applied sciences

MDPI

Article

A Fast Self-Learning Subspace Reconstruction Method for Non-Uniformly Sampled Nuclear Magnetic Resonance Spectroscopy

Zhangren Tu [1], Huiting Liu [2], Jiaying Zhan [1] and Di Guo [1,*]

[1] School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China; tzr12308@163.com (Z.T.); skylarzhan@outlook.com (J.Z.)

[2] Department of Electronic Science, Xiamen University, Xiamen 361005, China; madkm@163.com

* Correspondence: guodi@xmut.edu.cn

Received: 4 May 2020; Accepted: 2 June 2020; Published: 5 June 2020

Abstract: Multidimensional nuclear magnetic resonance (NMR) spectroscopy is one of the most crucial detection tools for molecular structure analysis and has been widely used in biomedicine and chemistry. However, the development of NMR spectroscopy is hampered by long data collection time. Non-uniform sampling empowers rapid signal acquisition by collecting a small subset of data. Since the sampling rate is lower than that of the Nyquist sampling ratio, undersampling artifacts arise in reconstructed spectra. To obtain a high-quality spectrum, it is necessary to apply reasonable prior constraints in spectrum reconstruction models. The self-learning subspace method has been shown to possess superior advantages than that of the state-of-the-art low-rank Hankel matrix method when adopting high acceleration in data sampling. However, the self-learning subspace method is time-consuming due to the singular value decomposition in iterations. In this paper, we propose a fast self-learning subspace method to enable fast and high-quality reconstructions. Aided by parallel computing, the experiment results show that the proposed method can reconstruct high-fidelity spectra but spend less than 10% of the time required by the non-parallel self-learning subspace method.

Keywords: nuclear magnetic resonance spectroscopy; non-uniform sampling; self-learning subspace; matrix factorization; acceleration

1. Introduction

Multidimensional nuclear magnetic resonance (NMR) spectroscopy plays an important role in the fields of biomedicine and chemistry [1–3]. However, long data acquisition time [4–6] has to be solved, and one of the effective approaches to reduce the time is to acquire partial data with non-uniform sampling (NUS) [7–12]. To obtain a full spectrum, reconstruction needs to incorporate priors, and the most common ones are sparsity [13,14] and low rankness [15–17]. Sparsity [18–20] assumes the spectrum has few non-zero values in the spectrum but will not be satisfied at low-intensity broad peaks [21]. The low rank approach transforms the time-domain NMR signal (called free induction decay (FID)) into a Hankel matrix, and explores the low rankness of this matrix. Therefore, it is also called a low-rank Hankel matrix (LRHM) method. Since the peak width will not affect the rank, LRHM offers better reconstructions for these challenging peaks, even though low-intensity peaks, referring to small singular values in the matrix rank minimization, may be compromised or even lost in LRHM reconstructions [22]. This problem also exists in general low-rank matrix reconstruction and has been alleviated by introducing the truncated nuclear norm (TNN) to preserve the small singular values in the reconstruction [23].

Recently, we proposed a self-learning subspace (SLS) [24] method to mitigate the spectra degradation of LRHM at high acceleration factors. Beyond the singular values, we found that the subspace of the Hankel matrix corresponds to spectral peaks, and the TNN provides one approach to incorporate the subspace information. With the help of TNN, we establish the concept of signal subspace reconstruction in the low-rank Hankel matrix. The signal space is divided into two subspaces, an obvious, easily-restored strong signal subspace and an unstable, hard-to-estimate weak signal subspace [24]. This method includes two iterations loops: The outer loops update the signal subspace with prior constraints, and the inner one performs updates of signal reconstruction under the given subspace. Experimental results show that SLS achieves as high-quality reconstructions as does LRHM, and notably, produces better reconstructions of low-intensity peaks [24].

The SLS approach, however, requires intensive computations of singular value decomposition (SVD). In the iterative reconstruction, SVD will be called approximately 100 to 200 times, which is very time-consuming. In addition, since the computational complexity of SVD is proportional to the three power of the matrix size [25], the lengthy SVD computations may be impractical due to the data size and dimensions rapidly increasing in high-dimensional NMR spectroscopy.

In this work, we propose an accelerated algorithm for the state-of-the-art self-learning subspace method by introducing matrix factorization [26,27] to avoid SVD. Results on synthetic and realistic NMR data show that compared with the SLS, the fast SLS approach saves considerable time without sacrificing spectrum quality and enables faster reconstruction with parallel computing.

2. Related Work

Mathematically, a 1D FID can be modeled as the finite sum of damping exponential functions as follows [28–31]:

$$x_n = \sum_{j=1}^{J} \left(A_j e^{i\phi_j} \right) e^{-\frac{n\Delta t}{\tau_j}} e^{in\Delta t\omega_j}, \tag{1}$$

where J is the number of spectral peaks, Δt is the time interval, and $A_j, \phi_j, \tau_j, \omega_j$ are the amplitude, phase, decay time, and frequency of the jth spectral peak, respectively. If there are N sampled data points, the sampled signal can be represented as a vector $\mathbf{x} = [x_1, x_2, x_3, \cdots, x_n, \cdots, x_N]^T$, where $\bullet^T$ represents vector transpose.

