Fast Superiorization Using a Dual Perturbation Scheme for Proton Computed Tomography

R. Davidi[1], R. W. Schulte[2], Y. Censor[3] and L. Xing[1]

[1]*Department of Radiation Oncology, Stanford University, Stanford, CA 94305, rdavidi@stanford.edu*
[2]*Department of Radiation Medicine, Loma Linda University Medical Center, Loma Linda, CA 92354*
[3]*Department of Mathematics, University of Haifa, Mt. Carmel, 31905 Haifa, Israel*

INTRODUCTION

The purpose of proton Computed Tomography (pCT) is to provide an image that accurately captures the relative stopping power (RSP) needed for proton treatment planning calculations. Series-expansion type methods [1] have been successfully employed in this area and have demonstrated good results [2,3]. These methods are used for solving a linear system of equations of the form

$$Ax = b, \tag{1}$$

where A is the system matrix with I rows and J columns, b is a vector with I elements representing the integral RSP measurements, and x is a J-dimensional vector representing the image which we need to reconstruct. The system (1) is often very large and requires substantial resources for solving it. It has the characteristic of also being sparse which makes it ideal when solved with iterative projection methods. These types of methods accept an initial point as part of an input and perform iterative updates to the image vector x. The updates are performed according to some projection scheme, that could be sequential or parallel (or any mix of the two, such as String-Averaging Projections (SAP) or Block Iterative Projection (BIP) methods). In pCT the parallel schemes are preferred since the rows of the matrix can then be partitioned into groups (blocks) where the rows in each block can be processed simultaneously, and achieve faster convergence as a result.

The projection methods discussed above belong to the family of *feasibility-seeking* algorithms. Their main characteristic is that they will accept any solution as long as it is consistent with the system (1). On the other hand, there are problems that require, in addition to finding a feasible solution, that the solution should also be a minimum (or a maximum) of some well-defined merit function; methods for doing this are referred to as *optimization* methods since they seek an optimal solution. A new methodology, called *superiorization*, was recently developed as a framework for algorithms that lie conceptually between feasibility-seeking and optimization algorithms. It is a heuristic tool that does not guarantee to find the optimum value of a given functional, rather it obtains a solution that is *superior* (with respect to the merit function) to what would be achieved if superiorization would not have been used (i.e., superior to a feasible solution). The advantage of an algorithm that uses superiorization as the driving tool (as opposed to optimization) is that it requires less computational resources while providing comparable solutions, from the point of view of real-world applications, to those that one would get with algorithms that use optimization. Superiorization relies on the general principle that many classes of projection methods (such as SAP and BIP) are perturbation resilient, meaning that every step in the iterative process can be perturbed and convergence to a feasible solution can still be retained. Mathematically we mean that if an algorithm, whose iterative step has the form

$$x^{k+1} = P(x^k), \tag{2}$$

where P is the algorithmic operator and k is a positive integer representing the current iteration index, converges, then so will the perturbed algorithm whose iterative step is either

$$x^{k+1} = P(x^k + \beta_k v^k)$$
$$\text{or} \tag{3}$$
$$x^{k+1} = P(x^k) + \beta_k v^k,$$

where β_k are nonnegative real numbers such that $\sum_{k=0}^{\infty} \beta_k < \infty$ and v^k are bounded vectors [4,5]. Superiorization uses these (bounded) perturbations to steer the iterates towards a solution that is *superior* with respect to a given merit function ϕ so that a superior feasible solution is found with respect to ϕ as opposed to just any feasible solution. This is achieved, for example, if at each iterative step the value of the function gets reduced as the iterations proceed.

Superiorization with total variation as the ϕ was successfully implemented for pCT. Here we take the approach one step further and suggest to have a *dual perturbation* scheme. The first perturbation of the scheme aims at superiorizing the ϕ while the second aims at improving our control of the balance between the work done to achieve feasibility and the work done to reduce the merit function. We discuss the details of our proposed method in the next section and then show its usefulness for pCT.

