

TOMOGRAPHIC RECONSTRUCTION WITH ADAPTIVE SPARSIFYING TRANSFORMS

Luke Pfister and Yoram Bresler

Department of Electrical and Computer Engineering and the Coordinated Science Laboratory,
University of Illinois, Urbana-Champaign, IL 61801, USA

ABSTRACT

A central problem in computed tomography (CT) imaging is to obtain useful, high-quality images from low-dose measurements. Methods that exploit the sparse representations of tomographic images have long been known to improve the quality of reconstructions from low-dose data. Recent work has shown that sparse representations learned directly from the data can outperform traditional, fixed representations, but are prohibitively expensive for practical use in CT. We propose a new method for tomographic reconstruction from low-dose data by combining the statistically weighted data fidelity term with an adaptive sparsifying transform regularizer. This regularizer can be fit to the data at lower cost than competing methods. Our algorithm alternates between reconstructing the image and learning the sparsifying transform. The Alternating Direction Method of Multipliers technique is used to provide an efficient solution to the statistically weighted minimization problem. Numerical experiments on data from clinical CT reconstructions indicate that adaptive sparsifying transform regularization outperforms synthesis sparsity methods at speeds rivaling total-variation regularization.

Index Terms— Sparsifying transform learning, Sparse representations, CT dose reduction, iterative reconstruction

1. INTRODUCTION

A major challenge in computed tomography (CT) is the reduction of harmful x-ray dose while maintaining the quality of reconstructed images. Key to achieving this goal are advancements in image reconstruction. Unlike the industry-standard filtered backprojection (FBP) algorithm, iterative reconstruction algorithms can produce high-quality images from low-dose data by incorporating detailed models of the data acquisition process, noise statistics, and the signal being reconstructed. These algorithms often reconstruct an image $x \in \mathbb{R}^N$ from (approximately) photon count data, $p \in \mathbb{R}^M$, by solving the penalized weighted-least squares problem [1]

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

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

The data vector $y \in \mathbb{R}^M$ contains the log of measured photon counts p , and the quantity $[Ax]_k$ represents a forward projection of x along the k th ray. The weighting matrix W is diagonal with entries $w_i = \exp(-y_i)$, and can be motivated as a quadratic approximation to the negative log-likelihood of the image given photon counts. The first term in (1) represents a statistically weighted fidelity measure between the data y and the reprojected image Ax .

The regularization functional $J : \mathbb{R}^N \rightarrow \mathbb{R}$ improves the conditioning of (1), and encourages solutions that satisfy a particular signal model. Signal models in which the data is assumed to have a sparse representation have shown enormous success, and this assumption of sparsity is at the heart of popular regularizers such as total-variation (TV). These methods 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.

Edge-preserving and total-variation regularization have been shown to be effective for both low-dose and limited data tomography [2, 3, 4]. These methods generally promote images that are piecewise constant, and can replace complex texture by patchy, uniform regions. More sophisticated regularization using shearlets has been shown to better preserve complex texture at the expense of performance on uniform, flat regions [5].

Recent work has shown the promise of adaptive, rather than fixed, sparse representations. A popular approach is to assume that small, overlapping patches of the signal x can be represented as the linear combination of a few columns of a (possibly overcomplete) dictionary. Several algorithms have been proposed to fit such a dictionary to a given set of data, and algorithms that alternate between learning a dictionary and image reconstruction have been shown to outperform traditional regularization techniques in both limited-angle [6] and low-dose [7, 8] tomography.

The past few years have seen an increase in development of algorithms that adaptively learn analysis operators. Algorithms have been proposed based on modifications of existing dictionary learning algorithms [9], variable-splitting methods [10], and manifold methods [11].

