Inverse Problems and Imaging doi:10.3934/ipi.2020027
Volume 14, No. 4, 2020, 583–606

IMAGE RESTORATION FROM NOISY INCOMPLETE FREQUENCY DATA BY ALTERNATIVE ITERATION SCHEME

Xiaoman Liu

School of Mathematics, Southeast University
Nanjing, 210096, China
College of Sciences, Nanjing Agricultural University
Nanjing, 210095, China

Jijun Liu*

School of Mathematics/S.T.Yau Center of Southeast University, Southeast University
Nanjing, 210096, China

(Communicated by Xiaoqun Zhang)

Abstract. Consider the image restoration from incomplete noisy frequency data with total variation and sparsity regularizing penalty terms. Firstly, we establish an unconstrained optimization model with different smooth approximations on the regularizing terms. Then, to weaken the amount of computations for cost functional with total variation term, the alternating iterative scheme is developed to obtain the exact solution through shrinkage thresholding in inner loop, while the nonlinear Euler equation is appropriately linearized at each iteration in exterior loop, yielding a linear system with diagonal coefficient matrix in frequency domain. Finally the linearized iteration is proven to be convergent in generalized sense for suitable regularizing parameters, and the error between the linearized iterative solution and the one gotten from the exact nonlinear Euler equation is rigorously estimated, revealing the essence of the proposed alternative iteration scheme. Numerical tests for different configurations show the validity of the proposed scheme, compared with some existing algorithms.

1. **Introduction.**

Image restorations from incomplete measurement data are typically ill-posed mathematically, in the sense that the restoration is generally neither unique nor stable. In order to overcome this ill-posedness, the regularizing penalty terms should be added into the image restoration models in some suitable way.

Let $f = (f_{m,n}) \in \mathbb{R}^{N\times N}$ be the gray scale matrix for an image in the domain $\Omega := \bigcup_{m,n=1}^{N} \Omega_{m,n}$, with $f_{m,n} \in [0,1]$ the constant grey value at each pixel $\Omega_{m,n}$. The image restoration from the noisy measurement data g^δ requires some generalized solution to

$$\mathcal{K}[f] = g^\delta, \tag{1}$$

where $\mathcal{K}$ is some linear operator representing the known blurring process for exact image f, and g^δ are specified noisy data, including either additive noise or multiplied noise as well as Possion noise.

2010 *Mathematics Subject Classification.* Primary: 65M32, 68U10; Secondary: 65K10, 65T50.
Key words and phrases. Image restoration, alternating iteration, regularization, optimization, convergence.

* Corresponding author: Prof. Dr. Jijun Liu.

Generally, the noisy measurement data may be incomplete, leading to the non-unique reconstruction on the desired image f. Moreover, in some engineering areas such as signal transmission and MRI, the incomplete measurement data may be specified in frequency domain in terms of some linear sampling operator $\mathcal{P}$. Then (1) is replaced by

$$\mathcal{P}\mathcal{F}[f] = \mathcal{P}[\hat{g}^\delta], \tag{2}$$

where $\hat{g}^\delta$ is the noisy frequency data satisfying $\|\hat{g}^\delta - \hat{g}\| \le \delta$, $\mathcal{K} := \mathcal{F}$ is the two-dimensional discrete Fourier transform (DFT) operator converting the spatial matrix f into frequency matrix $\mathcal{F}[f] = F^T f F := \hat{f} \in \mathbb{C}^{N\times N}$, where $F = (e^{-i\frac{2\pi}{N}mn}) \in \mathbb{C}^{N\times N}$ is the unitary Fourier transform matrix. $\mathcal{P}$ represents some linear sampling operator projecting the full frequency data in $\mathbb{C}^{N\times N}$ into a lower dimensional space $\mathbb{C}^{M_1\times M_2}$ with $M_1, M_2 \ll N$, generating incomplete data.

In many situations, the number of salient features hidden in an image is much fewer than its resolution, which means that the image f is usually sparse or compressible under some suitable basis by compressive sensing (CS) theory [5, 6, 9]. In other words, by the standard arguments in CS such as the wavelet expansions of a signal, the original image could be considered as sparse. Due to the high Nyquist sampling rate in wide applications such as in digital images and dynamic images, the compressions for an image are necessary prior to its storage and transmission [29], which leads to the sparse representation of an image. The basic CS theory [6, 9] has justified that it is of high probability to reconstruct an image signal accurately from its sparse or compressible information.

In the mathematical aspect, a sparse signal can be reconstructed from its small projections onto certain subspace with $K(K \ll N)$ dominant components. The sparse signal is called $K-$sparse under the basis Ψ, if the number of nonzeros in the sparse signal is no more than K. To recover the sparsity of an image, the most useful method is to minimize $\|\cdot\|_{l^1}$ approximately [32, 34]. Considering the piecewise smooth property of an image, the restoration model with double regularization terms based on (2) and CS theory is considered in a recent work [14], by minimizing the unconstraint cost functional

$$J_\alpha^{gen}(f) := \frac{1}{2}\left\|\mathcal{P}\mathcal{F}[f] - \mathcal{P}[\hat{g}^\delta]\right\|_F^2 + \alpha_1\|\Psi^{-1}[f]\|_{l^1} + \alpha_2|f|_{TV} \tag{3}$$

with regularizing parameters $\alpha := (\alpha_1, \alpha_2) > 0$, where $\|\cdot\|_F$ is the Frobenius norm of an $N\times N$ matrix, and Ψ is a known $N\times N$ matrix determined by the base function $\{\psi_{m,n} : m, n = 1, \cdots, N\}$, mapping the sparse representation $\tilde{\mathbf{f}} \in \mathbb{R}^{N^2\times 1}$ into $f \in \mathbb{R}^{N\times N}$ by $\tilde{\mathbf{f}} = \mathbf{vect}[\tilde{f}]$ for $\tilde{f} := (\tilde{f}_{m,n}) \in \mathbb{R}^{N\times N}$ with $\tilde{f}_{m,n} = \langle f, \psi_{m,n}\rangle_{\mathbb{R}^{N\times N}}$ [8]. The operator **vect** maps any $K\times L$ matrix (not necessarily $K = L$) as a $KL-$dimensional column vector by standard way. We would like to emphasize that two penalty terms in (3) are necessary from the point of view of image process, since we should consider both the sparsity and the piecewise smoothness of an image. Of course, the cost functional with two penalty terms requires large amount of computations as compared with that with only one penalty term.

The sparse basis Ψ is difficult to specify for expanding a general image, which is another important topic for image representation, even though the orthogonal (biorthogonal) basis can be applied. However, in sparse representation of an image, the specified bases like Danbechies wavelet bases can be considered as a bounded operator mapping the image to the expansion coefficients, and consequently the norm $\left\|\Psi^{-1}[f]\right\|_{l^1}$ has similar behavior to $\|\mathbf{vect}[f]\|_{l^1}$ in the cost functional theoretically, although their values are different. Hence, for optimization procedure, with some modified parameters $\alpha_1, \alpha_2 > 0$, instead of the functional (3), we consider

$$J_\alpha^{sim}(f) := \frac{1}{2}\left\|\mathcal{P}\mathcal{F}[f] - \mathcal{P}[\hat{g}^\delta]\right\|_F^2 + \alpha_1\|\mathbf{vect}[f]\|_{l^1} + \alpha_2|f|_{TV} \tag{4}$$

directly for the simplicity of our theoretical analysis.

The image restoration models with multiple regularizing terms have obtained considerable attentions, with main focuses on the efficient implementations on optimizing (4). Due to the non-differentiability of the penalty terms, the optimization of (4) is realized by solving its convex and smooth surrogates. The most general smooth approximation to the penalty term is the Charbonnier approximation which adds small perturbation on absolute value function $|x|$ [24], while the surrogate $|x|_\alpha$ called smooth L^1 approximation is proposed to approach $|x|$ in [20], which takes advantage of the non-negative projection operator $(x)_+ := \max\{x, 0\} \approx p(x,\alpha) = x + 1/\alpha \ln(1 + e^{-\alpha x})$. Since the approximation $|x|_\alpha := p(x,\alpha) + p(-x,\alpha) \to |x|$ requires $\alpha \to \infty$, the optimization algorithm is hard to implement in numerics.

Instead of using the surrogate functions to deal with the non-smooth penalty terms like TV, some implementable and efficient algorithms were proposed, for example, the weighted total variation measure TV_{l_1} which exploited the continuity and sparsity simultaneously in the partial gradient domain [22]; the Chambolle's projection method [7] where the cost function in terms of dual variables is often continually differentiable by the duality-based approach. Although the algorithms of iterative type with one regularizing term are popular for CS algorithms, the strict math theory about the convergence property are quite limited, to say nothing of the optimization for the cost functional with multiple penalty terms. Moreover, to implement the iterative schemes efficiently, some linearized process are introduced, which will lead to extra errors for the algorithms.

There have been many schemes to deal with the non-differentiability of the penalty term, such as the Gauss-Seidel algorithm [21], the Grafting algorithm [18], the shooting algorithm [10], and the Bregman iterations [11, 17, 30, 31], which belong to the category of sub-gradient method. In the recent work, we did the comparison between the Bregman iteration and the direct method (DM) proposed in [14] and gave the benefits of DM.

For the implementations of image restorations based on the optimization of regularizing cost functionals, the main schemes are iteration algorithms for finding the minimizer of approximately. Except for the extensive studies on the efficient realizations of the restoration algorithms, the convergence properties of the iterative process have also been considered. In [25], the authors establish strong convergence properties for fast TV algorithm for image denoising, including finite convergence for some variables and $q-$linear convergence for the others. They also give the similar convergence for multichannel image restoration model called multi-channel TV (MTV) [26], and the model with MTV-L^1 data-fitting term [28]. In [27], a splitting operator method on two penalty terms (TV and L^1 based on orthogonal wavelet frame) has been proposed with the convergence analysis for each variables. In [19], the authors give the convergence of the proposed iteration scheme for image restoration model with two different sparse frames, and prove that this method can be understood as special cases of the Douglas-Rachford split algorithm. However, the convergence analysis on the iteration process for minimizing the cost functionals in image restorations is still in its initial stage, since the convergence depends heavily on both the cost functional and also the algorithms for finding the solution.

In this paper, we consider two smooth approximations for the penalty terms for our algorithm implementations, namely, the Charbonnier approximation and Huber approximation. Our main contributions contain three points. Firstly, we derive the image restoration model with multi-regularization and different approximation surrogates. Secondly, we propose the alternating iterative scheme, where the shrinkage soft-thresholding process is carried out in inner iterative step, and the linearizing version of the nonlinear Euler equation in outer iteration is established, for which the diagonal coefficient matrix is derived for efficiently solving the iterative solution. Finally, we prove the convergence property of the linearized alternative iteration scheme rigorously, the error between the linearized iterative solution and the one from solving the nonlinear equation is also established. Numerical implementations are presented to show the validity of our proposed scheme in section 4, by comparing our results with the existing schemes given in [14] and [29].

2. Reformulation of the image restoration model.

In this section, we derive the surrogates for cost functional (4) with two smoothing functions. The corresponding derivatives for the cost functional and the alternative iteration scheme based on the linearized Euler equation are also proposed.

2.1. Smooth surrogates for cost functional.

For the cost functional (4), two penalty terms have the representation

$$\|\mathbf{vect}[f]\|_{l^1} := \sum_{m,n=1}^{N} |f_{m,n}|, \quad |f|_{TV} := \sum_{m,n=1}^{N} \|(\nabla f)_{m,n}\|_{l^2}$$

with $(\nabla f)_{m,n} := ((\nabla^{x_1} f)_{m,n}, (\nabla^{x_2} f)_{m,n})^T \in \mathbb{R}^{2\times 1}$ of the following components

$$(\nabla^{x_1} f)_{m,n} = \begin{cases} f_{m+1,n} - f_{m,n}, & \text{if } m < N, \\ f_{1,n} - f_{m,n}, & \text{if } m = N, \end{cases} \tag{5}$$

$$(\nabla^{x_2} f)_{m,n} = \begin{cases} f_{m,n+1} - f_{m,n}, & \text{if } \ n < N, \\ f_{m,1} - f_{m,n}, & \text{if } \ n = N \end{cases} \tag{6}$$

for $m, n = 1, \cdots, N$ due to the periodic boundary condition for f. Notice that the TV regularizing term is respect to the gradient vector $(\nabla f)_{m,n} \in \mathbb{R}^{2\times 1}$.

To overcome the non-differentiability of l^1 and TV penalty terms in (4), we use Charbonnier function [1] and Huber function [12] to amend the non-differentiable absolute value function, which are defined as

$$\phi_\beta^C(s) := \sqrt{s^2 + \beta}, \quad \phi_\epsilon^H(s) := \begin{cases} \frac{s^2}{2\epsilon}, & |s| \le \epsilon, \\ |s| - \frac{\epsilon}{2}, & |s| > \epsilon, \end{cases} \tag{7}$$

with small perturbation $\beta, \epsilon > 0$ respectively. Clearly, the Huber function with Lipschitz C^1 continuity can be considered as a better smooth approximation to $|x|$ [15], compared with the Charbonnier approximation with C^∞ continuity.

Although Charbonnier approximation is almost the most general smooth approximation for absolute value function, Huber approximation, which is non-quadratic but convex, has both better theoretical approximation and better performance [13] than Charbonnier approximation. The advantage by using Huber approximation is that it could smooth small scale noise by the quadratic function for argument below a threshold ϵ, i.e., the derivative of Huber function is a quadratic function in $[-\epsilon,\epsilon]$, while preserve discontinuities at edge regions by the linear function part above the threshold $\epsilon>0$.

