EFFECT of PRIOR SMOOTHING on the CONVERGENCE of PROXIMAL ALGORITHMS for PET and SPECT RECONSTRUCTION

Sam Porter, Daniel Deidda, *Member, IEEE*, Simon Arridge and Kris Thielemans, *Senior Member, IEEE*

***Abstract*—Proximal-based optimisation algorithms have been developed to be able to handle nondifferentiable functions. They have been widely studied for image reconstruction and denoising with priors such as the popular Total Variation (TV). Relatively little work has been done in evaluating their convergence performance with smooth priors that are more commonly used in emission tomography. We investigated the effect of varying the magnitude of a smoothing parameter for one image-based and one anatomical-based TV-like prior on the convergence rate of two proximal and two gradient-based algorithms for PET and SPECT reconstruction. The results suggest that the smoothness of a prior has less effect on the convergence rate for proximal algorithms than gradient-based algorithms. As expected, a smoother function results in faster convergence for gradient-based algorithms. A smoother function results in a slightly decreased convergence rate for proximal algorithms. Over-smoothing of the function resulted in under-regularisation and the breakdown of convergence observations.**

THE problem of reconstruction in tomography can be formulated in terms of an objective function (or loss function) to be solved by minimisation:

$$\mathcal{L}(u, g) = \sum_j \mathcal{D}(\mathcal{A}_{ij} u_i, g_j) + \sum_i \alpha \mathcal{R}(u_i) \tag{1}$$

$$u^* = \arg\min_{u \geq 0} \{\mathcal{L}(u, g)\} \tag{2}$$

where A is the forward acquisition model applied to an image, u, g is the acquired data, $\mathcal{D}(Au, g)$ is a data fidelity term that describes how well the estimated data fit the acquired data, and $\mathcal{R}(u)$ is a regularisation term that contains prior information about the image, such as smoothness or structural information.

Traditionally in emission tomography priors have been formulated to be differentiable and optimisation problems have been solved by iterative Expectation Maximisation or gradient-based algorithms such as gradient descent or Quasi-Newton methods that provide an image update using the gradient of the prior at the current and sometimes previous image estimates. For non-differentiable priors, proximal algorithms have been developed which repeatedly apply one or more proximal maps, defined for a function $f(x)$ as

$$\text{prox}_f^\tau(x') = \arg\min_x \left\{ f(x) + \frac{1}{2\tau} ||x' - x||_2^2 \right\} \tag{3}$$

where τ is the proximal step size and controls the size of the image update by forcing the new estimate to be close to the current estimate [1]. This mapping removes the requirement for differentiability and, provided that $\text{prox}_f^\tau(u)$ is easily found, can lead to fast reconstruction. This has led to increasing use of non-differentiable edge-preserving priors such as Total Variation (TV). However, these priors are often "smoothed", for example to prevent the stair-casing effect where reconstructed images are piecewise constant and thus appear unnatural.

Little work has been done in order to investigate the effect of this smoothing on the convergence of proximal algorithms. This may be important with the recent introduction of stochastic proximal algorithms such as SPDHG [2] into the emission tomography sphere. This work investigated the effect of increasing the smoothness of the TV approximating Huber prior [3] on the rate of convergence near convergence of two gradient-based and two proximal algorithms. We have evaluated this for both PET and SPECT as the corresponding inverse problems have different conditioning, leading to images with different resolution and noise properties.

I. Methods

A. Evaluation

The effect of smoothing of the prior on the rate of convergence near the final convergent image was chosen for two reasons:

1) We can use an inaccurate but fast algorithm, such as OSEM-TV (Section IV.A) used in this investigation, to get us close to the converged image. We can then warm start the convergent algorithm using this reconstructed image.
2) Changing the smoothing parameter alters the objective function and thus the solution to the minimisation problem, meaning that number of iterations to convergence, or similar measures, will not provide accurate results.

B. Data

Simulated 60x60 pixel two-dimensional emission (with uptake ratios roughly corresponding to an FDG-PET), attenuation

Manuscript received December 1, 2022. This work was supported in part by the National Physics Laboratory through the National Measurement System of the UK Department of Business, Energy, and Industrial Strategy. The software used in this project is maintained by CCP SyneRBI (EPSRC grant EP/T026693/1) and CCPi (EPSRC grant (EP/T026677/1).