However, both synthesis and analysis learning algorithms suffer from an expensive sparse coding step, which scales poorly with data size, and as a result may be prohibitively

expensive for practical tomographic reconstruction. An alternative approach is to assume that our signal is *approximately* sparse when acted on by a linear transformation Φ . In particular, we assume that our signal satisfies the relationship $\Phi x = z + e$, where z is sparse and called the *transform sparse code*, and the residual e is small. This is known as *transform sparsity* and can be thought of as a generalization of the analysis sparsity model by allowing deviation from exact sparsity in the transform domain. Recently, several algorithms [12, 13, 14] have been developed to adapt sparsifying transforms to data. The permitted deviation from exact transform sparsity results in algorithms that are much faster than competing synthesis dictionary and analysis operator learning algorithms.

In this work, we combine the use of statistical reconstruction techniques with an adaptive sparsifying transform regularizer to produce a computationally efficient algorithm with state of the art performance for low-dose CT imaging. Numerical experiments are performed using reprojected clinical images. The results indicate that adaptive sparsifying transform regularization outperforms synthesis sparsity methods at speeds rivaling total variation regularization.

2. ALGORITHM

Our goal is to reconstruct an image $x \in \mathbb{R}^N$ from data $y \in \mathbb{R}^M$ while jointly finding a sparsifying transform $\Phi \in \mathbb{R}^{k \times k}$ that acts on $\sqrt{k} \times \sqrt{k}$ patches of x . To allow for variable sparsity levels in the transform sparse codes, our regularizer is a modified version of the cost in [12], in which the sparsity constraint has been relaxed to a penalty term. The overall problem, corresponding to the learning of Φ and update of both the image x and transform sparse codes z_j , is written

$$\min_{x, z, \Phi} \frac{1}{2} \|y - Ax\|_W^2 + \frac{\lambda}{2} \sum_j \|\Phi E_j x - z_j\|_2^2 + \lambda (\gamma \|z_j\|_0 + \alpha (\|\Phi\|_F^2 - \log \det \Phi)) \quad (\text{P1})$$

where λ and α are positive scalar parameters and $\|z\|_0$ is the ℓ_0 quasinorm that counts the number of nonzero elements in z . The matrix $E_j \in \mathbb{R}^{k \times N}$ extracts the j -th $\sqrt{k} \times \sqrt{k}$ vectorized patch and removes its mean. As the images arising in CT applications occupy a finite region and are typically surrounded by a region of zero attenuation, we are free to extract patches that wrap around the image boundary.

The second term in (P1) penalizes the sparsification error of patches from the reconstructed image, while the third term encourages sparsity in the transform sparse codes z_j . The Frobenius norm penalty enforces good scaling of the learned transform. The negative log determinant penalty acts as a barrier function to ensure that the learned transform is nonsingular, provided that the algorithm is initialized with a nonsingular starting point. Together, these two penalties ensure that the transform is full-rank and well-conditioned.

As (P1) is highly nonconvex, we employ an alternating minimization scheme. When updating Φ , we fix x and the z_j and solve

$$\arg \min_{\Phi} \sum_j \frac{1}{2} \|\Phi E_j x - z_j\|_2^2 + \alpha (\|\Phi\|_F^2 + \log \det \Phi). \quad (2)$$

The minimization problem (2) can be solved in closed-form [15], or by using a solver such as nonlinear conjugate gradient. For fixed Φ and x , (P1) is rewritten as

$$\arg \min_z \sum_j \frac{1}{2} \|\Phi E_j x - z_j\|_2^2 + \gamma \|z_j\|_0. \quad (3)$$

The solution of (3) is given in closed-form for each z_j by setting to zero all entries with magnitude less than $\sqrt{\gamma}$, an operation called *hard thresholding* and written as $z_j = \mathcal{T}_{\gamma}(\Phi E_j x)$. In practice, we alternate between updating Φ and z a few times before moving to update x . This ensures that Φ is a good sparsifying transform for the current image x .

