

I. PROOF OF LEMMA 1

Lemma 1 If \mathbf{W}_S has an upper bound, then we have

$$\|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \leq c\sigma^2 \quad (1)$$

for some universal constant c independent of σ .

Proof: From Eqn. (1), we have

$$\begin{aligned} & \|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \\ & \leq \|\mathbf{W}_S\|_2^2 \|\mathbf{W}_A \mathbf{x} - \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2. \end{aligned} \quad (2)$$

A convolutional layer with RealSN constraint [1] is utilized to generate the synthesis sparsifying transform, which makes $\|\mathbf{W}_S\|_2^2$ have an upper bound. Defined the upper bound is α , and the upper bound of Eqn. (2) can be further determined by

$$\begin{aligned} & \|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \\ & \leq \alpha \|\mathbf{W}_A \mathbf{x} - \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2. \end{aligned} \quad (3)$$

Let $(\mathbf{W}_A \mathbf{x})_i$ represent the i -th element of $\mathbf{W}_A \mathbf{x}$ and β_i denote the i -th element of \mathbf{b} . The soft thresholding operator $\text{Soft}[(\mathbf{W}_A \mathbf{x})_i, \beta_i]$ is defined as

$$\text{Soft}[(\mathbf{W}_A \mathbf{x})_i, \beta_i] = \begin{cases} (\mathbf{W}_A \mathbf{x})_i + \beta_i, & (\mathbf{W}_A \mathbf{x})_i < -\beta_i \\ 0, & |(\mathbf{W}_A \mathbf{x})_i| \leq \beta_i \\ (\mathbf{W}_A \mathbf{x})_i - \beta_i, & (\mathbf{W}_A \mathbf{x})_i > \beta_i, \end{cases} \quad (4)$$

then one of the following situations will happen.

CASE 1: Any element of $\mathbf{W}_A \mathbf{x}$ satisfies $(\mathbf{W}_A \mathbf{x})_i < -\beta_i$, then $\text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b}) = \mathbf{W}_A \mathbf{x} + \mathbf{b}$. Therefore, we have

$$\begin{aligned} & \|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \\ & \leq \alpha \|\mathbf{W}_A \mathbf{x} - \mathbf{W}_A \mathbf{x} - \mathbf{b}\|_2^2 \\ & = \alpha \|\mathbf{b}\|_2^2 \end{aligned} \quad (5)$$

CASE 2: Any element of $\mathbf{W}_A \mathbf{x}$ satisfies $|(\mathbf{W}_A \mathbf{x})_i| \leq \beta_i$, then $\text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b}) = 0$. Therefore, we have

$$\begin{aligned} & \|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \\ & \leq \alpha \|\mathbf{W}_A \mathbf{x}\|_2^2. \end{aligned} \quad (6)$$

Since $|(\mathbf{W}_A \mathbf{x})_i| \leq \beta_i$, then $\|\mathbf{W}_A \mathbf{x}\|_2^2 \leq \|\mathbf{b}\|_2^2$, it follows that

$$\begin{aligned} & \|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \\ & \leq \alpha \|\mathbf{b}\|_2^2. \end{aligned} \quad (7)$$

CASE 3: Any element of $\mathbf{W}_A \mathbf{x}$ satisfies $(\mathbf{W}_A \mathbf{x})_i > \beta_i$, then $\text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b}) = \mathbf{W}_A \mathbf{x} - \mathbf{b}$. Therefore, we have

$$\begin{aligned} & \|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \\ & \leq \alpha \|\mathbf{W}_A \mathbf{x} - \mathbf{W}_A \mathbf{x} + \mathbf{b}\|_2^2 \\ & = \alpha \|\mathbf{b}\|_2^2. \end{aligned} \quad (8)$$

CASE 4: CASE 4 is a union of the CASE 1, CASE 2 and CASE 3. Therefore, as long as we find the upper bound of $\|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2$ under the CASE 1-3, the upper bound of that under the CASE 4 will be determined. Hence, when CASE 4 happens, the upper bound of $\|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2$ is the maximum upper

bound of the first three situations. Based on Eqn. (5), Eqn. (7) and Eqn. (8), we have

$$\|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \leq \alpha \|\mathbf{b}\|_2^2 \quad (9)$$