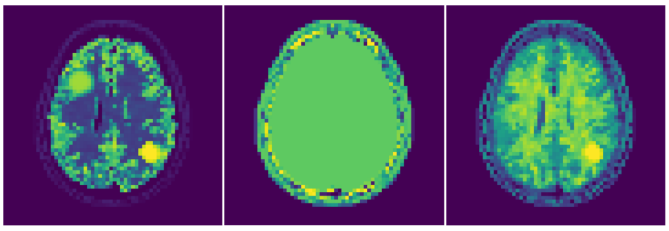
S. Porter is with the Institute of Nuclear Medicine, UCL, and the National Physical Laboratory, London, UK (e-mail: sam.porter.18@ucl.ac.uk).

D. Deidda is with the Institute of Nuclear Medicine, UCL, and the National Physical Laboratory, London, UK (e-mail: daniel.deidda@npl.co.uk).

K. Thielemans is with the Institute of Nuclear Medicine and the Centre for Medical Image Computing, UCL, London, UK (e-mail: k.thielemans@ucl.ac.uk).

S. Arridge is with the Centre for Inverse Problems, UCL, London, UK (e-mail: k.thielemans@ucl.ac.uk).

978-1-6654-8872-3/22/$31.00 ©2022 IEEE



(a) (b) (c)

Fig. 1: two-dimensional BrainWeb images. a) ground truth emission, b) attenuation image, c) T1-weighted MRI anatomical information

and T1-weighted magnetic resonance (MR) brain images were simulated using BrainWeb's brain database [4]. The emission and anatomical images were normalised to have a maximum pixel value of one. Two tumours were added to the image. One is present in both the emission and MRI images and one is present only in the emission image. The emission images were forward projected into data space using the Software for Tomographic Image Reconstruction (STIR) PET & SPECT projectors via the Synergistic Image Reconstruction Framework (SIRF) [5][1], including attenuation [6], [7]. A 2D PET scanner with similar characteristics as the Siemens mMR Biograph was chosen for PET projection with 504 detectors per ring, a 65.6 cm inner ring diameter and 126 projections. For the SPECT model, a 25.6 cm radius with 126 projections and a distant-dependent Gaussian PSF using a linear model (with $\sigma = 0.0163d + 0.1466cm$) was chosen. Poisson noise was added in the projection space. 10 noise realisations were used for the PET reconstructions and 1 noise realisation for the SPECT reconstructions due to time constraints.

C. Priors

The Huber function, defined as

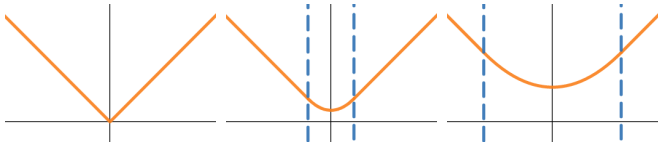
$$\mathcal{H}(\vec{x}) = \begin{cases} \frac{1}{2\theta}(\theta^2 + ||\vec{x}||_2^2) & ||\vec{x}||_2 \leq \theta \\ ||\vec{x}||_2 & ||\vec{x}||_2 > \theta \end{cases} \tag{4}$$

was used in this investigation (Fig. 2). Two variations, isotropic Huber TV (iHTV), where $\mathcal{R}(u) = \mathcal{H}(\nabla u)$, and anisotropic (or directional) Huber TV (dHTV), where $\mathcal{R}(u) = \mathcal{H}(Du)$. Here, D is a directional operator, defined as

$$D = (1 - \xi\xi^T), \quad \text{with} \quad \xi = \frac{\nabla v}{\sqrt{||\nabla v||_2^2 + \eta^2}}. \tag{5}$$

v is a registered image that contains structural information, such as the T1-weighted MR images used in this investigation, and η is a small constant to avoid division by zero in uniform regions. This operator has the effect of suppressing the prior when gradients in the emission and anatomical images align and reduces to Total Variation where the anatomical image is uniform [8].

[1]The same emission image was chosen for both PET and SPECT ground truth image to facilitate comparison of the performance of the algorithms for the two modalities.



(a) (b) (c)

Fig. 2: One-dimensional, one-pixel Huber functions. $\vec{x}$ on the horizontal axis and $\mathcal{H}(\vec{x})$ on the vertical axis. a) $\theta = 0$ (Total Variation), b) small θ, c) large θ

D. Algorithms

The effect of smoothing on the convergence rate of four algorithms was investigated. These were Preconditioned Gradient Projection with a back-tracking line search (PCGP) (Section IV.C), Low-memory Bounded Broyden-Fletcher-Goldfarb-Shannon (L-BFGS-B) (Section IV.B [9], 3-block Linearised Alternating Direction Method of Multipliers (L-ADMM) (Section IV.E) [1] and 3-block Primal-Dual Hybrid gradient (PDHG) (Section IV.D) [10].

