Application of incremental algorithms to CT image reconstruction for sparse-view, noisy data

Rose, Sean; Andersen, Martin Skovgaard; Sidky, Emil Y.; Pan, Xiaochuan

Published in:
Proceedings of the 3rd International Conference on Image Formation in X-Ray Computed Tomography

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Application of incremental algorithms to CT image reconstruction for sparse-view, noisy data

Sean Rose[1], Martin S. Andersen[2], Emil Y. Sidky[1], and Xiaochuan Pan[1]

***Abstract*—This conference contribution adapts an incremental framework for solving optimization problems of interest for sparse-view CT. From the incremental framework two algorithms are derived: one that combines a damped form of the algebraic reconstruction technique (ART) with a total-variation (TV) projection, and one that employs a modified damped ART, accounting for a weighted-quadratic data fidelity term, combined with TV projection. The algorithms are demonstrated on simulated, noisy, sparse-view CT data.**

I. Introduction

In iterative image reconstruction (IIR) there can be a large disconnect between practical iterative algorithms and the optimization problems that motivate their design. Particularly for image reconstruction from sparse-view CT data with its associated ill-conditioned linear system model, the number of required iterations for accurate solvers of relevant optimization problems can be much greater than 1,000. When iteration numbers are this large, the trajectory of the image estimates can be quite important because practical application of IIR dictates iteration numbers on the order of ten – well short of convergence.

The usual strategy for obtaining useful images rapidly is to employ algorithms that process the data sequentially [1,2]. In particular, for sparse-view CT we have been developing the adaptive-steepest-descent - projection-onto-convex-sets (ASD-POCS) algorithm [3], which is sequential in that it employs ART for the data agreement step. The algorithm has been shown to yield useful images at low iteration numbers [4]. While we have also used ASD-POCS for accurate solution of constrained TV-minimization, the algorithm is not guaranteed to converge and parameter selection is not straight-forward when accurate solution is desired.

Recently, an incremental framework [5,6] has been developed from which sequential iterative algorithms can be derived that both yield useful images at low iteration numbers and converge to the solution of a designed optimization problem. The reason why such a framework can be helpful for IIR algorithm development is that many design principles such as maximum entropy, maximum likelihood (ML), and sparsity exploitation are a form of optimization. It is not clear that truncating the iteration of the optimization problem solver will yield images that reflect the intentions of the designed optimization problem. With the incremental framework, where initial convergence is rapid, there may be a stronger link between early image estimates and the solution to the designed optimization.

In this work, we motivate and investigate the use of TV-constrained data-discrepancy minimization for sparse-view image reconstruction from noisy CT data. Two different data agreement terms are compared: a Euclidean distance between estimated data and input data, and a weighted quadratic where the weighting reflects the model used for generating the simulated noise. Accurate solution of the designed optimization, solved using the Chambolle-Pock (CP) algorithm, is compared to image estimates obtained at low iteration numbers for algorithms derived from the incremental framework.

II. Methods

Using a generic linear model for X-ray projection

$$\mathbf{g} = X\mathbf{f}, \tag{1}$$

ideal gradient magnitude image (GMI) sparsity exploiting image reconstruction is formulated as

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \|\nabla \mathbf{f}\|_1 \text{ such that } X\mathbf{f} = \mathbf{g}, \tag{2}$$

where X is the system matrix representing X-ray projection; $\mathbf{f}$ is the image, $\mathbf{g}$ is the sinogram; ∇ is a finite differencing implementation of the spatial gradient of the image; and the objective function $\|\nabla \mathbf{f}\|_1$ is the TV of the image $\mathbf{f}$. The specified optimization problem seeks, among all images that agree with the data perfectly, the one with minimum TV. When noise or other inconsistency is present in the data, the strict equality of Eq. (2) cannot be satisfied and this constraint must be relaxed

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \|\nabla \mathbf{f}\|_1 \text{ such that } \|X\mathbf{f} - \mathbf{g}\|_2 \leq \epsilon, \tag{3}$$

where ϵ is a parameter of the optimization that puts a tolerance on the allowable data discrepancy. While this

[1]The University of Chicago, Department of Radiology MC-2026, 5841 S. Maryland Avenue, Chicago IL, 60637.

