An Improved Method of Total Variation Superiorization Applied to Reconstruction in Proton Computed Tomography

Blake Schultze*, Yair Censor†, Paniz Karbasi*, Keith E. Schubert* *Senior Member, IEEE*, and Reinhard W. Schulte‡ *Member, IEEE*

*Department of Electrical and Computer Engineering, Baylor University, Waco, TX 76798, USA, email: Blake_Schultze@baylor.edu, Paniz_Karbasi@baylor.edu, Keith_Schubert@baylor.edu

†Department of Mathematics, University of Haifa, Haifa 3498838, Israel, email: yair@math.haifa.ac.il

‡Department of Basic Sciences, Division of Biomedical Engineering Sciences, Loma Linda University, Loma Linda, CA 92350 USA, email: rschulte@llu.edu

***Abstract*—Previous work showed that total variation superiorization (TVS) improves reconstructed image quality in proton computed tomography (pCT). The structure of the TVS algorithm has evolved since then and this work investigated if this new algorithmic structure provides additional benefits to pCT image quality. Structural and parametric changes introduced to the original TVS algorithm included: (1) inclusion or exclusion of TV reduction requirement, (2) a variable number, N, of TV perturbation steps per feasibility-seeking iteration, and (3) introduction of a perturbation kernel $0 < \alpha < 1$. The structural change of excluding the TV reduction requirement check tended to have a beneficial effect for $3 \leq N \leq 6$ and allows full parallelization of the TVS algorithm. Repeated perturbations per feasibility-seeking iterations reduced total variation (TV) and material dependent standard deviations for $3 \leq N \leq 6$. The perturbation kernel α, equivalent to $\alpha = 0.5$ in the original TVS algorithm, reduced TV and standard deviations as α was increased beyond $\alpha = 0.5$, but negatively impacted reconstructed relative stopping power (RSP) values for $\alpha > 0.75$. The reductions in TV and standard deviations allowed feasibility-seeking with a larger relaxation parameter λ than previously used, without the corresponding increases in standard deviations experienced with the original TVS algorithm. This work demonstrates that the modifications related to the evolution of the original TVS algorithm provide benefits in terms of both pCT image quality and computational efficiency for appropriately chosen parameter values.**

***Index Terms*—feasibility-seeking algorithms, image reconstruction, perturbations, proton computed tomography (pCT), superiorization, total variation superiorization (TVS)**

I. Introduction

PROTON computed tomography (pCT) is a relatively new imaging modality that has been developed from early beginnings [1], [2], [3], [4] towards a recent preclinical realization of a pCT scanner [5], [6]. The main motivation of pCT has been to improve the accuracy of proton therapy dose planning due to more accurate maps of relative stopping power (RSP) with respect to water, which determines how protons lose energy in human tissues in reference to water as a medium. The same method can also be used to image the patient immediately before treatment to verify the accuracy of the treatment plan about to be delivered. Proton therapy, like therapy with other heavy charged particles, e.g., carbon ions, is very susceptible to changes in tissue RSP, and small differences of a few percent can lead to range errors exceeding the desired limit of 1-2 mm. Thus, the planner of proton and ion therapy must increase margins around the target, which leads to unwanted exposure of normal tissues to high dose.

The faithful reconstruction of proton RSP maps is an important part of successful implementation of pCT. The approach that has been selected as the most promising in recent years, although technologically demanding, is to track individual protons through the patient and to predict their most likely path (MLP) [7], [8] in addition to measuring the energy loss of each proton and converting it to water-equivalent pathlength (WEPL). This has led to pCT reconstruction algorithms that are based on solving large and sparse linear equation systems, where each equation has the linear combination of intersection lengths of tracked protons through individual object voxels and the unknown RSP of those voxels on the left-hand side of the equation and the measured WEPL on the right-hand side of the equation. A solution of such large systems can be found with algorithms using projections onto convex sets and solving them iteratively as shown previously [9].

The superiorization method (SM) is another relatively recent development that has found its place between feasibility-seeking and constrained optimization in medical physics applications [10]. The superiorization method has also been tested as a technique to improve the image quality of pCT images when combined with the DROP algorithm [11]. Superiorization reduces, not necessarily minimizes, the value of a target function while seeking constraints-compatibility. This is done by taking a solely feasibility-seeking algorithm, analyzing its perturbation resilience, and proactively perturbing its iterates accordingly to steer them toward a feasible point with reduced value of the target function. When the perturbation steps are computationally efficient, this enables generation of a superior result with approximately the same computational cost as that of the original feasibility-seeking algorithm.

The mathematical principles of the SM over general consistent "problem structures" with the notion of bounded perturbation resilience were formulated in [12]. The framework of the SM was extended to the inconsistent case by using the

notion of strong perturbation resilience in [13], [14]. In [14], the efficacy of the SM was also demonstrated by comparing it with the performance of the projected subgradient method for constrained minimization problems.

A comprehensive overview of the state of the art and current research on superiorization appears in our continuously updated bibliography Internet page, which currently contains 74 items [15]. Research works in this bibliography include a variety of reports ranging from new applications to new mathematical results of the foundation of superiorization. A special issue entitled: "Superiorization: Theory and Applications" of the journal Inverse Problems appeared in [16].

Recently published works also attest to the advantages of the superiorization methodology in x-ray CT image reconstruction. These include reconstruction of CT images from sparse-view and limited-angle polyenergetic data [17], statistical tomographic image reconstruction [18], CT with total variation and with shearlets [19], and superiorization-based multi-energy CT image reconstruction [20].

In this work, we report on improvements of image quality and computational efficiency when applying novel modifications of superiorization to pCT reconstruction.

II. Motivation

Iterative projection methods seeking feasible solutions have been shown to be an effective image reconstruction technique for pCT [21], but reconstructed images exhibit local RSP fluctuations that cannot be removed by the reconstruction process alone. Inelastic electronic and nuclear events result in a statistical distribution of energy loss and, consequently, of the water-equivalent path length (WEPL) values calculated from measurements. These statistical variations in WEPL manifest in the reconstructed image as correlated localized fluctuations in the reconstructed RSP values. Although iterative reconstruction algorithms [9] are less sensitive to these variations than reconstruction transform methods, such as filtered backprojection (FBP) [22], [23], there is a propagation and amplification of WEPL uncertainty with successive iterations. Hence, although accuracy tends to increase with each iteration, as reconstruction nears convergence, updates of the solution from subsequent iterations are increasingly dominated by growing fluctuations. Thus, beyond a certain number of iterations, image quality begins to degrade, placing a limit on the maximum number of useful iterations and preventing steady-state convergence. WEPL uncertainty is inherent in the physical process and cannot be avoided, but techniques have been developed to reduce fluctuations and limit their propagation during iterative reconstruction. Given the amplification of uncertainty in the iterative process, any reduction in local RSP variations may lead to improved convergence behavior and, therefore, increase the accuracy of reconstructed RSP values.

