Supplement material for "Deep sparse representation driven network for compressive imaging"

Baoshun Shi[a,b,*], Dan Li[a,b]

[a]School of Information Science and Engineering, Yanshan University, Qinhuangdao, 066004, Hebei Province, China
[b]Hebei Key Laboratory of Information Transmission and Signal Processing, Qinhuangdao, 066004, Hebei Province, China

1. Convergence analysis

Assumption *(Bounded assumption of the denoiser) Assume that the proposed DeSRNet satisfies the boundary denoiser condition, i.e., DeSRNet with a parameter σ is a function $\mathcal{D}(:, \sigma)$: $\mathbb{R}^n \to \mathbb{R}^n$ such that for any input $\boldsymbol{x} \in \mathbb{R}^n$,*

$$\|\mathcal{D}(\boldsymbol{x}, \sigma) - \boldsymbol{x}\|_2^2 \leqslant C\sigma^2 \tag{1}$$

for some universal constant C independent of σ.

1.1. The proof of ***Theorem 1***

Theorem 1 *(Convergence analysis of DUN-DeSRNet on SCI task) In SCI task, when the proposed prior network DeSRNet satisfies* **Assumption**, *as well as the value of σ satisfies the condition of sequentially diminishing as the unfolding stage t increases in the proposed DUN-DeSRNet algorithm, then both $\boldsymbol{x}^t$ and $\boldsymbol{v}^t$ will converge to fixed points, respectively.*

Proof: Following [1], we define $\boldsymbol{\theta}^t = (\boldsymbol{x}^t, \boldsymbol{v}^t)$, and let $\boldsymbol{\Theta}$ be the domain of $\boldsymbol{\theta}^t$ for all stages t. On the domain $\boldsymbol{\Theta}$, we define a distance function D: $\boldsymbol{\Theta} \times \boldsymbol{\Theta} \to \mathbb{R}$ such that

$$D(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}) = \frac{1}{\sqrt{n}}\left(\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2 + \left\|\boldsymbol{v}^t - \boldsymbol{v}^{t-1}\right\|_2\right). \tag{2}$$

In SCI task, given the measurement $\boldsymbol{y}$, and the sensing matrix $\boldsymbol{\Phi}$, where $\boldsymbol{\Phi}$ is sparse as well as a concatenation of diagonal matrices. Moreover, $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top$ is a diagonal matrix.

*Corresponding author
Email address: shibaoshun@ysu.edu.cn (Baoshun Shi)

From $\boldsymbol{x}^t = \boldsymbol{v}^{t-1} + \boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{v}^{t-1}\right)$, we have

$$\begin{aligned}
\boldsymbol{\Phi}\boldsymbol{x}^t &= \boldsymbol{\Phi}\boldsymbol{v}^{t-1} + \boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{v}^{t-1}\right)\\
&= \boldsymbol{\Phi}\boldsymbol{v}^{t-1} + \left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{v}^{t-1}\right)\\
&= \boldsymbol{y}.
\end{aligned} \tag{3}$$

Due to $\boldsymbol{x}^t - \boldsymbol{x}^{t-1} = \boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1} + \boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{v}^{t-1}\right)$, we have

$$\begin{aligned}
&\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2^2\\
=&\left\|\boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1} + \boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{v}^{t-1}\right)\right\|_2^2\\
=&\left\|\boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1} + \boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{\Phi}\boldsymbol{x}^{t-1} - \boldsymbol{\Phi}\boldsymbol{v}^{t-1}\right)\right\|_2^2\\
=&\left\|\boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1} + \boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Phi}\left(\boldsymbol{x}^{t-1} - \boldsymbol{v}^{t-1}\right)\right\|_2^2\\
=&\left\|\left(\boldsymbol{I} - \boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Phi}\right)\left(\boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1}\right)\right\|_2^2\\
=&\left\|\boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1}\right\|_2^2 - \left\|\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{T}\right)^{-\frac{1}{2}}\boldsymbol{\Phi}\left(\boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1}\right)\right\|_2^2.
\end{aligned} \tag{4}$$

