

# Model-Based Iterative Tomographic Reconstruction with Adaptive Sparsifying Transforms

Luke Pfister and Yoram Bresler

Dept. of Electrical and Computer Engineering  
University of Illinois at Urbana-Champaign

## ABSTRACT

Model based iterative reconstruction algorithms are capable of reconstructing high-quality images from low-dose CT measurements. The performance of these algorithms is dependent on the ability of a signal model to characterize signals of interest. Recent work has shown the promise of signal models that are learned directly from data. We propose a new method for low-dose tomographic reconstruction by combining adaptive sparsifying transform regularization within a statistically weighted constrained optimization problem. The new formulation removes the need to tune a regularization parameter. We propose an algorithm to solve this optimization problem, based on the Alternating Direction Method of Multipliers and FISTA proximal gradient algorithm. Numerical experiments on the FORBILD head phantom illustrate the utility of the new formulation and show that adaptive sparsifying transform regularization outperforms competing dictionary learning methods at speeds rivaling total-variation regularization.

**Keywords:** Sparsifying transform learning, Sparse representations, CT dose reduction, Iterative reconstruction

## 1. INTRODUCTION

A major requirement in computed tomography(CT) is the reduction of harmful x-ray dose while maintaining the quality of reconstructed images. Recently, the Model-based Iterative Reconstruction (MBIR) framework has proven to be effective in solving a variety of inverse problems, including low-dose x-ray CT<sup>1,2</sup> The MBIR approach include three main ingredients: a forward model that describes how measurements are formed, a noise model that describes how measurements are corrupted, and a signal model that captures prior information about the signals we wish to recover. These ingredients are then tied together into an objective function that is minimized using an iterative algorithm.

A popular approach to low-dose CT reconstruction is to combine the noise and system model together into a penalized weighted least-squares (PWLS) problem:

$$\underset{x}{\text{minimize}} \quad \frac{1}{2} \|y - Ax\|_W^2 + \lambda J(x). \quad (1)$$

Here, the data vector  $y \in \mathbb{R}^M$  contains the log of the received photon counts,  $x \in \mathbb{R}^N$  is an estimate of the image to reconstruct, and the system matrix  $A$  models the forward projection operation. The diagonal matrix  $W$  consists of statistical weights  $w_i$ . The quantity  $J(x) : \mathbb{R}^N \rightarrow \mathbb{R}$  is a regularization functional that promotes images that match our prescribed signal model. Common choices for this regularizer include the total-variation (TV) seminorm, the  $\ell_1$  norm of the wavelet coefficients of the image, or the qGGMRF regularization functional. A trait shared by these regularizers is that they promote images that are sparse under a particular representation; for example, TV promotes images that are sparse under a finite-differencing operator; such images are piecewise-constant. Although these regularizers have been shown to be effective for both low-dose and limited data tomography, regularizers that promote piecewise constant images can replace complex texture by patchy, uniform regions. More sophisticated regularizers, such as the combination of  $\ell_1$  and shearlets, have shown to preserve complex texture at the cost of poorer performance on uniform regions.<sup>3</sup>

Recent years have shown the promise of sparse representations that are directly adapted to a class of signals rather than analytically designed. A popular approach is to assume that small, overlapping image patches can be formed by the product of a matrix called a dictionary, and a vector called a sparse code. This is known as the synthesis sparsity model, and many algorithms have been proposed to find such a dictionary and sparse codes, given a set of training data. These algorithms tend to alternate between updating the dictionary and updating the sparse codes. Regularization based on the synthesis sparsity model has been shown to be effective for both low-dose and limited-angle CT.<sup>4-6</sup> Unfortunately, the sparse coding update step is NP-Hard and algorithms to approximate a solution generally scale poorly with data size. As a result, dictionary-learning based regularization can dramatically increase the computational cost of the already challenging problem of CT reconstruction.

An alternative sparse signal model is to assume that image patches become sparse when acted on by a linear operator called an analysis operator. Recently, there has been an increase in the development of algorithms that use the analysis sparsity model. Many of these algorithms use the *analysis model*, which suggests that a signal  $x$  should be exactly sparse when acted on by a linear transform  $\Omega$ . When  $x$  is corrupted by noise the model can be extended to the noisy analysis model by taking  $x = q + e$ , where  $e$  is a (small) error in the signal domain, and  $\Omega q = z$  is sparse.<sup>7</sup> Unfortunately, as in the synthesis sparsity model, algorithms to learn  $\Omega$  and find the corresponding sparse codes  $z$  are also NP-Hard and algorithms to find an approximate solution are computationally expensive.

Recently, Ravishankar and Bresler have proposed several algorithms to learn signal representations based on the *transform sparsity* model,<sup>8,9,10</sup> In this framework, data is assumed to satisfy  $Wx = z + \eta$ , where  $z$  is sparse and  $\eta$  is small. The signal  $x$  need only be approximately sparsified when acted on by the matrix  $W$ , which is called a *sparsifying transform*. Transform sparsity, named as such owing to its similarity with transform coding schemes,<sup>11</sup> can be viewed as an alternative approximation to the co-sparse analysis model. The primary distinction between the noisy analysis and transform signal models is that the latter allows for deviation from exact sparsity in the transform, rather than the signal, domain. This characteristic allows for the design of efficient algorithms to learn sparsifying transforms from data. Regularization with adaptive sparsifying transforms has been shown to provide state-of-the-art performance in low-dose CT<sup>12</sup> and MRI<sup>13</sup> imaging applications.

