Adaptive Sparsifying Transforms for Iterative Tomographic Reconstruction

Luke Pfister and Yoram Bresler

***Abstract*—A major challenge in computed tomography imaging is to obtain high-quality images from low-dose measurements. Key to this goal are computationally efficient reconstruction algorithms combined with detailed signal models. We show that the recently introduced adaptive sparsifying transform (AST) signal model provides superior reconstructions from low-dose data at significantly lower cost than competing dictionary learning methods. We further accelerate this technique for tomography by utilizing the Linearized Alternating Direction Method of Multipliers (L-ADMM) to remove the need to solve an expensive least-squares problem that requires computing multiple forward and backward projections. Numerical experiments on data from clinical CT images show that adaptive sparsifying transform regularization outperforms total-variation and dictionary learning methods, and combining our regularizer with L-ADMM provides for faster reconstructions than standard ADMM.**

I. Introduction

The increased use of x-ray computed tomography (CT) in medical imaging has been accompanied by increased concerns about the x-ray exposure to the patient population. The ability to reconstruct high quality images from low-dose data has therefore become a central problem in CT. Advances in image reconstruction algorithms are key to achieving this goal. Unlike the standard filtered backprojection (FBP) algorithm, iterative reconstruction algorithms incorporate detailed models of data acquisition, noise statistics, and the signal to reconstruct. These algorithms are commonly developed as the solution to the penalized weighted-least squares (PWLS) problem [1]

$$\min_x \frac{1}{2}\|y - Ax\|_W^2 + \lambda J(x). \tag{1}$$

The first term represents a statistically weighted fidelity measure between the data vector $y \in \mathbb{R}^M$, containing the log of recieved photon counts, and the reprojected image Ax. The diagonal weighting matrix W consists of statistical weights w_i and can be motivated as a quadratic approximation to the negative log-likelihood of the image given the photon counts.

The regularization functional $J : \mathbb{R}^N \to \mathbb{R}$ improves the conditioning of (1) and encourages solutions that obey a prescribed signal model. Signal models based on sparse representations have shown to be effective for both low-dose and limited data tomography. These models have classically been instances of *analysis* sparsity, in which the image becomes sparse when acted on by a fixed linear transformation called an analysis operator. Many regularizers, such as total-variation, promote images that are sparse under finite differencing operators and are thus piecewise constant. This type of regularization can replace complex texture by flat, patchy regions. Regularization with more sophisticated analysis operators, such as shearlets, has been shown to better preserve complex texture at the expense of performance on uniform regions [2].

Recent years have shown the promise of signal models that are directly adapted to the signal of interest. A popular approach is to represent small, overlapping patches of the image as the linear combination of a few columns of a dictionary. Many algorithms have been proposed to fit a dictionary to a given set of data. Algorithms that jointly learn a dictionary while reconstructing the image have been shown to outperform traditional regularization techniques in both low-dose [3] and limited-data tomography [4]. More recently, algorithms have been developed to adaptively learn analysis operators based on manifold methods [5], variable-splitting methods [6], and modifications of dictionary learning algorithms [7].

Unfortunately, both synthesis and analysis learning algorithms scale poorly with data size and are prohibitively expensive for practical tomographic reconstruction. An alternative signal model is to assume that our signal satisfies $\Phi x = z + e$, where z is sparse and e is small. This is called the *transform model* and stipulates that x should be approximately sparse when acted on by the matrix $\Phi \in \mathbb{R}^{k \times k}$, which we call a sparsifying transform. Algorithms have been proposed to learn sparsifying transforms directly from data [8], [9]. Allowing a small deviation from exact sparsity facilitates transform learning algorithms that are much faster than competing synthesis dictionary and analysis operator learning algorithms. Adaptive sparsifying transforms have been shown to be effective in MRI reconstruction from sparsely sampled data [10].