We now turn to the update of the image x . For fixed Φ and z , (P1) reduces to the weighted least-squares problem

$$\min_x \frac{1}{2} \|y - Ax\|_W^2 + \frac{\lambda}{2} \sum_j \|\Phi E_j x - z_j\|_2^2. \quad (4)$$

Iterative methods are required due to the enormous size of A . However, the large dynamic range in W causes the Hessian $A^T W A + \lambda \sum_j E_j^T \Phi^T \Phi E_j$ to be very poorly conditioned and many iterations are required. Further, the placement of W causes the Hessian to be highly shift-variant and prohibits the use of efficient Fourier preconditioners [16].

To remedy this problem, Ramani & Fessler proposed the use of the Alternating Direction Method of Multipliers (ADMM) algorithm to decouple the influence of W and A , facilitating the use of Fourier preconditioners while also allowing for the efficient use of non-differentiable regularizers such as total-variation [17].

We employ a variation of this technique to accelerate the solution of (4). As the regularization term is differentiable, we only need to split the data fidelity. By not splitting the regularization term, we eliminate the need to store two additional auxiliary variables. This memory savings may prove critical for extending (P1) to 3D tomographic reconstruction.

To utilize the ADMM approach, we first introduce an auxiliary variable u and rewrite the unconstrained problem (4) in the constrained form

$$\min_{x, u} \frac{1}{2} \|y - Ax\|_W^2 + \frac{\lambda}{2} \sum_j \|\Phi E_j x - z_j\|_2^2 \quad \text{s.t. } u = Ax. \quad (5)$$

We next form the augmented Lagrangian function of (5), written in “scaled form” [18] as

$$\mathcal{L}(x, u, \eta) = \frac{1}{2} \|u - y\|_W^2 + \frac{\lambda}{2} \sum_j \|\Phi E_j x - z_j\|_2^2 + \frac{\mu}{2} \|u - Ax - \eta\|_2^2 - \frac{\mu}{2} \|\eta\|^2. \quad (6)$$

The parameter $\mu > 0$ affects the rate of convergence of the algorithm but not the overall solution, and the vector $\eta \in \mathbb{R}^M$ is a scaled version of the Lagrange multiplier of the constraint equation $u = Ax$. Our goal is now to find a saddle point of (6) by optimizing over x , u , and η in an alternating fashion.

At the k th iteration, we solve the following subproblems:

$$x^{k+1} = \arg \min_x \mathcal{L}(x, u^k, \eta^k) \quad (7)$$

$$u^{k+1} = \arg \min_u \mathcal{L}(x^{k+1}, u, \eta^k) \quad (8)$$

$$\eta^{k+1} = \eta^k - (u^{k+1} - Ax^{k+1}). \quad (9)$$

Subproblem (7) is an unweighted least-squares problem in x , with solution found by solving the linear system of equations

$$Hx^{k+1} = \mu A^T(u^k - \eta^k) + \lambda \sum_j E_j^T \Phi^T z_j \quad (10)$$

where $H \triangleq \mu A^T A + \lambda \sum_j E_j^T \Phi^T \Phi E_j$. If E_j extracts all patches from the image x and wraps around the image boundary, then $E_j^T \Phi^T \Phi E_j$ is a shift-invariant operator. As the effect of W is no longer present, H is approximately shift-invariant and well approximated using Fourier preconditioners. The patch mean removal behavior embedded in E_j causes $\lambda \sum_j E_j^T \Phi^T \Phi E_j$ to have high-pass behavior and thus has a nullspace that is disjoint from that of the low-pass $A^T A$, implying that a well-conditioned Φ is an effective regularizer. In practice, we do not require an exact solution to (10) and instead settle for a few preconditioned conjugate-gradient (PCG) iterations.

The u update problem (8) can be exactly solved by

$$u^{k+1} = (W + \mu I)^{-1} (Wy + \mu(Ax^{k+1} + \eta^k)). \quad (11)$$

We note that $W + \mu I$ is a diagonal matrix and is easily inverted.

