Reconstruction of CT Images Using Iterative Least-Squares Methods with Nonnegative Constraint

Hiromasa Kohno\textsuperscript{1}, Yuichi Tanji\textsuperscript{1}, Ken’ichi Fujimoto\textsuperscript{1}, Hiroyuki Kitajima\textsuperscript{1}, Yo Horikawa\textsuperscript{1} and Norikazu Takahashi\textsuperscript{2}

\textsuperscript{1} Faculty of Engineering and Design, Kagawa University
2217-20 Hayashi-cho, Takamatsu, Kagawa 761-0396, Japan
\textsuperscript{2} Graduate School of Natural Science and Technology, Okayama University
3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan
E-mail: \textsuperscript{1}\{tanji, fujimoto, kitaji, horikawa\}@eng.kagawa-u.ac.jp, \textsuperscript{2}takahashi@cs.okayama-u.ac.jp

Abstract The basic mathematical problem in computed tomography (CT) is defined as finding a nonnegative solution to a rectangular linear system with a sparse coefficient vector and matrix. The coefficient vector and matrix respectively correspond to acquired projection data and a projection operator based on a discrete Radon transform. As an approach to find a nonnegative solution corresponding to a tomographic image, a continuous method using simultaneous ordinary differential equations has been proposed. The continuous method can produce high-quality images even from an insufficient number of projection data. However, it requires a huge computational cost to obtain a high-quality image because of numerical integration. In this paper, to reduce the image-reconstruction time, we proposed two iterative methods without numerical integration that enable us to reconstruct high-quality CT images without negative pixels. The methods were produced on the basis of a hierarchical alternating least-squares algorithm that is known as a fast solver for nonnegative matrix factorization and conjugate gradient methods that are effective for a linear system. Through numerical experiments, we also discussed the performances of the proposed methods in terms of both the quality of obtained images and the time required to obtain high-quality images.

Keywords: computed tomography, nonnegative matrix factorization, hierarchical alternating least-squares, conjugate gradient method, preconditioner

1. Introduction

X-ray computed tomography (CT) apparatus is widely used in clinics to determine whether there is a tumor inside human bodies. CT apparatus yields tomographic images reconstructed from projection data acquired with X-ray sensors.

The basic problem in CT is defined as finding a solution vector to a rectangular linear system with projection data and a sparse projection operator that is obtained by a discrete Radon transform. A filtered back-projection (FBP) method \cite{1, 2}, which is a transform method and widely used in conventional CT apparatus, generates a high-quality image only when the number of projections satisfies the sampling theorem for a full scan of 360\degree. To be specific, the FBP method generates low-quality images with streak artifacts if the number of projections is insufficient. To obtain a high-quality image without artifacts, a large amount of projection data is needed, which means high X-ray doses are needed.

Fujimoto and co-workers \cite{3, 4} have proposed a continuous method using a continuous-time dynamical system defined by simultaneous ordinary differential equations (ODEs). Compared with the FBP method, the continuous method can produce a high-quality image corresponding to a solution to ODEs even from a small amount of projection data, which means an underdetermined linear system. The method also has an advantage for image reconstruction, in which images without negative pixels must be generated. The method, however, requires a huge computational cost to generate a high-quality image because numerical integration with a high computational cost must be performed to find the locus of a solution to ODEs precisely.

As other methods without numerical integration that can generate high-quality images even from a small number of projection data, iterative methods such as algebraic reconstruction techniques (ART) \cite{2} and the expectation-maximization (EM) method \cite{5} are also well known. The computational costs of the methods are lower than that of
the continuous method because of the absence of numerical integration; however, they do not necessarily generate images with nonnegative pixels.

Compared with the FBP method, the continuous and iterative methods can reduce X-ray doses for reconstructing high-quality images. Acquired projection data under low X-ray doses contains much noise. The FBP method reconstructs noisy, low-quality images for noisy projection data because the method is a transform method for projection data. In contrast, the continuous and iterative methods are based on numerical approximation and give true tomographic images. Therefore, they can reconstruct high-quality images even from noisy projection data.