[2]Technical University of Denmark, Department of Applied Mathematics and Computer Science, Lyngby, Denmark

optimization problem can be used for GMI sparsity-exploiting image reconstruction, it is somewhat inconvenient in that the parameter ϵ must be searched.

For this work, we formulate the GMI sparsity exploiting optimization with a constraint on the TV instead of the data discrepancy

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \frac{1}{2}\|X\mathbf{f} - \mathbf{g}\|_2^2 \text{ such that } \|\nabla \mathbf{f}\|_1 \leq \gamma, \quad (4)$$

where now the parameter γ constrains the image TV. This form is particularly convenient for phantom studies to test the effectiveness of GMI sparsity-exploiting image reconstruction, because γ can be set to the value derived from the test phantom. In this way, both ideal and noisy data studies can be conducted using Eq. (4). In the ideal case the data fidelity term can be driven to zero, while in the noisy case the objective minimum will likely be nonzero and positive. Moreover, employing a TV constraint allows easy comparison of GMI sparsity-exploiting image reconstruction with different data fidelity terms. As an example of such an alternative, which we investigate below, we employ the maximum-likelihood data fidelity for uncorrelated Gaussian noise

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \frac{1}{2}(X\mathbf{f} - \mathbf{g})^T \text{diag}(\mathbf{v})^{-1}(X\mathbf{f} - \mathbf{g}) \text{ such that } \|\nabla \mathbf{f}\|_1 \leq \gamma, \quad (5)$$

where $\mathbf{v}$ represents the variance of the data noise model.

All of the optimizations stated above can be solved by a first-order algorithm such as that of Chambolle and Pock [7,8], but the required iteration number for a useful image can be large. One can also employ the incremental framework to derive sequential algorithms that solve both Eqs. (4) and (5). Following Ref. [6], we write down in Algorithm 1 an instance of an incremental algorithm for the optimization problem in Eq. (5).

III. Results

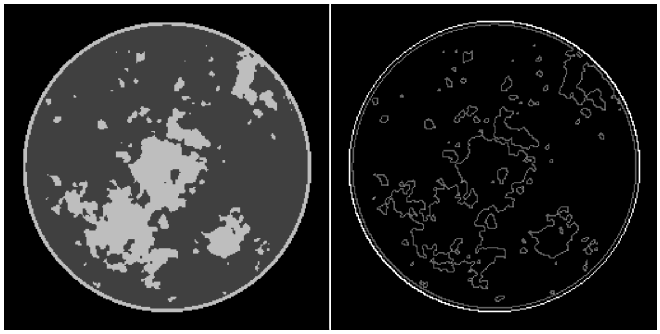


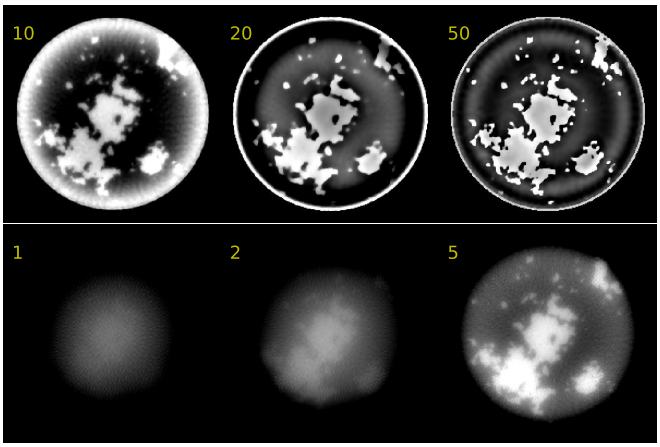
Fig. 1. Breast phantom for CT and its corresponding gradient-magnitude image (GMI). Left is the linear attenuation map of the phantom in the gray scale window [0.174,0.253] cm^{-1}. Right is the GMI [0.0,0.1] cm^{-1} that serves to illustrate that the test phantom is sparse in the GMI.

