

A MAJORIZE-MINIMIZE MEMORY GRADIENT ALGORITHM APPLIED TO X-RAY TOMOGRAPHY

Emilie Chouzenoux¹, Fiona Zolyniak¹, Emmanuelle Guillard² and Hugues Talbot¹

¹ Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge,
5, Boulevard Descartes, Champs-sur-Marne,
77454 Cedex 2, Marne-la-Vallée, France
first.last@univ-paris-est.fr

² Unité mixte CNRS/Saint-Gobain Surface du Verre et Interfaces
Saint-Gobain Recherche, 39 quai Lucien Lefranc
93303 Aubervilliers Cedex, France
emmanuelle.guillard@nsup.org

ABSTRACT

Tomography is an image reconstruction task that may be viewed as a linear inverse problem akin to deconvolution. Recent progresses in optimization methods have made it possible to formulate this task so that fewer projections and higher amounts of noise can be dealt with, making use of a-priori information and domain constraints. In this article, we investigate 3MG, a new optimization method that is highly flexible and effective. In particular, we propose and compare convex and non-convex regularization potentials on both synthetic and real images. We further investigate the possibility to deal with continuous angular integration, i.e. where projections rays are no longer straight lines, but cones. This is encountered in a variety of real-life situations, but is difficult or impossible to deal with exactly using traditional reconstruction algorithms. We show that in this situation it may be beneficial to acquire fewer projections than would be required using classical methods.

Index Terms— Image Reconstruction, Tomography, Optimization, Spatial Regularization, Segmentation.

1. INTRODUCTION

X-ray tomography [1] reconstructs dense 3-D volumes of objects from a set of projections measured at different angles. The measurements $\mathbf{y} \in \mathbb{R}^Q$ and the sought absorption image $\mathbf{x} \in \mathbb{R}^N$ obey the linear relation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \quad (1)$$

where \mathbf{w} is the measurement noise, that we assume i.i.d. Gaussian with variance σ^2 . The tomography matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ is sparse and encodes the geometry of the measurements. In this paper, we focus on the case when \mathbf{H} models parallel projections of a 2-D object \mathbf{x} . Moreover, two different acquisition schemes, shown in Fig. 1, are compared. In the discrete acquisition scheme (Fig. 1-(a)), measures are acquired at fixed and regularly sampled rotational positions

between the sample and the detector so that H_{mn} models the intersection length between the m th light-ray and the n th pixel. In the continuous acquisition scheme, the detector is continuously moving while the measures are acquired over an angular sector $[\theta, \theta + \Delta\theta)$ (Fig. 1-(b)). The latter configuration, which has been little-studied in tomographic reconstruction literature, is nonetheless commonly used in materials science to reduce acquisition time while retaining a sufficient signal to noise ratio [2]. In such a model, H_{mn} corresponds to the intersection area between the m th angular light-beam and the n th pixel. Let us call N_θ the number of different angular positions of the detector in Fig. 1 and L the linear size of the detector; the size of the image is then $N = L^2$ and the number of measurements $Q = L \times N_\theta$. In practice, the angular positions are regularly distributed on $[0, \pi)$ for both schemes, so that, in the continuous acquisition scheme, $\Delta\theta = \pi/N_\theta$.

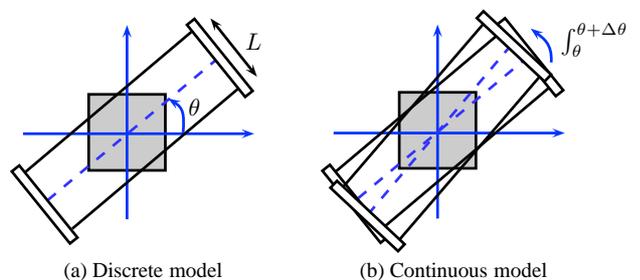


Fig. 1. Considered tomographic acquisition models

Traditional reconstruction methods such as the Filtered Back-Projection [3] require the linear system (1) to be sufficiently determined for good results, i.e., $N_\theta \sim L$. However, several applications could benefit from a smaller number of projections, either in order to reduce the total dose for medical applications [4], or to reduce the total acquisition time for in-situ experiments where the sample is evolving [5, 6]. The under-determinacy of Eq. (1) can be overcome by adding supplementary constraints to the data-fidelity term, which express prior information on \mathbf{x} . Tomographic reconstruction is expressed as an optimization problem. Since many samples

