# ADAPTIVE DICTIONARY LEARNING IN SPARSE GRADIENT DOMAIN FOR CT RECONSTRUCTION

Qiegen Liu[1], Minghui Zhang[1], and Jun Zhao[2,*] *Member, IEEE*

[1]Department of Electronic Information Engineering, Nanchang University, Nanchang 330031, China
[2]School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

## ABSTRACT

Image recovery from undersampled data has always been a challenging and fascinating task due to its implicit ill-posed nature and significance accompanied with the emerging compressed sensing (CS) theory. This paper proposes a novel Gradient based Dictionary Learning method for CT image Reconstruction (GradDL-CT), which alleviates the drawback of the popular total variation (TV) regularization by employing dictionary learning technique. Specifically, we firstly train dictionaries from the horizontal and vertical gradients of the image respectively, and then reconstruct the desired image using the sparse representations of both derivatives, exploiting gradient magnitude image sparsity for reduction in the number of projections or the X-ray dose. Preliminary results on phantom and real CT images demonstrate that the proposed method can efficiently recover images and presents advantages over the current state-of-the-art reconstruction approaches.

***Index Terms*****—**CT reconstruction; dictionary learning; sparse representation; gradient magnitude image; alternating direction method

## 1. INTRODUCTION

Computed Tomography (CT) is a technology that obtains the tomogram of the observed object. In the biomedical applications, lower radiation dose have been constantly pursued. In order to shorten scanning time and reduce radiation dose, the scanning views are within an angular range that is often both limited and sparsely sampled. Until quite recently, the traditional filtered back projection (FBP) method has been commonly exploited in CT reconstruction. However, using FBP method may prolong scanning time and cumulate with a high dose of radiation being harmful to human body. Otherwise, with fewer projections, it will produce reconstructions with serious blurring and artifacts [1].

The compressed sensing (CS) theory as a fundamental and newly developed methodology has gained much attention [2]. Compared to traditional algebraic reconstruction technique (ART) algorithms, CS-based iterative algorithms usually minimize the L1-norm of the sparse image in certain domain/transform as the constraint factor for the iteration procedure, to reconstruct images from substantially reduced projection data and to reduce the impact of artifacts introduced into the CT reconstructed image due to insufficient projections. Recent work in iterative image reconstruction in CT has focused on some forms of total variation (TV) minimization [3], and one of the motivations for employing TV minimization is exploiting sparsity in the gradient magnitude image to reduce sampling requirements.

Regretfully, TV prior prefers cartoon-like images which are piecewise constant, thus it would not be an ideal option. There are two research directions for alleviating the drawback. Some authors turn to the non-convex penalty alternative to the L1-norm. Ramirez-Giraldo presented the non-convex prior image constrained compressed sensing algorithm for dynamic CT [4]. Some researchers devoted to integrating the nonlocal and similarity property into the TV model [5-8]. For instance, Lou et al. [5] incorporated a semi-nonlocal priority into the TV-minimization (NLTV) to enhance the tomographic reconstruction; some methods that utilize sparsity prior under adaptive transform/dictionary were developed [7-8]. Xu et al. [8] applied synthesis dictionary learning as a regularizer for CT reconstruction from low-dose projections.

Motivated by our recently work on dictionary learning (DL) in gradient domain for CS [9], we present a new method for CT image reconstruction from few-projection data. The method combines an iterative reconstruction framework with an adaptive sparsifying transform penalty. An alternating minimization approach is used to jointly reconstruct the image while learning a sparsifying transform adapted to the gradient magnitude of particular image being reconstructed. The alternating direction method (ADM) is used to provide a computationally efficient solution to the minimization problem. Numerical experiments performed on phantom data and clinical CT images indicate its superior to other existing methods.

*Correspondence to: Jun Zhao (junzhao@sjtu.edu.cn). This work was supported in part by the NSFC under Nos. 61362001, 61262084, 81371634, 20132BAB211030 and 973 Program of China 2010CB834302.

## 2. ALGORITHM GRADDL-CT

In this section, after the traditional CT reconstruction methods were briefly reviewed, the proposed GradDL-CT was detailed derived by interpreting the proposed model and then we employed alternating algorithm to tackle with the resulting minimization.

### 2.1. Reviews of conventional CT reconstruction

Let vector $u$ denote one cross-section of patient anatomical information, and $f$ represent another vector on x-ray detectors in parallel beam geometry. These two are related by $Pu = f$ , where $P$ represents an x-ray projection matrix. A CT reconstruction problem is formulated as the retrieval of the vector $u$ based on the observation $f$ given the projection matrix $P$ .

