Nonconvex optimization for improved exploitation of gradient sparsity in CT image reconstruction

Emil Y. Sidky[1], Rick Chartrand[2], Jakob S. Jørgensen[3], and Xiaochuan Pan[1]

***Abstract*—A nonconvex optimization algorithm is developed, which exploits gradient magnitude image (GMI) sparsity for reduction in the projection view angle sampling rate. The algorithm shows greater potential for exploiting GMI sparsity than can be obtained by convex total variation (TV) based optimization. The nonconvex algorithm is demonstrated in simulation with ideal, noiseless data for a 2D fan-beam computed tomography (CT) configuration, and with noisy data for a 3D circular cone-beam CT configuration.**

I. Introduction

Much recent work in iterative image reconstruction in computed tomography (CT) has focused on some form of total variation (TV) minimization, and one of the motivations for employing TV minimization is exploiting sparsity in the gradient magnitude image (GMI) to reduce sampling requirements for the CT system. TV-minimization has been demonstrated, in simulations and with real scanner data, to be effective at allowing for projection view sampling reduction. There is, however, potential to take the sparsity-exploiting principle further, because TV-minimization is an ℓ_1-based convex relaxation of an ideal, nonconvex, sparsity-exploiting optimization based on the ℓ_0-norm. To approach more closely the ℓ_0-based minimization, we develop a GMI sparsity-exploiting algorithm for CT based on an ℓ_p-norm where $p \in (0,1)$. Section II summarizes the theory and algorithm, and Sec. III shows results based on 2D and 3D CT simulations.

II. Constrained, nonconvex optimization by reweighting

We briefly state the rationale and methods for GMI exploiting CT image reconstruction with nonconvex optimization. We write the CT data model generically as a linear system

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

where $\mathbf{f}$ is the image vector comprised of voxel coefficients, X is the system matrix generated by some approximation to projection of the voxels, and $\mathbf{g}$ is the data vector containing the estimated projection samples. The model can be applied equally to 2D and 3D geometries, and we note that there are many specific forms to this linear system depending on sampling, image expansion elements, and approximation of continuous fan- or cone-beam projection.

[1]The University of Chicago, Department of Radiology MC-2026, 5841 S. Maryland Avenue, Chicago, IL 60637. Corresponding author: Emil Y. Sidky, E-mail: sidky@uchicago.edu. [2]Theoretical Division, T-5, MS B284, Los Alamos National Laboratory, Los Alamos, NM 87545. [3]Technical University of Denmark, Department of Applied Mathematics and Computer Science, Matematiktorvet, bygning 303B, 2800 Kgs. Lyngby, Denmark.

For the present work, we focus on CT configurations with sparse angular sampling, where the sampling rate is too low for Eq. (1) to have a unique solution. In this situation, there has been much interest in exploiting GMI sparsity of the object to narrow the solution space of Eq. (1) and potentially obtain an accurate reconstruction from under-sampled data. The formulation of this idea results in a nonconvex constrained optimization:

$$\mathbf{f}^\circ = \arg\min_{\mathbf{f}} \left\| \sqrt{(\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right\|_0 \quad \text{such that } \mathbf{g}_{\text{data}} = X\mathbf{f}, \tag{2}$$

where the argument of the ℓ_0-norm is the voxel-wise magnitude of the image spatial gradient; the linear operators ∂_x, ∂_y, and ∂_z are matrices representing finite differencing in their respective labeled directions; the numerical gradient of the image is formed by, $\nabla \mathbf{f} = [\partial_x \mathbf{f}, \partial_y \mathbf{f}, \partial_z \mathbf{f}]^T$ (2D is obtained by deleting the third component); the ℓ_0-norm counts the number of non-zero components in the argument vector; and $\mathbf{g}_{\text{data}}$ is the available projection data. In words, this optimization seeks the image $\mathbf{f}$ with the lowest GMI sparsity while agreeing exactly with the data.

The optimization problem in Eq. (2) does not lead directly to a practical image reconstruction algorithm, because, as of yet, no large scale solver is available for this problem. Also, the equality constraint, requiring perfect agreement between the available and estimated data, makes no allowance for data inconsistency. In working toward developing a practical image reconstruction algorithm different relaxations of Eq. (2) have been considered. One such relaxation is

$$\mathbf{f}^\circ = \arg\min_{\mathbf{f}} \left\| \sqrt{(\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right\|_p^p \quad \text{such that } \|\mathbf{g}_{\text{data}} - X\mathbf{f}\|_2 \le \epsilon, \tag{3}$$

where the ℓ_0-norm is replaced by the ℓ_p-norm,

$$\|\mathbf{v}\|_p^p \equiv \sum_i |v_i|^p,$$

and the data equality constraint is relaxed to an inequality constraint with data-error tolerance parameter ϵ. An important strategy, which has been studied extensively in Compressive Sensing [1], is to set $p = 1$, which corresponds to TV-minimization. This, on the one hand, maintains some of the sparsity seeking features of Eq. (2) and, on the other hand, leads to a convex problem, which has convenient features for algorithm development. For example, a local minimizer is a global minimizer in convex optimization.