In this paper, we propose two iterative methods that generate images without negative pixels on the basis of least-squares methods for sparse rectangular linear systems [6, 7]. One is based on the hierarchical alternating least-squares (HALS) method [8]–[11], which is known as a fast method for nonnegative matrix factorization (NMF) [12]–[14]. The other is a restarted preconditioned conjugate gradient least-squares (PCGLS) method with a nonnegative constraint (NC) and repeatedly restarts the PCGLS algorithm after negative elements of a transitional solution are reset to zeros forcibly and periodically. The idea is based on the facts that (1) CG methods are effective for solving a sparse linear system [15, 16], (2) all elements of a transitional solution are reset to zeros in every iteration in the HALS method. Through numerical experiments, we also discuss the performances of the proposed methods in terms of both the quality of obtained images and the time required to obtain high-quality images.

2. Problem Description

The basic problem in CT is to find a solution vector, \( \mathbf{x}^* \in \mathbb{R}^N_+ \), which corresponds to a tomographic image, to the rectangular linear system defined as

\[
A \mathbf{x}^* = \mathbf{y}
\]  

(1)

Here, \( A \in \mathbb{R}^{P \times N}_+ \) expresses a sparse matrix obtained with a discrete Radon transform, which corresponds to a projection operator, and \( \mathbf{y} \in \mathbb{R}^P_+ \) represents projection data acquired from X-ray detectors of CT apparatus. The symbols \( P, N, \) and \( \mathbb{R}_+ \) correspond to the number of projections, the number of pixels, and a nonnegative real space, respectively.

In this paper, we treat the case of \( P < N \), i.e., this linear problem is underdetermined and ill-posed [17]. If the square matrix \( A^\top A \), where \( A^\top \) denotes the transpose of \( A \), is non singular and the space that solutions exist in is \( \mathbb{R}^N_+ \), we can find the solution such that the Euclidean norm of the solution becomes minimum by directly solving the normal equation [15] defined as

\[
A^\top A \mathbf{x}^* = A^\top \mathbf{y}
\]  

(2)

i.e., \( \mathbf{x}^* = (A^\top A)^{-1} A^\top \mathbf{y} \). However, because the space of solutions is restricted to \( \mathbb{R}^N_+ \), there is no direct method of solving Eq. (2) as far as the authors know. Although singular value decomposition may be applied to find the least-squares solution to Eq. (1), the nonnegativity of obtained solutions is not necessarily guaranteed.

To find a nonnegative solution to the linear system on the basis of ideas of least-squares methods, we replace Eq. (1) with a least-squares optimization problem described as

\[
\min_{\mathbf{x} \in \mathbb{R}^N_+} V(\mathbf{x})
\]  

(3)

\[
V(\mathbf{x}) := \frac{1}{2} \|A \mathbf{x} - \mathbf{y}\|^2_2
\]

where \( \mathbf{x} \in \mathbb{R}^N_+ \) corresponds to an approximation of \( \mathbf{x}^* \) and \( \|\cdot\|_2 \) expresses the Euclidean norm for a vector.

As an approach to finding nonnegative solutions to Eq. (3), a continuous method using a continuous-time dynamical system described as ODEs has been proposed [3, 4]. The dynamical system described later can reconstruct high-quality images without negative pixels even for underdetermined problems. However, it takes a long time to obtain a high-quality image.

3. Continuous Method

To find a nonnegative solution to Eq. (3), Fujimoto and co-workers [3, 4] have proposed a continuous-time dynamical system defined as

\[
\frac{dx(t)}{dt} = -\text{diag}(\mathbf{x}(t)) A^\top (A \mathbf{x}(t) - \mathbf{y})
\]  

(4)

where \( \text{diag}(\mathbf{x}(t)) \) indicates the diagonal matrix of order \( N \times N \) in which the corresponding diagonal elements are elements of \( \mathbf{x}(t) \).

An equilibrium point \( \mathbf{x}_e \) in Eq. (4) corresponds to a least-squares solution to Eq. (1) and is also a minimizer of the Kullback–Leibler (KL) divergence between the current solution \( \mathbf{x}_e \) and \( \mathbf{x}^* \) [4]. Moreover, the values of \( x_n(t), n = 1, 2, \ldots, N \), emanating from nonnegative values never become negative because \( dx_n(t)/dt = 0 \) at a point such that \( x_n(t) = 0 \) in the \( n \)th ODE. These facts indicate that the continuous method using Eq. (4) generates “high-quality” images in terms of the KL divergence becoming minimum and all pixel values in a reconstructed image being nonnegative.