The two measures that were adopted to quantify the prevalence and magnitude of these RSP fluctuations are total variation (TV) and standard deviation. For an introduction to TV for image analysis see, e.g., [24]. Standard deviation is a commonly used measure of variability around the mean in statistics. In image analysis, it is often employed to characterize the amount of fluctuations present in a region of interest that is known to present a homogeneous material. These measures provide a basis for comparing the effectiveness of techniques developed to address this problem. A technique for diminishing the impact of WEPL uncertainty and reducing RSP fluctuations is incorporating total variation superiorization (TVS) into reconstruction, with steepest descent steps of TV interlaced between iterations of an iterative feasibility-seeking algorithm.

Feasibility-seeking tends to accentuate RSP variations present due to WEPL uncertainty. Although this sharpens edges between different material regions, it also results in an amplification of RSP fluctuations during iterative image reconstruction. Performing TV reduction steps between consecutive feasibility-seeking iterations slows the growth of RSP variations. This permits more feasibility-seeking iterations before fluctuations grow to dominate updates of the solution. Hence, although the reduction in TV is itself an important aspect of TVS, another important aspect is the increased number of useful iterations made possible by the reduction in the amplification of RSP fluctuations.

III. Methods

A. TVS Algorithms

The efficacy of TVS for image reconstruction in pCT has been demonstrated in previous work [11]. In recent years, the algorithmic structure of the superiorization method has undergone some evolution in ways that offer several potential benefits in pCT. The details of this evolution can be found in the Appendix of [25], titled "The algorithmic evolution of superiorization". In addition, there were certain aspects of the original TVS algorithm, here referred to as OTVS, that were not previously investigated.

The new version of the TVS algorithm, here referred to as NTVS, investigated both the structural changes and those aspects previously not investigated of the OTVS algorithm. The notation of both algorithms and other algorithmic details can be found in the Appendix.

B. NTVS Algorithm

The NTVS algorithm investigated in this work has the following properties that were not present in previous work of applying TVS to pCT:

(1) Exclusion of the TV reduction verification step (step (10) of the OTVS algorithm in Appendix B).
(2) Usage of powers of the perturbation kernel α to control the step-sizes β_k in the TV perturbation steps.
(3) Incorporation of the user-chosen integer N (step (8) of the NTVS algorithm in Appendix C) that specifies the number of TV perturbation steps between consecutive feasibility-seeking iterations.
(4) Incorporation of a new formula for calculating the power ℓ_k, $\ell_k = \text{rand}(k, \ell_{k-1})$, used to calculate the step-size $\beta_k = \alpha^{\ell_k}$ at iteration k of feasibility-seeking (step (6) of the NTVS algorithm in Appendix C).

The step verifying the reduction of TV (step (10) of the OTVS algorithm in Appendix B) is not time consuming, but

such decision-controlled branches present their own challenges with respect to computational efficiency. Although there are technically a few computations with data dependencies (e.g., norm calculations), in each case, these can either be rearranged/reformulated or simply repeated separately to generate data independent calculations, making parallel computation of the algorithm possible. Hence, if the branching introduced by the TV reduction verification can be removed without compromising image quality, the NTVS algorithm can be incorporated into the existing parallelization scheme, providing up to a 30% reduction in its computational cost. This option was therefore explored in this work.

The OTVS algorithm, initializing TVS with $\beta_0 = 1$ and simply halving the perturbation magnitude each time through the TV perturbation loop prevented access to one of the most influential variables of TVS: the perturbation kernel α. With the magnitude of the perturbations given by $\beta_k = \alpha^{\ell_k}$, convergence is maintained by requiring $0 < \alpha < 1$. The primary purpose of α is to control the rate at which β_k converges to zero. In OTVS, $\beta_0 = 1$ and $\alpha = 0.5$ results in a relatively modest initial perturbation and a rapidly decreasing β such that little to no perturbation is applied after the first few feasibility-seeking iterations. Hence, OTVS perturbations applied after subsequent feasibility-seeking iterations are unlikely to have a meaningful impact on the amplification of RSP variations. This results in an overall under-utilization of TV perturbations. Thus, NTVS provides direct control of α, and its performance for various values of α was investigated in this work.

With the ability to increase the perturbation kernel α, larger reductions in TV can be generated; this also produces slower-decaying perturbations, which may not be desired. Alternatively, larger reductions in TV can also be generated by applying perturbations multiple times per feasibility-seeking iteration without increasing the magnitude of individual perturbations. Hence, NTVS introduces a variable N controlling the number of repetitions of TV perturbations between feasibility-seeking iterations.

Since the exponent ℓ increases after each of the N applied perturbations, reducing the perturbation coefficient $\beta_k = \alpha^{\ell}$, an increase in N causes the perturbation magnitude to converge to zero earlier in reconstruction. To preserve meaningful perturbations in later iterations, the exponent ℓ is adjusted between feasibility-seeking iterations by decreasing it to a random integer between its current (potentially large) value and the (potentially much smaller) iteration number k, i.e., $\ell_k = \text{rand}(k, \ell_{k-1})$. This update was suggested and justified in [26, page 38], [27, page 36] and subsequently used in [28] for maximum likelihood expectation maximization (MLEM) algorithms and in the linear superiorization (LinSup) algorithm [25, Algorithm 4]. Although the particular value of ℓ_k is random, on average, the corresponding perturbation coefficient β experiences a sizeable increase. This slows the rate at which β_k converges towards zero while preserving the convergence property given that the iteration number k, which increases sequentially, is set as the lower limit.

C. Input Data Set

The data set used as input to image reconstruction, containing about 120M proton histories, was generated from a Geant4 [29] simulated scan of the Catphan® CTP404 phantom module (The Phantom Laboratory Incorporated, Salem, NY). The phantom is a 15 cm diameter by 2.5 cm tall cylinder composed of an epoxy material with an RSP ≈ 1.0, which was set to RSP $= 1.0$ for simulation purposes. The phantom has three geometric types of contrasting material inserts embedded with the centers of each arranged in evenly spaced circular patterns of varying diameter. See Figure 1.