This work was partially funded by the ANR grant EDDAM (ANR-11-BS09-027).

consist of a limited number of phases, with known and near-constant absorption characteristics, a common prior is to encourage piecewise-constant results. This can be achieved by enforcing the sparsity of the image in a wavelet basis [7], or of its vertical and horizontal gradients [8]. In order to use convex optimization techniques, the ℓ_1 norm is often used as the convex relaxation of the ℓ_0 norm encoding sparsity. As far as gradient sparsity is concerned, a common strategy is to minimize the total variation semi-norm of the image [7, 8].

In this work, we apply a recently-developed memory-gradient algorithm [9] to the problem of tomographic reconstruction. The versatility of the algorithm allows us to compare different types of spatial regularization, both convex and non-convex. The paper is organized as follows: Section 2 presents the penalized criterion we considered, and introduces the memory-gradient algorithm. Then, Section 3 illustrates the applicability of this algorithm through a set of experiments in image reconstruction.

2. PROPOSED RECONSTRUCTION METHOD

2.1. Penalized criterion

An efficient strategy to address the reconstruction problem is to define \mathbf{x} as a minimizer of an appropriate cost function $F(\mathbf{x})$. More specifically, we focus on the following penalized least-squares criterion:

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 + R(\mathbf{x}), \quad (2)$$

where R is a regularization function incorporating a-priori assumptions to guarantee the robustness of the solution with respect to noise. A hybrid regularization function $R = R_1 + R_2$, is considered. First, since in practice, a reasonable guess of the dynamic range $[x_{\min}, x_{\max}]$ of the expected absorption image is available, we define:

$$R_1(\mathbf{x}) = \eta \sum_{n=1}^N d_{[x_{\min}, x_{\max}]}^2(x_n), \quad (3)$$

where $d_{[x_{\min}, x_{\max}]}$ denotes the distance to the closed convex interval $[x_{\min}, x_{\max}]$, and $\eta > 0$. Note that, when η goes to infinity, R_1 becomes the indicator function of the convex domain $[x_{\min}, x_{\max}]^N$. Furthermore, in order to promote images formed by smooth regions separated by sharp edges, we set

$$R_2(\mathbf{x}) = \sum_{n=1}^N \psi \left(\sqrt{(\mathbf{V}_n^h \mathbf{x})^2 + (\mathbf{V}_n^v \mathbf{x})^2} \right), \quad (4)$$

where $(\mathbf{V}_n^h)^\top \in \mathbb{R}^N$ (resp. $(\mathbf{V}_n^v)^\top \in \mathbb{R}^N$) corresponds to a horizontal (resp. vertical) gradient operator with zero-boundaries assumption, and ψ is a potential function fulfilling the following conditions:

- (i) ψ is a differentiable function.

- (ii) $\psi(\sqrt{\cdot})$ is concave on $[0, +\infty)$.

- (iii) There exists $\bar{\omega} \in [0, +\infty)$ such that $(\forall t \in (0, +\infty))$ $0 \leq \dot{\psi}(t) \leq \bar{\omega}t$ where $\dot{\psi}$ is the derivative of ψ . In addition, $\lim_{t \rightarrow 0} \omega(t) \in \mathbb{R}$ with $\omega(t) \triangleq \dot{\psi}(t)/t$.

