梁文轩, 胡广书. 关于不充分投影数据下的二维扇形束 CT 重建[J]. CT 理论与应用研究, 2010, 19(3): 1-12.
Liang WX, Hu GS. On the 2D fan-beam CT reconstruction from insufficient projection data[J]. CT Theory and Applications, 2010, 19(3): 1-12.

On the 2D Fan-beam CT Reconstruction from Insufficient Projection Data

LIANG Wen-xuan, HU Guang-shu✉

(Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China)

Abstract: CT reconstruction from insufficient data is important in many medical and other applications. Recently to solve this under-determined problem, the Total Variation minimization method has regained popularity due to the newly-emerging compressed sensing theory. In this paper TV-minimization is extended to a more general framework for CT reconstuction from insufficient data. The framework proposed is also based on the compressed sensing theory, but is more flexible and amenable to various applications. Analysis and simulations are conducted for an initial special case and promising results are achieved.

Key words: CT reconstruction; compressed sensing; few-view reconstruction; limited-range reconstruction

Article ID: 1004-4140（2010）03-0001-12 **CLC number:** TP 301.6 **Document code:** A

1 Introduction and motivation

Among the various forms of Computed Tomography (CT), one of the important problems on image formation concerns on reconstruction from insufficient projection data. Situation of insufficient data arises frequently out of the constraints from the imaging hardware and scanning geometry (as in tomosynthesis and retrospectively gated cardiac micro-CT[1]), or the purpose to reduce the radiation exposure imposed on the sample scanned[2].

Generally there exist two main categories of insufficient data problems. One class can be referred to as few-view sampling, where the projection data are obtained at sparse angles that are uniformly distributed. The other class is named limited-range sampling, where the projection data are gathered in some limited-angular range. In both situations, since the Nyquist sampling law is not satisfied, the application of standard filtered back-projection (FBP) algorithm will result in perceptible streaks or artifacts in the result reconstructed.

Mathematically the formulation of projection data can be modeled as line integrals of the tomographic image. Since in practice we are concerned with discrete tomography, i.e., what we want to reconstruct is an image $\boldsymbol{f}$ composed of N by N pixels, the reconstruction problem reduces to how to invert the following linear transform (or solve the following linear equation)

$$\boldsymbol{g} = \boldsymbol{M}\boldsymbol{f} \tag{1}$$

where the system matrix $\boldsymbol{M}$ represents the contributions of each pixel in $\boldsymbol{f}$ to the projection data and is usually derived via the ray-driven model. Since the linear equation (1) is highly

Received date: 2009-05-31.
Foundation item: National Basic Research Program of China (973) under Grant No.2006CB705700; Texas Instruments Innovation Funds.

under-determined, there are infinite solutions satisfying the equation. Thus the reconstruction procedure actually boils down to how to search for the tomographic image we want from all the feasible solutions.

1.1 TV-minimization

To complete such a search in the solution space, other prior information or model assumption should be imposed. And usually some iterative optimization algorithm has to be adopted. One popular assumption which has been widely used is that in medical applications, tomographic images are close to piecewise constant, so the total variation (TV) of the solution should be quite small. Here the TV of a discrete 2D image $\boldsymbol{f}_{x,y}$ is defined as

$$\mathrm{TV}(\boldsymbol{f}) = \sum_{x,y}\left(\left(\boldsymbol{f}_{x,y} - \boldsymbol{f}_{x-1,y}\right)^2 + \left(\boldsymbol{f}_{x,y} - \boldsymbol{f}_{x,y-1}\right)^2\right)^{\frac{1}{2}} \tag{2}$$

By minimizing the TV of the solution, one can pick up an almost piecewise constant image from all the feasible solutions[3-4]. Traditionally such solving methods were implemented via TV regularized iterative optimization algorithm. And so there is a trade-off between the data fidelity and the TV regularity, i.e., to obtain an image with smaller TV, one usually has to violate the data fidelity constraints more. Also there is no profound theory analysis concerning the existence and uniqueness of the feasible solution.

Recently, the emergence of the compressed sensing theory brings new insight into the reason why the TV-minimization method could succeed. Candes et al.[5] proved that with overwhelming probability a discrete signal $\boldsymbol{f} \in C^N$ can be exactly reconstructed from highly incomplete frequency information provided that the support of $\boldsymbol{f}$ is sparse enough. For a 2D image $\boldsymbol{f}_{x,y}$ we can actually derive a complex gradient signal $\nabla\boldsymbol{f}_{x,y}$ whose real and imaginary components are the two directional derivatives along the $\boldsymbol{x}$ and $\boldsymbol{y}$ directions respectively, i.e.

$$\nabla\boldsymbol{f}_{x,y} = \left(\boldsymbol{f}_{x,y} - \boldsymbol{f}_{x-1,y}\right) + j\cdot\left(\boldsymbol{f}_{x,y} - \boldsymbol{f}_{x,y-1}\right) \tag{3}$$

Then it is easy to see that $\mathrm{TV}(\boldsymbol{f})$ is the l_1 norm of signal $\nabla\boldsymbol{f}_{x,y}$. Also note that the frequency information of $\nabla\boldsymbol{f}_{x,y}$ can be inferred from that of signal $\boldsymbol{f}_{x,y}$. So if the gradient signal of an image is sparse enough, or equivalently if the image $\boldsymbol{f}_{x,y}$ is piecewise constant enough, it can also be accurately reconstructed from highly incomplete frequency information. In the simulations in Candes et al[5], the Shepp-Logan phantom can be accurately reconstructed from the parallel-beam projections gathered under as few as 22 angles.