Another interesting option for GMI sparsity-exploiting image reconstruction is to consider Eq. (3) for $0 < p < 1$. Such a choice for p leads to nonconvex optimization, which can allow for greater sampling reduction than the $p = 1$ case while maintaining highly accurate image reconstruction. These gains intuitively stem from the fact $p < 1$ is closer to the ideal sparsity-exploiting case of $p = 0$; the catch, however, is on the algorithmic side where one has to deal with potential local minima, which are not part of the global solution set. Despite this potential difficulty, practical algorithms based on this nonconvex principle are available [2,3], and gains in sampling reduction for various imaging systems have been reported for both simulated and real data cases. For X-ray tomography, use of this nonconvex strategy has shown promising results [4,5], but the algorithms proposed in those works for CT are only motivated by the optimization problem in Eq. (3) and are not accurate solvers of this problem. An accurate solver is important for theoretical studies of CT image reconstruction with under-sampled data and may also aid in developing algorithms for limited-data tomographic devices.

For CT, one of the barriers to developing an efficient and accurate solver for Eq. (3) in the nonconvex $p < 1$ case, is that it is already challenging to develop such a solver for the convex $p = 1$ case. In order to handle the latter convex, but non-smooth case, we have been interested in an alternate line of optimization problems, where the salient image metrics are written as constraints instead of in an objective function. It is a strategy similar to the set theoretic approach presented in Ref. [6]; the algebraic reconstruction technique (ART) is a specific realization of this strategy; and this type of approach can be useful for nonconvex constraint sets [7]. For the alternate, constraint-based optimization problem there are efficient, large-scale solvers recently available [8,9].

Returning to GMI sparsity-exploiting image reconstruction, we employ an approach developed in Ref. [9] and alter Eq. (3) to the following

$$\mathbf{f}^\circ = \arg\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{f} - \mathbf{f}_{\text{prior}}\|_2^2 \text{ such that } \|\mathbf{g}_{\text{data}} - X\mathbf{f}\|_2 \leq \epsilon \text{ and } \left\| \sqrt{(\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right\|_p^p \leq \gamma, \quad (4)$$

which seeks the image $\mathbf{f}$ closest to a prior image $\mathbf{f}_{\text{prior}}$ while respecting constraints on the ℓ_p-norm of the GMI and data-error tolerance. We do not consider, here, the availability of a prior image and set $\mathbf{f}_{\text{prior}} = 0$, keeping this vector only for generality. Consider, first, the case of $p = 1$; the constraint on the GMI becomes a constraint directly on the image TV. Constrained minimization of image TV is known to encourage GMI sparsity. We do not directly minimize TV, rather we independently select parameters γ and ϵ. For sparsity-exploiting image reconstruction, both of these parameters are chosen to have small values: small ϵ forces tight agreement with the data, and small γ encourages GMI sparsity. We note that $\epsilon = 0$ corresponds to a data equality constraint, which may allow no solutions when inconsistencies are present in the data. For $p = 1$, the optimization problem in Eq. (4) is convex and the algorithm presented in Ref. [9] can be used directly to obtain the solution.

For this abstract, we are interested in developing an algorithm for $0 < p < 1$, where the GMI constraint becomes nonconvex. The issue then becomes how to solve Eq. (4) for $p < 1$, because the algorithm in Ref. [9] applies only to convex problems. The approach taken involves approximating Eq. (4) with a convex problem employing a weighted ℓ_1-norm:

$$\mathbf{f}^\circ = \arg\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{f} - \mathbf{f}_{\text{prior}}\|_2^2 \text{ such that } \|\mathbf{g}_{\text{data}} - X\mathbf{f}\|_2 \leq \epsilon \text{ and } \left\| \mathbf{w}\sqrt{(\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right\|_1 \leq \gamma, \quad (5)$$

where the GMI constraint involves only the ℓ_1-norm and a non-negative weight vector $\mathbf{w}$. For a given $\mathbf{w}$ this optimization problem is convex and can be solved efficiently using the algorithm in Ref. [9]. To attack the nonconvex problem, we employ a reweighting technique, where there are two loops: an inner loop where Eq. (5) is solved given parameters γ, ϵ, and weight vector $\mathbf{w}$, and an outer loop where the weight vector is adjusted based on the solution of the inner loop:

$$\mathbf{w} = \left(\sqrt{\eta + (\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right)^{p-1}.$$

