Supplementary Materials of “HLRTF: Hierarchical Low-Rank Tensor Factorization for Inverse Problems in Multi-Dimensional Imaging”

Yisi Luo\(^1\) Xile Zhao\(^1\)* Deyu Meng\(^2,3\) Taixiang Jiang\(^4\)
\(^1\)University of Electronic Science and Technology of China, Chengdu, China
\(^2\)Xi’an Jiaotong University, Xi’an, China
\(^3\)Peng Cheng Laboratory, Shenzhen, China
\(^4\)Southwestern University of Finance and Economics, Chengdu, China

\(^*\)Corresponding author

\texttt{yisiluo1221@foxmail.com, xizhao122003@163.com, dymeng@mail.xjtu.edu.cn, taixiangjiang@gmail.com}

Proof of Theorem 2

\textit{Proof.} We first prove (i). Assume that \(\text{rank}_h(\mathcal{X}) = r\). According to Definition 3, we have \(\text{rank}(f(\mathcal{X})^{(i)}) = r_i \leq r\) for \(i = 1, 2, \ldots, n_3\). Therefore, \(f(\mathcal{X})^{(i)}\) can be factorized as \(f(\mathcal{X})^{(i)} = A_i B_i\), where \(A_i \in \mathbb{R}^{n_1 \times n_i}, B_i \in \mathbb{R}^{r_i \times n_2}\) and they meet \(\text{rank}(A_i) = \text{rank}(B_i) = r_i\). Let \(\bar{A}_i = [A_i, 0] \in \mathbb{R}^{n_1 \times r}\) and \(\bar{B}_i = [B_i; 0] \in \mathbb{R}^{r \times n_2}\), then we have \(f(\mathcal{X})^{(i)} = \bar{A}_i \bar{B}_i\). Let \(A = g(\bar{A}) \in \mathbb{R}^{n_1 \times r \times n_3}\) and \(B = g(\bar{B}) \in \mathbb{R}^{r \times n_2 \times n_3}\) be two tensors, where \(g(\cdot)\) is the inverse DNN of \(f(\cdot)\), \(\bar{A}^{(i)} = \bar{A}_i\), and \(\bar{B}^{(i)} = \bar{B}_i\) \((i = 1, 2, \ldots, n_3)\). Then, we can have

\[
\mathcal{A} \ast f \mathcal{B} = g(f(A) \triangle f(B)) = g(g(\bar{A})) \triangle f(g(\bar{B}))) = g(f(\mathcal{X})) = \mathcal{X}.
\]

Since \(\text{rank}_h(\mathcal{X}) = r\), there exists \(j \in \{1, 2, \ldots, n_3\}\) such that \(\text{rank}(f(\mathcal{X})^{(j)}) = r\) holds. Thus, \(\text{rank}(\bar{A}_j) = \text{rank}(\bar{B}_j) = r\) holds. According to Definition 3, we have \(\text{rank}_h(\mathcal{A}) = \text{rank}_h(\mathcal{B}) = r\) holds.

Then, we prove the property (ii). Assume that \(\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}\) and \(\mathcal{Z} \in \mathbb{R}^{n_2 \times n_4 \times n_3}\). Then,

\[
\text{rank}_h(\mathcal{Y} \ast f \mathcal{Z}) = \max_{i=1,2,\ldots,n_3} \{\text{rank}(f(\mathcal{Y} \ast f \mathcal{Z})^{(i)})\}
\]
\[
= \max_{i=1,2,\ldots,n_3} \{\text{rank}(f(\mathcal{Y}) \triangle f(\mathcal{Z})^{(i)})\}
\]
\[
= \max_{i=1,2,\ldots,n_3} \{\text{rank}(f(\mathcal{Y})^{(i)} f(\mathcal{Z})^{(i)})\}
\]
\[
\leq \max_{i=1,2,\ldots,n_3} \{\min \{\text{rank}(f(\mathcal{Y})^{(i)}), \text{rank}(f(\mathcal{Z})^{(i)})\}\}
\]
\[
\leq \max_{i=1,2,\ldots,n_3} \{\text{rank}(f(\mathcal{Y})^{(i)})\} = \text{rank}_h(\mathcal{Y}).
\]

Similarly, we can have \(\text{rank}_h(\mathcal{Y} \ast f \mathcal{Z}) \leq \text{rank}_h(\mathcal{Z})\). Thus, the following inequality holds: \(\text{rank}_h(\mathcal{Y} \ast f \mathcal{Z}) \leq \min\{\text{rank}_h(\mathcal{Y}), \text{rank}_h(\mathcal{Z})\}\). ❄

Proof of Lemma 2

\textit{Proof.} Assume that \(L(\mathcal{X}, \mathcal{O}) = \| (\mathcal{X} - \mathcal{O})_{\Omega} \|_{F}^{2}\), where \(\mathcal{X} = g(\hat{\mathcal{A}} \triangle \hat{\mathcal{B}})\). Here, we use \(\mathcal{X}(i, j, k)\) or \(\mathcal{X}_{ijk}\) to denote the \(i, j, k\)-th element of \(\mathcal{X}\).

