Dynamic SPECT reconstruction from few projections: a sparsity enforced matrix factorization approach

This content has been downloaded from IOPscience. Please scroll down to see the full text.
2015 Inverse Problems 31 025004
(http://iopscience.iop.org/0266-5611/31/2/025004)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 202.121.182.12
This content was downloaded on 28/01/2015 at 07:58

Please note that terms and conditions apply.
Dynamic SPECT reconstruction from few projections: a sparsity enforced matrix factorization approach

Qiaqiao Ding1, Yunlong Zan2,3, Qiu Huang2,3 and Xiaoqun Zhang1,4

1 Department of Mathematics, MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China
2 School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, People’s Republic of China
3 Rui Jin Hospital, School of medicine, Shanghai Jiao Tong University, Shanghai 200030, People’s Republic of China
4 Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China

E-mail: xqzhang@sjtu.edu.cn

Received 11 September 2014, revised 11 November 2014
Accepted for publication 1 December 2014
Published 21 January 2015

Abstract

The reconstruction of dynamic images from few projection data is a challenging problem, especially when noise is present and when the dynamic images are vary fast. In this paper, we propose a variational model, sparsity enforced matrix factorization (SEMF), based on low rank matrix factorization of unknown images and enforced sparsity constraints for representing both coefficients and bases. The proposed model is solved via an alternating iterative scheme for which each subproblem is convex and involves the efficient alternating direction method of multipliers (ADMM). The convergence of the overall alternating scheme for the nonconvex problem relies upon the Kurdyka–Łojasiewicz property, recently studied by Attouch et al (2010 Math. Oper. Res. 35 438) and Attouch et al (2013 Math. Program. 137 91). Finally our proof-of-concept simulation on 2D dynamic images shows the advantage of the proposed method compared to conventional methods.

Keywords: dynamic SPECT, low rank matrix factorization, Kurdyka–Łojasiewicz property, sparsity, total variation

(Some figures may appear in colour only in the online journal)
1. Introduction

Single photon emission computed tomography (SPECT) is a nuclear medicine diagnostic technique which measures the three-dimensional activity distribution of a radioactively-labeled pharmaceutical injected into the body prior to measurement. As compared to standard imaging techniques such as computed tomography (CT), the significance of SPECT lies in the fact that it reveals the variation of the function rather than the structure. For example, if a radio-pharmaceutical is rejected by healthy tissue and absorbed by unhealthy tissue, SPECT will show the unhealthy tissue as a bright region, which helps in clinical diagnosis. Similar to another nuclear imaging technique, positron emission tomography (PET) [1–4], SPECT has been used in early clinical diagnosis with a higher detection sensitivity than CT or magnetic resonance imaging (MRI). A SPECT camera works by rotating around the patient and detecting gamma rays emitted from the radio tracer in the targeted tissue. The resulting images on the camera are 2D projections of the original 3D activity distribution in the patient.

It is commonly assumed that in dynamic SPECT imaging the radioactivity concentration is static during the acquisition period of each projection rotation. Many works [5–8] have been proposed to monitor the tracer concentration functions over time under this assumption. However, physiological processes in the body are usually dynamic, and some organs (kidney, heart) show a significant decay of activity due to wash out, especially when the measuring procedure takes a considerable amount of time. Hence, application of the static assumption in the dynamic case often leads to serious motion artifacts in the reconstruction [9]. Being able to trace activity in space and time is therefore of importance and expected to significantly enhance diagnostic possibilities. With a significant decay of the radioisotope over the acquisition period, the filtered back projection method (FBP) is never good in SPECT due to the spatial attenuation of the emitted photons, and new approaches have to be developed. In early work [10, 11], a procedure was first developed in which a dynamic image sequence was reconstructed independently for each frame, followed by Wiener filtering across the frames for noise reduction. Subsequently, Jin et al proposed in [12, 13] a joint reconstruction approach for dynamic imaging, where the dynamic images of the different times are treated collectively as a single signal through motion compensation, and determined from the acquired data by using maximum a posteriori (MAP) estimation. Carson [14] and Formiconi [15] put forward a method to extract the time activity curves (TACs) of regions of interest. After that, there was a great deal of research into the extraction of the TACs of the radionuclide tracer in different tissues from a time series of reconstructed dynamic images or directly from projections [16–18] and then estimating metabolic parameters through a compartmental model [19]. In related work, many approaches have been developed to extract the TACs directly from SPECT projections, such as the ‘dSPECT method’ [20, 21], the ‘d2EM method’ presented by Humphries et al [22] and factor analysis of dynamic structures (FADS) [23]. In [9], motion of the organs has been taken into account for the dynamic CT reconstruction.

The reconstruction of the dynamic SPECT image is an ill-posed inverse problem, especially when very few views can be acquired. Nowadays, sparsity regularization has been widely adopted in diverse imaging applications [24], along with rapid development of the field of compressive sensing (CS) [25–27]. For instance, by modeling unknown images as piecewise constant functions, total variation (TV) regularization, proposed in [28], has been demonstrated to be effective for tomographic reconstruction from undersampled and limited projection data [29–32]. In the dynamic setting, a low rank matrix factorization based method is proposed in [33] for 4D cone beam CT, where sparsity, such as spatial wavelet sparsity and temporal periodic structure for breathing cycles, are penalized for the factorization matrices.
In addition, a temporal nonlocal regularization method has also been developed in [34] for 4D cone beam CT reconstruction from few projections data. In [35], a blurred piecewise constant image model is proposed for reconstructing dynamic images from few projection data.

In this paper, we aim to reconstruct the dynamic radioisotope change in a SPECT image from very few projection data (such as two projections per time interval in a pinhole camera), under the assumption that the organs of interest are nearly static during the acquisition procedure. Inspired by the work on dynamic image reconstruction with matrix factorization with sparsity, such as in [33] and the concept of blind compressive sensing in [36], we propose a matrix factorization model based on the enforcement of local and global coherence in the temporal-spatial domain to reduce the degrees of freedom in the desired solution. More specifically, global temporal-spatial coherence is modeled as a low rank matrix representation \( U = \alpha B^T \), where the coefficient \( \alpha \) and the basis \( B \) carry designed physical properties. The matrix \( B \) corresponds to the basis used in the compartmental model, from which the radioisotope concentration distribution is mixed. We also enforce the group sparsity of coefficients by assuming that the concentration distribution is mixed from few bases for each voxel. Meanwhile, the representation coefficients \( \alpha \) are assumed to have small spatial and temporal local variation according to the spatial piecewise constant model. Furthermore, the smoothness of basis TACs \( B \) is enforced through a regularization functional along the temporal direction. Finally, the proposed variational model is composed of a data fidelity term and several regularization elements, together with other physical constraints, such as non-negativity.

The overall model can be alternately solved for \( \alpha \) and \( B \), for which each subproblem is convex and solved by the popular \( l_1 \) based first order splitting method, such as the alternating direction method of multipliers (also known as the split Bregman method [37]). Our another contribution is to provide a convergence analysis for the alternating scheme for the nonconvex problem. The proposed algorithm can be viewed as a proximal regularization of the two-blocks Gauss-Seidel method. Recently, the proximal block coordinate method for nonconvex and nonsmooth functions have been studied by many researchers, such as [38–41]. In [38], it has been shown that if a nonconvex and nonsmooth function satisfies the sufficient decrease condition and relative inexact optimality condition, the proximal block descent method will have been proved to converge to a critical point without any assumption of convexity. We will show the convergence of our scheme under this framework. Finally, our numerical experiments on the simulated phantom have shown the feasibility of the proposed model for reconstruction from highly undersampled data with noise, compared to the conventional methods such as the FBP method and the least squares method. In terms of computational complexity, the problem size has been significantly reduced due to the low rank factorization and the subproblems are easy to implement, which allows a potential application to real large scale reconstruction in practice.