Algorithm 1 Pseudocode for N steps of an incremental algorithm instance for solving the TV-constrained optimization problem in Eq. (5). Note that this algorithm also applies to Eq. (4) by setting $\mathbf{v} = 1$. The integer M is the total number of measurements in the sinogram. The parameters t_0, ρ and α all affect rate of convergence, but for the simulations here where iteration numbers are low, ρ and α are fixed to one and zero, respectively. The parameter t_0 is determined by the value that yields the smallest data discrepancy within a fixed number of iterations. The parameter γ belongs to the optimization problem, and for the simulations presented here it is always set to the TV of the test phantom. The projection in line 14, which finds the image closest to $\mathbf{h}_M$ with TV less than or equal to γ, is carried out by the CP algorithm [7,8].

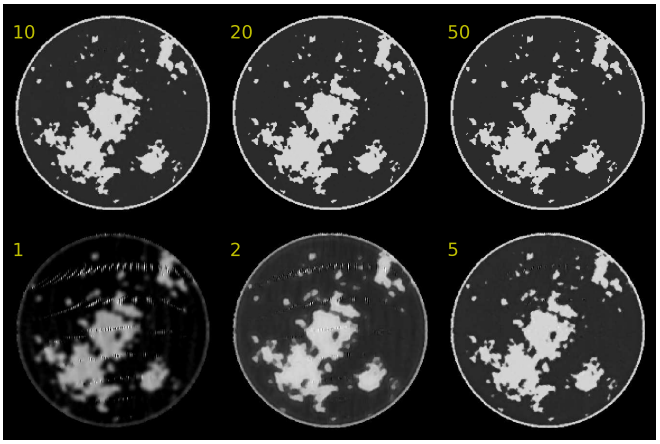
1: select algorithm parameters:
2: $t_0 \in (0, \infty);\ \rho \in (0, 2);\ \alpha \in [0, 1]$
3: select TV constraint parameter γ
4: initialize $\mathbf{f}_0$
5: $n \leftarrow 0$
6: **repeat**
7: $\quad \mathbf{h}_0 \leftarrow \mathbf{f}_n$
8: $\quad t_n \leftarrow t_0/(n+1)^\alpha$
9: $\quad i \leftarrow 0$
10: $\quad$ **repeat**
11: $\quad\quad \mathbf{h}_{i+1} \leftarrow \mathbf{h}_i - \rho \mathbf{x}_i \frac{(\mathbf{x}_i^T \mathbf{h}_i - g_i)}{\|\mathbf{x}_i\|_2^2 + v_i/t_n}$
12: $\quad\quad i \leftarrow i + 1$
13: $\quad$ **until** $i \geq M$
14: $\quad \mathbf{f}_{n+1} \leftarrow \rho\, \text{proj}_{\{\mathbf{f} | \|\nabla \mathbf{f}\|_1 \leq \gamma\}}(\mathbf{h}_M) + (1 - \rho)\mathbf{h}_M$
15: $\quad n \leftarrow n + 1$
16: **until** $n \geq N$

To demonstrate the utility of the Algorithm 1, we conduct image reconstruction studies with ideal and noise sparse-view projection data. The digital phantom shown in Fig. 1 emulates breast CT and it consists of 256×256 pixel array. For the ideal data study, the projection data are obtained by use of Eq. (1) so that the system matrix employed in data generation and image reconstruction are identical. The projection data consist of 100 projections onto a 512-bin linear detector array. The source-detector and source-isocenter distances are modeled to be 72cm and 36cm, respectively. The sampling of this configuration is clearly not sufficient for direct or implicit inversion of Eq. (1) because the number of image pixels exceeds the number of measurements.

a) Noiseless study: Exploiting GMI sparsity with TV-constrained optimization, we solve Eq. (4) setting $\gamma = \gamma_0$, the true value obtained from the test phantom. We state without showing results that the true phantom can be recovered exactly in the numerical sense under these



Chambolle-Pock



incremental

Fig. 2. Progression of image estimates for both CP and incremental algorithms for the case of noiseless projection data. The iteration number is indicated in each panel.