In this paper, we show that adaptive sparsifying transform (AST) regularization outperforms both TV and synthesis dictionary methods for low-dose CT imaging, while operating at a speed rivaling that of TV [11]. Our algorithms use the Alternating Direction Method of Multipliers (ADMM) to provide a computationally efficient solution to the resulting optimization problem. The computational cost of the algorithms is dominated by solving an unweighted least squares problem that requires many applications of A and A^T. We propose the use of the Linearized ADMM algorithm to circumvent this least-squares problem and further accelerate AST-regularized tomographic reconstruction. Numerical experiments are performed using reprojected clinical images. The results show that AST regularization outperforms dictionary learning and total variation methods and that Linearized ADMM provides an appreciable decrease in computation time.

This work was supported in part by the National Science Foundation (NSF) under grants CCF-1018660 and CCF-1320953.

II. Algorithm

Our goal is to reconstruct our image $x \in \mathbb{R}^N$ from noisy projection data $y \in \mathbb{R}^M$ while simultaneously learning a sparsifying transform Φ that acts on $\sqrt{k} \times \sqrt{k}$ patches of x. We accomplish this by solving

$$\begin{aligned} \min_{x,z,\Phi} \ & \frac{1}{2}\|y - Ax\|_W^2 + \frac{\lambda}{2}\sum_j \|\Phi E_j x - z_j\|_2^2 \\ & + \lambda\left(\gamma\|z_j\|_0 + \alpha(\|\Phi\|_F^2 - \log\det\Phi)\right), \end{aligned} \tag{P1}$$

where λ, γ and α are positive scalar parameters, and $\|z_j\|_0$ is the ℓ_0 quasinorm that counts the number of nonzero elements in z_j. The matrix $E_j \in \mathbb{R}^{k \times N}$ extracts the jth $\sqrt{k} \times \sqrt{k}$ vectorized patch and removes its mean. The second term in $(P1)$ penalizes the sparsification error of patches from x, while the third term encourages sparsity in the z_j. The final two terms ensure that the learned transform is both non-singular and well conditioned. The minimization problem $(P1)$ can be viewed as a PWLS problem where the regularizer is itself the solution of a minimization problem.

We solve (P1) using an alternating minimization algorithm. With x and the z_j fixed, we update Φ by solving

$$\Phi^{k+1} = \arg\min_{\Phi} \sum_j \|\Phi E_j x - z_j\|_2^2 + \alpha(\|\Phi\|_F^2 + \log\det\Phi), \tag{2}$$

which can be solved in closed-form [12], requiring only three products of $k \times N$ and $N \times k$ matrices and one Cholesky decomposition and one SVD of $k \times k$ matrices. Typically $k << N$ and so the solution of (2) is cheap.

With x and Φ fixed, we update each z_j by solving

$$z_j^{k+1} = \arg\min_{z_j} \gamma\|z_j\|_0 + \frac{1}{2}\|\Phi E_j x - z_j\|_2^2. \tag{3}$$

The solution of (3) is given in closed form by setting to zero all entries with magnitude less than $\sqrt{\gamma}$, an operation known as *hard thresholding*. We will write this update as $z_j = \mathcal{T}_\gamma(\Phi E_j x)$.

In practice, we ensure that Φ is a good transform for the current image by alternating between updating Φ and the z_j a few times before proceeding to the image update phase.

The image update phase begins by fixing Φ and the z_j. Then, (P1) reduces to the weighted least squares problem

$$x^{k+1} = \arg\min_x \frac{1}{2}\|y - Ax\|_W^2 + \lambda\sum_j \|\Phi E_j x - z_j\|_2^2. \tag{4}$$

Owing to the size of A, direct inversion is impossible and we must resort to iterative methods. The large dynamic range in W causes the Hessian $A^T W A + \lambda\sum_j E_j^T \Phi^T \Phi E_j$W to be poorly conditioned and many iterations are required, while the placement of W makes this problem highly shift-variant and prohibits the use of efficient Fourier preconditioners [13].