Consequently, we are led to the following unconstraint cost functional

$$J^Z_{\alpha,\nu}(f):=\frac{1}{2}\left\|\mathcal{P}\mathcal{F}[f]-\mathcal{P}[\hat{g}^\delta]\right\|^2_F+\alpha_1\|\mathbf{vect}[f]\|_{l^1,\phi^Z_\nu}+\alpha_2|f|_{TV}, \tag{8}$$

which approximates the l^1 norm by $\|\mathbf{vect}[f]\|_{l^1,\phi^Z_\nu}=\sum\limits_{m,n=1}^{N}\phi^Z_\nu(f_{m,n})$ for $(Z,\nu)=(C,\beta)$ or $(Z,\nu)=(H,\epsilon)$. Use $\mathbf{f}\equiv\mathbf{vect}[f]$ by the definition and introduce the $N\times N$ matrix

$$D_-:=\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0\\ 0 & -1 & 1 & \cdots & 0 & 0 & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & \cdots & -1 & 1 & 0\\ 0 & 0 & 0 & \cdots & 0 & -1 & 1\\ 1 & 0 & 0 & \cdots & 0 & 0 & -1 \end{pmatrix}, \tag{9}$$

then $\nabla\mathbf{f}:=(\nabla^{x_1}\mathbf{f},\nabla^{x_2}\mathbf{f})^T\in\mathbb{R}^{2\times N^2}$ has the representation in terms of its components by

$$(\nabla^{x_1}\mathbf{f})_j:=((I\otimes D_-)\mathbf{f})_j,\ (\nabla^{x_2}\mathbf{f})_j:=((D_-\otimes I)\mathbf{f})_j,\quad j=1,\cdots,N^2, \tag{10}$$

where $I\otimes D_-, D_-\otimes I\in\mathbb{R}^{N^2\times N^2}$ are block-circulant-circulant-block (BCCB) matrices [24] with tensor product $\otimes$, representing two first-order forward finite difference operators for f with periodic boundary conditions along x_1,x_2 directions [16]. So the penalty term in (8) can be expressed as

$$|\mathbf{f}|_{TV}=\sum_{j=1}^{N^2}\|((\nabla^{x_1}\mathbf{f})_j,(\nabla^{x_2}\mathbf{f})_j)\|_{l^2},\quad \|\mathbf{f}\|_{l^1,\phi^Z_\nu}=\sum_{j=1}^{N^2}\phi^Z_\nu(\mathbf{f}_j), \tag{11}$$

where $j:=j(m,n)=(n-1)\times N+m$ for $m,n=1,\cdots,N$.

Consider the data-matching term in (4). The two-dimensional discrete Fourier transform (DFT) operating on image vector $\mathbf{f}$ has the representation $\hat{\mathbf{f}}=(F\otimes F)\mathbf{f}:=\mathbf{F}\mathbf{f}$, where $\mathbf{F}\in\mathbb{C}^{N^2\times N^2}$ is the two-dimensional DFT matrix. The partial frequency data can be generated from different sampling operators $\mathcal{P}$, such as and random band sampling (RBS) method $\mathcal{P}_*$ [14], radial sampling method [29] and band sampling method [33]. Denoted by $P\in\mathbb{R}^{N\times N}$ the corresponding sampling matrix for $\hat{g}^\delta\in\mathbb{C}^{N\times N}$, which is generated from identity matrix I by setting its $(N-M)$ rows as null vectors with $M\ll N$. The corresponding sampling matrix acting on the vector $\mathbf{g}^\delta\in\mathbb{C}^{N^2\times 1}$ is denoted by $\mathbf{P}\in\mathbb{R}^{N^2\times N^2}$ [14]. More over, we use R_{total}, R_{center} to represent the sampling ratio and the efficient elements among all the sampling elements [14].

By using the argument $\mathbf{f}$ instead of f in both the data-fitting term and penalty terms, we finally establish the following unconstrained optimization models for image restoration

$$\min_{\mathbf{f}}\left\{J^Z_{\alpha,\nu}(\mathbf{f}):=\frac{1}{2}\left\|\mathbf{P}\mathbf{F}\mathbf{f}-\mathbf{P}\hat{\mathbf{g}}^\delta\right\|^2_{l^2}+\alpha_1\|\mathbf{f}\|_{l^1,\phi^Z_\nu}+\alpha_2|\mathbf{f}|_{TV}\right\} \tag{12}$$

with $(Z,\nu)=(C,\beta)$ or $(Z,\nu)=(H,\epsilon)$ using Charbonnier function and Huber function to smoothen the penalty terms, respectively.

2.2. **The alternative iteration scheme.** The TV penalty term $|\mathbf{f}|_{TV}$ in (12) is inconvenient for establishing the gradient type iteration algorithms. To overcome this difficulty, we introduce the new argument $\mathbf{w}$ to represent $\nabla\mathbf{f}$ and consider the corresponding alternative iteration scheme for optimizing (12).

Let $\mathbf{w}=(\mathbf{w}_1,\cdots,\mathbf{w}_{N^2})\in\mathbb{R}^{2\times N^2}$ with $\mathbf{w}_j=(\mathbf{w}_j^1,\mathbf{w}_j^2)^T\in\mathbb{R}^{2\times 1}$. We rewrite (12) as the following constrained one

$$\begin{cases}\min\limits_{\mathbf{w},\mathbf{f}} & \widetilde{J}^Z_{\alpha,\nu}(\mathbf{w},\mathbf{f})\\ \text{s.t.} & \mathbf{w}_j-(\nabla\mathbf{f})_j=(0,0)^T,\quad j=1,\cdots,N^2\end{cases}\tag{13}$$

with the cost functional defined by

$$\widetilde{J}^Z_{\alpha,\nu}(\mathbf{w},\mathbf{f}):=\tfrac{1}{2}\left\|\mathbf{PFf}-\mathbf{P}\hat{\mathbf{g}}^\delta\right\|^2_{l^2}+\alpha_1\|\mathbf{f}\|_{l^1,\phi^Z_\nu}+\alpha_2\sum_{j=1}^{N^2}\|\mathbf{w}_j\|_{l^2}.\tag{14}$$

To minimize (13), we use the alternating iterative scheme by augmented Lagrange method with multiplier method [23] in inner iteration for updating $\mathbf{w}$ and then solve the Euler equation of the cost functional in outer iteration for updating $\mathbf{f}$, instead of using the classic alternating direction method with multiplier [4].

In alternative iteration scheme, we fix one argument either $\mathbf{w}$ or $\mathbf{f}$ and optimize the functional with respect to the other argument at each iteration step. Firstly, for fixed $\mathbf{f}^{(k)}$, which is just an approximation to the minimizer of (13), we minimize (13) with respect to $\mathbf{w}$, which is equivalent to minimize the sub-problem

$$\begin{cases}\min\limits_{\mathbf{w}} & \sum\limits_{j=1}^{N^2}\|\mathbf{w}_j\|_{l^2}\\ \text{s.t.} & \left\|\mathbf{w}_j-(\nabla\mathbf{f}^{(k)})_j\right\|_{l^2}\le\varepsilon_{tol},\quad j=1,\cdots,N^2,\end{cases}\tag{15}$$

where ε_{tol} is some specified error indicating the approximation level $\mathbf{f}^{(k)}$. In the sequel, we use $\nabla\mathbf{f}^{(k)}_j:=(\nabla\mathbf{f}^{(k)})_j$ for the simplicity of notations.

By the K-T Theorem, the sub-problem (15) can be solved by the first-order optimal condition [23] for its Lagrange functional

$$\mathcal{L}^{\boldsymbol{\lambda}}(\mathbf{w})=\sum_{j=1}^{N^2}\|\mathbf{w}_j\|_{l^2}-(\boldsymbol{\lambda}_1^T,\cdots,\boldsymbol{\lambda}_{N^2}^T)\left((\mathbf{w}_1-\nabla\mathbf{f}_1^{(k)})^T,\cdots,(\mathbf{w}_{N^2}-\nabla\mathbf{f}_{N^2}^{(k)})^T\right)^T$$

with the Lagrange multiplier $\boldsymbol{\lambda}:=(\boldsymbol{\lambda}_1,\cdots,\boldsymbol{\lambda}_{N^2})\in\mathbb{R}^{2\times N^2}$. However, to make the iterative process stable, we consider the enhanced Lagrange method, i.e., instead of minimizing $\mathcal{L}^{\boldsymbol{\lambda}}(\mathbf{w})$, we consider the unconstrained optimization problem for

$$\begin{aligned}\mathcal{L}^{\boldsymbol{\lambda},\tau}(\mathbf{w}):&=\mathcal{L}^{\boldsymbol{\lambda}}(\mathbf{w})+\frac{\tau}{2}\left\|\left((\mathbf{w}_1-\nabla\mathbf{f}_1^{(k)}),\cdots,(\mathbf{w}_{N^2}-\nabla\mathbf{f}_{N^2}^{(k)})\right)\right\|^2_{l^2}\\ &\equiv\sum_{j=1}^{N^2}\left(\|\mathbf{w}_j\|_{l^2}+\frac{\tau}{2}\left\|\mathbf{w}_j-\nabla\mathbf{f}_j^{(k)}-\frac{1}{\tau}\boldsymbol{\lambda}_j\right\|^2_{l^2}-\frac{1}{2\tau}\|\boldsymbol{\lambda}_j\|^2_{l^2}\right)\end{aligned}\tag{16}$$

with some weight $\tau>0$. The motivation on considering $\mathcal{L}^{\boldsymbol{\lambda},\tau}(\mathbf{w})$ is to ensure the approximate minimizer of (13) can be reached by minimizing $\mathcal{L}^{\boldsymbol{\lambda},\tau}(\mathbf{w})$ within finite iterative steps, by weakening the morbidity of the Hessian matrix of $\mathcal{L}^{\boldsymbol{\lambda}}(\mathbf{w})$ [3].

Since the optimal Lagrange multiplier $\boldsymbol{\lambda}^*$ satisfying the first-order optimal conditions for $\mathcal{L}^{\lambda}(\mathbf{w})$ is unknown, it is required to update $\boldsymbol{\lambda}$ from some initial multiplier $\boldsymbol{\lambda}^{(k),0}$ such that $\boldsymbol{\lambda}^{(k),l} \to \boldsymbol{\lambda}^*$ as $l \to \infty$ when updating the cost functional $\mathcal{L}^{\lambda,\tau}(\mathbf{w})$ for fixed $\tau > 0$. So the constrained optimization problem (13) is reformulated as minimizing the augmented Lagrangian with multiplier method using the inner iteration process, i.e., the unconstrained problem

$$\min_{\mathbf{w}} \widetilde{J}^{Z,\boldsymbol{\lambda},\tau}_{\alpha,\nu}(\mathbf{w}, \mathbf{f}^{(k)}), \tag{17}$$

where the cost functional is defined as

$$\begin{aligned}\widetilde{J}^{Z,\boldsymbol{\lambda},\tau}_{\alpha,\nu}(\mathbf{w}, \mathbf{f}^{(k)}) \;&=\; \frac{1}{2}\left\|\mathbf{P}\mathbf{F}\mathbf{f}^{(k)} - \mathbf{P}\hat{\mathbf{g}}^{\delta}\right\|^2_{l^2} + \alpha_1\|\mathbf{f}^{(k)}\|_{l^1,\phi^Z_\nu} + \\ &\alpha_2 \sum_{j=1}^{N^2}\left(\|\mathbf{w}_j\|_{l^2} + \frac{\tau}{2}\left\|\mathbf{w}_j - \nabla\mathbf{f}^{(k)}_j - \frac{1}{\tau}\boldsymbol{\lambda}_j\right\|^2_{l^2} - \frac{1}{2\tau}\|\boldsymbol{\lambda}_j\|^2_{l^2}\right)\end{aligned} \tag{18}$$

with the multi-regularizing parameters $\alpha > 0$, smoothing factor $\nu > 0$, multiplier parameter $\boldsymbol{\lambda}$ and the penalty factor $\tau > 0$.