METHODS

Our suggested scheme uses two kinds of perturbations. The algorithm in each iterative step performs the first kind, followed by a projection operator (that is known to be bounded perturbation resilient), followed by the second kind of perturbations, where the output of the latter is considered the starting point for the next iteration. The first kind of perturbations aims at reducing the value of a given merit function ϕ. Here we use the Total Variation (TV) defined as

$$TV(w) = \sum_{r=1}^{n-1}\sum_{c=1}^{n-1}\sqrt{\left(w_{r+1,c} - w_{r,c}\right)^2 + \left(w_{r,c+1} - w_{r,c}\right)^2}, \quad (4)$$

where w is the image vector x of (1) represented as a 2D $n\times n$ array and r and c represent the index of the of rows and columns of w, respectively. The perturbations scheme for reducing the TV was chosen as in [4]. The essence of it relies on the fact that TV is a convex function and that moving in the negative (normalized) subgradient direction with a given step size insures that the value of TV be superiorized. Due to the normalization, the perturbation vector is bounded. The vector for the k^{th} iteration is v^k. The step size β_k is initially set to 1. To insure that the merit function is reduced, we check its value before and after the perturbation. If the value of the function of the new perturbed point is not reduced (superiorized), then the size of β_k is halved and the process repeats (with the same v^k) until the condition succeeds. The resulting sequence of β_k is summable, as required for bounded perturbations resilience.

The projection operator is then applied to the newly perturbed point. In this work we have chosen the DROP operator [6], which belongs to the BIP algorithmic family. In DROP, the rows of the system matrix A in (1) are partitioned into B blocks, where I_t is the set of row indices of the t^{th} block and the total number of rows is $I = I_1 \cup I_2 \cup \cdots \cup I_B$. The DROP operator was chosen as in [1] but here the blocks were partitioned so that the number of proton histories was not necessarily equal in each block (although their sizes are approximately the same). The DROP algorithm is defined as

$$x^{k+1} = x^k + \lambda_k U_{t(k)} \sum_{i\in I_{t(k)}} \frac{b_i - \langle a^i, x^k\rangle}{\|a^i\|^2} a^i, \quad (5)$$

where x^k and x^{k+1} are the current and next image vectors, respectively, λ_k is a sequence of user-chosen relaxation parameters, $U_{t(k)}$ is a diagonal matrix with an element on its diagonal equivalent to the minimum between 1 and the reciprocal of the number of nonzero intersections of the I^{th} row in $I_{t(k)}$ and the J^{th} pixel in x^k, and b_i and a^i correspond to the I^{th} row and element of A and b in the block $I_{t(k)}$, respectively. Here we have adopted the cyclic control, where $t(k) = k \bmod B + 1$ controls the order by which the blocks are picked and projected onto.

The resulting point of the projection operator is then perturbed for the second time. The aim of the second kind of perturbations is to enable us to better control the trade-off between feasibility-seeking and merit function reduction. Too intensive use of the feasibility-seeking projection method might give the superiorization algorithm insufficient time to properly reduce the merit function values in the process. Too little use of the feasibility-seeking projections method might be counterproductive to the quest for a solution that will agree with the constraints. Therefore, we introduce a "second" perturbation into the process. This second perturbation is periodically done with respect to a fixed perturbation vector, i.e., $v^k = v$ for all $k \geq 0$, but with another sequence of step-sizes β'_k. In our implementation, we have chosen a Filtered Backprojection (FBP) reconstruction as the single perturbation vector. In pCT, since the calculation of the history paths of the protons is considered an intense computational effort, it was suggested [3] to use an FBP reconstruction in order to provide the outer contour of the object and calculate the proton most likely paths only inside the object as opposed to the entire reconstruction region (outside the object straight lines are assumed when formulating the matrix A). Since an FBP reconstruction is already calculated as part of formulating the system matrix, there is no added computational cost for using it in our proposed method. Furthermore, the FBP reconstruction is based on a transform method that still relies on the data provided by the measurements. The intuition behind the choice of the second perturbation (fixed) vector as the FBP is the following: Superiorization was demonstrated in the literature [5] to provide superior results when the starting point, x^0, is further away from the sought after feasible solutions set defined by the constraints. The reason is that starting from a point that is further away from it allows for more perturbations to be interlaced between the iterative steps and, as a result, a better value for the merit function can be achieved when the stopping criterion is met. On the other hand, starting the process from a point that is closer to the feasible solutions set defined by the constraints will reach a stopping point faster, however the value of the merit function will not be very close to a desired value. The proposed second perturbation (within an iteration step) is therefore suggested to overcome the need to balance these contradicting aims between speed and quality of the reconstruction, allowing the process to start with a point further away while not penalizing the number of iterations.