The reconstruction model of the LRHM method is [21]:

$$\min_{\mathbf{x}} \|\mathcal{R}\mathbf{x}\|_* + \frac{\lambda}{2}\|\mathbf{y} - \mathcal{U}\mathbf{x}\|_2^2 \tag{2}$$

where $\mathcal{R}$ denotes an operator converting the undersampled FID $\mathbf{x}$ into a Hankel matrix $\mathcal{R}\mathbf{x}$, $\mathcal{U}$ is an undersampling operator, $\|\bullet\|_*$ represents the matrix nuclear norm as the sum of singular values, $\|\bullet\|_2$ represents the l_2 norm that measures data consistency, and λ is a regularization parameter that balances the low rankness and the data consistency.

However, the LRHM model may sacrifice low-intensity peak recovery, which is related to small singular values. As we know, the information of a spectral peak includes its intensity, central frequency, and line shape. The nuclear norm only focuses on intensity. In order to utilize the extra information, including central frequency and line shape, hidden in the Hankel matrix, the SLS method was proposed [24]. The SLS method built a subspace framework to separate the original signal into a relatively strong signal subspace and a weak one. Then, peaks that lie in the weak subspace are particularly protected by enforcing them to have prior spectral peak shapes. The self-learning subspace reconstruction model is expressed as [24]:

$$\min_{\mathbf{x}} P_\Omega(\mathbf{x}) + \frac{\lambda}{2}\|\mathbf{y} - \mathcal{U}\mathbf{x}\|_2^2, \tag{3}$$

where:

$$P_{\Omega}(\mathbf{x}) = \|\mathcal{R}\mathbf{x}\|_{*} - Tr\left(\mathbf{A}_{\Omega}^{H}\mathcal{R}\mathbf{x}\mathbf{B}_{\Omega}\right), \tag{4}$$

where Ω represents strong signal space, $Tr(\cdot)$ is the trace function and defined as the sum of the main diagonal elements, $\mathbf{A}_{\Omega} \in \mathbb{C}^{r\times P}$ and $\mathbf{B}_{\Omega} \in \mathbb{C}^{r\times Q}$ are two matrices formed by the first *r* columns of the left and right unitary matrixes, satisfying SVD $\mathcal{R}\mathbf{x}=\mathbf{A}\Lambda\mathbf{B}^{H}$. The function $P_{\Omega}(\mathbf{x})$ sums up the least *r* singular values. We can estimate $\mathbf{A}_{\Omega}$ and $\mathbf{B}_{\Omega}$ from the initial solution $\mathbf{x}_0$, and we can improve the accuracy of the subspace as it iterates continuously. For example, the subspace $\mathbf{A}_{\Omega,l}, \mathbf{B}_{\Omega,l}$ of the lth outer iteration is more accurate than the initial estimate. Machine learning typically includes supervised learning, semi-supervised learning, and unsupervised learning, which use all, partial, and non-labeled samples, respectively. Transduction learning [32] is another type that predicts specific test samples by learning specific training samples.

The proposed method belongs to unsupervised learning since no labeled samples or specific test samples are used. The singular value decomposition is adopted to learn the subspace of the Hankel matrix. The self-learning process means no prior or label information is provided in advance, and we only learn the subspace space information from the initial time-domain signal that is filled with zeros, and we update this information from the intermediate reconstruction. In practice, prior or label spectra are not accessible in the biomedical magnetic resonance experiment. Thus, unsupervised self-learning is a reasonable choice in this application.

Nevertheless, the application of SLS to higher dimensional NMR spectroscopy experiments is impeded by the expensive computation of SVD demanded by computing the nuclear norm term. Now, to reduce the computation time, we will introduce the matrix factorization and derive the fast algorithm.

3. Methods

For a given matrix, which can be factorized into two matrices, $\mathbf{P}\in\mathbb{C}^{m\times r}$ and $\mathbf{Q}\in\mathbb{C}^{n\times r}$ (size r is unlimited), its nuclear norm can be computed as the following [33,34]:

$$\|\mathbf{X}\|_{*} = \min_{\mathbf{P},\mathbf{Q}} \frac{1}{2}\left(\|\mathbf{P}\|_{F}^{2} + \|\mathbf{Q}\|_{F}^{2}\right) \qquad s.t. \mathbf{X} = \mathbf{P}\mathbf{Q}^{H}, \tag{5}$$

where $\|\bullet\|_{F}$ denotes the Frobenius norm of a matrix and $\bullet^{H}$ means conjugate transpose.

In this work, we propose a self-learning subspace matrix factorization (SLSMF) method to accelerate the state-of-the-art SLS method:

$$\min_{\mathbf{x},\mathbf{P},\mathbf{Q}} \frac{1}{2}\left(\|\mathbf{P}\|_{F}^{2} + \|\mathbf{Q}\|_{F}^{2}\right) - Tr\left(\mathbf{A}_{\Omega,l}^{H}\mathcal{R}\mathbf{x}\mathbf{B}_{\Omega,l}\right) + \frac{\lambda}{2}\|\mathbf{y} - \mathcal{U}\mathbf{x}\|_{F}^{2} \qquad s.t. \mathcal{R}\mathbf{x} = \mathbf{P}\mathbf{Q}^{H}, \tag{6}$$

It requires no SVD computations in the inner loop, which greatly speeds up the reconstruction.