The so-called few-projection CT reconstruction problem is well known to be highly under-determined in that there are infinitely many solution vectors $u$ satisfying the equation $Pu = f$ . In order to single out an ideal CT image $u$ , additional information needs to be imposed properly. For this purpose, regularization models are usually used to reconstruct a desirable CT image. As such, one considers the optimization problem

$$u = \arg\min_{u} J(u) + \nu_1 \|Pu - f\|_2^2 \tag{1}$$

where the first term $J(u)$ is a regularization term. The second term ensures the consistency between the reconstructed CT image $u$ and the observation $f$ . $\nu_1 > 0$ is a penalty parameter. The well known example is the TV regularization $J(u) = \|\nabla u\|_1$ [3]. Recently, more sophisticated regularization penalties such as wave-lets [10], tight frames (TF) [11], NLTV [6], and DL [12] have been proposed.

### 2.2. Algorithm GradDL-CT

Motivated by the strong ability of utilizing adaptive dictionary to representation patch structure, and to reconstruct image using dictionaries learned from the gradients [12, 9], we propose a new model as follows:

$$\begin{aligned} &\min_{u, D^{(i)}, \Gamma^{(i)}} \left\{ \sum_{i=1}^{2} \sum_{l} \left\| D^{(i)} \alpha_l^{(i)} - R_l(\nabla^{(i)} u) \right\|_2^2 + \nu_1 \|Pu - f\|^2 \right\} \\ &s.t. \quad \left\| \alpha_l^{(i)} \right\|_0 \le T_0, \forall l, i \end{aligned} \tag{2}$$

where $\nabla^{(1)}$ and $\nabla^{(2)}$ denote the difference operators in horizontal and vertical directions. The first term in the cost function captures the sparse prior of the gradient-image patches with respect to dictionaries $\{D^{(i)}, i = 1,2\}$ . $R_l(\nabla^{(i)}u)$ denotes a vectored form of the $\sqrt{M} \times \sqrt{M}$ patch extracted from the image $\nabla^{(i)}u$ of size $\sqrt{N} \times \sqrt{N}$ and $\Gamma = [\alpha_1, \alpha_1, \cdots, \alpha_L]$ denotes the sparse coefficient matrix of these patches. $T_0$ controls the sparsity of the patch representation.

The proposed model in (2) can be viewed as an extension and adaptive strategy of the model in (1). Specifically, the former can be degraded to the latter by setting the dictionary to be fixed basis (i.e. identity matrix $D = I_{M \times M}$ ) and relaxing $l_0$ quasi-norm to be $l_1$ -norm. Therefore, with Lagrangian formulation, Eq. (2) has the unconstrained version:

$$\min_{u, \Gamma^{(i)}} \left\{ \sum_{i=1}^{2} (\mu_1 \sum_{l} \left\| \alpha_l^{(i)} \right\|_1 + \mu_3 \left\| I \alpha_l^{(i)} - R_l(\nabla^{(i)} u) \right\|_2^2) + \|Pu - f\|^2 \right\} \tag{3}$$

Eq. (3) approximates the model in (1) with TV regularization as $\mu_3 \to \infty$ . In this case, by restricting the basis to be fixed, the local-adaptive patch based sparsity is degraded to the global TV-based image sparsity. Hence, it can be expected that the proposed model (2) has the potential to successfully recover image with complex structure and details by adaptively learning basis/atom from the gradient images to match local feature, while TV model cannot when the measurements are highly undersampled or contaminated with heavy noise.

### 2.3. ADM-based solver

Recently, splitting methods such as ADM or Split-Bregman have become popular for solve optimization problem [9, 10, 13]. Here we also resort to the splitting technique. Specifically, an augmented Lagrangian/ Bregman iterative technique is employed and an algorithm called GradDL-CT is developed. The algorithm alternately updates sparse representation of the gradient image patches, recovers the horizontal and vertical gradients, and estimates the desired image from both gradients.