E. Reconstruction Routine

1) Warm-start the reconstruction using 5 full iterations of 21-subset OSEM-TV.
2) Reconstruct the "converged" image used for comparison with 1000 iterations of L-BFGS-B. This is much more than is required for convergence to machine precision for two-dimensional images.
3) Reconstruct the image using the desired algorithm until the mean squared error (normalised to the maximum of the ground truth image) between the current image and the "converged" comparison image is less than a threshold. This was chosen to be 0.0001.
4) The line connecting the final 10 iterations when plotted on a graph of mean squared error against iteration was observed to be linear. The slope of this line was measured and was used as the rate near convergence (see Fig. 3),

$$C = \frac{\partial \text{ MSE}}{\partial \text{ iterations}}. \tag{6}$$

This value was then compared for different algorithms and smoothing parameters.

II. Results

A. Small smoothing parameter

For values of θ less than approximately 0.01 (the red dashed line in Fig. 4), gradient-based algorithms observed an increase in convergence rate when the smoothing parameter was increased. This was observed for both modalities and both priors.

Proximal algorithms stayed more stable with increased smoothing. The convergence rate decreased marginally with increasing θ.

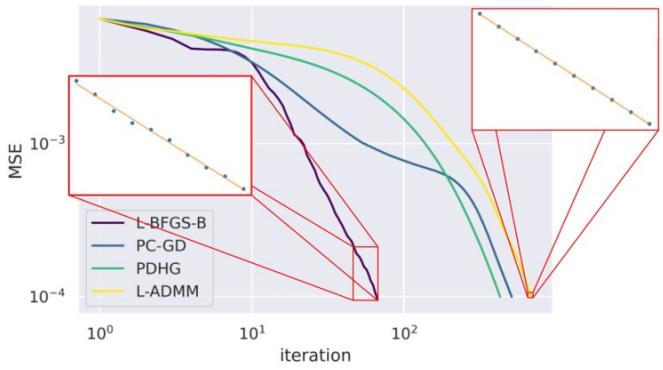


Fig. 3: Illustration of calculation of convergence rate, C. Inset graphs have linear y-axis

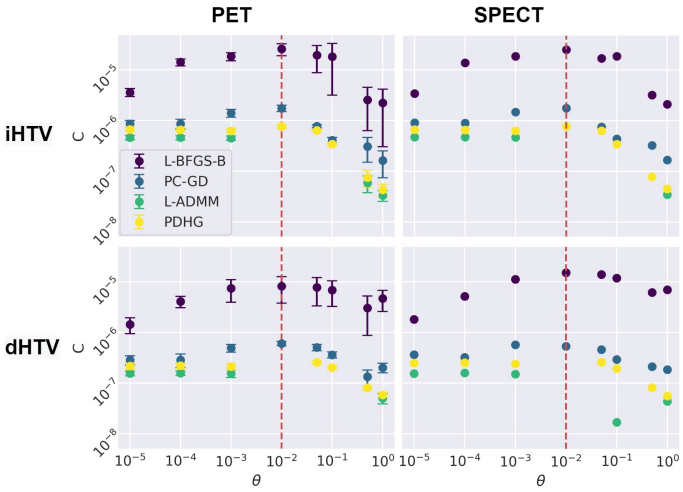


Fig. 4: Effect of varying smoothing parameter, θ, on the convergence rate of four algorithms

B. Large smoothing parameter

When the smoothing parameter was sufficiently large, approximately greater than 0.01, the convergence results broke down and all algorithms were observed to have a reduced convergence rate. The spread of results also increased significantly for PET reconstructions with different noise realisations, suggesting a greater dependence on noise.

L-ADMM was unable to converge in less than 10,000 iterations for smoother priors (Fig. 5).

III. Discussion

Despite the different system models and well-posedness of the PET and SPECT reconstruction problem, we have observed very similar trends in the PET and SPECT cases.

It has long been known that less-smooth functions provide a difficult environment for gradient-based algorithms. Large differences in gradient can cause overshooting or small step sizes and result in slow convergence. Therefore, an improvement in algorithm performance was expected for smoother priors, and this was indeed observed.