For the second perturbation vector, we have chosen the (positive) FBP vector for all k. For the sequence of the step sizes (denoted here as β'_k) we chose the same sequence of β_k from the first kind of perturbation step, multiplied by a small positive real number. The sequence of the β'_k is summable since the β_k are summable.

The stopping criterion for the algorithm was chosen in our implementation as in [5]; it makes use of a user-specified ε that sets a threshold on the residual between the system matrix and the measurements, given the current image. Since our simulated data are noisy, the phantom (i.e., the object that we need to reconstruct) has a residual as well. From the point of view of what we are trying to do (estimating the RSP values of the phantom), there is no need to find a solution that has a smaller residual than the phantom itself. In the Appendix we provide the pseudo-code of the algorithm described above and in the next section we illustrate the usefulness of our method on pCT-simulated data.

RESULTS

We have simulated data from 180 projections over 360 degrees of the Herman Head phantom (shown in Fig. 1(a)) using the simulation tool Geant 4. For the DROP algorithm, we have set λ_k to 0.04 and the number of blocks to 180, associating all the histories that belong to a specific projection angle with a block. We set the stopping criterion to 736 since this was the residual of the phantom with the dataset. We compared our newly proposed method with two other reconstructions: ones that use only one perturbation scheme that aims at superiorizing TV, where the only difference between the two are the starting points x^0, with the first being the FBP point as was done in [2] and the second being the ZERO point (i.e., a vector with all elements zero) [5]. We first examine the reconstruction that started with FBP. As pointed out in the previous section, since the residual of the FBP point (737) is very close to the residual of the phantom (736) it required only one iteration to get below it. While getting to the stopping criterion faster is generally a good thing, as indicated previously, the superiorization algorithm does not have enough iterations to steer the process towards a superior point. This can be seen when the TV values are compared. The TV of the phantom is 1287 while the one of the FBP is 2441 and the reconstructed image obtained after one iteration was 1944. If we allow more iterations that satisfy the stopping criterion (with a lower residual), then after 8 iterations the algorithm that started with the FBP point reduces the TV to 1479 (the lowest TV value for this reconstruction), which is not as low as the TV of the phantom; the image presented in Figure 1(b) displays this reconstruction. In Figure 1(c) we show the reconstruction of the algorithm when the starting point is the ZERO point. The TV of the reconstruction is now lower than the TV of the phantom with 1262. The number of iterations, however, to reach the stopping criterion was 10.

Figure 1(d) shows the image obtained with our newly proposed scheme, when the starting point is the ZERO point, the first kind of perturbations is aimed at superiorizing TV (similar to the other two reconstructions) but here we also use the FBP point for the second kind of perturbations, with β'_k proportional to β_k with a multiplier of 0.25 (i.e., $\beta'_k = \beta_k/4$). The TV of the reconstructed image is 1278 (lower than the phantom) but took only 4 iterations to reach a residual set by the stopping criterion. In Figure 2 we present the relative stopping power distributions of the phantom and the three reconstructions.

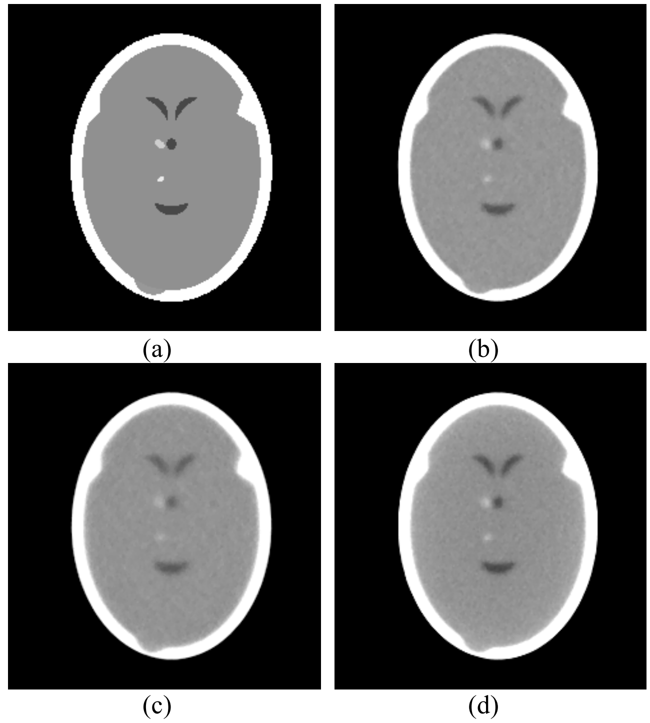

Figure 1: The Herman Head phantom, TV=1287 is shown in (a) and superiorization reconstructions aiming at superior TV with (b) FBP as the starting point (8 iterations, TV =1479) and (c) ZERO as the starting point (10 iterations, TV =1262). (d) Superiorization reconstruction with dual perturbations scheme (4 iterations, TV =1278) starting from the ZERO point.