Next, we will solve the optimization problem of Equation (6). The augmented Lagrange form of Equation(6) is:

$$\begin{aligned} L(\mathbf{x},\mathbf{P},\mathbf{Q},\mathbf{D}) = \min_{\mathbf{x},\mathbf{P},\mathbf{Q}} \max_{\mathbf{D}} \frac{1}{2}\|\mathbf{P}\|_{F}^{2} + \frac{1}{2}\|\mathbf{Q}\|_{F}^{2} - Tr\left(\mathbf{A}_{\Omega,l}^{H}\mathcal{R}\mathbf{x}\mathbf{B}_{\Omega,l}\right) + \\ \frac{\lambda}{2}\|\mathbf{y} - \mathcal{U}\mathbf{x}\|_{2}^{2} + \left\langle \mathbf{D}, \mathcal{R}\mathbf{x} - \mathbf{P}\mathbf{Q}^{H}\right\rangle + \frac{\beta}{2}\left\|\mathcal{R}\mathbf{x} - \mathbf{P}\mathbf{Q}^{H}\right\|_{F}^{2}, \end{aligned} \tag{7}$$

where $\mathbf{D}$ is a dual variable used to improve the convergence speed of the algorithm, $\langle\cdot,\cdot\rangle$ is the inner product in the Hilbert space of matrices, and $\beta > 0$ is the regularization parameter.

The proposed algorithm consists of two main loops. The outer one iteratively updates the signal subspace $\mathbf{A}_{\Omega,l}, \mathbf{B}_{\Omega,l}$ to determine the best subspace under the prior information. Once the subspace

is obtained, in the inner one, in order to solve Equation(7), it is converted into four sub-problems by adopting alternating direction methods of multipliers (ADMM) [35]:

$$\begin{cases} \mathbf{x}_{k+1} = arg\,\underset{\mathbf{x}}{min}\, L\left(\mathbf{x}, \mathbf{P}_k, \mathbf{Q}_k, \mathbf{D}_k\right) \\ \mathbf{P}_{k+1} = arg\,\underset{\mathbf{P}}{min}\, L\left(\mathbf{x}_{k+1}, \mathbf{P}, \mathbf{Q}_k, \mathbf{D}_k\right) \\ \mathbf{Q}_{k+1} = arg\,\underset{\mathbf{Q}}{min}\, L\left(\mathbf{x}_{k+1}, \mathbf{P}_{k+1}, \mathbf{Q}, \mathbf{D}_k\right) \\ \mathbf{D}_{k+1} = \mathbf{D}_k + \left(\mathbf{R}\mathbf{x}_{k+1} - \mathbf{P}_{k+1}\mathbf{Q}_{k+1}^H\right) \end{cases}, \tag{8}$$

(1) Fixing $\mathbf{P}_k, \mathbf{Q}_k, \mathbf{D}_k$, $\mathbf{x}_{k+1}$ is obtained by solving:

$$\underset{\mathbf{x}}{min}\frac{\lambda}{2}\left\|\mathbf{y} - \mathcal{U}\mathbf{x}\right\|_2^2 - \mathrm{Tr}\left(\mathbf{A}_r^H \mathcal{R}\mathbf{x}\mathbf{B}_r\right) + \left\langle \mathbf{D}, \mathcal{R}\mathbf{x} - \mathbf{P}\mathbf{Q}^H \right\rangle + \frac{\beta}{2}\left\|\mathcal{R}\mathbf{x} - \mathbf{P}\mathbf{Q}^H\right\|_F^2. \tag{9}$$

The solution is:

$$\mathbf{x}_{k+1} = \left(\lambda\mathcal{U}^H\mathcal{U} + \beta\mathcal{R}^H\mathcal{R}\right)^{-1}\left[\lambda\mathcal{U}^H\mathbf{y} + \beta\mathcal{R}^H\left(\mathbf{P}_k\mathbf{Q}_k^H + \frac{\mathbf{A}_{\Omega,l}^H\mathcal{R}\mathbf{x}\mathbf{B}_{\Omega,l} - \mathbf{D}_k}{\beta}\right)\right]. \tag{10}$$

(2) Fixing $\mathbf{x}_{k+1}, \mathbf{Q}_k, \mathbf{D}_k$, $\mathbf{P}_{k+1}$ is obtained by solving:

$$\underset{\mathbf{P}}{min}\frac{1}{2}\left\|\mathbf{P}\right\|_F^2 + \left\langle \mathbf{D}, \mathcal{R}\mathbf{x} - \mathbf{P}\mathbf{Q}^H \right\rangle + \frac{\beta}{2}\left\|\mathcal{R}\mathbf{x} - \mathbf{P}\mathbf{Q}^H\right\|_F^2. \tag{11}$$

The solution is:

$$\mathbf{P}_{k+1} = \left(\beta\mathcal{R}\mathbf{x}_{k+1} + \mathbf{D}_k\right)\mathbf{Q}_k\left(\beta\mathbf{Q}_k^H\mathbf{Q}_k + \mathbf{I}\right)^{-1}. \tag{12}$$

(3) Fixing $\mathbf{x}_{k+1}, \mathbf{P}_{k+1}, \mathbf{D}_k$, $\mathbf{Q}_{k+1}$ is obtained by solving:

$$\underset{\mathbf{Q}}{min}\frac{1}{2}\left\|\mathbf{Q}\right\|_F^2 \left\langle \mathbf{D}, \mathcal{R}\mathbf{x} - \mathbf{P}\mathbf{Q}^H \right\rangle + \frac{\beta}{2}\left\|\mathcal{R}\mathbf{x} - \mathbf{P}\mathbf{Q}^H\right\|_F^2. \tag{13}$$

