# Supplementary Material

## I. Proof of Lemma 1

**Lemma 1: Bound of the trainable denoiser**. *For any input $\boldsymbol{x} \in \mathbb{R}^N$ whose elements admit $x_i \in [0,1]$ and some universal constant $L = c_{max}^2$ independent of $M$, the proposed trainable denoiser based on dual tight frames* is bounded such that

$$||\boldsymbol{x} - \boldsymbol{\Phi} T_{\varepsilon(\sigma)}(\boldsymbol{\Psi x})||_2^2 \le M\sigma^2 L. \tag{1}$$

***Proof***: In Eqn. (1), $\boldsymbol{\Psi} \in \mathbb{R}^{M\times N}$ and $\boldsymbol{\Phi} \in \mathbb{R}^{N\times M}$ are dual tight frames satisfying the dual property $\boldsymbol{\Phi\Psi} = \boldsymbol{I}$ and tight property $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} = \boldsymbol{I}$, $\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi} = \boldsymbol{I}$. Therefore, we have

$$\begin{aligned} &||\boldsymbol{x} - \boldsymbol{\Phi} T_{\varepsilon(\sigma)}(\boldsymbol{\Psi x})||_2^2 \\ &= ||\boldsymbol{\Phi\Psi x} - \boldsymbol{\Phi} T_{\varepsilon(\sigma)}(\boldsymbol{\Psi x})||_2^2 \\ &\le ||\boldsymbol{\Phi}||_2^2 ||\boldsymbol{\Psi x} - T_{\varepsilon(\sigma)}(\boldsymbol{\Psi x})||_2^2. \end{aligned} \tag{2}$$

Due to $||\boldsymbol{\Phi}||_2^2 = \lambda_{max}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}) = 1$. Therefore, the upper bound of Eqn. (2) can be further determined as follows:

$$||\boldsymbol{x} - \boldsymbol{\Phi} T_{\varepsilon(\sigma)}(\boldsymbol{\Psi x})||_2^2 \le ||\boldsymbol{\Psi x} - T_{\varepsilon(\sigma)}(\boldsymbol{\Psi x})||_2^2. \tag{3}$$

Eqn. (3) can be recast as

$$||\boldsymbol{x} - \boldsymbol{\Phi} T_{\varepsilon(\sigma)}(\boldsymbol{\Psi x})||_2^2 \le ||\boldsymbol{\Psi x} - T(\boldsymbol{\Psi x}, \boldsymbol{e})||_2^2. \tag{4}$$

Here, we use $\boldsymbol{e} \in \mathbb{R}^M$ to represent the threshold vector whose elements are utilized for shrinking $\boldsymbol{\Psi x}$. Let $(\boldsymbol{\Psi x})_i$ represents the *i*-th element of $\boldsymbol{\Psi x}$ and $\varepsilon_i$ denotes the *i*-th element of $\boldsymbol{e}$. The soft thresholding operator $T[(\boldsymbol{\Psi x})_i, \varepsilon_i]$ is defined as

$$T[(\boldsymbol{\Psi x})_i, \varepsilon_i] = \begin{cases} (\boldsymbol{\Psi x})_i + \varepsilon_i, & (\boldsymbol{\Psi x})_i < -\varepsilon_i \\ 0, & |(\boldsymbol{\Psi x})_i| \le \varepsilon_i \\ (\boldsymbol{\Psi x})_i - \varepsilon_i, & (\boldsymbol{\Psi x})_i > \varepsilon_i. \end{cases} \tag{5}$$

We consider the shrinking process $T(\boldsymbol{\Psi x}, \boldsymbol{e})$, one of the following situations will happen.

**S1**: Any element of $\boldsymbol{\Psi x}$ satisfies $(\boldsymbol{\Psi x})_i < -\varepsilon_i$, then $T(\boldsymbol{\Psi x}, \boldsymbol{e}) = \boldsymbol{\Psi x} + \boldsymbol{e}$. Therefore,

$$||\boldsymbol{\Psi x} - T(\boldsymbol{\Psi x}, \boldsymbol{e})||_2^2 = ||\boldsymbol{\Psi x} - \boldsymbol{\Psi x} - \boldsymbol{e}||_2^2 = ||\boldsymbol{e}||_2^2. \tag{6}$$

**S2**: Any element of $\boldsymbol{\Psi x}$ satisfies $|(\boldsymbol{\Psi x})_i| \le \varepsilon_i$, then $T(\boldsymbol{\Psi x}, \boldsymbol{e}) = \boldsymbol{0}$. Therefore,

$$||\boldsymbol{\Psi x} - T(\boldsymbol{\Psi x}, \boldsymbol{e})||_2^2 = ||\boldsymbol{\Psi x} - \boldsymbol{0}||_2^2 \le ||\boldsymbol{e}||_2^2. \tag{7}$$

**S3**: Any element of $\boldsymbol{\Psi x}$ satisfies $(\boldsymbol{\Psi x})_i > \varepsilon_i$, then $T(\boldsymbol{\Psi x}, \boldsymbol{e}) = \boldsymbol{\Psi x} - \boldsymbol{e}$. Therefore,

$$||\boldsymbol{\Psi x} - T(\boldsymbol{\Psi x}, \boldsymbol{e})||_2^2 = ||\boldsymbol{\Psi x} - \boldsymbol{\Psi x} + \boldsymbol{e}||_2^2 = ||\boldsymbol{e}||_2^2. \tag{8}$$

**S4**: Any two or all three of the above situations occur. **S4** is a union of the **S1**, **S2**, and **S3**. Therefore, as long as we find the upper bound of $||\boldsymbol{\Psi x} - T(\boldsymbol{\Psi x}, \boldsymbol{e})||_2^2$ under the **S1**-**S3**, the upper bound of that under the **S4** will be determined. Hence, when **S4** happens, the upper bound of $||\boldsymbol{\Psi x} - T(\boldsymbol{\Psi x}, \boldsymbol{e})||_2^2$ is the maximum upper bound of the first three situations. Based on Eqn. (6), Eqn. (7), and Eqn. (8), we have

$$||\boldsymbol{\Psi x} - T(\boldsymbol{\Psi x}, \boldsymbol{e})||_2^2 \le ||\boldsymbol{e}||_2^2. \tag{9}$$

Recall the definition of threshold vector $\boldsymbol{e} = \boldsymbol{c} \odot \boldsymbol{m}$, and each element of proportional constant vectors $\boldsymbol{c}$ has a limited range $c_i \in [c_{min}, c_{max}]$. Let $\varepsilon_{max} = c_{max} \cdot \sigma$ denote the maximum element of $\boldsymbol{e}$. Thus,

$$||\boldsymbol{e}||_2^2 \le M\varepsilon_{max}^2 \le Mc_{max}^2\sigma^2. \tag{10}$$

According to Eqn. (4), Eqn. (9), and Eqn. (10), we can get

$$||\boldsymbol{x} - \boldsymbol{\Phi} T_{\varepsilon(\sigma)}(\boldsymbol{\Psi x})||_2^2 \le M\sigma^2 c_{max}^2 \le M\sigma^2 L. \tag{11}$$

Here, $L = c_{max}^2$.