Two main families of such functions ψ are commonly used for image reconstruction, namely ℓ_2 - ℓ_1 functions, i.e. convex, continuously differentiable, asymptotically linear functions with a quadratic behavior near 0 [10], and ℓ_2 - ℓ_0 functions, i.e. asymptotically constant functions with a quadratic behavior near 0 [11].

2.2. Majorize-Minimize Memory Gradient algorithm

The Majorize-Minimize Memory Gradient (3MG) algorithm [9, 12] is suitable for the minimization of (2). It consists of building a sequence $(\mathbf{x}_k)_{k \in \mathbb{N}}$ according to the following update scheme:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{D}_k \mathbf{u}_k, \quad (5)$$

where \mathbf{D}_k is the Memory Gradient set of directions:

$$\mathbf{D}_k = \begin{cases} -\nabla F(\mathbf{x}_0) & \text{if } k = 0 \\ [-\nabla F(\mathbf{x}_k) \quad \mathbf{x}_k - \mathbf{x}_{k-1}] & \text{if } k > 0 \end{cases} \quad (6)$$

and $\mathbf{u}_k \in \mathbb{R}^2$ is a multivariate stepsize that aims at partially minimizing $\mathbf{u} \mapsto f_k(\mathbf{u}) = F(\mathbf{x}_k + \mathbf{D}_k \mathbf{u})$. The convergence of the recursive update equation (5) requires the design of a proper strategy to determine the stepsize \mathbf{u}_k . The computation of \mathbf{u}_k relies on the Majorization-Minimization (MM) principle. Let $\mathbf{x}' \in \mathbb{R}^N$. Function $Q(\cdot, \mathbf{x}')$ is said to be a *tangent majorant* for F at \mathbf{x}' , if

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad Q(\mathbf{x}, \mathbf{x}') \geq F(\mathbf{x}) \quad \text{and} \quad Q(\mathbf{x}', \mathbf{x}') = F(\mathbf{x}').$$

Following [12], the following quadratic function is a convex quadratic tangent majorant of (2) at \mathbf{x}' ,

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad Q(\mathbf{x}, \mathbf{x}') = F(\mathbf{x}') + \nabla F(\mathbf{x}')^\top (\mathbf{x} - \mathbf{x}') + \frac{1}{2} (\mathbf{x} - \mathbf{x}')^\top \mathbf{A}(\mathbf{x}') (\mathbf{x} - \mathbf{x}'),$$

if, for every $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{A}(\mathbf{x}) \in \mathbb{R}^{N \times N}$ is the symmetric positive matrix:

$$\mathbf{A}(\mathbf{x}) = \mathbf{H}^\top \mathbf{H} + \mathbf{V}^{h\top} \text{Diag}\{\mathbf{b}(\mathbf{x})\} \mathbf{V}^h + \mathbf{V}^{v\top} \text{Diag}\{\mathbf{b}(\mathbf{x})\} \mathbf{V}^v + 2\eta \mathbf{I},$$

and $\mathbf{b}(\mathbf{x}) \in \mathbb{R}^N$ with $b_n(\mathbf{x}) = \omega \left(\sqrt{(\mathbf{V}_n^h \mathbf{x})^2 + (\mathbf{V}_n^v \mathbf{x})^2} \right)$. The MM stepsize is then obtained by minimizing the quadratic function $\mathbf{u} \mapsto Q(\mathbf{x}_k + \mathbf{D}_k \mathbf{u}, \mathbf{x}_k)$, leading to the following update rule:

$$\mathbf{u}_k = - (\mathbf{D}_k^\top \mathbf{A}(\mathbf{x}_k) \mathbf{D}_k)^{-1} \mathbf{D}_k^\top \nabla F(\mathbf{x}_k). \quad (7)$$

The 3MG algorithm runs through steps (6)-(7) and (5) for $k \geq 0$. The convergence of 3MG to a critical point of (2) was established in [12], under mild assumptions on the penalization term R . In particular, some non-convex potential functions ψ are suitable. In practice, the algorithm is run until the following stopping criterion is fulfilled:

$$\|\nabla F(\mathbf{x}_k)\|/\sqrt{N} \leq \epsilon, \quad (8)$$

where, typically, $\epsilon = 10^{-4}$.