The solution is:

$$\mathbf{Q}_{k+1} = \left(\beta\mathcal{R}\mathbf{x}_{k+1} + \mathbf{D}_k\right)^H \mathbf{P}_{k+1}\left(\beta\mathbf{P}_{k+1}^H\mathbf{P}_{k+1} + \mathbf{I}\right)^{-1}. \tag{14}$$

(4) Fixing $\mathbf{x}_{k+1}, \mathbf{P}_{k+1}, \mathbf{Q}_{k+1}$, the solution of $\mathbf{D}_{k+1}$ is:

$$\mathbf{D}_{k+1} = \mathbf{D}_k + \left(\mathcal{R}\mathbf{x}_{k+1} - \mathbf{P}_{k+1}\mathbf{Q}_{k+1}^H\right). \tag{15}$$

The alternating iterations in Equation(8) stop if the number of iterations k reaches the maximal number K, or the normalized successive difference $\eta_{k+1} = \left\|\mathbf{x}_{k+1} - \mathbf{x}_k\right\| / \left\|\mathbf{x}_k\right\|$ is smaller than that of a given tolerance η_{tol}. The pseudocode of this reconstruction algorithm is shown in Table 1.

Table 1. The reconstruction algorithm of self-learning subspace matrix factorization (SLSMF).

Initialization: Input $\mathbf{y}$, $\mathcal{R}$, $\mathcal{U}$, set outer maximal iterations times L=5, convergence condition $\eta_{\mathrm{tol}} = 10^{-6}$, and maximal inner number of iterations $K = 10^3$. Initialize the solution $\mathbf{x}_0 = \mathcal{U}^T\mathbf{y}$, the dual variable $\mathbf{D}_0 = \mathbf{1}$, the number of iterations $k = 0$, and $\eta_0 = 1$.
Main:
While ($\eta_l \geq \eta_{\mathrm{tol}}$) or ($l < L$), **do:**
1) Estimate $\mathbf{A}_{\Omega,l}$ and $\mathbf{B}_{\Omega,l}$ from the first r columns of $\mathbf{A}_l$ and $\mathbf{B}_l$, such that $\mathcal{R}\mathbf{x}_l = \mathbf{A}_l\mathbf{\Lambda}\mathbf{B}_l^H$

2) **While** ($\eta_k \geq \eta_{\text{tol}}$) or ($k < K$), **do;**

   a) Update $\mathbf{x}_{k+1}$ according to Equation(10);
   b) Update $\mathbf{P}_{k+1}$ according to Equation(12);
   c) Update $\mathbf{Q}_{k+1}$ according to Equation(14);
   d) Update $\mathbf{D}_{k+1}$ according to Equation(15);
   e) Compute $\eta_{k+1} = \|\mathbf{x}_{k+1} - \mathbf{x}_k\| / \|\mathbf{x}_k\|$ and $k \leftarrow k+1$;

   End while;

3) Set $k = 0$, $\eta_k = 1$ and $l \leftarrow l+1$;
4) Update $\mathbf{x}_{l+1} \leftarrow \mathbf{x}_{k+1}$, $\eta_{l+1} = \|\mathbf{x}_{l+1} - \mathbf{x}_l\| / \|\mathbf{x}_l\|$;

End for;

Output: The reconstructed FID $\hat{\mathbf{x}} \leftarrow \mathbf{x}_{l+1}$.

A 2D NMR spectrum contains two dimensions, one of which is the direct dimension t_2, and the other is the indirect dimension t_1 on which almost all the experiment time is spent. In the case of the 2D NMR spectrum, reconstruction can be performed on the 1D reconstruction of each slice one by one. Thus, it is possible to accelerate the computation by making use of the proposed parallel architecture. In detail, the reconstruction tasks for different slices are assigned to multiple processing cores [36] and can be computed simultaneously (Figure 1).

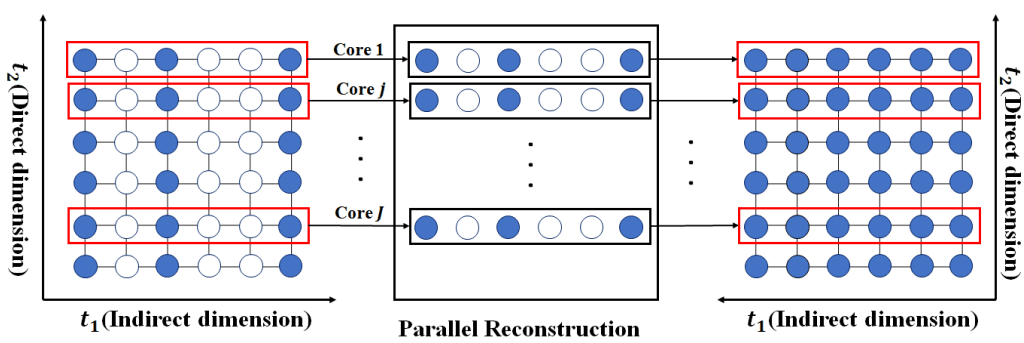

Figure 1. Schematic diagram of parallel computing of NMR spectroscopy reconstruction. Note: The blue dot represents the sampled data point and the white one represents unsampled data point; "Core j" represents the jth CPU (Central Processing Unit) core of a multi-core CPU for parallel computation. The non-uniform sampling (NUS) is performed along the indirect dimension (t_1), and each 1D vector (in red rectangle) located at each data point in the direct dimension (t_2) can be reconstructed separately.