This completes the proof.

## II. proof of Lemma 2

**Lemma 2: Bound of the BM3D denoiser**. *We assume that there exists two monotone decreasing functions $f(\sigma)$ and $\varepsilon(\sigma)$ admitting $\lim_{\sigma\to 0} f(\sigma) = p$ and $\lim_{\sigma\to 0} \varepsilon(\sigma) = 0$. For any input $\boldsymbol{x} \in \mathbb{R}^N$ whose elements admit $x_i \in [0,1]$ and some universal constant $\widetilde{L} = pC_{max}^2$*, the BM3D denoiser is bounded such that

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}} H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le [\varepsilon(\sigma)]^2[p - f(\sigma)] \le \sigma^2\widetilde{L}. \tag{12}$$

***Proof***: Firstly, we introduce the matrix and frame representations of analysis and synthesis BM3D operations. Let $\boldsymbol{x}_n \in \mathbb{R}^N$ be a noisy image, and $\boldsymbol{R}_i\boldsymbol{x}_n \in \mathbb{R}^n$ be the $i$th image patch extracted from the noisy image $\boldsymbol{x}_n$. Here, matrix $\boldsymbol{R}_i \in \mathbb{R}^{n\times N}$ represents the operator that extracts the patch from the whole image. The total number of image patches in each group is a fixed number $K$, and the total number of groups is $R$. Let $J_r = \{i_{r,1}, \dots, i_{r,k}\}$ be the set of image patch indexes in the $r$th group, thus grouping can be defined by $J = \{J_r : r = 1, \dots, R\}$. Based on these notations, the explicit matrix representation of the BM3D analysis operation can be written as

$$\boldsymbol{w} = \boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}_n = \begin{bmatrix} \boldsymbol{\Psi}_{\mathrm{BM3D}_1} \\ \cdots \\ \boldsymbol{\Psi}_{\mathrm{BM3D}_R} \end{bmatrix} \boldsymbol{x}_n \tag{13}$$

where $\boldsymbol{w}$ represents the joint 3D groupwise spectrum, and $\boldsymbol{\Psi}_{\mathrm{BM3D}_R}$ is defined as

$$\boldsymbol{\Psi}_{\mathrm{BM3D}_R} = \sum_{i\in J_r} \boldsymbol{d}_i \otimes [(\boldsymbol{D}_2 \otimes \boldsymbol{D}_2)\boldsymbol{R}_i] \tag{14}$$

where $\otimes$ represents the Kronecker matrix product. The 3D decorrelating transform is constructed as a separable combination of 2D intrablock and 1D interblock transforms. Here, each block is an image patch. In Eqn. (14), $\boldsymbol{d}_i$ is the $i$th column of 1D interblock transform $\boldsymbol{D}_1$, and $\boldsymbol{D}_2$ represents the 1D transform that constitutes the separable 2D intrablock transform.

The synthesis matrix is derived similarly. The estimated image is the weighted mean of the groupwize estimation, and the weights are defined as $g_r > 0$. The synthesis representation model for a noisy image can be represented as

$$\boldsymbol{x}_n = \boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{w} = \boldsymbol{W}^{-1}[g_1\boldsymbol{\Phi}_{\mathrm{BM3D}_1}, \ldots, g_R\boldsymbol{\Phi}_{\mathrm{BM3D}_R}]\boldsymbol{w} \tag{15}$$

where $\boldsymbol{W} = \sum\limits_r g_r \sum\limits_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i$. The explicit matrix representation of the BM3D synthesis operation can be formulated as

$$\boldsymbol{\Phi}_{\mathrm{BM3D}_R} = \sum_{i\in J_r} \boldsymbol{d}_i^{\mathrm{T}} \otimes [\boldsymbol{R}_i^{\mathrm{T}}(\boldsymbol{D}_2 \otimes \boldsymbol{D}_2)^{\mathrm{T}}]. \tag{16}$$

**Remark** Based on the definition of the analysis and synthesis BM3D operations, the matrix $\boldsymbol{\Psi}_{\mathrm{BM3D}}$ admits: $\boldsymbol{\Psi}_{\mathrm{BM3D}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathrm{BM3D}} = \sum\limits_r \sum\limits_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i > 0$ and the matrix $\boldsymbol{\Phi}_{\mathrm{BM3D}}$ admits: $\boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{\Phi}_{\mathrm{BM3D}}^{\mathrm{T}} = \boldsymbol{W}^{-2}\sum\limits_r {g_r}^2 \sum\limits_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i > 0$. Therefore, the following equation hold for the underlying or clean image $\boldsymbol{x} \in \mathbb{R}^N$ :

$$\boldsymbol{x} = \boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}. \tag{17}$$

Based on the matrix representations of analysis and synthesis BM3D operations, the analysis and synthesis BM3D frames are defined as follows. Definition of analysis BM3D frame: It follows from Remark that rows of $\boldsymbol{\Psi}_{\mathrm{BM3D}}$ constitute frame in $\{\boldsymbol{\Psi}_{\mathrm{BM3D}_k}\}$; Definition of synthesis BM3D frame: It follows from Remark that columns of $\boldsymbol{\Phi}_{\mathrm{BM3D}}$ constitute frame in $\{\boldsymbol{\Phi}_{\mathrm{BM3D}_k}\}$. It follows the Remark that the BM3D frames admit dual property

$$\boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{\Psi}_{\mathrm{BM3D}} = \boldsymbol{I}. \tag{18}$$

According to the analysis and synthesis BM3D frames, the BM3D denoiser can be described as

$$D_{\mathrm{BM3D}}(\bullet;\sigma) = \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}[\boldsymbol{\Psi}_{\mathrm{BM3D}}(\bullet)] \tag{19}$$

where $H_{\varepsilon(\sigma)}$ is the hard thresholding operator using the thresholding value $\varepsilon(\sigma)$. The hard thresholding operator is defined as $H_{\varepsilon(\sigma)}(u) = u$ if $|u| > \varepsilon(\sigma)$, and $H_{\varepsilon(\sigma)}(u) = 0$, otherwise. The parameter $\varepsilon(\sigma)$ is correlated with the input noise standard deviation $\sigma$. In general, the thresholding value $\varepsilon(\sigma)$ is a monotone decreasing function of the input parameter $\sigma$, and this function admits $\lim\limits_{\sigma\to 0} \varepsilon(\sigma) = 0$.

Then, in order to analyze the bounded property of the BM3D denoiser, we give the following assumption.