The continuous method, however, takes a long time to obtain a high-quality image. This is because we need an ODE solver with a high computational cost such as the Runge–Kutta method to find the locus of a solution to Eq. (4) precisely. In addition, the values of \( dx(t)/dt \) approach zero as \( x(t) \) approaches an equilibrium point. A method of accelerating the convergence of \( x(t) \) to an equilibrium point has not yet been proposed.
4. Proposed Iterative Methods

4.1 HALS method

NMF [12]–[14] is a method of factor analysis [9] and deals with only nonnegative matrices. Here, we introduce an NMF method that can be used for the reconstruction of CT images.

We now consider an equation defined as

\[ AX = Y \] (5)

where \( A \in \mathbb{R}^{P \times N} \), \( X \in \mathbb{R}^{N \times Q} \), and \( Y \in \mathbb{R}^{P \times Q} \) are matrices without negative elements. The original problem in NMF is to factorize a given nonnegative \( Y \) to the product of \( A \) and \( X \), and the HALS method [8] has been proposed to solve the original problem. However, since the problem that we treat here is to reconstruct CT images, we assume that both \( A \) and \( Y \) are given and the problem is to find \( X \) that satisfies Eq. (5). In the use of the HALS method, this assumption is reasonable. In fact, although \( A \) and \( X \) are alternately updated in the original HALS algorithm to solve NMF problems, only \( X \) is updated under the assumption that \( A \) is constant in CT problems.

In accordance with the assumption that both \( A \) and \( Y \) are given, we replace Eq. (5) with a least-squares optimization problem described as

\[
\min_{X \in \mathbb{R}^{N \times Q}} W(X) = \frac{1}{2} \| AX - Y \|_F^2, \tag{6}
\]

where \( \| \cdot \|_F \) expresses the Frobenius norm for a matrix. We also assume \( Q = 1 \); then the problem described in Eq. (6) is completely replaced with the CT problem described in Eq. (3).

We again consider the least-squares problem described in Eq. (3) and apply the HALS method to solve the problem. Here, we express the \( j \)th column vector of \( A \) and the \( j \)th element of \( x \) as \( a_j \) and \( x_j \), respectively. We also assume that \( a_j \neq 0 \) for all \( j \). Then, we can obtain \( x_j(n) \) by solving the \( j \)th subproblem of Eq. (3) in order from \( j = 1 \) to \( j = N \), where \( n \) is a nonnegative integer and denotes the discrete time or iteration number. For example, the \( j \)th subproblem is defined as

\[
\min_{x_j(n)} \| a_j \cdot x_j(n) - R_j(n) \|_2^2 \tag{7}
\]

where

\[
R_j(n) = y - \sum_{k=1,k\neq j}^N x_k(n) \cdot a_k \tag{8}
\]

i.e., when we find \( x_j(n) \), we assume that all the elements of \( x(n) \) except for \( x_j(n) \) are constant. To be specific, the update rule of \( x_j \) is derived from the normal equation \( A_j^\top a_j x_j(n) = A_j^\top R_j(n) \) of Eq. (7) and is defined as

\[
x_j(n + 1) = \frac{[R_j(n)]_+ \cdot a_j}{\| a_j \|_2^2} \tag{9}
\]

where \( [a]_+ := \max(a, 0) \). Here, we must pay attention to what is needed to update \( R_{j+1}(n) \) before finding \( x_{j+1}(n + 1) \). The values of \( R_{j+1}(n) \) can be computed with Eq. (8); however, the updating process is not efficient in terms of computational cost. To reduce the cost of computing \( R_{j+1}(n) \), we use another formula [8] defined as

\[
R_{j+1}(n) = R_j(n) - x_j(n+1) \cdot a_j + x_{j+1}(n) \cdot a_{j+1} \tag{10}
\]

4.2 Restarted PCGLS methods with nonnegative constraint

We now treat a normal equation defined as \( A^\top A x = A^\top y \). CG methods can be applied to the normal equation and are effective for solving the linear system even though \( A^\top A \) is generally dense [18].