Since the lower bound of $\left\|\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{T}\right)^{-\frac{1}{2}}\boldsymbol{\Phi}\left(\boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1}\right)\right\|_2^2$ is 0, we rewrite Eqn. (4) as

$$\begin{aligned}
&\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2^2\\
\leq&\left\|\boldsymbol{v}^{t-1} - \boldsymbol{x}^{t-1}\right\|_2^2\\
=&\left\|\mathcal{D}(\boldsymbol{x}^{t-1}, \sigma) - \boldsymbol{x}^{t-1}\right\|_2^2.
\end{aligned} \tag{5}$$

By utilizing the bounded property defined in Eqn. (1), we have

$$\begin{aligned}
&\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2^2\\
\leq&\left\|\mathcal{D}(\boldsymbol{x}^{t-1}, \sigma) - \boldsymbol{x}^{t-1}\right\|_2^2\\
\leq& nC\left[\sigma^{t-1}\right]^2.
\end{aligned} \tag{6}$$

Based on this, we have

$$
\begin{aligned}
&\left\|\boldsymbol{v}^{t}-\boldsymbol{v}^{t-1}\right\|_{2}^{2} \\
=&\left\|\mathcal{D}_{\sigma}\left(\boldsymbol{x}^{t}\right)-\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)\right\|_{2}^{2} \\
=&\|\mathcal{D}_{\sigma}\left(\boldsymbol{x}^{t}\right)-\boldsymbol{x}^{t}+\boldsymbol{x}^{t}-\boldsymbol{x}^{t-1}+\boldsymbol{x}^{t-1}-\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)\|_{2}^{2} \\
\leq&\left\|\mathcal{D}_{\sigma}\left(\boldsymbol{x}^{t}\right)-\boldsymbol{x}^{t}\right\|_{2}^{2}+\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{t-1}\right\|_{2}^{2}+\left\|\boldsymbol{x}^{t-1}-\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)\right\|_{2}^{2}.
\end{aligned} \tag{7}
$$

The inequality holds true due to the triangle inequality. By utilizing Eqns. (1) and (6), we have

$$
\begin{aligned}
&\left\|\boldsymbol{v}^{t}-\boldsymbol{v}^{t-1}\right\|_{2}^{2} \\
\leq&\left\|\mathcal{D}_{\sigma}\left(\boldsymbol{x}^{t}\right)-\boldsymbol{x}^{t}\right\|_{2}^{2}+\left\|\boldsymbol{x}^{t}-\boldsymbol{x}^{t-1}\right\|_{2}^{2}+\left\|\boldsymbol{x}^{t-1}-\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)\right\|_{2}^{2} \\
\leq& nC\left[\sigma^{t}\right]^{2}+nC\left[\sigma^{t-1}\right]^{2}+nC\left[\sigma^{t-1}\right]^{2} \\
\leq& 3nC\left[\sigma^{t-1}\right]^{2}.
\end{aligned} \tag{8}
$$

In this case, it is easily to see from Eqn. (6) and Eqn. (8) that

$$
\begin{aligned}
D\left(\boldsymbol{\theta}^{t},\boldsymbol{\theta}^{t-1}\right)\leq&\ \sqrt{C}\sigma^{t-1}+\sqrt{3C}\sigma^{t-1} \\
=&\ C'\sigma^{t-1}
\end{aligned} \tag{9}
$$

where $C'=\sqrt{C}+\sqrt{3C}$. The above convergence analysis also fits the noisy measurement in SCI task, we illustrate it in **Remark 1**.