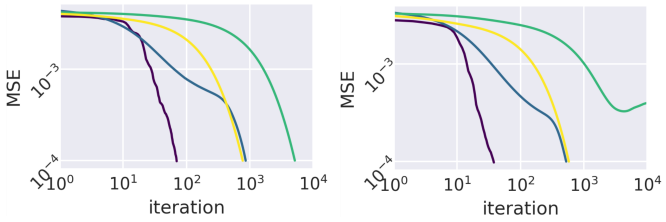
Proximal algorithms were developed for non-smooth functions and so were expected to perform better on priors with a small smoothing parameter. The effect of the quadratic term "pulling" the current and next estimates together was expected to increase for smoother functions, resulting in slower convergence. We anticipate that this effect would be larger near the minimum of the prior term. In the future, this could be better investigated using separate regions of interest for high-gradient and low-gradient parts of the image and investigating the effect of smoothing on the convergence rate separately.

In this investigation, we have chosen the normalisation of the Huber prior (Eq. 4) such that the behaviour of the prior was independent of θ at large edges. However, this choice means that for large θ the prior became very flat for most of the image gradients. This resulted in under-regularisation (Fig. 5), meaning that the optimisation problem was less well posed. Due to this, the convergence rate became much more varied (demonstrated by the increased size of the error bars in Fig. 4) and the convergence rate decreased.

Proximal algorithms consistently performed more poorly than gradient-based algorithms throughout this investigation, both in terms of convergence rate near the solution (Fig. 4) and the required number of iterations (Fig.5), especially L-BFGS-B which produced "converged" images in an order of magnitude fewer iterations than its proximal counterparts. The gradient-based algorithms were also much easier to fine-tune, with L-BFGS-B being practically plug-and-play. However, as the aim of this investigation was not to compare the convergence of these algorithms, we did not do an exhaustive search on optimal parameter settings. We speculate that this could also have been caused by the fact that both gradient-based algorithms had step sizes (and directions in the case of L-BFGS-B) based on line searches and iteration-dependent preconditioners, whereas the proximal algorithms had (preconditioned) proximal step sizes based on the warm-start images that were fixed at the start of the reconstruction process. This warrants further investigation, especially for more complex algorithms such as stochastic variants and with iteration-dependent preconditioners and step sizes. At present, however, gradient-based algorithms seem to be preferable over proximal algorithms for fully differentiable objective functions. Nonetheless, the only small reduction in convergence rate for low smoothing of the Huber prior is reassuring that, should prior smoothing be desired, the convergence of proximal algorithms should not be overly affected.

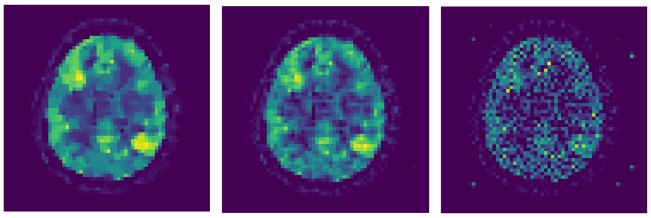
IV. Conclusion

The convergence rate near the final solution, C, depends on the amount of smoothing of the prior, with similar trends observed for PET and SPECT as well as for isotropic and directional variants of the Huber prior. For low smoothing parameter, θ, values, more smoothing resulted in improvements in the convergence rate of gradient-based algorithms and marginal reductions in the convergence rate of proximal algorithms. For very large smoothing parameters, convergence rate becomes more unstable and was reduced for all algorithms due to under-regularisation.



(a) $\theta = 1 \times 10^-5$ (b) $\theta = 0.1$

Fig. 5: Convergence of four algorithms. Colours are the same as in Fig. 4. Note the log scale for the horizontal axis, and the fact that L-ADMM (green) fails to converge with high smoothing parameter.



(a) $\theta = 1 \times 10^-5$ (b) $\theta = 0.05$ (c) $\theta = 1$

Fig. 6: L-BFGS-B PET reconstructed image with a iHTV prior for three different smoothing parameters

ACKNOWLEDGMENT

Software used for PET & SPECT projectors, Synergistic Tomographic Image Reconstruction Software [6], [7]. Software used for Reconstruction, Core Imaging Library [11], [12].

