Impact of the Non-Negativity Constraint in Model-Based Iterative Reconstruction from CT Data

Viktor Haase, Katharina Hahn, Harald Schöndube, Karl Stierstorfer, and Frédéric Noo

***Abstract*—Model-based iterative reconstruction is a promising approach to achieve dose reduction without affecting image quality in diagnostic X-ray computed tomography. In the problem formulation, it is common to enforce non-negative values, which is motivated by physics but narrows down the choice of optimization algorithm. In this work, we report on experiments assessing the impact of the non-negativity constraint on image quality and reconstruction speed. The assessment is performed under eight scenarios that challenge the usefulness of the constraint. These include reconstructions from full and sparse view sampling, with quadratic or edge-preserving regularization, for two different objects. Our results show that improvements due to the non-negativity constraint are strongly scenario-dependent, and likely negligible for conventional full view CT imaging. This implies that for specific reconstructions, the non-negativity constraint could be disregarded to simplify the optimization problem.**

I. Introduction

A lot of the research in clinical computed tomography (CT) is driven by the aim to reduce radiation dose while maintaining image quality. One promising way to achieve this goal is model-based iterative reconstruction (MBIR). Its potential for diagnostic CT imaging was shown in recent studies [1]–[4]. A popular MBIR formulation is penalized least squares reconstruction [5], which includes two key components: (i) the data fidelity term, which is characterized by the choice of a forward projection model and the option of a statistical weighting of the projections; (ii) the penalty term, which defines a regularization process with a potential function and additional incorporation of a priori knowledge, such as the non-negativity constraint.

To get the most out of the MBIR approach, it is valuable to understand the impact of each component and its subparts. For that reason, Thibault et al. have examined different potential functions for the regularization when they first introduced the concept of MBIR [5]. As another example, Hahn et al. have compared linear interpolation models for iterative CT reconstruction in various imaging scenarios [6]. In our previous work, we focused on the effect of statistical weights, which was analyzed using a lesion detection study with human observers [7].

The purpose of this paper is to look at the impact of the non-negativity constraint. Since the attenuation coefficient of X-rays is known to be positive, this constraint appears very natural, and its use is reinforced by knowledge that reconstruction without it leads to non-physical negative attenuation values, likely due to inaccuracy of the forward projection model and noise in the projection data. Our primary question of interest is about quantifying the impact of the constraint in regions of the image that are within the object. To our knowledge, this question has not been thoroughly addressed in the context of MBIR from real CT data.

V. Haase, K. Hahn, H. Schöndube, and K. Stierstorfer are with Siemens Healthcare GmbH, Forchheim, Germany. V. Haase is also with the Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. F. Noo is with the Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, Utah, USA.

II. Background

MBIR from CT data is formulated as optimization of a convex objective function. The optimization requires a dedicated algorithm that is consistent with the characteristics of this function. These two aspects are covered in the following subsections.

A. Objective Function

Let $\vec{x} \in \mathbb{R}^N$ be a discrete vector for the 3-D reconstruction volume and $\vec{p} \in \mathbb{R}^M$ a discrete vector for the measured projection data. Our objective function F consists of three parts: the data fidelity term f, the regularization term g with its hyper parameter $\beta \geq 0$, and the indicator function $\iota_{\mathbb{R}_+}$:

$$F(\vec{x}) = f(\vec{x}) + \beta\, g(\vec{x}) + \iota_{\mathbb{R}_+}(\vec{x}). \tag{1}$$

To ensure the data fidelity, we use the squared residual between the forward projected reconstruction and the projection data:

$$f(\vec{x}) = \frac{1}{2}\|A\vec{x} - \vec{p}\|_2^2. \tag{2}$$

The forward projection is modeled using Joseph's method [8]. It is a ray driven approach that provides a good compromise between accuracy and computational cost [6]. The forward projection process is symbolized by matrix A. The matching back projection operator is written as its transpose, A^T.