Remark 1. *In the SCI task, we utilize the GAP solver and the proof is independent of the noise. This is because the update equation of $\boldsymbol{x}^t$ always satisfies $\boldsymbol{y}=\boldsymbol{\Phi}\boldsymbol{x}^t$. Consider the noisy measurement, i.e., $\boldsymbol{y}=\boldsymbol{\Phi}\boldsymbol{x}^*+\boldsymbol{n}$, where $\boldsymbol{n}$ denotes the measurement noise. Though hereby the measurement $\boldsymbol{y}$ is different from the noise free case, we still enforce $\boldsymbol{y}=\boldsymbol{\Phi}\boldsymbol{x}^t$ in each iteration.*

*1.2. The proof of **Theorem 2***

Lemma 1. *(The bounded gradient of the CSMRI data-fidelity function) Consider the sampling model $\boldsymbol{y}=\boldsymbol{F}_u\boldsymbol{x}+\boldsymbol{n}$ and $\|\boldsymbol{n}\|_2\leq\epsilon$, then the function $f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{F}_u\boldsymbol{x}\|_2^2$ has the bounded gradient, i.e., $\|\nabla f(\boldsymbol{x})\|_2\leq 2\sqrt{N}+\epsilon$.*

Proof: See Section 2.

Theorem 2 *(Convergence analysis of DUN-DeSRNet on CSMRI task) Based on Lemma 1, in CSMRI task, when the proposed prior network DeSRNet satisfies* **Assumption**, *as well as the value of σ satisfies the condition of sequentially diminishing as the unfolding stage t increases in the proposed DUN-DeSRNet algorithm, then both $\boldsymbol{x}^t$ and $\boldsymbol{v}^t$ will converge to fixed points, respectively.*

Proof: In CSMRI task, consider the proposed formulation

$$\boldsymbol{x}^t = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{F}_u\boldsymbol{x}\|_2^2 + \beta\|\boldsymbol{x} - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma)\|_2^2, \tag{10}$$

based on the first order optimality of Eqn. (10), we have

$$\nabla f(\boldsymbol{x}) + \beta(\boldsymbol{x} - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma)) = 0. \tag{11}$$

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

$$\boldsymbol{x}^t - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma) = -\frac{\nabla f(\boldsymbol{x}^t)}{\beta^{t-1}}. \tag{12}$$

At the $t-$th iteration, the value of β for updating image is β^{t-1}. Using the **Lemma 1**, we have

$$\|\boldsymbol{x}^t - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma)\|_2 \leq \frac{(2\sqrt{N} + \epsilon)}{\beta^{t-1}}. \tag{13}$$

Based on this bound and **Assumption**, we obtain

$$\begin{aligned}
&\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\|_2 \\
=&\|\boldsymbol{x}^t - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma) + \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma) - \boldsymbol{x}^{t-1}\|_2 \\
\leq&\|\boldsymbol{x}^{(t)} - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma)\|_2 + \|\mathcal{D}(\boldsymbol{x}^{t-1}, \sigma) - \boldsymbol{x}^{t-1}\|_2 \\
\leq&\frac{(2\sqrt{N} + \epsilon)}{\beta^{t-1}} + C\sigma^{t-1}.
\end{aligned} \tag{14}$$

Based on this, we have

$$\begin{aligned}
&\left\|\boldsymbol{v}^t - \boldsymbol{v}^{t-1}\right\|_2 \\
=&\left\|\mathcal{D}_\sigma\left(\boldsymbol{x}^t\right) - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma)\right\|_2 \\
=&\|\mathcal{D}_\sigma\left(\boldsymbol{x}^t\right) - \boldsymbol{x}^t + \boldsymbol{x}^t - \boldsymbol{x}^{t-1} + \boldsymbol{x}^{t-1} - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma)\|_2 \\
\leq&\left\|\mathcal{D}_\sigma\left(\boldsymbol{x}^t\right) - \boldsymbol{x}^t\right\|_2 + \left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2 + \left\|\boldsymbol{x}^{t-1} - \mathcal{D}(\boldsymbol{x}^{t-1}, \sigma)\right\|_2.
\end{aligned} \tag{15}$$

The inequality holds true due to the triangle inequality. By utilizing Eqns. (1) and (14), we have

$$
\begin{aligned}
&\left\|\boldsymbol{v}^t-\boldsymbol{v}^{t-1}\right\|_2 \\
\leq&\left\|\mathcal{D}_\sigma\left(\boldsymbol{x}^t\right)-\boldsymbol{x}^t\right\|_2+\left\|\boldsymbol{x}^t-\boldsymbol{x}^{t-1}\right\|_2+\left\|\boldsymbol{x}^{t-1}-\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)\right\|_2 \\
\leq&C\sigma^t+\frac{(2\sqrt{N}+\epsilon)}{\beta^{t-1}}+C\sigma^{t-1}+C\sigma^{t-1} \\
\leq&3C\sigma^{t-1}+\frac{(2\sqrt{N}+\epsilon)}{\beta^{t-1}}.
\end{aligned}
\tag{16}
$$