Regardless of the choice of regularizer, it can be difficult to set the PWLS regularization parameter  $\lambda$ , which controls the bias-variance tradeoff induced by the regularizer. The MAP-optimal choice of  $\lambda$  requires knowledge of the value of  $J(x)$  evaluated on the true image, information that we do not have access to. Instead, the usual approach is to tune  $\lambda$  to provide the best empirical reconstruction. This generally involves performing reconstructions for many values of  $\lambda$  followed by selecting the best reconstruction. Such an approach is unacceptable for practical CT reconstruction, where even a single reconstruction requires significant time and computational resources. The difficulty of parameter selection has motivated several automatic parameter tuning strategies for linear inverse problems; for example, those based on Stien’s Unbiased Risk Estimator.<sup>14</sup>

Niu & Zhu<sup>15</sup> proposed to eliminate the problem of parameter selection for CT reconstruction by replacing the PWLS problem (1) with the constrained problem

$$\underset{x}{\text{minimize}} J(x) \text{ s. t. } \|y - Ax\|_2^2 \leq \varepsilon, \tag{2}$$

where the tolerance level  $\varepsilon$  is chosen using the Poisson statistical noise model. The constrained problem is transformed into an unconstrained problem through the use of a log-barrier barrier function, and the unconstrained minimization problem is solved using gradient projection with a Barzilai-Borwein step selection procedure. This constrained problem requires that the overall deviation between  $y$  and the reprojection  $Ax$  of the reconstructed image be less than the noise variance. However, this constraint fails to account for the variance on a per-projection basis, which is crucial in low-dose CT as the variance of each detector reading varies greatly depending on the path travelled by the x-ray.

In Section 2, we develop a we develop a modified version of the constrained formulation (2) that incorporates per-projection variance information through a weighted norm. The new constrained formulation is combined with adaptive sparsifying transform regularization. In Section 3, we propose to solve this optimization problem using the Split Augmented Lagrangian Shrinkage Algorithm (SALSA)<sup>16</sup> in combination with the FISTA proximal gradient algorithm. Simulations in Section 4 show that the inclusion of the statistical weighting improves

reconstruction from low-dose measurements. We show that adaptive sparsifying transform regularization does not suffer from the patchy artifacts of total-variation regularization, and facilitates reconstruction while enable reconstruction much faster than with dictionary learning regularization.

## 2. PROBLEM FORMULATION

### 2.1 Adaptive Sparsifying Transform Regularization

Adaptive sparsifying transform regularization is based on the principle that we can find a matrix,  $\Phi \in \mathbb{R}^{k \times k}$ , that will nearly sparsify patches from the image  $x$ . We define a matrix  $E_j \in \mathbb{R}^{k \times N}$  to have the form  $E_j = (I_k - k^{-1}\mathbf{1}^T\mathbf{1})R_j$ , where  $\mathbf{1} \in \mathbb{R}^k$  is a vector of all ones and  $R_j \in \mathbb{R}^{k \times N}$  is a matrix that extracts and vectorizes the  $\sqrt{k} \times \sqrt{k}$  patch whose top-leftmost pixel is the  $j$ -th index of the vectorized image. The matrix  $(I_k - k^{-1}\mathbf{1}^T\mathbf{1})$  removes the mean of the image patch, which is necessary to ensure that patches that differ by a constant offset will be sparsified in the same way.

The transform model stipulates that  $\Phi E_j x = z_j + e_j$ , where the vector  $z_j \in \mathbb{R}^k$  is sparse and the deviation from sparsity,  $e_j \in \mathbb{R}^k$ , is small. We will promote sparsity by penalizing the  $\ell_1$  norm of the transform coefficients  $z_j$ . Given an image  $x$ , we will search for both  $\Phi$  and the  $z_j$  by solving an optimization problem. Viewed in the MBR framework, this corresponds to a regularizer that is itself the solution of a minimization problem:

$$J(x) \triangleq \min_{z, \Phi} \sum_j \|\Phi E_j x - z_j\|_2^2 + \gamma \|z_j\|_1 + \alpha (\|\Phi\|_F^2 - \log \det \Phi) \quad (3)$$

The first term penalizes the sparsification error of the mean-removed patches of  $x$ , while the second term promotes sparsity in  $z_j$ . The final two terms promote well-conditioned  $\Phi$  by ensuring that it is both well-scaled and non-singular.

### 2.2 Constrained Optimization Problem

Adopting the usual Poisson model for transmission CT, we model the number of photons received at the detector,  $g_k$ , corresponding to the  $k$ -th projection  $[Ax]_k$ , as a Poisson random variable with  $g_k \sim \text{Poi}\{I_0 e^{-[Ax]_k}\}$ . The photon flux  $I_0$  represents the number of photons received at the  $k$ -th detector when no object is present in the x-ray path. We assume this quantity is known and constant for all projections. For simplicity, we neglect both electronic readout noise and background events.

We gather the received photon counts into a random vector  $g \in \mathbb{R}^M$ . Adopting the usual assumption that the  $g_k$  are statistically independent, a quadratic approximation to the negative-log likelihood  $L(g|x)$  yields the familiar weighted-norm term

$$L(g|x) \approx -\frac{1}{2} \|y - Ax\|_W^2, \quad (4)$$

where  $W$  is a diagonal matrix with  $w_k = g_k$ , and  $y_k = -\log(g_k/I_0)$ .