Finally, the update step for the scaled dual variable η is cheap, requiring only vector additions.

ADMM has changed the difficult strictly convex problem (4) into a series of subproblems that are much easier to solve numerically. These subproblems are iterated until the convergence. However, as ADMM is used to solve a subproblem of the larger optimization problem (P1), there is no need to force ADMM to fully converge. Still, as a fixed number of iterations may not reduce the overall cost function, we adopt the following heuristic strategy: the ADMM iterations are repeated until we observe a decrement in the cost (P1) and then perform three additional iterations. We further impose a minimum number of 10 ADMM iterations.

The overall algorithm, which we call AST-CT, is presented as Algorithm 1. We initialize the sparsifying transform stage with the Hamming-weighted FBP reconstruction of the data y , and take our initial Φ to be a separable approximation of the finite differencing matrix. The auxiliary variables u and η are reinitialized at the start of each ADMM sequence.

Algorithm 1 AST-CT

INPUT: Initial transform Φ , observed data y

OUTPUT: Reconstructed image x

```

1:  $x^0 \leftarrow \text{FBP}(y)$ 
2:  $z_j^0 \leftarrow \mathcal{T}_\gamma(\Phi E_j x^0) \forall j$ 
3: repeat
4:   repeat
5:     Update  $\Phi$  by solving (2)
6:      $z_j^k \leftarrow \mathcal{T}_\gamma(\Phi E_j x) \forall j$ 
7:   until Halting condition
8:    $i \leftarrow 0, u^0 \leftarrow Ax^k, v^0 \leftarrow \vec{0}$ 
9:   repeat
10:    Use PCG to find approximate solution
    of  $H\tilde{x}^{i+1} = \mu A^T(u^i - v^i) + \lambda \sum_j E_j^T \Phi^T z_j^i$ 
11:     $u^{i+1} \leftarrow (W + \mu I)^{-1} (Wy + \mu(A\tilde{x}^{i+1} + v^i))$ 
12:     $v^{i+1} \leftarrow v^i - (u^{i+1} - A\tilde{x}^{i+1})$ 
13:     $i \leftarrow i + 1$ 
14:  until Halting condition
15:   $x^{k+1} \leftarrow \tilde{x}^{i+1}$ 
16: until Halting condition
```

3. EXPERIMENTS

Simulations were implemented in Matlab R2012b on a computer containing an Intel i5-2520m processor and 6GB of RAM. The matrix-vector products Ax and $A^T y$ were performed using a multithreaded C implementation of the distance-driven projector and backprojector, which ensures a matched projector and back-projector pair [19]. Projections were taken with a parallel beam geometry, although the algorithm can easily incorporate fan-beam measurements. In all simulations, the attenuation coefficient of water was taken to be 1.83 mm^{-1} , corresponding to a 80 keV source.

We compare the performance of the algorithms using two metrics. The first is the root mean square error (RMSE), defined for an image with K pixels as $RMSE = \sqrt{\sum_{k=1}^K (x_k - \bar{x}_k)^2 / K}$, where x_k is a pixel from the reconstructed image and \bar{x}_k is the value of a reference image. The second metric is the Structural Similarity Index which has been shown to be consistent with qualitative visual appearance [20]. The SSIM ranges from 0 to 1, with higher values indicating a larger degree of similarity.

We compare the performance of AST-CT against FBP reconstruction and two iterative reconstruction schemes. The first, TV-CT, uses ADMM to split both the data fidelity and regularization term [17]. The second, DL-CT, is a modified version of Algorithm 1 where regularization term in (1) is given as $\sum_j 0.5 \|E_j x - Da_j\|_2^2 + \gamma \|a_j\|_0$, and the ADMM equations are appropriately modified. In particular, the update for a will be a synthesis sparse coding problem, which we solve using the efficient OMPbox Matlab toolbox¹. The dictionary is updated using the K-SVD algorithm. We use the same statistical weighting and preconditioning for AST-CT, DL-CT, and TV-CT. The only difference in the algorithms is