The problem in (2) can be rewritten as follows by introducing auxiliary variables $w^{(i)}$ , $i = 1,2$ ,

$$\begin{aligned} &\min_{u, w, D^{(i)}, \Gamma^{(i)}} \left\{ \sum_{i=1}^{2} \sum_{l} \left\| D^{(i)} \alpha_l^{(i)} - R_l(w^{(i)}) \right\|_2^2 + \nu_1 \|Pu - f\|^2 \right\} \\ &s.t. \quad \left\| \alpha_l^{(i)} \right\|_0 \le T_0, \forall l, i; \; w^{(i)} = \nabla^{(i)} u, \; \forall i; \end{aligned} \tag{4}$$

Then, by applying the Bregman technique and denoting $\nabla = \begin{bmatrix} \nabla^{(1)} \\ \nabla^{(2)} \end{bmatrix}$, $b = \begin{bmatrix} b^{(1)} \\ b^{(2)} \end{bmatrix}$ and $w = \begin{bmatrix} w^{(1)} \\ w^{(2)} \end{bmatrix}$, we can obtain a sequence of constrained subproblems as follows:

$$\begin{aligned} &\{u^{k+1}, w^{k+1}, (D^{(i)})^{k+1}, (\alpha_l^{(i)})^{k+1}\} \\ &= \arg\min_{u,w,D,\Gamma} \left\{ \begin{aligned} &\sum_{i=1}^{2} \sum_{l} \left\| D^{(i)} \alpha_l^{(i)} - R_l(w^{(i)}) \right\|_2^2 + \nu_1 \|Pu - f\|^2 + \nu_2 \left\| b^k + \nabla u - w \right\|_2^2 \\ &s.t. \quad \left\| \alpha_l^{(i)} \right\|_0 \le T_0, \forall l, i \end{aligned} \right. \end{aligned} \tag{5}$$

$$b^{k+1} = b^k + \nabla u^{k+1} - w^{k+1}$$
(6)

where $\nu_2$ denotes the positive penalty parameter. In order to address the minimization of Eq. (5) with respect to $u$ , $w$ and the sparse representation variables $D$ and $\alpha$ , the ADM was used in our work. This technique carries out approximation via alternating minimization with respect to one variable while keeping other variables fixed.

*Updating the solution* $u$

At the $k$ -th iteration, $w$ , $D^{(i)}$ and $\alpha_l^{(i)}$ are fixed and the objective function is minimized over $u$

$$u^{k+1} = \arg\min_u \left\{ \nu_1 \left\| Pu - f \right\|_2^2 + \nu_2 \left\| b^k + \nabla u - w^k \right\|_2^2 \right\} \tag{7}$$

This is a minimization problem of a quadratic of variable $u$ and can be solved by simple gradient descent or conjugate gradient algorithm. We follow the implementation as that in ref. [14].

*Updating the gradient image variables* $w^{(i)}, i = 1,2$

The minimization in Eq. (5) with respect to $w^{(1)}$ and $w^{(2)}$ is decoupled, and then can be solved separately. It yields:

$$(w^{(i)})^{k+1} = \arg\min_{w^{(i)}} \left\{ \begin{array}{l} \sum_l \left\| (D^{(i)})^k (\alpha_l^{(i)})^k - R_l(w^{(i)}) \right\|_2^2 \\ + \nu_2 \left\| (b^{(i)})^k + (\nabla^{(i)} u)^{k+1} - w^{(i)} \right\|_2^2 \end{array} \right\} \tag{8}$$

Since this is also a least squares problem, we can follow the implementation as that in ref. [9].

*Sparse representation for gradient patches with respect to variables* $D^{(i)}$ *and* $\alpha_l^{(i)}$ , $l = 1, 2, \cdots, L$

The minimization (5) with respect to dictionary and coefficient variables of the gradient images in horizontal and vertical directions is also decoupled thus they can be solved separately. It yields:

$$\{(D^{(i)})^{k+1}, (\alpha_l^{(i)})^{k+1}\} = \arg\min_{D^{(i)}, \Gamma^{(i)}} \left\{ \begin{array}{l} \sum_l \left\| D^{(i)} \alpha_l^{(i)} - R_l (w^{(i)})^{k+1} \right\|_2^2 \\ s.t. \ \left\| \alpha_l^{(i)} \right\|_0 \le T_0, \forall l \end{array} \right., \quad i = 1,2 \tag{9}$$

The strategy to solve (9) is alternatively updating the dictionary $D^{(i)}$ and coefficient matrix $\alpha_l^{(i)}$ , the same as that used in K-SVD algorithm [7, 12]. At sparse coding step, seeking the solution of Eq. (9) with respect to a fixed dictionary $D^{(i)}$ is achieved by greedy algorithm- orthogonal matching pursuit (OMP). While at the dictionary updating step, the columns of the dictionary are updated sequentially one at a time using the singular value decomposition (SVD) to minimize the approximation error.