To reduce the noise in the reconstruction, a regularization term is used that is defined as

$$g(\vec{x}) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\psi(x_i - x_j), \tag{3}$$

with $w_{ij} = 1$ for the neighbors of each voxel found in the three Cartesian directions and $w_{ij} = 0$ otherwise. ψ is the potential function that assigns a cost to the difference between each voxel and one of its neighbors. We use two different potential functions for two different reconstruction scenarios, namely a quadratic and an edge-preserving potential function that can both be fine-tuned with $\delta > 0$:

$$\begin{aligned} &\text{(i)} \quad \psi(t,\delta) = \frac{t^2}{2\delta}, \\ &\text{(ii)} \quad \psi(t,\delta) = \sqrt{t^2 + \delta^2} - \delta. \end{aligned} \tag{4}$$

Parameter δ controls the importance given to differences between neighboring voxel values. Both choices for ψ are convex and differentiable, which implies that they are easy to handle for an optimization algorithm.

When we enforce non-negative voxels for our reconstruction result, the indicator function $\iota_{\mathbb{R}_+}$ is part of the objective function. This non-negativity constraint is defined as

$$\iota_{\mathbb{R}_+}(\vec{x}) = \begin{cases} 0 & \text{if } x_i \in \mathbb{R}_+ \\ +\infty & \text{otherwise} \end{cases} . \tag{5}$$

B. Optimization Algorithm

Because the indicator function is non-smooth, it has to be processed via its proximal operator. This means the optimization algorithm needs to be able to minimize a convex objective function that consists of a smooth and a proximable part. The fast iterative shrinkage-thresholding algorithm, also known as FISTA, meets these demands [9]. It requires only one gradient evaluation per iteration. The pseudocode of FISTA applied to our reconstruction problem is shown in Algorithm 1. We chose the version of FISTA with a fixed step size λ. The convergence is guaranteed for $\lambda < 1/L$, where L is the Lipschitz constant for the gradient of the smooth part, $f + \beta\, g$, of our objective function. The non-negativity constraint is enforced in line 3 of the algorithm. Because the proximal operator of an indicator function of a given set is the orthogonal projection operator onto the same set,

$$\operatorname{prox}_{\iota_{\mathbb{R}_+}}(\vec{x}) = \begin{cases} x_i & \text{if } x_i \in \mathbb{R}_+ \\ 0 & \text{otherwise} \end{cases} . \tag{6}$$

In the pseudocode we use the shorter symbol $(\cdot)_+$ instead of $\operatorname{prox}_{\iota_{\mathbb{R}_+}}(\cdot)$.

Algorithm 1: FISTA with constant step size

Input: Parameters $\beta \geq 0, \lambda > 0$ and initial image $\vec{x}_0$.

1 $\vec{y}_1 = \vec{x}_0$, $t_1 = 1$

2 **for** $k = 1, 2, \ldots$ **do**

3 $\quad \vec{x}_k = (\vec{y}_k - \lambda(A^T(A\vec{y}_k - \vec{p}) + \beta(\nabla g)(\vec{x})))_+$

4 $\quad t_{k+1} = \dfrac{1 + \sqrt{1 + 4t_k^2}}{2}$

5 $\quad \vec{y}_{k+1} = \vec{x}_k + (\dfrac{t_k - 1}{t_{k+1}})(\vec{x}_k - \vec{x}_{k-1})$

If we do not apply the non-negativity constraint, $\iota_{\mathbb{R}_+}$ is not used and our objective function is purely smooth. In this case, the algorithm simplifies itself to Nesterov's accelerated gradient descent [10]. The only change in the presented pseudocode is that the operation $(\cdot)_+$ disappears from line 3. The convergence condition for the step size is the same as in FISTA.

III. Experimental Setup

To evaluate our reconstruction method, experiments were carried out on a state-of-the-art clinical CT system. An overview of the scanner geometry can be found in Table I. The X-ray tube was operated at $80\,\mathrm{kV}$ and $500\,\mathrm{mAs}$. We have used a circular trajectory scan with a flying-focal-spot (FFS) in the x,y-plane that records 2304 projection images distributed over $360°$. We refer to this setting as full view sampling. To create a second scenario that is more challenging for the reconstruction algorithm, we also consider using only every 4th projection image taken from one of the two focus positions. This results in $^1/_8$th of the original projection data and represents a sparse view sampling.