This oft-used quadratic approximation to  $L(g|x)$  implies that each  $y_k$  is well-modelled as a Gaussian random variable with mean  $[Ax]_k$  and variance  $1/w_k$ . Thus, the quantity  $\sqrt{w_k} \cdot (y_k - [Ax]_k)$  is a standard normal random variable and the sum  $\sum_{k=1}^M w_k (y_k - [Ax]_k)^2$  are distributed according to the central  $\chi_M^2$  distribution. It follows that the mean and variance of  $\|y - Ax\|_W^2$  are  $M$  and  $2M$ , respectively. This motivates the new constrained problem

$$\underset{x}{\text{minimize}} J(x) \text{ s. t. } \|y - Ax\|_W^2 \leq \varepsilon, \quad (5)$$

where  $\varepsilon = c \cdot (M + 2\sqrt{2M})$ . The scalar  $c$  is a user-tunable parameter to account for measurement errors that are not captured by the Poisson noise model, such as discretization errors, scatter, or beam hardening effects. For the remainder of this work we will assume  $c = 1$ . In this case, if the measurement model holds exactly and  $x$  equals to the true image, applying the Central Limit Theorem for large  $M$  yields that the constraint in (5) will be satisfied with probability of 98%.

The constraint in (5) improves on (2) by incorporating our knowledge of the variance of each projection. Further, when  $J(x)$  is a convex regularizer, there exists a particular parameter  $\lambda$  for which the unconstrained PWLS problem (1) and the constrained formulation (5) are equivalent.<sup>17</sup>

We will combine the AST regularizer (3) with the constrained formulation in (5) to form our reconstruction problem

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \left\{ \min_{z, \Phi} \sum_j \|\Phi E_j x - z_j\|_2^2 + \gamma \|z_j\|_1 + \alpha (\|\Phi\|_F^2 - \log \det \Phi) \right\} \\ & \text{s. t.} \quad \|y - Ax\|_W^2 \leq \varepsilon. \end{aligned} \tag{P1}$$

### 3. ALGORITHM

We solve (P1) using a block alternating minimization scheme. We alternate between updating the regularizer by minimizing over the  $z_j$  and  $\Phi$ , followed by reconstructing the image by minimizing with respect to  $x$ .

#### 3.1 Regularizer Update

We update the regularizer by minimizing over  $\Phi$  and  $z$  in an alternating fashion. With  $x$  and  $z$  fixed, the transform  $\Phi$  is updated by solving

$$\Phi^{k+1} = \arg \min_{\Phi} \sum_j \|\Phi E_j x - z_j\|_2^2 + \alpha (\|\Phi\|_F^2 - \log \det \Phi), \tag{6}$$

which can be solved in closed-form.<sup>9</sup> The closed-form solution requires three products of  $k \times N$  and  $N \times k$  matrices, as well as one Cholesky decomposition and one SVD of matrices of size  $k \times k$ .

With  $\Phi$  and  $x$  fixed, each  $z_j$  is updated by solving

$$z_j^{k+1} = \arg \min_{z_j} \|\Phi E_j x - z_j\|_2^2 + \gamma \|z_j\|_1. \tag{7}$$

This minimization problem can be solve in closed-form by soft thresholding each component of  $\Phi E_j x$  at level  $\gamma$ . For a scalar  $a$ , soft thresholding at level  $\lambda$  is defined as

$$\mathcal{T}_\lambda(a) = \begin{cases} 0, & |a| \leq \lambda \\ \left(1 - \frac{\lambda}{|a|}\right) a, & \text{otherwise.} \end{cases} \tag{8}$$

To ensure that  $\Phi$  is a good sparsifying transform for the current image, we alternate between updating all  $z_j$  and  $\Phi$  a few times before proceeding.

#### 3.2 Image Update

With the regularizer updated we can now move to the image reconstruction step. Fixing  $\Phi$  and  $z$ , we form the new image estimate by solving

$$x^{k+1} = \arg \min_x \sum_j \|\Phi E_j x - z_j\|_2^2 \quad \text{s. t.} \quad \|y - Ax\|_W^2 \leq \varepsilon. \tag{9}$$

We propose to solve this constrained minimization problem using the SALSA algorithm. SALSA involves transforming the constrained minimization problem (9) into an equivalent unconstrained problem, then attacking the unconstrained problem using the Alternating Direction Method of Multipliers (ADMM). We do this by first defining the indicator function for our constraint set  $\mathcal{C} = \{v \in \mathbb{R}^M : \|y - v\|_W^2 \leq \varepsilon\}$  as

$$I_{\mathcal{C}}(v) \triangleq \begin{cases} 0, & \|y - v\|_W^2 \leq \varepsilon \\ \infty, & \text{else,} \end{cases} \tag{10}$$

and rewrite the  $x$ -update step (9) in the unconstrained form

$$x^{k+1} = \arg \min_x \sum_j \|\Phi E_j x - z_j\|_2^2 + I_{\mathcal{C}}(Ax). \tag{11}$$

We now use the ADMM variable splitting method to break the minimization problem (11) into several smaller, more manageable optimization problems. We accomplish this by introducing writing a problem equivalent to (11),

$$\underset{x}{\text{minimize}} \sum_j \|\Phi E_j x - z_j\|_2^2 + I_{\mathcal{C}}(v) \quad \text{s. t.} \quad v = Ax. \quad (12)$$