## 3. EXPERIMENTAL RESULTS

To evaluate the proposed GradDL-CT, Digital NCAT phantom and real clinical images with size $256 \times 256$ reported in ref. [6] were used in our experiment, which are shown in Fig. 1(a)(b). Parallel beams with the beam of 363 X-rays were simulated to generate the complete projection data. The projections were uniformly distributed in the angle range of $0 \sim 180$ degrees, and the sparse projections were simulated by generating the specified number of uniformly distributed projections from the test images. Our proposed method GradDL-CT was compared with the TV-CT proposed by Sidky et al. [3] and DL-CT developed by Liao et al. [7]. The setting of these parameters is very similar to that in refs. [7, 9]. We do not discuss here due to the paper limit. The quality of the reconstruction is quantified using the Mean Square Error (MSE).

Fig. 1(c)(d) list the MSE values of the three methods versus sparse projections ranging from 40 to 150. It can be observed that our method constantly outperforms the other two approaches at all sparse levels. Compared to the TV-CT method, GradDL-CT substantially improves the quality of the reconstructed image when the projections is sparse, indicating that enhancing the sparse representation of gradient domain. On the other hand, compared to the DL-CT method, GradDL-CT achieves comparable results when the projections is sparse, and substantially improves the quality of the reconstructed image when the number of the project-ions is relatively high. Specially, as shown in Fig. 1(d), the MSE value of the GradDL-CT image reconstructed from 70 projections is comparable to the value of the DL-CT result from 100 projections and TV-CT from 120 projections.

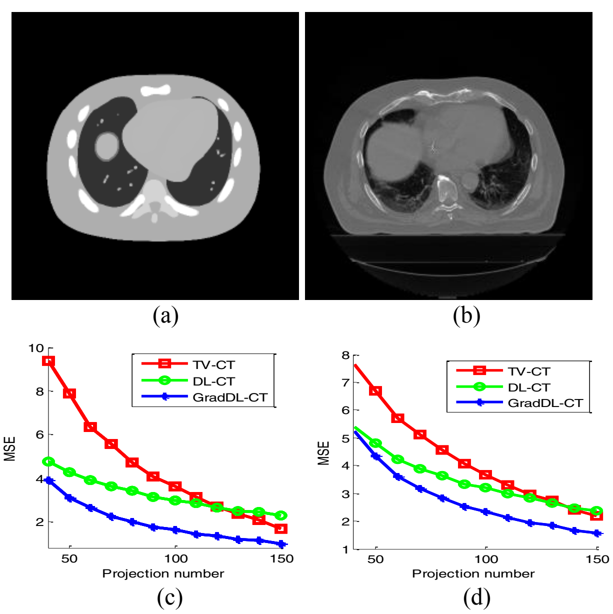


Fig. 1. Reconstruction errors of TV-CT, DL-CT and GradDL-CT versus the number of projections. Result (c) corresponds to image (a), and (d) corresponds to image (b).

Figs. 2 and 3 show comparisons of results generated by the three methods. Regardless of the phantom with simple anatomical structure or the real image containing detailed structures, it can be observed from the error images in the second row, GradDL-CT is better in reducing aliasing artifacts and maintaining fine details than TV-CT and DL-CT. The reconstructions with GradDL-CT method exhibit higher resolution than the other two methods.

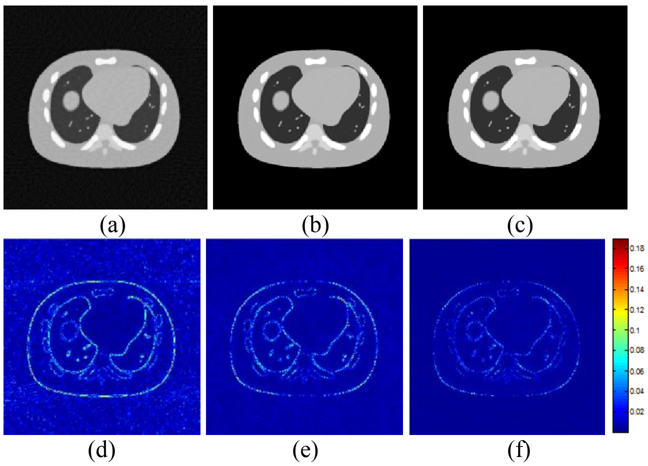


Fig. 2. Reconstruction of NCAT phantom by TV-CT (a), DL-CT (b) and GradDL-CT (c), in which the number of projections was 80. (d), (e) and (f) are error images for (a), (b) and (c), respectively.