Ramani & Fessler [14] proposed the use ADMM to mitigate these problems by introducing an auxiliary variable $u \in \mathbb{R}^M$. The constraint $u = Ax$ is used to split the data fidelity term and separate the projection operator A from the statistical weighting matrix W. Applying ADMM to the new constrained optimization problem results in an algorithm consisting of the following update steps:

$$x^{k+1} = \arg\min_x \frac{\lambda}{2}\sum_j \|\Phi E_j x - z_j\|_2^2 + \frac{\mu}{2}\|u^k - \eta^k - Ax\|_2^2 \tag{5}$$

$$u^{k+1} = \arg\min_u \frac{1}{2}\|y - u\|_W^2 + \frac{\mu}{2}\|u^k - \eta^k - Ax\|_2^2 \tag{6}$$

$$\eta^{k+1} = \eta^k - (u^{k+1} - Ax^{k+1}). \tag{7}$$

These subproblems are solved in an alternating fashion. Subproblem (6) is solved as

$$u^{k+1} = (W + \mu I)^{-1}(Wy + \mu(Ax^{k+1} + \eta^k)), \tag{8}$$

which is computationally inexpensive as $W + \mu I$ is diagonal. Similarly, (7) requires only vector additions and is cheap.

The bulk of computation occurs during the solution of subproblem (5). This is an unweighted least-squares problem and is amenable to Fourier-based preconditioners. However, even when an efficient preconditioner is used, the solution of (5) requires many products with A and A^T and thus remains computationally expensive.

We propose to accelerate this step by using the Linearized ADMM (L-ADMM) approach. This technique has been developed under many names and in multiple contexts; see [15] and the references within. We begin by introducing a new inertia term into the x-update subproblem (5):

$$x^{k+1} = \arg\min_x \frac{\lambda}{2}\sum_j \|\Phi E_j x - z_j\|_2^2 + \frac{\mu}{2}\|u^k - \eta^k - Ax\|_2^2 + \|x - x^k\|_Q^2 \tag{9}$$

where Q is a positive-definite matrix that can be chosen to improve the conditioning of this least squares problem. In particular, we take $Q = \delta I - \mu A^T A$ and the updated image is now given by the solution of

$$Gx^{k+1} = \lambda\sum_j E_j^T \Phi^T z_j + \delta x^k + \mu A^T(u^k - \eta^k - Ax^k) \tag{10}$$

where $G \triangleq \lambda\sum_j E_j^T \Phi^T \Phi E_j + \delta I$. We have used the matrix Q to eliminate the influence of $A^T A$ from the Hessian G.

The solution of (10) depends on the boundary conditions present in the patch extraction operators E_j. As the region of interest in CT images is surrounded by air, which provides zero attenuation, we are free to extract patches that wrap around the image boundary without incurring distortion. In this case, the quantity $\sum_j E_j^T \Phi^T \Phi E_j + \delta I$ is circularly shift-invariant and can be diagonalized using a 2D DFT. During the image update phase, Φ is held constant, so this diagonalization can be computed and stored. The solution of (10) can be computed in closed form and only requires a product with A and A^T as well as a 2D FFT/IFFT pair. This represents a significant improvement over the multiple products with $A^T A$ that are required to solve (5) using the conjugate gradient method. Note that as we do not solve a least squares problem, we do not need to precondition $A^T A$. This may prove especially beneficial for high angle cone beam CT and other geometries where circulant preconditioners are less effective.

Algorithm 1 L-AST-CT

INPUT: Initial transform Φ, observed data y
OUTPUT: Reconstructed image x

Set γ by using power iteration on A^TA
$x^0 \leftarrow \text{FBP}(y)$
$z_j^0 \leftarrow \mathcal{T}_\gamma\left(\Phi E_j x^0\right)\ \forall j$
repeat
  repeat
    Update Φ by solving (2)
    $z_j^k \leftarrow \mathcal{T}_\gamma\left(\Phi E_j x\right)\ \forall j$
  until Halting condition
  $\tilde{G} \leftarrow$ diagonalization of $\lambda \sum_j E_j^T \Phi^T \Phi E_j + \delta I$
  $i \leftarrow 0$, $u^0 \leftarrow Ax^k$, $v^0 \leftarrow \vec{0}$
  repeat
    $\zeta \leftarrow \texttt{FFT2}\left[\lambda \sum_j E_j^T \Phi^T z_j + \delta x^k + \mu A^T(u^k - \eta^k - Ax^k)\right]$
    $\tilde{x}^{i+1} \leftarrow \texttt{IFFT2}\left[\tilde{G}^{-1}\zeta\right]$
    $u^{i+1} \leftarrow (W + \mu I)^{-1}\left(Wy + \mu(A\tilde{x}^{i+1} + v^i)\right)$
    $v^{i+1} \leftarrow v^i - \left(u^{i+1} - A\tilde{x}^{i+1}\right)$
    $i \leftarrow i + 1$
  until Halting condition
  $x^{k+1} \leftarrow \tilde{x}^{i+1}$
until Halting condition