In generating the minimizer $\mathbf{w}^{(k+1)}$ from (17) by inner iteration, the artificially introduced parameters $\boldsymbol{\lambda}$ and τ should be specified. For simplicity, we take a fixed value $\tau > 0$ and only update $\boldsymbol{\lambda}$ in the iteration process. For $\boldsymbol{\lambda}^{(k),l}$ with $l = 0, 1, \cdots$ and $\boldsymbol{\lambda}^{(k),0} := \boldsymbol{\lambda}^{(k)}$ at the $l-$th inner iteration step, the minimization of $\widetilde{J}^{Z,\boldsymbol{\lambda}^{(k),l},\tau}_{\alpha,\nu}(\mathbf{w}, \mathbf{f}^{(k)})$ with respect to $\mathbf{w}$ can be carried out easily, because all $\mathbf{w}_j$ are separated each other. The Euler equation for $\widetilde{J}^{Z,\boldsymbol{\lambda}^{(k),l},\tau}_{\alpha,\nu}(\mathbf{w}, \mathbf{f}^{(k)})$ with respect to $\mathbf{w}$ can be written in terms of each component $\mathbf{w}_j$ as

$$\frac{\mathbf{w}_j}{\|\mathbf{w}_j\|_{l^2}} + \tau(\mathbf{w}_j - \mathbf{t}^{(k),l}_j) = \mathbf{0}, \quad j = 1, \cdots, N^2 \tag{19}$$

with $\mathbf{t}^{(k),l}_j := \nabla\mathbf{f}^{(k)}_j + \boldsymbol{\lambda}^{(k),l}_j/\tau \in \mathbb{R}^{2\times 1}$. The solution to (19) is

$$\mathbf{w}^{(k),l+1}_j = \max\left\{1 - \frac{1}{\tau}\frac{1}{\left\|\mathbf{t}^{(k),l}_j\right\|_{l^2}}, 0\right\}\mathbf{t}^{(k),l}_j. \tag{20}$$

Once we have $\mathbf{w}^{(k),l+1}_j$, we then update $\boldsymbol{\lambda}^{(k),l}$ in the inner iteration by

$$\begin{aligned}\boldsymbol{\lambda}^{(k),l+1}_j : \;&=\; \boldsymbol{\lambda}^{(k),l}_j - \tau(\mathbf{w}^{(k),l+1}_j - \nabla\mathbf{f}^{(k)}_j) \\ &=\; \begin{cases}\dfrac{\mathbf{w}^{(k),l+1}_j}{\left\|\mathbf{w}^{(k),l+1}_j\right\|_{l^2}}, & \mathbf{w}^{(k),l+1}_j \neq \mathbf{0}, \\ \boldsymbol{\lambda}^{(k),l}_j + \tau\nabla\mathbf{f}^{(k)}_j, & \mathbf{w}^{(k),l+1}_j = \mathbf{0}\end{cases}\end{aligned} \tag{21}$$

due to (19). By (20), $\mathbf{w}^{(k),l+1}_j = \mathbf{0}$ means $\left\|\mathbf{t}^{(k),l}_j\right\|_{l^2} \leq \frac{1}{\tau}$, i.e., we always have $\left\|\boldsymbol{\lambda}^{(k),l+1}_j\right\| \leq 1$ for all $l = 1, \cdots$ at any fixed k and $j = 1, \cdots, N^2$.

The advantage of Lagrange multiplier method is the weight τ should not be large enough. On the other words, convergence in Lagrange multiplier method can usually be attained without the need to increase τ to infinity thereby alleviating the ill-posed conditioning problem that plagues the penalty method. The inner iteration will be stopped, if

$$\left\|\mathbf{w}^{(k),l+1}-\nabla \mathbf{f}^{(k)}\right\|_{l^2} \le \varepsilon_{tol} \tag{22}$$

at some step $l = L(k)$ for specified small tolerance $\varepsilon_{tol} > 0$. Then the main loop is going on with

$$\mathbf{w}^{(k+1)} := \mathbf{w}^{(k),L(k)+1}, \quad \boldsymbol{\lambda}^{(k+1)} := \boldsymbol{\lambda}^{(k),L(k)+1}. \tag{23}$$

With the above updated $\mathbf{w}^{(k+1)}$ and $\boldsymbol{\lambda}^{(k+1)}$, the AIS leads to

$$\min_{\mathbf{f}} \widetilde{J}_{\alpha,\nu}^{Z,\boldsymbol{\lambda}^{(k+1)},\tau}(\mathbf{w}^{(k+1)}, \mathbf{f}) \tag{24}$$

for updating $\mathbf{f}^{(k)}$. Define the $N \times N$ matrix $D_+ := -D_-^T$. Using the identity

$$\frac{\partial}{\partial \mathbf{f}_l} \sum_{j=1}^{N^2} (\boldsymbol{\lambda}_j)^T(\mathbf{w}_j - \nabla \mathbf{f}_j) = \frac{\partial}{\partial \mathbf{f}_l} \sum_{j=1}^{N^2} \left[(\lambda_j^1(\mathbf{w}_j^1 - (\nabla^{x_1}\mathbf{f})_j) + \lambda_j^2(\mathbf{w}_j^2 - (\nabla^{x_2}\mathbf{f})_j))\right]$$

for $l = 1, \cdots, N^2$ and defining $\boldsymbol{\lambda}^i = \left(\lambda_1^i, \cdots, \lambda_{N^2}^i\right)$ for $i = 1, 2$, the vector representation of the derivative for the Lagrange multiplier term is

$$\begin{aligned} \nabla_{\mathbf{f}} \left(\sum_{j=1}^{N^2} (\boldsymbol{\lambda}_j)^T(\mathbf{w}_j - \nabla \mathbf{f}_j) \right) &= (I \otimes D_+)(\boldsymbol{\lambda}^1)^T + (D_+ \otimes \ I)(\boldsymbol{\lambda}^2)^T \\ &= (I \otimes D_+, D_+ \otimes I)\vec{\boldsymbol{\lambda}}, \end{aligned} \tag{25}$$

where $\vec{\boldsymbol{\lambda}} := (\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2)^T \equiv \mathbf{vect}[\boldsymbol{\lambda}^T] \in \mathbb{R}^{2N^2\times 1}$.

Analogously, the gradient of the penalty term $\sum_{j=1}^{N^2} \|\mathbf{w}_j - \nabla \mathbf{f}_j\|_{l^2}^2$ with respect to $\mathbf{f}$ has the form

$$\nabla_{\mathbf{f}} \left(\sum_{j=1}^{N^2} \|\mathbf{w}_j - \nabla \mathbf{f}_j\|_{l^2}^2 \right) = 2\left(I \otimes D_+, D_+ \otimes I\right) \left[\vec{\mathbf{w}} - \begin{pmatrix} I \otimes D_- \\ D_- \otimes I \end{pmatrix} \mathbf{f}\right], \tag{26}$$

where $\vec{\mathbf{w}} := \left(\mathbf{w}_1^1, \cdots, \mathbf{w}_{N^2}^1, \mathbf{w}_1^2, \cdots, \mathbf{w}_{N^2}^2\right)^T \equiv \mathbf{vect}[\mathbf{w}^T] \in \mathbb{R}^{2N^2\times 1}$.

For smoothing function $\|\mathbf{f}\|_{l^1,\phi_\nu^Z}$, it follows that

$$\frac{\partial}{\partial \mathbf{f}_{j'(m',n')}} \|\mathbf{f}\|_{l^1,\phi_\beta^C} = \frac{\mathbf{f}_{j'(m',n')}}{\sqrt{|\mathbf{f}_{j'(m',n')}|^2 + \beta}}, \tag{27}$$

$$\frac{\partial}{\partial \mathbf{f}_{j'(m',n')}} \|\mathbf{f}\|_{l^1,\phi_\epsilon^H} = \begin{cases} \mathbf{f}_{j'(m',n')}/\epsilon, & |\mathbf{f}_{j'(m',n')}| \le \epsilon \\ \operatorname{sgn}(\mathbf{f}_{j'(m',n')}), & |\mathbf{f}_{j'(m',n')}| > \epsilon \end{cases} \tag{28}$$

for $j' = 1, \cdots, N^2$ which are defined as $j' := j'(m', n') = (n' - 1)N + m'$ for $m', n' = 1, \cdots, N$. Therefore, for $l = 1, \cdots, N^2$, we have

$$\frac{\partial}{\partial \mathbf{f}_l} \|\mathbf{f}\|_{l^1,\phi_\nu^Z} = a_l^Z[\mathbf{f}]\mathbf{f}_l, \tag{29}$$

where

$$a_l^C[\mathbf{f}] := \frac{1}{\sqrt{|\mathbf{f}_l|^2 + \beta}}, \qquad a_l^H[\mathbf{f}] := \begin{cases} 1/\epsilon, & |\mathbf{f}_l| \le \epsilon, \\ \operatorname{sgn}(\mathbf{f}_l)/\mathbf{f}_l, & |\mathbf{f}_l| > \epsilon. \end{cases}$$

So the gradient form under each smoothing function is

$$\nabla_{\mathbf{f}} \|\mathbf{f}\|_{l^1,\phi_\nu^Z} = \Lambda^Z[\mathbf{f}]\mathbf{f}, \tag{30}$$

where $\Lambda^Z[\mathbf{f}] := \operatorname{diag}(a_1^Z[\mathbf{f}], a_2^Z[\mathbf{f}], \cdots, a_{N^2}^Z[\mathbf{f}])$ for $Z = C, H$.

Combining all the above analysis, the Euler equation for the cost functional $\widetilde{J}^{Z,\boldsymbol{\lambda}^{(k+1)},\tau}_{\alpha,\nu}(\mathbf{w}^{(k+1)},\mathbf{f})$ with respect to $\mathbf{f}$ in vector form is

$$\begin{aligned}&\overline{\mathbf{F}}^T\mathbf{P}^T(\mathbf{P}\mathbf{F}\mathbf{f}-\mathbf{P}\hat{\mathbf{g}}^\delta)-\alpha_2(I\otimes D_+,D_+\otimes I)\vec{\boldsymbol{\lambda}}^{(k+1)}+\\ &\alpha_2\tau\,(I\otimes D_+,D_+\otimes I)\left[\vec{\mathbf{w}}^{(k+1)}-\begin{pmatrix} I\otimes D_-\\ D_-\otimes I\end{pmatrix}\mathbf{f}\right]+\alpha_1\Lambda^Z[\mathbf{f}]\mathbf{f}=\mathbf{0}\end{aligned}\tag{31}$$

for $Z=C,H$ respectively. Introduce $A^Z_{(k)}:=\max\{a^Z_j[\mathbf{f}^{(k)}]:\ j=1,\cdots,N^2\}$. We propose to linearize the nonlinear term $\Lambda^Z[\mathbf{f}]\mathbf{f}$ of (31) near $\mathbf{f}^{(k)}$ by

$$\Lambda^Z[\mathbf{f}]\mathbf{f}\approx A^Z_{(k)}\mathbf{f}+\Lambda^Z[\mathbf{f}^{(k)}]\mathbf{f}^{(k)}-A^Z_{(k)}\mathbf{f}^{(k)},$$

which leads to the linear equation

$$\mathbf{L}^{(k)}\mathbf{f}^{(k+1)}=\mathbf{b}^{(k)}\tag{32}$$

for solving $\mathbf{f}^{(k+1)}$ from $\mathbf{f}^{(k)}$, where

$$\begin{cases}\mathbf{L}^{(k)}=-\alpha_2\tau\,(I\otimes D_+,D_+\otimes I)\begin{pmatrix} I\otimes D_-\\ D_-\otimes I\end{pmatrix}+\alpha_1A^Z_{(k)}\mathbf{I}+\overline{\mathbf{F}}^T\mathbf{P}^T\mathbf{P}\mathbf{F},\\ \mathbf{b}^{(k)}=\alpha_2\,(I\otimes D_+,D_+\otimes I)\left(-\tau\vec{\mathbf{w}}^{(k+1)}+\vec{\boldsymbol{\lambda}}^{(k+1)}\right)+\overline{\mathbf{F}}^T\mathbf{P}^T\mathbf{P}\hat{\mathbf{g}}^\delta+\\ \qquad\alpha_1(A^Z_{(k)}\mathbf{I}-\Lambda^Z[\mathbf{f}^{(k)}])\mathbf{f}^{(k)}.\end{cases}$$

The motivation to replace the nonlinear Euler equation (31) by (32) is that we try to find the minimizer in an efficient way, which can be realized by solving (32). In fact, applying two-dimensional DFT operator $\mathcal{F}$ on (32), we obtain

$$\widetilde{\mathbf{L}}^{(k)}\hat{\mathbf{f}}^{(k+1)}=\hat{\mathbf{b}}^{(k)},\tag{33}$$

where $\hat{\mathbf{f}}^{(k+1)}=\mathcal{F}[\mathbf{f}^{(k+1)}]=\mathbf{F}\mathbf{f}^{(k+1)}$ and

$$\begin{aligned}\hat{\mathbf{b}}^{(k)}\ &=\ \mathbf{F}\left[\alpha_2(I\otimes D_+,D_+\otimes I)\left(-\tau\vec{\mathbf{w}}^{(k+1)}+\vec{\boldsymbol{\lambda}}^{(k+1)}\right)\right]+\mathbf{P}^T\mathbf{P}\hat{\mathbf{g}}^\delta+\\ &\qquad \alpha_1\mathbf{F}(A^Z_{(k)}\mathbf{I}-\Lambda^Z[\mathbf{f}^{(k)}])\mathbf{f}^{(k)},\end{aligned}\tag{34}$$

$$\begin{aligned}\widetilde{\mathbf{L}}^{(k)}\ &=\ \mathbf{F}\mathbf{L}^{(k)}\mathbf{F}^{-1}\\ &=\ -\alpha_2\tau\mathbf{F}\left[(I\otimes D_+)(I\otimes D_-)+(D_+\otimes I)(D_-\otimes I)\right]\mathbf{F}^{-1}+\\ &\qquad\alpha_1A^Z_{(k)}\mathbf{I}+\mathbf{P}^T\mathbf{P}\\ &:=\ -\alpha_2\tau\mathbf{F}\left(\mathbb{D}_1+\mathbb{D}_2\right)\mathbf{F}^{-1}+\alpha_1A^Z_{(k)}\mathbf{I}+\mathbf{P}^T\mathbf{P},\end{aligned}\tag{35}$$

where $\mathbb{D}_i(i=1,2)$ are the $N^2\times N^2$ symmetric block-circulate-circulate-block (BCCB) matrices generated by $\mathbb{D}_1:=\mathbf{bccb}\circ D_*,\mathbb{D}_2:=\mathbf{bccb}\circ D_*^T$ for $N\times N$ matrix $D_*=(\mathbf{d}_*,\mathbf{0},\cdots,\mathbf{0})$, where

$$\mathbf{d}_*=(-2,2)^T\text{ for }N=2,\ \mathbf{d}_*=(-2,1,\underbrace{0,\cdots,0}_{N-3},1)^T\text{ for }N=3,4,\cdots,\tag{36}$$

where **bccb** is the standard recursion operator, see [24]. With the commutativity of Fourier transform on BCCB matrix $\mathbb{D}_i$, it follows that $\mathbf{F}\left(\mathbb{D}_1+\mathbb{D}_2\right)=-\left(\mathbb{L}_1+\mathbb{L}_2\right)\mathbf{F}$, where $\mathbb{L}_i(i=1,2)$ are diagonal matrices, with the elements being the negative eigenvalues of $\mathbb{D}_i(i=1,2)$ (Prop. 5.31 in [24]). More precisely, by straightforward computations on $\mathbb{L}_1+\mathbb{L}_2=-\mathbf{F}\left(\mathbb{D}_1+\mathbb{D}_2\right)\overline{\mathbf{F}}^T$, it follows

$\mathbb{L}_1 + \mathbb{L}_2 = \text{diag}(\mathbf{vec}[\mathbb{L}])$ with the elements of $\mathbb{L}$ being

$$\mathrm{l}_{m,n} = 4 - 2\left(\cos\frac{2\pi}{N}(m-1) + \cos\frac{2\pi}{N}(n-1)\right), \quad m,n = 1,\cdots,N. \tag{37}$$

Therefore $\widetilde{\mathbf{L}}^{(k)}$ has the expression

$$\widetilde{\mathbf{L}}^{(k)} = \alpha_2\tau\left(\mathbb{L}_1 + \mathbb{L}_2\right) + \alpha_1 A^Z_{(k)}\mathbf{I} + \mathbf{P}^T\mathbf{P}. \tag{38}$$

The last term $\mathbf{P}^T\mathbf{P}$ is also a diagonal matrix for both random band sampling and radiative sampling [14]. So for (33) with diagonal coefficient matrix, it can be solved simply to yield $\hat{\mathbf{f}}^{(k+1)}$ in frequency domain, and then generate $\mathbf{f}^{(k+1)} = \overline{\mathbf{F}}^T\hat{\mathbf{f}}^{(k+1)}$ by the inverse Fourier transform.