We can then write the augmented Lagrangian function for this problem as

$$\mathcal{L}(x, v, \eta) = \sum_j \|\Phi E_j x - z_j\|_2^2 + I_{\mathcal{C}}(v) + \frac{\mu}{2} \|v - Ax - \eta\|_2^2 - \frac{\mu}{2} \|\eta\|_2^2, \quad (13)$$

where the vector  $\eta \in \mathbb{R}^M$  can be viewed as a scaled version of the Lagrange multiplier corresponding to the constraint  $v = Ax$ , and the scalar  $\mu > 0$  is a parameter that affects the rate of convergence, but not the overall solution, to the optimization problem.<sup>18</sup>

We must now solve a saddle point problem involving  $\mathcal{L}(x, v, \eta)$ , alternating between minimizing over  $x$  and  $v$  followed by a gradient step to maximize with respect to  $\eta$ . At the  $k$ -th iteration, we must solve

$$x^{k+1} = \arg \min_x \sum_j \|\Phi E_j x - z_j\|_2^2 + \frac{\mu}{2} \|v^k - \eta^k - Ax\|_2^2, \quad (14)$$

$$v^{k+1} = \arg \min_v I_{\mathcal{C}}(v) + \frac{\mu}{2} \|v - (\eta^k + Ax^{k+1})\|_2^2, \quad (15)$$

$$\eta^{k+1} = \eta^k - v^{k+1} + Ax^{k+1}. \quad (16)$$

We will examine the  $x$  and  $v$  subproblems in detail.

The  $x$ -update step (14) is a linear least-squares problem with solution given by

$$x^{k+1} = \left( \mu A^T A + \sum_j E_j^T \Phi^T \Phi E_j \right)^{-1} \left( \mu A^T (v^k - \eta^k) + \sum_j E_j^T \Phi^T z_j \right). \quad (17)$$

As the Hessian  $H \triangleq \mu A^T A + \sum_j E_j^T \Phi^T \Phi E_j$  is of size  $M \times M$ , direct inversion is infeasible and iterative methods, such as the conjugate gradient algorithm, must be used. The convergence rate of conjugate gradient can be greatly improved through the use of preconditioning. A popular choice is to take advantage of the (approximate) shift-invariance in  $A^T A$  by using a circulant preconditioner.<sup>19</sup> Such preconditioners can be efficiently implemented using the FFT algorithm and considerably accelerate convergence. If  $E_j$  extracts overlapping patches that wrap-around the image boundary, the term  $\sum_j E_j \Phi^T \Phi E_j$  corresponds to a circularly shift-invariant operator and thus  $H^{-1}$  can be well-approximated using circulant preconditioners.

The  $v$ -update step can be viewed as the proximal operator of  $I_{\mathcal{C}}$  evaluated at  $Ax^{k+1} + \eta^k$ , or equivalently as the projection of the point  $Ax^{k+1} + \eta^k$  onto the ellipsoidal constraint set  $\mathcal{C}$ . The problem of projecting a point onto an ellipse is as a quadratic optimization problem, and there have been several algorithms proposed for its solution. We will use a change of variables to reformulate the projection into a quadratic optimization problem with an unweighted  $\ell_2$ -norm constraint, and then use the FISTA accelerated proximal gradient algorithm to solve the reformulated problem.

Let  $I_{\tilde{\mathcal{C}}}$  be the indicator function for the unweighted  $\ell_2$  norm ball of radius  $\sqrt{\varepsilon}$ , defined as

$$I_{\tilde{\mathcal{C}}}(x) = \begin{cases} 0, & \|x\|_2^2 \leq \varepsilon \\ \infty, & \text{else.} \end{cases} \quad (18)$$

This new indicator function is related to  $I_{\mathcal{C}}$  through the definition of the weighted norm, with  $I_{\mathcal{C}}(v) = I_{\tilde{\mathcal{C}}}(W^{\frac{1}{2}}(v - y))$ . Here, the diagonal entries of  $W^{\frac{1}{2}}$  are given by  $\sqrt{w_{ii}}$ , which is well-defined as the photon counts are non-negative. We can now pose the  $v$ -update problem as the solution of

$$\arg \min_v I_{\tilde{\mathcal{C}}}(W^{\frac{1}{2}}(v - y)) + \frac{\mu}{2} \|v - (\eta^k + Ax^{k+1})\|_2^2. \quad (19)$$

It is the placement of  $W^{\frac{1}{2}}$  within the indicator function that prohibits a closed-form solution to this proximity mapping. We can untangle these terms by exploiting the invertibility of  $W^{\frac{1}{2}}$ . We define the change of variables  $\zeta = W^{\frac{1}{2}}(v - y)$  and rewrite the problem as

$$\arg \min_{\zeta} I_{\tilde{\mathcal{C}}}(\zeta) + \frac{\mu}{2} \|W^{-\frac{1}{2}}\zeta - (\eta^k + Ax^{k+1} - y)\|_2^2. \quad (20)$$

This is no longer in the form of a proximal mapping, but is instead the sum of a convex, lower semi-continuous indicator function and a quadratic term. Problems of this form can be efficiently solved using proximal gradient methods. We make use of the FISTA algorithm,<sup>20</sup> chosen for its ease of implementation and rapid convergence. The complete algorithm for solving (15) is listed as Algorithm 1. The dominant computations in the  $v$ -update step are multiplication by  $W^{-\frac{1}{2}}$  and the evaluation of the  $\ell_2$  norm. The required multiplication is inexpensive operation as  $W$  is a diagonal matrix. As FISTA is used to solve a subproblem of the overall optimization problem, we do not require that FISTA reach convergence, and instead use a fixed number of 15 iterations.