APPENDIX

A. EM-TV

To warm-start the reconstructions we used a simple algorithm consisting of an OSEM step followed by a weighted denoising step and then applied a positivity constraint.

Algorithm 1 EM-TV

$u_0 \leftarrow$ uniform image
$s = \mathcal{A}^T\mathbf{1}$
for k iterations **do**
  $u_k^{EM} \leftarrow$ OSEM step
  $w_k \leftarrow \frac{s}{u_k}$
  $f(u_k) = \sum_i w_i ||u_k - u_k^{EM}||^2$
  $\bar{u}_{k+1} \leftarrow \arg\min_x \{\mathcal{R}(u_k) + f(u_k)\}$
  $u_{k+1} \leftarrow \text{Proj}_{\{\geq 0\}}(\bar{u}_{k+1})$

B. L-BFGS-B

We use the scipy implementation of L-BFGS-B [13]. This was altered to allow the algorithm to run without a stopping criteria.

C. PC-GD

Algorithm 2 PC-GD

$u_0 \leftarrow u_{init}$
$s \leftarrow \mathcal{A}^T\mathbf{1}$
Initialise $\lambda, c > 0$
for k iterations **do**
  $g_k \leftarrow \nabla\mathcal{L}$
  $p_k \leftarrow \frac{u_k}{s}$
  $t_k \leftarrow cg_k \cdot p_k$
  $\bar{u}_{k+1} \leftarrow u_k - \lambda g_k$
  $u_{k+1} \leftarrow \text{Proj}_{\{\geq 0\}}(\bar{u}_{k+1})$
  if $\mathcal{L}(u_{k+1}) - \mathcal{L}(u_k) < \lambda t_k$ **then**
    $\lambda \leftarrow 0.5\lambda$
    repeat iteration
  else
    continue
  end if

The initial step size, λ, was chosen to be 1 and c was chosen to be 0.0001

D. PDHG

We used a 3-block algorithm, inspired by work on the stochastic variant of PDHG [2]

Algorithm 3 PDHG

$u_0, \leftarrow u_{init}, y_0 \leftarrow Ku, z_0 \leftarrow \mathcal{A}u$
$\theta \in [0, 1]$
Initialise $\sigma_\mathcal{D}, \sigma_\mathcal{R}, \tau > 0$
for k iterations **do**
  $y_k = \beta \,\text{prox}_{R*}^{\sigma_\mathcal{R}/\beta}((y_{k-1} + \sigma_\mathcal{R} K u_{k-1})/\beta)$
  $z_k = \text{prox}_{\mathcal{D}^*}^{\sigma_\mathcal{D}}(z_{k-1} + \sigma_\mathcal{D}\mathcal{A}u_{k-1})$
  $u_k = \text{Proj}_{\{\geq 0\}}(u_{k-1} - \tau K^T y_k - \tau\mathcal{A}^T z_k)$
  $\bar{u} = u_k + \theta(u_k - u_{k-1})$

where K is ∇ for iHTV or D for dHTV. Step sizes were chosen to be $\sigma_\mathcal{D} = |u|/|\mathcal{A}|$, $\sigma_\mathcal{R} = |u|/|K|$, and $\tau = 2 \cdot \min\{1/(|u||\mathcal{A}|), 1/(|u||K|)\}$.

E. L-ADMM

We use 3-block ADMM with the first term linearised using a small approximation to give an explicit proximal in Step 5 [14].

Algorithm 4 L-ADMM

$u_0 \leftarrow u_{init}, y_0 \leftarrow Ku, z_0 \leftarrow \mathcal{A}u$
$p_0, q_0 \leftarrow 0$
Initialise $\sigma_\mathcal{D}, \sigma_\mathcal{R}, \tau > 0$
for k iterations **do**
  $u_k \leftarrow \text{Proj}_{\{\geq 0\}\ R}\big[u - \tau\big($
    $(1/\sigma_\mathcal{D})(\mathcal{A}^T(\mathcal{A}u_{k-1} - y_{k-1} + q_{k-1})) -$
    $(1/\sigma_\mathcal{R})(K^T(Ku_{k-1} - z_{k-1} + p_{k-1}))\big)\big]$
  $y_k \leftarrow \text{prox}_R^{\beta_\mathcal{R}\sigma}(Ku + p)$
  $z_k \leftarrow \text{prox}_\mathcal{D}^{\sigma_\mathcal{D}}(Au + q)$
  $p_k \leftarrow p_k + Ku_k - y_k$
  $q_k \leftarrow q_k + Au_k - z_k$

where K is ∇ for iHTV or D for dHTV. Step sizes were chosen to be $\sigma_{\mathcal{D}} = |u|/|\mathcal{A}|$, $\sigma_{\mathcal{R}} = |u|/|K|$, and $\tau = 2 \cdot \min\{\sigma_{\mathcal{D}}/(|\mathcal{A}|^2), \sigma_{\mathcal{R}}/(|K|^2)\}$.