Since $\mathbf{f}$ represents the grey values of an image, we always embed the *a-priori* information $\mathbf{f} \in [0,1]$ into our iterative algorithm, i.e., we further set the cut-off process

$$\mathbf{f}_j^{(k+1)} = \begin{cases} \mathbf{f}_j^{(k+1)}, & \text{if } 0 \le \mathbf{f}_j^{(k+1)} \le 1, \\ 0, & \text{if } \mathbf{f}_j^{(k+1)} < 0, \\ 1, & \text{if } \mathbf{f}_j^{(k+1)} > 1. \end{cases} \tag{39}$$

Then the alternating iteration process is stopped under the condition

$$\text{either } \left\|\mathbf{P}\hat{\mathbf{f}}^{(k+1)} - \mathbf{P}\hat{\mathbf{g}}^\delta\right\|_{l^2} \le \delta \text{ or } \|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_{l^2} \le \varepsilon \text{ or } k \le K_0, \tag{40}$$

where δ is the noise level, and K_0 is the maximum iteration number. Under the reasonable restriction (39), we find that (40) can be realized efficiently, although the cut-off process (39) is rarely excited.

Algorithm Alternative iteration scheme

Input: noisy frequency data $\{\hat{g}^\delta_{m',n'} : \quad m',n' = 1,\cdots,N\}$, sampling matrix $\mathbf{P} \in \mathbb{R}^{N^2\times N^2}$, parameters $\alpha_1, \alpha_2, \beta, \epsilon, \tau, K_0$, tolerance ε_{tol}.
Set initial value $\mathbf{f}^{(0)} = \mathbf{0} \in \mathbb{R}^{N^2\times 1}, \boldsymbol{\lambda}^{(0)} = \mathbf{0} \in \mathbb{R}^{2\times N^2}$.
Do exterior loop from $k = 1,2,\cdots$
While $\left\|\mathbf{P}\mathbf{F}\mathbf{f}^{(k)} - \mathbf{P}\hat{\mathbf{g}}^\delta\right\|_{l^2} > \delta$ or $k < K_0$
  Do inner loop from $l = 0,1,\cdots$ with $\boldsymbol{\lambda}^{(k),0} = \boldsymbol{\lambda}^{(k-1)} \in \mathbb{R}^{2\times N^2}$.
  { Compute:
    Determine $\mathbf{w}_j^{(k),l+1}$ by (20) for all j;
    Update: $\boldsymbol{\lambda}_j^{(k),l+1} \leftarrow \boldsymbol{\lambda}_j^{(k),l} - \tau(\mathbf{w}_j^{(k),l+1} - \nabla\mathbf{f}_j^{(k)})$ by (21) for all j;
    If $\left\|\mathbf{w}^{(k),l+1} - \nabla\mathbf{f}^{(k)}\right\|_{l^2} \le \varepsilon_{tol}$ Break. }
  $\mathbf{w}^{(k+1)} := \mathbf{w}^{(k),l+1}$, $\boldsymbol{\lambda}^{(k+1)} := \boldsymbol{\lambda}^{(k),l+1}$.
  Determine $\mathbf{f}^{(k+1)}$ by solving (33) and then taking IFFT.
Modify $\mathbf{f}^{(k+1)}$ by (39).
End while
$f^{(k+1)} \in \mathbb{R}^{N\times N} \leftarrow \mathbf{f}^{(k+1)} \in \mathbb{R}^{N^2\times 1}$ and output.
End

To sum up, we propose to solve the unconstrained optimization problem $\min\limits_{\mathbf{f}} J^Z_{\alpha,\nu}(\mathbf{f})$ by solving the constrained optimization problem $\min\limits_{\mathbf{w},\mathbf{f}} \widetilde{J}^Z_{\alpha,\nu}(\mathbf{w},\mathbf{f})$ under the constraint $\mathbf{w} = \nabla\mathbf{f}$. By applying the augmented Lagrange multiplier method, the AIS is applied to solve the unconstrained problem $\min\limits_{\mathbf{w},\mathbf{f}} \widetilde{J}^{Z,\boldsymbol{\lambda},\tau}_{\alpha,\nu}(\mathbf{w},\mathbf{f})$ with artificially introduced weights $\boldsymbol{\lambda},\tau$ by two iterations. The first iteration is to update $(\boldsymbol{\lambda},\mathbf{w})$ for which the solution can be expressed explicitly, and the second iteration is to update $\mathbf{f}$ by solving a linear equation in frequency domain, which has a diagonal coefficient matrix and consequently can be solved efficiently. Such a nice iterative equation is generated by further modifying the linearized Euler equation for the cost functional.

The essence of the proposed simplified scheme for yielding $\mathbf{f}$ is to solve the nonlinear Euler equation approximately from its linearizing version. Consequently, we need to deal with the convergence property of this iterative sequence $\{(\mathbf{w}^{(k)},\mathbf{f}^{(k)}) : k\in\mathbb{N}\}$ as well as its error with the minimizer by solving the nonlinear Euler equation, which will be the topic in the next section.

3. **Convergence property of iteration process.** In this section, we will establish the convergence property of the iterative process for the model with Charbonnier approximation, but the analysis is also applicable to Charbonnier approximation. However, the numerical tests in section 4 will be implemented for these two approximations. To simplify the notations, we use $\|\cdot\| := \|\cdot\|_{l^2}$ to represent the norm of a vector, while $\|\cdot\|_2$ represents the norm of an matrix.

By representing the grey function of an image in vector form, we take the minimizer $\mathbf{f}^{\alpha,\beta}_*$ of $J^C_{\alpha,\beta}(\mathbf{f})$ as our approximation to the image. Then this unconstrained optimization problem is rewritten as (13), the constrained one, which is solved by AIS for $\mathbf{w}$ and $\mathbf{f}$.

The main difficulties for the convergence come from two points. Firstly, the constrained problem (13) is solved at each fixed outer iteration step by the augmented Lagrange multiplier scheme for $\mathbf{w}$ through inner iteration, where both $\mathbf{w}$ and the artificially introduced weight $\boldsymbol{\lambda}$ are updated by (20)-(21). Secondly, when we update $\mathbf{f}^{(k)}$ in outer iteration, the exact nonlinear Euler function (31) for the cost functional $\widetilde{J}^{C,\boldsymbol{\lambda}^{(k+1)},\tau}_{\alpha,\beta}(\mathbf{w}^{(k+1)},\mathbf{f})$ is linearized and decomposed as (33) for simplifying our computations in frequency domain.

We will firstly show the convergence property of $\{(\mathbf{w}^{(k)},\mathbf{f}^{(k)}) : k\in\mathbb{N}\}$ from our linearized iterative scheme. In fact we can only establish some "almost convergent" property for the iterative sequence.

Definition 3.1. We call the sequence $\{a^{(k)} :\ k\in\mathbb{N}\}\subset X$, where X is a complete normed space, is almost convergent, if for $\forall\varepsilon>0$, there exists a positive integer N such that for all $m,n>N$, it holds

$$\left\|a^{(m)}-a^{(n)}\right\| < \varepsilon+\varepsilon_{tol},$$

where $\varepsilon_{tol}>0$ is some known small constant.

In case of $\varepsilon_{tol}=0$, this definition ensures that $\{a^{(k)} :\ k\in\mathbb{N}\}$ is a Cauchy sequence in X, and therefore is convergent in the classical sense. So the difference between the classical convergence and the above proposed "almost convergent" is that here $\{a^{(k)} :\ k\in\mathbb{N}\}$ is just a Cauchy sequence approximately due to $\varepsilon_{tol}>0$.

Theorem 3.2. *For any fixed $\alpha_1,\alpha_2>0$, if we take $\tau>0$ small and $\beta>0$ large appropriately, the iterative sequences $\{\mathbf{f}^{(k)} : k\in\mathbb{N}\}$ from the proposed AIS almost converges for small tolerance $\varepsilon_{tol}>0$.*

Proof. For any fixed k and $j = 1, \cdots, N^2$, it follows $\left\|\boldsymbol{\lambda}_j^{(k),l+1}\right\|_{l^2} \leq 1$ from (21) for all $l = 0, 1, \cdots$. So there exists a subsequence of $\{\boldsymbol{\lambda}_j^{(k),l} : l \in \mathbb{N}\}$, still denoted by $\{\boldsymbol{\lambda}_j^{(k),l} : l \in \mathbb{N}\}$ such that $\boldsymbol{\lambda}_j^{(k),l} \to \boldsymbol{\lambda}_j^{(k),*}$ as $l \to \infty$, which leads to $\mathbf{w}_j^{(k),l+1} - \nabla \mathbf{f}_j^{(k)} \to 0$ from (21) as $l \to \infty$. This convergence means (22) can be satisfied for given small $\varepsilon_{tol} > 0$.

Now for $\mathbf{w}^{(k+1)}, \boldsymbol{\lambda}^{(k+1)}$ given by (23) from inner iteration, we update $\mathbf{f}^{(k)}$ by solving (24) approximately using the linearized scheme with the restriction (39). In terms of the expressions of $\tilde{\mathbf{L}}^{(k)}$ and $\hat{\mathbf{b}}^{(k)}$, (33) can be written as

$$
\begin{aligned}
& \left(\alpha_2 \tau \left(\mathbb{L}_1 + \mathbb{L}_2\right) + \alpha_1 A_{(k)}^C \mathbf{I} + \mathbf{P}^T \mathbf{P}\right) \hat{\mathbf{f}}^{(k+1)} \\
= \; & \alpha_2 \mathbf{F}(I \otimes D_+, D_+ \otimes I) \left(-\tau \vec{\mathbf{w}}^{(k+1)} + \vec{\boldsymbol{\lambda}}^{(k+1)}\right) + \mathbf{P}^T \mathbf{P} \hat{\mathbf{g}}^\delta + \\
& \alpha_1 \mathbf{F}(A_{(k)}^C \mathbf{I} - \Lambda^C[\mathbf{f}^{(k)}]) \mathbf{f}^{(k)},
\end{aligned}
\tag{41}
$$

which generates

$$
\begin{aligned}
& \left(\alpha_2 \tau \left(\mathbb{L}_1 + \mathbb{L}_2\right) + \alpha_1 A_{(k)}^C \mathbf{I} + \mathbf{P}^T \mathbf{P}\right) (\hat{\mathbf{f}}^{(k+1)} - \hat{\mathbf{f}}^{(k)}) \\
= \; & \alpha_2 \mathbf{F}(I \otimes D_+, D_+ \otimes I) \left(-\tau (\vec{\mathbf{w}}^{(k+1)} - \vec{\mathbf{w}}^{(k)}) + (\vec{\boldsymbol{\lambda}}^{(k+1)} - \vec{\boldsymbol{\lambda}}^{(k)})\right) + \\
& \alpha_1 \mathbf{F}(A_{(k-1)}^C \mathbf{I} - \Lambda^C[\mathbf{f}^{(k)}])(\mathbf{f}^{(k)} - \mathbf{f}^{(k-1)}) + \\
& \alpha_1 \mathbf{F}(\Lambda^C[\mathbf{f}^{(k-1)}] - \Lambda^C[\mathbf{f}^{(k)}]) \mathbf{f}^{(k-1)}.
\end{aligned}
\tag{42}
$$