The parameter η is needed to prevent the singularity at voxels with zero GMI when $p < 1$. For all simulations in this abstract $\eta = 10^{-6}$. With a reweighting approach, an important question is how accurately does the intermediate weighted problem need to be solved in the inner loop so that overall convergence of the outer loop is attained. It turns out for the present reweighting scheme it suffices to have only one inner iteration. Thus, the complete algorithm is derived from the algorithm in Ref. [9], and the weights are recomputed at every iteration based on the current image estimate $\mathbf{f}$.

III. Results

To demonstrate the new image reconstruction algorithm, we perform two sets of experiments. In the first, we employ the algorithm on ideal, noiseless fan-beam CT data where it is possible to recover the exact image. With this ideal simulation, we demonstrate the potential for angular sampling reduction. In the second simulation, we apply the algorithm to circular, cone-beam CT projections with noise. The purpose of the latter simulation is to demonstrate that the algorithm can indeed by applied to 3D CT, and to illustrate the impact of the nonconvex algorithm on data inconsistency.

A. Ideal fan-beam CT simulation

For the 2D simulation we employ the breast phantom shown in Fig. 1. In the figure, the phantom GMI is also shown, which is seen to have many more zeros than the original phantom. It is this sparsity in the GMI, which we seek to exploit in order to reduce angular sampling. The phantom is discretized on a 128×128 pixel array, which is 20 cm on a side. Only the pixels within the largest inscribed circle are allowed to vary, and pixels outside this 20 cm diameter circle are fixed to zero. The fan-beam CT simulation models an X-ray source 40 cm from the isocenter, and a 80 cm source-to-detector distance.

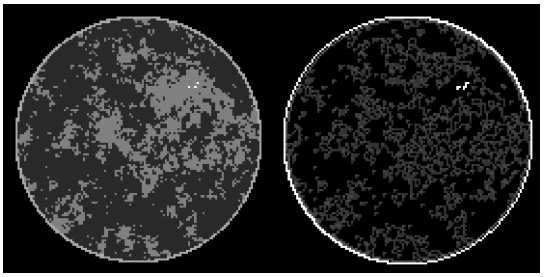
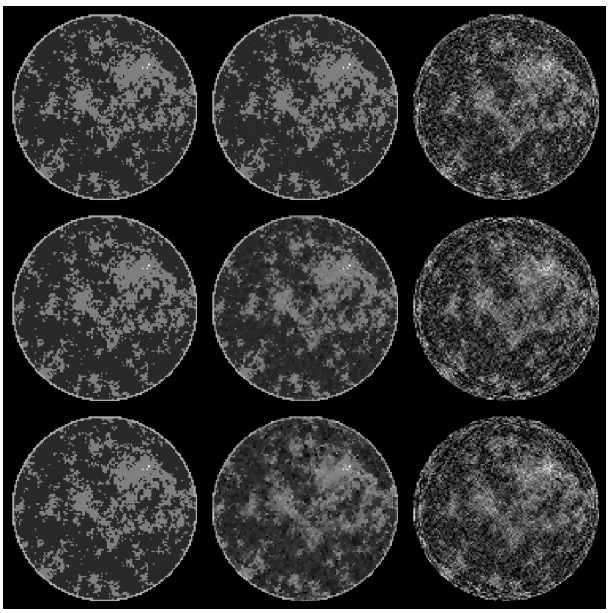


Fig. 1. Left: computerized breast phantom shown in a gray scale window [0.95, 1.25]. Right: gradient magnitude image (GMI), which has greater sparsity than the original phantom.



$p = 0.8$ $p = 1.0$ $p = 2.0$

Fig. 2. Reconstructed images for nonconvex $p = 0.8$, left column, compared with convex $p = 1.0$, middle column, and $p = 2.0$, right column. The number of views covering 360 degrees is 35, 30, and 25 for the top, middle, and bottom rows, respectively. The gray scale window is [0.95, 1.25].

The detector consists of 256 bins in a linear configuration, which is long enough to capture the projection of the 20 cm diameter pixel array. We consider only 360 degree scans, but allow the number of projections to vary.