Recall the definition of threshold vectors $\mathbf{b} = \mathbf{c} \otimes \mathbf{s}$, and each element of proportional constant vectors \mathbf{c} has a limited range $c_i \in [c_{\min}, c_{\max}]$. Let $\beta_{\max} = c_{\max} \cdot \sigma$ denote the maximum element of \mathbf{b} . Thus,

$$\|\mathbf{W}_S \mathbf{W}_A \mathbf{x} - \mathbf{W}_S \text{Soft}(\mathbf{W}_A \mathbf{x}, \mathbf{b})\|_2^2 \leq \alpha c_{\max}^2 \sigma^2 \quad (10)$$

II. PROOF OF LEMMA 2

Lemma 2 (The bounded gradient of data fidelity function) Consider the sampling model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ and $\mathbf{y} = \mathbf{F}_u \mathbf{x} + \mathbf{n}$, where $\|\mathbf{n}\|_2 \leq \epsilon$, then the functions $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$ and $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{F}_u \mathbf{x}\|_2^2$ have the bounded gradient, respectively.

Proof:

(In the SCI task) The gradient of $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$ in the SCI formulation is

$$\nabla f(\mathbf{x}) = \mathbf{H}^\top (\mathbf{H}\mathbf{x} - \mathbf{y}) \quad (11)$$

thus, we have

$$\begin{aligned} \|\nabla f(\mathbf{x})\|_2 & \leq \|\mathbf{H}^\top \mathbf{H}\mathbf{x}\|_2 + \|\mathbf{H}^\top \mathbf{y}\|_2 \\ & \leq \|\mathbf{H}^\top\|_2 (\|\mathbf{H}\mathbf{x}\|_2 + \|\mathbf{y}\|_2). \end{aligned} \quad (12)$$

Since the values in the sensing matrix \mathbf{H} and the pixel values in the image \mathbf{x} are normalized into $[0, 1]$, we obtain

$$\|\mathbf{H}\mathbf{x}\|_2 \leq \|\mathbf{H}\|_2 \|\mathbf{x}\|_2 \leq \sqrt{\lambda_1 N}. \quad (13)$$

where λ_1 denotes the maximum eigenvalue of $\mathbf{H}^\top \mathbf{H}$, and N represents the pixel values of the image. In Eqn. (13), $\|\mathbf{y}\|_2 = \|\mathbf{H}\mathbf{x} + \mathbf{n}\|_2$ and $\|\mathbf{n}\|_2 \leq \epsilon$, here, \mathbf{x} is the image whose pixel values are normalized into $[0, 1]$. Therefore,

$$\|\mathbf{y}\|_2 = \|\mathbf{H}\mathbf{x} + \mathbf{n}\|_2 \leq \|\mathbf{H}\mathbf{x}\|_2 + \|\mathbf{n}\|_2 \leq \sqrt{\lambda_1 N} + \epsilon. \quad (14)$$

Let λ_2 denote the maximum eigenvalue of $\mathbf{H}\mathbf{H}^\top$, then we have

$$\|\mathbf{H}^\top\|_2 \leq \sqrt{\lambda_2}. \quad (15)$$

Substituting Eqn. (13) and Eqn. (14) Eqn. (15) into Eqn. (12), we get

$$\begin{aligned} \|\nabla f(\mathbf{x})\|_2 & \leq \|\mathbf{H}^\top\|_2 (\|\mathbf{H}\mathbf{x}\|_2 + \|\mathbf{y}\|_2) \\ & \leq \sqrt{\lambda_2} (2\sqrt{\lambda_1 N} + \epsilon). \end{aligned} \quad (16)$$

This completes the proof of **Lemma 2** in the SCI task.

(In the CSMRI task) See [2] for detailed proof.

III. PROOF OF THEOREM 2

Theorem 2 (Convergence of the proposed algorithms) Based on **Lemma 2** and **Theorem 1**, as well as the value of σ satisfies the condition of sequentially diminishing as the unfolding stage t increases in the proposed DUN-BSTNet algorithm, then both \mathbf{x}^t and \mathbf{v}^t will converge to a fixed point, respectively.

Proof: Following [3], we define $\boldsymbol{\theta}^t = (\mathbf{x}^t, \mathbf{v}^t)$, and let Θ be the domain of $\boldsymbol{\theta}^t$ for all stages t . On the domain Θ , we define a distance function $D: \Theta \times \Theta \rightarrow \mathbb{R}$ such that

$$D(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}) = \|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 + \|\mathbf{v}^t - \mathbf{v}^{t-1}\|_2. \quad (17)$$

(In the SCI task) Consider the proposed formulation

$$\mathbf{x}^t = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \gamma \|\mathbf{x} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2^2, \quad (18)$$

based on the first order optimality of Eqn. (18), we have

$$\nabla f(\mathbf{x}) + \gamma (\mathbf{x} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)) = 0. \quad (19)$$