**Assumption:** The number of coefficients in $H_{\varepsilon(\sigma)}[\boldsymbol{\Psi}_{\mathrm{BM3D}}\ \boldsymbol{x}^{(t-1)}]$ is denoted as $m$, and the number of coefficients in $\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}^{(t-1)}$ is $p$, we assume that there exists a decreasing function $f(\sigma)$ such that $m = f(\sigma)$. The BM3D denoiser is asymptotically invariant in the sense that it ensures $\boldsymbol{x}^{(t-1)} = \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}[\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}^{(t-1)}]$ as $\sigma \to 0$, i.e., $\lim\limits_{\sigma\to 0} f(\sigma) = p$.

Finally, based on the Eqn. (18) and Eqn. (19), we have

$$\begin{aligned}&||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \\ &= ||\boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \\ &\le ||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2||\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x} - H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2.\end{aligned} \tag{20}$$

Here, $||\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x} - H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2$ represents the square of the $l_2$ norm of the filtered spectrum or filtered coefficient vector. Among these lost coefficients, the bound of these filtered coefficients is the threshoulding value $\varepsilon(\sigma)$. Therefore,

$$\begin{aligned}&||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \\ &\le ||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2[\varepsilon(\sigma)]^2[p - f(\sigma)]\end{aligned} \tag{21}$$

where $[p - f(\sigma)]$ is the number of the lost or filtered coefficients. Due to $||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2 = \lambda_{max}(\boldsymbol{\Phi}_{\mathrm{BM3D}}^{\mathrm{T}}\boldsymbol{\Phi}_{\mathrm{BM3D}})$ (here, $\lambda_{max}(\bullet)$ represents the maximum eigenvalue), we have

$$||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2 = \lambda_{max}\left(\frac{\sum\limits_r {g_r}^2 \sum\limits_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i}{\sum\limits_r g_r \sum\limits_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i \bullet \sum\limits_r g_r \sum\limits_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i}\right). \tag{22}$$

Letting $\boldsymbol{A} = \sum\limits_r {g_r}^2 \sum\limits_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i$ and $\boldsymbol{B} = \sum\limits_r g_r \sum\limits_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i$, these two matrices are diagonal matrices. In general, it is assumed that for each pixel there is at least one image patch containing the pixel and entering in some group. Moreover, the weights $g_r$ are in the range of $(0, 1]$. Based on these assumptions, the element values in $\boldsymbol{A}$ are in the range of $[{g_r}^2, {g_r}^2KR]$, and the element values in $\boldsymbol{B}$ are in the range of $[g_r, g_rKR]$. Therefore, the matrix $\frac{\boldsymbol{A}}{\boldsymbol{B}\boldsymbol{B}}$ is a diagonal matrix whose diagonal elements are in the range of $[\frac{1}{KR}, 1]$. The maximum eigenvalue of the $\boldsymbol{\Phi}_{\mathrm{BM3D}}^{\mathrm{T}}\boldsymbol{\Phi}_{\mathrm{BM3D}}$ is 1, i.e.,

$$||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2 \le 1. \tag{23}$$

Substituting Eqn. (23) into Eqn. (21), we have

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le [\varepsilon(\sigma)]^2[p - f(\sigma)]. \tag{24}$$

According to $p - f(\sigma) \le p$, the upper bound of Eqn. (24) can be further determined as follows:

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le [\varepsilon(\sigma)]^2 p. \tag{25}$$

Due to the $\varepsilon(\sigma)$ is a monotone decreasing function of $\sigma$, there must exist a function $C(\sigma)$ satisfying $\varepsilon(\sigma) = C(\sigma)\sigma$. For a practical BM3D denoiser, the maximum value of $C(\sigma)$ must exist, and we term it as $C_{max}$. Then Eqn. (25) can be recast as

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le [C(\sigma)\sigma]^2 p \le C_{max}^2\sigma^2 p. \tag{26}$$

Let $\widetilde{L} = pC_{max}^2$, we have

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le \sigma^2\widetilde{L}. \tag{27}$$

This completes the proof.

## III. Proof of Lemma 3

**Lemma 3: The bounded gradient of the CSMRI data-fidelity function.** Consider the sampling model $\boldsymbol{y} = \boldsymbol{F}_u\boldsymbol{x} + \boldsymbol{n}$ and $||\boldsymbol{n}||_2 \le \kappa$, then, the function $f(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{F}_u\boldsymbol{x}||_2^2 : [0,1]^N \to \mathbb{R}$ has the bounded gradient, i.e., $||\nabla f(\boldsymbol{x})||_2 \le 2\sqrt{N} + \kappa$.

***Proof***: The gradient of $f(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{F}_u\boldsymbol{x}||_2^2$ in the CSMRI formulation is

$$\nabla f(\boldsymbol{x}) = \boldsymbol{F}_u^{\mathrm{H}}(\boldsymbol{F}_u\boldsymbol{x} - \boldsymbol{y}), \tag{28}$$

thus, we have

$$\begin{aligned}||\nabla f(\boldsymbol{x})||_2 &\leq ||\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{x}||_2 + ||\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{y}||_2 \\ &\leq ||\boldsymbol{F}_u^{\mathrm{H}}||_2[||\boldsymbol{F}_u\boldsymbol{x}||_2 + ||\boldsymbol{y}||_2].\end{aligned} \tag{29}$$

Since $||\boldsymbol{F}_u\boldsymbol{x}||_2 \leq ||\boldsymbol{F}\boldsymbol{x}||_2 = ||\boldsymbol{x}||_2$ and the pixel values in the image $\boldsymbol{x}$ are normalized into $[0, 1]$, we obtain

$$||\boldsymbol{F}_u\boldsymbol{x}||_2 \leq \sqrt{N}. \tag{30}$$

In Eqn. (29), $||\boldsymbol{y}||_2 = ||\boldsymbol{F}_u\boldsymbol{x}_{ori} + \boldsymbol{n}||_2$ and $||\boldsymbol{n}||_2 \leq \kappa$, here, $\boldsymbol{x}_{ori}$ is the original image whose pixel values are normalized into [0,1]. Therefore,

$$||\boldsymbol{y}||_2 = ||\boldsymbol{F}_u\boldsymbol{x}_{ori} + \boldsymbol{n}||_2 \leq ||\boldsymbol{F}_u\boldsymbol{x}_{ori}||_2 + ||\boldsymbol{n}||_2 \leq \sqrt{N} + \kappa. \tag{31}$$