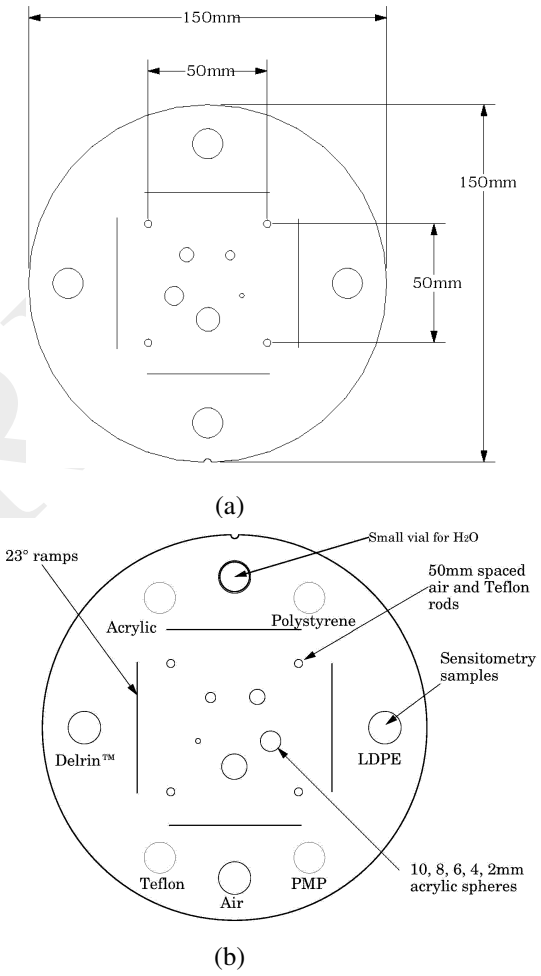


Fig. 1: Catphan CTP404 phantom: (a) dimensions and (b) composition and geometry of the material inserts.

D. Data Processing and Image Reconstruction

Details of the pCT data preprocessing, calibration, and image reconstruction have been presented previously [30], [31]. For the purposes of this work, feasibility-seeking was performed using the diagonally-relaxed orthogonal projections (DROP) algorithm of [32] with blocks containing 3200 proton histories. This block-size was chosen since smaller block-sizes are more sensitive to WEPL uncertainty. Image reconstruction was performed within a $20 \times 20 \times 5$ cm volume with each

voxel representing a volume of $1.0 \times 1.0 \times 2.5$ mm, yielding 200×200 image matrix for each slice.

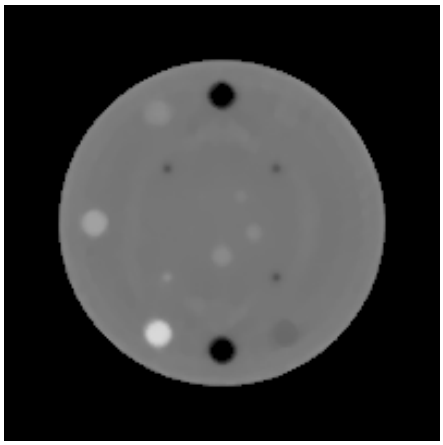


Fig. 2: Representative reconstruction of the central slice.

E. Reconstruction Parameter Space

The following describes the choices for reconstruction parameters that were systematically investigated in this work.

1) Inclusion or exclusion of TV reduction requirement: The primary structural change to the OTVS algorithm is the option to exclude the requirement that a perturbation reduces image TV, thereby eliminating the need to calculate and compare image TV before and after perturbations. Hence, NTVS was investigated with and without this check.

2) The number of TV perturbations per feasibility-seeking iteration: After initial investigations with increasing N, results were found to degrade as N increased beyond $N \approx 10$. Therefore, in this work the values of N chosen were between 1 and 12, in increments of 1.

3) The perturbation kernel coefficient: Since the configuration of the OTVS algorithm effectively used $\alpha = 0.5$ and the resulting perturbations did not negatively effect RSP accuracy [11], this work only investigated with $\alpha > 0.5$ to determine how large it can be set without affecting RSP accuracy. The values of α investigated in this work were $\alpha = 0.5, 0.65, 0.75, 0.85,$ and 0.95.

4) The choice of relaxation parameter in the feasibility-seeking algorithm: In previous unpublished work $\lambda = 0.0001$ yielded optimal results for a block-size containing 3200 proton histories; increasing λ beyond this value results in increased standard deviations. To investigate the interaction between TVS parameters and λ and determine if NTVS is capable of reducing the increase in standard deviations, the values of λ investigated in this work were $\lambda = 0.0001, 0.00015,$ and 0.0002.

IV. Computational Hardware and Performance Analysis

Image reconstruction was executed on a single node of a compute cluster with input data read from a local solid state drive and the bulk of computation was performed in parallel on a single NVIDIA k40 GPU. The parallel computational efficiency of the DROP algorithm increases as the number of histories per block increases since this permits better GPU utilization, but even with only 3200 histories per block, the total computation time from reading of input data from disk through the writing of reconstructed images to disk was only about 6 minutes.

The central phantom slices containing the spherical inserts have the most complicated material composition and represent the greatest challenge to reconstruction. Consequently, the data acquired for protons passing through these slices will have a greater variance in paths and WEPL values, which manifests in the corresponding slices of the reconstructed images as an increase in noise. Hence, analysis of these slices provides a better basis for comparing the NTVS and OTVS algorithms. Since the coplanar centers of the spherical inserts lie in the central slice, the comparative analyses performed in this work focused on this slice. A representative reconstruction of this slice is shown in Figure 2.

The image analysis program ImageJ2 1.51r [33] was used to perform quantitative analyses of reconstructed image quality. Its ovular selection and measurement tool was used to select a 7 mm diameter circular region of interest within the boundary of each cylindrical insert and calculate the mean and standard deviation in reconstructed RSP, with identical region selection and analyses performed for all images reconstructed using the OTVS and NTVS algorithms. RSP error was defined as the percentage difference between the mean measured RSP in a region of interest and the RSP defined for the material in the Geant4 simulation.

V. Results

In the following we present results from an investigation of the multi-parameter space, including potential interactions between parameters.

A. Number of TVS steps (N)

The number of TV perturbations per feasibility-seeking iteration, N, was varied between 1 and 12 in increments of 1. Figure 3 shows the dependence of TV as a function of N for each of the first four feasibility-seeking iterations with the TV reduction requirement excluded. As will be shown later, a similar pattern was observed with the TV reduction requirement included. The general effect of increasing N was a reduction in TV that leveled off after $N \geq 5$ steps, as best seen in the $k = 1$ iteration plot (top left of Figure 3). An irregular oscillation in TV as a function of increasing N appears for $k \geq 2$ and increases in magnitude as the number of feasibility-seeking iterations k increases. This oscillatory behavior can be understood to be the result of the opposing effects on TV of the alternating applications of TVS and feasibility-seeking. A detailed discussion of this phenomenon is outside the scope of this work.