conditions. The more important point for our purpose is how quickly can a useful image be obtained. We solve Eq. (4) by use of both CP and the incremental instance shown in Algorithm 1, setting $\mathbf{v} = 1$. Intermediate iterates are shown up to 50 iterations in Fig. 2. One can immediately see the advantage of Algorithm 1 as a visually accurate reconstruction appears already at 10 iterations while the CP results are not close to the phantom for any of the shown images. This difference is not specific to these two algorithms. Rather it stems from a well-known feature of algorithms that process the data sequentially versus those that do not. The point of interest here is that we have the rapid initial convergence with Algorithm 1, and as the iterations continue the image estimate is guaranteed to converge to a solution of Eq. (4).

b) Noisy study: For the remainder of the results, image reconstruction is performed on simulated data including noise. The same discrete-to-discrete model generates the mean sinogram, but noise realizations are drawn from a Gaussian distribution, where the covariance is taken to be diagonal and the variance at each sample is the one over the transmitted number of photons. The integrated incident number of photons per view per detector bin is modeled to be $2{\times}10^5$, corresponding to a fairly low intensity that might be used in an actual breast CT scan. In this simulation the only source of inconsistency is due to the noise model, and we know exactly what probability distribution function governs the noise realization selection. All other sources of inconsistency: continuous object model, beam hardening, scatter, partial volume averaging, etc., which would be present in actual CT data are suppressed. In this way, we can isolate and address two questions: (1) within the parameters of the simulation what is the impact of using the ML motivated weighted quadratic in Eq. (5) as opposed to the isotropic quadratic of Eq. (4) in terms of the solution to the respective optimization problems, and (2) if there is difference between solutions of these problems, will this difference be reflected in the images generated by Algorithm 1 when the iteration is severely truncated, i.e. 10 to 20 iterations.

isotropic: $\mathbf{v} = 1$ weighted: $\mathbf{v} = 1/N_{ph}$

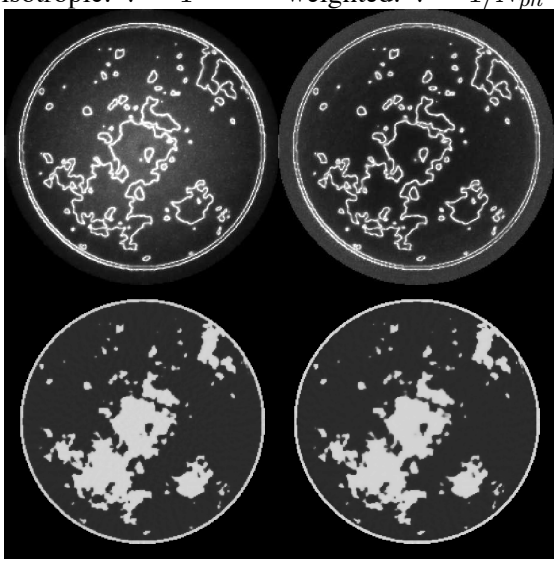


Fig. 3. Image standard deviation (top row, [0.0,0.005] cm^{-1}) and mean (bottom row, [0.174,0.253] cm^{-1}) estimated from accurate solution of Eq. (5) for 1000 realizations of noisy data. The left and right columns correspond respectively to an isotropic data fidelity and a one-over-transmission-intensity ($1/N_{ph}$) weighted quadratic motivated by the ML principle. The mean image RMSE from the truth is $2.19{\times}10^{-3}$ cm^{-1}, left, and $2.10{\times}10^{-3}$ cm^{-1}, right.

Employing the CP algorithm, the mean and standard deviation of the solution to Eq. (5) using $\gamma = \gamma_0$, the true TV of the phantom, are estimated from 1000 noise realizations and are shown in Fig. 3. The bias for both isotropic and ML weighting is low as both image means are visually close to the phantom. The estimated standard deviation images are more interesting, showing structure that reflects the object. In both cases the image standard deviation is reduced dramatically by the use of the TV constraint except for at pixels near to those corresponding to nonzero values of the phantom's GMI.

This phenomenon was also observed in the fully sampled case in Ref. [9]. We do observe a difference in the background of these images and it does appear that the ML weighting yields a lower standard deviation in the middle of the image. Going toward the image periphery, the isotropic weighting appears to result in a slightly lower standard deviation. Overall, the ML weighting lowers the image standard deviation with respect to isotropic weighting without increasing bias.