TABLE I
Parameters of scanner geometry

Source to detector distance	$108.56\,\mathrm{cm}$
Source trajectory radius	$59.5\,\mathrm{cm}$
Anode angle	$7°$
Number of detector channels	736
Angular detector width	$0.067\,864°$
Number of detector rows	8
Detector row height at isocenter	$0.06\,\mathrm{cm}$
Number of projections	2304 (full view sampling) 288 (sparse view sampling)

The ACR CT accreditation phantom (model 464, Gammex-RMI, Middleton, WI, USA) was scanned as a test object. It has a cylindrical shape with a $20\,\mathrm{cm}$-diameter and a length of $16\,\mathrm{cm}$. The phantom is divided into four different modules of which we have looked at two for our study. The first one, called module A, has five cylinders representing the attenuation behavior of bone, polyethylene, water, acrylic, and air, respectively. Also two ramps are included that consists of small bars which are visible in $0.5\,\mathrm{mm}$ z-axis increments. The module can be used to assess positioning and CT number accuracy. The second one, called module D, contains eight aluminum bar patterns with up to $12\,\mathrm{lp/cm}$. The bar patterns provide very high contrast relative to the background and are used to assess the spatial resolution for high contrast objects. During the scan, each module was centered on the rotation axis, and the plane of the source trajectory passed through the middle of the module.

All reconstructions were done on a grid of $512 \times 512 \times 16$ voxels centered at the isocenter of the scanner. The voxel size was $0.1\,\mathrm{cm}$ in x and y, and $0.06\,\mathrm{cm}$ in z. The radius for the field of view (FOV) in the x, y-plane was set to $25\,\mathrm{cm}$. 5000 iterations of the reconstruction algorithm were calculated and after every 25th iteration the intermediate result was saved. As initial reconstruction volume we used $\vec{x}_0 = \vec{0}$. The chosen step size λ was based on the Lipschitz constant of the data fidelity term, $L(f)$. This can be calculated as the largest eigenvalue of $A^T A$ using the power method [11]. To account for g, we used $\lambda = 0.95/L(f)$ and the knowledge that $L(g) \ll L(f)$. For the full view sampling this means $\lambda = 0.000065$, and for the sparse view sampling, it is $\lambda = 0.00052$. The hyper parameter β is also linked to the size of the projection data. For the full view sampling we used $\beta = 0.1$; for the sparse sampling, which has $^1/_8$th of the full projection data, we used $\beta = 0.0125$ so that the amount of applied regularization is comparable for the different data sets. The value for the parameter of the potential function was empirically chosen as $\delta = 0.0005$

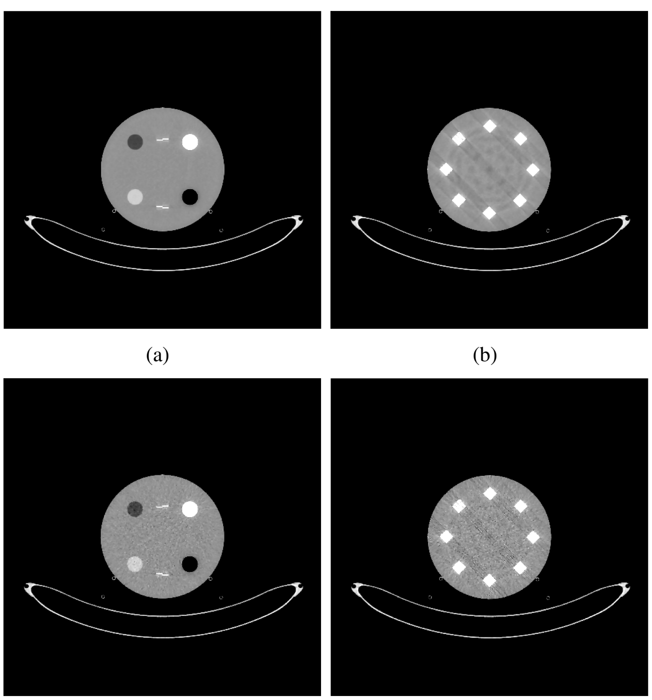
for the quadratic regularization and $\delta = 0.001$ for the edge-preserving regularization

To summarize, we have four different projection data sets that differ in the scanned object and the number of projections, and for reconstruction we use two different regularizers. This results in eight reconstruction scenarios for which we compare the result with and without the non-negativity constraint.