The paper is organized as followed: section 2 briefly introduces the concept of sparse representation with known and unknown bases. Section 3 presents the proposed model and the numerical algorithm. The convergence is provided in section 4. Finally, section 5 demonstrates numerical results on simulated 2D dynamic images, in the settings of static boundary, limited projections and moving boundary.

### 2. Sparse representation

Before presenting our model, we review the notion of sparse representation. Sparse representation or regularization has been largely used in signal and image reconstruction, machine
learning and the recently developed field of compressive sensing (CS) [25–27]. The goal here is to reconstruct a vector \( u \in \mathbb{R}^n \) from measurements \( b = Au \), where \( A \in \mathbb{R}^{m \times n} \). This problem is ill-posed in general and therefore has infinitely many possible solutions. A fundamental assumption in CS is that we seek the sparsest solution:

\[
\hat{u} = \arg \min_u \|u\|_0 \quad \text{s.t.} \quad f = Au
\]  

(1)

where \( \|\cdot\|_0 \) denotes the number of nonzero elements of the vector. This idea can be generalized to the case in which \( u \) is sparse under a given basis, for example Fourier or wavelet bases, so that there is a sparse vector \( \alpha \) such that \( u = B\alpha \). Problem (1) then becomes

\[
\hat{\alpha} = \arg \min_\alpha \|\alpha\|_0 \quad \text{s.t.} \quad f = AB(\alpha)
\]  

(2)

and the reconstructed signal is \( \hat{u} = B(\hat{\alpha}) \). When the maximal number of nonzero elements in \( \alpha \) is known equal to \( k \), we may consider the objective

\[
\hat{\alpha} = \arg \min_\alpha \|f - AB(\alpha)\|_2^2 \quad \text{s.t.} \quad \|\alpha\|_0 \leq k.
\]  

(3)

Even when the recovery property is achieved in CS, the knowledge of the basis \( B \) is required for the reconstruction process. In literature, dictionary learning or sparse coding methods are often designed to simultaneously recover both the sparse coefficients and the unknown basis by solving the followed model

\[
\left( \hat{\alpha}, \hat{B} \right) = \arg \min_{\alpha, B} \|f - AB(\alpha)\|_2^2 \quad \text{s.t.} \quad \|\alpha\|_0 \leq k.
\]  

(4)

For instance, K-means singular value decomposition (K-SVD) proposed in [42] is a popular method for learning a patch based image representation. A similar idea is also proposed as blind compressive sensing (BCS) in [36, 43, 44]. A data driven tight frame is proposed in [41, 45, 46] for natural image de-noising, which can greatly reduce the computational burden of the dictionary based method. All these methods are based on the assumption that the desired signal \( U \) is a product of a sparse coefficient matrix \( \alpha \) and a dictionary \( B \). In conventional sparse representation, we often have a large pool of dictionary or analytic bases, where the degrees of freedom of the coefficients are still huge, even with the sparsity assumption.

3. Sparsity enforced matrix factorization model

3.1. Low rank representation

The traditional tomographic problem is that \( f \) is the data observed and \( A \) is the projection matrix and we want to reconstruct unknown image \( u \) under the condition \( Au = f \), where the 2D/3D image \( u \) is stacked as a vector. In dynamic SPECT, we need to reconstruct a sequence of \( T \) images \( u_1, u_2, \ldots, u_T \) from \( f_1, f_2, \ldots, f_T \), where \( f_t \) is projection data at time interval \( t \) with different view angles. We can denote \( A_1, A_2, \ldots, A_T \) as \( T \) corresponding projection matrices and

\[
A_i u_t = f_t, \quad t = 1, 2, \ldots, T.
\]

Assume that the number of pixels/voxels of each image \( u_t \) is \( M \) and the dynamic image vector \( U \in \mathbb{R}^{M \times T}; \left( u_1, u_2, \ldots, u_T \right) \). For ease of notation, we represent the sequence of equations
as a linear transform \( AU = f \), where \( f \) is the matrix formed by the projection vectors.

Tracer kinetic modeling is widely used in biological research to make tracks for the dynamic processes of blood flow, tissue perfusion, metabolism and receptor binding [16]. The compartment model establishes the connection between the dynamic of the tracer with the image and gives a mathematical description of the pathway and the dynamic behavior of the biological tissue of interest. Between the different compartments of the model, that is the physical space such as an organ, there exist the transportation [47] and mixture. In other words, it assumes that the concentration distribution of the radioisotope \( u(x) \), for each pixel/voxel \( x \in \Omega \) at time \( t \) can be represented as a linear combination of some basis TACs:

\[
    u_t(x) = \sum_{k=1}^{K} \alpha_k(x) B_k(t),
\]

where \( B_k(t) \) denotes the TAC for the \( k \)th compartment at time \( t \), and \( \alpha_k(x) \) denotes the mixed coefficients. This can be reduced in matrix form as

\[
    U = aB^T,
\]

where \( a \in \mathbb{R}^{M \times K} \) and \( B \in \mathbb{R}^{T \times K} \) with \( K \) as the number of compartments. Since, in general, \( K \) is a small number compared to the number of time intervals \( T \), we naturally obtain a low rank matrix representation for \( U \).

Note that low rank matrix factorization is a straightforward way for dimension reduction. In particular, nonnegative matrix factorization [48–51] has been considered in many machine learning applications to reconstruct the inline pattern or feature of the data. In general, depending on data and applications, \( K \) can be either equal to or greater than the real rank of the matrix. In this paper, we aim to enforce more constraints for the basis \( B \) and the coefficient matrix \( a \) by exploring the physical meaning of these two variables in the setting of dynamic SPECT images. Our proposed model imposes the low-rank structure of the dynamic images by assuming that the unknown concentration distribution is a sparse linear combination of few temporal basis functions which represent the TACs of different compartments. The dictionary and the sparse coefficient are simultaneously recovered from the under-sampled measurements. Our model can be regarded as data driven basis learning as well as a low rank matrix factorization approach, with further imposed constraints.

### 3.2. Variational model

In real problems, the observed projection data are often inevitably accompanied by noise. In this paper, for simplicity, we consider white Gaussian noise. But we point out that the model can be easily adapted to Poisson noise with a different data fidelity term. Suppose the data noise is \( \epsilon \), we have

\[
    AU + \epsilon = (A_1 u_1, A_2 u_2, \ldots A_T u_T) + \epsilon = (f_1, f_2, \ldots f_T) = f, \tag{6}
\]

where \( U = aB^T \). The straightforward MAP estimator is given by solving the least squares formulation

\[
    \sum_{i=1}^{T} ||A_i u_t - f_t||^2_2 \approx ||AaB^T - f||^2_2.
\]
The generic model we propose takes the following form

\[
\{ \hat{a}, \hat{B} \} = \arg \min_{a \geq 0, B \geq 0} \| AaB^T - f \|_F^2 + \epsilon \phi(a) + \lambda \psi(aB^T) + \sigma \varphi(B) \tag{7}
\]

where \( \epsilon, \lambda, \sigma \) are positive parameters. Here we impose the nonnegativity constraints for both basis \( B \) and coefficients \( a \) so that the reconstructed image \( U = aB^T \) is positive. The three terms \( \phi(\cdot), \psi(\cdot) \) and \( \varphi(\cdot) \) correspond to the \( a \) priori constraints that we can impose for the representing coefficients \( a \), the dynamic image \( U = aB^T \) and the basis \( B \). In the following, we will explain specifically how these terms are chosen.