Substituting Eqn. (30) and Eqn. (31) into Eqn. (29), we get

$$\begin{aligned}||\nabla f(\boldsymbol{x})||_2 &\leq ||\boldsymbol{F}_u^{\mathrm{H}}||_2(2\sqrt{N} + \kappa) \\ &= \sqrt{\lambda_{max}(\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}})}(2\sqrt{N} + \kappa)\end{aligned} \tag{32}$$

where $\lambda_{max}(\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}})$ represents the maximum eigenvalue of $\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}}$. Now, we focus on $\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}}$. The sampling operator $\boldsymbol{F}_u$ can be regarded as $\boldsymbol{F}_u = \boldsymbol{P}\boldsymbol{F}$, here, $\boldsymbol{P} \in \mathbb{R}^{M\times N}$ is a mask matrix whose element 1 indicates the corresponding index of the sampling point. Hence,

$$\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}} = \boldsymbol{P}\boldsymbol{F} \bullet (\boldsymbol{P}\boldsymbol{F})^{\mathrm{H}} = \boldsymbol{P}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{P}^{\mathrm{H}} = \boldsymbol{P}\boldsymbol{P}^{\mathrm{H}} = \boldsymbol{I} \tag{33}$$

where $\boldsymbol{I} \in \mathbb{R}^{M\times M}$ is an identify matrix. Therefore, we obtain $\lambda_{max}(\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}}) = 1$, this leads to

$$||\nabla f(\boldsymbol{x})||_2 \leq 2\sqrt{N} + \kappa. \tag{34}$$

This completes the proof.

## IV. Proof of Lemma 4

**Lemma 4**: At the $t$th iteration, if Case 1 holds, then

$$||\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}||_2 \leq \frac{C_3}{\sqrt{\beta^{(t-1)}}} \tag{35}$$

for some constants $C_3$.

***Proof***: Let us consider the proposed formulation

$$\boldsymbol{x}^{(t)} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{y} - \boldsymbol{F}_u\boldsymbol{x}||_2^2 + \beta||\boldsymbol{x} - D_{\mathrm{BMDual}}(\boldsymbol{x}^{(t-1)}; \sigma)||_2^2. \tag{36}$$

Based on the first order optimality, we have

$$\nabla f(\boldsymbol{x}) + \beta(\boldsymbol{x} - D_{\mathrm{BMDual}}(\boldsymbol{x}^{(t-1)}; \sigma)) = 0. \tag{37}$$

We assume that the optimal solution is $\boldsymbol{x}^{(t)}$ at the $t$th iteration. Substituting this solution into the above equation yields

$$\boldsymbol{x}^{(t)} - D_{\mathrm{BMDual}}(\boldsymbol{x}^{(t-1)}; \sigma) = -\frac{\nabla f(\boldsymbol{x}^{(t)})}{\beta^{(t-1)}}. \tag{38}$$

At the $t$th iteration, the value of $\beta$ for updating image is $\beta^{(t-1)}$. Using the Lemma 3, we have

$$\begin{aligned}||\boldsymbol{x}^{(t)} - D_{\mathrm{BMDual}}(\boldsymbol{x}^{(t-1)}; \sigma)||_2 &= ||-\frac{\nabla f(\boldsymbol{x}^{(t)})}{\beta^{(t-1)}}||_2 \\ &\leq \frac{(2\sqrt{N} + \kappa)}{\beta^{(t-1)}}.\end{aligned} \tag{39}$$

Based on this bound and the Theorem 1, we obtain

$$\begin{aligned}&||\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}||_2 \\ &= ||\boldsymbol{x}^{(t)} - D_{\mathrm{BMDual}}(\boldsymbol{x}^{(t-1)}; \sigma) \\ &+ D_{\mathrm{BMDual}}(\boldsymbol{x}^{(t-1)}; \sigma) - \boldsymbol{x}^{(t-1)}||_2 \\ &\leq ||\boldsymbol{x}^{(t)} - D_{\mathrm{BMDual}}(\boldsymbol{x}^{(t-1)}; \sigma)||_2 \\ &+ ||D_{\mathrm{BMDual}}(\boldsymbol{x}^{(t-1)}; \sigma) - \boldsymbol{x}^{(t-1)}||_2 \\ &\leq \frac{(2\sqrt{N} + \kappa)}{\beta^{(t-1)}} + \sqrt{k^2[\sigma^{(t)}]^2\widetilde{L} + (1-k)^2M[\sigma^{(t)}]^2L} \\ &\leq \frac{(2\sqrt{N} + \kappa)}{\beta^{(t-1)}} + k\sigma^{(t)}\sqrt{\widetilde{L}} + (1-k)\sigma^{(t)}\sqrt{ML}.\end{aligned} \tag{40}$$

The second inequality holds true due to the triangle inequality. Letting $C_1 = k\sqrt{\widetilde{L}} + (1-k)\sqrt{ML}$, $C_2 = 2\sqrt{N} + \kappa$, and using $\sigma^{(t)} = \sqrt{\lambda/\beta^{(t-1)}}$, thus, (40) can be rewritten as

$$\begin{aligned}||\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}||_2 &\leq \frac{C_2}{\beta^{(t-1)}} + C_1\sigma^{(t)} \\ &= \frac{C_2}{\beta^{(t-1)}} + C_1\sqrt{\frac{\lambda}{\beta^{(t-1)}}} \\ &= [\frac{C_2}{\sqrt{\beta^{(t-1)}}} + C_1\sqrt{\lambda}]\frac{1}{\sqrt{\beta^{(t-1)}}} \\ &\leq [\frac{C_2}{\sqrt{\beta^{(0)}}} + C_1\sqrt{\lambda}]\frac{1}{\sqrt{\beta^{(t-1)}}} \\ &\leq \frac{C_3}{\sqrt{\beta^{(t-1)}}}\end{aligned} \tag{41}$$

where $C_3 = \frac{C_2}{\sqrt{\beta^{(0)}}} + C_1\sqrt{\lambda}$.

This completes the proof.

## V. Proof of Lemma 5

**Lemma 5**: Regardless which of **S1**-**S3**, for any $t$, the sequence $\left\{\boldsymbol{x}^{(t)}\right\}_{t=1}^{\infty}$ always satisfies

$$d(\boldsymbol{x}^{(t)}, \boldsymbol{x}^{(t-1)}) \leq C_6\delta^t \tag{42}$$

for some constants $C_6$ and $0 < \delta < 1$.

***Proof***: For any case **S1**-**S3**, at the $t$th iteration, it holds that

$$\begin{aligned}d(\boldsymbol{x}^{(t)}, \boldsymbol{x}^{(t-1)}) &\leq max(\frac{C_3}{\sqrt{\gamma}^{t-T_1}\sqrt{\beta^{(T_1-1)}}}, \eta^{t-T_2}\frac{C_3}{\sqrt{\beta^{(T_2-1)}}}) \\ &= max\{C_4(\frac{1}{\sqrt{\gamma}})^t, C_5\eta^t\}\end{aligned} \tag{43}$$

where $C_4 = \frac{C_3}{\sqrt{\gamma}^{-T_1}\sqrt{\beta^{(T_1-1)}}}$ and $C_5 = \eta^{-T_2}\frac{C_3}{\sqrt{\beta^{(T_2-1)}}}$. Letting $C_6 = max(c_5, c_6)$ and $\delta = (\frac{1}{\sqrt{\gamma}}, \eta)$, we have

$$d(\boldsymbol{x}^{(t)}, \boldsymbol{x}^{(t-1)}) \leq C_6\delta^t \tag{44}$$

where $0 < \delta < 1$.