---

**Algorithm 1** FISTA for  $v$ -update

---

- 1:  $u^0 \leftarrow W^{\frac{1}{2}}(v^k - y)$
  - 2:  $t^0 \leftarrow 1$
  - 3:  $L \leftarrow 1/\|W\|_{\infty}$
  - 4: **repeat**
  - 5:    $\tilde{\zeta}^i \leftarrow u^i - \frac{\mu}{L} \left( W^{-1}u^i - W^{-\frac{1}{2}}(\eta^k + Ax^{k+1} - y) \right)$
  - 6:    $\zeta^i \leftarrow \sqrt{\varepsilon} \cdot \tilde{\zeta}^i / \|\tilde{\zeta}^i\|_2$
  - 7:    $t^{i+1} \leftarrow (1 + \sqrt{1 + 4(t^i)^2})/2$
  - 8:    $u^{i+1} \leftarrow \zeta^i + \frac{t^i - 1}{t^{i+1}}(\zeta^i - \zeta^{i-1})$
  - 9:    $i \leftarrow i + 1$
  - 10: **until** Halting condition
  - 11:  $v^{k+1} \leftarrow W^{-\frac{1}{2}}\zeta^i + y$
- 

### 3.3 Overall Algorithm and Parameter Selection

The overall algorithm, which we call C-AST-CT, is listed as Algorithm 2. The algorithm is initialized using an FBP reconstruction of the noisy measurements.

We have three main parameters that must be chosen. The parameter  $\alpha$  is set empirically to a value that results in well conditioned  $\Phi$ . As the algorithm is somewhat robust to the choice of this parameter we can search for  $\alpha$  by learning transforms on the FBP reconstructed image. The sparsity parameter  $\gamma$  is more difficult to select. The parameter should be large enough to reject noise, which is not sparsified well, but low enough to preserve low-contrast features in the image. Unlike the synthesis dictionary learning setting, there is no direct link between noise rejection and sparsity level unless  $\Phi$  is an orthonormal transform. As such,  $\gamma$  must be tuned empirically for good performance.

Finally, we must select the ADMM parameter  $\mu$ . We follow the suggestion of Goldstein and Osher<sup>21</sup> and select  $\mu$  to ensure that the linear system in (17) is well conditioned. As we cannot efficiently estimate the conditioning of  $H$ , we instead minimize the condition number of the circulant approximation to  $H^{-1}$ . This is a simple scalar minimization problem with low computational cost.

If  $J(x)$  is a closed, proper, and convex function, Algorithm 2 is guaranteed to converge so long as the matrix  $[W^{\frac{1}{2}}, A]^T$  has full column rank.<sup>22</sup> The latter condition clearly holds as  $W^{\frac{1}{2}}$  is invertible. Although our regularizer as defined in (3) is nonconvex, it becomes convex for fixed  $z_j$  and  $\Phi$ . Empirically, both  $\Phi$  and the  $z_j$  change little after roughly 10 iterations of Algorithm 2. We use this property to motivate a heuristic stopping criterion: once a suitable  $z_j$  and  $\Phi$  are found, we cease updating them and run the SALSA image update steps until the algorithm has been determined to converge.

---

**Algorithm 2** C-AST-CT

---

**INPUT:** Initial transform  $\Phi$ , observed data  $y$ **OUTPUT:** Reconstructed image  $x$ 

```
1:  $x^0 \leftarrow \text{FBP}(y)$ 
2:  $z_j^0 \leftarrow \mathcal{T}_\gamma(\Phi E_j x^0) \quad \forall j$ 
3: repeat
4:   repeat
5:     Update  $\Phi$  by (6)
6:      $z_j^k \leftarrow \mathcal{T}_\gamma(\Phi E_j x) \quad \forall j$ 
7:   until Halting condition
8:    $i \leftarrow 0$ 
9:    $u^0 \leftarrow Ax^k$ 
10:   $v^0 \leftarrow \vec{0}$ 
11:  repeat
12:    Use preconditioned CG to find approximate solution
    of  $H\tilde{x}^{i+1} = \mu A^T(v^i - \eta^i) + \sum_j E_j^T \Phi^T z_j^i$ 
13:    Use Algorithm 1 to solve
     $v^{i+1} \leftarrow \arg \min_v I_C(v) + \frac{\mu}{2} \|v - (Ax^{i+1} + \eta^i)\|_2^2$ 
14:     $\eta^{i+1} \leftarrow \eta^i - v^{i+1} + A\tilde{x}^{i+1}$ 
15:     $i \leftarrow i + 1$ 
16:  until Halting condition
17:   $x^{k+1} \leftarrow \tilde{x}^{i+1}$ 
18: until Halting condition
```

---

## 4. EXPERIMENTS

The algorithms were implemented using NumPy 1.8 and SciPy 0.13 on a computer containing an Intel i5-2520m processor with 6GB of RAM. The projection  $Ax$  and back projection  $A^T y$  were performed using a multithreaded C implementation of the distance-driven projector and backprojector, which ensures a matched projector and backprojector pair.<sup>23</sup> The system matrix  $A$  simulates the GE Lightspeed geometry,<sup>24</sup> with 888 detector bins and 984 projections evenly spaced between 0 and 360°. All simulations assume a 80 keV monoenergetic source and the linear attenuation coefficient of water is set to 1.83 mm<sup>-1</sup>.