The parameter δ must be chosen to make $\delta I - \mu A^T A$ positive definite, so we require that $\delta > \mu\|A^TA\|_2$. We estimate this lower bound by performing power iteration on A^TA, and we take δ to be slightly larger than this estimate.

Note that the update (10) can alternatively be derived by linearizing the Augmented Lagrangian term in (5) about the point x^k and adding additional quadratic regularization. We choose to use the Q-norm notation as it clearly shows the required lower bound for δ. The overall algorithm, which we call L-AST-CT, is presented as Algorithm 1. We initialize the algorithm by taking x^0 to be a Hamming-weighted FBP of the data y. Our initial sparsifying transform is a separable approximation of the 2D finite differencing matrix.

III. Experiments

The algorithm was implemented using NumPy 1.8 and SciPy 0.13 on a computer containing an Intel i5-2520m processor with two cores and 6GB of RAM. The projection operator A simulates the central slice of the GE Lightspeed geometry, with 888 detector bins and 984 projections spaced between 0 and $360°$. Forward and back projections were performed using a multithreaded C implementation of the distance-driven projector and backprojector to ensure a matched projector and backprojector pair.

Our error metric is the root mean square error (RMSE), defined for $x \in \mathbb{R}^N$ as $\text{RMSE} = \sqrt{\sum_{k=1}^N (x_k - \bar{x}_k)^2/N}$, where x_k is the k-th index of x and $\bar{x}$ is the ground truth image. We compare the performance of AST-CT to Hamming-weighted FBP reconstruction and two iterative reconstruction algorithms. The first uses total-variation regularization. We use ADMM to split the non-differentiable regularization term and refer to this algorithm as TV-CT. The second, which we call DL-CT, uses a regularizer of the form $J(x) = \min_{D,a_j} \sum_j \|E_j x - Da_j\|_2^2 + \gamma\|a_j\|_0$. The update for a is solved using orthogonal matching pursuit from the efficient SPAMS [1] toolbox, and the dictionary update is performed using K-SVD. We use ADMM to separate W and A. In both TV-CT and DL-CT, we do not linearize the x-update step, but instead use a circulant preconditioner to improve the rate of convergence.

The parameters λ and γ were determined empirically by sweeping over a large range of values and choosing the parameter that corresponded to the lowest RMSE. For L-AST-CT and DL-CT, the ADMM parameter μ was chosen to ensure that the Hessian is well-conditioned, and for TV-CT it was chosen according to the strategy in [14]. The dictionary for DL-CT is of size 64×121 and is initialized with a DCT matrix, while in L-AST-CT the sparsifying transform is 64×64 and initialized with a separable approximation to the finite differencing matrix. Both algorithms use 8×8 patches with maximal overlap.

We evaluate the performance of the algorithm on data formed by reprojecting a 512×512 pixel clinical-dose CT image of a human abdomen. The clinical data consists of overlapping 0.9mm slices with 0.45mm overlap. The slices have noise standard deviation of 21 HU, as measured over a flat region in the liver. We form a ground truth image $\bar{x}$ by averaging together 5 consecutive slices of the clinical data. This reduces streaking present in our $\bar{x}$. We form clinical dose data by taking the k-th detector measurement to be $y_k = -log(\text{P}(I_0 \exp -[A\bar{x}]_k)/I_0)$, where $P(t)$ represents a Poisson random variable with mean t. Setting I_0 to 2.0×10^6 results in a noise level that matches that of the original slices. We synthesize low-dose data by taking $I_0 = 5 \times 10^5$ which corresponds to dose reduction by a factor of 4.