We set

\[
\phi(a) = \sum_{k=1}^{K} TV(a_k) + \kappa \| a \|_{1,\infty} \tag{8}
\]

to recover the desired structure in the spatial and temporal domain, where \( a_k \) is the \( k \)th column of the coefficient \( a \), and each element of \( a_k \) represents the contribution of the \( k \)th basis to the current pixel. The \( \ell_{1,\infty} \) norm is defined as

\[
\| a \|_{1,\infty} = \sum_{j=1}^{K} \max_i a_{i,j} \tag{9}
\]

The total variation \( TV(a_k) \) is applied to the image form reshaped from the column vector \( a_k \). Recall that the total variation of an image \( u: \Omega \rightarrow \mathbb{R} \) is defined as

\[
TV(u) = \int_{\Omega} \sqrt{V_x u_x^2 + V_y u_y^2} \, dx,
\]

where \( V \) is the gradient operator. For short notation, we denote \( \phi \) as

\[
\phi(a) = \| Va \|_{1,2,1} + \kappa \| a \|_{1,\infty} \tag{10}
\]

where \( \| v \|_{1,2,1} = \sum_{k=1}^{K} \sum_{i,j}^{N} \sqrt{v_{2,k}^2(i,j) + v_{2,k}^2(i,j)} \) for \( v = Va \).

The first term in (8) is used to recover the piecewise constant structure since the representing coefficients from the same structure are assumed to be the same. The second term is designed to select as few as possible bases to represent the images \( u_t \). This type of column sparsity was previously studied in [52] and in [50] for hyperspectral image classification.

Considering images are piecewise constant, the nonzero gradient of the image can be used to describe the boundary of organs. Assuming that the organs only present minor movement or stay static, we can use the piecewise smooth structure of the edges of each image along time. In particular, we set

\[
\psi(aB^T) = \| V_a V (aB^T) \|_{F}^2 \tag{11}
\]

where \( \| \cdot \|_F \) is the Frobenius-norm of \( V_a V (aB^T) \). To be more specific, denote

\[
V (aB^T) = VU = (Vu_1, Vu_2, \ldots Vu_T), \tag{12}
\]
where $V_{u_t} = (V_{u_1}, V_{u_T})$ and denote

$$V_i \left( \begin{bmatrix} \alpha B^T \end{bmatrix} \right) = V_i(VU) = V_i \left( \begin{bmatrix} V_{u_1} & \ldots & V_{u_T} \\ V_{u_1} & \ldots & V_{u_T} \end{bmatrix} \right) \triangleq \left( \begin{array}{ccc} w_{11} & \ldots & w_{1T} \\ w_{21} & \ldots & w_{2T} \end{array} \right) \triangleq W_i.$$ 

Suppose $u_t$ is an image with resolution $M = N \times N$, and the square of the Frobenius-norm of $W$ is defined as follows,

$$\|W\|_F^2 = \sum_{i,j} \left( \sum_{t=1}^T w_{1,t}(i,j) + \sum_{t=1}^T w_{2,t}(i,j) \right). \quad (13)$$

Finally, since the decay of radioactive distribution is smooth over time, we can set

$$\varphi(B) = \sum_{k=1}^K \sum_{t=1}^T (B(t+1,k) - B(t,k))^2 \quad (14)$$

where $B(t,k)$ denotes the $k$th basis at time $t$.

Taking all we have mentioned above into consideration, we can write out the model of the problem as

$$\{ \hat{a}, \hat{B} \} = \arg \min_{a \in \mathbb{R}^{m \times k}, a \geq 0, B \in \mathbb{R}^{n \times k}, B \geq 0} \| AaB^T - f \|_F^2 + \varepsilon \sum_{k=1}^K \|TV(\alpha_k)\|_1 + \delta \|\alpha\|_\infty$$

$$+ \lambda \| V_i V_i \left( \alpha B^T \right) \|_F^2 + \sigma \sum_{k=1}^K \sum_{t=1}^T (B(t+1,k) - B(t,k))^2, \quad (15)$$

where $\varepsilon > 0$, $\delta > 0$, $\lambda > 0$, $\sigma > 0$ and $\delta = \varepsilon \delta$.

### 3.3. Algorithm

The optimization problem (15) is nonconvex and we solve it with a proximal Gauss–Seidel alternating scheme on updating the coefficient $\alpha$ and the basis $B$:

$$\alpha^{n+1} = \arg \min_{a \in \mathbb{R}^{m \times k}, a \geq 0} \| AaB^nT - f \|_F^2 + \varepsilon \phi(a) + \lambda \psi \left( \alpha(B^n)^T \right)$$

$$+ \frac{L_1^n}{2} \|\alpha - \alpha^n\|_F^2 \quad (16a)$$

$$B^{n+1} = \arg \min_{B \in \mathbb{R}^{n \times k}, B \geq 0} \| A\alpha^{n+1}B^T - f \|_F^2 + \lambda \psi \left( \alpha^{n+1}B^T \right) + \sigma \varphi(B)$$

$$+ \frac{L_2^n}{2} \|B - B^n\|_F^2. \quad (16b)$$

Here $L_1^n \geq 0$, $L_2^n \geq 0$ are chosen parameters. Note that this scheme reduces to the usual alternating method if the two parameters $L_1^n = L_2^n = 0$. This scheme has been studied in [38] and [53] for functions with the so-called Kurdyka–Łojasiewicz property. Some convergence results can be obtained, as we will discuss in detail in section 4. In [41] and [46], the convergence of this scheme was also discussed for data driven dictionary learning models. In the following, we will focus on solving the two convex subproblems (16a) and (16b).
3.3.1. Update of the coefficients $\alpha$. For convenience, we drop the subscript $n$ for the fixed basis matrix $B^\circ$. Denote the set characteristic function of $C = \{ \alpha \mid \alpha \geq 0 \}$ as

$$\chi_C(\alpha) = \begin{cases} 0 & \text{if } \alpha \geq 0, \\ \infty & \text{else.} \end{cases}$$ (17)

The first subproblem can be rewritten as

$$\alpha^{n+1} = \arg\min_{\alpha} \| AaB^T - f \|_F^2 + \epsilon \| V\alpha \|_{l,2,1} + \delta \| \beta \|_{l,1,\infty} + \lambda \| V_i V (aB^T) \|_F^2$$

$$+ \chi_C(\alpha) + \frac{L_n}{2} \| \alpha - \alpha^n \|_F^2.$$ (18)

The optimization model can be solved by the alternating direction method of multipliers (ADMM) [54–56], also known as the split Bregman method [37] in imaging science. By introducing the auxiliary variables, the model is reformulated as:

$$\alpha^{n+1} = \arg\min_{\alpha} \| AaB^T - f \|_F^2 + \epsilon \| V\alpha \|_{l,2,1} + \delta \| \beta \|_{l,1,\infty} + \lambda \| V_i V (aB^T) \|_F^2$$

$$+ \chi_C(\beta) + \frac{L_n}{2} \| \alpha - \alpha^n \|_F^2.$$ (19)

The constraints can be written in the linear transform:

$$P\alpha = (V\alpha, \alpha)^T = (\nu, \beta)^T.$$ (20a)

Then (18) is a convex separable optimization problem with equality constraints and can be simplified as

$$\arg\min_{\alpha} D(\alpha) + R(\nu, \beta) \quad \text{s.t. } P\alpha = (\nu, \beta)^T,$$ (19)

where $D(\alpha) = \| AaB^T - f \|_F^2 + \lambda \| V_i V (aB^T) \|_F^2 + \frac{L_n}{2} \| \alpha - \alpha^n \|_F^2,$

$$R(\nu, \beta) = \epsilon \| \nu \|_{l,2,1} + \delta \| \beta \|_{l,1,\infty} + \chi_C(\beta).$$

For $\mu > 0$, the augmented Lagrangian function of the optimization problem (19) is defined as:

$$\mathcal{L}(\alpha, \nu, \beta; p) = D(\alpha) + R(\nu, \beta) + \langle p, P\alpha - (\nu, \beta)^T \rangle + \frac{\mu}{2} \| P\alpha - (\nu, \beta)^T \|_2^2$$

where $p = (p_1, p_2)^T$ is dual variable, with the same size as $(\nu, \beta)^T$. Note that we can also select a vector $\mu = (\mu_1, \mu_2)$ for the two constraints’ quadratic penalty, while we use the same scalar $\mu$ here for simplicity of notation.