We first prove (i). Suppose that \(i, b, c \notin \Omega\) for arbitrary \(b, c\). Then for arbitrary \(v, w\), the gradient of \(L\) on the \(i, v, w\)-th element of \(\hat{A}\) is

\[
\frac{\partial L(\mathcal{X}, \mathcal{O})}{\partial \hat{A}(i, v, w)} = \sum_{a,b,c} \frac{\partial L}{\partial (g(\hat{\mathcal{A}} \triangle \hat{\mathcal{B}}))_{abc}} \frac{\partial (g(\hat{\mathcal{A}} \triangle \hat{\mathcal{B}}))_{abc}}{\partial \hat{A}(i, v, w)}
\]
\[
= \sum_{(a,b,c) \in \Omega} 2((g(\hat{\mathcal{A}} \triangle \hat{\mathcal{B}}) - \mathcal{O})_{abc}) \frac{\partial (g(\hat{\mathcal{A}} \triangle \hat{\mathcal{B}}))_{abc}}{\partial \hat{A}(i, v, w)}.
\]

Note that

\[
(g(\hat{\mathcal{A}} \triangle \hat{\mathcal{B}}))_{abc} = \sum_{j_1=1}^{n_3} \mathbf{H}_{k_{j_1}} \sigma(\sum_{j_2=1}^{n_3} \mathbf{H}_{k_{j_2}} \cdots \sum_{j_{k-1}=1}^{n_3} \mathbf{H}_{k_{j_{k-1}}} \cdots \sum_{j_{k-2}=-1}^{n_3} \mathbf{H}_{k_{j_{k-2}}} \cdots \sum_{j_{k+1}=-1}^{n_3} \mathbf{H}_{k_{j_{k+1}}} b_{j} b_{j_{k+1}})
\]
\[
\cdots \sum_{j_{k-1}=1}^{n_3} \mathbf{H}_{k_{j_{k-1}}} b_{j} b_{j_{k}})
\]
\[
\cdots \sum_{j_{k+1}=-1}^{n_3} \mathbf{H}_{k_{j_{k+1}}} b_{j} b_{j_{k+1}}).
\]

Therefore, for arbitrary \(a\) such that \(a \neq i\), we have

\[
\frac{\partial (g(\hat{\mathcal{A}} \triangle \hat{\mathcal{B}}))_{abc}}{\partial \hat{A}(i, v, w)} = 0.
\]
Thus,
\[
\frac{\partial L(X, O)}{\partial A(i,v,w)} = \sum_{(i,b,c)\in \Omega} 2((g(\hat{A}\triangle \hat{B}) - O)_{abc})
\]
\[
\frac{\partial (g(\hat{A}\triangle \hat{B}))_{abc}}{\partial A(i,v,w)}.
\]
(6)

Since \((i, b, c) \notin \Omega\) for arbitrary \(b, c\), we have
\[
\frac{\partial L(X, O)}{\partial A(i,v,w)} = 0, v = 1, 2, \ldots, r, w = 1, 2, \ldots, n_3.
\]
(7)

Then, we prove (ii). Suppose that \((a, i, c) \notin \Omega\) for arbitrary \(a, c\). Then for arbitrary \(u, w\), the gradient of \(L\) on the \(u, i, w\)-th element of \(\hat{B}\) is
\[
\frac{\partial L(X, O)}{\partial B(u,i,w)} = \sum_{(a,b,c)\in \Omega} 2((g(\hat{A}\triangle \hat{B}) - O)_{abc})
\]
\[
\frac{\partial (g(\hat{A}\triangle \hat{B}))_{abc}}{\partial B(u,i,w)}.
\]
(8)

According to Eq. (4), for arbitrary \(b\) such that \(b \neq i\), we have
\[
\frac{\partial (g(\hat{A}\triangle \hat{B}))_{abc}}{\partial B(u,i,w)} = 0.
\]
(9)