This completes the proof.

## VI. Proof of Lemma 6

**Lemma 6:** The $\{\boldsymbol{x}^{(t)}\}_{t=1}^{\infty}$ is a Cauchy sequence under the **S4** case.

***Proof***: Concretely, we firstly demonstrate that $\Delta^{(t)}$ can be bounded by a piecewise geometric sequence (PGS), which vanishes sufficiently fast. Then, we will show that $\{\boldsymbol{x}^{(t)}\}_{t=1}^{\infty}$ is a Cauchy sequence under this scenario.

**Definition 1**. (**Piecewise geometric sequence**) A positive sequence $\{z_t\}_{t\geq 1}$ is a piecewise geometric sequence (PGS) if there exists $0<\rho<1$ and indices $n_1<n_2<\dots$ such that

- for $j\geq 1$, the terms $z_{n_j},\dots,z_{n_{j+1}-1}$ are in geometric progression with the rate $\rho$, i.e., for $t=n_j,\dots,n_{j+1}-1$,

$$z_t = z_{n_j}\rho^{t-n_j}.$$

- the subsequence $\{z_{n_j}\}_{j\geq 1}$ that consists of peaks (the first term in each chunk of a PGS) is in geometric progression with rate $\rho$, i.e., for $j\geq 2$,

$$z_{n_j} = z_{n_1}\rho^{j-1}.$$

The PGS is divided into chunks. Each chunk is a geometric progression with rate $\rho$, and the peaks of each chunk is also a geometric progression with the same rate. Compared to a geometric sequence under the sample rate, the PGS is slower to decay to zero. In Lemma 7, we demonstrate the sequence of $\Delta^{(t)}$ can be bounded by a PGS under the **S4** case. Namely, there exists a PGS $\{z_t\}_{t\geq 1}$ such that $\Delta^{(t)}\leq z_t$ for all $t$. Moreover, in Lemma 8, we show that if $\{z_t\}_{t\geq 1}$ is a PGS, then $\sum_{t=1}^{\infty} z_t$ converges. By using Lemma 7 and 8, we can conclude that $\{\boldsymbol{x}^{(t)}\}_{t=1}^{\infty}$ is a Cauchy sequence. This result holds because for $j>t$,

$$\begin{aligned} d(\boldsymbol{x}^{(t)},\boldsymbol{x}^{(j)}) &\leq d(\boldsymbol{x}^{(t)},\boldsymbol{x}^{(t+1)})+\dots+d(\boldsymbol{x}^{(j-1)},\boldsymbol{x}^{(j)}) \\ &= \Delta^{(t+1)}+\dots+\Delta^{(j)}. \end{aligned} \tag{45}$$

Since $\Delta^{(t)}$ is bounded by the PGS $\{z_t\}_{t\geq 1}$ whose series itself converges, we can conclude that $\Delta^{(1)}+\Delta^{(2)}+\dots$ converges. The partial sums of $\Delta^{(t)}$ form a Cauchy sequence.

This completes the proof.

## VII. Proof of Lemma 7

**Lemma 7**: For the residuals $\Delta^{(t)}=d(\boldsymbol{x}^{(t)},\boldsymbol{x}^{(t-1)})$ under the **S4** case, there exists a PGS $\{z_t\}_{t\geq 1}$ such that $\Delta^{(t)}\leq z_t$ for all $t$.

***Proof***: The **S4** case claims that both Case 1 and Case 2 occur infinitely many $t$ often. In other words, both Case 1 and Case 2 are visited infinitely often. Therefore, Case 1 may hold for finitely many $t$, and then Case 2 holds for finitely many $t$. Alternatively, Case 2 holds for finitely many $t$, and then Case 1 holds for finitely many $t$. We call the first time that both Case 1 and Case 2 occurs as the first round, and call the second time as the second round, and so on. Let $n_1$ be the iteration when the Case 1 occurs for the first time, and $m_1>n_1$ be the iteration when the Case 2 occurs for the first time. Let $n_2>m_1$ be the iteration when the Case 1 occurs for the first time after $m_2$, and $m_2>n_2$ be the iteration when the Case 2 occurs for the first time after $n_2$. At the first round, Case 1 is visited for the iterations $n_1,n_1+1,\dots,m_1-1$ and Case 2 is visited for the iterations $m_1,m_1+1,\dots,n_2-1$. For each round $j\geq 1$, Case 1 occurs at iterations $t\in\{n_j,n_j+1,\dots,m_j-1\}$ and Case 2 occurs at iterations $t\in\{m_j,m_j+1,\dots,n_{j+1}-1\}$. To complete the proof, we should demonstrate that there exists a PGS $\{z_t\}_{t\geq 1}$ such that $\Delta^{(t)}\leq z_t$. Now, we discuss this bound for Case 1 and Case 2 respectively.

1) Case 1

For each round $j$, Case 1 holds at iterations $t\in\{n_j,n_j+1,\dots,m_j-1\}$, and based on the adaptive update role, we get

$$\beta^{(t)}=\gamma^{t-n_j+1}\beta^{(n_j-1)}, t\in\{n_j,n_j+1,\dots,m_j-1\}. \tag{46}$$

Using Lemma 4, we have

$$\begin{aligned} \Delta^{(t)} &\leq \frac{C_3}{\sqrt{\beta^{(t-1)}}}=\frac{C_3}{\sqrt{\gamma^{t-n_j}\beta^{(n_j-1)}}} \\ &, t\in\{n_j,n_j+1,\dots,m_j-1\}. \end{aligned} \tag{47}$$

Let $\alpha=\frac{1}{\sqrt{\gamma}}$ and $L_j=\frac{C_3}{\sqrt{\beta^{(n_j-1)}}}$, we can rewrite (47) as

$$\Delta^{(t)}\leq L_j\alpha^{t-n_j}, t\in\{n_j,n_j+1,\dots,m_j-1\}. \tag{48}$$

Now, we derive the relation of $L_j$ bounds at different rounds. Since Case 2 holds at iterations $t\in\{m_j,m_j+1,\dots,n_{j+1}-1\}$, we have

$$\begin{aligned} &\beta^{(n_{j+1}-1)}=\beta^{(n_{j+1}-2)}=\dots\beta^{(m_j)} \\ &=\beta^{(m_j-1)}=\gamma^{m_j-n_j}\beta^{(n_j-1)}. \end{aligned} \tag{49}$$