4. Experiments and Results

We compare the proposed SLSMF method with the other three state-of-the-art NMR reconstruction methods, including the LRHM, the LRHM with matrix factorization (LRHMF) [26], and the SLS. Spectra reconstruction experiments were conducted on both synthetic 1D NMR spectra (Figure 2) and real 2D NMR spectra (Figure 3-5). All reconstruction algorithms were performed using MATLAB 2017b (Mathworks Inc., Natick, MA, USA) on a computational server with two E5-2650v4 CPUs (12 core/CPU) and 160 GB RAM.

The important spectra parameters, including two 2D spectra of small, large, and intrinsically disordered proteins and one 2D spectrum of solid NMR, are listed in Table 2. More details can be found in the experimental descriptions below.

Table 2. Important parameters of experimental 2D spectra.

Type	Protein	Molecular Weight	Spectrometer Frequency	Sampling Type	Date Point ($t_2 \times t_1$)	References
HSQC	GB1	~8.0 kDa	600 MHz	Full	1466×170	Figure 3
TROSY	Ubiquitin	~8.6 kDa	800 MHz	Full	683×128	Figure 4
Solid NMR		~68 Da	900 MHz	Full	249×150	Figure 5

Note: The NUS and reconstruction are performed on the indirect dimension, which is noted as t_1.

The 2D ^{1}H-^{15}N HSQC spectrum of GB1 was prepared with following conditions: The GB1 object sample is 2 mM U-^{15}N, 20%-^{13}C GB1 in 25 mM PO_4, pH 7.0 with 150 mM NaCl, and 5% D_2O. Data were collected using a phase-cycle selected HSQC (hsqcfpf3gpphwg in Bruker library) at 298 K on a Bruker Avance 600 MHz spectrometer using a room temp HCN TXI probe, equipped with a z-axis gradient system. The fully sampled spectrum consists of 1676 × 170 complex points; the direct dimension (^{1}H) has 1676 data points, while the indirect dimension (^{15}N) has 170 data points.

The 2D ^{1}H-^{15}N best-TROSY spectrum of ubiquitin was acquired at 298.2 K temperature on an 800 MHz Bruker spectrometer and was described in a previous paper [37]. The fully sampled spectrum consists of 683 × 128 complex points; the direct dimension (^{1}H) has 683 data points, while the indirect dimension (^{15}N) has 128 data points.

The solid-state NMR spectrum was acquired from an imidazole sample at room temperature on a 21.1 T Bruker AVANCE-III spectrometer. The sample was spinning at 60 kHz inside a 1.3 mm probe. The 2D ^{1}H-^{1}H double-quantum spectrum was recorded using a symmetry-based R1225 pulse sequence with a recycle delay of 5 s. Data were recorded by co-adding 16 transients with 300 t1 increments. The number of fully sampled data points in the indirect dimension is 150.

In addition, to quantify the reconstruction error, a relative l_2-norm error (RLNE) defined as:

$$\mathrm{RLNE}\left(\hat{\mathbf{x}}\right) = \frac{\left\|\hat{\mathbf{x}} - \mathbf{x}\right\|_2}{\left\|\mathbf{x}\right\|_2} \tag{16}$$

where $\hat{\mathbf{x}}$ is the reconstructed signal and $\mathbf{x}$ is the fully sampled signal, is adopted.

4.1. Reconstruction of Synthetic 1D NMR Spectra

The synthetic 1D NMR spectrum is generated according to Equation(1), and its specific parameter settings are shown in Table 3. The total number of data points is 512, and 31 complex data points are acquired (means 6% NUS data). Then, simulated Gaussian noise with a mean of zero and a standard deviation of 0.005 is added to the FID. Here, the maximum value of the noise is lower than that of the lowest spectral peak.

The reconstruction results are shown in Figure 2. Under very limited sampled data, the first and second reconstructed spectral peaks are severely damaged in LRHM low-fidelity reconstructions, while the SLS and the proposed method reconstruct the spectral peaks with high fidelity (see the supplementary materials Table S1 for the correlation of each peak in Figure 2). In Table 4, it is clearly shown that the robustness of the proposed method is consistent with SLS, and the Pearson correlation coefficient is much higher than LRHM at low intensity spectral peaks. The proposed method can accelerate the reconstruction of the best method, SLS, without compromising the spectra quality. Thus, it is correct and with the expectation that no significant difference between Figure 2f, the reconstructed spectra using SLSMF, and Figure 2e, the reconstructed spectra using SLS. Accordingly, LRHMF only accelerates the algorithm of LRHM, thus Figure 2c,d should be the same. In addition, in the supplementary materials, we displayed more reconstruction results of traditional methods with moderate-fidelity (Figure S2) and high-fidelity (Figure S3) from 100 simulation data trails and

the correlation of their corresponding spectral peaks. Since the 1D reconstruction is fast (several seconds) for all compared methods, there is no need to compare the computation time.

Table 3. Parameter settings of the synthetic 1D NMR spectrum.