3. RESULTS AND DISCUSSION

3.1. Comparison of regularization strategies

We first consider a set of 2 circular images (Fig. 2) in order to compare the reconstruction results with both smooth convex (SC) ℓ_2 - ℓ_1 penalty and smooth non-convex (SNC) ℓ_2 - ℓ_0 penalty, given, respectively, by $\psi(t) = \lambda\sqrt{1+t^2/\delta^2}$, and $\psi(t) = \lambda(1 - (1+t^2/2\delta^2)^{-1})$, $(\lambda, \delta) > 0$. Note that, when $\delta \rightarrow 0$, the regularization term (4) with SC penalty becomes the total variation [13].

Projections were acquired using the discrete acquisition model from Fig. 1-(a), with $L = 128$ and $N_\theta = 16, 32, 64$ or 128 angle values. The image diameter is chosen equal to the detector size L , so that the number of pixels to estimate is given by:

$$N = \text{Card} \left\{ (i, j) \in \mathbb{Z}^2; (i+1/2)^2 + (j+1/2)^2 \leq L^2/4 \right\}, \\ \approx L^2\pi/4.$$

Reconstruction was performed for $x_{\min} = 0$, $x_{\max} = 1$, $\eta = 10^4$, parameters (λ, δ) were adjusted to maximize the Signal to Noise Ratio (SNR) between the original image $\bar{\mathbf{x}}$ and the reconstructed image $\hat{\mathbf{x}}$, defined as:

$$\text{SNR} = 20 \log_{10} (\|\bar{\mathbf{x}}\|/\|\hat{\mathbf{x}} - \bar{\mathbf{x}}\|). \quad (9)$$

The 3MG algorithm was able to reach the stopping criterion (8) in about 300 iterations. Computational time was around 10s, using the EPD 7.3 Python distribution on an IBM HX5 blade computer with four 1.86GHz CPUs and 144GB of RAM running Linux CentOS 6.2. Fig. 3 shows the evolution of the reconstruction error, evaluated in terms of SNR, as a function of the undersampling rate N_θ/L , for noise standard deviation $\sigma = 1$ or 2, and both SC and SNC regularizations. In Fig. 5-(a) we show the relative classification error obtained after thresholding of the reconstructed binary images.

As expected, the reconstruction quality decreases with a smaller number of projections. Reconstruction quality is better for the binary image than with the natural image, probably because the former is piecewise-constant at a larger scale, with all pixels saturating the interval constraint (3). Note that SNC regularization leads to higher SNR than SC for the binary image, especially for high values of N_θ/L . On the other

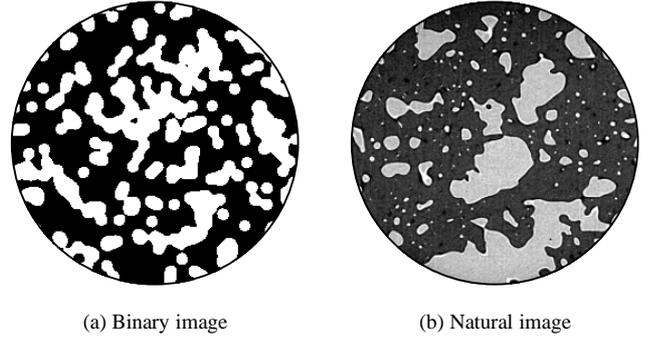


Fig. 2. Original images. The natural image is a phase-separated barium borosilicate glass imaged at the ESRF synchrotron (courtesy of David Bouttes).