Using the distance function defined in Eqn. (2), it is easily to see from Eqn. (14) and Eqn. (16) that

$$
\begin{aligned}
D\left(\boldsymbol{\theta}^t,\boldsymbol{\theta}^{t-1}\right)&\leq\frac{(2\sqrt{N}+\epsilon)}{\beta^{t-1}}+C\sigma^{t-1}+3C\sigma^{t-1}+\frac{(2\sqrt{N}+\epsilon)}{\beta^{t-1}} \\
&=\frac{(4\sqrt{N}+2\epsilon)}{\beta^{t-1}}+4C\sigma^{t-1}.
\end{aligned}
\tag{17}
$$

Since $\beta^t=\eta\beta^{t-1}=\eta^{t-1}\beta^0$ and σ diminishes grandually, where $\eta>1$, when $t\to\infty$, $\beta^t\to\infty$ and $\sigma\to 0$, then $\|\boldsymbol{x}^t-\boldsymbol{x}^{t-1}\|_2\to 0$, $\|\boldsymbol{v}^t-\boldsymbol{v}^{t-1}\|_2\to 0$, and $D\left(\boldsymbol{\theta}^t,\boldsymbol{\theta}^{t-1}\right)\to 0$. In particular, we present the proof when the measurement $\boldsymbol{y}$ does not contain noise in **Remark 2**.

Remark 2. *In the CSMRI task, noise-free as a special case still satisfies the above analysis.*

Proof: Consider the noise-free case, i.e., $\boldsymbol{y}=\boldsymbol{F}_u\boldsymbol{x}^t$, and given the partial Fourier transform $\boldsymbol{F}_u$. Since the Fourier transform $\boldsymbol{F}$ is an unitary matrix admitting $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F}=\boldsymbol{I}$, calculating $\left\|\boldsymbol{x}^t-\boldsymbol{x}^{t-1}\right\|_2$ is equivalent to calculating $\left\|\boldsymbol{F}\boldsymbol{x}^t-\boldsymbol{F}\boldsymbol{x}^{t-1}\right\|_2$ in the frequency domain. Following $\boldsymbol{v}^t=\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)$ and $\boldsymbol{x}^t=\boldsymbol{F}^{\mathrm{H}}\left[\frac{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{y}+\beta\boldsymbol{F}\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}}+\beta I}\right]$, we have

$$
\begin{aligned}
&\left\|\boldsymbol{x}^t-\boldsymbol{x}^{t-1}\right\|_2^2 \\
=&\left\|\frac{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{y}+\beta\boldsymbol{F}\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}}+\beta I}-\boldsymbol{F}\boldsymbol{x}^{t-1}\right\|_2^2 \\
=&\left\|\frac{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{y}+\beta\boldsymbol{F}\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)-\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F}\boldsymbol{x}^{t-1}-\beta\boldsymbol{F}\boldsymbol{x}^{t-1}}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}}+\beta I}\right\|_2^2 \\
=&\left\|\frac{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{y}-\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{x}^{t-1}+\beta\boldsymbol{F}\mathcal{D}(\boldsymbol{x}^{t-1},\sigma)-\beta\boldsymbol{F}\boldsymbol{x}^{t-1}}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}}+\beta I}\right\|_2^2.
\end{aligned}
\tag{18}
$$