Partly inspired by [5], TV minimization has regained popularity in tomographic reconstruction. Pan et al[6-7] extends similar methods to the fan-beam 2D CT and 3D cone-beam CT reconstruction, where the formulation is to minimize TV of the object image while conforming to the projection fidelity and image positivity as,

$$\min \mathrm{TV}(\boldsymbol{f}) \quad s.t. \quad \begin{cases} \|\boldsymbol{M}\boldsymbol{f} - \boldsymbol{g}\|^2 \le \varepsilon \\ \boldsymbol{f} \ge \boldsymbol{0} \end{cases} \tag{4}$$

Note that the second constraint should be understood to hold component-wise. To solve the optimization problem, the authors proposed an solving procedure which alternated between the Algebraic Reconstruction Technique (ART) iteration and the gradient descent iteration of the TV

function. In Song[1], an equivalent TV-regularized linear least squares formulation is adopted for the retrospectively gated cardiac micro-CT, where the insufficient projection data also has irregular angle distribution. The non-linear conjugate gradient (CG) method is used to solve the optimization problem. In Duan et al[8], an iterative reconstruction-reprojection algorithm is integrated with minimizing TV to solve a similar optimization problem. Despite the different solving approaches, the kernel formulations in [1] and [6]～[8] are actually the same. The underlying principle actually boils down to minimizing the l_1 norm of the gradient signal so as to restore an image with sparse gradient, which matches the philosophy of compressed sensing.

However, the gradient of the real-world CT images is hardly sparse. Though tomographic images are relatively constant within some volumes like in an organ, abundant details in the image usually lead to dense gradient support. That partly explains why when applying TV-minimization method to practical CT images, usually more projections that what is suggested in simulation are necessary [1-2, 9]. Moreover, to the best knowledge of the authors, apart from the original theorems in [5] no other explicit theoretical analysis or simulations have been done to investigate how sparse the gradient of the image should be so that it could be restored by how much projection data.

1.2 Compressed sensing and RIP

Starting from [5], the compressed sensing theory has been extended from recovery from incomplete frequency information to other more general linear transforms. In [10] and [11], a series of powerful results have been developed about the exact recovery conditions of a sparse signal $x \in R^m$ from a limited number of observations. We will introduce some of the main results in this section.

Consider recovering an unknown sparse signal $\boldsymbol{x}_0 \in R^m$ whose support $T_0 = \{t \mid \boldsymbol{x}_0(t) \neq 0\}$ is assumed to have small cardinality. And all we know about $\boldsymbol{x}_0$ are n linear measurements which are denoted by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ where $\boldsymbol{A}$ is the linear sensing matrix. Here the linear equation system is under-determined since we have fewer observations than the unknowns, i.e. $n < m$.

It turns out that if the signal $\boldsymbol{x}_0$ is sparse enough, it can be recovered by solving the convex program below

$$\min \|\boldsymbol{x}\|_1 \quad s.t. \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{5}$$

provided that the matrix $\boldsymbol{A}$ obeys the “restricted isometry property” (RIP) .

To introduce the notion of RIP, let $\boldsymbol{A}_T$, $T \subseteq \{1, 2, \cdots, m\}$ denote the sub-matrix formed by extracting $|T|$ columns from $\boldsymbol{A}$ corresponding to the indices contained in T. Then the S-restricted isometry constant δ_S of $\boldsymbol{A}$ is defined [10] to be the smallest quantity so that

$$(1-\delta_S)\|\boldsymbol{c}\|^2 \leq \|\boldsymbol{A}_T\boldsymbol{c}\|^2 \leq (1+\delta_S)\|\boldsymbol{c}\|^2 \tag{6}$$

holds for all subsets T satisfying $|T| \leq S$. Intuitively, this S-restricted isometry constant reflects how well a subset of the columns with cardinality less than S behaves like an orthonormal system, i.e., all the columns should be nearly unit-length and as incoherent as possible.

It was proved[10] that if the RIP of the sensing matrix $\boldsymbol{A}$ is good enough so that

$$\delta_S + \delta_{2S} + \delta_{3S} < 1 \tag{7}$$

then solving (5) retrieves exactly any sparse signal x_0 whose support size satisfies $|T| \le S$.