For $3 \leq N \leq 6$, there was a benefit from NTVS compared to OTVS, which persisted throughout all twelve feasibility-seeking iterations (see Figure 4(a)). However, for $N \geq 7$ the benefits of NTVS were increasingly lost as N and k increased. This can be explained by the decreasing magnitude

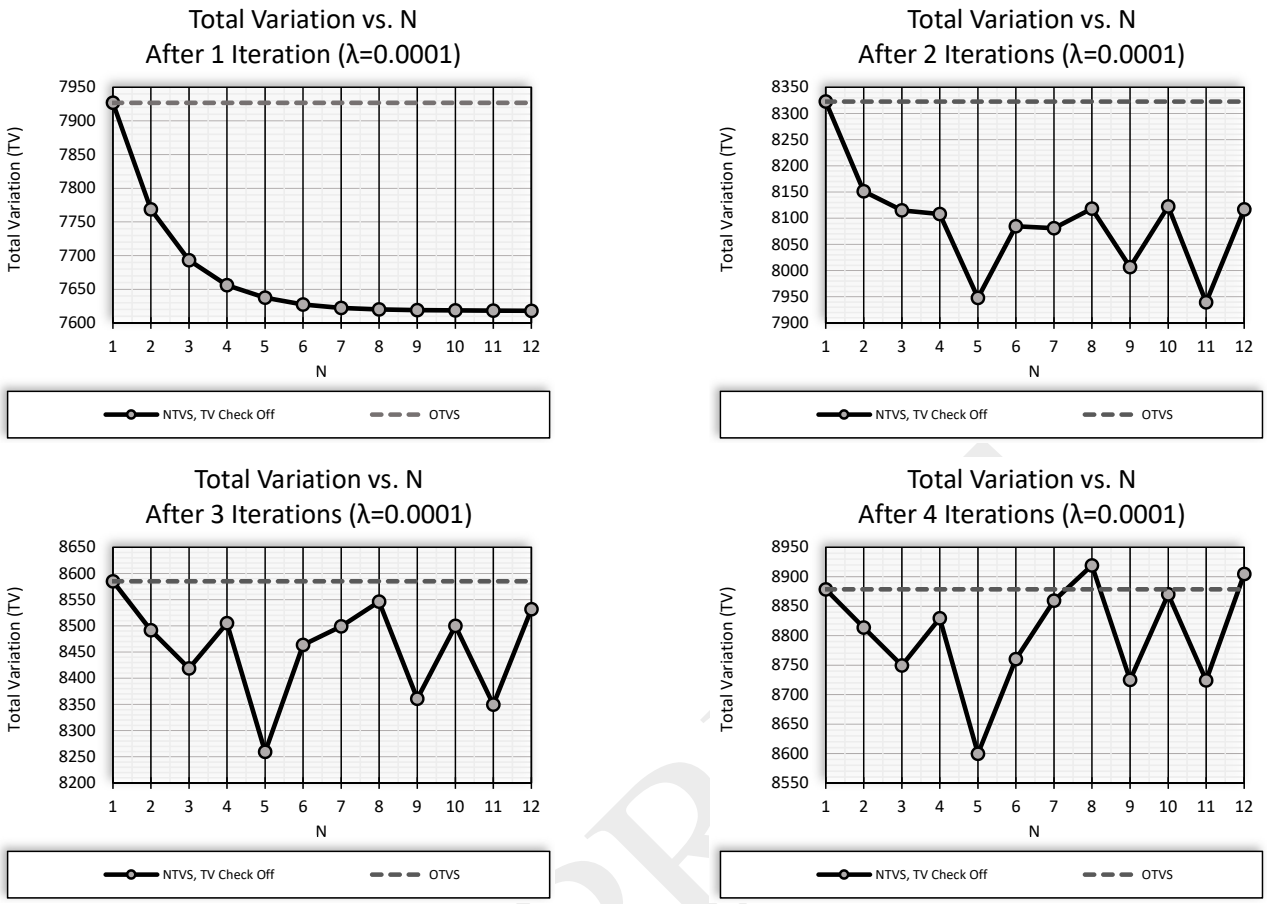


Fig. 3: TV as a function of N after each of the first 4 feasibility-seeking iterations.

of TV reducing perturbations with increasing N and the overall increase in TV from each feasibility-seeking iteration. Although not shown here, a similar dependence on N and k was seen for regional standard deviations. However, the benefit of NTVS in terms of standard deviation was consistently seen, including for $N \geq 7$, after twelve feasibility-seeking iterations (see, e.g., Figure 4(b)).

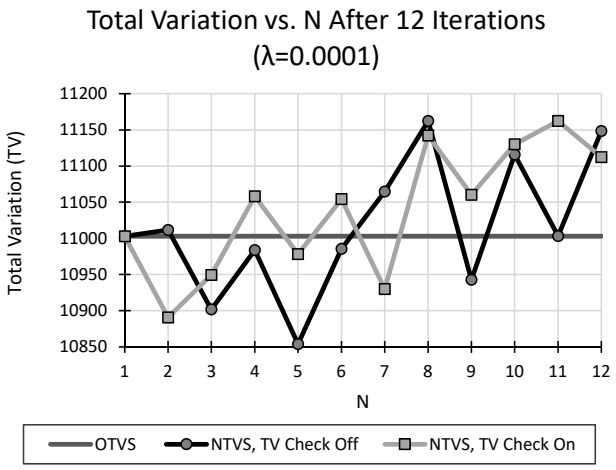
B. Inclusion/Exclusion of TV Reduction Requirement

Figures 4(a) and 4(b) show the comparison of TV and standard deviation, respectively, for OTVS and NTVS with relaxation parameter $\lambda = 0.0001$, median filter radius $r = 2$ applied to the initial iterate [30], and 12 feasibility-seeking iterations. In each plot, the results for NTVS with and without inclusion of the TV reduction requirement are shown as a function of N. The horizontal line corresponds to the result of OTVS ($N = 1$, $\alpha = 0.5$).

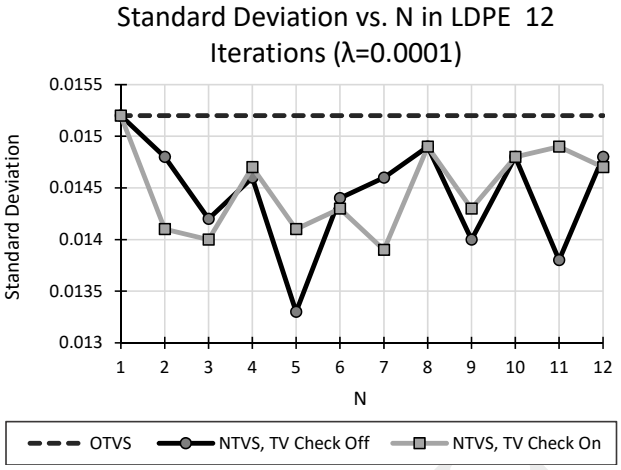
In the range of $3 \leq N \leq 6$, including the TV reduction requirement had practically no benefit, whereas its removal yields up to a 5.7% reduction in the standard deviation in RSP within the LDPE material insert and up to a 1.2% reduction in overall TV. Similar results were obtained for other values of α, λ, and, in the case of standard deviation, for different materials. One can conclude that imposing the TV reduction requirement does not provide a consistent benefit in terms of TV and standard deviation. Therefore, for the remainder of the parameter space exploration, the TV reduction requirement was excluded.