Our error metric is the root mean square error (RMSE), defined for an image with  $N$  pixels as  $RMSE = \sqrt{\sum_{k=1}^N (x_k - \bar{x})^2 / N}$ , where  $x_k$  is pixel of the reconstructed image and  $\bar{x}$  is the ground truth image.

We compare the performance of regularization with adaptive sparsifying transforms against FBP reconstruction, as well as two constrained iterative reconstruction schemes. The first, which we refer to as C-TV-CT, takes  $J(x) = \|x\|_{TV}$ . We use ADMM to handle this non-differentiable regularizer in the same manner as in the unconstrained case.<sup>25</sup> The second algorithm, which we call C-DL-CT, combines our constrained formulation with dictionary learning regularizer. This is accomplished by setting

$$J(x) = \sum_j \frac{1}{2} \|E_j x - D a_j\|_2^2 + \gamma \|a_j\|_0 \quad (21)$$

and suitably modifying Algorithm (2). The update for  $a_j$  becomes a synthesis sparse coding problem, which we solve using the efficient SPAMS toolbox\*. The dictionary  $D$  is updated using the K-SVD algorithm.<sup>26</sup> We use the same constraint set and preconditioning strategy for C-AST-CT, C-DL-CT, and C-TV-CT, so that the only difference in the algorithms comes from the choice of regularizer.

We evaluate the performance of the algorithms on 350 × 350 pixel slice of the FORBILD head phantom, occupying a FOV 25.6 cm in diameter. Projection data  $P_\theta(t)$  was formed by sampling analytic line integrals<sup>27</sup>

---

\*Available: <http://spams-devel.gforge.inria.fr>

Table 1. RMSE of low-dose phantom reconstructions. (units: HU)

FBP	C-AST-CT $\ell_2$ norm	C-AST-CT Weighted norm	C-DL-CT Weighted norm	C-TV-CT Weighted norm
81	78	71	72	68

through the phantom at 0.128 mm intervals. To account for the finite width of the x-ray, we approximate strip integrals by combining 8 consecutive samples according to

$$\hat{P}_\theta[t] = -\log\left(\frac{1}{8}\sum_{i=-4}^3 \exp(-P_\theta[t+i+\delta/2])\right). \quad (22)$$

The resulting projections  $\hat{P}_\theta[t]$  correspond to 1.0239 mm detectors. The simulated photon counts are generated from the strip integrals using the Poisson model with

$$y_\theta[t] = -\log\left(\frac{\text{Poi } I_0 \exp(-\hat{P}_\theta[t])}{I_0}\right). \quad (23)$$

The data vector is then formed by collecting all photon counts into the vector  $y \in \mathbb{R}^M$ .

We form ‘clinical dose’ data by setting  $I_0 = 3.5 \times 10^6$ . This results in a noise level of roughly 3.5 HU in a Hamming-weighted FBP reconstruction, where the noise level is estimated over a flat region of the phantom. We form low-dose data with  $I_0 = 1.75 \times 10^6$ , corresponding to dose reduction by a factor of 2. A discretized version of the phantom is taken to be  $\bar{x}$ .

For C-AST-CT, we update alternate between updating the  $z_j$  and  $\Phi$  five times before moving to the image update step. We then perform 30 SALSA iterations. This process is repeated 10 times, after which we fix  $\Phi$  and  $z_j$  and perform 150 SALSA update iterations to form the final reconstructed image. We use the same strategy for C-DL-CT. For C-TV-CT, we perform 500 SALSA iterations. In all algorithms, the  $v$ -update step is solved using a fixed number of 15 FISTA iterations. We initialize  $\Phi$  in C-AST-CT using a separable approximation to the finite differencing operator. In C-DL-CT, the initial dictionary is a  $64 \times 121$  overcomplete DCT matrix. Both algorithms use  $8 \times 8$  image patches.

Figure 1 clearly illustrates the benefit of the per-projection weighting in (5) over the unweighted constraint in (2). Both C-AST-CT reconstructions were performed using the same number of iterations, but the reconstruction performed with a weighted norm constraint shows significantly less noise throughout the image.

Figure 2 shows reconstructed images using each of the regularizers and the weighted norm constraint. As expected, TV regularization performs well on this piecewise-constant phantom. However, there are noticeably patchy regions throughout the image. The images reconstructed using C-AST-CT and C-DL-CT regularization exhibit no patchy regions, but there is a noticeable smoothing effect. This may be remedied by more careful tuning of the sparsity parameter. The DL reconstruction has strong streaking artifacts near the ear. It is possible that the learning algorithm characterized these streaks as features to be learned, rather than rejected. Although present, the streaking in the C-AST-CT reconstruction is much less pronounced.

Table 2 illustrates the average amount of time needed to complete one outer-loop iteration of each algorithm. For C-DL-CT, the Dictionary Update time corresponds to the amount of time needed for 5 K-SVD iterations. For C-AST-CT,  $\Phi$  Update refers to the amount of time needed for 5 updates of the sparsifying transforms. For all algorithms, Image Update is the amount of time needed for 30 SALSA image update iterations. These results show that C-AST-CT performs at nearly the same speed as C-TV-CT, while C-DL-CT increases the computation time by roughly 30%. This computational advantage will be further amplified as the image size grows.