Thus,
\[
\frac{\partial L(X, O)}{\partial B(u,i,w)} = \sum_{(a,b,c)\in \Omega} 2((g(\hat{A}\triangle \hat{B}) - O)_{aie})
\]
\[
\frac{\partial (g(\hat{A}\triangle \hat{B}))_{aie}}{\partial B(u,i,w)}.
\]
(10)

Since \((a, i, c) \notin \Omega\) for arbitrary \(a, c\), we have
\[
\frac{\partial L(X, O)}{\partial B(u,i,w)} = 0, u = 1, 2, \ldots, r, w = 1, 2, \ldots, n_3.
\]
(11)

Then, we prove (iii). Suppose that \((a, b, i) \notin \Omega\) for arbitrary \(a, b\). Then for arbitrary \(v\), the gradient of \(L\) on the \(i, v\)-th element of \(H_k\) is
\[
\frac{\partial L(X, O)}{\partial H_k(i,v)} = \sum_{(a,b,c)\in \Omega} 2((g(\hat{A}\triangle \hat{B}) - O)_{abc})
\]
\[
\frac{\partial (g(\hat{A}\triangle \hat{B}))_{abc}}{\partial H_k(i,v)}.
\]
(12)

According to Eq. (4), for arbitrary \(c\) such that \(c \neq i\), we have
\[
\frac{\partial (g(\hat{A}\triangle \hat{B}))_{abc}}{\partial H_k(i,v)} = 0.
\]
(13)

Thus,
\[
\frac{\partial L(X, O)}{\partial H_k(i,v)} = \sum_{(a,b,i)\in \Omega} 2((g(\hat{A}\triangle \hat{B}) - O)_{abi})
\]
\[
\frac{\partial (g(\hat{A}\triangle \hat{B}))_{abi}}{\partial H_k(i,v)}.
\]
(14)

Since \((a, b, i) \notin \Omega\) for arbitrary \(a, b\), we have
\[
\frac{\partial L(X, O)}{\partial H_k(i,v)} = 0, v = 1, 2, \ldots, n_3.
\]
(15)

\[\Box\]

**Proof of Theorem 3**

The proof follows from the inverse function of the LeakyReLU function [3], we have \(\sigma^{-1}(0) = 0\). Since \(\sigma^{-1}(\cdot)\) is Lipschitz continuous, there exists \(P \geq 0\) such that \(\|\sigma^{-1}(A)\|_{\ell_1} = \|\sigma^{-1}(A) - \sigma^{-1}(0)\|_{\ell_1} \leq P\|\Delta A - 0\|_{\ell_1} = P\|\Delta A\|_{\ell_1}\) holds for arbitrary \(\Delta A\). Then
\[
\|\nabla_x \hat{A}\|_{\ell_1} = \|\hat{A}(1:\{n_1-1,...,\}) - \hat{A}(2:\{n_1,...,\})\|_{\ell_1} = \|\sigma^{-1}(\cdot)\|_{\ell_1} \leq P\|\nabla_x \hat{A}\|_{\ell_1}.
\]
(16)

Since \(\{H_j\}_{j=1}^k\) and \(\hat{B}\) are bounded, we have
\[
\|\nabla_x X\|_{\ell_1} = J_1\|\nabla_x \hat{A}\|_{\ell_1},
\]
(17)

where \(J_1 = P^{k-1}\|H_1\|_{\ell_1}\|H_2\|_{\ell_1} \cdots \|H_k\|_{\ell_1}\|\hat{B}\|_{\ell_1}\|\hat{A}\|_{\ell_1}\) is a constant.
Similar to (16), we have
\[
\|\nabla_x x\|_{\ell_1} \leq \sum_{k=1}^{K-1} \|\mathbf{H}_1\|_{\ell_1} \|\mathbf{H}_2\|_{\ell_1} \cdots \|\mathbf{H}_k\|_{\ell_1} \|\hat{\mathbf{A}}\|_{\ell_1} \|\hat{\mathbf{B}}\|_{\ell_1} = J_2 \|\nabla_x x\|_{\ell_1},
\]
where \( J_2 = \sum_{k=1}^{K-1} \|\mathbf{H}_1\|_{\ell_1} \|\mathbf{H}_2\|_{\ell_1} \cdots \|\mathbf{H}_k\|_{\ell_1} \|\hat{\mathbf{A}}\|_{\ell_1} \|\hat{\mathbf{B}}\|_{\ell_1} \) is a constant.