Parameters \ Peaks ID	1	2	3	4	5
Amplitude (A_j)	0.3	0.4	0.5	1	1
Damping factor (τ_j)	0.01	0.02	0.03	0.04	0.08

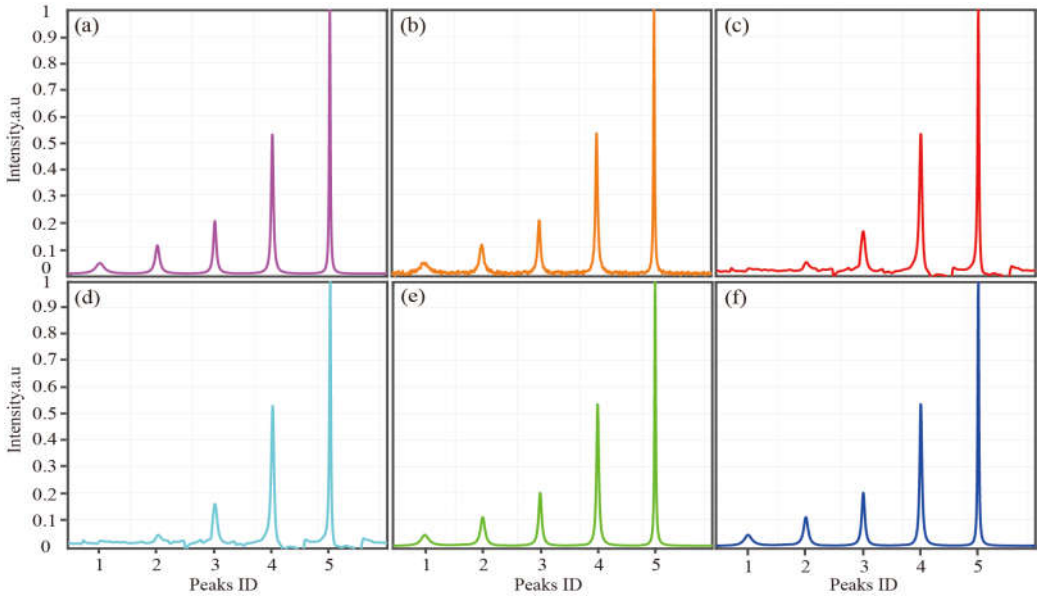

Figure 2. Reconstructed synthetic 1D spectra with five peaks from 6% acquired data. (**a**) The fully sampled spectrum; (**b**) the noisy fully sampled spectrum; (**c**–**f**) reconstructed spectra using the low-rank Hankel matrix (LRHM), LRHM with matrix factorization (LRHMF), self-learning subspace (SLS) and SLSMF, respectively.

Table 4. Statistics of peaks correlations and standard deviations.

Method \ Peak ID	1	2	3	4	5
LRHM	0.597 ± 0.384	0.859 ± 0.203	0.944 ± 0.090	0.991 ± 0.012	0.997 ± 0.004
LRHMF	0.600 ± 0.384	0.857 ± 0.203	0.943 ± 0.090	0.991 ± 0.012	0.997 ± 0.004
SLS	0.935 ± 0.183	0.974 ± 0.152	0.992 ± 0.039	0.999 ± 0.003	0.999 ± 0.001
SLSMF	0.936 ± 0.181	0.974 ± 0.152	0.993 ± 0.035	0.999 ± 0.002	0.999 ± 0.001

Note: Peak intensity correlations and error bars for the reconstructed spectra when 6% of data are acquired. The error bars are the standard deviations of correlations over 100 sampling trials. The numerical form of the table is $C \pm D$, where C is the mean of correlation, and D is the standard deviation.

4.2. Reconstruction of 2D NMR Spectra

Reconstruction experiments with realistic NMR are conducted on 15% NUS 2D ^{1}H-^{15}N HSQC spectra of GB1, 20% NUS 2D ^{1}H-^{15}N best-TROSY spectra of ubiquitin, and 15% NUS 2D ^{1}H-^{1}H solid-state spectra.

Figures 3a, 4a, and 5a are fully sampled 2D ^{1}H-^{15}N HSQC spectra of GB1, 2D ^{1}H-^{15}N best-TROSY spectra of ubiquitin, and 2D ^{1}H-^{1}H solid-state spectrum, respectively. Results show that all methods can reconstruct a majority of spectral peaks well. However, some peaks in best-TROSY experiments

shown by arrows in Figure 4b,c are lost in LRHM and LRHMF. LRHM and LRHMF also produce some pseudo peaks in the reconstructed GB1 and solid spectra, as shown by arrows in Figures 3b,c and 5b,c. By contrast, SLS and SLSMF methods provide faithful reconstructions.

What is more, Figure 6 illustrates that the proposed method performs better in reconstructing low-intensity peaks. For example, LRHM and LRHMF weaken the reconstructed spectral peak near 125 ppm (Figure 6b–1,c–1)), while SLSMF obtains the spectrum with peak intensity and line shape close to the fully sampled spectrum (Figure 6e–1).

Regression analyses (Figures 7–9) also exhibit that the SLSMF outperforms LRHM and LRHMF. It is shown that the SLSMF method gets a higher correlation than LRHM and LRHMF, especially in low-intensity peaks (Figures 7e–h, 8e–h, and 9e–h). Besides, lower RLNEsby SLSMF can be achieved than in LRHM and LRHMF, and RLNEs of SLSMF decrease much faster (Figure 10). For instance, we can see that after the update of the subspace, the RLNEs of SLSMF drop almost linearly, implying that the introduction of the signal subspace can help reconstruct high-fidelity spectra more quickly and accurately with the same sampling rate. The above observations indicate that in terms of reconstruction time and reconstruction quality, the proposed approach permits considerable improvements from the SLS and low-rank Hankel matrix methods.