Since under the noise-free case, $\boldsymbol{y}$ can be denoted as $\boldsymbol{y} = \boldsymbol{F}_u\boldsymbol{x}^t$. Eqn. (18) can be rewritten as

$$
\begin{aligned}
&\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2^2 \\
=&\left\|\frac{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{y} - \boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{x}^{t-1} + \beta\boldsymbol{F}\mathcal{D}(\boldsymbol{x}^{t-1},\sigma) - \beta\boldsymbol{F}\boldsymbol{x}^{t-1}}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}} + \beta I}\right\|_2^2 \\
=&\left\|\frac{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{x}^t - \boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{x}^{t-1} + \beta\boldsymbol{F}\mathcal{D}(\boldsymbol{x}^{t-1},\sigma) - \beta\boldsymbol{F}\boldsymbol{x}^{t-1}}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}} + \beta I}\right\|_2^2 \\
\le&\left\|\frac{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u(\boldsymbol{x}^t - \boldsymbol{x}^{t-1})}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}} + \beta I}\right\|_2^2 + \left\|\frac{\beta\boldsymbol{F}(\mathcal{D}(\boldsymbol{x}^{t-1},\sigma) - \boldsymbol{x}^{t-1})}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}} + \beta I}\right\|_2^2.
\end{aligned} \tag{19}
$$

Since $\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}}$ is a diagonal matrix whose elements are ones or zeros, then $\left\|\frac{\boldsymbol{I}}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}} + \beta I}\right\|_F^2$ can be written as $\frac{1}{1+\beta} \in (0, 1)$, and by utilizing the bounded property defined in Eqn. (1), Eqn. (19) can be written as

$$
\begin{aligned}
&\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2^2 \\
\le&\left\|\frac{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u(\boldsymbol{x}^t - \boldsymbol{x}^{t-1})}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}} + \beta I}\right\|_2^2 + \left\|\frac{\beta\boldsymbol{F}(\mathcal{D}(\boldsymbol{x}^{t-1},\sigma) - \boldsymbol{x}^{t-1})}{\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}} + \beta I}\right\|_2^2 \\
\le&\frac{1}{1+\beta}\left\|\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\right\|_2^2\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2^2 + \frac{\beta}{1+\beta}\left\|\mathcal{D}(\boldsymbol{x}^{t-1},\sigma) - \boldsymbol{x}^{t-1}\right\|_2^2 \\
\le&\frac{1}{1+\beta}\left\|\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\right\|_2^2\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2^2 + \frac{\beta}{1+\beta}C_1(\sigma^{t-1})^2.
\end{aligned} \tag{20}
$$

Since $\boldsymbol{F}$ is an unitary matrix and $\boldsymbol{F}\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{F}^{\mathrm{H}}$ is a diagonal matrix whose elements are ones or zeros, we have $\left\|\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\right\|_2^2 = 1$. Following this, we have

$$
\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2^2 \le C_1(\sigma^{t-1})^2. \tag{21}
$$

Based on this, we have

$$
\begin{aligned}
&\left\|\boldsymbol{v}^t - \boldsymbol{v}^{t-1}\right\|_2^2 \\
=&\left\|\mathcal{D}(\boldsymbol{x}^{t-1},\sigma) - \mathcal{D}_\sigma\left(\boldsymbol{x}^{t-2}\right)\right\|_2^2 \\
=&\|\mathcal{D}(\boldsymbol{x}^{t-1},\sigma) - \boldsymbol{x}^{t-1} + \boldsymbol{x}^{t-1} - \boldsymbol{x}^{t-2} + \boldsymbol{x}^{t-2} - \mathcal{D}_\sigma\left(\boldsymbol{x}^{t-2}\right)\|_2^2 \\
\le&\left\|\mathcal{D}(\boldsymbol{x}^{t-1},\sigma) - \boldsymbol{x}^{t-1}\right\|_2^2 + \left\|\boldsymbol{x}^{t-1} - \boldsymbol{x}^{t-2}\right\|_2^2 + \left\|\boldsymbol{x}^{t-2} - \mathcal{D}_\sigma\left(\boldsymbol{x}^{t-2}\right)\right\|_2^2 \\
\le&C_2\left[\sigma^{t-1}\right]^2 + C_3\left[\sigma^{t-2}\right]^2 + C_4\left[\sigma^{t-2}\right]^2.
\end{aligned} \tag{22}
$$

Since the value of σ satisfies the condition of sequentially diminishing as the unfolding stage t increases, i.e., $\sigma^{t-1} \le$