For $\vec{\mathbf{w}}^{(k+1)} = \mathbf{vect}[(\mathbf{w}^{(k+1)})^T] \in \mathbb{R}^{2N^2 \times 1}$ with $\mathbf{w}^{(k+1)} \in \mathbb{R}^{2 \times N^2}$, using

$$
\mathbf{w}^{(k+1)} = \nabla \mathbf{f}^{(k)} + q_k \varepsilon_{tol}
\tag{43}
$$

with $\|q_k\| \leq 1$ and the representation of $\nabla \mathbf{f}^{(k)}$, we have

$$
\vec{\mathbf{w}}^{(k+1)} = \begin{pmatrix} I \otimes D_- \\ D_- \otimes I \end{pmatrix} \mathbf{f}^{(k)} + \tilde{q}_k \varepsilon_{tol},
$$

which says

$$
\vec{\mathbf{w}}^{(k+1)} - \vec{\mathbf{w}}^{(k)} = \begin{pmatrix} I \otimes D_- \\ D_- \otimes I \end{pmatrix} (\mathbf{f}^{(k)} - \mathbf{f}^{(k-1)}) + (\tilde{q}_k - \tilde{q}_{k-1}) \varepsilon_{tol}
\tag{44}
$$

with $\|\tilde{q}_k\| = \|q_k\| \leq 1$. Therefore (42) becomes

$$
\begin{aligned}
& \left(\alpha_2 \tau \left(\mathbb{L}_1 + \mathbb{L}_2\right) + \alpha_1 A_{(k)}^C \mathbf{I} + \mathbf{P}^T \mathbf{P}\right) (\hat{\mathbf{f}}^{(k+1)} - \hat{\mathbf{f}}^{(k)}) \\
= \; & \alpha_2 \tau (\mathbb{L}_1 + \mathbb{L}_2)(\hat{\mathbf{f}}^{(k)} - \hat{\mathbf{f}}^{(k-1)}) - \alpha_2 \tau \varepsilon_{tol} \mathbf{F}(\tilde{q}_k - \tilde{q}_{k-1}) + \\
& \alpha_2 \mathbf{F}(I \otimes D_+, D_+ \otimes I)(\vec{\boldsymbol{\lambda}}^{(k+1)} - \vec{\boldsymbol{\lambda}}^{(k)}) + \\
& \alpha_1 \mathbf{F}(A_{(k-1)}^C \mathbf{I} - \Lambda^C[\mathbf{f}^{(k)}])(\mathbf{f}^{(k)} - \mathbf{f}^{(k-1)}) + \\
& \alpha_1 \mathbf{F}(\Lambda^C[\mathbf{f}^{(k-1)}] - \Lambda^C[\mathbf{f}^{(k)}]) \mathbf{f}^{(k-1)}.
\end{aligned}
\tag{45}
$$

On the other hand, the updating process $\boldsymbol{\lambda}_j^{(k+1)} := \boldsymbol{\lambda}_j^{(k)} - \tau(\mathbf{w}_j^{(k+1)} - \nabla \mathbf{f}_j^{(k)})$ means $\boldsymbol{\lambda}^{(k+1)} - \boldsymbol{\lambda}^{(k)} = -\tau(\mathbf{w}^{(k+1)} - \nabla \mathbf{f}^{(k)})$. So we have

$$
\vec{\boldsymbol{\lambda}}^{(k+1)} - \vec{\boldsymbol{\lambda}}^{(k)} = \mathbf{vect}[(\boldsymbol{\lambda}^{(k+1)} - \boldsymbol{\lambda}^{(k)})^T] = -\tau \mathbf{vect}[(\mathbf{w}^{(k+1)} - \nabla \mathbf{f}^{(k)})^T],
$$

which generates

$$
\left\|\vec{\boldsymbol{\lambda}}^{(k+1)} - \vec{\boldsymbol{\lambda}}^{(k)}\right\| \leq \tau \varepsilon_{tol}
\tag{46}
$$

from (43). Moreover, we have

$$\left\|\Lambda^C[\mathbf{f}^{(k-1)}]-\Lambda^C[\mathbf{f}^{(k)}]\right\|,\ \left\|\Lambda^C[\mathbf{f}^{(k-1)}]-\Lambda^C[\mathbf{f}^{(k)}]\right\|\le\frac{1}{\sqrt{\beta^3}}\left\|\mathbf{f}^k-\mathbf{f}^{k-1}\right\|$$

from the expression of $\Lambda^C[\mathbf{f}]$ and $\left\|\mathbf{f}^{(k)}\right\|,\left\|\mathbf{f}^{(k-1)}\right\|\le 1$. Consequently, using $\|\mathbf{f}\|=\left\|\hat{\mathbf{f}}\right\|$, it follows from (45) and (46) that

$$\begin{aligned}\left\|\mathbf{f}^{(k+1)}-\mathbf{f}^{(k)}\right\| \quad\le\quad & \frac{8\alpha_2\tau}{\alpha_1 A^C_{(k)}}\left\|\mathbf{f}^{(k)}-\mathbf{f}^{(k-1)}\right\|+\\ & \frac{1}{\alpha_1 A^C_{(k)}}\left[C\alpha_2\tau\varepsilon_{tol}+\alpha_1\frac{C}{\sqrt{\beta^3}}\left\|\mathbf{f}^{(k)}-\mathbf{f}^{(k-1)}\right\|\right]\end{aligned}\tag{47}$$

using $\left\|\left(\alpha_2\tau\left(\mathbb{L}_1+\mathbb{L}_2\right)+\alpha_1 A^C_{(k)}\mathbf{I}+\mathbf{P}^T\mathbf{P}\right)^{-1}\right\|_\infty\le\frac{1}{\alpha_1 A^C_{(k)}}$ from $0\le l_j\le 8$ due to (37) and $p_j=0,1$. So we have

$$\left\|\mathbf{f}^{(k+1)}-\mathbf{f}^{(k)}\right\|\le\frac{\sqrt{1+\beta}}{\alpha_1}\left[\left(8\alpha_2\tau+\frac{C\alpha_1}{\sqrt{\beta^3}}\right)\left\|\mathbf{f}^{(k)}-\mathbf{f}^{(k-1)}\right\|+C\alpha_2\tau\varepsilon_{tol}\right]$$

due to $\frac{1}{\sqrt{1+\beta}}\le A^C_{(k)}\le\frac{1}{\sqrt{\beta}}$. If we take the parameters $\alpha_1,\alpha_2,\beta,\tau>0$ such that

$$q_1:=\frac{\sqrt{1+\beta}}{\alpha_1}\left(8\alpha_2\tau+\frac{C\alpha_1}{\sqrt{\beta^3}}\right)\in(0,1),\qquad q_2:=\frac{C\sqrt{1+\beta}}{\alpha_1}\alpha_2\tau\in(0,1),\tag{48}$$

then we finally have

$$\left\|\mathbf{f}^{(k+1)}-\mathbf{f}^{(k)}\right\|\le q_1^k\left\|\mathbf{f}^{(1)}-\mathbf{f}^{(0)}\right\|+\frac{1}{1-q_1}\varepsilon_{tol}.\tag{49}$$

This estimate means, if $\varepsilon_{tol}=0$, then $\{\mathbf{f}^{(k)}:k\in\mathbb{N}\}$ is a Cauchy sequence. However, in our inner iteration, it is very hard to make $\varepsilon_{tol}=0$ which means $\mathbf{w}^{(k+1)}=\nabla\mathbf{f}^k$. But when the tolerance $\varepsilon_{tol}>0$ is small enough, $\{\mathbf{f}^{(k)}:k\in\mathbb{N}\}$ is almost a Cauchy series. On the other hand, since the convergence of $\{\mathbf{f}^{(k)}:k\in\mathbb{N}\}$ is based on the convergence of subsequence of $\{\boldsymbol{\lambda}^{(k),l}:l\in\mathbb{N}\}$ in inner iteration, the convergence of $\{\mathbf{f}^{(k)}:k\in\mathbb{N}\}$ is also in the sense of convergence for some subsequence. We call this phenomena as *almost convergence*. The proof is complete. □

In Theorem 3.1, we only prove the convergence of our linearized AIS process, where we can solve $\hat{\mathbf{f}}^{(k+1)}$ efficiently from an linear equation (33) with diagonal matrix. Obviously, the limitation of $\{\mathbf{f}^{(k)}:k\in\mathbb{N}\}$ by this iteration process is only the approximation of $\mathbf{f}_*$, the minimizer of $\widetilde{J}^{C,\boldsymbol{\lambda},\tau}_{\alpha,\beta}(\mathbf{w},\mathbf{f})$. The iterative sequence $\{\mathbf{f}^{(k)}:k\in\mathbb{N}\}$ can not approach to $\mathbf{f}_*$ up to arbitrary accuracy by taking large k, the error always has the lower bound $O(\varepsilon_{tol})$, which reveals the characteristics of our alternative iteration scheme for $(\mathbf{w}^{(k+1)},\mathbf{f}^{(k+1)})$, since the inner iteration for yielding $\mathbf{w}^{(k+1)}$ which approaches to $\nabla\mathbf{f}^{(k)}$ always stops with the error ε_{tol}. Because the AIS algorithm finds the minimizer of (18) iteratively, denote by $\{\mathbf{f}_E^{(k+1)}:k\in\mathbb{N}\}$

the sequence generated by solving the nonlinear Euler equation

$$\begin{aligned}&\overline{\mathbf{F}}^T(\mathbf{P}^T\mathbf{P}\mathbf{F}\mathbf{f}_E^{(k+1)}-\mathbf{P}\hat{\mathbf{g}}^\delta)-\alpha_2(I\otimes D_+,D_+\otimes I)\vec{\boldsymbol{\lambda}}_E^{(k+1)}+\\&\alpha_2\tau\,(I\otimes D_+,D_+\otimes I)\left[\vec{\mathbf{w}}_E^{(k+1)}-\begin{pmatrix}I\otimes D_-\\D_-\otimes I\end{pmatrix}\mathbf{f}_E^{(k+1)}\right]+\\&\alpha_1\Lambda^C[\mathbf{f}_E^{(k+1)}]\mathbf{f}_E^{(k+1)}=\mathbf{0}\end{aligned}\tag{50}$$

from known $\vec{\mathbf{w}}_E^{(k+1)},\vec{\boldsymbol{\lambda}}_E^{(k+1)}$ determined from outer iteration (23). Since this equation is nonlinear, we need to clarify the relation between $\{\mathbf{f}^{(k)}:k\in\mathbb{N}\}$ solving the linear system and $\{\mathbf{f}_E^{(k)}:k\in\mathbb{N}\}$ solving (50).

Theorem 3.3. *If $\{\mathbf{f}^{(k)}:k\in\mathbb{N}\}$ and $\{\mathbf{f}_E^{(k)}:k\in\mathbb{N}\}$ are generated from the same initial guess $\mathbf{f}^{(0)}$, then for small $\alpha_2,\tau,\varepsilon_{tol}>0$ and large $\alpha_1,\beta>0$, it follows*

$$\lim_{k\to\infty}\left\|\mathbf{f}^{(k)}-\mathbf{f}_E^{(k)}\right\|\approx 0\tag{51}$$

up to the accuracy $O(\alpha_2+\tau\varepsilon_{tol})$, where $\lim_{k\to\infty}\mathbf{f}^{(k)}$ is the minimizer of the cost functional $\lim_{k\to\infty}\widetilde{J}^{C,\boldsymbol{\lambda}^{(k)},\tau}_{\alpha,\beta}(\nabla\mathbf{f},\mathbf{f})$ related to $\mathbf{f}^{(0)},\boldsymbol{\lambda}^{(0)}$.

Proof. It follows from direct computations that $\mathbf{f}^{(k)}-\mathbf{f}_E^{(k)}$ meets

$$\begin{aligned}&\overline{\mathbf{F}}^T\mathbf{P}^T\mathbf{P}\mathbf{F}(\mathbf{f}^{(k+1)}-\mathbf{f}_E^{(k+1)})-\alpha_2(I\otimes D_+,D_+\otimes I)(\vec{\boldsymbol{\lambda}}^{(k+1)}-\vec{\boldsymbol{\lambda}}_E^{(k+1)})+\\&\alpha_2\tau(\mathbb{D}_1+\mathbb{D}_2)\left((\mathbf{f}^{(k)}-\mathbf{f}_E^{(k)})+(\tilde{q}_k-\tilde{q}_{k,E})\varepsilon_{tol}-(\mathbf{f}^{(k+1)}-\mathbf{f}_E^{(k+1)})\right)+\\&+\alpha_1A^C_{(k)}\mathbf{f}^{(k+1)}+\alpha_1(\Lambda^C[\mathbf{f}^{(k)}]-A^C_{(k)})\mathbf{f}^{(k)}-\alpha_1\Lambda^C[\mathbf{f}_E^{(k+1)}]\mathbf{f}_E^{(k+1)}=\mathbf{0}.\end{aligned}\tag{52}$$