In practice what we can usually obtain is limited to somewhat noisy measurements as $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e}$, where $\boldsymbol{e}$ is the unknown noise assumed to be bounded by $\|\boldsymbol{e}\|_2 < \varepsilon$. To retrieve the original signal, we can alternatively solve the following program,

$$\min \|\boldsymbol{x}\|_1 \quad s.t. \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2 < \varepsilon \tag{8}$$

Theorem 1.1 Assume that for matrix $\boldsymbol{A}$ there exists S such that $\delta_{3S} + \delta_{4S} < 2$. Then for any signal $\boldsymbol{x}_0$ supported on T_0 such that $|T_0| < S$ and any perturbation $\boldsymbol{e}$ satisfying $\|\boldsymbol{e}\|_2 \le \varepsilon$, the solution $\boldsymbol{x}^{\#}$ to (8) obeys

$$\|\boldsymbol{x}^{\#} - \boldsymbol{x}_0\|_2 \le C_S \cdot \varepsilon \tag{9}$$

where the constant C_S depends only on δ_{4S} and is well behaved for δ_{4S} with reasonable value.

Theorem 1.1, which is proved in [11], ensures the stableness of the solution of (5) in the sense that small disturbance in the observation only results in small changes in the signal recovered. So the l_1 minimization method can be used robustly for the imperfect measurements.

1.3 This paper

Since the success of TV in CT reconstruction from insufficient data is just one special case indicated by the compressed sensing theory, we want to investigate a more general form of this problem. Unlike some special case in some specific application [12], the CT images in reality are hardly sparse; however, it is possible to find some (over-complete) basis $\boldsymbol{W}$ on which the expansion coefficients of the image $\boldsymbol{x}$ is sparse, i.e. $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{\alpha}$ with sparse $\boldsymbol{\alpha}$. So (5) turns into

$$\min \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \boldsymbol{A}\boldsymbol{W}\boldsymbol{\alpha} = \boldsymbol{y} \tag{10}$$

And accordingly problem (8) turns into

$$\min \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \|\boldsymbol{A}\boldsymbol{W}\boldsymbol{\alpha} - \boldsymbol{y}\|_2 \le \varepsilon \tag{11}$$

There exist many possible bases on which the expansion coefficients of $\boldsymbol{x}$ can be sparse. The selection of appropriate basis depends highly on specific practical application. For example, the basis can be derived from orthogonal wavelet transform or some over-complete dictionary. Research in compressed sensing theory also suggests the combination of two orthogonal bases like discrete cosine transform (DCT) and wavelets. How to construct or learn a feasible dictionary to approximate the signals of interest sparsely and effectively is still an open problem, and is beyond the scope of this paper.

For simplicity, in this paper we will take $\boldsymbol{W}$ to be the 3-level Haar wavelet basis. We will investigate the sparse extent of the wavelet coefficients of the images that can be recovered accurately by solving (11). Such investigation, even just numerically, can give us a quantitative comprehension about the reconstruction ability we can hope for and the applicability of the compressed sensing theory in the CT reconstruction. Similar ideas and preliminary study were also represented in [13]. However, our paper is deeper and more comprehensive in various aspects. For example, we test the recovery performance of (11) in both few-view and limited-range cases. Also the solving method we adopt is more amenable to large-scale practical CT reconstruction.

The rest of this paper is structured as follows. In Section 2, we will detail on the analysis of the RIP of the combined sensing matrix. Also we will describe the solving method adopted. Section 3 presents the simulation results and relative analysis. Conclusions and future work are discussed in Section 4.

2 RIP analysis and reconstruction methods

In this paper, we focus on the 2D fan-beam CT reconstruction (though the method proposed can be extended to 3D case). The system matrix $\boldsymbol{M}$ in (1) is constructed via the well-known ray-driven projection model.

Since some extent of noise is inevitable in practice, formulation (11) is adopted. Replacing the matrix $\boldsymbol{A}$ in (11) by the system matrix $\boldsymbol{M}$, we rewrite (11) as below.

$$\min \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \|\boldsymbol{MW\alpha} = \boldsymbol{y}\|_2 \le \varepsilon \tag{12}$$

It is easily to see that sensing matrix is now the composition of wavelet synthesizing and projecting. Since the RIP of the sensing matrix plays a crucial role in sparse, we first give out some preliminary analysis of the RIP of the combined sensing matrix $\boldsymbol{MW}$. In the second part of this section, we will detail on the algorithm used to solve (12).

2.1 RIP of the combined sensing matrix

Replacing $\boldsymbol{A}$ in (6) by the combined sensing matrix $\boldsymbol{MW}$, we immediately have

$$(1-\delta_S)\|\boldsymbol{c}\|^2 \le \|\boldsymbol{MW}_T\boldsymbol{c}\|^2 \le (1+\delta_S)\|\boldsymbol{c}\|^2 \tag{13}$$

Note that $\boldsymbol{W}$ is orthonormal; so $\|\boldsymbol{c}\|_2 = \|\boldsymbol{W}_T\boldsymbol{c}\|_2$ for any index set $T \subseteq \{1,2,\cdots,m\}$ and any vector $\boldsymbol{c} \in R^{|T|}$. This immediately leads to the following theorem.