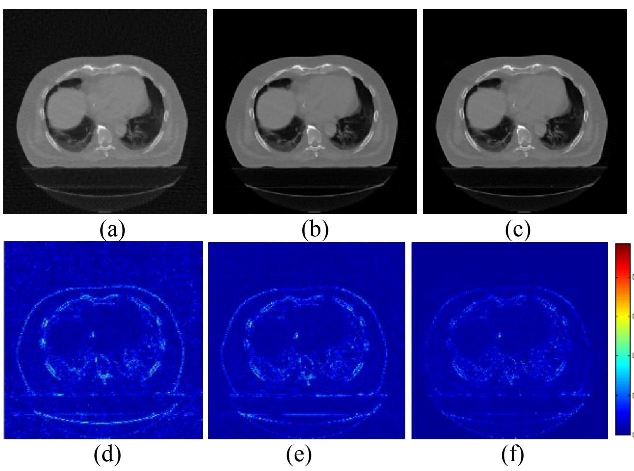


Fig. 3. Reconstruction of a real clinical image by TV-CT (a), DL-CT (b) and GradDL-CT (c). The number of projections was 80. (d), (e) and (f) are error images for (a), (b) and (c), respectively.

## 4. CONCLUSION

We present a new gradient-domain dictionary learning method for CT reconstruction from few-view data. The proposed GradDL-CT method improves the TV-CT model to handle with not only piecewise constant images but also texture-rich images. It enables local features existed in the gradient images to be captured effectively. Preliminary experimental results on both simulated and real images have consistently demonstrated the superior performance of the algorithm. We will extend the proposed framework to deal with more complicated statistical model including weighted norm on the data fidelity term like refs. [8, 13].

## 5. REFERENCES


1. H. Fang, Q.B. Zhang, W. Sui, “Method of image reconstruction based on very sparse random projections. *Comput. Eng. Appl.*, vol. 42, no 22, pp. 25–27, 2007.
2. E. J. CandÃ¨s , J. K. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489 -509, 2006.
3. E. Y. Sidky, C.-M. Kao, and X. Pan, “Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT,” *J. X-ray Sci. Tech.*, vol. 14, pp. 119–139, 2006.
4. J. Ramirez-Giraldo, J. Trzasko, S. Leng, L. Yu, A. Manduca, and C. H. McCollough, “Nonconvex prior image constrained compressed sensing (NCPICCS): Theory and simulations on perfusion CT”, *Med. Phys.* vol. 38, pp. 2157-2167, 2011.
5. Y. Lou, X. Zhang, S. Osher and A. Bertozzi, “Image Recovery via Nonlocal Operators,” *J. of Sci. Comput.*, vol. 42, no. 2, pp. 185-197, 2010.
6. X. Jia, Y. Lou, B. Dong, et al., "4D computed tomography reconstruction from few-projection data via temporal non-local regularization," *Med Image Comput Comput Assist Interv*., vol. 13, pp. 143-150, 2010.
7. H. Y. Liao and G. Sapiro, “Sparse representations for limited data tomography,” in *IEEE International Symposium on Biomedical Imaging*, pp. 1375–1378, 2008
8. Q. Xu, H. Yu, X. Mou, D. Zhang, J. Hsieh, and G. Wang, “Low-dose X-ray CT reconstruction via dictionary learning,” *IEEE Trans on Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, 2012.
9. Q. Liu, S. Wang, L. Ying , X. Peng, Y. Zhu, D. Liang, “Adaptive dictionary learning in sparse gradient domain for image recovery,” *IEEE Trans on Image Process.*, vol. 22, no. 12, pp. 4652-4663, 2013.
10. S. Ramani and J. Fessler, “Statistical x-ray CT reconstruction using a splitting-based iterative algorithm with orthonormal wavelets,” in 9th *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1008–1011, 2012.
11. B. Zhao, H. Gao, H. Ding, and S. Molloi, “Tight-frame based iterative image reconstruction for spectral breast CT, *Med. Phys.* vol. 40, no. 3, 031905, 2013.
12. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311-4322, 2006.
13. S. Ramani, J. A. Fessler, “A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction,” *IEEE Trans Med Imag.,* vol. 31, no. 3, pp. 677-688, 2012.
14. Z. Tian, X. Jia, B. Dong, et al., "Low-dose 4DCT reconstruction via temporal nonlocal means," *Med. Phys.*, vol. 38, pp. 1359-1365, 2011.