σ^{t-2}, we have

$$
\begin{aligned}
&\left\|\boldsymbol{v}^{t}-\boldsymbol{v}^{t-1}\right\|_{2}^{2} \\
\leq & C_{2}\left[\sigma^{t-1}\right]^{2}+C_{3}\left[\sigma^{t-2}\right]^{2}+C_{4}\left[\sigma^{t-2}\right]^{2} \\
\leq & (C_{2}+C_{3}+C_{4})\sigma^{t-2} \\
\leq & M_{1}\sigma^{t-2}
\end{aligned} \tag{23}
$$

where $M_1 = C_2 + C_3 + C_4$. In this case, it is easily to see from Eqn. (21) and Eqn. (23) that

$$
\begin{aligned}
D\left(\boldsymbol{\theta}^{t}, \boldsymbol{\theta}^{t-1}\right) &\leq \sqrt{C_{1}}\sigma^{t-1}+\sqrt{M_{1}}\sigma^{t-2} \\
&\leq \sqrt{C_{1}}\sigma^{t-2}+\sqrt{M_{1}}\sigma^{t-2} \\
&= M\sigma^{t-2}
\end{aligned} \tag{24}
$$

where $M = \sqrt{C_1} + \sqrt{M_1}$.

This completes the proof.

2. The proof for Lemma 1

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

$$
\nabla f(\boldsymbol{x}) = \boldsymbol{F}_u^{\mathrm{H}}(\boldsymbol{F}_u\boldsymbol{x} - \boldsymbol{y}), \tag{25}
$$

and thus we have

$$
\|\nabla f(\boldsymbol{x})\|_2 \leq \|\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{F}_u\boldsymbol{x}\|_2 + \|\boldsymbol{F}_u^{\mathrm{H}}\boldsymbol{y}\|_2 \leq \|\boldsymbol{F}_u^{\mathrm{H}}\|_2(\|\boldsymbol{F}_u\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2). \tag{26}
$$

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

$$
\|\boldsymbol{F}_u\boldsymbol{x}\|_2 \leq \sqrt{N}. \tag{27}
$$

In Eqn. (26), $\|\boldsymbol{y}\|_2 = \|\boldsymbol{F}_u\boldsymbol{x}_{ori} + \boldsymbol{n}\|_2$ and $\|\boldsymbol{n}\|_2 \leq \epsilon$; here, $\boldsymbol{x}_{ori}$ is the original image whose pixel values are normalized into $[0, 1]$. Therefore,

$$
\|\boldsymbol{y}\|_2 = \|\boldsymbol{F}_u\boldsymbol{x}_{ori} + \boldsymbol{n}\|_2 \leq \|\boldsymbol{F}_u\boldsymbol{x}_{ori}\|_2 + \|\boldsymbol{n}\|_2 \leq \sqrt{N} + \epsilon. \tag{28}
$$

Substituting Eqn. (27) and Eqn. (28) into Eqn. (26), we get

$$\|\nabla f(\boldsymbol{x})\|_2 \leq \|\boldsymbol{F}_u^{\mathrm{H}}\|_2(2\sqrt{N}+\epsilon) = \sqrt{\lambda_{max}(\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}})}(2\sqrt{N}+\epsilon) \tag{29}$$

where $\lambda_{max}(\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}})$ represents the maximum eigenvalue of $\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}}$. 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,

$$\boldsymbol{F}_u\boldsymbol{F}_u^{\mathrm{H}} = (\boldsymbol{P}\boldsymbol{F})(\boldsymbol{P}\boldsymbol{F})^{\mathrm{H}} = \boldsymbol{I} \tag{30}$$

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

$$\|\nabla f(\boldsymbol{x})\|_2 \leq 2\sqrt{N}+\epsilon. \tag{31}$$

This completes the proof of **Lemma 1**.

References

[1] S. H. Chan, X. Wang, O. A. Elgendy, Plug-and-play ADMM for image restoration: Fixed-point convergence and applications, IEEE Transactions on Computational Imaging. 3 (1) (2017) 84–98. doi:10.1109/TCI.2016.2629286.