C. Perturbation Kernel (α)

Further investigations were performed to determine the effect of the perturbation kernel α (see step (10) of the NTVS algorithm in Appendix C) on TV and standard deviation for $0.5 \leq \alpha \leq 0.95$ and $1 \leq N \leq 12$. Increasing Increasing α produces larger perturbations and results in the perturbation magnitude β_k converging to zero more slowly. Thus, one can expect a larger reduction of TV and standard deviation for larger values of α. Figures 5(a) and 5(b) demonstrate this effect. Figures 6(a) and 6(b) show the effect of α on the accuracy of reconstructed RSP values in the Delrin and Polystyrene inserts, respectively. These two materials were chosen because they were most affected by the value of α. From these plots, one can see that for $\alpha > 0.75$, perturbations have a growing effect on RSP accuracy as α and N increase. This leads to changes in error greater than 1% for Delrin and greater than 0.5% for Polystyrene. Although increasing α to decrease TV and standard deviation is a worthwhile goal, one cannot do so without considering its effect on RSP error. On the other hand, increasing α from $\alpha = 0.5$ to $\alpha = 0.75$ yielded up to a 39.3% reduction in the standard deviation in RSP within the LDPE material insert and up to an 8.2% reduction in overall TV without negatively impacting RSP error.

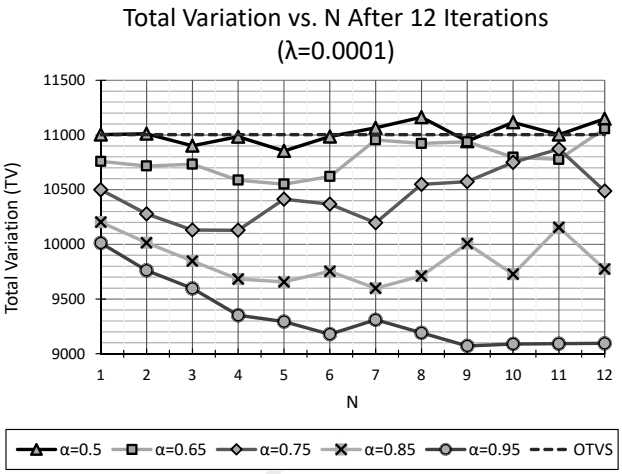


(a)

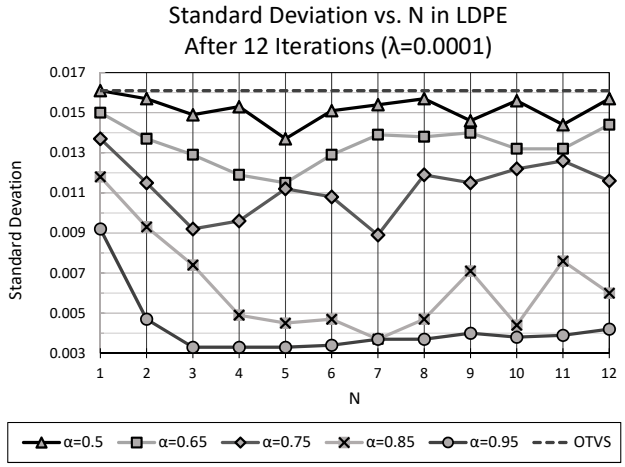


(b)

Fig. 4: (a) TV and (b) standard deviation after 12 DROP iterations for $\lambda = 0.0001$, with the OTVS algorithm and NTVS with and without the TV reduction requirement each plotted as a function of $1 \leq N \leq 12$.



(a)



(b)

Fig. 5: (a) TV and (b) standard deviation after 12 DROP iterations for $\lambda = 0.0001$, plotted for OTVS and NTVS as a function of $1 \leq N \leq 12$ with each curve representing a different value of the perturbation kernel α.

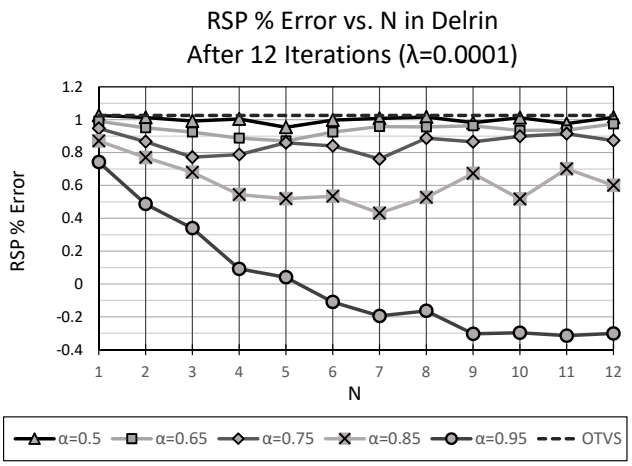
D. Relaxation Parameter (λ)

Increasing the relaxation parameter accelerates the rate of convergence of the feasibility-seeking algorithm. To investigate the impact of NTVS independent of convergence rate, the number of iterations was adjusted for $\lambda = 0.00015$ and $\lambda = 0.0002$ to obtain the same RSP accuracy as for $\lambda = 0.0001$ and 12 iterations. For this comparison, $\alpha = 0.75$ was chosen. Figures 7(a) and 7(b) show TV and standard deviation of the LDPE material, respectively, for $\lambda = 0.0001, 0.00015,$ and $\lambda = 0.0002$ after $k = 12, 8,$ and 6 iterations, respectively. For most values of N, the relative improvements in TV and standard deviation increase as λ increases. Note that the trend for standard deviation was not as pronounced for other materials, but increasing λ consistently produced comparable or larger reductions in TV and standard deviation in each material region.