Thus the scheme solving (18) by ADMM is as follows:

$$\alpha^{k+1} = \arg\min_{\alpha} \mathcal{L}(\alpha, v^k, \beta^{k+1}; p^k),$$ (20a)

$$\left(v^{k+1}, \beta^{k+1}\right) = \arg\min_{\nu, \beta} \mathcal{L}(\alpha^{k+1}, \nu, \beta; p^k).$$ (20b)

$$p^{k+1} = p^k + \mu \left( P\alpha^{k+1} - (v^{k+1}, \beta^{k+1})^T \right).$$ (20c)

where $k \in \mathbb{N}$ symbolizes the iteration steps of (20). The update for $p$ is straightforward. In the following, we present in detail the solution of the first two subproblems.
For subproblem (20a),
\[ \arg\min_\alpha \mathcal{L}(\alpha, v^k, \beta^k; p^k) = \arg\min_\alpha D(\alpha) + \left\langle p^k, \mathcal{P}\alpha - (v^k, \beta^k)^\top \right\rangle + \frac{\mu}{2} \| \mathcal{P}\alpha - (v^k, \beta^k)^\top \|^2_{\mathcal{F}}. \] (21)

It follows that
\[ \alpha^{k+1} = \arg\min_\alpha \| A\alpha B^\top - f \|^2_{\mathcal{F}} + \lambda \| \nabla \mathcal{F}(\alpha B^\top) \|^2_{\mathcal{F}} + \frac{L_k}{2} \| \alpha - \alpha^k \|^2_{\mathcal{F}} + \left\langle p_1^k, \nabla\alpha - v^k \right\rangle + \frac{\mu}{2} \| \nabla\alpha - v^k \|^2_{\mathcal{F}} + \left\langle p_2^k, \alpha - \beta^k \right\rangle + \frac{\mu}{2} \| \alpha - \beta^k \|^2_{\mathcal{F}}. \]

The problem is quadratic and differentiable on \( \alpha \) and the solution can be obtained by conjugate gradient (CG) method. 

To solve (20b) for the auxiliary variables \( v^{k+1} \) and \( \beta^{k+1} \):
\[ (v^{k+1}, \beta^{k+1}) = \arg\min_{v, \beta} \mathcal{L} \left( \alpha^{k+1}, v, \beta; p^k \right) \]
\[ = \arg\min_{v, \beta} R(v, \beta) + \left\langle p^k, \mathcal{P}\alpha^{k+1} - (v, \beta)^\top \right\rangle + \frac{\mu}{2} \| \mathcal{P}\alpha^{k+1} - (v, \beta)^\top \|^2_{\mathcal{F}}. \]

Recall that \( R(v, \beta) = \varepsilon \| v \|_{2,1} + \delta \| \beta \|_{1,\infty} + \chi_{\mathcal{C}}(\beta) \) is separable on \( v \) and \( \beta \). Then, \( v, \beta \) can be solved separately by
\[ v^{k+1} = \arg\min_v \mathcal{L} \left( \alpha^{k+1}, v, \beta; p^k \right), \] (22a)
\[ \beta^{k+1} = \arg\min_{\beta} \mathcal{L} \left( \alpha^{k+1}, v, \beta; p^k \right). \] (22b)

For solving (22a),
\[ v^{k+1} = \arg\min_v \mathcal{L} \left( \alpha^{k+1}, v, \beta; p^k \right) \]
\[ = \arg\min_v \varepsilon \| v \|_{2,1} + \frac{\mu}{2} \| v - v^k \|^2_{\mathcal{F}} \] (23)
where \( v^k := \frac{1}{\mu}(\mu \nabla\alpha^{k+1} + p_1^k) \). Recall that \( \| v \|_{2,1} \) is defined as (10) and the solution is obtained by soft shrinkage
\[ v_{t+1}^{(i, j)} = \max \left( \frac{\| v_t^{(i, j)} \|_2 - \frac{\varepsilon}{\mu}}{\| v_t^{(i, j)} \|_2}, 0 \right) \frac{v_t^{(i, j)}}{\| v_t^{(i, j)} \|_2} \]
\[ = \begin{cases} \frac{v_t^{(i, j)}}{\| v_t^{(i, j)} \|_2} & \text{if } \| v_t^{(i, j)} \|_2 > \frac{\varepsilon}{\mu} \\ \frac{v_t^{(i, j)}}{\| v_t^{(i, j)} \|_2} & \text{if } \| v_t^{(i, j)} \|_2 \leq \frac{\varepsilon}{\mu} \end{cases} \]
where \( t = 1, \ldots, K \).
To solve (22b), the subproblem is simplified as
\[ \beta^{k+1} = \arg \min_{\beta} \mathcal{L}(\alpha^{k+1}, \nu, \beta; p^k) \]
\[ = \arg \min_{\beta} \delta \sum_{j=1}^{K} \max_{i} \beta_{i,j} + \chi_{C}(\beta) + \frac{\mu}{2} \| \beta - \tilde{\beta}^{k} \|_{F}^{2} . \]
where \( \tilde{\beta}^{k} := \frac{1}{\mu}(\mu \alpha^{k+1} + p_{j}^{k}) \).

From the formulation above, we can see that \( \beta \) is separable by column. Thus we can solve for each column \( \beta_{j} \):
\[ \beta_{j}^{k+1} = \arg \min_{\beta_{j}} \delta \max_{i} \left( \beta_{i,j} \right) + \chi_{C}(\beta_{j}) + \frac{\mu}{2} \| \beta_{j} - \tilde{\beta}_{j}^{k} \|_{F}^{2} . \] (24)

The solution of this problem is given by Moreau decomposition [57–60]. The Moreau decomposition of a convex function \( J \) in \( \mathbb{R}^{n} \) is defined as
\[ x = \arg \min_{u \in \mathbb{R}^{n}} J(u) + \frac{1}{2\alpha} \| u - x \|_{2}^{2} + \alpha \arg \min_{p \in \mathbb{R}^{n}} J^{\ast}(p) + \frac{\alpha}{2} \| p - \frac{x}{\alpha} \|_{2}^{2} , \]
where \( J^{\ast}(p) \) is the conjugate function of \( J \) [58, 61]. Let \( J(\beta) \triangleq \delta \max_{i} (\beta_{i,j}) + \chi_{C}(\beta_{j}) \), then
\[ J^{\ast}(p) = \sup_{\beta_{j}} \left( \max_{i} \beta_{i,j} \right) \left( \| \max (p, 0) \|_{1} - \delta \right) \]
\[ = \left\{ \begin{array}{ll}
\infty & \text{otherwise} \\
0 & \text{if } \| \max (p, 0) \|_{1} \leqslant \delta .
\end{array} \right. \] (25)

From this result, we can see that \( J^{\ast}(p) \) is the set characteristic function of the convex set \( C_{\delta} = \{ p \in \mathbb{R}^{n}, \| \max (p, 0) \|_{1} \leqslant \delta \} \). By Moreau decomposition, \( \beta^{k+1} \) is then given by
\[ \beta^{k+1} = \tilde{\beta}^{k} - \Pi_{C_{\delta}}(\tilde{\beta}^{k}) , \]
where \( \Pi_{C_{\delta}}(\tilde{\beta}^{k}) \) is the orthogonal projection of each column of \( \tilde{\beta}^{k} \) onto \( C_{\delta} \) and \( \tau = \frac{\delta}{p} \).