The solutions to the normal equation and least-squares problem of

\[
f(x) = \frac{1}{2} \langle x, A^\top A x \rangle - \langle x, A^\top y \rangle \tag{11}
\]

are the same, where \( \langle \cdot, \cdot \rangle \) represents the inner product for two vectors. The CGLS method solves the least-squares minimization problem of Eq. (11) and its algorithm is described in Algorithm 1 [15, 18].

According to this algorithm, we can obtain the sequence of \( x \) emanating from a user-defined initial point \( x_0 \). However, the convergence speed of a solution for the CGLS method depends on the condition number of \( A^\top A \). Therefore, its convergence can be improved using an appropriate preconditioner [16]. In this study, we choose

\[
M = \text{diag}(A^\top A)^{-1} \tag{12}
\]

as a preconditioner because \( \text{diag} \ (A^\top A) \) necessarily becomes nonsingular, although \( A^\top A \) in CT problems may not be of full column rank. Since the matrix \( M \) becomes approximately \( (A^\top A)^{-1} \), \( MA^\top A \) is similar to the identity matrix. In addition, it is expected that the condition number of \( MA^\top A \) is lower than that of \( A^\top A \). The PCGLS method solves the least-squares problem for the preconditioned normal equation described by

\[
MA^\top Ax = MA^\top y \tag{13}
\]

Algorithm 1 The CGLS method

1: Set \( T \): the maximum number of iterations
2: Set \( x_0 \): initial values of \( x \)
3: \( r_0 = A^\top (y - Ax_0) \), \( p_0 = r_0 \)
4: for \( n = 1, 2, \ldots, T \) do
5: \( \alpha_n = \frac{r_n^\top r_n}{(p_{n-1}^\top A^\top p_{n-1})} \)
6: \( x_n = x_{n-1} + \alpha_n p_{n-1} \)
7: \( r_n = r_{n-1} - \alpha_n p_{n-1} \)
8: \( \beta_n = \frac{r_n^\top r_n}{(r_{n-1}^\top r_{n-1})} \)
9: \( p_n = r_n + \beta_n p_{n-1} \)
10: end for
and is generally faster than the CGLS method when solving the original normal equation. However, it is not necessarily guaranteed that a solution obtained with the PCGLS method does not take negative values.

To solve the problem with nonnegative solutions, we append a periodic replacement so that solutions to the PCGLS method never have a negative value, i.e., a nonnegative constraint (NC). This additional processing originates from the idea based on Eq. (9) in the HALS method and replaces all negative elements of a transitional solution with zeros every certain number of iterations. From another perspective, one can say that the PCGLS method is repeatedly restarted after all negative elements of a transitional solution are set to zero. Therefore, we call this the “restarted PCGLS method with NC”.

We show its algorithm in Algorithm 2. In this algorithm, we first set the maximum number of iterations \( T \) and the number of iterations before restarting \( Q \) such that \( T \) becomes a multiple of \( Q \). In line 14, the function \( \max(r, 0) \), where \( r \in \mathbb{R}^N \) and 0 represents a zero vector with \( N \) elements, returns a vector with \( N \) elements that are obtained by \( \max(r_j, 0), j = 1, 2, \ldots, N \). That is, the function returns a nonnegative vector.

### Algorithm 2 Restarted PCGLS method with NC

1: Set \( T \): the maximum number of iterations
2: Set \( Q \): the number of iterations before restarting
3: \( x_0 = MA^\top y \)
4: for \( m = 1, 2, \ldots, T/Q \) do
5: \( r_0 = A^\top(y - Ax_0), p_0 = Mr_0 \)
6: for \( n = 1, 2, \ldots, Q \) do
7: \( \alpha_n = (r_{n-1}^\top, Mr_{n-1}) \)
8: \( x_n = x_{n-1} + \alpha_n p_{n-1} \)
9: \( r_n = r_{n-1} - \alpha_n p_{n-1} \)
10: \( \tilde{r}_n = M r_n \)
11: \( \beta_n = (\tilde{r}_n^\top, \tilde{r}_{n-1}) \)
12: \( p_n = r_n + \beta_n p_{n-1} \)
13: end for
14: \( x_0 = \max(x_Q, 0) \)
15: end for

5. Experimental Results

5.1 Preliminaries

We used a 256 × 256 Shepp–Logan phantom image in Fig. 1 imitating a human skull. Here, for display purposes, intensities of 0 and 1 correspond to pixel values of 0 (black) and 255 (white), respectively.