## 5. CONCLUSIONS

Model-based iterative reconstruction methods combine physical and statistical models of image acquisition with detailed signal models to achieve state of the art results in low-dose CT imaging. The performance of these



Table 2. Computation time for one outer iteration for each of the algorithms. Units: seconds

	$D/\Phi$ Update	$a/z$ Update	Image Update	Total
C-TV-CT	0	0	257.7	257.7
C-DL-CT	54.5	20.3	249.1	323.9
C-AST-CT	2.1	0.1	254.4	256.6

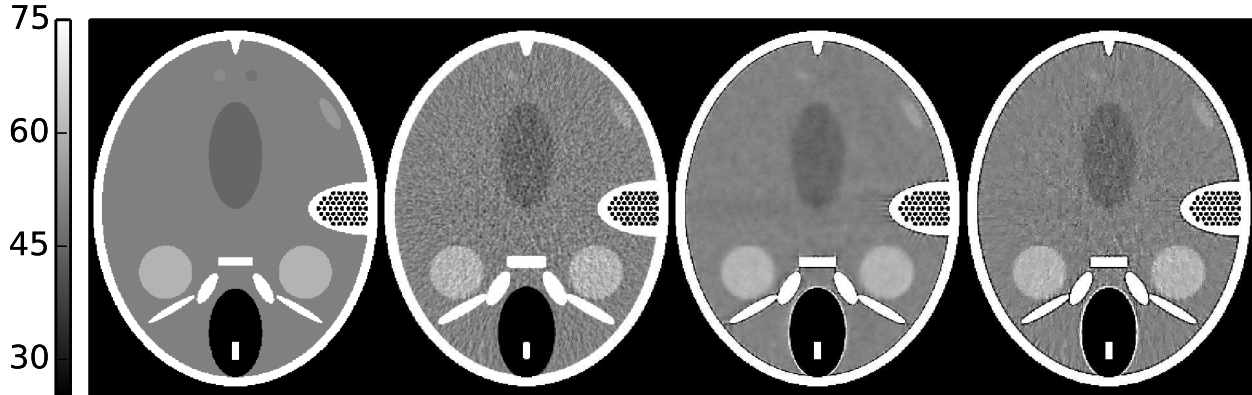


Figure 1. From left to right: discretized phantom, FBP reconstruction, C-AST-CT reconstruction with weighted norm, C-AST-CT reconstruction with  $\ell_2$  norm. All units in HU.

algorithms depend on the signal model and the choice of regularization parameter. We have proposed the use of regularization based on adaptive sparsifying transforms, in conjunction with a constrained optimization problem that removes the need to tune the regularization parameter while tightly incorporating the statistical signal model. Although we focused on the use of adaptive sparsifying transform regularization, the new constrained formulation and optimization scheme can be adapted for any regularizer.

## ACKNOWLEDGMENTS

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

## REFERENCES

- [1] Sauer, K. and Bouman, C., “A local update strategy for iterative reconstruction from projections,” *IEEE Trans. Signal Process.* **41**(2), 534–548 (1993).
- [2] Thibault, J.-B., Sauer, K. D., Bouman, C. A., and Hsieh, J., “A three-dimensional statistical approach to improved image quality for multislice helical CT,” *Med. Phys.* **34**(11), 4526 (2007).
- [3] Vandeghinste, B., Goossens, B., Van Holen, R., Vanhove, C., Pizurica, A., Vandenberghe, S., and Staels, S., “Iterative CT reconstruction using shearlet-based regularization,” *Proc. SPIE 8313 Medical Imaging* **8313**, 83133I–83133I–7 (2012).
- [4] Liao, H. Y. and Sapiro, G., “Sparse representations for limited data tomography,” in [2008 5<sup>th</sup> IEEE International Symposium on Biomedical Imaging: From Nano to Macro], 1375–1378, IEEE (May 2008).
- [5] Shtok, J., Elad, M., and Zibulevsky, M., “Sparsity-based sinogram denoising for low-dose computed tomography,” in [2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)], 569–572, IEEE (May 2011).
- [6] Xu, Q., Yu, H., and Mou, X., “Low-dose x-ray CT reconstruction via dictionary learning,” *IEEE Trans. Med. Imag.* **31**, 1682–1697 (Sept. 2012).
- [7] Rubinstein, R., Peleg, T., and Elad, M., “Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model,” *IEEE Trans. Signal Process.* **61**, 661–677 (Feb. 2013).