The convergence of ADMM has been well studied. Numerically, we can check the primal and dual residual for the scheme (20) as the stopping criteria [62, 63]
\[ n_{k} = \left( \nu^{k} - V \alpha^{k} \right) , \quad d_{k} = \left( \frac{\mu}{2} \right) \left( \nu^{k} - \nu^{k-1} \right) , \]
\[ = \left( \frac{\mu}{2} \right) \left( \nu^{k} - \nu^{k-1} \right) . \]

3.3.2. Update of the basis \( B \). The update of the basis \( B \) is similar to the previous section. Let \( \alpha \) be the fixed nonnegative matrix \( \alpha^{n+1} \) in (16b). The second subproblem is rewritten as
\[ B^{n+1} = \arg \min_{B} \| AaB^{T} - f \|_{F}^{2} + \lambda \| V_{i} V_{aB}^{T} \|_{F}^{2} + \sigma \| V_{i}B \|_{F}^{2} + \chi_{C}(B) + \frac{L_{a}^{n}}{2} \| B - B^{n} \|_{F}^{2} . \]

Here we abuse the notation for the set characteristic function of the nonnegativity constraint for \( B \) and denote \( \| V_{i}B \|_{F}^{2} = \sum_{k=1}^{K} \sum_{t=1}^{T} (B(t + 1, k) - B(t, k))^{2} \) for simplicity of
presentation. Analogously, we use ADMM to solve the problem by introducing an auxiliary variable \( Q = B \). The problem is reduced to

\[
\arg\min_{B} D(B) + R(Q)
\]

s.t. \( B = Q \),

(27)

where

\[
D(B) = \| AaB^T - f \|_F^2 + \lambda \| V_i V_i^T \|_F^2 + \sigma \| V_i B \|_F^2 + \frac{L_2^2}{2} \| B - B^n \|_F^2,
\]

\[R(Q) = \chi_c(Q).\]

For conciseness of presentation, we ignore the detail and present the numerical scheme directly as follows:

\[
B^{k+1} = \arg\min_{B} \| AaB^T - f \|_F^2 + \lambda \| V_i V_i^T \|_F^2 + \sigma \| V_i B \|_F^2 + \frac{L_2^2}{2} \| B - B^k \|_F^2 + \langle \nu, B - Q^k \rangle + \frac{\nu}{2} \| B - Q^k \|_F^2. \tag{28a}
\]

\[
Q^{k+1} = \arg\min_{Q} \chi_c(Q) + \frac{\nu}{2} \| Q - \frac{1}{\nu} (\nu B^{k+1} + P^k) \|_F^2, \tag{28b}
\]

\[
P^{k+1} = P^k + \nu (B^{k+1} - Q^{k+1}). \tag{28c}
\]

where \( P \) is the dual variable.

Similarly, the subproblem (28a) is quadratic and can be solved by the conjugate gradient method. For (28b), define \( \frac{1}{\nu} (\nu B^{k+1} + Q^k) \) as \( \tilde{Q}^k \),

\[
Q^{k+1} = \arg\min_{Q} \chi_c(Q) + \frac{\nu}{2} \| Q - \tilde{Q}^k \|_F^2 = \begin{cases} \tilde{Q}^k & \text{if } \tilde{Q}^k \geq 0, \\ 0 & \text{else,} \end{cases} \]

\( \Delta = P_c \left( \tilde{Q}^k \right). \tag{29} \)

The primal and dual residual of the subproblem (16b) is as follows,

\[
r'_k = Q^k - B^k, \\
d'_k = \nu (Q^k - Q^{k-1}).
\]

Finally, we summarize the algorithm of SEMF model:

**Algorithm 1.**

**Step 1.** Initial value, \( \alpha_0, B_0, \nu_0, \beta_0, Q_0, \varepsilon, \delta, \lambda, \sigma, \mu, \nu, n = 0 \). The initial dual variables \( p_0, \beta_0 \) and the stop condition \( N, \delta_0 \)

while \( \| r_k \| > \delta_0, \| d_k \| > \delta_0, \| r'_k \| > \delta_0, \| d'_k \| > \delta_0 \) or \( n < N \) do

**Step 2. Update the coefficient \( \alpha \)**

while \( \| r_k \| > \delta_0, \| d_k \| > \delta_0 \) do

Solve (21) for \( \alpha^{k+1} \) by CG.

\[
\nu^{k+1} = \text{prox}_{\delta_0/2} \left( \nu^k + \frac{\mu}{2} \right), \text{ where } \nu^k = \frac{1}{\mu} (\nu \alpha^{k+1} + p_1^k).
\]

\[
\beta^{k+1} = \beta^k - \Pi_{\frac{\mu}{3}} \left( \tilde{\beta}^k \right), \text{ where } \tilde{\beta}^k = \frac{1}{\mu} (\mu\alpha^{k+1} + p_2^{k+1}).
\]

\[
p_1^{k+1} = p_1^k + \mu (V a^{k+1} - \nu^{k+1}).
\]

\[
p_2^{k+1} = p_2^k + \mu (a^{k+1} - \beta^{k+1}).
\]

endwhile

**Step 3. Update the basis \( B \)**

\[
\beta^k = \beta^k + \mu (V a^{k+1} - \nu^{k+1}).
\]

endwhile

11
while \( \| \mathbf{r}_k \| > \delta_0, \| \mathbf{d}_k \| > \delta_0 \)

Solve (28a) for \( \mathbf{B}^{k+1} \) by CG.

\[ Q^{k+1} = \mathcal{P}_L(\tilde{Q}^k), \]

where \( \tilde{Q}^k = \frac{1}{\nu} (\nu \mathbf{B}^{k+1} + \mathbf{P}^k) \).

end while

\[ n = n + 1. \]

end while

4. Convergence analysis

In this section, we will discuss the convergence of algorithm 1. In this whole section, we write

our problem (14) in a simplified form

\[ \min_{\alpha, \mathbf{B}} F(\alpha, \mathbf{B}) = \min_{\alpha, \mathbf{B}} f(\alpha, \mathbf{B}) + \eta(\alpha) + r_2(\mathbf{B}), \]

where \( f(\alpha, \mathbf{B}) \) is a function of \( \alpha \) and \( \mathbf{B} \), \( \eta(\alpha) \), \( r_2(\mathbf{B}) \) are functions of \( \alpha \) and \( \mathbf{B} \) respectively. In fact, the function \( \| \mathbf{A} \alpha \mathbf{B}^T - f \|^2_H + \| \mathbf{V} \alpha \mathbf{B}^T \|^2_H \) can be seen as \( f(\alpha, \mathbf{B}) \). The regularization terms on \( \alpha \), \( \mathbf{B} \) can be symbolized by \( \eta(\alpha), r_2(\mathbf{B}) \), respectively. It is easy to see that \( f(\alpha, \mathbf{B}) \) is a smooth function on \( \alpha \) and \( \mathbf{B} \) and \( \eta(\alpha) \) and \( r_2(\mathbf{B}) \) are convex on \( \alpha \) and \( \mathbf{B} \) respectively. Consequently, algorithm 1 can be simplified as:

**Algorithm 2. Proximal alternating minimization**

**Initialization:** \( a^0, B^0, L_1^0 > 0, L_2^0 > 0. \)

for \( k = 0, 1, 2 \ldots \) do

\[ a^{k+1} \in \text{arg min}_{a} F(a, B^k) + \frac{\nu_1}{2} \| a - a^k \|^2_H. \]

\[ B^{k+1} \in \text{arg min}_{B} F(a^{k+1}, B) + \frac{\nu_2}{2} \| B - B^k \|^2_H. \]

end for

Note that this algorithm has been recently studied in [38, 53] under the framework of the Kurdyka–Lojasiewicz property (see definition 2). The Kurdyka–Lojasiewicz property was first introduced by Lojasiewicz on the real analysis function [64], and Kurdyka extended the property to the functions on o-minimal structure [65]. The property was recently extended to the nonsmooth subanalytic function [66].

In the following, we summarize some convergence results of algorithm 1. Denote \( Z := (\alpha, \mathbf{B}) \) and \( Z^k := (\alpha^k, \mathbf{B}^k), \forall k \geq 0. \) We first present some notations and preliminaries.

**Definition 1.** Let \( f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{ \infty \} \) be a proper lower semicontinuous function.