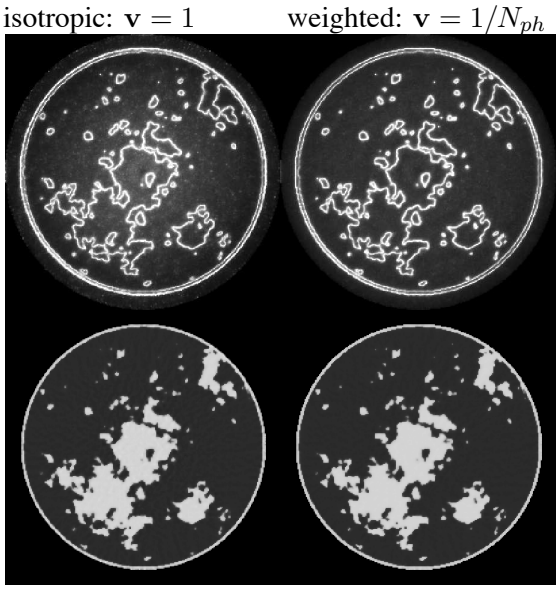


Fig. 4. Same as Fig. 3 except that results are obtained after 20 iterations of Algorithm 1. For both isotropic and ML weighting, the algorithm t_0 is selected to yield the lowest value of the data fidelity at the last iteration. The mean image RMSE from the truth is 3.14×10^{-3} cm^{-1}, left, and 2.21×10^{-3} cm^{-1}, right.

Turning to the use of Algorithm 1, we have verified (results not shown) using a single noise realization that we obtain an accurate solution to Eq. (5) in 1000 iterations. The interest here, however, is in use of Algorithm 1 at low iteration number. A parallel set of results are shown in Fig. 4 obtained by use of Algorithm 1 and stopping at 20 iterations. Again the TV constraint is set to the phantom TV, but because the iteration is truncated an additional parameter t_0 has a large effect on the reconstructed image. For this preliminary work we have selected this parameter as described in the figure caption, but there is a bias-variance trade-off associated with t_0, which would need to be fully explored for a more complete understanding. Nevertheless we do observe an effect of the different weightings even though the iteration is severely truncated. The ML weighting for the given t_0 settings yields a visually lower standard deviation background and a lower RMSE between mean and phantom.

IV. Summary

We have applied the incremental framework of Refs. [5,6] to generate an algorithm instance of TV constrained, data discrepancy minimization for CT image reconstruction. The algorithm yields an accurate solution to the designed optimization problem at large iteration numbers and can provide a useful image at low iteration numbers. To demonstrate the utility of this framework, we apply the algorithm to two different TV constrained optimization problems with different data discrepancy terms. We observe with preliminary results from a controlled simulation that at low iteration numbers we may be able to (1) accurately recover the image from under-sampled data by use of the TV constraint and (2) obtain a more favorable variance-bias trade-off by use of the weighted quadratic term.

V. Acknowledgment

MSK was supported by Grant No. ERC-2011-ADG 20110209 from the European Research Council. This work was also supported in part by NIH R01 Grant Nos. CA158446, CA120540, and EB000225. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theor. Biol.*, vol. 29, pp. 471–481, 1970.

[2] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.*, vol. 51, no. 15, pp. R541–R578, 2006.

[3] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008.

[4] E. Y. Sidky, X. Pan, I. Reiser, R. M. Nishikawa, R. H. Moore, and D. B. Kopans, "Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms," *Med. Phys.*, vol. 36, pp. 4920–4932, 2009.

[5] D. P. Bertsekas, "Incremental proximal methods for large scale convex optimization," *Math. Program.*, vol. 129, pp. 163–195, 2011.

[6] M. S. Andersen and P. C. Hansen, "Generalized row-action methods for tomographic imaging," *Numer. Alg.*, pp. 1–24, 2013, http://dx.doi.org/10.1007/s11075-013-9778-8.

[7] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, pp. 120–145, 2011.

[8] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm," *Phys. Med. Biol.*, vol. 57, pp. 3065–3091, 2012.

[9] T. Köhler and R. Proksa, "Noise properties of maximum likelihood reconstruction with edge-preserving regularization in transmission tomography," in *Proceedings of the 10th International Meeting on Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2011, pp. 263–266.