The TV-CT algorithm was run for 300 iterations. For DL-CT, each outer-loop iteration consists of five dictionary and sparse code updates followed by updating the image by performing 10 ADMM iterations. For L-AST-CT, the outer-loop consists of 10 sparsifying transform and sparse code updates followed by updating the image with a minimum of 30 L-ADMM iterations. In both cases, we repeat the ADMM/L-ADMM steps until the cost function has decreased. We use a total of 30 outer-loop iterations for L-AST-CT and DL-CT.

Figure 1 show the images reconstructed with each algorithm and the magnitude of the difference between the reconstructed images and $\bar{x}$. FBP suffers from the expected streaking behavior due to the reduction in dose. The TV-CT reconstruction shows no streaking artifacts, but has patchy artifacts in textured areas such as the bone. Reconstructions with adaptive regularization show low error in both the bone and soft tissue. We see that L-AST-CT outperforms DL-CT in the bone regions.

Table II illustrates the amount of time spent on updating the signal models in single outer-loop iteration of DL-CT and L-AST-CT. Even with the highly efficient multithreaded C implementations provided by SPAMS, the dictionary learning steps are significantly more expensive than the sparsifying transform updates. As an alternative benchmark, the time to perform 10 ADMM image update iterations in DL-CT is

[1] Available: http://spams-devel.gforge.inria.fr/

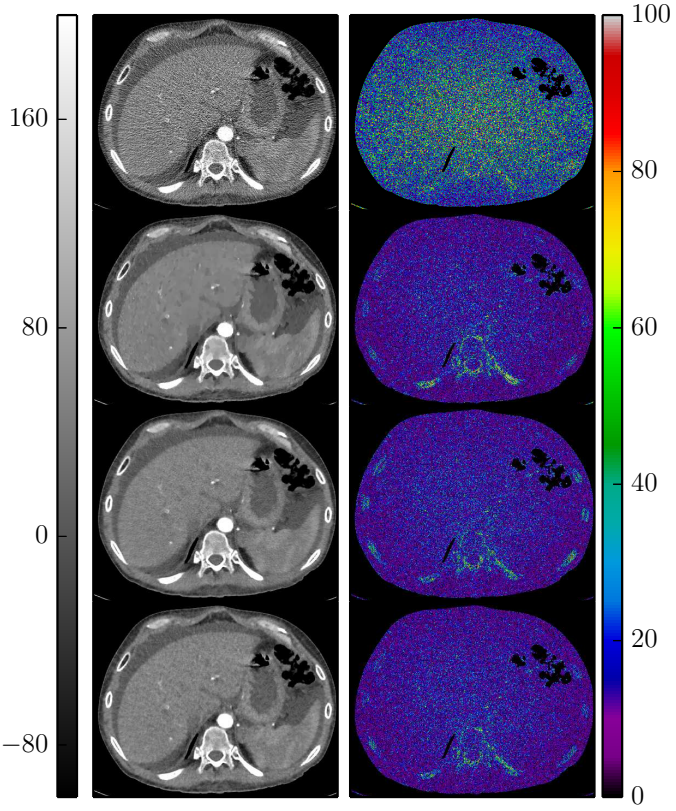

Fig. 1: Left column: Reconstruction from low-dose projections. Right column: Magnitude of error between reconstruction and $\bar{x}$. From top to bottom: FBP, TV-CT, DL-CT, L-AST-CT. All units in HU

TABLE I: RMSE in HU of reconstructions from low-dose data.

	Clinical dose FBP	Low dose FBP	L-AST-CT	DL-CT	TV-CT
RMSE	19	33	16	18	18

85 seconds. This shows that the use of dictionary learning regularization has nearly tripled the reconstruction time, while the overhead incurred by AST regularization is negligible.