(I) (Fréchet subdifferential) For each \( x \in \text{dom}(f) \), the Fréchet subdifferential of \( f \) at \( x \) is the set of vectors \( x^* \in \mathbb{R}^n \), which satisfies

\[ \lim_{y \rightarrow x, y \neq x} \inf \frac{1}{\| x - y \|} \left[ f(y) - f(x) - \left\langle x^*, y - x \right\rangle \right] \geq 0 \]

and can be written as \( df \).
(II) (limiting-subdifferential [67]) The limiting-subdifferential of $f$ at $x \in \text{dom}(f)$, written as $$\partial f(x) := \left\{ x^* \in \mathbb{R}^n : \exists x_n \to x, f(x_n) \to f(x), x_n^* \in \tilde{\partial} f(x_n) \to x^* \right\}.$$ 

**Remark 1.**

(a) From the definition we can conclude that $\partial f(x)$ is a closed set. In fact, let $(x_k, x_k^*)_{k \in \mathbb{N}}$ be a sequence such that $(x_k, x_k^*) \in \text{Graph} \partial f$ for all $k \in \mathbb{N}$. If $(x_k, x_k^*) \to (x, x^*)$ and $f(x_k) \to f(x)$, then $(x, x^*) \in \text{Graph} \partial f$.

(b) If $0 \in \partial f(\tilde{x})$, $\tilde{x}$ is a critical point of $f(x)$.

**Definition 2 (The Kurdyka–Łojasiewicz (KL) property).** (a) The function $f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is said to have the Kurdyka–Łojasiewicz property at $x^* \in \text{dom} \partial f$ if there exists $\eta \in (0, +\infty)$, a neighborhood $U$ of $x^*$ and a continuous function $\varphi: [0, \eta) \to \mathbb{R}^+$ such that:

(I) $\varphi(0) = 0$;

(II) $\varphi$ is $C^1$ on $(0, \eta)$;

(III) for all $s \in (0, \eta)$, $\varphi'(s) > 0$;

(IV) for all $x \in U \cap \{f(x^*) < f < f(x^*) + \eta\}$, the Kurdyka–Łojasiewicz inequality holds

$$\varphi'(f(x) - f(x^*)) \text{dist}(0, \partial f(x)) \geq 1.$$ 

(b) If $f$ satisfies the KL property at each point of $\text{dom}(\partial f)$, then $f$ is called a KL function.

For all semi-algebraic functions that are KL functions, we give the definition of a semi-algebraic function.

**Definition 3 (Semi-algebraic sets and functions).** (a) A subset $S$ of $\mathbb{R}^n$ is a real semi-algebraic set if there exists a finite number of real polynomial functions$ g_{i,j}, h_{i,j}: \mathbb{R}^n \to \mathbb{R}$ such that

$$S = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \left\{ x \in \mathbb{R}^n : g_{i,j}(x) = 0 \quad \text{and} \quad h_{i,j}(x) < 0 \right\}.$$ 

(b) A function $f: \mathbb{R}^n \to (-\infty, +\infty)$ is called semi-algebraic if its graph,

$$\left\{ (x, t) \in \mathbb{R}^{n+1} : f(x) = t \right\},$$

is a semi-algebraic subset of $\mathbb{R}^{n+1}$.

**Lemma 1 (square summable $\|Z^k - Z^{k-1}\|$).** Let $\{Z^k\}_{k \in \mathbb{N}}$ be a sequence generated by algorithm 2 and $0 < l_1 \leq L_1^k \leq L_1 < \infty$, $0 < l_2 \leq L_2^k \leq L_2 < \infty$. Then the sequence $F(Z^k)_{k \in \mathbb{N}}$ is non-increasing and satisfies
where \( l = \min \{ l_1, l_2 \} \). Furthermore, we can deduce
\[
\sum_{k=1}^{\infty} ||Z_k - Z_{k-1}||^2 < \infty.
\]

**Remark 2.** From lemma 1, we conclude that the objective function is decreasing for \( l > 0 \).
This is theorem 3.1 in [38] and the property with \( l > 0 \) is called sufficient decrease.

**Lemma 2.** \( Vf(\alpha, B) \) is \( L_G \)-Lipschitz continuous on a bounded set.

**Proof.** In the proposed SEMF model, the function \( f(\alpha, B) = \|A\alpha B^T - f\|^2_F + \lambda \|V\alpha\|_F^2 \) is a smooth function. The gradient of \( f(\alpha, B) \) is as follows:
\[
Vf(\alpha, B) = \left( V_{\alpha} f(\alpha, B), V_{B} f(\alpha, B) \right) = \left( 2A\alpha A^T + 2\lambda V^T V\alpha V, \alpha^T (2A\alpha A^T + 2\lambda V^T V\alpha V) \right).
\]
\( Vf(\alpha, B) \) is Lipschitz constant on any bounded set. In other words, for any bounded set \( S \), there exists a constant \( L_G > 0 \) such that for any \( \{ (\alpha_1, B_1), (\alpha_2, B_2) \} \subseteq S \),
\[
\| Vf(\alpha_1, B_1) - Vf(\alpha_2, B_2) \| \leq L_G \| (\alpha_1, B_1) - (\alpha_2, B_2) \|.
\]
Then, \( V_{\alpha} f(\alpha, B) \) is also \( L_G \)-Lipschitz continuous on the bounded set.

**Lemma 3 (Relative error).** \( \{ Z_k \}_{k \in \mathbb{N}} \) is a sequence generated by algorithm 2 and \( V_{\alpha} f \) is \( L_G \)-Lipschitz continuous on a bounded set.
\[
\| V_{\alpha} f(\alpha_1, B) - V_{\alpha} f(\alpha_2, B) \| \leq L_G \| \alpha_1 - \alpha_2 \|, \quad \forall B \in \text{dom}(f).
\]
Then we have
\[
\text{dist}(0, \partial f(\alpha^*, B^*)) \leq (L + L_G) \| Z^k - Z^{k-1} \|,
\]
where \( L = \max \{ L_1, L_2 \} \).

**Proof.** See [39] theorem 3.4.

**Lemma 4.** In our SEMF model, \( F(\alpha, B) = \|A\alpha B^T - f\|^2_F + \epsilon \phi(\alpha) + \lambda \psi(\alpha B^T) + \varphi(B) + \chi_C(\alpha) + \chi_C(B) \) satisfies the KL property.

**Proof.** For \( \|A\alpha B^T - f\|^2_F \), \( \|V\alpha B^T\|^2_F \) and \( \|V_i B\|^2_F \) are real polynomial functions, \( \|A\alpha B^T - f\|^2_F \), \( \psi(\alpha B^T) \) and \( \varphi(B) \) are semi-algebraic functions [39].

The nonnegative orthant \( \mathbb{R}_{+}^{M \times K} = \{ \alpha \in \mathbb{R}^{M \times K}; \alpha_{ij} \geq 0, \forall i, j \} \), \( \mathbb{R}_{+}^{K} = \{ B \in \mathbb{R}^{M \times K}; B_{ij} \geq 0, \forall i, j \} \) are semi-algebraic sets. For indicator functions of semi-algebraic sets that are semi-algebraic, \( \chi_C(\alpha) \) and \( \chi_C(B) \) are semi-algebraic functions.

Inverse Problems 31 (2015) 025004 Q Ding et al

14
For the function \( \phi_\alpha \varepsilon \alpha \lambda \delta \alpha \), we denote \( \nu = V \alpha \), then, 
\[
\| v \|_{2,1} = \sum^K_{i=1} \sum^N_{j=1} v^2_{i,j} + v^2_{2,j}.
\]
The graph of \( \sqrt{v^2_{i,j} + v^2_{2,j}} \) is
\[
\text{Graph}\left( \sqrt{v^2_{i,j} + v^2_{2,j}} \right) \in \mathbb{R}^3,
\]
\[
= \left\{ (v_{1,j}, v_{2,j}, t) \in \mathbb{R}^3 : \sqrt{v^2_{i,j} + v^2_{2,j}} = t \right\}.
\]
It is easy to know that \( \text{Graph}(\sqrt{v^2_{i,j} + v^2_{2,j}}) \) is a semi-algebraic set. For finite sums and products of semi-algebraic functions that are semi-algebraic functions, we know \( \| v \|_{2,1} \) is a semi-algebraic function. Then, we conclude that \( \| v \|_{2,1} \) is a semi-algebraic function for the composition of semi-algebraic functions or mappings are semi-algebraic. For \( \| \alpha \|_{1,\infty} = \sum^K_{j=1} \max \alpha_{i,j} \), the graph of \( \max \alpha_{i,j} \) is
\[
\text{Graph}\left( \max \alpha_{i,j} \right) = \left\{ (\alpha(\cdot, j), t) : \max \alpha_{i,j} = t \right\} = \bigcup_{t \in \mathbb{N}^+} \left\{ (\alpha(\cdot, j), t) : \alpha_{i,j} = t, \alpha_{k,j} < t, \forall k \neq i \right\},
\]
which is a semi-algebraic set. \( \| \alpha \|_{1,\infty} \) is a semi-algebraic function since it is finite sums of semi-algebraic functions.