We know that $\frac{L_{j+1}}{L_j}=\frac{\sqrt{\beta^{(n_j-1)}}}{\sqrt{\beta^{(n_{j+1}-1)}}}$, $0<\alpha<1$, and $m_j-n_j\geq 1$, thus,

$$\begin{aligned} L_{j+1} &= \sqrt{\frac{\beta^{(n_j-1)}}{\beta^{(n_{j+1}-1)}}}L_j=\frac{1}{\sqrt{\gamma^{m_j-n_j}}}L_j \\ &=\alpha^{m_j-n_j}L_j\leq L_j\alpha. \end{aligned} \tag{50}$$

Letting $\rho=max\{\alpha,\eta\}$, and applying the inequality (50) recursively, we get

$$L_j\leq L_1\alpha^{j-1}\leq L_1\rho^{j-1}. \tag{51}$$

Hence, (48) can be rewritten as

$$\Delta^{(t)}\leq\overline{L_j}\alpha^{t-n_j}\leq\overline{L_j}\rho^{t-n_j}, t\in\{n_j,n_j+1,\dots,m_j-1\} \tag{52}$$

where $\overline{L_j}=L_1\rho^{j-1}$.

2) Case 2

For each round $j$, Case 2 holds at iterations $t\in\{m_j,m_j+1,\dots,n_{j+1}-1\}$. Using the adaptive update role, we have

$$\Delta^{(t)}\leq\eta^{(t-m_j)}\Delta^{(m_j)}, t\in\{m_j,m_j+1,\dots,n_{j+1}-1\}. \tag{53}$$

Since $\Delta^{(m_j)}\leq\eta\Delta^{(m_j-1)}$ holds at the $m_j$th iteration, we can substitute this inequality into (53). Hence, we get

$$\begin{aligned} \Delta^{(t)} &\leq \eta^{(t-m_j+1)}\Delta^{(m_j-1)} \\ &\leq \rho^{(t-m_j+1)}\overline{L_j}\rho^{m_j-n_j-1} \\ &= \overline{L_j}\rho^{t-n_j}, t\in\{m_j,m_j+1,\dots,n_{j+1}-1\}. \end{aligned} \tag{54}$$

The second inequality is based on the fact that $\eta < \rho$ and using (52) with $t = m_j - 1$.

From (52) and (54), we have

$$\Delta^{(t)} \leq \overline{L_j}\rho^{t-n_j}, t \in \{n_j, n_j+1, \ldots, n_{j+1}-1\}. \tag{55}$$

Therefore, for the residuals $\Delta^{(t)} = d(\boldsymbol{x}^{(t)}, \boldsymbol{x}^{(t-1)})$ under the **S4** case, there exists a sequence $\{z_t\}_{t\geq 1}$ such that $\Delta^{(t)} \leq z_t$ for all $t$. The sequence $\{z_t\}_{t\geq 1}$ is defined as

$$z_t = \overline{L_j}\rho^{t-n_j} = (L_1\rho^{j-1})\rho^{t-n_j} \tag{56}$$

where $j \geq 1$ is the number of round and $t \in \{n_j, n_j+1, \ldots, n_{j+1}-1\}$ is the iteration index. According to Definition 1, we can see that the sequence $\{z_t\}_{t\geq 1}$ in (56) is a PGS. In summary, $\Delta^{(t)}$ is bounded by a PGS with rate $\rho$.

This completes the proof.

## VIII. Proof of Lemma 8

**Lemma 8:** If $\{z_t\}_{t\geq 1}$ for $t \in \{n_j, n_j+1, \ldots, n_{j+1}-1\}$ is a PGS, then $\sum\limits_{t=1}^{\infty} z_t$ converges.

***Proof***: For each round $j$, we firstly show the bound of the sum of $z_t$ from $n_j$ to $n_{j+1}-1$. According to the definition of a PGS, we have

$$\begin{aligned} z_{n_j} + \cdots + z_{n_{j+1}-1} &= z_{n_j}(1 + \rho + \cdots + \rho^{n_{j+1}-n_j-1}) \\ &= z_{n_j} \times \frac{1-\rho^{n_{j+1}-n_j}}{1-\rho} \\ &< \frac{z_{n_j}}{1-\rho} = \frac{z_{n_1}\rho^{j-1}}{1-\rho}. \end{aligned} \tag{57}$$

The inequality holds due to $0 < \rho^{n_{j+1}-n_j} < 1$.

Fix an integer $T$ such that

$$\rho^{T-1} < \frac{\varepsilon(1-\rho)^2}{z_{n_1}} \tag{58}$$

where $\varepsilon > 0$. This inequality is possible since the values at two hands of (58) are positive. Now, we show $N = n_T + 1$ such that

$$z_n + \cdots + z_m < \varepsilon \tag{59}$$

for $m > n \geq N$. Letting $m > n \geq n_T$, and supposing $m \in \{n_j, n_j+1, \ldots, m_{j+1}-1\}$ for some $j > T$. Hence, $n_T \leq n < m \leq n_{j+1} - 1$. Based on (57), we have

$$\begin{aligned} \sum_{t=n}^{m} z_t &\leq \sum_{t=n_T}^{n_{j+1}-1} z_t \\ &= \sum_{t=n_T}^{n_{T+1}-1} z_t + \cdots + \sum_{t=n_j}^{n_{j+1}-1} z_t \\ &< \frac{z_{n_1}\rho^{T-1}}{1-\rho} + \cdots + \frac{z_{n_1}\rho^{j-1}}{1-\rho} \\ &= \frac{z_{n_1}}{1-\rho}\frac{\rho^{T-1}}{1-\rho}(1-\rho^{j-T+1}) \\ &< \frac{z_{n_1}\rho^{T-1}}{(1-\rho)^2}. \end{aligned} \tag{60}$$