Define $\mathbf{z}^{(k+1)}:=\mathbf{f}^{(k+1)}-\mathbf{f}_E^{(k+1)}$. Replacing k as $k-1$ in this relation and doing substraction, we have

$$\begin{aligned}&\overline{\mathbf{F}}^T\mathbf{P}^T\mathbf{P}\mathbf{F}(\mathbf{z}^{(k+1)}-\mathbf{z}^{(k)})-\\&\alpha_2(I\otimes D_+,D_+\otimes I)\left((\vec{\boldsymbol{\lambda}}^{(k+1)}-\vec{\boldsymbol{\lambda}}^{(k)})-(\vec{\boldsymbol{\lambda}}_E^{(k+1)}-\vec{\boldsymbol{\lambda}}_E^{(k)})\right)+\\&\alpha_2\tau(\mathbb{D}_1+\mathbb{D}_2)\left((\mathbf{z}^{(k)}-\mathbf{z}^{(k-1)})+(\tilde{Q}_k-\tilde{Q}_{k,E})\varepsilon_{tol}-(\mathbf{z}^{(k+1)}-\mathbf{z}^{(k)})\right)+\\&\alpha_1\left(A^C_{(k)}\mathbf{f}^{(k+1)}+(\Lambda^C[\mathbf{f}^{(k)}]-A^C_{(k)}\mathbf{I})\mathbf{f}^{(k)}\right)-\\&\alpha_1\left(A^C_{(k-1)}\mathbf{f}^{(k)}+(\Lambda^C[\mathbf{f}^{(k-1)}]-A^C_{(k-1)}\mathbf{I})\mathbf{f}^{(k-1)}\right)+\\&\alpha_1\left(\Lambda^C[\mathbf{f}_E^{(k+1)}]\mathbf{z}^{(k+1)}-\Lambda^C[\mathbf{f}_E^{(k)}]\mathbf{z}^{(k)}+\Lambda^C[\mathbf{f}_E^{(k)}]\mathbf{f}^{(k)}-\Lambda^C[\mathbf{f}_E^{(k+1)}]\mathbf{f}^{(k+1)}\right)=\mathbf{0}.\end{aligned}$$

We can rewrite it as

$$\begin{aligned}&(\mathbf{P}^T\mathbf{P}+\alpha_2\tau(\mathbb{L}_1+\mathbb{L}_2)+\alpha_1\mathbf{F}\Lambda^C[\mathbf{f}_E^{(k+1)}]\overline{\mathbf{F}}^T)(\hat{\mathbf{z}}^{(k+1)}-\hat{\mathbf{z}}^{(k)})\\=\;&\alpha_2\mathbf{F}(I\otimes D_+,D_+\otimes I)\left((\vec{\boldsymbol{\lambda}}^{(k+1)}-\vec{\boldsymbol{\lambda}}^{(k)})-(\vec{\boldsymbol{\lambda}}_E^{(k+1)}-\vec{\boldsymbol{\lambda}}_E^{(k)})\right)+\\&\alpha_2\tau(\mathbb{L}_1+\mathbb{L}_2)(\hat{\mathbf{z}}^{(k)}-\hat{\mathbf{z}}^{(k-1)})+\alpha_2\tau(\mathbb{L}_1+\mathbb{L}_2)(\hat{\tilde{Q}}_k-\hat{\tilde{Q}}_{k,E})\varepsilon_{tol}+\\&\alpha_1\mathbf{F}\left((A^C_{(k-1)}\mathbf{I}-\Lambda^C[\mathbf{f}_E^{(k)}])\mathbf{f}^{(k)}+(\Lambda^C[\mathbf{f}^{(k-1)}]-A^C_{(k-1)}\mathbf{I})\mathbf{f}^{(k-1)}\right)-\\&\alpha_1\mathbf{F}\left((A^C_{(k)}\mathbf{I}-\Lambda^C[\mathbf{f}_E^{(k+1)}])\mathbf{f}^{(k+1)}+(\Lambda^C[\mathbf{f}^{(k)}]-A^C_{(k)}\mathbf{I})\mathbf{f}^{(k)}\right)-\\&\alpha_1\mathbf{F}\left(\Lambda^C[\mathbf{f}_E^{(k+1)}]-\Lambda^C[\mathbf{f}_E^{(k)}]\right)\mathbf{z}^{(k)}.\end{aligned}$$

For matrix $\mathbb{B} := \mathbf{P}^T\mathbf{P} + \alpha_2\tau(\mathbb{L}_1 + \mathbb{L}_2) + \alpha_1\mathbf{F}\Lambda^C[\mathbf{f}_E^{(k+1)}]\overline{\mathbf{F}}^T$, since

$$\mathbb{A} := \alpha_1\mathbf{F}\Lambda^C[\mathbf{f}_E^{(k+1)}]\overline{\mathbf{F}}^T \sim \alpha_1\Lambda^C[\mathbf{f}_E^{(k+1)}]$$

which is diagonal, by Bauer-Fick theorem $|\lambda(\mathbb{B}) - \lambda(\mathbb{A})| \leq \|\mathbb{B} - \mathbb{A}\|_2$ [2], we can estimate $\lambda(\mathbb{B})$, the eigenvalues of $\mathbb{B}$, by

$$|\lambda(\mathbb{B})| \geq |\lambda(\mathbb{A})| - \|\mathbb{B} - \mathbb{A}\|_2 \geq \frac{\alpha_1}{\sqrt{1+\beta}} - (1 + 8\alpha_2\tau),$$

notice that $\mathbb{B} - \mathbb{A} = \mathbf{P}^T\mathbf{P} + \alpha_2\tau(\mathbb{L}_1 + \mathbb{L}_2)$ is diagonal with the elements between $[0, 1 + 8\alpha_2\tau]$. Using the norm equivalence for the matrix and the known estimate

$$\left\|\vec{\boldsymbol{\lambda}}^{(k+1)} - \vec{\boldsymbol{\lambda}}^{(k)}\right\|, \ \left\|\vec{\boldsymbol{\lambda}}_E^{(k+1)} - \vec{\boldsymbol{\lambda}}_E^{(k)}\right\| \leq \tau\varepsilon_{tol}$$

together with the bound $\frac{1}{\sqrt{\beta^3}}\alpha_1$ on the last three terms containing α_1, we have

$$\begin{aligned}
\left\|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\right\| &\leq \frac{1}{|\lambda(\mathbb{B})|}\left[\alpha_2\tau\varepsilon_{tol} + \alpha_2\tau\left\|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\right\| + \frac{\alpha_1}{\sqrt{\beta^3}}\right] \\
&\leq \frac{\alpha_2\tau}{\frac{\alpha_1}{\sqrt{1+\beta}} - (1 + 8\alpha_2\tau)}\left\|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\right\| + \\
&\quad \frac{1}{\frac{\alpha_1}{\sqrt{1+\beta}} - (1 + 8\alpha_2\tau)}\left(\alpha_2\tau\varepsilon_{tol} + \frac{\alpha_1}{\sqrt{\beta^3}}\right).
\end{aligned} \tag{53}$$

So, for the parameters satisfying $\alpha_1 > (1 + 8\alpha_2\tau)\sqrt{1+\beta}$ with small $\tau\varepsilon_{tol} > 0$ and large $\beta > 0$, it follows that $\left\|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\right\| \leq q_1\left\|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\right\| + q_2$ for constants $q_1, q_2 \in (0, 1)$, which leads to

$$\left\|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\right\| \leq q_1^k\left\|\mathbf{z}^{(1)} - \mathbf{z}^{(0)}\right\| + \frac{1}{1 - q_1}q_2,$$

i.e., $\{\mathbf{z}^{(k)} : k \in \mathbf{N}\}$ is almost a Cauchy sequence and consequently almost converges. Therefore $\{\mathbf{f}_E^{(k)} : k \in \mathbf{N}\}$ is almost convergent. Denote by

$$\mathbf{f}_E^{(k)} \to \mathbf{f}_E, \quad \vec{\boldsymbol{\lambda}}_E^{(k+1)} \to \vec{\boldsymbol{\lambda}}_E, \quad \mathbf{f}^{(k)} \to \mathbf{f}, \quad \vec{\boldsymbol{\lambda}}^{(k+1)} \to \vec{\boldsymbol{\lambda}}$$

as $k \to \infty$. By taking limit in (52), we have

$$\begin{aligned}
&\left(\overline{\mathbf{F}}^T\mathbf{P}^T\mathbf{P}\mathbf{F} + \alpha_1(\Lambda^C[\mathbf{f}] + \Lambda^C[\mathbf{f}_E])\right)(\mathbf{f} - \mathbf{f}_E) \\
= \ &\alpha_2(I \otimes D_+, D_+ \otimes I)(\vec{\boldsymbol{\lambda}} - \vec{\boldsymbol{\lambda}}_E) - \\
&\alpha_1(\Lambda^C[\mathbf{f}_E] - \Lambda^C[\mathbf{f}])\mathbf{f} - \alpha_2\tau(\mathbb{D}_1 + \mathbb{D}_2)\lim_{k\to\infty}(\tilde{q}_k - \tilde{q}_{k,E})\varepsilon_{tol}.
\end{aligned} \tag{54}$$

By updating process for $\lambda_j^{(k)}$ at each inner iteration, we have

$$\boldsymbol{\lambda}_j^{(k+1)} = \boldsymbol{\lambda}_j^{k,L(k)+1} = \boldsymbol{\lambda}_j^{k,L(k)} - \tau(\mathbf{w}_j^{k,L(k)+1} - \nabla\mathbf{f}_j^{(k)}) = \boldsymbol{\lambda}_j^{k,L(k)} - \tau(\mathbf{w}_j^{(k+1)} - \nabla\mathbf{f}_j^{(k)}),$$

and consequently $\vec{\boldsymbol{\lambda}}^{(k+1)} = \mathbf{vect}[(\boldsymbol{\lambda}^{k,L(k)})^T - \tau(\mathbf{w}^{(k+1)} - \nabla\mathbf{f}^{(k)})^T]$. Analogously, we also have $\vec{\boldsymbol{\lambda}}_E^{(k+1)} = \mathbf{vect}[(\boldsymbol{\lambda}_E^{k,L(k)})^T - \tau(\mathbf{w}_E^{(k+1)} - \nabla\mathbf{f}_E^{(k)})^T]$. Since $\left\|\boldsymbol{\lambda}^{k,L(k)}\right\|, \left\|\boldsymbol{\lambda}_E^{k,L(k)}\right\| \leq 1$, $\left\|\mathbf{w}^{(k+1)} - \nabla\mathbf{f}^{(k)}\right\|, \left\|\mathbf{w}_E^{(k+1)} - \nabla\mathbf{f}_E^{(k)}\right\| \leq \varepsilon_{tol}$, we have

$$\left\|\vec{\boldsymbol{\lambda}}^{(k+1)} - \vec{\boldsymbol{\lambda}}_E^{(k+1)}\right\| \leq C(1 + \tau\varepsilon_{tol}). \tag{55}$$

We again apply the Bauer-Fick theorem to get

$$\begin{aligned}&\left|\lambda\left(\overline{\mathbf{F}}^T\mathbf{P}^T\mathbf{P}\mathbf{F}+\alpha_1(\Lambda^C[\mathbf{f}]+\Lambda^C[\mathbf{f}_E])\right)\right|\\ \geq\ &\left|\lambda\left(\alpha_1(\Lambda^C[\mathbf{f}]+\Lambda^C[\mathbf{f}_E])\right)\right|-\left\|\overline{\mathbf{F}}^T\mathbf{P}^T\mathbf{P}\mathbf{F}\right\|_2\geq\frac{\alpha_1}{\sqrt{\beta+1}}-1\end{aligned}$$

due to $p_j = 0$ or 1, which leads to

$$\left\|\left(\overline{\mathbf{F}}^T\mathbf{P}^T\mathbf{P}\mathbf{F}+\alpha_1(\Lambda^C[\mathbf{f}]+\Lambda^C[\mathbf{f}_E])\right)^{-1}\right\|_2\leq\frac{1}{\frac{\alpha_1}{\sqrt{\beta+1}}-1}. \tag{56}$$

On the other hand, we also have

$$\left\|\Lambda^C[\mathbf{f}]-\Lambda^C[\mathbf{f}_E]\right\|\leq\frac{C}{\sqrt{\beta^3}}\left\|\mathbf{f}-\mathbf{f}_E\right\|. \tag{57}$$

Inserting (55)-(57) into (54) yields

$$\|\mathbf{f}-\mathbf{f}_E\|\leq C\frac{\alpha_1\sqrt{\beta+1}}{\alpha_1-\sqrt{\beta+1}}\left[\frac{1}{\sqrt{\beta^3}}\left\|\mathbf{f}-\mathbf{f}_E\right\|+\frac{\alpha_2}{\alpha_1}(1+\tau\varepsilon_{tol})\right]. \tag{58}$$

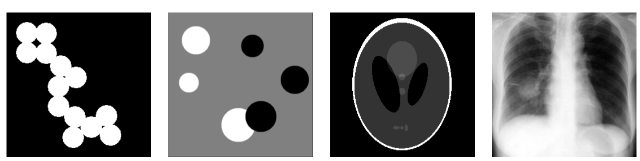
So for $\alpha_1>\sqrt{\beta+1}$ with large $\beta>0$ and small $\tau\varepsilon_{tol},\alpha_2>0$, we finally have

$$\|\mathbf{f}-\mathbf{f}_E\|\leq C(\alpha_2+\tau\varepsilon_{tol}).$$

The proof is complete. □

4. Numerical experiments.

In this section, we implement the AIS algorithm from noisy frequency sampling data numerically for different images with grey values between $[0,1]$, which are shown in Fig.1, where Fig.1(A) and Fig.1(B) are model images with grey values $\{0,1\},\{0,\frac{128}{255},1\}$ respectively, while Fig.1(C) is a standard phantom from Matlab, and Fig.1(D) is an MRI chest image. The number of pixels for these four images is $256\times256, 512\times512, 256\times256, 512\times512$, respectively.



(A) (B) (C) (D)

FIGURE 1. Object images: (A) circles under black background; (B) circles under gray background; (C) a phantom from Matlab; (D) an MRI chest image.