For all of the above, \( F(\alpha, B) \) is the finite sum of the semi-algebraic functions and satisfies the KL property.

**Theorem 4** (Lemma 2.6 [40]). \( \{Z^k\} \subseteq \mathbb{N} \) is a sequence generated by proximal alternating minimization and \( 0 < l_1 \leq L_1 \leq L_1 < \infty, 0 < l_2 \leq L_2 < \infty \). Moveover, the following hold:

(i) \( F \) is \( \lambda L \)-Lipschitz continuous on a bounded set;

(ii) \( F \) satisfies the KL inequality at \( Z \);

(iii) \( Z_0 \) is sufficiently close to \( Z \), and \( F(Z^k) > \bar{F} \) for \( k > 0 \).

Then there exists \( N(\bar{Z}, \rho) \subseteq U \) \( \cap \) dom(\( \partial F \)), such that \( \{Z^k\} \subseteq N \) and \( Z^k \to Z^* \), for \( k \to \infty \) and

\[
\sum^\infty_{k=0} \| Z^{k+1} - Z^k \| < \infty,
\]

\[
F(Z^k) \to F(\bar{Z}), \text{ as } k \to \infty,
\]
where \( U \) is defined in definition 2 relative to \( F \) and \( \bar{Z} \).

**Theorem 5** (convergence to a critical point). Let \( \{Z^k\} \subseteq \mathbb{N} \) be a sequence generated by algorithm 2 and \( 0 < l_1 \leq L_1 \leq L_1 < \infty, 0 < l_2 \leq L_2 \leq L_2 < \infty \). \( Z_0 \) is sufficiently close to \( \bar{Z} \), and \( F(Z) > \bar{F} \) for \( k > 0 \). Let \( \rho > 0 \) such that \( N(\bar{Z}, \rho) \subseteq U \), then \( \{Z^k\} \subseteq N \) and \( Z^k \) converges to \( Z^* \in N \). Furthermore, if \( Z^k \) is bounded, then \( Z^k \) converges to a critical point \( Z^* \) of \( F(Z) \).
Proof. From lemma 4, our SEMF model satisfies the KL property and from lemma 2, $V f'$ is $L_G$-Lipschitz continuous on a bounded set. Then by theorem 4, we can conclude that $\mathcal{N}(Z, \rho) \in U \cap \text{dom}(\partial F)$, such that $\{Z^k\} \subset \mathcal{N}$ and $Z^k$ converges to $Z^* \in \mathcal{N}$.

From above, $Z^* = (\alpha^*, B^*)$ is a limit point of $\{Z^k\}_{k \in \mathbb{N}} = \{(\alpha^k, B^k)\}_{k \in \mathbb{N}}$. Then, there exists a subsequence $\{(\alpha^{k_i}, B^{k_i})\}_{i \in \mathbb{N}}$ such that $(\alpha^{k_i}, B^{k_i}) \to (\alpha^*, B^*)$. For our problem, we can easily obtain that $\lim_{\eta \to \infty} F(\alpha^{k_i}, B^{k_i}) = F(\alpha^*, B^*)$. Furthermore, by the property of relative error in lemma 3, we know that $\partial F(\alpha^{k_i}, B^{k_i}) \to (0, 0)$. For $\partial F$ is a close convex set and by the definition limiting-subdifferential, we can get $(0, 0) \in \partial F(\alpha^*, B^*)$. Consequently, $Z^*$ is a critical point of $F(Z)$.

Define the difference measure of two sets $\mathcal{X}, \mathcal{Y}$

$$\text{diff}(\mathcal{X}, \mathcal{Y}) = \max \left( \sup_{a \in \mathcal{X}} \inf_{B \in \mathcal{Y}} \|a - B\|, \sup_{B \in \mathcal{Y}} \inf_{a \in \mathcal{X}} \|a - B\| \right).$$

The set of feasible points of $F(\alpha, B), \mathcal{X}$, is a closed and block multiconvex subset of $\mathbb{R}^n$, which can be decomposed into two blocks. We define

$$\mathcal{X}_1(B) \triangleq \{\alpha: (\alpha, B) \in \mathcal{X}\}, \quad \mathcal{X}_2(\alpha) \triangleq \{B: (\alpha, B) \in \mathcal{X}\}.$$

Definition 6 (Nash point). A point $\bar{x}$ is called the Nash point of a proper lower semicontinuous function $f(x_1, x_2, \cdots, x_s): \mathbb{R}^n \to \mathbb{R}$, if for $i = 1, \cdots, s$ and feasible point set $\mathcal{X}$,

$$f(\bar{x}_1, \cdots, \bar{x}_{i-1}, \bar{x}_i, \bar{x}_{i+1} \cdots \bar{x}_s) \leq f(\bar{x}_1, \cdots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1} \cdots \bar{x}_s) \quad \forall x_i \in \bar{X}_i$$

or equivalently,

$$\partial_i f(\bar{x}), x_i - \bar{x}_i \geq 0 \quad \forall x_i \in \bar{X}_i,$$

where $\bar{X}_i = \{x_i \in \mathbb{R}^{n_i}: (\bar{x}_1, \cdots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1} \cdots \bar{x}_s) \in \mathcal{X}\}$ and $n_i$ is the dimension of the $i$th block.

Theorem 7. If $Z^k, Z \in \mathcal{X}$ and $Z^k \to Z$ implies

$$\lim_{k \to \infty} \text{diff}(\mathcal{X}_1(B^k), \mathcal{X}_1(B)) = 0 \lim_{k \to \infty} \text{diff}(\mathcal{X}_2(\alpha^k), \mathcal{X}_2(\alpha)) = 0.$$

Then if $\|Z^{k+1} - Z^k\| \to 0$ holds, any limit point of $\{Z^k\}_{k \in \mathbb{N}}$ is a Nash point.

Proof. See theorem 2.3 in [40].

Since each subproblem of algorithm 1 is convex, we can also conclude that

Theorem 8 (limit point is a Nash point). Let $\{Z^k\}_{k \in \mathbb{N}}$ be a sequence generated by algorithm 2. If $Z^k$ is bounded, then any limit point of $Z^*$ is a Nash point.

Proof. Let $Z^*$ be a limit point of $Z^k$, then there exists a subsequence $Z^{k_i} \to Z^*$. For convenience, we write $Z^{k_i}$ as $Z^k$. We can get that
\[ \lim_{k \to \infty} \text{diff} \left( \mathcal{X}_1(B^k), \mathcal{X}_1(B) \right) = \text{diff} \left( \alpha^k, \alpha^s \right) = 0, \]
\[ \lim_{k \to \infty} \text{diff} \left( \mathcal{X}_2(a^k), \mathcal{X}_2(\alpha) \right) = \text{diff} \left( B^k, B^s \right) = 0. \]

From lemma 1, \( \sum_{k=1}^{\infty} \| Z^k - Z^{k-1} \|^2 < \infty \), then we have \( \| Z^k - Z^{k-1} \| \to 0 \). Referring to theorem 7, we obtain the conclusion that any limit point of \( Z^k \) is a Nash point.