We next evaluate the influence of the linearized x update step. We initialized Φ to be a separable finite differencing matrix and calculated the resulting z_j. We then fixed these variables and ran 4000 iterations of the usual ADMM algorithm to solve (4). The resulting image is denoted x^*. We then solved (4) using ADMM and L-ADMM and evaluate the distance from convergence as the RMSE between x^k and x^*. For ADMM, the least squares problem is solved by CG with a circulant preconditioner to the Hessian to accelerate convergence. Figure 2 shows the rate of convergence of both ADMM and L-ADMM. The quantity ϵ illustrates the RMSE of the ADMM reconstruction after performing the minimum 10 iterations. These results show that while L-ADMM requires more iterations to reach the value ϵ, it does so in roughly half the time as ADMM.

REFERENCES

[1] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 534–548, 1993.

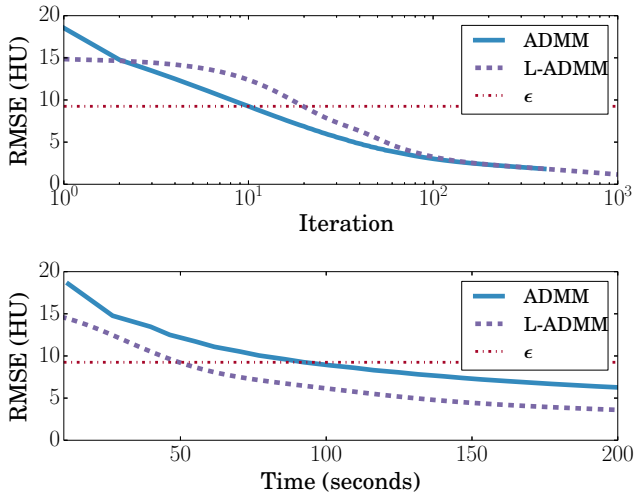

Fig. 2: Rate of convergence for ADMM and L-ADMM.

TABLE II: Amount of time to perform dictionary and transform learning steps. Units: seconds

	D/Φ Update	a/z Update	Total
DL-CT	87.5	60.3	147.8
L-AST-CT	4.4	0.2	4.6

[2] B. Vandeghinste, B. Goossens, R. Van Holen, C. Vanhove, A. Pizurica, S. Vandenberghe, and S. Staelens, "Iterative CT reconstruction using shearlet-based regularization," *Proc. SPIE 8313 Medical Imaging*, vol. 8313, pp. 83 133I–83 133I–7, 2012.

[3] Q. Xu, H. Yu, and X. Mou, "Low-dose x-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.

[4] H. Y. Liao and G. Sapiro, "Sparse representations for limited data tomography," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, May 2008, pp. 1375–1378.

[5] S. Hawe, M. Kleinsteuber, and K. Diepold, "Analysis operator learning and its application to image reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2138–2150, June 2013.

[6] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies, "Constrained overcomplete analysis operator learning for cosparse signal modelling," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2341–2355, May 2013.

[7] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[8] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, 2013.

[9] S. Ravishankar and Y. Bresler, "Learning doubly sparse transforms for image representation," in *2012 19th IEEE International Conference on Image Processing*, no. 2. IEEE, Sep. 2012, pp. 685–688.

[10] S. Ravishankar and Y. Bresler, "Sparsifying transform learning for compressed sensing MRI," in *International Symposium on Biomedical Imaging*, 2013.

[11] L. Pfister, "Tomographic reconstruction with adaptive sparsifying transforms," M.S. thesis, University of Illinois at Urbana-Champaign, 2013.

[12] S. Ravishankar and Y. Bresler, "Closed-form solutions within sparsifying transform learning," in *Acoustics Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013.

[13] J. Fessler and S. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 688–99, Jan. 1999.

[14] S. Ramani and J. Fessler, "A splitting-based iterative algorithm for accelerated statistical x-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–688, Mar. 2012.

[15] X. Zhang, M. Burger, and S. Osher, "A unified primal-dual algorithm framework based on bregman iteration," *Journal of Scientific Computing*, vol. 46, no. 1, pp. 20–46, Jan 2011.