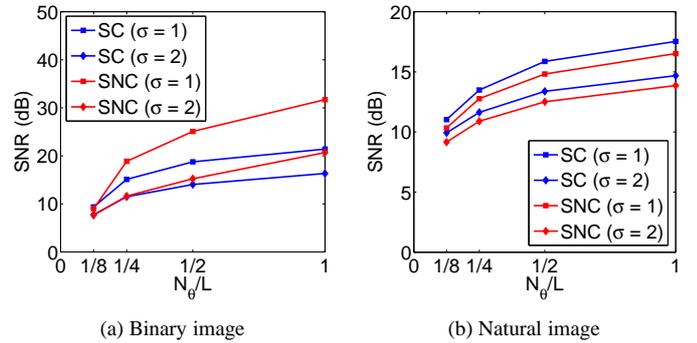


Fig. 3. Quality of reconstruction, in terms of SNR, for SC and SNC potential functions, $L = 128$.

hand, SC regularization appears to yield better results for the natural image. Regarding segmentation, different behaviors are observed: While for a low noise level ($\sigma = 1$), the SNC strategy gives the best results, for higher noise level ($\sigma = 2$), SC penalty leads to a smallest number of misclassified pixels.

The absolute reconstruction error $|\hat{\mathbf{x}} - \bar{\mathbf{x}}|$ is shown in Fig. 4. No particular spatial structure is visible for the synthetic image; because the non-convex penalization results in an almost binary image, the main contribution to the error comes from isolated pixels that are wrongly labeled (absolute error of the order of 1). On the other hand, a systematic error is present on the boundary of the domains for the natural image. The natural image was acquired with some phase contrast [14, 15] at a synchrotron light source (ESRF), resulting in narrow contrasted fringes along the domain boundaries. The reconstruction error is concentrated in these regions where the gradient is larger.

We note that similar results to the SC penalty, both in terms of SNR and segmentation error, were obtained with a standard total-variation penalization associated with range constraints on $[x_{\min}, x_{\max}]$, the minimization of (2) being performed in that case with the proximal-based algorithm from [16].

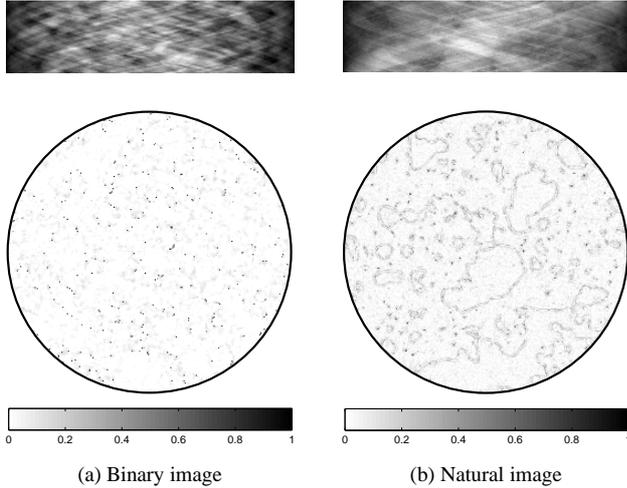


Fig. 4. Noisy sinograms (top) and absolute reconstruction error (bottom) using the discrete model, $L = 256$, $N_\theta/L = 1/4$, $\sigma = 2$. SNC (resp. SC) penalty is employed for the binary (resp. natural) image.

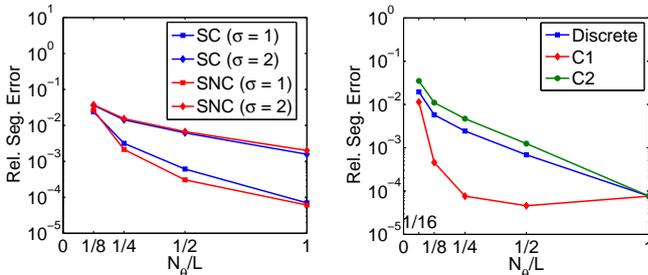


Fig. 5. Segmentation errors in the reconstruction of the binary image, for the experiments settings of Sec. 3.1 (left) and Sec. 3.2 (right).