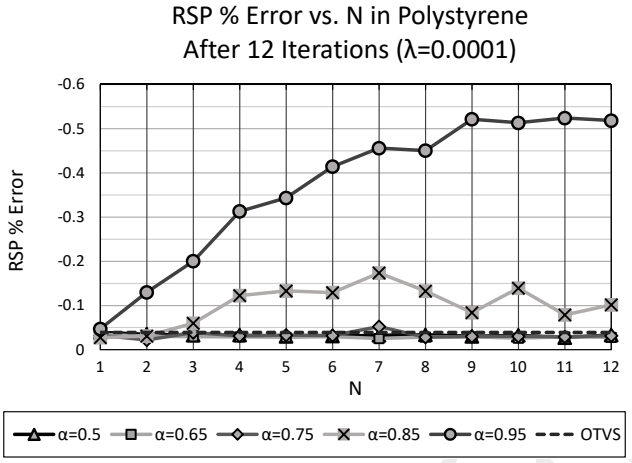
VI. Discussion

In this work, several innovations were introduced to the original TVS algorithm as applied to pCT, leading to an algorithm referred to here as NTVS. Compared to the original TVS algorithm, referred to here as OTVS, the NTVS algorithm yields larger reductions in TV and standard deviations within specific material regions. The NTVS algorithm also provides an opportunity to reduce computation time.

The benefit of NTVS was dependent on the number of perturbations per feasibility-seeking iteration, N, with the largest benefit consistently attained for $3 \leq N \leq 6$. For $N > 6$, these benefits decreased as the number of feasibility-seeking iterations, k, increased. This can be understood as the effect that larger N have on the magnitude of perturbations β_k as k increases. With ℓ_k increasing by 1 after each of the N perturbations, increasing N results in $\beta_k = \alpha^{\ell_k}$ decreasing more quickly as k increases. Hence, for larger

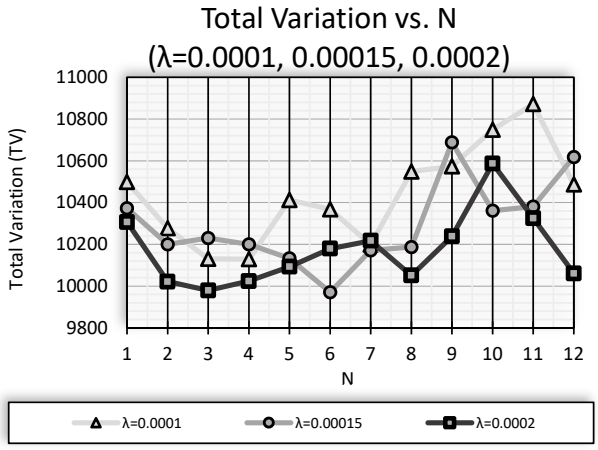


(a)

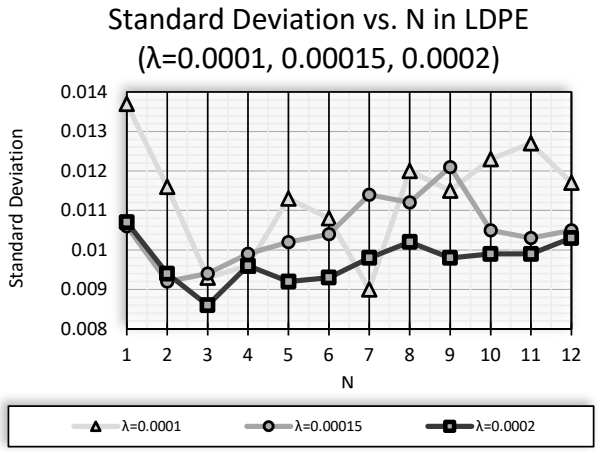


(b)

Fig. 6: RSP error in the (a) Delrin and (b) Polystyrene material inserts after 12 DROP iterations for $\lambda = 0.0001$, plotted for NTVS as a function of $1 \leq N \leq 12$ with each curve representing a different value of the perturbation kernel α.



(a)



(b)

Fig. 7: (a) TV and (b) standard deviation after an equivalent # of DROP iterations of the OTVS and NTVS with $\alpha = 0.75$, plotted as a function of $1 \leq N \leq 12$ with each curve representing a different relaxation parameter λ.

N, meaningful perturbations persist for a smaller number of feasibility-seeking iterations.

Whereas the inclusion of the TV reduction requirement in NTVS had no consistent discernable benefit, excluding this requirement had a tendency to improve results for $3 \leq N \leq 6$. Removing the TV reduction requirement also improves computation time by eliminating the conditional branch that prevents full parallelization of the superiorization algorithm. There are global calculations within the TVS algorithm, such as the ℓ_2 (discrete-space) norm used to normalize perturbation vectors, which act as a bottleneck in an explicit and direct implementation. However, such data dependencies can be eliminated by performing these calculations in each thread rather than communicating these from a central location. Hence, there are no real data dependencies and the parallelization made possible by removing the conditional branch reduces NTVS computation time by up to 30% (based on a count of the reduced number of sequential computational operations).

An appealing aspect of NTVS is the added ability to control the perturbation kernel α, which was previously held constant at a value of $\alpha = 0.5$ in OTVS. Increasing the perturbation kernel α yields larger reductions in TV and standard deviations. However, it was found that as α increased beyond $\alpha \approx 0.75$, perturbations begin to affect reconstructed RSP values in an unpredictable and material dependent manner. Thus, $\alpha = 0.75$ was chosen, as this retained the benefits of a larger reduction in TV and standard deviations without the undesirable effect on reconstructed RSP values.

Another benefit of NTVS is that it allows feasibility-seeking to be performed with a larger relaxation parameter λ than was optimal with OTVS ($\lambda = 0.0001$, $k = 12$). It was found that with NTVS the same RSP error can be obtained with $\lambda = 0.0002$ and $k = 6$ without producing larger standard deviations, as previously observed with OTVS with these parameter values. Arriving at an acceptable solution in $k = 6$ feasibility-seeking iterations offers substantial computational benefit.

In the work presented here, each of the TVS parameter values were held fixed throughout reconstruction. One possible direction to explore in future work is investigating how parameter values can be varied *during* reconstruction to produce greater benefits with NTVS.

Another interesting question to explore is if the diminishing benefits for $N \geq 7$ are due to an excessive use of TVS per feasibility-seeking iteration or if this is simply a consequence of β decreasing too quickly as a function of k, perhaps resulting in an under-utilization of TVS at larger values of k. Note that the value of the perturbation kernel α determines not only the initial perturbation magnitude $\beta(k = 1)$, but also the rate at which β decreases after each perturbation. One would like the ability to control the initial perturbation magnitude $\beta(N = 1)$ as a function of k while independently determining the rate at which β decreases between each of the N perturbations per feasibility-seeking iteration. This is not possible with $\beta = \alpha^{\ell}$ since ℓ is a function of both N and k.