5. Numerical results

5.1. 2D dynamic images with static boundary

Now, we present numerical results to verify our model and algorithm. The algorithm is tested on numerical phantoms for a proof of concept study. We simulate 90 image frames of size 64 \( \times \) 64 and two projections per frame. Three time activity curves (TAC) for blood, liver and myocardium, previously used in [18] (see figure 1), are used to simulate the dynamic images. The first simulated dynamic phantom is composed of two ellipses. In the temporal direction, the positions of the two ellipses are stationary while the intensity in 90 frames within the region of each ellipse is generated according to the TAC of blood or liver. Then the projections are generated by Radon transforms, sequentially performed for each frame.

We compare our method with the filtered back projection (FBP) method, the solutions by solving two least square models
\[ \alpha \parallel - \parallel \alpha ^\top B f \arg \min_F \| Au - f \|^2_B, \]
with a given fixed B-spline basis \( B \).

As for the initial values of \( \alpha \) and \( B \), we use 16 uniform B-spline bases to solve
\[ \arg \min_{\alpha} \| AaB^T - f \|^2_B, \]
where the knots of the B-spline basess are uniformly distributed over 0–90 s. Then, the solution \( \alpha \) obtained by solving \( \arg \min_{\alpha} \| AaB^T - f \|^2_B \) and 16 uniform B-spline bases, \( B \), are used as the initialization of \( \alpha \) and \( B \) to solve our SEMF model.

In the first test, projections at two orthogonal angles are simulated for every frame to mimic two-head camera data collection. The projection angles increase sequentially by 1° along the temporal direction. For example, at frame 1, projections are simulated at angle 1° and 91°, and at frame 2, angle 2° and 92°, etc. Finally, 10% white Gaussian noise is added to the projection data. Reconstruction results using different methods are shown in figure 2.

Since the number of projections is very limited for each frame, the traditional FBP and least
Figure 2. From top to bottom: true image frames; reconstructed by FBP; a solution solving $\text{argmin}_f \| AU - f \|^2_F$; a solution using a fixed basis $B$, solving $\text{argmin}_f \| AaB^T - f \|^2_F$; reconstructed by our method.

Figure 3. The TACs of blood and liver. (a) is the TAC of blood and (b) is the TAC of liver. The solid lines are the true TACs and the dash lines are the reconstructed TACs.
Figure 4. The TACs used in the rats’ liver image.

Figure 5. From top to bottom: true image frames; reconstruction by FBP; a solution solving $\min_u \|AU - f\|_2^2$; a solution using a fixed basis $B$, solving $\min_u \|AuB^T - f\|_2^2$; reconstructed by our method.
squares methods cannot reconstruct the images satisfactorily, while the proposed method is capable of reconstructing the images effectively.

Figure 3 illustrates the comparison of the TACs of blood and liver. The solid lines are the true TACs and the dash lines are the ones extracted from the reconstruction images by our method. Due to the weak signal and fast change of radioisotope at the beginning frames in the liver TAC, the reconstructed one does not fit closely to the true one.

The second numerical experiment is performed on a synthetic image simulating a rat’s abdomen, where the bright region represents the liver of a rat. We use the TAC in figure 4 to simulate the dynamic images. The same procedure is applied to generate projection data. Figure 5 compares the frames reconstructed by different methods. Clearly, the traditional FBP method and least squares method cannot reconstruct the dynamic images with very few
Figure 8. The objective function of the model. The left is the first numerical experiment and the right is the second one.

Figure 9. The result of two-dimensional images with one projection angle. The images are reconstructed using SEMF and every frame has one projection uniformly sampled from $[0, 180]$. The first row is the result of ellipse and the second row is the result of the rat’s liver.

Figure 10. The result of two-dimensional images with the projection angular space $[30, 150]$. The images are reconstructed by our SEMF model and every image has two projection angles uniformly sampled from $[30, 150]$. The first row is the result of the ellipse and the second row is the result of the rat’s liver.
projections; however, the proposed method reconstructs the images quite accurately. Similar to the ellipse phantom, figure 6 illustrates the comparison of the true TACs and those reconstructed by the SEMF method. We can see that two of them are quite accurate while the other two present relatively larger errors.

Figure 7 demonstrates the relative error \( \| U_{\text{true}} - U_{\text{rec}} \| \) for the \( t \)th frame, where \( U_{\text{rec}} \) is the reconstructed frame by the proposed method and \( U_{\text{true}} \) is the ground truth image. We can see that the relative error is larger for the first frames. This is partly due to the fact that the intensity of the image is in a much smaller scale and the decay of activity is more significant for these frames. Figure 8 demonstrates the evolution of the total energy of the model. It has been proved that the energy is decreasing in lemma 1 and numerically we can observe the global decrease with small numerical errors.

In the following, we test the proposed method on two more challenging settings: one is with only one projection per frame and the other is with limited projection angles which does not cover the whole range of \([0, 180]\). For the first test, the total number of projections is 90 compared to 180 in the previous tests for 90 frames, while the angles are uniformly sampled from the angular range \([0, 180]\); and for the second test there are two projections per frame with angular distance \(60^\circ\) and the angles are uniformly sampled from the range \([30, 150]\). We add a further 1% noise to the projection data of the ellipse and 10% noise to the one of the rat’s liver, respectively. Figures 9 and 10 show the reconstruction results on the two test images by the SEMF method, from one-projection-per-frame and partial angles settings respectively. We can see that with even less data, our method can recover most of the information, especially for those frames at the stable stage. In figure 10, the images suffer more stripe effects due to the limited angles, while the default is partially overcome by the symmetricity for the ellipse image.

5.2. 2D dynamic images with moving boundaries

In practice, the boundary of the observed organ is moving with time due to the respirator motion. In this section, we verify the feasibility of the proposed model for such a scenario. Similar to the previous experiment, we simulate 90 frames of images composed of two ellipses with the intensity changing according to the TACs. Furthermore, we simulate a periodic motion for the boundary of the ellipses. In particular, the semi-major and semi-minor axes of the ellipses change cyclically. The area variation of the two ellipses is shown in figure 11. The projection data of each image in the two angles are simulated and 10% white Gaussian noise is also added.

\[ \text{Figure 11. The area variation of the two ellipses by time.} \]
The comparison of the results of the different reconstruction methods are shown in figure 12. We can see that our model can reconstruct the images even with motion present, while the other methods fail. For this test case, similar curves can be obtained for the the TACs, relative error and the decreasing energy as shown in figures 3, 7 and 8.

6. Conclusion

In this paper, we presented a novel reconstruction model for dynamic SPECT that combines low rank decomposition and temporal and spatial regularization. The model greatly reduces the degree of freedom of the solution, which allows the reconstruction of the dynamic images with very few and incomplete projections. We also provide a convergence analysis for the alternating scheme to solve the proposed nonconvex model. Our reconstruction results on a 2D phantom with three examples, two with static boundaries and one with moving boundaries, indicate that our algorithm outperforms the conventional FBP-type reconstruction.
algorithm and least squares method. This study has verified the feasibility of such a regularization model for dynamic image reconstruction from few projections, and allows for the future implementation and adaptation to real 3D reconstruction with Poisson noise.

Acknowledgments

We are grateful to the anonymous reviewers for their valuable comments and advice. Q Ding and X Zhang were partially supported by the NSFC (grant nos. NSFC11101277 and NSFC91330102) and the 973 program (no. 2015CB856004). Y Zan and Q Huang were partially supported by the NSFC (grant nos. NSFC81201114 and NSFC91330102) and the Innovation Program of Shanghai Municipal Education Commission (grant no. 13ZZ017).

References

[40] Xu Y and Yin W 2013 A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion SIAM J. Imag. Sci. 6 1758–89
[58] Lojasiewicz S 1993 Sur la géométrie semi-et sous-analytique Annales de l’institut Fourier vol 43 (Grenoble: Institut Fourier) pp 1575–95
[60] Bao C, Ji H, Quan Y and Shen Z 2013 L0 norm based dictionary learning by proximal methods with global convergence Proc. IEEE Conf. on Computer Vision and Pattern Recognition pp 3858–65