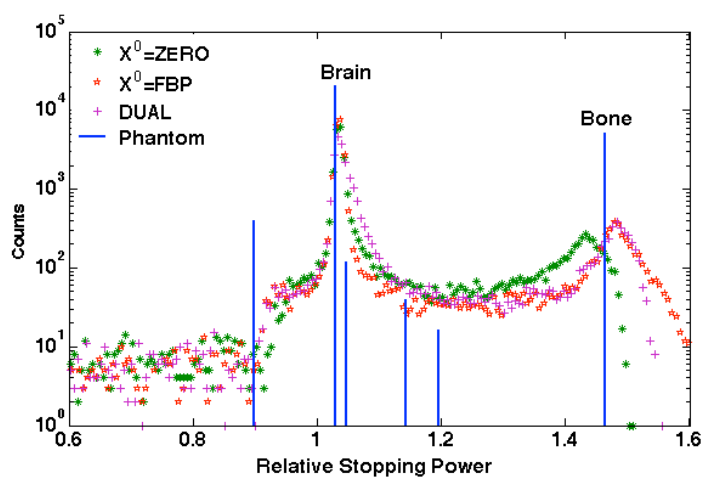

Figure 2: Relative stopping power distribution of the phantom and the three reconstructions. ZERO and FBP are the superiorization reconstructions that started from the zero and FBP points, and DUAL is the newly proposed superiorization method that uses the zero point as the starting point and FBP as a second perturbation vector.

CONCLUSIONS

We presented a new superiorization scheme that uses two kinds of perturbations with the first aiming at reducing a merit function and the second to steer the iterates faster towards a solution by controlling the balance between the activities of feasibility-seeking and merit function reduction. We illustrated the usefulness of the newly proposed method over previous published results and showed their applicability to pCT with realistic simulated data reconstructing the Herman Head phantom.

APPENDIX: PSEUDOCODE OF THE SUPERIORIZATION ALGORITHM USING A DUAL PERTURBATIONS SCHEME

```
set $k = 0$
set $x^k = ZERO$
set flag = true
while (flag)
    set logic = true
    while (logic)
        set $z = x^k - \beta_k v^k$
        if $\left(TV(z) \leq TV\left(x^k\right)\right)$
            then
                set $y = \mathbf{P}z$
                if $(residual(y) < \varepsilon)$
                    then
                        set $x^{k+1} = y$
                        set flag = false
                    else
                        set $x^{k+1} = y + \beta'_k FBP$
                set logic = false
            else set $\beta_k = \beta_k/2$
        set $k = k + 1$
```

ACKNOWLEDGEMENT

This work is supported by Grant Number 2009012 from the United States – Israel Binational Science Foundation (BSF) and by the US Department of Army award number W81XWH-10-1-0170.

REFERENCES

1. G. T. HERMAN, "Fundamentals of Computerized Tomography: Image Reconstruction from Projections," Springer, NY, 2nd Edition (2009)
2. S. N. PENFOLD, R. W. SCHULTE, Y. CENSOR, A. B. ROSENFELD, "Total variation superiorization schemes in proton computed tomography image reconstruction," Med. Phys. **37**, 5887–5895 (2010)
3. S. N. PENFOLD, "Image Reconstruction and Monte Carlo Simulations in the Development of Proton Computed Tomography for Applications in Proton Radiation Therapy," Ph.D. Thesis, University of Wollongong (2010)
4. D. BUTNARIU, R. DAVIDI, G. T. HERMAN, I.G. KAZANTSEV, "Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems," IEEE J. Selected Topics Signal Process. **1**, 540–547 (2007)
5. Y. CENSOR, R. DAVIDI, G. T. HERMAN, "Perturbation resilience and superiorization of iterative algorithms," Inverse Probl. **26**, 065008 (2010)
6. Y. CENSOR, T. ELFVING, G. T. HERMAN, and T. NIKAZAD, "On diagonally-relaxed orthogonal projection methods," *SIAM Journal of Scientific Computing*, **30**, 473–504 (2008)