Hence, an interesting direction to explore is the introduction of another parameter γ that controls the rate at which β decreases as a function of k. The parameter α would then control only the rate at which β decreases between each of the $n = 1...N$ perturbation steps. By reformulating the perturbation magnitude as $\beta = \alpha^n \gamma^{f(k)}$, with $0 < \alpha, \gamma < 1$ and $f(k)$ chosen such that $\lim_{k\to\infty} f(k) = \infty$ (e.g., $f(k) = k$), the rate at which β decreases as a function of N and k can then be controlled independently while preserving the superiorization requirement that $\lim_{k\to\infty} \beta = 0$.

VII. Conclusions

The investigations performed in this work demonstrate that the modifications implemented by the NTVS algorithm provide clear advantages over the OTVS algorithm in terms of both quality and computational cost. Future work should include investigating whether varying parameters during reconstruction or controlling the decrease of the perturbation magnitude independently during iterations and repeated perturbation steps can further increase the advantages of the NTVS algorithm.

Appendix A
Definition of Terms

The list below defines the terms and the corresponding mathematical notation used in describing the OTVS and NTVS algorithms:

- k : overall cycle #, i.e., k-th iteration of feasibility-seeking and TV perturbations.
- n : TV perturbation step #, $1 \leq n \leq N$.
- x^k : image vector x at cycle k.
- $\bar{x}$: initial iterate x^0 of image reconstruction.
- N : # of TV perturbation steps per feasibility-seeking iteration.
- α : perturbation kernel, $0 < \alpha < 1$.
- ℓ_k : perturbation kernel exponent.
- $\beta_{k,n}$: perturbation coefficient $\beta_{k,n} = \alpha^{\ell_k}$ at TV perturbation step n and feasibility-seeking iteration k.
- ϕ : the target function to which superiorization is applied; here, $\phi =$ TV, the total variation of the image vector.
- $\phi(x^{k,n})$: TV of image vector $x^{k,n}$ at TV perturbation step n and feasibility-seeking iteration k.
- $v^{k,n}$: normalized non-ascending perturbation vector for ϕ at $x^{k,n}$, i.e.,

$$v^{k,n} = -\frac{\nabla\phi(x^{k,n})}{||\nabla\phi(x^{k,n})||} = \phi'(x^{k,n})$$

.
- P_T : projection operator representative of an iterative feasibility-seeking algorithm.

Appendix B
OTVS Algorithm

A pseudocode definition of the OTVS algorithm is written as follows:

set $k = 0$
set $\ell = 0$
set $\beta = 1$
set $x^k = \bar{x}$
while *true* **do**
    set $v^k = \phi'(x^k)$
    set $loop = true$
    while *loop* **do**
        set $z = x^k + \beta v^k$
        if $\phi(z) \leq \phi(x^k)$ **then**
            set $x^k = z$
            set $loop = false$
        end if
        set $\ell = \ell + 1$
        set $\beta = (\frac{1}{2})^{\ell}$ (originally $\beta \leftarrow \beta/2$)
    end while
    set $x^{k+1} = P_T(x^k)$
    set $k = k + 1$
end while

Appendix C
NTVS Algorithm

A pseudocode definition of the NTVS algorithm is written as follows:

set $k = 0$
set $\ell_{-1} = 0$
set $x^k = \bar{x}$
while *true* **do**
    set $n = 0$
    set $\ell_k =$ rand(k, ℓ_{k-1})
    set $x^{k,n} = x^k$
    while $\boldsymbol{n < N}$ **do**
        set $v^{k,n} = \phi'(x^{k,n})$
        set $\beta_{k,n} = \alpha^{\ell_k}$
        set $x^{k,n+1} = x^{k,n} + \beta_{k,n} v^{k,n}$
        set $n = n + 1$
        set $\ell_k = \ell_k + 1$
    end while
    set $x^{k+1} = P_T(x^{k,N})$
    set $k = k + 1$
end while

ACKNOWLEDGMENT

The research in proton CT was supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institute of Health (NIH) and the National Science Foundation (NSF) award number R01EB013118, and the United States - Israel Binational Science Foundation (BSF) grant no. 2009012, and is currently supported by BSF grant no. 2013003. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of NBIB or NIH. The support of UT Southwestern and State of Texas through a Seed Grants in Particle Therapy award is gratefully acknowledged.

REFERENCES

[1] A. M. Cormack and A. M. Koehler, "Quantitative proton tomography: preliminary experiments," *Physics in Medicine & Biology*, vol. 21, no. 4, pp. 560–569, 1976.

[2] K. M. Hanson, J. N. Bradbury, T. M. Cannon, R. L. Hutson, D. B. Laubacher, R. Macek, M. A. Paciotti, and C. A. Taylor, "The application of protons to computed tomography," *IEEE Transactions on Nuclear Science*, vol. 25, no. 1, pp. 657–660, Feb 1978.

[3] K. M. Hanson, "Proton computed tomography," *IEEE Transactions on Nuclear Science*, vol. 26, no. 1, pp. 1635–1640, Feb 1979.

[4] K. M. Hanson, J. N. Bradbury, T. M. Cannon, R. L. Hutson, D. B. Laubacher, R. J. Macek, M. A. Paciotti, and C. A. Taylor, "Computed tomography using proton energy loss," *Physics in Medicine & Biology*, vol. 26, no. 6, pp. 965–983, 1981.

[5] R. Johnson, V. Bashkirov, G. Coutrakon, V. Giacometti, P. Karbasi, N. T. Karonis, C. Ordonez, M. Pankuch, H. F. W. Sadrozinski, K. Schubert, and R. W. Schulte, "Results from a prototype proton-CT head scanner," *Conference on the Application of Accelerators in Research and Industry, CAARI 2016, 30 October 4 November 2016, Ft. Worth, TX, USA*, 07 2017. [Online]. Available: https://arxiv.org/pdf/1707.01580

[6] V. A. Bashkirov, R. P. Johnson, H. F.-W. Sadrozinski, and R. W. Schulte, "Development of proton computed tomography detectors for applications in hadron therapy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 809, pp. 120 – 129, 2016.

[7] R. Schulte, S. Penfold, J. Tafas, and K. Schubert, "A maximum likelihood proton path formalism for application in proton computed tomography," *Med. Phys.*, vol. 35, pp. 4849–4856, November 2008.

[8] C.-A. C. Fekete, P. Doolan, M. F. Dias, L. Beaulieu, and J. Seco, "Developing a phenomenological model of the proton trajectory within a heterogeneous medium required for proton imaging," *Physics in Medicine & Biology*, vol. 60, no. 13, pp. 5071–5082, 2015.