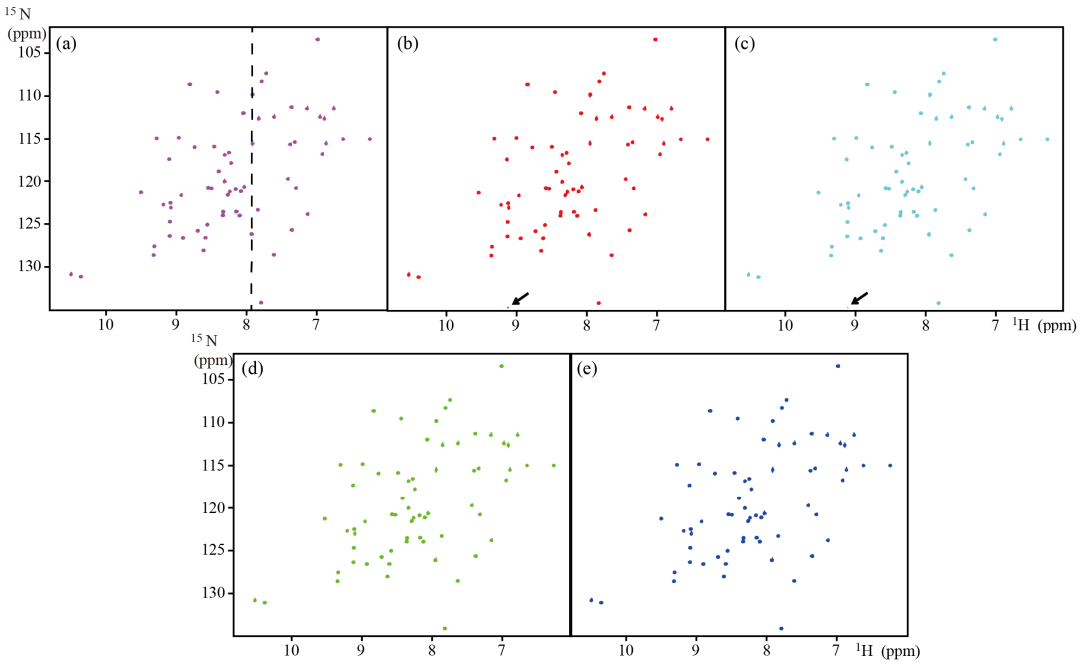

Figure 3. Reconstructed 2D HSQC spectra of GB1 with 15% NUS data. (**a**) The fully sampled spectrum; (**b**–**e**) reconstructed spectra using the LRHM, LRHMF, SLS, and SLSMF, respectively. Note: The contours of spectra are at the same level.

From Figure 11, we can see that compared to the SLS approach, the proposed method significantly speeds up the reconstruction time about 2–4 times (Figure 11a). If equipped with parallel computing architecture, the acceleration in the reconstruction time of SLSMF would be striking. Take the GB1 experiment as an example: With the parallel computing SLSMF, the time spent on reconstructing spectra only requires 4.6% of the computation of the SLS without parallelism. However, if the data dimension is too small, parallel computing is not recommended because it must pay for a certain time on switching the parallel pool, i.e., a mechanism used to control multiple processing cores in software. For example, for the reconstruction of the solid-state spectra, using the LRHMF only takes about 22 s, but it takes 26 s under parallel computing, whereas switching the parallel pool spends about 24 s. We also ran methods on the personal computer and the computation time, as reported in Figure 12. The calculation time with the personal computer also confirms that the

proposed method can accelerate the reconstruction and can further reduce the time by parallel computing.

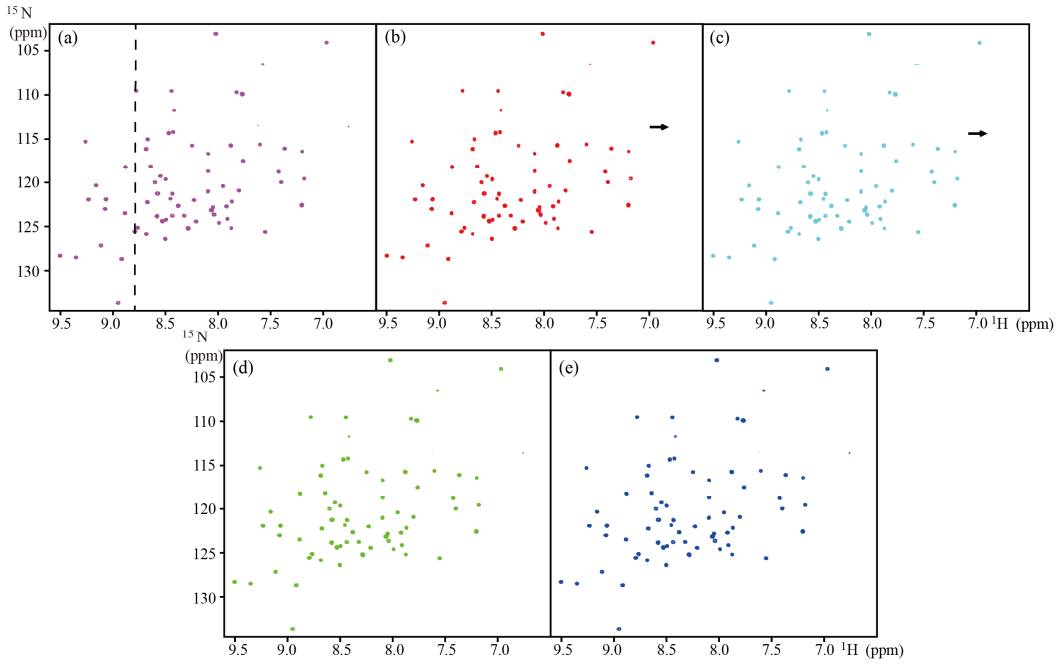

Figure 4. Reconstructed 2D ^{1}H–^{15}N best-TROSY spectra of ubiquitin with 20% NUS data. (**a**) The fully sampled spectrum; (**b**–**e**) reconstructed spectra using the LRHM, LRHMF, SLS, and SLSMF, respectively. Note: The contours of spectra are at the same level.