For the size of the phantom image, the number of virtual X-ray detectors was set to 367 in accordance with the “radon” function in MATLAB®. We also set the angular range of projections to \([0, 180)°\) and the number of projection directions to 100, i.e., we assumed that an X-ray was exposed every 1.8°. From the prescribed parameter values, \( N = 65, 536 \) and \( P = 36, 700 \).

In the continuous method, we chose the Dormand–Princean explicit Runge–Kutta method [19], i.e., the “ode45” function in MATLAB®, to solve Eq. (4) and we set the maximum number of executions of the Runge–Kutta method to 200. We also set the maximum number of iterations to 200 for the restarted PCGLS method with NC.

To evaluate the quality of a reconstructed image \( x \), we used the peak signal-to-noise ratio (PSNR) defined by

\[
\text{PSNR} = 20 \log_{10} \left( \frac{255}{\sqrt{\text{MSE}}} \right)
\]

One can conclude that the quality of an obtained image \( x \) with a high PSNR is good because the PSNR takes a large value if the mean squared error (MSE) is small. We also used the residual norm defined as \( \|Ax - y\|_2 \) to show the convergence of solutions.

The computer used in our experiments had an Intel® Core™ i5-7400 CPU with 3.00 GHz and a main memory of 16.0 GB. The version of MATLAB® that we used was R2017a.

5.2 Continuous method

For the case of using the continuous method, we show the time evolution of the PSNR and residual norm in Fig. 2. The residual norm monotonically decreases with time in Fig. 2(b). The results were consistent with the results of analysis [3]. On the other hand, we confirmed that the PSNR gradually increased and reached 33.3 dB at 200 iterations in Fig. 2(a). The computational time for 200 iterations was 4651.29 s.

5.3 HALS method

Figure 3 show the time evolution of the PSNR and residual norm in the case of using the HALS method. The residual norm gradually decreases with time in Fig. 3(b); however, we confirmed a nonsmooth locus within 60 iterations after the initial state. We also confirmed that the PSNR did not increase monotonically within 30 iterations.
after the initial state in Fig. 3(a). The nonsmooth loci were caused by replacing negative values with zeros using the function $\lfloor \cdot \rfloor_+$ in Eq. (9).

The PSNR gradually increased except for in the aforementioned period of computational time and became approximately 29.6 dB at 200 iterations. The method required 836.36 s for 200 iterations.

5.4 Restarted PCGLS method with NC

We also carried out image reconstruction using the restarted PCGLS method with NC. In the experiments, the number of iterations before restarting $Q$ was set to 5, 10, and 20. Figure 4 shows the time evolution of the PSNR and residual norm in this method. In Fig. 4(a), the PSNRs with $Q = 5$ and $Q = 10$ monotonically increased and were almost the same values (32 dB) after 200 iterations. The PSNR with $Q = 20$ also monotonically increased and exceeded 30 dB after 200 iterations; however, it was slightly lower than those in the other cases.

In Fig. 4(b), the values of residual norms in the three cases decreased every iteration. For the cases of $Q = 10$ and $Q = 20$, the values of the residual norms at 200 iterations were considerably different, although their PSNRs at 200 iterations were almost the same.

On the other hand, the method with $Q = 5$, 10, and 20 required 9.9656, 9.3835, and 9.1272 s for 200 iterations, respectively. As an inevitable consequence, the method with $Q = 5$ needed more time than the other cases because of the large amount of processing for restarting.

6. Discussion

Referring to the results, we discuss the performances of the proposed methods in terms of both the quality of obtained images and the time required to obtain high-quality images.

Figure 5 shows the PSNRs for the continuous method, HALS method, and PCGLS method with $Q = 5$. From the comparison of the PSNRs after 200 iterations, one can conclude that the quality of the image obtained with the continuous method was the highest. Note that the PCGLS method also generated a high-quality image similarly to the continuous method because the PSNR for the PCGLS method at 200 iterations was slightly less than that for the continuous method. The PSNR for the HALS method was the lowest among the methods.
In addition, we visually inspected the quality of the reconstructed images. To compare the quality of images at a glance, we show the phantom image of Fig. 1 again and images reconstructed by the three methods in Fig. 6. The quality of the images reconstructed by the continuous and PCGLS methods was almost the same, and we can see only slight artifacts in the two reconstructed images. In contrast, streak artifacts were clearly visible in the image reconstructed by the HALS method. Note that intensities of more than one in the reconstructed images were matched to 255 (white) to ensure the consistency of the display with the phantom image.