REFERENCES

[1] N. Parikh and S. Boyd, "Proximal Algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, jan 2014. [Online]. Available: https://doi.org/10.1561/2400000003

[2] M. J. Ehrhardt, P. Markiewicz, A. Chambolle, P. Richtárik, J. Schott, and C.-B. Schönlieb, "Faster PET reconstruction with a stochastic primal-dual hybrid gradient method," in *Wavelets and Sparsity XVII*, Y. M. Lu, D. V. D. Ville, and M. Papadakis, Eds., vol. 10394, International Society for Optics and Photonics. SPIE, 2017, p. 103941O. [Online]. Available: https://doi.org/10.1117/12.2272946

[3] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73 – 101, 1964. [Online]. Available: https://doi.org/10.1214/aoms/1177703732

[4] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans, "BrainWeb: Online Interface to a 3D MRI Simulated Brain Database," *NeuroImage*, 1997.

[5] E. Ovtchinnikov, R. Brown, C. Kolbitsch, E. Pasca, C. da Costa-Luis, A. G. Gillman, B. A. Thomas, N. Efthimiou, J. Mayer, P. Wadhwa, M. J. Ehrhardt, S. Ellis, J. S. Jørgensen, J. Matthews, C. Prieto, A. J. Reader, C. Tsoumpas, M. Turner, D. Atkinson, and K. Thielemans, "SIRF: Synergistic Image Reconstruction Framework," *Computer Physics Communications*, vol. 249, p. 107087, Apr. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010465519303984

[6] K. Thielemans, C. Tsoumpas, S. Mustafovic, T. Beisel, P. Aguiar, N. Dikaios, and M. Jacobson, "STIR: Software for tomographic image reconstruction release 2," *Physics in medicine and biology*, vol. 57, pp. 867–83, 02 2012.

[7] B. M. Fuster, C. Falcon, C. Tsoumpas, L. Livieratos, P. Aguiar, A. Cot, D. Ros, and K. Thielemans, "Integration of advanced 3D SPECT modeling into the open-source STIR framework," *Medical Physics*, vol. 40, no. 9, p. 092502, 2013. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4816676

[8] M. J. Ehrhardt and M. M. Betcke, "Multicontrast MRI Reconstruction with Structure-Guided Total Variation," *SIAM Journal on Imaging Sciences*, vol. 9, no. 3, pp. 1084–1106, 2016. [Online]. Available: https://doi.org/10.1137/15M1047325

[9] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A Limited Memory Algorithm for Bound Constrained Optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995. [Online]. Available: https://doi.org/10.1137/0916069

[10] Chambolle, Antonin and Pock, Thomas, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, 05 2011.

[11] J. S. Jørgensen, E. Ametova, G. Burca, G. Fardell, E. Papoutsellis, E. Pasca, K. Thielemans, M. Turner, R. Warr, W. R. B. Lionheart, and P. J. Withers, "Core Imaging Library - Part I: a versatile Python framework for tomographic imaging," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2204, p. 20200192, 2021. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0192

[12] E. Papoutsellis, E. Ametova, C. Delplancke, G. Fardell, J. S. Jørgensen, E. Pasca, M. Turner, R. Warr, W. R. B. Lionheart, and P. J. Withers, "Core Imaging Library - Part II: multichannel reconstruction for dynamic and spectral tomography, journal = Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences," vol. 379, no. 2204, p. 20200193, 2021. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0193

[13] "SciPy," May 2022, original-date: 2011-03-09T18:52:03Z. [Online]. Available: https://github.com/scipy/scipy/blob/2204e3a39c0d6c2d5292eaeae1a752e5565d9eb9/CITATION.bib

[14] E. Esser, X. Zhang, and T. F. Chan, "A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science," *SIAM Journal on Imaging Sciences*, vol. 3, no. 4, pp. 1015–1046, 2010. [Online]. Available: https://doi.org/10.1137/09076934X