To assess the image quality, difference images between the result with and without non-negativity constraint were calculated. Where needed, profile plots through the difference image were created. Note that the result of the reconstruction is a 3-D volume, of which we only analyze one of the central slices for simplification. As figure of merit for convergence, we used the root-mean-square error (RMSE). To ignore the irrelevant structures outside of the phantom, a binary mask was applied with a radius of $10.5\,\mathrm{cm}$ that only contains the ACR module including its edges. The RMSE was calculated within the binary mask for the central six z-slices of the reconstruction. The final result after 5000 iterations was used as ground truth.

IV. Results

The results are separated into two parts. We first present the outcome of the experiments for the projection data with full view sampling, and then the results for the sparse view projection data. The results are focused on difference images and convergence speed. Fig. 1 shows how some reconstructions appear prior to computing differences.

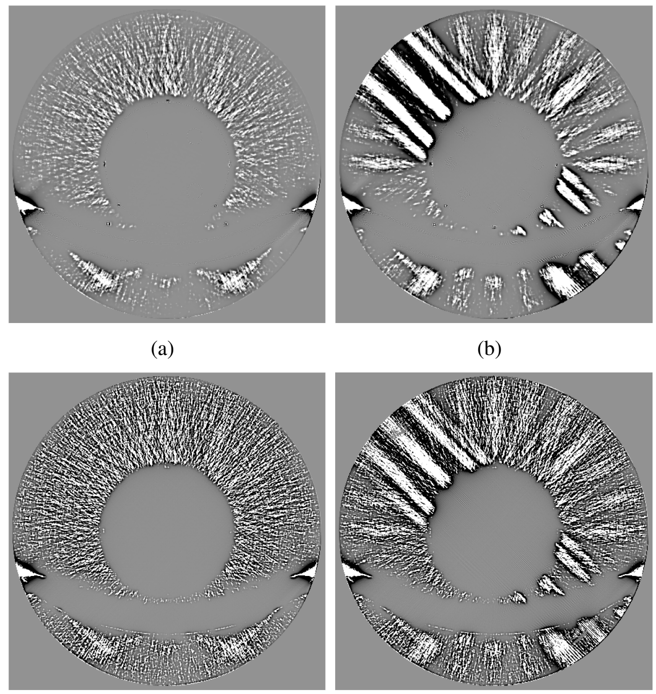


(a) (b) (c) (d)

Fig. 1. Ground truth for reconstruction with edge-preserving regularization and non-negativity constraint after 5000 iterations. (Top row) full view sampling, (bottom row) sparse view sampling. Module A (left) and module D (right) are both displayed with a grayscale of $[-200,\ 200]$ HU.

A. Full View Sampling of Projection Data

Fig. 2 shows the difference images between the reconstruction with non-negativity constraint and without. On the basis of the displayed grayscale window of $[-2,\ 2]$ HU, no significant differences are observed inside the phantom, for both phantom modules and both potential functions. The summary values given in Table II confirm this visual impression.



(a) (b) (c) (d)

Fig. 2. Difference images for reconstruction with and without non-negativity constraint. These images are for *full view sampling* of module A (left) and module D (right). The applied potential function is either edge-preserving (top) or quadratic (bottom). Grayscale window: $[-2, 2]$ HU.

Convergence according to the RMSE value is presented in Fig. 3. The results for the different scanned modules are similar. For the first 100 iterations, there is no difference in the convergence behavior with and without non-negativity. In the later iterations, differences in the RMSE can be observed when using the quadratic potential, whereas the curves essentially remain the same when using the edge-preserving potential. For the quadratic potential, we thus see a gain in convergence speed when enforcing non-negativity; this gain is observed for a small improvement in an already small RMSE.