Theorem 2.1 Assume that the columns of $\boldsymbol{W}$ are orthonormal. Then the operator l_2 norm (i.e., the maximum singular value) of the system matrix $\boldsymbol{M}$ provides an upper bound of the S-restricted isometry constants of $\boldsymbol{MW}$. Exactly, denoting the maximum singular value of $\boldsymbol{M}$ by $\sigma_{\boldsymbol{M}}$, we have

$$\delta_S \le \min\left\{\left|\sigma_{\boldsymbol{M}}{}^2 - 1\right|, \quad 1\right\} \tag{14}$$

Proof The proof is straightforward from the definition of S-restricted isometry constants in (6) and the basic conclusions of the linear algebraic. What need to note is that since we usually have $|T| < n < m$, $\boldsymbol{W}_T\boldsymbol{c}$ cannot transverse the whole R^m space.

Theorem 2.1 actually indicates that to make δ_S small for larger S, the operator l_2 norm of the system matrix $\boldsymbol{M}$ should be as close to 1 as possible. However, simple numerical experiment manifests that the practical system matrix created via the ray-driven model seriously violates this condition (actually the maximum singular value of $\boldsymbol{M}$ can be as large as several thousand in our problem scale); however, solving (12) still recovers the sparse coefficients $\boldsymbol{\alpha}_0$ accurately in our simulations. Motivations to seek reasonable explanations for this lead us to the following two lemmas.

Lemma 2.2 For given sensing matrix $\boldsymbol{A}$, assume $\boldsymbol{D}$ to be a given non-singular matrix with appropriate size. If $\boldsymbol{x}_0$ is the solution to (5), then it is also the solution to the following problem,

$$\min \|\boldsymbol{x}\|_1 \quad s.t. \quad \boldsymbol{DAx} = \boldsymbol{Dy} \tag{15}$$

Proof The proof is direct since the feasible region of the convex optimization problem (15) is the same with the feasible region of (5).

Lemma 2.3 Given sensing matrix $\boldsymbol{A}$, sparse signal $\boldsymbol{x}_0$, and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e}$ with $\|\boldsymbol{e}\|_2 \le \varepsilon$, the solution to problem (8) and the solution to the following problem

$$\min \|\boldsymbol{x}\|_1 \quad s.t. \quad \|\boldsymbol{DAx} - \boldsymbol{Dy}\|_2 < \|\boldsymbol{D}\|_2 \varepsilon \tag{16}$$

are within almost the same distance from $\boldsymbol{x}_0$.

Lemma 2.3 is brought up here without profound proof; however, it leads to reasonable explanations to why we can recover sparse signals by solving (12) directly. Such explanation, starting from the following theorem, is verified and supported by numerical simulations.

Theorem 2.4 If the rows of the system matrix $\boldsymbol{M}$ are nonzero and orthogonal, then by solving (12) we can retrieve the sparse signals with errors of reasonable bound, though the RIP of the original combined sensing matrix $\boldsymbol{MW}$ is not good enough.

Proof Denote the i-th row of the system matrix $\boldsymbol{M}$ by $\boldsymbol{M}_i$. Let $\boldsymbol{D}$ be a diagonal matrix with the i-th diagonal element to be $\|\boldsymbol{M}_i\|_2^{-1}$, then we have $\boldsymbol{DM}(\boldsymbol{DM})^{\mathrm{T}} = \boldsymbol{I}$ where $\boldsymbol{I}$ denotes the unity matrix. By Theorem 2.1, the combined sensing matrix $\boldsymbol{DMW}$ obviously has good RIP. So solving the following optimization problem

$$\min \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \|\boldsymbol{DMW\alpha} - \boldsymbol{D}\,\boldsymbol{y}\|_2 \le \|\boldsymbol{D}\|_2 \varepsilon \tag{17}$$

retrieves the original sparse coefficients $\boldsymbol{\alpha}_0$ with reasonably small error. Then by Lemma 2.3, solving (12) will also retrieve $\boldsymbol{\alpha}_0$ with similar reasonable error.

Actually theorem 2.4 points out that, if the system matrix $\boldsymbol{M}$ is good structured in the sense that by imposing some non-singular square matrix $\boldsymbol{D}$ its singular values will approach 1, then we can still use the combined matrix $\boldsymbol{MW}$ to retrieve sparse coefficient effectively.

In the practical ray-driven model, every row of the system matrix $\boldsymbol{M}$ denotes a group of weights whose inner product with the image vector leads to the integral value along the corresponding ray. Since intersection of those rays is inevitable, it is easy to see that these rows of weights could not be strictly orthogonal. So after scaling all the diagonal entries to 1, the weighted symmetrical matrix $\boldsymbol{DM}(\boldsymbol{DM})^{\mathrm{T}}$ will also have many nonzero non-diagonal entries. In this case, based on the well-known Gershgorin Disc Theorem, all the eigenvalues of $\boldsymbol{DM}(\boldsymbol{DM})^{\mathrm{T}}$ will fall into an interval on the number axis whose center is 1, with the radius of the interval bounded by the maximum of the sums of the magnitude of the non-diagonal entries in one row. To make the eigenvalues as close to 1 as possible, the magnitude of the non-diagonal entries should be suppressed as small as possible.