To illustrate the potential of nonconvex optimization for sparsity-exploiting image reconstruction, we compare solutions of Eq. (4) for $p = 0.8$, $p = 1.0$, and $p = 2.0$. The latter two values lead to a convex problem, which can be solved with the algorithm in Ref. [9], and the first value leads to a nonconvex problem solved by the proposed reweighting algorithm using Eq. (5). For values $p = 1.0$ and $p = 2.0$, we have a direct convergence check, but for the nonconvex case we cannot claim to have found a global solution to Eq. (4). Instead, we can verify that Eq. (5) is solved for the weights $\mathbf{w}$ that are settled upon.

In applying the constraint-based optimization problem in Eq. (4), we need to specify two parameters ϵ and γ. The data used in this simulation are ideal, and accordingly we employ a tight data-error constraint and use a value for ϵ corresponding to an root-mean-square-error (RMSE) of 10^{-5}. For the image TV constraint we set γ to the value of the

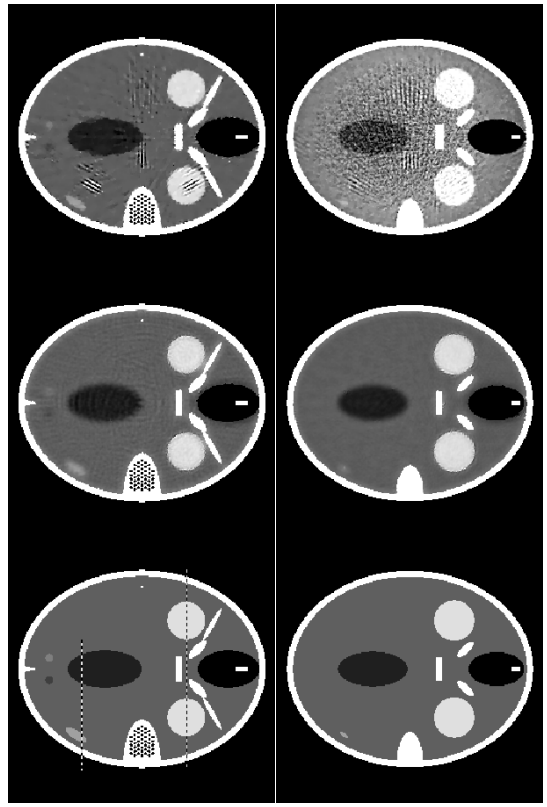


Fig. 3. Images reconstructed from noisy projections of the FORBILD head phantom. The rows show the results for $p = 1.0$, top, $p = 0.8$, middle, and the phantom, bottom in a gray scale window of [1.0425, 1.0625]. The first column shows the midplane, and the second column shows a transaxial plane near the top of the bony structure at the ear. The dashed lines in the phantom midplane slice indicate the locations of the profiles for Figs. 4 and 5.

ℓ_p-norm of the actual phantom GMI to the pth power. We note that in actual application, access to this information is unavailable and selection of γ would need to be based on different image quality metrics. Here, however, we are exploring the theoretical potential of the proposed algorithm.

Shown in Fig. 2, are image reconstruction results for 25, 30, and 35 simulated projections. The $p = 1.0$ case has some potential to reduce angular sampling by exploiting GMI sparsity. This is evident in the comparison with $p = 2.0$, which does not exploit GMI sparsity; the $p = 1.0$ results show visually accurate reconstruction for 35-view projection data while the $p = 2.0$ results do not show accurate reconstruction for any of the projection data sets. The nonconvex $p = 0.8$ results, however, extend the visually accurate reconstruction range down to 25-view projection data.

B. Circular cone-beam CT simulation with noisy projections

For the 3D circular cone-beam CT simulation, we scale up the problem approaching the scale of a realistic volume CT system, and we include noise on the CT projections. The phantom used for this simulation is the FORBILD head phantom, which has many low contrast objects, with gray level variations ranging from 0.25% to 1% of the phantom background, together with complex high-contrast bony structures. This phantom is quite challenging, because even minor streaks from the bony structures can interfere strongly with imaging the low-contrast objects.

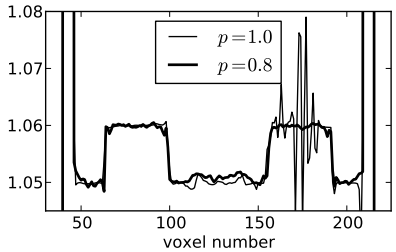


Fig. 4. Profile comparison corresponding to the images in Fig. 3 along a line in the midplane, through the eyes.