3.2. Comparison of acquisition models

We now compare the results obtained with the discrete and continuous acquisition models. To this end, synthetic acquisitions were simulated from the two images of Fig. 2, for $L = 256$, and $N_\theta = 16, 32, 64, 128$ or 256 projections. Due to angular integration, for a given image \mathbf{x} , the energy of $\mathbf{H}\mathbf{x}$ is L/N_θ times greater in the continuous case than in the discrete case. Two scenarios were considered for the tuning of the noise standard deviation σ for the continuous model: The first strategy, denoted after by C1, is to consider the same value σ for both the discrete and continuous cases. It implies that the input SNR is higher in the continuous case than in the discrete case. It corresponds to high flux cases when the readout noise dominates. The second strategy, denoted after by C2, is to increase the value of σ in the continuous case, so that the input SNR is the same for both models. The reconstruction error as a function of N_θ/L is illustrated on Fig. 6. The SC penalty was used for the natural image, while SNC was chosen for

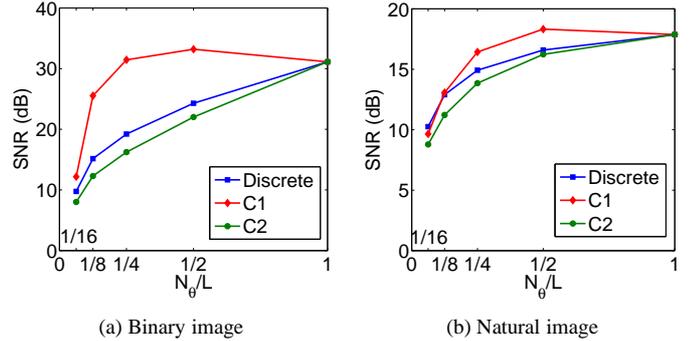


Fig. 6. Quality of reconstruction, in terms of SNR, for discrete and continuous acquisition models, $L = 256$.

the binary image. The noise standard noise deviation was set to $\sigma = 2$ for the discrete case. For a constant noise level on the measurements (model C2), we observe that continuous acquisition degrades the reconstruction vs. the discrete case, for both images. This can likely be attributed to the angular integration which can be viewed as an averaging of neighboring pixels inside a cone of angular width $\Delta\theta$. On the other hand, model C1 yields better results than the discrete acquisition for both images, because the input signal to noise ratio is improved. Interestingly, the best reconstruction is observed not with the maximum number of projections ($N_\theta/L = 1$) but with $N_\theta/L = 1/2$. This expresses the fact that an optimal trade-off must exist between the number of measurements and the noise level in individual projections for a given global acquisition time. Segmentation results in Fig. 5-(b) in the case of the binary image lead to the same ranking between discrete, C1 and C2 acquisition models.

4. CONCLUSION

This paper addresses the problem of tomography reconstruction formulated as a penalized least-squares optimization approach. The proposed 3MG algorithm is highly flexible. It is capable of solving convex or non-convex problem formulations; it can be used with domain constraints; it can cope with various acquisition settings and it is not restricted to one particular type of image (e.g. binary data). We show on experimental results that a non-convex penalization is preferable in the case of a high gradient sparsity level, which confirms the conclusions drawn in [17, 18] for ℓ_p penalties with $0 < p < 1$. For images presenting some phase contrast effect, the convex strategy seems to provide better results. Furthermore, through a simulation of the effects of a continuous rotation acquisition model, we exhibit the existence of an optimal strategy to calibrate the angular acquisitions, which may not correspond, surprisingly, to a maximization of the number of measurements. Future work will be directed at pursuing this study for a larger set of test images, considering various sparsity patterns and more realistic acquisition simulations.