Based on this, an obvious good way is to make the rows of $\boldsymbol{M}$ as uncorrelated as possible. For fixed number of projection angles available, this can be achieved by spreading the projecting rays more separately. We can also process $\boldsymbol{M}$ ahead by excluding those rows that are too correlated with some other row. Both approaches will be verified in simulations described in Section 3.

2.2 Solving via the fast iterative shrinkage-thresholding algorithm

To solve the program (12), we convert it to the following unconstrained l_1 norm regularized linear least squares problem as

$$\boldsymbol{\alpha}^{\#} = \arg\min_{\boldsymbol{\alpha}} \quad \lambda \left\| \boldsymbol{\alpha} \right\|_1 + \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{MW\alpha} \right\|_2^2 \tag{18}$$

Classical optimization theory points out that, the formulations in (18) and (12) are equivalent under some appropriate values of parameters, though it is hard to derive explicit relationship between λ and ε for general problems. However, from the optimality condition of (18), we can still analyze the reasonable value of λ to make these two problems achieve similar solutions.

Since the object function is convex, it is easy to see that the optimal solution $\boldsymbol{\alpha}^{\#}$ satisfies

$$\lambda \cdot \partial \left\| \boldsymbol{\alpha}^{\#} \right\|_1 + \left(\boldsymbol{MW} \right)^{\mathrm{T}} \left(\boldsymbol{MW\alpha}^{\#} - \boldsymbol{y} \right) = \boldsymbol{0} \tag{19}$$

where $\partial \left\| \boldsymbol{\alpha}^{\#} \right\|_1$ denotes the subgradient of the l_1 norm function at the point $\boldsymbol{\alpha}^{\#}$. Since we hope that $\boldsymbol{\alpha}^{\#} = \boldsymbol{\alpha}_0$ and the absolute value of $\partial \left\| \boldsymbol{\alpha}^{\#} \right\|_1$ is no more than 1, it easy to see that the parameter λ should be at the same scale with the entries in $\left(\boldsymbol{MW} \right)^{\mathrm{T}} \boldsymbol{e}$ when (19) holds. So in practice based on the estimation of the magnitude of the error $\boldsymbol{e}$, we can set the value of λ to be

$$\lambda = \left\| \left(\boldsymbol{MW} \right)^{\mathrm{T}} \boldsymbol{e} \right\|_{\infty} \tag{20}$$

To solve the unconstrained optimization itself, we adopt the fast iterative shrinkage-thresholding algorithm (FISTA)[14]. This algorithm has the merit of easy implementation since it is the extension of the optimal gradient descent method. Also this algorithm is appropriate for solving large-scale problems without special requirement of the matrix $\boldsymbol{MW}$.

FISTA algorithm requires the Lipschitz constant of the gradient of the smooth part in the object function, i.e. $\frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{MW\alpha} \right\|^2$ here. In our implementation the constant can be estimated via numerical singular value decomposition (SVD) of the system matrix $\boldsymbol{M}$.

Another acceleration measures adopted here is based on the fact that $\boldsymbol{W}$ is derived from the orthonormal wavelet transform. So the left multiplication of $\boldsymbol{W}$ and $\boldsymbol{W}^{\mathrm{T}}$ is realized via the wavelet transform rather than direct matrix multiplication.

3 Simulations

We first describe the system configuration of our simulations. The object tomographic image is the high-contrast Shepp-Logan phantom with the resolution fixed at 256×256. We assume the pixel width to be 1, and the orbit radius of the ray source is 640. On the virtual detector, there are 512 bins with the bin size fixed to 0.75. So under every angle, 512 transverse rays are simulated. Some of these rays don't transverse the phantom and so the corresponding empty rows of weights are excluded from the system matrix $\boldsymbol{M}$.

As stated in Section 2.1, excluding those rows of M that are too correlated with some other row helps improve the RIP of the combined sensing matrix. This procedure is done by checking the row vectors one by one. For every incoming row of $\boldsymbol{M}$, its correlation coefficients with all retained rows are calculated. If any correlation coefficient surpasses the threshold we set, the new row is discarded.

With the system matrix in hand, the ideal projection data is generated simply by $\boldsymbol{y} = \boldsymbol{Mx}_0$ where $\boldsymbol{x}_0$ is the ideal phantom image. Then we add some Gaussian noise to the projection data on

purpose with the SNR set to 80db. The level of Haar wavelet we used here is always three. Then the number of non-zero Haar coefficients of the 256×256 phantom is 3 823, which correspond to 5.83% of all the 65 536 coefficients.