On the other hand, the time required for 200 iterations in the continuous method was longer than those in the HALS and PCGLS methods. The restarted PCGLS method with $Q=5$ was the quickest among the three methods. Let us compare the reconstruction times between the continuous and PCGLS methods. In Fig. 5, the PSNR with the continuous method reached 32 dB after 45 iterations, which corresponds to about 1,050 s. In comparison, the PSNR for the PCGLS method reached 32 dB after 200 iterations and it took only 10 s to obtain an image of the same quality. The results indicated that the PCGLS method was approximately 105 times quicker than the continuous method.

7. Conclusion

On the basis of iterative least-squares approaches, we proposed the HALS and PCGLS methods to reconstruct high-quality images without negative pixels. The restarted PCGLS method with NC generated a high-quality image whose PSNR was similar to that for the continuous method. In addition, the time required to reconstruct a high-quality image was about $1/105$ that of the continuous method. On the other hand, the PSNR for the HALS method was lower than that for the PCGLS method, and the reconstruction time also took longer than that for the PCGLS method. From the results, we conclude that the restarted PCGLS method with NC is the best among the three methods including the continuous method. In future,
we will examine the tolerance of the proposed methods to noisy projection data.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant Number 17H01816.

References

[3] K. Fujimoto, O. M. Abou Alola and T. Yoshinaga: Continuous-time image reconstruction using differential equations for computed tomo-
Lyapunov function based on Kullback-Leibler divergence for a
tensor factorizations: A unified view based on block coordinate de-
HALS algorithm for nonnegative matrix factorization, 2015 IEEE
6th Int. Workshop Computational Advances in Multi-Sensor Adap-
negative factor model with optimal utilization of error estimates of
[13] D. D. Lee and H. S. Seung: Learning the parts of objects by non-
[18] M. R. Hestenes and E. Stiefel: Methods of conjugate gradients

Hiromasa Kohno received his B.E. degree from Kagawa University in 2017. He is currently studying for his M.E. degree at Kagawa University.

Yuichi Tanji received his B.E., M.E., and Ph.D. degrees from the University of Tokushima, Tokushima, Japan, in 1993, 1995, and 1998, re-
spectively. After graduation, he held Research Associate positions at Sophia University and Kagawa University, and is currently a Professor at the Faculty of Engineering and Design, Kagawa Uni-
versity. His research interests are in electronics design automation, artificial neural networks, and image and signal processing. Dr. Tanji is a member of IEEE and IEICE.

Ken’ichi Fujimoto received his M.E. degree from Tokyo Institute of Technology in 1997 and his Ph.D. de-
gree from the University of Tokushima in 2000. He is currently an Associate Professor at Kagawa University. His research interest spans interdisciplinary basic research on the dynamics of non-
linear dynamical systems and its appli-
cations to medical engineering as tomo-
graphic image reconstruction. He is a
member of IEEE, IEICE, IEEJ, SICE,
and ISCIE.
Hiroyuki Kitajima received his B.E., M.E. and D.E. degrees in electrical and electronic engineering from Tokushima University, in 1993, 1995, and 1998, respectively. He is a Professor at the Faculty of Engineering and Design, Kagawa University. He is interested in bifurcation problems.

Yo Horikawa is a Professor in the Faculty of Engineering at Kagawa University, Japan. He received his B. Eng., M. Eng., and Ph.D. (Eng.) degrees in mathematical engineering and information physics from the University of Tokyo in 1983, 1985, and 1994, respectively. His research interests include nonlinear dynamical systems and statistical pattern recognition.

Norikazu Takahashi received his B.E., M.E., and D.E. degrees from Kyushu University, Japan, in 1991, 1993, and 1996, respectively. He is currently a Professor in the Department of Computer Science, Okayama University, Japan. His research interests include optimization theory, nonlinear systems, multiagent systems, graph theory, and neural networks. He is a member of IEEE, IEICE, and Japanese Neural Network Society.

(Received August 30, 2018; revised October 31, 2018)