5. REFERENCES

- [1] A. Kak and M. Slaney, *Principles of computerized tomographic imaging*, Classics in Applied Mathematics. SIAM, Philadelphia, 1988.
- [2] O. Ludwig, M. Dimichiel, L. Salvo, M. Suéry, and P. Falus, “In-situ three-dimensional microstructural investigation of solidification of an al-cu alloy by ultrafast x-ray microtomography,” *Metallurgical and Materials Transactions A*, vol. 36, no. 6, pp. 1515–1523, Jun. 2005.
- [3] G. N. Ramachandran and A. V. Lakshminarayanan, “Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 68, no. 9, pp. 2236–2240, Sep. 1971.
- [4] J. Hausleiter, T. Meyer, M. Hadamitzky, E. Huber, M. Zankl, S. Martinoff, A. Kastrati, and A. Schömig, “Radiation dose estimates from cardiac multislice computed tomography in daily practice impact of different scanning protocols on effective dose estimates,” *Circulation*, vol. 113, no. 10, pp. 1305–1310, Mar. 2006.
- [5] J. Baruchel, J. Y. Buffiere, P. Cloetens, M. Di Michiel, E. Ferrie, W. Ludwig, E. Maire, and L. Salvo, “Advances in synchrotron radiation microtomography,” *Scripta Materialia*, vol. 55, no. 1, pp. 41–46, Jul. 2006.
- [6] J.-Y. Buffiere, E. Maire, J. Adrien, J.-P. Masse, and E. Boller, “In situ experiments with X-ray tomography: an attractive tool for experimental mechanics,” *Experimental Mechanics*, vol. 50, no. 3, pp. 289–305, Mar. 2010.
- [7] N. Pustelnik, C. Chaux, and J.-C. Pesquet, “Parallel proximal algorithm for image restoration using hybrid regularization,” *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2450–2462, Sep. 2011.
- [8] E. Y. Sidky and X. Pan, “Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization,” *Physics in medicine and biology*, vol. 53, no. 17, pp. 4777, 2008.
- [9] E. Chouzenoux, J.-C. Pesquet, H. Talbot, and A. Jezierska, “A memory gradient algorithm for ℓ_2 - ℓ_0 regularization with applications to image restoration,” in *18th IEEE International Conference on Image Processing (ICIP 2011)*, 11–14 Sep. 2011, pp. 2717–2720.
- [10] M. Allain, J. Idier, and Y. Goussard, “On global and local convergence of half-quadratic algorithms,” *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1130–1142, May 2006.
- [11] A. H. Delaney and Y. Bresler, “Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography,” *IEEE Transactions on Image Processing*, vol. 7, no. 2, pp. 204–221, Feb. 1998.
- [12] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, “A majorize-minimize subspace approach for ℓ_2 - ℓ_0 image regularization,” *SIAM Journal on Imaging Science*, vol. 6, no. 1, pp. 563–591, 2013.
- [13] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.
- [14] A. Snigirev, I. Snigireva, V. Kohn, S. Kuznetsov, and I. Schelokov, “On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation,” *Review of Scientific Instruments*, vol. 66, no. 12, pp. 5486–5492, 1995.
- [15] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, “Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources,” *Nature Physics*, vol. 2, no. 4, pp. 258–261, 2006.
- [16] H. Raguet, J. Fadili, and G. Peyré, “Generalized Forward-Backward splitting,” Tech. Rep., 2011, <http://hal.archives-ouvertes.fr/hal-00613637>.
- [17] J. C. Ramirez-Giraldo, J. Trzasko, S. Leng, L. Yu, A. Manduca, and C. H. McCollough, “Nonconvex prior image constrained compressed sensing (NCPICCS): Theory and simulations on perfusion CT,” *Medical Physics*, vol. 38, pp. 2157, 2011.
- [18] E. Y. Sidky, R. Chartrand, and X. Pan, “Image reconstruction from few views by non-convex optimization,” in *IEEE Nuclear Science Symposium Conference Record (NSS 2007)*, 2007, vol. 5, pp. 3526–3530.