[9] S. N. Penfold, R. W. Schulte, Y. Censor, V. Bashkirov, S. Macallister, K. E. Schubert, and A. B. Rosenfeld, "Block-iterative and string-averaging projection algorithms in proton computed tomography image reconstruction," in *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems*, Y. Censor, M. Jiang, and G. Wang, Eds., The Huangguoshu International Interdisciplinary Conference. Medical Physics Publishing, Madison, WI, USA, 2010, pp. 347–367.

[10] G. T. Herman, E. Garduo, R. Davidi, and Y. Censor, "Superiorization: An optimization heuristic for medical physics," *Medical Physics*, vol. 39, no. 9, pp. 5532–5546, 2012. [Online]. Available: http://dx.doi.org/10.1118/1.4745566

[11] S. Penfold, R. Schulte, Y. Censor, and A. Rosenfeld, "Total variation superiorization schemes in proton computed tomography image reconstruction," *Med. Phys.*, vol. 37, pp. 5887–5895, 2010.

[12] Y. Censor, R. Davidi, and G. T. Herman, "Perturbation resilience and superiorization of iterative algorithms," *Inverse problems*, vol. 26, p. 65008, June 2010.

[13] Y. Censor, "Weak and strong superiorization: Between feasibility-seeking and minimization," *Analele Stiintifice ale Universitatii Ovidius Constanta, Seria Matematica*, vol. 23, pp. 41–54, October 2014.

[14] Y. Censor, R. Davidi, G. T. Herman, R. W. Schulte, and L. Tetruashvili, "Projected subgradient minimization versus superiorization," *Journal of Optimization Theory and Applications*, vol. 160, no. 3, pp. 730–747, March 2014. [Online]. Available: https://doi.org/10.1007/s10957-013-0408-3

[15] Y. Censor, "Superiorization and perturbation resilience of algorithms: A bibliography compiled and continuously updated," http://math.haifa.ac.il/yair/bib-superiorization-censor.html.

[16] Y. Censor, G. T. Herman, and M. Jiang, "Superiorization: theory and applications," Special Issue of *Inverse Problems*, vol. 33, no. 4, p. 040301, 2017. [Online]. Available: http://stacks.iop.org/0266-5611/33/i=4/a=040301

[17] T. Humphries, J. Winn, and A. Faridani, "Superiorized algorithm for reconstruction of ct images from sparse-view and limited-angle polyenergetic data," *Physics in Medicine & Biology*, vol. 62, no. 16, p. 6762, 2017. [Online]. Available: http://stacks.iop.org/0031-9155/62/i=16/a=6762

[18] E. S. Helou, M. V. W. Zibetti, and E. X. Miqueles, "Superiorization of incremental optimization algorithms for statistical tomographic image reconstruction," *Inverse Problems*, vol. 33, no. 4, p. 044010, 2017. [Online]. Available: http://stacks.iop.org/0266-5611/33/i=4/a=044010

[19] E. Garduo and G. T. Herman, "Computerized tomography with total variation and with shearlets," *Inverse Problems*, vol. 33, no. 4, p. 044011, 2017. [Online]. Available: http://stacks.iop.org/0266-5611/33/i=4/a=044011

[20] Q. Yang, W. Cong, and G. Wang, "Superiorization-based multi-energy ct image reconstruction," *Inverse Problems*, vol. 33, no. 4, p. 044014, 2017. [Online]. Available: http://stacks.iop.org/0266-5611/33/i=4/a=044014

[21] S. Penfold and Y. Censor, "Techniques in iterative proton CT image reconstruction," *Sensing and Imaging*, vol. 16, no. 1, p. 19, October 2015.

[22] R. Bracewell and A. Riddle, "Inversion of fan beam scawns in radio astronomy," *Astrophysics Journal*, vol. 150, pp. 427–434, 1967.

[23] G. Ramanchandran and A. Lakshminarayanan, "Three dimensional reconstructions from radiographs and electron micrographs: Application of convolution instead of Fourier transforms," *Proc. Natl. Acad. Sci. USA*, vol. 68, pp. 2236–2240, 1971.

[24] A. Chambolle, M. Novaga, D. Cremers, and T. Pock, "An introduction to total variation for image analysis," in *Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter*, 2010.

[25] Y. Censor, "Can linear superiorization be useful for linear optimization problems?" *Inverse Problems*, vol. 33, no. 4, p. 044006, 2017. [Online]. Available: http://stacks.iop.org/0266-5611/33/i=4/a=044006

[26] O. Langthaler, "Incorporation of the superiorization methodology into biomedical imaging software," Salzburg University of Applied Sciences, Salzburg, Austria, and the Graduate Center of the City University of New York, NY, USA, Marshall Plan Scholarship Report, September 2014, 76 pages.

[27] B. Prommegger, "Verification and evaluation of superiorized algorithms used in. biomedical imaging: Comparison of iterative algorithms with and without superiorization for image reconstruction from projections," Salzburg University of Applied Sciences, Salzburg, Austria, and the Graduate Center of the City University of New York, NY, USA, Marshall Plan Scholarship Report, October 2014, 84 pages.

[28] C. Havas, "Revised implementation and empirical study of maximum likelihood expectation maximization algorithms with and without superiorization in image reconstruction," Salzburg University of Applied Sciences, Salzburg, Austria, and the Graduate Center of the City University of New York, NY, USA, Marshall Plan Scholarship Report, October 2016, 49 pages.

[29] S. Agostinelli, J. Allison, K. Amako *et al.*, "Geant4a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900203013688

[30] B. Schultze, P. Karbasi, V. Giacometti, T. Plautz, K. E. Schbert, and R. W. Schulte, "Reconstructing highly accurate relative stopping powers in proton computed tomography," in *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, Oct 2015, pp. 1–3.

[31] V. Giacometti, V. A. Bashkirov, P. Piersimoni, S. Guatelli, T. E. Plautz, H. F.-W. Sadrozinski, R. P. Johnson, A. Zatserklyaniy, T. Tessonnier, K. Parodi, A. B. Rosenfeld, and R. W. Schulte, "Software platform for simulation of a prototype proton CT scanner," *Medical Physics*, vol. 44, no. 3, pp. 1002–1016, 2017. [Online]. Available: http://dx.doi.org/10.1002/mp.12107

[32] Y. Censor, T. Elfving, G. T. Herman, and T. Nikazad, "On diagonally relaxed orthogonal projection methods," *SIAM Journal on Scientific Computing*, vol. 30, no. 1, pp. 473–504, 2008. [Online]. Available: https://doi.org/10.1137/050639399

[33] C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri, "Imagej2: Imagej for the next

generation of scientific image data," *BMC Bioinformatics*, vol. 18, no. 1, p. 529, Nov 2017. [Online]. Available: https://doi.org/10.1186/s12859-017-1934-z