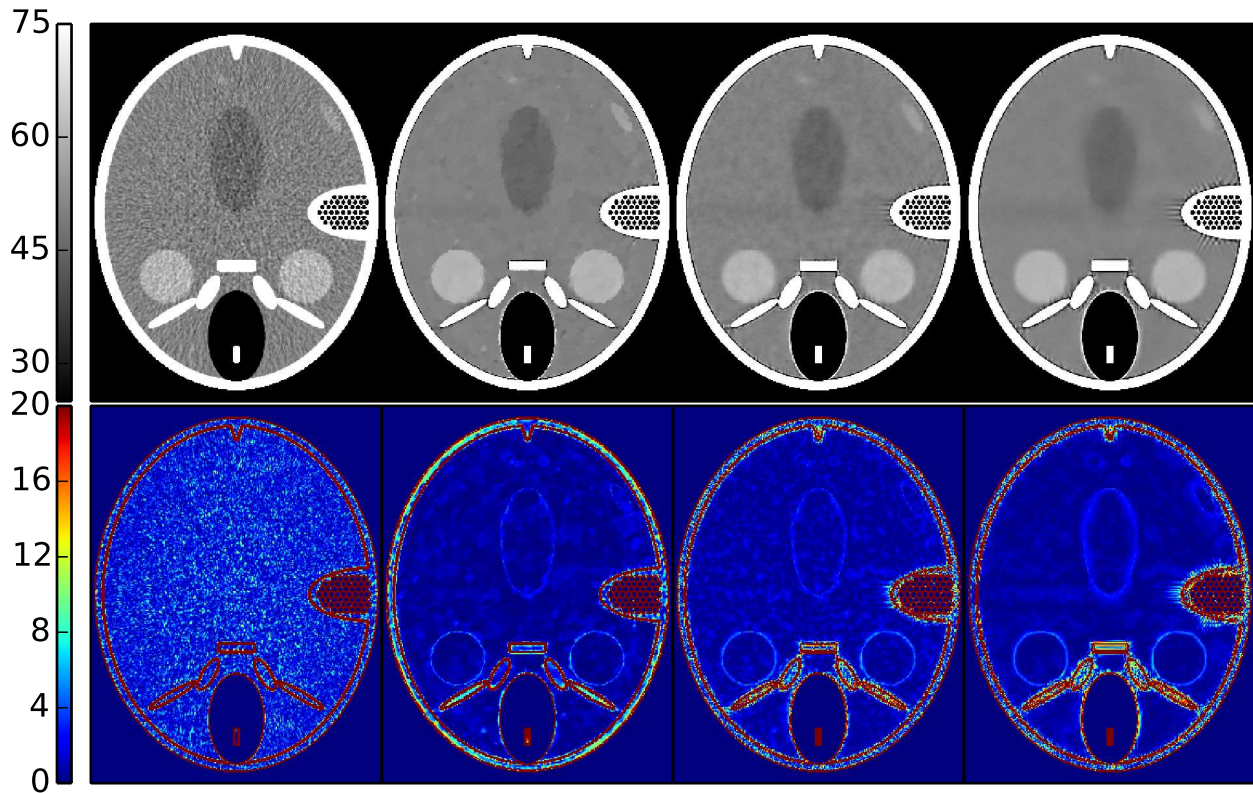


Figure 2. Top row: Reconstruction from low-dose projections. Bottom row: Magnitude of error between reconstruction and  $\bar{x}$ . From left to right: FBP, C-TV-CT, C-AST-CT, C-DL-CT. All units in HU.

- [8] Ravishankar, S. and Bresler, Y., “Learning sparsifying transforms,” *IEEE Trans. Signal Process.* **61**(5), 1072–1086 (2013).
- [9] Ravishankar, S. and Bresler, Y., “Closed-form solutions within sparsifying transform learning,” in [*Acoustics Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*], (2013).
- [10] Ravishankar, S. and Bresler, Y., “Learning doubly sparse transforms for image representation,” in [*2012 19<sup>th</sup> IEEE International Conference on Image Processing*], 685–688, IEEE (Sept. 2012).
- [11] Pratt, W., Kane, J., and Andrews, H., “Hadamard transform image coding,” *Proc. IEEE* **57**(1) (1969).
- [12] Pfister, L., *Tomographic Reconstruction with Adaptive Sparsifying Transforms*, Master’s thesis, University of Illinois at Urbana-Champaign (2013).
- [13] Ravishankar, S. and Bresler, Y., “Sparsifying transform learning for compressed sensing MRI,” in [*International Symposium on Biomedical Imaging*], (2013).
- [14] Ramani, S., Liu, Z., Rosen, J., Nielsen, J., and Fessler, J., “Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods,” *IEEE Trans. Image Process.* , 1–16 (Apr. 2012).
- [15] Niu, T. and Zhu, L., “Accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT: Phantom studies,” *Med. Phys.* **39**(7), 4588 (2012).
- [16] Afonso, M. V., Biucas-Dias, J. M., and Figueiredo, M. A. T., “An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems,” *IEEE Trans. Image Process.* **20**(3), 681–695 (2011).
- [17] Boyd, S. and Vandenberghe, L., [*Convex Optimization*], Cambridge University Press (2004).
- [18] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning* **3**(1), 1–122 (2010).

- [19] Fessler, J. and Booth, S., “Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction,” *IEEE Trans. Image Process.* **8**, 688–99 (Jan. 1999).
- [20] Beck, A. and Teboulle, M., “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences* **2**, 183–202 (Jan. 2009).
- [21] Goldstein, T. and Osher, S., “The split bregman method for l1-regularized problems,” *SIAM Journal on Imaging Sciences* **2**, 323–343 (Jan. 2009).
- [22] Eckstein, J. and Bertsekas, D., “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming* **55**(1-3), 293–318 (1992).
- [23] Basu, S. and De Man, B., [*Branchless distance driven projection and backprojection*], 60650Y–60650Y–8, SPIE - International Society for Optical Engineering (Feb. 2006).
- [24] Wang, J., Li, T., Lu, H., and Liang, Z., “Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography,” *IEEE Trans. Med. Imag.* **25**, 1272–1283 (Oct. 2006).
- [25] Ramani, S. and Fessler, J., “A splitting-based iterative algorithm for accelerated statistical x-ray CT reconstruction,” *IEEE Trans. Med. Imag.* **31**, 677–688 (Mar. 2012).
- [26] Aharon, M., Elad, M., and Bruckstein, A., “K-SVD : An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006).
- [27] Yu, Z., Noo, F., Dennerlein, F., Wunderlich, A., Lauritsch, G., and Hornegger, J., “Simulation tools for two-dimensional experiments in x-ray computed tomography using the FORBILD head phantom,” *Phys. Med. Biol.* **57**, N237–52 (July 2012).