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

$$\mathbf{x}^t - \mathcal{D}(\mathbf{x}^{t-1}, \sigma) = -\frac{\nabla f(\mathbf{x}^t)}{\gamma^{t-1}}. \quad (20)$$

At the t -th iteration, the value of γ for updating image is γ^{t-1} . Using the **Lemma 2**, we have

$$\|\mathbf{x}^t - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 \leq \frac{\sqrt{\lambda_2} (2\sqrt{\lambda_1 N} + \epsilon)}{\gamma^{t-1}}. \quad (21)$$

Based on this bound and the bounded property of BSTNet, we obtain

$$\begin{aligned} & \|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 \\ &= \|\mathbf{x}^t - \mathcal{D}(\mathbf{x}^{t-1}, \sigma) + \mathcal{D}(\mathbf{x}^{t-1}, \sigma) - \mathbf{x}^{t-1}\|_2 \\ &\leq \|\mathbf{x}^t - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 + \|\mathcal{D}(\mathbf{x}^{t-1}, \sigma) - \mathbf{x}^{t-1}\|_2 \\ &\leq \frac{\sqrt{\lambda_2} (2\sqrt{\lambda_1 N} + \epsilon)}{\gamma^{t-1}} + C\sigma^{t-1}. \end{aligned} \quad (22)$$

Based on this, we have

$$\begin{aligned} & \|\mathbf{v}^t - \mathbf{v}^{t-1}\|_2 \\ &= \|\mathcal{D}_\sigma(\mathbf{x}^t) - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 \\ &= \|\mathcal{D}_\sigma(\mathbf{x}^t) - \mathbf{x}^t \\ &\quad + \mathbf{x}^t - \mathbf{x}^{t-1} + \mathbf{x}^{t-1} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 \\ &\leq \|\mathcal{D}_\sigma(\mathbf{x}^t) - \mathbf{x}^t\|_2 \\ &\quad + \|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 + \|\mathbf{x}^{t-1} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2. \end{aligned} \quad (23)$$

The inequality holds true due to the triangle inequality. Based on the bounded property of BSTNet and (22), we have

$$\begin{aligned} & \|\mathbf{v}^t - \mathbf{v}^{t-1}\|_2 \\ &\leq \|\mathcal{D}_\sigma(\mathbf{x}^t) - \mathbf{x}^t\|_2 + \|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 \\ &\quad + \|\mathbf{x}^{t-1} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 \\ &\leq C\sigma^t + \frac{\sqrt{\lambda_2} (2\sqrt{\lambda_1 N} + \epsilon)}{\gamma^{t-1}} + C\sigma^{t-1} + C\sigma^{t-1} \\ &\leq 3C\sigma^{t-1} + \frac{\sqrt{\lambda_2} (2\sqrt{\lambda_1 N} + \epsilon)}{\gamma^{t-1}}. \end{aligned} \quad (24)$$

Using the distance function defined in Eqn. (17), it is easily to see from Eqn. (22) and Eqn. (24) that

$$\begin{aligned} D(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}) &\leq \frac{\sqrt{\lambda_2} (2\sqrt{\lambda_1 N} + \epsilon)}{\gamma^{t-1}} + C\sigma^{t-1} \\ &\quad + 3C\sigma^{t-1} + \frac{\sqrt{\lambda_2} (2\sqrt{\lambda_1 N} + \epsilon)}{\gamma^{t-1}} \\ &= \frac{\sqrt{\lambda_2} (4\sqrt{\lambda_1 N} + 2\epsilon)}{\gamma^{t-1}} + 4C\sigma^{t-1}, \end{aligned} \quad (25)$$

Since $\frac{1}{1+\gamma^t} < 1$ and σ diminishes grandually, when $t \rightarrow \infty$, $\gamma^t \rightarrow \infty$ and $\sigma \rightarrow 0$, then $\|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 \rightarrow 0$, $\|\mathbf{v}^t - \mathbf{v}^{t-1}\|_2 \rightarrow 0$, and $D(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}) \rightarrow 0$.