¹Available: <http://www.cs.technion.ac.il/~ronrubin/software.html>

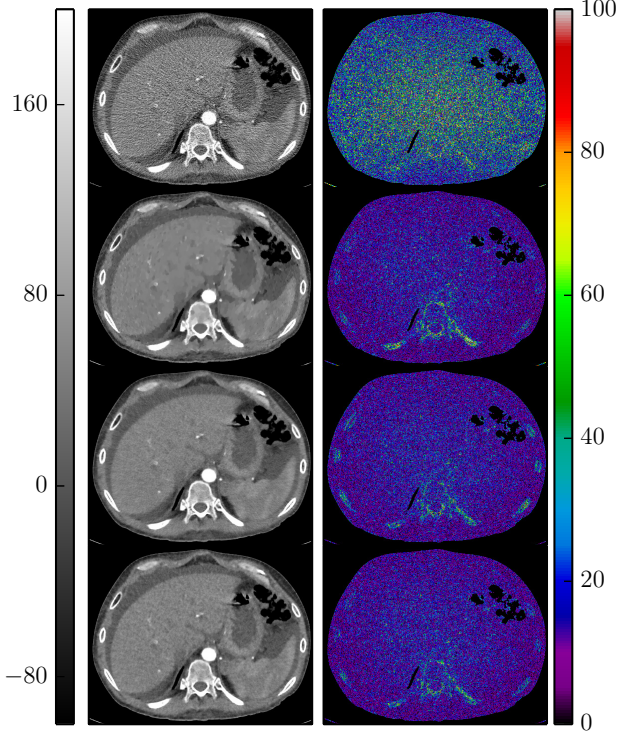


Fig. 1: Left column: Reconstruction from low-dose projections. Right column: Magnitude of error between reconstruction and \bar{x} . From top to bottom: FBP, TV-CT, DL-CT, AST-CT. All units in HU

the choice of regularization function.

For each algorithm, the parameter λ was empirically chosen by sweeping over a large range of values and choosing the parameter that corresponded to the reconstruction with lowest RMSE. For AST-CT and DL-CT, a parameter sweep over γ was also performed. For AST-CT and DL-CT, the ADMM parameter μ was chosen to ensure that H is well conditioned, and chosen according to the strategy in [21] for TV-CT. For AST-CT, the size of the sparsifying transform was chosen to be 64×64 , while the dictionary in DL-CT is 64×121 . Both algorithms use 8×8 patches with maximal overlap. DL-CT is trained using 20% of the available patches, chosen at random in each outer loop iteration. In contrast, AST-CT uses all available patches.

Empirically, both AST-CT and DL-CT benefit from performing multiple transform/dictionary and sparse code updates before updating the image. For AST-CT, we execute 10 sparsifying transform and sparse code update steps before performing a minimum of 10 ADMM iterations. For DL-CT, we perform five dictionary and sparse code updates before executing a minimum of 10 ADMM iterations. For both AST-CT and DL-CT, we repeat the ADMM iterations until we observe a decrement in the cost function. We run each algorithm for a fixed number of overall iterations. For TV-CT, we use a total of 300 iterations, while for AST-CT and DL-CT we use 30 outer-loop iterations.

We evaluate the performance of the algorithm on real

Table 1: RMSE & SSIM of reconstructions from low-dose projections. RMSE given in HU.