For testing our algorithm, we add the additive random noise in the frequency data of the image to yield the the noisy frequency data by

$$\hat{g}^{\delta}_{m',n'}=\hat{f}^R_{m',n'}+\delta\times rand(e_{m',n'})+i\cdot(\hat{f}^I_{m',n'}+\delta\times\widetilde{rand}(e_{m',n'})), \tag{59}$$

with fixed noise level $\delta=0.01$ and $rand(e_{m',n'}),\widetilde{rand}(e_{m',n'})\in[-1,1]$ for $m',n'=1,\cdots,N$, where $\hat{f}^R_{m',n'}+i\cdot\hat{f}^I_{m',n'}$ are the frequency data of exact image f. Then the inversion input data are sampling from (59) by some sampling operator $\mathcal{P}$. More precisely, the sampling masks $\mathcal{P}\{\hat{g}^{\delta}_{m',n'}:m',n'=1,\cdots,N\}$ for each image are specified as follows.

- $\mathcal{P}$ for Fig.1(A) and Fig.1(B)(random band sampling):
  For Fig.1(A), we sample 20 rows and 20 columns with $R_{center} = 0.3$ and sampling ratio $R_{total} = 15.02\%$, the sampling matrix is

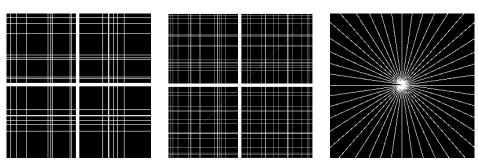
$$\begin{aligned} P \quad = \quad & \operatorname{diag}(0, \cdots, 0, p_9, 0, \cdots, 0, p_{19}, 0, \cdots, 0, p_{37}, 0, \cdots, 0, p_{75}, 0, \cdots, \\ & 0, p_{80}, 0, 0, p_{83}, 0 \cdots, 0, p_{114}, 0, 0, p_{117}, 0, \cdots, 0, p_{125}, p_{126}, p_{127}, \\ & p_{128}, p_{129}, p_{130}, 0, \cdots, 0, p_{172}, 0, \cdots, 0, p_{182}, p_{183}, 0, \cdots, 0, \\ & p_{191}, 0, \cdots, 0, p_{198}, 0, \cdots, p_{210}, 0, \cdots), \end{aligned} \tag{60}$$

  where $p_i = 1$ for $i = 9, 19, 37, \cdots, 210$.
  For Fig.1(B), we sample 40 rows and 40 columns with $R_{center} = 0.3$ and sampling ratio $R_{total} = 7.66\%$. The sampling matrix is

$$\begin{aligned} P \quad = \quad & \operatorname{diag}(0, \cdots, 0, p_9, 0, \cdots, 0, p_{19}, 0, \cdots, 0, p_{37}, 0, \cdots, 0, p_{75}, 0, \cdots, 0, \\ & p_{80}, 0, 0, p_{83}, 0 \cdots, 0, p_{114}, 0, 0, p_{117}, 0, \cdots, 0, p_{166}, 0, \cdots, 0, p_{176}, \\ & p_{177}, 0, \cdots, 0, p_{185}, 0, \cdots, 0, p_{192}, 0, \cdots, 0, p_{204}, 0, \cdots, 0, p_{250}, p_{251}, \\ & p_{252}, p_{253}, p_{254}, p_{255}, p_{256}, p_{257}, p_{258}, p_{259}, p_{260}, p_{261}, 0, \cdots, 0, p_{282}, \\ & 0, \cdots, 0, p_{293}, 0, \cdots, 0, p_{332}, 0, \cdots, 0, p_{346}, 0, p_{348}, 0, \cdots, 0, p_{376}, \\ & 0, \cdots, 0, p_{392}, 0, \cdots, 0, p_{424}, 0, p_{426}, 0, p_{428}, 0, \cdots, 0, p_{444}, 0, \cdots, 0, \\ & p_{476}, 0, \cdots, 0, p_{496}, 0, \cdots, 0, p_{501}, 0, \cdots, 0), \end{aligned} \tag{61}$$

  where $p_i = 1$ for $i = 9, 19, 37, \cdots, 501$.
- $\mathcal{P}$ for Fig.1(C) and Fig.1(D)(radial sampling):
  For Fig.1(C), we sample 22 lines with $R_{total} = 9.36\%$, while we sample sampling 44 lines with $R_{total} = 9.64\%$.

The sampling masks in frequency domain for the above four configurations are shown in Fig.2, respectively.



(A) (B) (C) (D)

FIGURE 2. Masks: (A) random sampling with 20 rows and 20 columns; (B) random sampling with 40 rows and 40 columns; (C) radial sampling with 22 lines; (D) radial sampling with 44 lines.

For the multi-regularizing pars, we choose regularizing parameters $\alpha_1 = \delta^3, \alpha_2 = \delta^2, \tau = 10$, the small approximation threshold $\beta = 0.01$ (or $\epsilon = 0.1$), the inner iteration for $\mathbf{w}$ is stopped with $\varepsilon_{tol} = 10^{-3}$, while the maximum iteration number for outer recursion is $K_0 = 100$. All numerical implementations are performed in MATLAB R2017b on a laptop with 1.6GHz Intel Core i5 processor and 8 GB of memory.

We introduce a new SNR index called improved signal to noise ratio (ISNR) together with the relative error (ReErr) to measure the reconstruction quality for an image, which are defined as

$$\text{ISNR} = 10\lg\left(\frac{\|\mathbf{f}^* - \mathcal{F}^{-1}\circ\mathbf{P}\hat{\mathbf{g}}^\delta\|_{l^2}}{\left\|\mathbf{f}^{(k)} - \mathbf{f}^*\right\|_{l^2}}\right),\ \text{ReErr} = \frac{\left\|\mathbf{f}^{(k)} - \mathbf{f}^*\right\|_{l^2}}{\|\mathbf{f}^*\|_{l^2}}, \tag{62}$$

where $\mathbf{f}^{(k)}$ and $\mathbf{f}^*$ are the reconstructed and exact image, respectively. Essentially, ISNR measures the reconstruction performance in terms of the ratio between input data error and reconstruction error, while ReErr is the standard relative error. The iteration number (IterNum) and the CPU time are used to evaluate the computational costs roughly.

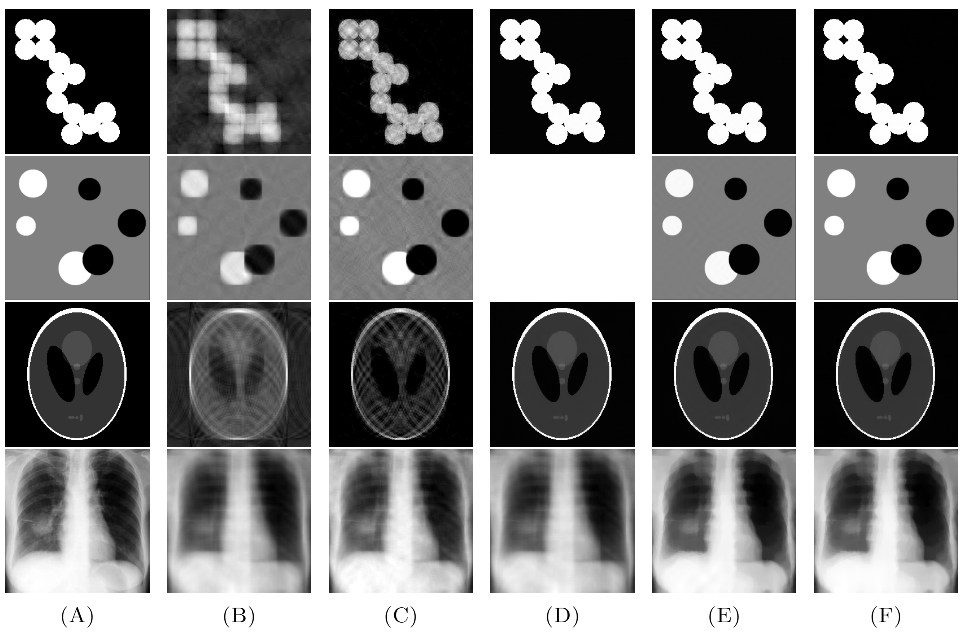


FIGURE 3. Reconstructions by random band sampling. From left to right: exact images, images by back projections, images by DM, images by RecPF, images by C-SMRM, images by H-SMRM.

To show the performances of our proposed scheme, we firstly give the reconstructions $\mathcal{F}^{-1}[\mathbf{P}\hat{\mathbf{g}}^\delta]$ (the second column in Fig.3 and Fig.5), which take the inverse Fourier transform on noisy sampling data directly without any iterative scheme. We call this method as back projection scheme. Then we compare our numerics with the iterative type schemes, i.e., direct method (DM) in [14] and the reconstruction from partial Fourier image data (RecPF) in [29] for each simplified multi-regularization model with Charbonnier approximation (C-SMRM) and Huber approximation (H-SMRM), respectively. With the stopping rules in (40), our experiments for alternating iterative algorithm stopped at minimum of $\left\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\right\|_{l^2}$ within maximum iterations K_0. The direct method (DM) considered the computations with wavelet sparsity of an image, by solving the Euler equation for $\mathbf{f}$ directly, without the introduction of alternative iterations, while RecPF method used the alternating direction method with multipliers (ADMM) to obtain the image restoration with an outstanding performance. The source codes of above methods in [14, 29] are publicly downloaded from the website. For fair comparisons, we have carefully adjusted the parameters in their algorithms so that their best reconstructions are compared with ours.

We firstly check the reconstruction performances for random band sampling. The reconstructions for different images are shown in Fig.3, and the error distributions are shown in Fig.4. The computational costs are given in Tab.1.

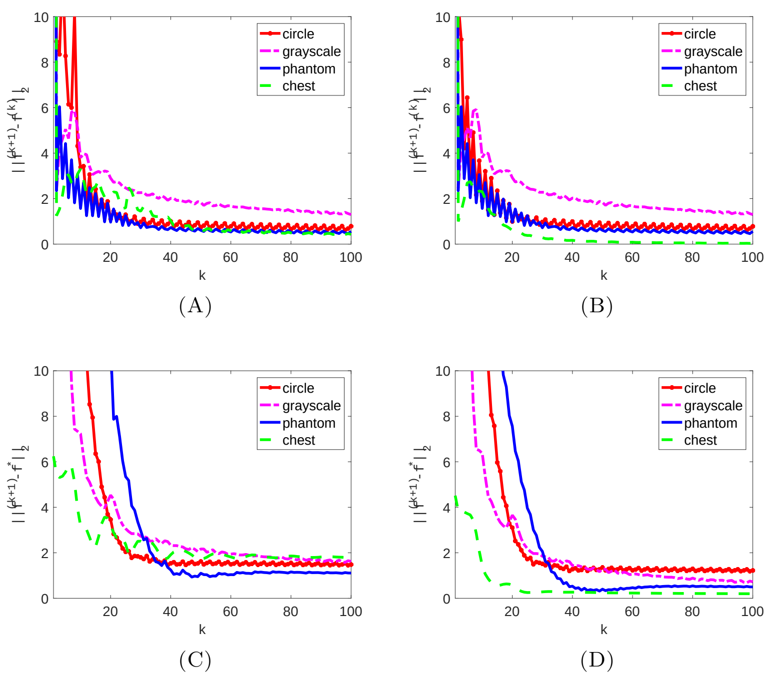


Figure 4. Errors by random band sampling: $\|\mathbf{f}^{(k+1)}-\mathbf{f}^{(k)}\|_{l^2}$ (top line); $\|\mathbf{f}^{(k+1)}-\mathbf{f}^{*}\|_{l^2}$ (bottom line). The first column is for C-SMRM method, while the second column is for H-SMRM method.

From the above figures, it can be found that all the schemes can reconstruct the objective image from the incomplete noisy frequency data. The direct method shown in Fig.3(C) can restore the image, but the grey values 1 for the circles part is not recovered well, and the reconstructed MRI image is no value for treatment, even though the direct method is much efficient to restore the image with multiplicative noise. Meanwhile, the circles image and gray-scale image as well as phantom image obtained by RecPF method is more clear especially for edge-preservation as shown in Fig.3(D), while the MRI chest image result obtained by RecPF is indistinct and lose much structured details. With the stopping rules in (40), the reconstruction results by both approximation approaches could be smaller than the results on the last step, i.e., the cost functionals are decreasing with respect to iteration times.

It is well known that the RecPF method with ADMM algorithm is a very competitive method for compressive sensing (CS) image restoration due to its remarkable performance. However, for these tested images in Fig.3, the performance of RecPF

TABLE 1. Computational costs for random band sampling.