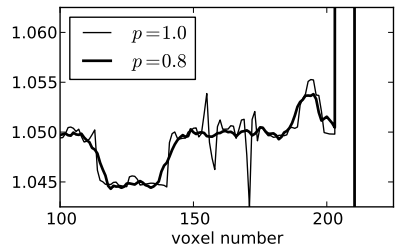


Fig. 5. Profile comparison corresponding to the images in Fig. 3 along a line in the midplane, through the ventricle and subdural hematoma.

The middle section of the head phantom is voxelized in a $256\times256\times32$ volume array, and the projection data simulate 100 projections onto a 512×80 bin flat-panel detector. Noise on the projections is modeled by employing independent 1D Gaussian distributions for each line-integration data value. The mean of each Gaussian distribution is the value of the corresponding line-integration over the phantom, and the standard deviation is taken to be 1% of this mean. The parameters of the simulation are such that it only makes sense to compare algorithms that exploit GMI sparsity, and accordingly we show results from Eq. (4) for $p = 0.8$ and $p = 1.0$. Larger p results in images that are heavily polluted with streak artifacts. For the constraint parameters, we employ an ϵ corresponding to an RMSE of 0.01, and for γ we use the value derived from the test phantom.

For the specified parameters, the image reconstruction results are shown in Fig. 3 together with corresponding slices in the phantom. The gray scale display window is 1% of the phantom complete dynamic range; and streak artifacts are difficult to avoid due to the rapidly oscillating bone structures near the ear at the bottom of the images. The results for $p = 1.0$, in the top row of the figure, show such streaks, even though this value for p does exploit GMI sparsity. The middle row shows results for the nonconvex case of $p = 0.8$, but the streak artifacts are nearly completely removed.

Inspection of the nonconvex results shows a rather interesting behavior in that the image regularization is highly non-uniform. The structures with the contrast of the eyes and greater ($\geq 1\%$ of phantom background) appear to have sharp edges, while the lower contrast structures are visible, yet, are blurred relative to the same structures in the $p = 1.0$ images. This visual impression is borne out quantitatively in vertical profile plots shown through the eyes, in Fig. 4, and through the ventricle and subdural hematoma, in Fig. 5. In the former profile, the nonconvex result has as sharp a transition at the eye border as the convex $p = 1.0$ result without the oscillations from the streaks. The latter lower contrast structures show fewer oscillation for the nonconvex result, but there is also a clear blurring as the transitions at the ventricle and hematoma borders are more gradual for $p = 0.8$ than for $p = 1.0$. This feature of the proposed nonconvex optimization can be understood from inspecting Eq. (5) where we see that the image TV term has a spatially dependent weighting. During the iteration of the nonconvex algorithm the weighting $\mathbf{w}$ evolves in such a way that less weight, and hence less smoothing, is applied to voxels with large gradient-magnitude.

IV. Summary

We have demonstrated GMI sparsity-exploiting image reconstruction by a nonconvex optimization algorithm. Under ideal conditions we have shown that the algorithm is capable of obtaining accurate image recovery with fewer projections than convex TV-based image reconstruction. The algorithm can also be applied to 3D cone-beam CT systems, and preliminary results indicate that the nonconvex algorithm can be effective in controlling steak artifacts resulting from a combination of projection view under-sampling and the presence of complex high-contrast structures.

V. Acknowledgments

This research was supported in part by the U.S. Department of Energy through the LANL/LDRD Program. This work is also part of the project CSI: Computational Science in Imaging, supported by grant 274-07-0065 from the Danish Research Council for Technology and Production Sciences. This work was also supported in part by NIH R01 grants 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] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, 2008.

[2] Rick Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

[3] Rick Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, pp. 5810–5819, 2012.

[4] Emil Y. Sidky, Rick Chartrand, and Xiaochuan Pan, "Image reconstruction from few views by non-convex optimization," in *IEEE Nuclear Science Symposium Conference Record*, Honolulu, HI, 2007, pp. 3526 – 3530.

[5] E. Y. Sidky, X. Pan, I. S. 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.

[6] P. L. Combettes, "The foundations of set theoretic estimation," *IEEE Proc.*, vol. 81, pp. 182–208, 1993.

[7] X. Han, J. Bian, E. L. Ritman, E. Y. Sidky, and X. Pan, "Optimization-based reconstruction of sparse images from few-view projections," *Phys. Med. Biol.*, vol. 57, pp. 5245–5274, 2012.

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

[9] E. Y. Sidky, J. S. Jørgensen, and X. Pan, "First-order convex feasibility algorithms for X-ray CT," *Med. Phys.*, 2013, Accepted. Arxiv preprint (http://arxiv.org/abs/1209.1069).