W., “Coordinate Friendly Structures, Algorithms and Applications,” *Annals of Mathematical Sciences and Applications* **1**(7), 1–54 (2016).
- [12] Kaczmarz, S., “Angenäherte Auflösung von Systemen linearer Gleichungen (Approximate Solution for Systems of Linear Equations),” *Bulletin International de l’Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles. Série A, Sciences Mathématiques* **35**, 355–357 (1937).
- [13] Strohmer, T. and Vershynin, R., “A Randomized Kaczmarz Algorithm with Exponential Convergence,” *J Fourier Anal Appl* **15**(2), 262–278 (2009).
- [14] Censor, Y., Herman, G. T., and Jiang, M., “A Note on the Behavior of the Randomized Kaczmarz Algorithm of Strohmer and Vershynin,” *J Fourier Anal Appl* **15**(4), 431–436 (2009).
- [15] Gower, R. M. and Richtárik, P., “Stochastic Dual Ascent for Solving Linear Systems,” (2015).
- [16] Gordon, R., Bender, R., and Herman, G. T., “Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography,” *J Theor Biol* **29**(3), 471–481 (1970).
- [17] Hsiao, I.-t., Rangarajan, A., and Gindi, G., “A Provably Convergent OS-EM Like Reconstruction Algorithm for Emission Tomography,” in [*Proc SPIE*], **4684**, 10–19 (2002).
- [18] Ahn, S. and Fessler, J. A., “Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms,” *IEEE Trans Med Imag* **22**(5), 613–626 (2003).
- [19] Hudson, H. M. and Larkin, R. S., “Accelerated Image Reconstruction Using Ordered Subsets of Projection Data,” *IEEE Trans Med Imag* **13**(4), 601–609 (1994).
- [20] Nien, H. and Fessler, J. A., “Relaxed Linearized Algorithms for Faster X-Ray CT Image Reconstruction,” *IEEE Trans Med Imag* **35**(4), 1090–1098 (2016).
- [21] Nien, H. and Fessler, J. A., “Fast X-ray CT Image Reconstruction using the Linearized Augmented Lagrangian Method with Ordered Subsets,” *IEEE Trans Med Imag* **34**(2), 21 (2015).
- [22] Tseng, P., “An Incremental Gradient(-Projection) Method with Momentum Term and Adaptive Stepsize Rule,” *SIAM J Optim* **8**(2), 506–531 (1998).
- [23] Blatt, D., Hero, A. O., and Gauchman, H., “A Convergent Incremental Gradient Method with a Constant Step Size,” *SIAM J Optim* **18**(1), 29–51 (2007).
- [24] Bertsekas, D. P., “Incremental proximal methods for large scale convex optimization,” *Math Prog* **129**(2), 163–195 (2011).
- [25] Schmidt, M., Le Roux, N., and Bach, F., “Minimizing Finite Sums with the Stochastic Average Gradient,” *Math Prog*, 1–30 (2016).
- [26] de Oliveira, R. M., Helou, E. S., and Costa, E. F., “String-Averaging Incremental Subgradients for Constrained Convex Optimization with Applications to Reconstruction of Tomographic Images,” *Inverse Probl* **32**(11), 115014 (2016).
- [27] Balamurugan, P. and Bach, F., “Stochastic Variance Reduction Methods for Saddle-Point Problems,” (2016).
- [28] Fercoq, O. and Bianchi, P., “A Coordinate Descent Primal-Dual Algorithm with Large Step Size and Possibly Non-Separable Functions,” (2015).

- [29] Pesquet, J.-C. and Repetti, A., “A Class of Randomized Primal-Dual Algorithms for Distributed Optimization,” (2015).
- [30] Valkonen, T., “Block-proximal methods with spatially adapted acceleration,” (2016).
- [31] Zhong, L. W. and Kwok, J. T., “Fast Stochastic Alternating Direction Method of Multipliers,” *J Mach Learn Res* **32**, 46–54 (2014).
- [32] Gao, X., Xu, Y., and Zhang, S., “Randomized Primal-Dual Proximal Block Coordinate Updates,” (2016).
- [33] Rudin, L. I., Osher, S., and Fatemi, E., “Nonlinear Total Variation based Noise Removal Algorithms,” *Physica D* **60**(1), 259–268 (1992).
- [34] Burger, M. and Osher, S., “A Guide to the TV Zoo,” in [*Level Set and PDE Based Reconstruction Methods in Imaging*], *Lecture Notes in Mathematics* **2090**, 1–70, Springer International Publishing, Cham (2013).
- [35] Bauschke, H. H. and Combettes, P. L., [*Convex Analysis and Monotone Operator Theory in Hilbert Spaces*] (2011).
- [36] Richtárik, P. and Takáč, M., “Parallel coordinate descent methods for big data optimization,” *Math Prog* **156**(1), 433–484 (2016).
- [37] Qu, Z. and Richtárik, P., “Quartz: Randomized Dual Coordinate Ascent with Arbitrary Sampling,” in [*Proc NIPS*], (1), 1–34 (2015).