3.1 Few-view sampling reconstruction

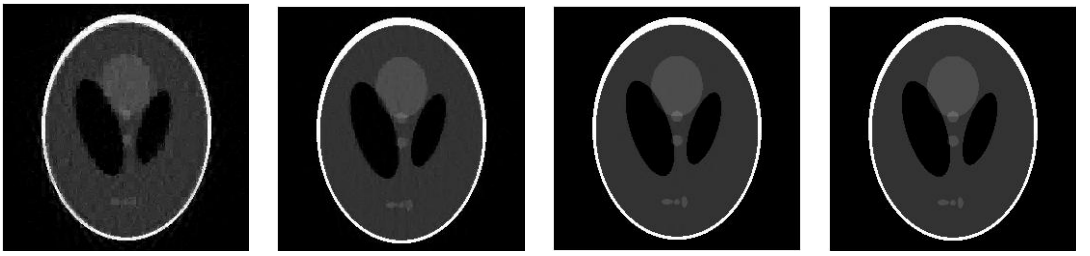
In this case, we generated the system matrix $\boldsymbol{M}$ under 4 different numbers of projection angles that sparsely cover the full 360-degree range. In the 90-angle, 60-angle and 45-angle cases, we further exclude those too-correlated rows using the method stated above. The threshold used and the final total number of rows in the system matrices are listed in Table 1. Note that throughout the paper, pattern A corresponds to excluding only those empty rows, while pattern B corresponds to excluding both empty and correlated rows from the original system matrix.

Table 1 The total number of rows of the system matrix in different cases

Number of Projection Angles	90	60	45	30
Pattern A	40 500	26 988	20 246	13 492
Pattern B	27 023	20 112	13 421	—

The threshold used for 90-B and 60-B is 0.33, and for 45-B is 0.27. These thresholds are selected so that the number of rows in case 90-B, 60-B, and 45-B is similar to that in case 60-A, 45-A, and 30-A respectively.

We then test the 7 different system matrices in the reconstruction simulation. The images reconstructed are shown in Fig.1, while the relative l_1 errors of the recovered wavelet coefficients are listed in Table 2.



(a) (b) (c) (d)

Fig.1 The reconstructed phantoms under some different full-range cases: (a) 30-A; (b) 45-B; (c) 45-A; (d) The ground truth. The results under cases 90-A/B and 60-A/B are not shown here, since they are visually indistinguishable with both (c) and (d)

Table 2 The relative l_1 error of the recovered coefficient in different full-range cases

Full-range cases	90−A	90−B	60−A	60−B	45−A	45−B	30−A
$\|\alpha^{\#}-\alpha_0\|_1/\|\alpha_0\|_1$	2.462×10^{-3}	2.778×10^{-3}	4.236×10^{-3}	4.268×10^{-3}	1.760×10^{-2}	1.186×10^{-1}	3.320×10^{-1}

From the results above, we have two conclusions about the RIP of the combined sensing matrix for CT reconstruction. First, sensing matrix with more rows tends to have small S-restricted isometry constant for larger S. This observation is straightforward from the definition of δ_S in (6), and is reflected by the improved quality from Fig.1 (a) to Fig.1 (c). Generally speaking, the sensing matrix should be wide but not too wide so that the support of signals it can recover could be larger.

The second conclusion is more relative to the analysis presented in Section 2.1. Comparing the case 45-B with the case 30-A, the heights of the two system matrices are almost the same; however, the quality of reconstructed phantom is improved perceivably from Fig.1 (a) to Fig.1 (b). Similar improvement can also be seen from the case 45-A to case 60-B as well as from case 60-A to case 90-B in the relative recovery error in Table 2. These results above strongly verified our previous analysis.

3.2 Limited range reconstruction

We also test the reconstruction performance of the method proposed under the limited-range circumstances. This time the number of projection angles is fixed to 60, while the coverage of these angles ranging from 60 to 240 degrees (with an interval of 30 degrees). For each spanning range the projection angles are still distributed uniformly. The total numbers of rows of all these system matrices are displayed in Table 3, from which it is easy to tell that the total number of rows doesn't change a lot with the spanning range varying.

Table 3　The number of rows of the system matrix in different coverage ranges

Coverage range in degree	60	90	120	150	180	210	240	360
Number of rows	27397	26986	26767	27135	26988	26847	27082	26988

This time the number of projection angles is fixed to 60, and we didn't exclude those rows that are too correlated with some other.

As in the last section, the reconstructed phantoms are shown in Fig.2. Also shown for comparison is the Fig.1 (b) where the 60 projection angles span the full 360 degrees. The relative l_1 recovery errors of the recovered coefficients are listed in Table 4.