B. Sparse View Sampling of Projection Data

The difference images for the sparse view sampling are shown in Fig. 4. For the edge-preserving regularization, in Figs. 4(a) and (b), the difference inside the phantom is a noisy pattern with little to no structural information. For the quadratic regularization, in Figs. 4(c) and (d), the differences inside the phantom show a noisy pattern with enhancement of the edges with the sharp objects. Compared to the results with full view sampling, a larger grayscale window was required for display. This visual impression is confirmed by the summary values in Table II. The profiles through the edge of the air

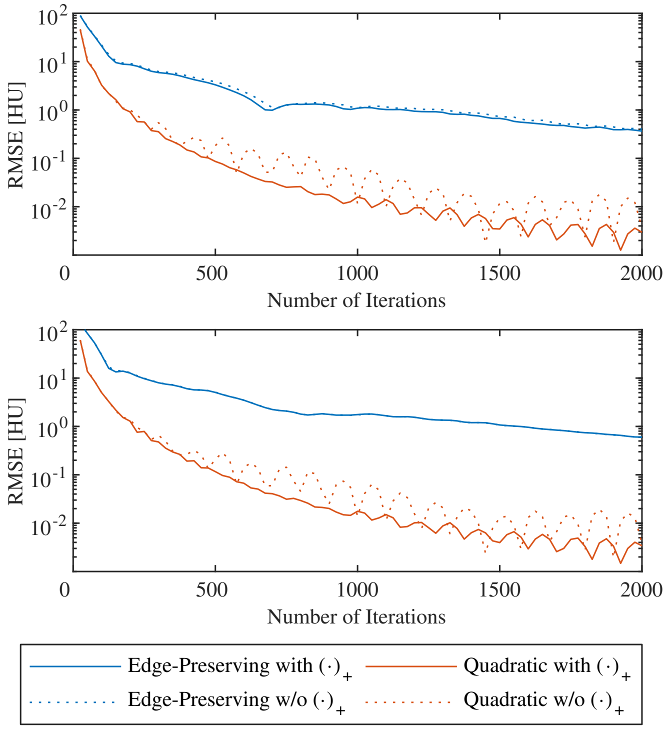


Fig. 3. Convergence behavior for reconstruction with *full view sampling* of module A (top) and module D (bottom). A solid (resp. dotted) line is used for reconstruction with and without the non-negativity constraint (resp.). The potential function is either edge-preserving (blue) or quadratic (orange).

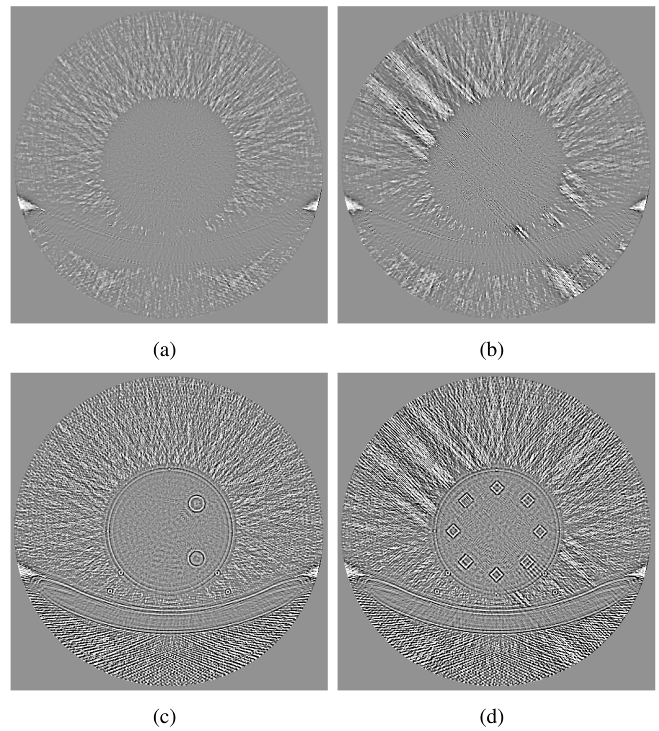


Fig. 4. Difference images for reconstruction with and without non-negativity constraint. These images are for *sparse view sampling* of module A (left) and module D (right). The applied potential function is either edge-preserving (top) or quadratic (bottom). Grayscale window: $[-20, 20]$ HU.

cylinder in module A are compared in Fig. 5, where we see that the non-negativity constraint leads to a slightly sharper profile, though the difference is tiny and likely not significant for a human observer.

The convergence behavior is depicted in Fig. 6. It has the same characteristics as seen in Fig. 3, showing that the non-negativity constraint does not impact convergence speed for the edge-preserving regularization, but does impact it, at a similar rate, for the quadratic regularization.