	Clinical dose FBP	Low dose FBP	AST-CT	DL-CT	TV-CT
RMSE	19	33	15	16	16
SSIM	0.63	0.40	0.58	0.56	0.55

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

	D/Φ Update	a/z Update	Image Update	Total
FBP	0	0	2.3	2.3
TV-CT	0	0	91.3	91.3
DL-CT	87.5	60.3	85.4	233.3
AST-CT	4.4	0.2	88.4	93.0

tomographic images by synthesizing low-dose data from clinical dose 512×512 pixel CT images of a human abdomen. The clinical data consists of 0.9mm thick overlapping slices with 0.45mm separation between slices. These slices have a noise standard deviation of 21 HU, as estimated over a flat region of the liver. To reduce streaking artifacts, we averaged 5 consecutive of clinical dose images to form a ground truth image \bar{x} with an effective slice thickness of 2.7mm. Projection data was synthesized by reprojecting the image \bar{x} using the distance driven projector. Low dose data was synthesized with k -th detector measurement $y_k = -\log(P(I_0 \exp[-A\bar{x}]_k)/I_0)$, where $P(t)$ is a Poisson random variable with mean t . Our synthetic clinical dose data was formed using $I_0 = 2 \times 10^6$. The resulting Hamming-weighted FBP reconstruction has a noise standard deviation of 20 HU, as estimated over a flat region of the liver, which matches that of the original thin slices. Low dose data is formed using $I_0 = 5.0 \times 10^5$, which corresponds to dose reduction by a factor of 4. Reconstructed images from this data and the magnitude of the difference between the low-dose reconstructions and \bar{x} are shown in Figure 1.

We see that FBP suffers from the usual noise and streaking artifacts as a consequence of the reduction in dose. TV-CT effectively removes the streaking artifacts, but suffers from the patchy, staircase-like regions that are typical of this regularization scheme. The TV-CT reconstruction also has high error and loss of texture in the bone regions, which are not well modelled as piecewise-constant. In contrast, both DL-CT and AST-CT have low error in both the bone and soft tissue regions, with AST-CT being the better of the two reconstructions. The RMSE and SSIM values of the reconstructions are given in Table 1 and show that the adaptive regularization schemes outperform TV, with AST-CT outperforming DL-CT.

Table 2 lists the run time of a single outer loop iteration of each algorithm. Although DL-CT uses only 20% of the available patches, the dictionary and sparse code update steps more than double the computation time of TV-CT. In contrast, AST-CT suffers negligible overhead compared to TV-CT.

4. 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.
- [2] A. Delaney and Y. Bresler, *Efficient edge-preserving regularization for limited-angle tomography*. Institute of Electrical and Electronics Engineers, 1995, pp. 176–179.
- [3] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Jan. 1997.
- [4] G. T. Herman and R. Davidi, "On image reconstruction from a small number of projections," *Inverse problems*, vol. 24, no. 4, pp. 45 011–45 028, Aug. 2008.
- [5] 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.
- [6] 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.
- [7] J. Shtok, M. Elad, and M. Zibulevsky, "Sparsity-based sinogram denoising for low-dose computed tomography," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, no. 1. IEEE, May 2011, pp. 569–572.
- [8] 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.
- [9] 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.
- [10] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies, "Constrained overcomplete analysis operator learning for cosparsity signal modelling," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2341–2355, May 2013.
- [11] S. Hawe, M. Kleinstuber, 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.
- [12] S. Ravishanker and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, 2013.
- [13] S. Ravishanker and Y. Bresler, "Learning overcomplete sparsifying transforms for signal processing," in *Acoustics, Speech and Sig. Proc (ICASSP)*, 2013, pp. 3088–3092.
- [14] S. Ravishanker 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.
- [15] S. Ravishanker and Y. Bresler, "Closed-form solutions within sparsifying transform learning," in *Acoustics Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013.
- [16] 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.
- [17] S. Ramani and J. Fessler, "Statistical x-ray CT reconstruction using a splitting-based iterative algorithm with orthonormal wavelets," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, May 2012, pp. 1008–1011.
- [18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [19] B. De Man and S. Basu, "Distance-driven projection and backprojection," *2002 IEEE Nuclear Science Symposium Conference Record*, vol. 3, pp. 1477–1480, 2003.
- [20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–12, Apr. 2004.
- [21] 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.