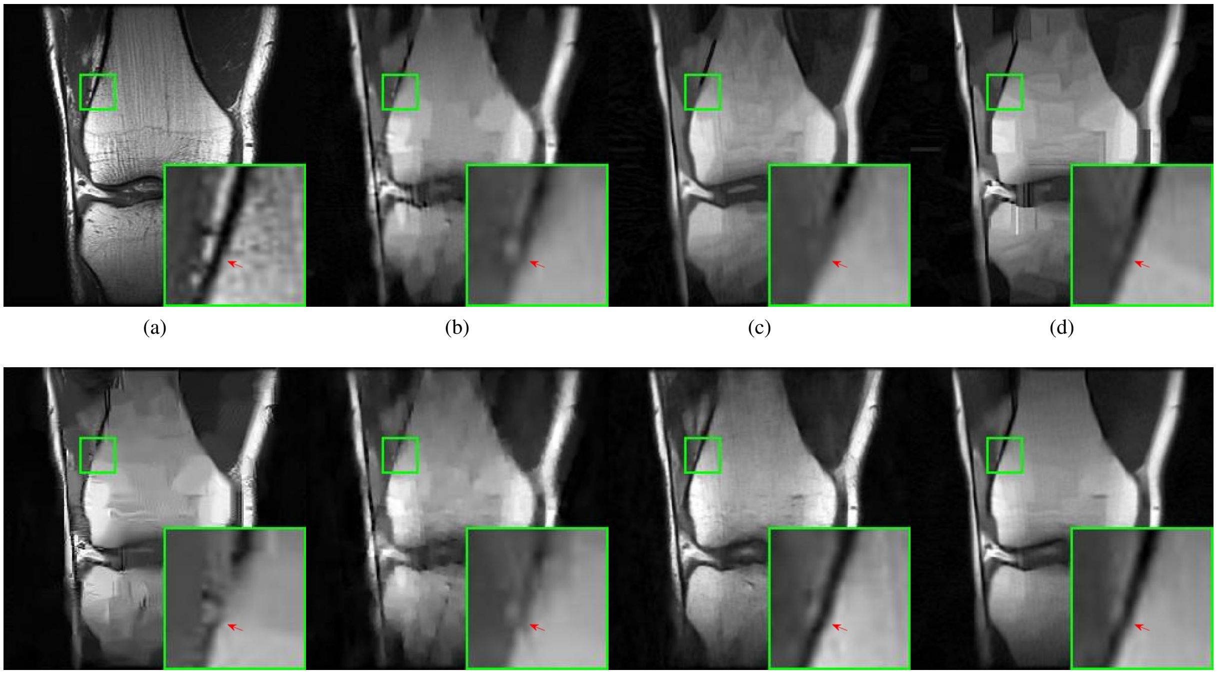
Substituting the inequality (58) into (60), we have

$$\sum_{t=n}^{m} z_t < \varepsilon. \tag{61}$$

Therefore, we find an index $N$ such that $N = n_T + 1$ satisfies the Cauchy criterion. Using the Cauchy criterion, we conclude that $\sum\limits_{t=1}^{\infty} z_t$ converges.

This completes the proof.

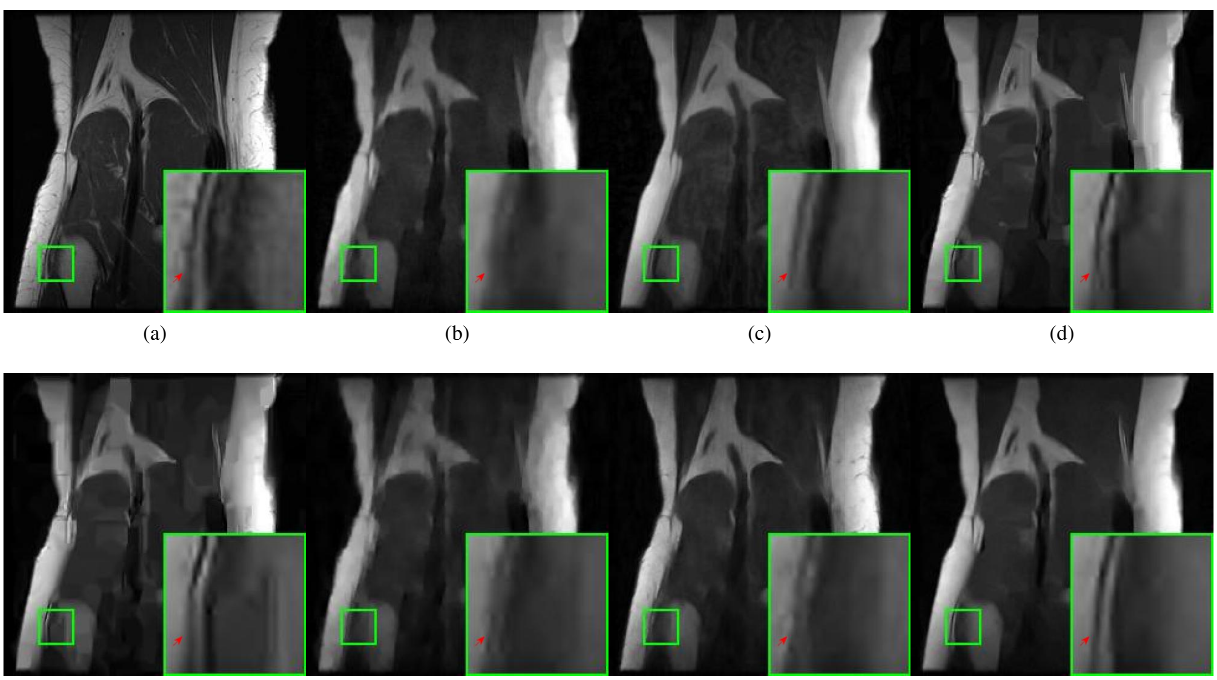
## IX. Reconstructions by the CSMRI algorithms



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

(e) (f) (g) (h)

Figure 1. The c-1176-1343 images recovered by the CSMRI algorithms at the case of 15 radial lines, i.e., sampling ratio 0.058. From left to right, and top to bottom: the original image, and the images recovered by (b) DLMRI (PSNR=23.24 dB, SSIM=0.6778); (c) NLR (PSNR=24.15 dB, SSIM=0.6689); (d) BM3D-MRI (PSNR=23.56 dB, SSIM=0.6587); (e) BM3D-AMP (PSNR=23.72 dB, SSIM=0.7107); (f) HQS-Net (PSNR=22.69 dB, SSIM=0.7824); (g) Deep-Cascade (PSNR=24.65 dB, SSIM=0.8360); (h) CSMRI-ReBMDual (PSNR=25.61 dB, SSIM=0.8497).



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

(e) (f) (g) (h)

Figure 2. The c-1167-1090 images recovered by the CSMRI algorithms at the case of 20 radial lines, i.e., sampling ratio 0.076. From left to right, and top to bottom: the original image, and the images recovered by (b) DLMRI (PSNR=27.14 dB, SSIM=0.7676); (c) NLR (PSNR=27.37 dB, SSIM=0.7451); (d) BM3D-MRI (PSNR=27.75 dB, SSIM=0.7839); (e) BM3D-AMP (PSNR=26.90 dB, SSIM=0.7558); (f) HQS-Net (PSNR=24.63 dB, SSIM=0.8401); (g) Deep-Cascade (PSNR=28.17 dB, SSIM=0.8960); (h) CSMRI-ReBMDual (PSNR=29.59 dB, SSIM=0.9079).