Next, we prove the third inequality. Since
\[
\|\nabla_x x\|_{\ell_1} = \|\mathbf{X}_0(x;1:n_3-1) - \mathbf{X}_0(x;2:n_3)\|_{\ell_1} = \left\| -\sigma^{-1}(\cdots - \sigma^{-1}(\sigma^{-1}(\hat{\mathbf{A}}\hat{\mathbf{B}}) \times \mathbf{H}_1) \times \mathbf{H}_2) \cdots \times \mathbf{H}_{k-1} \right\|_{\ell_1}
\]
\[
\leq \left\| -\sigma^{-1}(\cdots - \sigma^{-1}(\sigma^{-1}(\hat{\mathbf{A}}\hat{\mathbf{B}}) \times \mathbf{H}_1) \times \mathbf{H}_2) \cdots \times \mathbf{H}_{k-1} \right\|_{\ell_1} \leq J_3 \|\nabla_x \mathbf{H}_k\|_{\ell_1},
\]
we have
\[
\|\nabla_x x\|_{\ell_1} \leq J_3 \|\nabla_x \mathbf{H}_k\|_{\ell_1},
\]
where \( J_3 = \sum_{k=1}^{K-1} \|\mathbf{H}_1\|_{\ell_1} \|\mathbf{H}_2\|_{\ell_1} \cdots \|\mathbf{H}_k\|_{\ell_1} \|\hat{\mathbf{A}}\|_{\ell_1} \|\hat{\mathbf{B}}\|_{\ell_1} \) is a constant. The proof is completed.

**Implementation Details**

In this work, all experiments are conducted on the PyTorch and MATLAB 2019a platform with an i5-9400f CPU, an RTX 3060 GPU, and 16GB RAM. As for the proposed method, we set the number of network layers \( k \) as 2 in the experiments. The rank \( r \) and the trade-off parameter \( \gamma \) are tuned based on the highest PSNR value. The hyperparameters of compared methods are tuned to obtain the highest PSNR value. For CNN-based methods, we use the pretrained model provided by the authors.

In this work, we consider the simple fully connected network. In future work, we can consider some advanced DNNs (e.g., attention) in our framework.

The limitation of our method lies in the manual selection of the hyperparameters, i.e., the rank \( r \) and the trade-off parameter \( \gamma \). In experiments, we select these parameters to obtain the best PSNR value. Numerical tests in Fig. 1 show that our method is relatively insensitive to the choice of \( r \) and \( \gamma \). Thus, it is not difficult to choose these hyperparameters in experiments.

**More Experimental Results**

Please see Figs. 3-7 for more visual results. Our method shows advantageous performance against compared state-of-the-art methods for different tasks. Moreover, in Fig. 2, we show the visualizations of the learned matrices \( \mathbf{H}_1 \) and \( \mathbf{H}_2 \) for HSIs WDC mall and Pavia with SR=0.1. Compared with the fixed DFT matrix, our learned nonlinear transform is more flexible to represent different data.

![Figure 1. Tensor completion results vs. \( \gamma \) and \( r \) (Pavia, SR=0.1).](image1)

![Figure 2. Given different data (e.g., the HSI WDC mall and Pavia), traditional t-SVD methods [6, 21] use fixed DFT matrix for multidimensional image recovery. In contrast, our DeepLRTF flexibly learns different nonlinear transforms (see the learned matrices \( \mathbf{H}_1 \) and \( \mathbf{H}_2 \)) for different data to help obtain better performance.](image2)

**References**


[7] Canyi Lu, Xi Peng, and Yunchao Wei. Low-rank tensor completion with a new tensor nuclear norm induced by invertible
Figure 3. The results of multi-dimensional image completion by different methods on HSI WDC mall, HSI Pavia, video Backdoor, and video Yard with SR=0.1.

Figure 4. The results of MSI denoising by different methods on HSI WDC mall, HSI Pavia, MSI Balloons, and MSI Fruits for Case 1.
Figure 5. The results of MSI denoising by different methods on MSIs Balloons, Fruits, Pool, and Doll for Case 2.

Figure 6. The results of snapshot spectral imaging by different methods on HSI WDC mall (SR = 0.5), HSI Pavia (SR = 0.5), MSI Toy (SR = 0.5), and MSI Flowers (SR = 0.3).
Figure 7. The results of multi-dimensional image completion by different methods on WDC mall with horizontal/lateral slice missing and frontal slice missing.


[14] Longhao Yuan, Chao Li, Danilo P. Mandic, Jiating Cao, and Qibin Zhao. Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion. In AAAI, 2019, pages 9151–9158, 2019. 4, 6