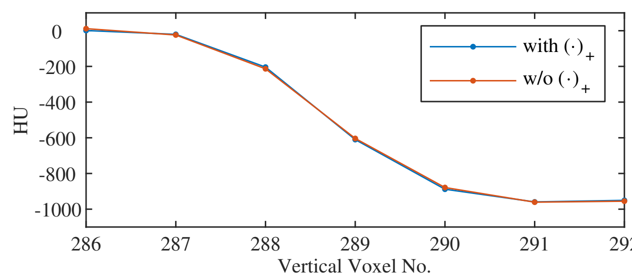


Fig. 5. Profiles through the edge of the air cylinder in module A for reconstruction with quadratic regularization and sparse view sampling. The profiles show the voxel value for the edge crossing from water to air, in HU.

TABLE II
MINIMUM, MAXIMUM, MEAN, AND STANDARD DEVIATION OF THE DIFFERENCE IMAGES. ALL VALUES ARE IN HU AND MEASURED INSIDE THE PHANTOM.

	Min.	Max.	Mean	SD
Full view sampling				
Module A, edge-preserving reg.	-1.50	1.85	-0.02	0.03
Module A, quadratic reg.	-0.20	0.05	-0.05	0.03
Module D, edge-preserving reg.	-1.61	1.36	-0.09	0.09
Module D, quadratic reg.	-0.79	0.35	-0.12	0.10
Sparse view sampling				
Module A, edge-preserving reg.	-13.02	11.29	-0.08	1.09
Module A, quadratic reg.	-29.17	26.45	-0.14	3.77
Module D, edge-preserving reg.	-42.90	38.98	-0.16	3.56
Module D, quadratic reg.	-47.39	54.89	-0.26	7.01

V. Discussion and Conclusions

In this work, we have reported results assessing the impact of the non-negativity constraint on image appearance and convergence speed under eight different scenarios based on real CT data. These scenarios were selected to challenge the potential usefulness of the non-negativity constraint.

Our observations in terms of image quality were as follows. For experiments with full view sampling, we observed that the use of the non-negativity constraint has only minimal impact on the image appearance inside the object. This was the case for both quadratic and edge-preserving regularization. The differences were on the order of 1-2 HU, and thus would likely have no effect on human observer performance.

For the more challenging problem of reconstruction from sparse view sampling, we observed more important differences: (i) enhancement of sharp edges, indicating a slight difference in resolution for sharp-contrast features, (ii) a noisy pattern inside the object. The difference in resolution is not likely to be significant for human observer performance but the noisy pattern might; this aspect requires further investigation.

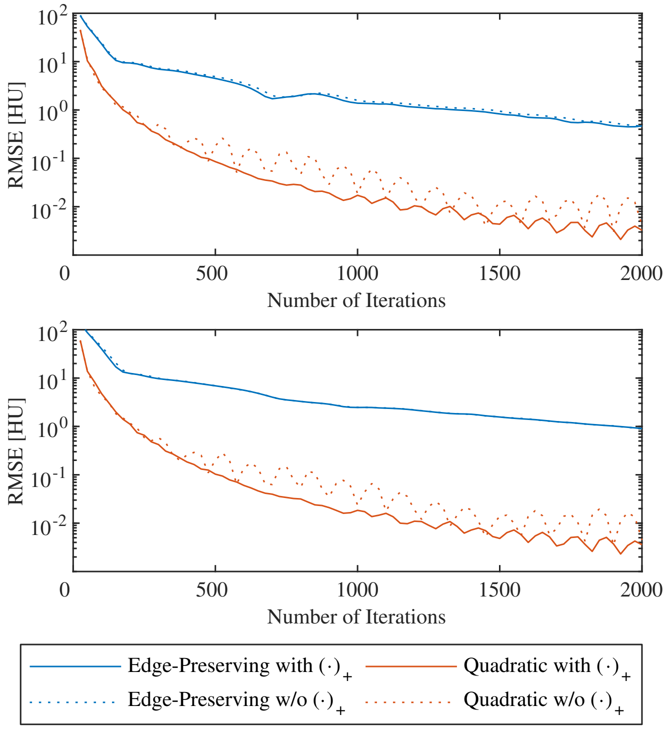


Fig. 6. Convergence behavior for reconstruction with *sparse view sampling* of module A (top) and module D (bottom). A solid (resp. dotted) line is used for reconstruction with and without the non-negativity constraint (resp.). The potential function is either edge-preserving (blue) or quadratic (orange).