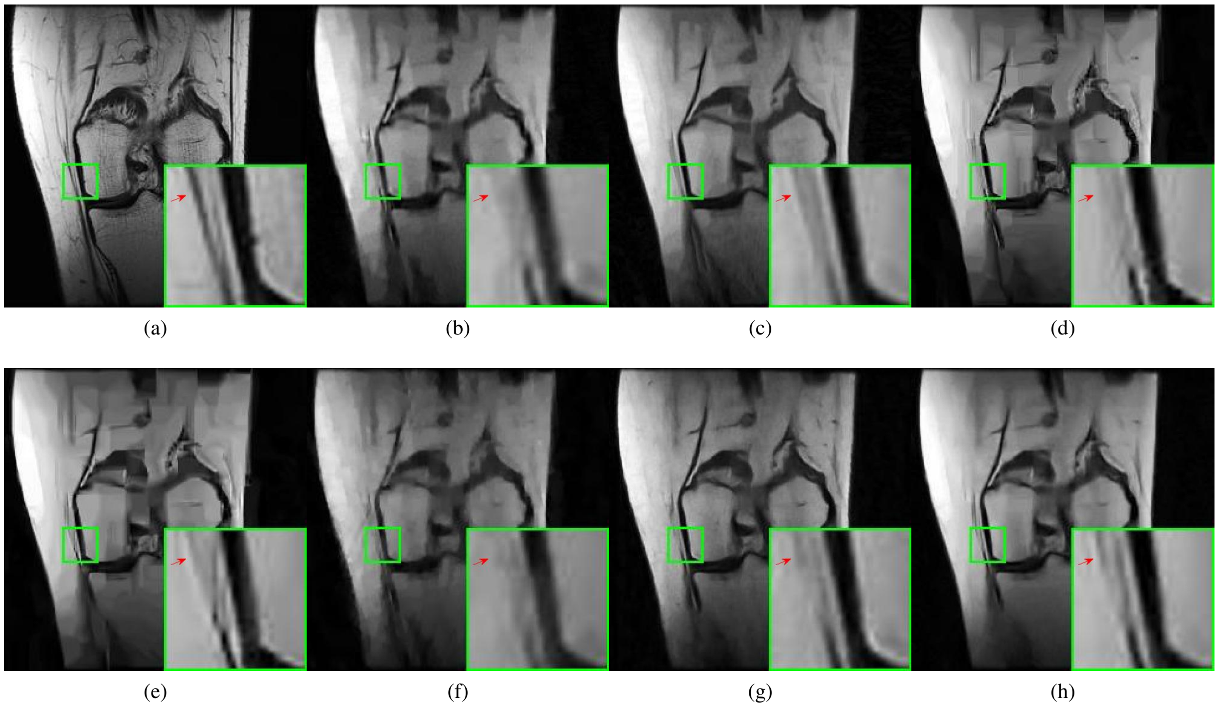



Figure 3. The c-1157-791 images recovered by the CSMRI algorithms at the case of 25 radial lines, i.e., sampling ratio 0.095. From left to right, and top to bottom: the original image, and the images recovered by (b) DLMRI (PSNR=24.76 dB, SSIM=0.7151); (c) NLR (PSNR=25.31 dB, SSIM=0.7140); (d) BM3D-MRI (PSNR=25.31 dB, SSIM=0.7279); (e) BM3D-AMP (PSNR=24.90 dB, SSIM=0.7192); (f) HQS-Net (PSNR=21.98 dB, SSIM=0.8035); (g) Deep-Cascade (PSNR=26.09 dB, SSIM=0.8768); (h) CSMRI-ReBMDual (PSNR=27.04 dB, SSIM=0.8760).

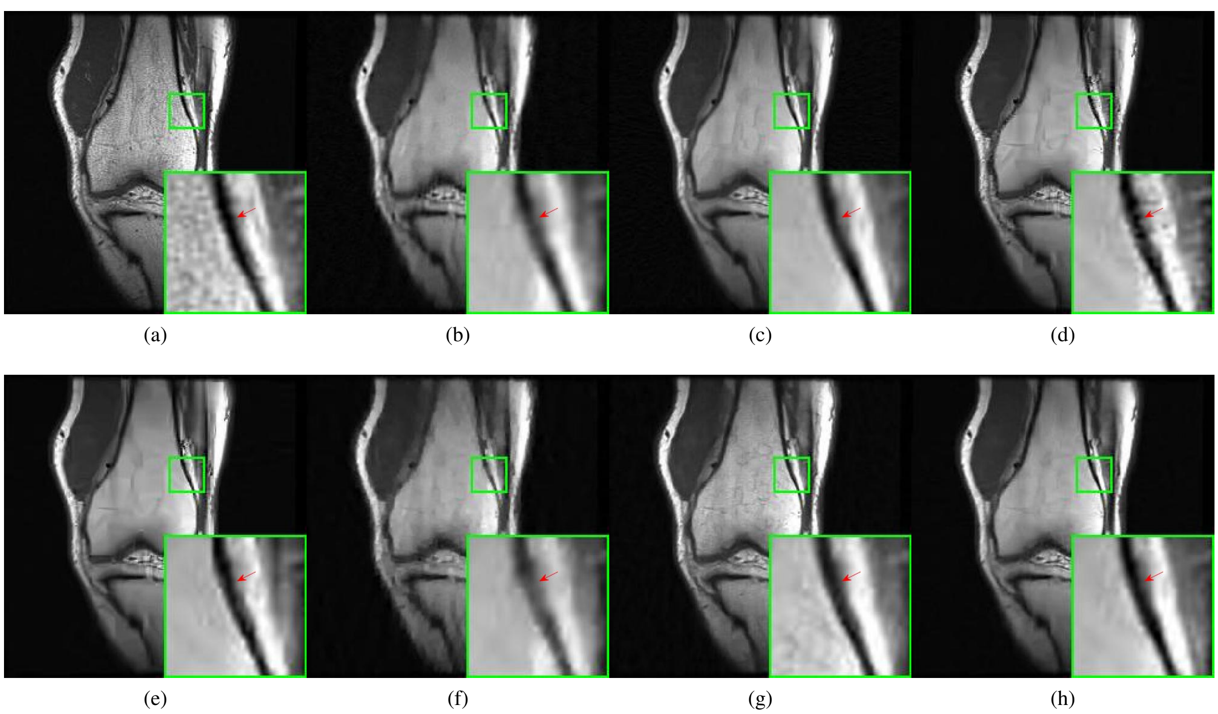



Figure 4. The c-1165-1037 images recovered by the CSMRI algorithms at the case of 30 radial lines, i.e., sampling ratio 0.113. From left to right, and top to bottom: the original image, and the images recovered by (b) DLMRI (PSNR=26.82 dB, SSIM=0.7957); (c) NLR (PSNR=28.12 dB, SSIM=0.8267); (d) BM3D-MRI (PSNR=27.71 dB, SSIM=0.8514); (e) BM3D-AMP (PSNR=27.37 dB, SSIM=0.8461); (f) HQS-Net (PSNR=26.18 dB, SSIM=0.8826); (g) Deep-Cascade (PSNR=27.36 dB, SSIM=0.9171); (h) CSMRI-ReBMDual (PSNR=28.53 dB, SSIM=0.9177).