- [38] Adler, J., Kohr, H., and Öktem, O., “Operator Discretization Library (ODL),” (2017).
- [39] Markiewicz, P. J., Ehrhardt, M. J., Erlandsson, K., Barnes, A., Schott, J. M., Atkinson, D., Arridge, S. R., Hutton, B. F., and Ourselin, S., “NiftyPET: A High-Throughput Software Platform for High Quantitative Accuracy and Precision PET Imaging and Analysis,” *submitted* (2017).
- [40] Lane, C. A., Parker, T. D., Cash, D. M., Macpherson, K., Donnachie, E., Murray-Smith, H., Barnes, A., Barker, S., Beasley, D. G., Bras, J., Brown, D., Burgos, N., Byford, M., Jorge Cardoso, M., Carvalho, A., Collins, J., De Vita, E., Dickson, J. C., Epie, N., Espak, M., Henley, S. M. D., Hoskote, C., Hutel, M., Klimova, J., Malone, I. B., Markiewicz, P., Melbourne, A., Modat, M., Schrag, A., Shah, S., Sharma, N., Sudre, C. H., Thomas, D. L., Wong, A., Zhang, H., Hardy, J., Zetterberg, H., Ourselin, S., Crutch, S. J., Kuh, D., Richards, M., Fox, N. C., and Schott, J. M., “Study protocol: Insight 46 – a neuroscience sub-study of the mrc national survey of health and development,” *BMC Neurology* **17**(1), 75 (2017).
- [41] Burgos, N., Cardoso, M. J., Thielemans, K., Modat, M., Dickson, J., Schott, J. M., Atkinson, D., Arridge, S. R., Hutton, B. F., and Ourselin, S., “Multi-Contrast Attenuation Map Synthesis for PET/MR Scanners: Assessment on FDG and Florbetapir PET tracers,” *EJNMMI* **42**(9), 1447–1458 (2015).
- [42] Modat, M., Cash, D. M., Daga, P., Winston, G. P., Duncan, J. S., and Ourselin, S., “Global Image Registration using a Symmetric Block-Matching Approach,” *Journal of Medical Imaging* **1**(2), 024003 (2014).
- [43] Markiewicz, P. J., Thielemans, K., Schott, J. M., Atkinson, D., Arridge, S. R., Hutton, B. F., and Ourselin, S., “Rapid Processing of PET List-Mode Data for Efficient Uncertainty Estimation and Data Analysis,” *Phys Med Biol* **61**(13), N322 (2016).
- [44] Markiewicz, P. J., Ehrhardt, M. J., Atkinson, D., Arridge, S. R., Hutton, B. F., and Ourselin, S., “Uniform Acquisition Modelling across PET Imaging Systems: Unified Scatter Modelling,” in [*Proc IEEE NSS-MIC*], (2016).
- [45] Ehrhardt, M. J., Markiewicz, P. J., Liljeroth, M., Barnes, A., Kolehmainen, V., Duncan, J., Pizarro, L., Atkinson, D., Hutton, B. F., Ourselin, S., Thielemans, K., and Arridge, S. R., “PET Reconstruction with an Anatomical MRI Prior using Parallel Level Sets,” *IEEE Trans Med Imag* **35**(9), 2189–2199 (2016).
- [46] Ehrhardt, M. J. and Betcke, M. M., “Multi-Contrast MRI Reconstruction with Structure-Guided Total Variation,” *SIAM J Imaging Sci* **9**(3), 1084–1106 (2016).
- [47] Bredies, K., Kunisch, K., and Pock, T., “Total Generalized Variation,” *SIAM J Imaging Sci* **3**(3), 492–526 (2010).
- [48] Bredies, K. and Holler, M., “A TGV-Based Framework for Variational Image Decompression, Zooming, and Reconstruction. Part I: Analytics,” *SIAM J Imaging Sci* **8**(4), 2814–2850 (2015).
- [49] Bredies, K. and Holler, M., “A TGV-Based Framework for Variational Image Decompression, Zooming, and Reconstruction. Part II: Numerics,” *SIAM J Imaging Sci* **8**(4), 2851–2886 (2015).
- [50] Kongskov, R. D., Dong, Y., and Knudsen, K., “Directional Total Generalized Variation Regularization,” (2017).

- [51] Estellers, V., Soatto, S., and Bresson, X., “Adaptive Regularization With the Structure Tensor,” *IEEE Trans Image Proc* **24**(6), 1777–1790 (2015).
- [52] Knoll, F., Holler, M., Koesters, T., Otazo, R., Bredies, K., and Sodickson, D. K., “Joint MR-PET Reconstruction using a Multi-channel Image Regularizer,” *IEEE Trans Med Imag* **0062**.
- [53] Rasch, J., Brinkmann, E.-M., and Burger, M., “Joint Reconstruction via Coupled Bregman Iterations with Applications to PET-MR Imaging,” (2017).
- [54] Rigie, D. and La Riviere, P., “Joint Reconstruction of Multi-Channel, Spectral CT Data via Constrained Total Nuclear Variation Minimization,” *Phys Med Biol* **60**, 1741–1762 (2015).