Also of importance is the observation that the differences are of a fairly smaller magnitude for reconstruction with edge-preserving regularization.

In terms of convergence speed, we only identified benefits in using the non-negativity constraint for reconstruction with the quadratic penalty. In this case, the same benefit was observed for both full and sparse view sampling. However, this benefit plays a role only for improving the convergence when the RMSE is already small (below 1 HU).

Overall, we conclude that the non-negativity constraint may not offer any benefit for conventional diagnostic CT imaging, but could possibly slightly help for reconstruction under challenging conditions like sparse view sampling. To further verify this conclusion, a wider range of objective functions and reconstruction algorithms should be examined. Also, the outcome of our experiments could be affected by the choice of the phantom. A more complex anthropomorphic object with several air cavities (e.g., as encountered in lung imaging of patients with COPD) should be the basis for further investigations.

From an algorithm viewpoint, it is important to know that there are situations where the non-negativity constraint has little impact on image quality, because, for these situations, the choice of the optimization algorithm would not be restricted to those that can handle the non-smooth indicator function for non-negative values. This can have an impact on reconstruction speed.

Disclaimer

The concepts and information presented in this paper are based on research and are not commercially available.

Acknowledgment

This project was partly supported by Siemens Healthcare GmbH and partly by the National Cancer Institute of the National Institutes of Health under R21CA211035.

References

[1] Vardhanabhuti, V., Riordan, R. D., Mitchell, G. R., Hyde, C., and Roobottom, C. A., "Image comparative assessment using iterative reconstructions: clinical comparison of low-dose abdominal/pelvic computed tomography between adaptive statistical, model-based iterative reconstructions and traditional filtered back projection in 65 patients," *Investigative radiology* **49**(4), 209–216 (2014).

[2] Ichikawa, Y., Kitagawa, K., Nagasawa, N., Murashima, S., and Sakuma, H., "CT of the chest with model-based, fully iterative reconstruction: comparison with adaptive statistical iterative reconstruction," *BMC medical imaging* **13**(1), 27 (2013).

[3] Desai, G. S., Uppot, R. N., Elaine, W. Y., Kambadakone, A. R., and Sahani, D. V., "Impact of iterative reconstruction on image quality and radiation dose in multidetector CT of large body size adults," *European radiology* **22**(8), 1631–1640 (2012).

[4] Pickhardt, P. J., Lubner, M. G., Kim, D. H., Tang, J., Ruma, J. A., del Rio, A. M., and Chen, G.-H., "Abdominal ct with model-based iterative reconstruction (MBIR): initial results of a prospective trial comparing ultralow-dose with standard-dose imaging," *American journal of roentgenology* **199**(6), 1266–1274 (2012).

[5] Thibault, J.-B., Sauer, K. D., Bouman, C. A., and Hsieh, J., "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Medical physics* **34**(11), 4526–4544 (2007).

[6] Hahn, K., Schöndube, H., Stierstorfer, K., Hornegger, J., and Noo, F., "A comparison of linear interpolation models for iterative CT reconstruction," *Medical physics* **43**(12), 6455–6473 (2016).

[7] Haase, V., Griffith, A., Guo, Z., Oktay, M., Hahn, K., Schöndube, H., Stierstorfer, K., and Noo, F., "Penalized least-square CT reconstruction without and with statistical weights: effect on lesion detection performance with human observers," in [*Proceedings of the 14th International Meeting on Fully Three-dimensional Image Reconstruction in Radiology and Nuclear Medicine*], 831–835 (2017).

[8] Joseph, P. M., "An improved algorithm for reprojecting rays through pixel images," *IEEE transactions on medical imaging* **1**(3), 192–196 (1982).

[9] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences* **2**(1), 183–202 (2009).

[10] Nesterov, Y., "A method for unconstrained convex minimization problem with the rate of convergence $\mathcal{O}(1/k^2)$," in [*Doklady AN USSR*], **269**, 543–547 (1983). (In Russian).

[11] Sidky, E. Y., Jørgensen, J. H., and Pan, X., "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Physics in Medicine & Biology* **57**(10), 3065 (2012).