(In the CSMRI task) Consider the proposed formulation

$$\mathbf{x}^t = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{F}_u \mathbf{x}\|_2^2 + \gamma \|\mathbf{x} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2^2, \quad (26)$$

based on the first order optimality of Eqn. (26), we have

$$\nabla f(\mathbf{x}) + \gamma (\mathbf{x} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)) = 0. \quad (27)$$

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

$$\mathbf{x}^t - \mathcal{D}(\mathbf{x}^{t-1}, \sigma) = -\frac{\nabla f(\mathbf{x}^t)}{\gamma^{t-1}}. \quad (28)$$

At the t -th iteration, the value of γ for updating image is γ^{t-1} . Using the **Lemma 2**, we have

$$\|\mathbf{x}^t - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 \leq \frac{(2\sqrt{N} + \epsilon)}{\gamma^{t-1}}. \quad (29)$$

Based on this bound and the bounded property of BSTNet, we obtain

$$\begin{aligned} & \|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 \\ &= \|\mathbf{x}^t - \mathcal{D}(\mathbf{x}^{t-1}, \sigma) + \mathcal{D}(\mathbf{x}^{t-1}, \sigma) - \mathbf{x}^{t-1}\|_2 \\ &\leq \|\mathbf{x}^t - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 + \|\mathcal{D}(\mathbf{x}^{t-1}, \sigma) - \mathbf{x}^{t-1}\|_2 \\ &\leq \frac{(2\sqrt{N} + \epsilon)}{\gamma^{t-1}} + C\sigma^{t-1}. \end{aligned} \quad (30)$$

Based on this, we have

$$\begin{aligned} & \|\mathbf{v}^t - \mathbf{v}^{t-1}\|_2 \\ &= \|\mathcal{D}_\sigma(\mathbf{x}^t) - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 \\ &= \|\mathcal{D}_\sigma(\mathbf{x}^t) - \mathbf{x}^t \\ &\quad + \mathbf{x}^t - \mathbf{x}^{t-1} + \mathbf{x}^{t-1} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 \\ &\leq \|\mathcal{D}_\sigma(\mathbf{x}^t) - \mathbf{x}^t\|_2 \\ &\quad + \|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 + \|\mathbf{x}^{t-1} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2. \end{aligned} \quad (31)$$

The inequality holds true due to the triangle inequality. By utilizing the bounded property of BSTNet and (30), we have

$$\begin{aligned} & \|\mathbf{v}^t - \mathbf{v}^{t-1}\|_2 \\ &\leq \|\mathcal{D}_\sigma(\mathbf{x}^t) - \mathbf{x}^t\|_2 + \|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 \\ &\quad + \|\mathbf{x}^{t-1} - \mathcal{D}(\mathbf{x}^{t-1}, \sigma)\|_2 \\ &\leq C\sigma^t + \frac{(2\sqrt{N} + \epsilon)}{\gamma^{t-1}} + C\sigma^{t-1} + C\sigma^{t-1} \\ &\leq 3C\sigma^{t-1} + \frac{(2\sqrt{N} + \epsilon)}{\gamma^{t-1}}. \end{aligned} \quad (32)$$

Using the distance function defined in Eqn. (17), it is easily to see from Eqn. (30) and Eqn. (32) that

$$\begin{aligned} D(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}) &\leq \frac{(2\sqrt{N} + \epsilon)}{\gamma^{t-1}} + C\sigma^{t-1} \\ &\quad + 3C\sigma^{t-1} + \frac{(2\sqrt{N} + \epsilon)}{\gamma^{t-1}} \\ &= \frac{(4\sqrt{N} + 2\epsilon)}{\gamma^{t-1}} + 4C\sigma^{t-1}, \end{aligned} \quad (33)$$

Since $\gamma^t = \eta\gamma^{t-1} = \eta^{t-1}\gamma^0$ and σ diminishes grandually, where $\eta > 1$, when $t \rightarrow \infty$, $\gamma^t \rightarrow \infty$ and $\sigma \rightarrow 0$, then $\|\mathbf{x}^t - \mathbf{x}^{t-1}\|_2 \rightarrow 0$, $\|\mathbf{v}^t - \mathbf{v}^{t-1}\|_2 \rightarrow 0$, and $D(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}) \rightarrow 0$.

This completes the proof of **Theorem 2** in different tasks.

REFERENCES

- [1] E. Ryu, J. L. Liu, S. C. Wang, X. H. Chen, Z. Y. Wang, W. T. Yin, Plug-and-play methods provably converge with properly trained denoisers, in: Proceedings of the International Conference on Machine Learning (ICML), 2019, pp. 5546–5557.
- [2] B. Shi, K. Liu, Regularization by multiple dual frames for compressed sensing magnetic resonance imaging with convergence analysis., IEEE/CAA Journal of Automatica Sinica 10 (11) (2023) 2136–2153. [doi:10.1109/JAS.2023.123543](https://doi.org/10.1109/JAS.2023.123543).
- [3] 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](https://doi.org/10.1109/TCI.2016.2629286).