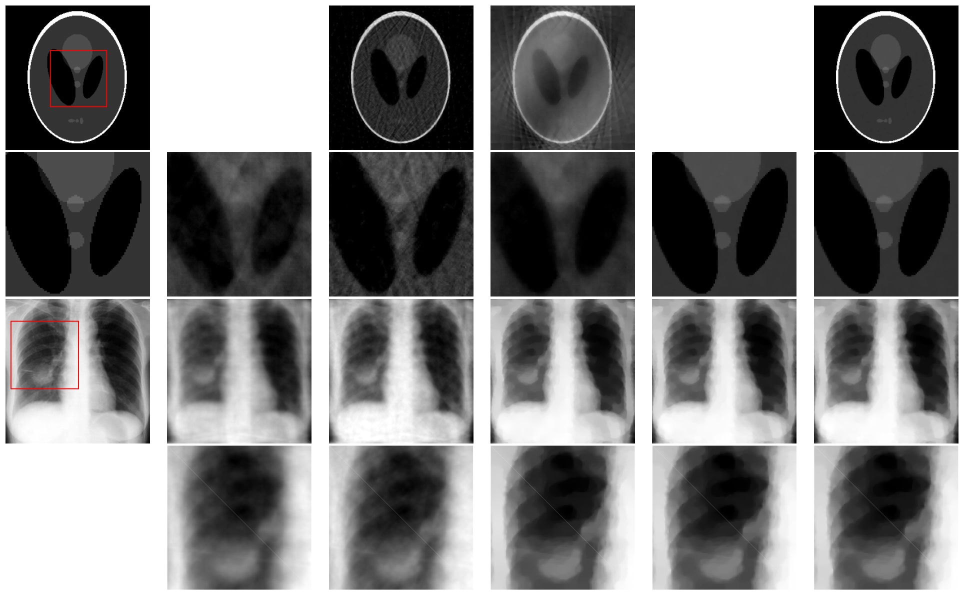
image	scheme	ISNR(dB)	ReErr(%)	CPU time(s)	IterNum
circles	direct	3.0437	17.0221	1.5734	40
	RecPF	14.6056	1.0492	**0.2680**	27
	C-SMRM	14.2219	1.1462	1.9248	100
	H-SMRM	**14.7205**	**1.0265**	1.3517	100
grayscale	direct	1.8183	5.7418	9.6594	40
	RecPF	11.9131	0.5245	**1.0943**	21
	C-SMRM	11.5221	0.5627	1.9255	100
	H-SMRM	**13.5264**	**0.2621**	39.1639	100
phantom	direct	1.7383	8.7405	2.1778	40
	RecPF	11.3009	1.9514	**0.3689**	38
	C-SMRM	11.4490	1.8190	1.2895	100
	H-SMRM	**14.4623**	**0.8047**	50.9566	100
chest	direct	2.6484	3.8296	2.1683	40
	RecPF	11.4839	1.0310	**0.2189**	21
	C-SMRM	11.5464	0.9734	1.7518	100
	H-SMRM	**17.4996**	**0.0164**	45.6364	100

method is undesirable (maybe due to the different sampling method). The alternative iteration scheme for both C-SMRM and H-SMRM can restore the image characteristics very well, i.e., it gives rise to better piecewise constant property and the image edges shown in the last two columns of Fig.3, than the details shown in Fig.3(C)-3(D). Obviously, our simplified model and approaches achieve the best visual quality and highest ISNR value among all test methods. This observation can be verified quantitatively from Tab.1, namely, H-SMRM from our algorithm achieves the best reconstruction performance with the index either ISNR or relative error, as compared with other schemes. From the definition of ISNR, the larger ISNR is, the smaller the error $\left\|\mathbf{f}^{(k)} - \mathbf{f}^*\right\|_{l^2}$, which means that the reconstruction error is smaller than the error obtained from using back projections.

Now we check the performances for radial sampling scheme for the complex situations, the phontam from Matlab and the MRI chest image. The reconstructions are given in Fig.5, while the decreasing behaviors of the errors with respect to the iteration times are given in Fig.6. The quantitative descriptions on the computational costs are shown in Tab.2.

From Fig.5 and Fig.6 together with Tab.2, we can reach the same observations for complex images (phantom from Matlab and MRI chest image) using radial sampling scheme as those using random band sampling data, i.e., H-SMRM from the AIS yield the best reconstructions, but with a little bitter large computational cost. The reconstruction differences between Charbonnier approximation and Huber approximation (shown in Fig.5(E) and Fig.5(F)) can be seen from Tab.2 quantitatively.

5. **Conclusion.** We consider an image reconstruction model from the data-fitting models with two regularizing terms. By modifying the total variational penalty term and l^1-norm penalty term using two smooth approximation, the corresponding optimization problem is solved by alternative iteration scheme (AIS) with augmented Lagrange penalty term. For this optimization problem, we propose an iterative scheme for updating two arguments as well as the weight of Lagrange penalty term



(A) (B) (C) (D) (E) (F)

FIGURE 5. Reconstructions by radial sampling. From left to right: exact images, images by back projections, images by DM, images by RecPF, images by C-SMRM, images by H-SMRM.

TABLE 2. Computational costs for radial sampling.

image	scheme	ISNR(dB)	ReErr(%)	CPU time(s)	IterNum
phantom	direct	3.8014	11.1883	2.4127	40
	RecPF	12.3966	2.7976	**0.4137**	36
	C-SMRM	12.0596	3.0233	1.3909	100
	H-SMRM	**13.0561**	**2.2250**	45.5967	100
chest	direct	4.0305	4.3061	2.2478	40
	RecPF	2.6372	4.5132	**0.2848**	22
	C-SMRM	11.4866	0.6724	1.3651	100
	H-SMRM	**12.5369**	**3.5845**	37.5576	100

by inner and outer iterations. The solution in the inner iteration has an explicit solution, while the solution in outer iteration can be determined by solving a linear system with diagonal coefficient matrix in frequency domain. Therefore the proposed AIS can be realized efficiently.

For this linearized iterative scheme, we further prove the convergence property of the iterative sequence in some generalized sense rigorously under suitably specified regularizing parameters. Such a generalized convergence reveals the essence of the AIS due to the finite accuracy approximation in inner iteration. Moreover, the error between the solution obtained from the linearize system and the solution from the nonlinear Euler equation at each step in outer iteration is quantitatively estimated.

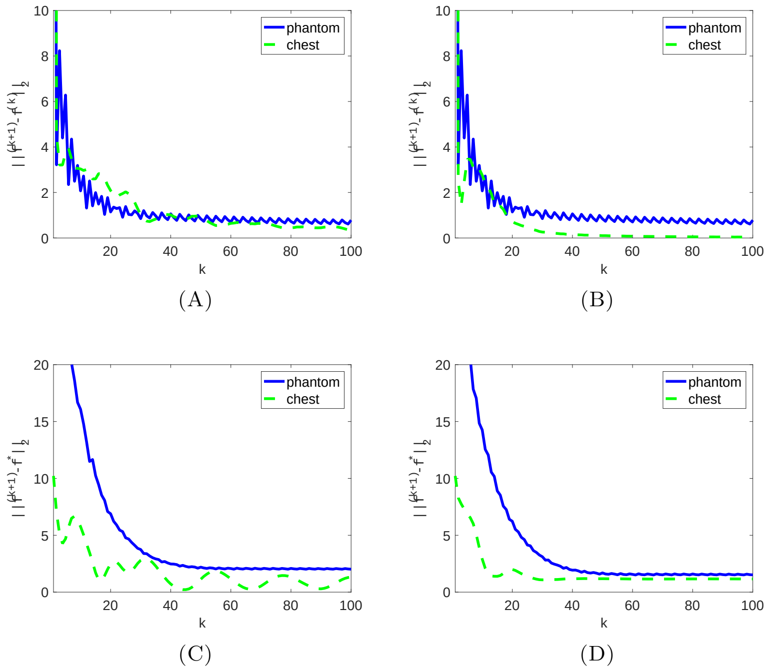


FIGURE 6. Error distributions by radial sampling. $\|\mathbf{f}^{(k+1)} - \mathbf{f}^{(k)}\|_{l^2}$ (top line), $\|\mathbf{f}^{(k+1)} - \mathbf{f}^{*}\|_{l^2}$ (bottom line). The first column is for C-SMRM, the second column is for H-SMRM.

We would like to emphasize that our work establishes the above theoretical analysis only on some type of images, and also the amount of computations is a little bitter heavy due to the introduction of two regularizing terms in the cost functional as compared with the cost functional with only one regularizing term. Such an extra cost is unavoidable, since both the sparsity and the discontinuity of an image are recovered in some balanced way.

As for the numerical performances, the alternative iteration scheme for both C-SMRM and H-SMRM model can improve the quality of restorations in an acceptable calculating time, noticing that our initial guess for the iteration is $\mathbf{f}^{(0)} = \mathbf{0}$. However, the relative error from proposed simplified model is smaller. For different smooth approximations, the optimization problem C-SMRM can yield a better reconstruction spending less time, as compared with the results by DM and RecPF. Additionally, the H-SMRM model can improve the quality of reconstructions obviously, even though the CPU time is longer than other approaches for identifying different structures of the objective images. Moreover, if we take the results by back projection scheme (shown in Fig.3(B) and Fig.5(B)) as our initial guess instead of $\mathbf{f}^{(0)} = \mathbf{0}$, the iterative process will converge quickly, leading to the decreasing of computational costs.

Acknowledgments. This work is supported by NSFC (No.11971104, 11531005, 11871149). The authors would like to thank the anonymous referees for their valuable comments and suggestions, which make our work much improved.

REFERENCES

[1] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*, Vol. 147, Applied Mathematical Sciences, 2^{nd} edition, Springer, New York, 2006.

[2] F. L. Bauer and C. T. Fike, Norms and exclusion theorems, *Numer. Math.*, **2** (1960), 137–141.

[3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1982.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends in Machine Learning*, **3** (2011), 1–122.

[5] E. J. Candès and J. Romberg, Sparsity and incoherence in compressive sampling, *Inverse Problems*, **23** (2007), 969–985.

[6] E. J. Candès, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory*, **52** (2006), 489–509.

[7] A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imaging Vision*, **20** (2004), 89–97.

[8] I. Daubechies, M. Defrise and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, **57** (2004), 1413–1457.

[9] D. L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, **52** (2006), 1289–1306.

[10] W. J. Fu, Penalized regressions: The bridge versus the lasso, *J. Comput. Graph. Statist.*, **7** (1998), 397–416.

[11] T. Goldstein and S. Osher, The split Bregman method for $L1$-regularized problems, *SIAM J. Imaging Sci.*, **2** (2009), 323–343.

[12] P. J. Huber, Robust estimation of a location parameter, *Ann. Math. Statist.*, **35** (1964), 73–101.

[13] E. M. Kalmoun, An investigation of smooth TV-like regularization in the context of the optical flow problem, *J. Imaging*, **4** (2018), 31.

[14] X. Liu and J. Liu, On image restoration from random sampling noisy frequency data with regularization, *Inverse Probl. Sci. Eng.*, **27** (2019), 1765–1789.

[15] K. Madsen and H. B. Nielsen, A finite smoothing algorithm for linear l_1 estimation, *SIAM J. Optim.*, **3** (1993), 223–235.

[16] M. K. Ng, R. H. Chan and W.-C. Tang, A fast algorithm for deblurring models with Neumann boundary conditions, *SIAM J. Sci. Comput.*, **21** (1999), 851–866.

[17] S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, An iterative regularization method for total variation-based image restoration, *Multiscale Model. Simul.*, **4** (2005), 460–489.

[18] S. Perkins, K. Lacker and J. Theiler, Grafting: Fast, incremental feature selection by gradient descent in function space, *J. Mach. Learn. Res.*, **3** (2003), 1333–1356.

[19] G. Plonka and J. Ma, Curvelet-wavelet regularized split Bregman iteration for compressed sensing, *Int. J. Wavelets Multiresolut. Inf. Process.*, **9** (2011), 79–110.

[20] M. Schmidt, G. Fung and R. Rosales, Fast optimization methods for $L1$ regularization: A comparative study and two new approaches, in *Machine Learning: ECML 2007*, Lecture Notes in Computer Science, 4701, Springer, Berlin Heidelberg, 2007, 286–297.

[21] S. K. Shevade and S. S. Keerthi, A simple and efficient algorithm for gene selection using sparse logistic regression, *Bioinformatics*, **19** (2003), 2246–2253.

[22] X. Shu and N. Ahuja, Hybrid compressive sampling via a new total variation TVL1, in *Computer Vision – ECCV 2010*, Lecture Notes in Computer Science, 6316, Springer, Berlin, Heidelberg, 2010, 393–404.

[23] W. Sun and Y.-X. Yuan, *Optimization Theory and Methods*, Optimization and Its Applications, 1, 1^{st} edition, Springer, New York, 2006.

[24] C. R. Vogel, *Computational Methods for Inverse Problems*, Frontiers in Applied Mathematics, 23, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

[25] Y. Wang, J. Yang, W. Yin and Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, *SIAM J. Imaging Sci.*, **1** (2008), 248–272.

[26] J. Yang, W. Yin, Y. Zhang and Y. Wang, A fast algorithm for edge-preserving variational multichannel image restoration, *SIAM J. Imaging Sci.*, **2** (2009), 569–592.

[27] J. Yang, Y. Zhang and W. Yin, A fast TVL1-L2 minimization algorithm for signal reconstruction from partial Fourier data, 2008.

[28] J. Yang, Y. Zhang and W. Yin, An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise, *SIAM J. Sci. Comput.*, **31** (2009), 2842–2865.
[29] J. Yang, Y. Zhang and W. Yin, A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data, *IEEE Journal of Selected Topics in Signal Processing*, **4** (2010), 288–297.
[30] W. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman iterative algorithms for I_1-minimization with applications to compressed sensing, *SIAM J. Imaging Sci.*, **1** (2008), 143–168.
[31] X. Zhang, M. Burger, X. Bresson and S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, *SIAM J. Imaging Sci.*, **3** (2010), 253–276.
[32] Y. Zhu and I.-L. Chern, Convergence of the alternating minimization method for sparse MR image reconstruction, *J. Inform. Comput. Sci.*, **8** (2011), 2067–2075.
[33] Y. Zhu and X. Liu, A fast method for L1-L2 modeling for MR image compressive sensing, *J. Inverse Ill-Posed Probl.*, **23** (2015), 211–218.
[34] Y. Zhu, Y. Shi, B. Zhang and X. Yu, Weighted-average alternating minimization method for magnetic resonance image reconstruction based on compressive sensing, *Inverse Probl. Imaging*, **8** (2014), 925–937.

Received March 2019; 1^{st} revision October 2019; 2^{nd} revision February 2020.

E-mail address: liuxm@njau.edu.cn
E-mail address: jjliu@seu.edu.cn