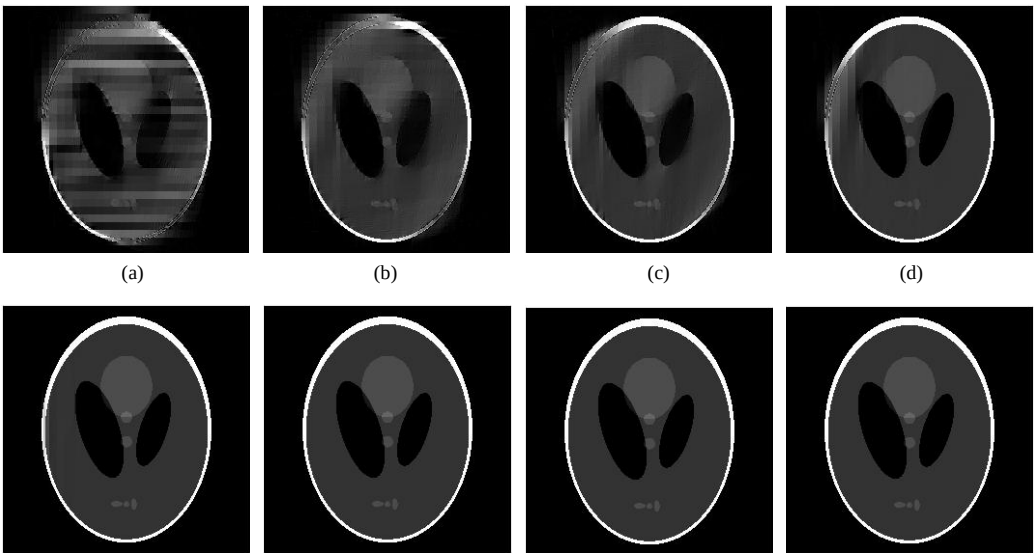


Fig.2　The reconstructed phantoms with different ranges covered by 60 projecting angles:　(a) 60 degrees; (b) 90 degrees; (c) 120 degrees; (d) 150 degrees; (e) 180 degrees; (f) 210 degrees; (g) 240 degrees; (h) 360 degrees

Table 4　The relative l_1 error of the recovered coefficient in the limited-range case

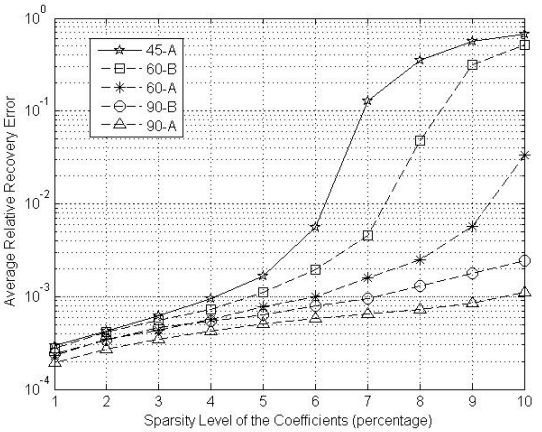
Coverage range in degree	60	90	120	150	180	210	240	360
$\|\alpha^{\#}-\alpha_0\|_1/\|\alpha_0\|_1$	5.981×10^{-1}	4.385×10^{-1}	3.017×10^{-1}	1.219×10^{-1}	1.398×10^{-2}	5.206×10^{-3}	4.357×10^{-3}	4.268×10^{-3}

Fig.2 illustrates that the reconstruction quality becomes better and better as the range of coverage increases. This tendency actually also reflects our analysis in Section 2.1. To clarify this relationship, we should first note that, the projecting rays are denser in the covering range when the range is smaller. So the rows of the system matrix will be more correlated to each other, deduced from which we know that the RIP of the combined sensing matrix will be worse than that with the angles spread more separately.

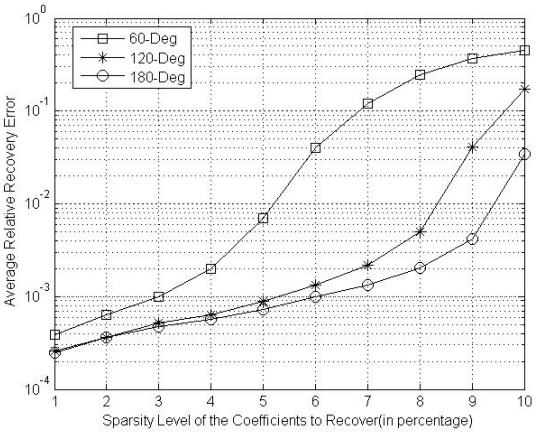
Another interesting observation obtained from Fig.2 and Table 4 has some relationship with the short-scan range in our simulation configuration. Since the fan angle here is of about 17 degrees, so the short-scan range is of 197 degrees. In Fig.2 (e) where the 180-degree range is covered, some minor artifact can still be identified; however, when the range covered surpasses 197 degrees as in Fig.2 (f) and (g), the reconstructed phantoms are visually the same with Fig.2 (h). The relative recovery errors in Table 4 match this observation too.

3.3 Simulations on random coefficients

The phantom used in the simulations above is just a special case of the general images that have sparse expansion coefficients under the Haar wavelet transform. From this special case, we have already seen the varying recovery performance in different system configurations.



(a)



(b)

Fig.3 Illustration of the average recovery error (in l_1 sense) against the varying sparsity level of coefficients to restore. (a) corresponds to those full-range cases with different projection angles, while (b) corresponds to those limited-range cases where the number of the projection angles is fixed and the range of coverage varies

To illustrate this difference more convincingly, we randomly generate some sparse wavelet coefficients with the sparsity level increasing from 1% to 10%. Under each sparsity level, ten simulations are run and the relative l_1 errors of recovery are recorded. In Fig.3 (a) the average relative errors are plotted against the sparsity level of the coefficients for different angle numbers under the full-range scanning. In Fig.3 (b) the same results are illustrated for different limited covering ranges.

Results in Fig. 3 further verified our previous analysis about the RIP of the combined sensing matrix. First the “taller” wide sensing matrix tends to be able to recover “denser” sparse signals; this is explicated shown in Fig.3 (a). Then larger coverage in the limited-range case and excluding rows that are too correlative with others help to make the row vectors of the system matrix more irrelative, thus improving the RIP of the combined sensing matrix.

4 Conclusion and future work

In this paper, we proposed a more general frame (12) than TV minimization for CT reconstruction from insufficient data. The basic idea is to try to find some basis on which the object tomographic image can be represented or effectively appropriated with sparse coefficients, so that the image can be retrieved accurately by small number of measurements.

To verify the feasibility of our proposal, we investigated one special instance where the expansion dictionary is fixed to be Haar wavelet basis. Conforming to the practical constraints of real CT system, we analyzed and brought up principles about how to improve the RIP of the combined sensing matrix. The principles were verified by following simulations.

For future work, one direction is to pursue deeper comprehension about the RIP of the combined sensing matrix. Such analysis is interesting in the sense that $\boldsymbol{M}$ is constrained by physically realizable hardware platform while $\boldsymbol{W}$ can be chosen from multiple options. Another direction is to apply the proposed method to real tomographic image reconstruction. For different reconstruction object, the selection of $\boldsymbol{W}$ can be very different. How to choose a suitable $\boldsymbol{W}$ is expected to depend highly on the specific applications.

References

[1] Song JY, Liu QH, Johnson GA, et al. Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT[J]. Medical Physics, 2007, 34(11): 4476-4483.

[2] Bian JG, Han X, Sidky EY, et al. Sparse-data reconstruction in flat-panel cone-beam CT for potential use in image-guided surgery[C]. Proceedings of the 10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 2009: 231-233.

[3] Persson M, Bone D, Elmqvist H. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography[J]. Phys Med Biol, 2001, 46: 853-866.

[4] Zhang XQ, Froment J. Total variation based Fourier reconstruction and regularization for computer tomography[C]. IEEE Nuclear Science Symposium Conference Record, 2005: 2332-2336.

[5] Candes E, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information[J]. IEEE Transactions on Information Theory, 2006, 52(2): 489-509.

[6] Sidky EY, Kao CM, Pan XC. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT[J]. Journal of X-ray Science and Techonology, 2006,14: 119-139.

[7] Sidky EY, Pan XC. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization[J]. Phys Med Biol, 2008, 53: 4777-4807.

[8] Duan XH, Zhang L, Xing YX, et al. Few-view projection reconstruction with an iterative

reconstruction-reprojection algorithm and TV constraint[J]. IEEE Transactions on Nuclear Science, 2009, 56(3): 1377-1382.

[9] Herman GT, Davidi R. Image reconstruction from a small number of projections[J]. Inverse Problems, 2008, 24: 1-17

[10] Candes E, Tao T. Decoding by linear programming[J]. IEEE Transactions on Information Theory, 2005, 51(12): 4203-4215.

[11] Candes E, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements[J]. Communications on Pure and Applied Mathematics, 2006, 59(8): 1207-1223.

[12] Li MH, Yang HQ, Kudo H. An accurate iterative reconstruction algorithm for sparse objects: Application to 3D blood vessel reconstruction from a limited number of projections[J]. Phys Med Biol, 2002, 47: 2599-2609.

[13] Wang H, Legoupil S, Desbat L. 2D X-ray tomographic reconstruction from few number of projections and 3D perspectives: Applications of compressed sensing theory[C]. Proceedings of the 10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 2009: 186-189.

[14] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problem[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.

关于不充分投影数据下的二维扇形束 CT 重建

梁文轩，胡广书✉

（清华大学生物医学工程系，北京 100084）

摘要：在医学和其他的很多 CT 应用中，一个重要问题是如何从不充分的投影数据中重建出断层图像。近年来由于新兴的压缩传感理论，Total Variation 最小化方法在求解这一问题中重新获得了重视。本文将该方法推广成一类更一般的不充分投影 CT 重建算法。这一类重建算法同样基于压缩传感理论，但是更加灵活，能适用于不同的应用背景。针对该类算法中的一个特例进行的理论分析和算法仿真展现了该算法的应用潜力。

关键词：CT 重建；压缩传感；少投影重建；有限角度重建

作者简介：梁文轩（1986－），男，清华大学生物医学工程系硕士研究生，主要从事基于 DSP 平台的 CT 成像算法加速、CT 成像算法理论等方面的研究，Tel：15120001076，E-mail：liangwenxuan1986@gmail.com；胡广书✉（1945－），男，清华大学生物医学工程系教授，主要从事生物医学信号的检测与处理、医学图像处理和数字化医疗仪器等方面的研究，Tel：010-62784568，E-mail：hgs-dea@tsinghua.edu.cn。