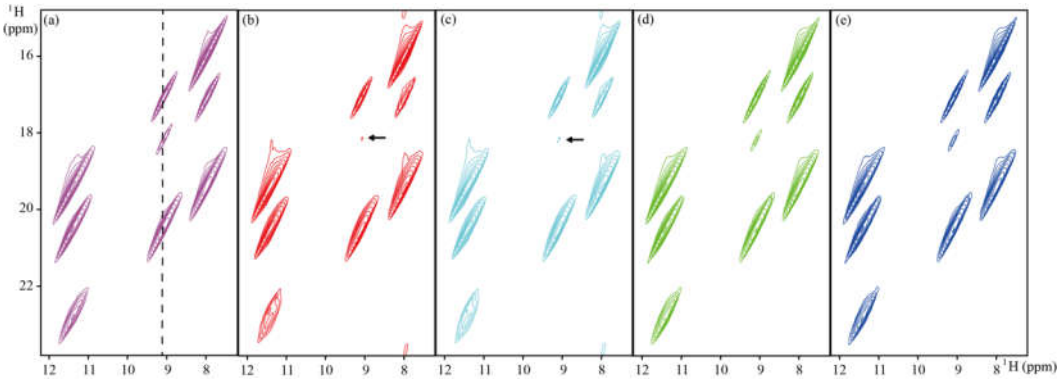

Figure 5. Reconstructed solid NMR spectra with 15% NUS data. (**a**) The fully sampled spectrum; (**b**–**e**) reconstructed spectra using the LRHM, LRHMF, SLS, and SLSMF, respectively. Note: The contours of spectra are at the same level.

In the fast NUS NMR, high-fidelity reconstruction of the spectrum is most important, followed by the rest of other performances, such as the computation time. Our work is to design a new algorithm to reduce the computation time while achieving the best spectra. To emphasize our proposed method's advantages in both reconstruction quality and speed, in Table 5 we list the advantages and disadvantages of the comparison method.

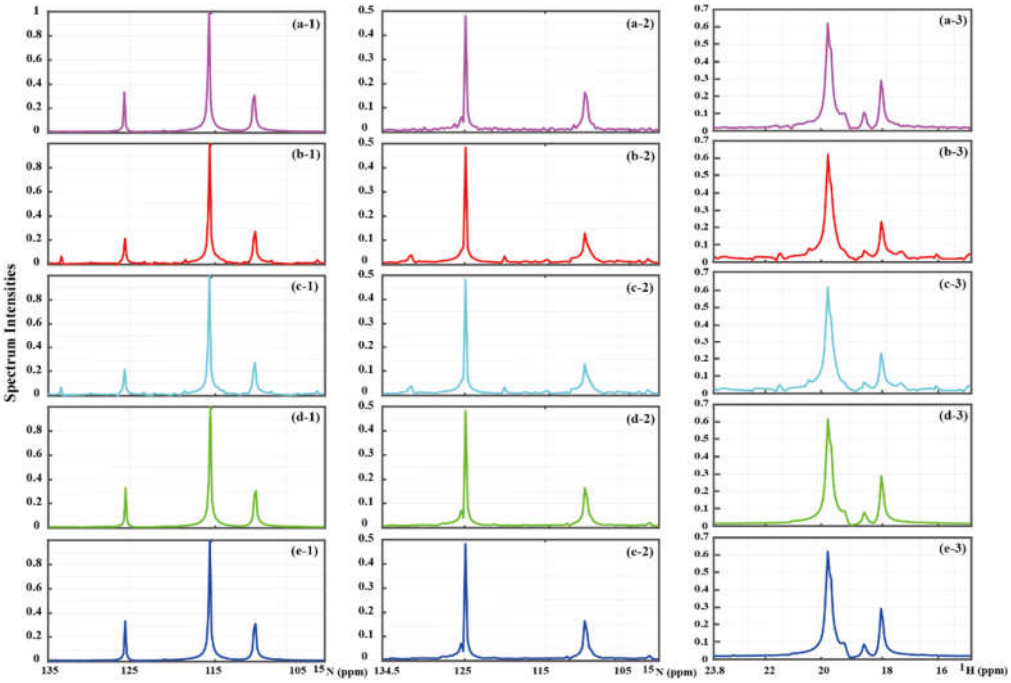

Figure 6. 1D traces of reconstructed 2D spectra (**a-1**)–(**a-3**) is the fully sampled spectrum of the 1D traces of HSQC spectra of GB1, the best-TROSY spectrum of ubiquitin, and solid NMR spectrum, respectively. (**b-1**)–(**b-3**), (**c-1**)–(**c-3**), (**d-1**)–(**d-3**), and (**e-1**)–(**e-3**) are reconstructed spectra using LRHM, LRHMF, SLS, and SLSMF, respectively. Note: (**a-1**)–(**a-3**) are fully sampled spectra of 1D traces, and their locations in the corresponding fully sampled spectrum are marked with a dotted line in Figures 3–5.

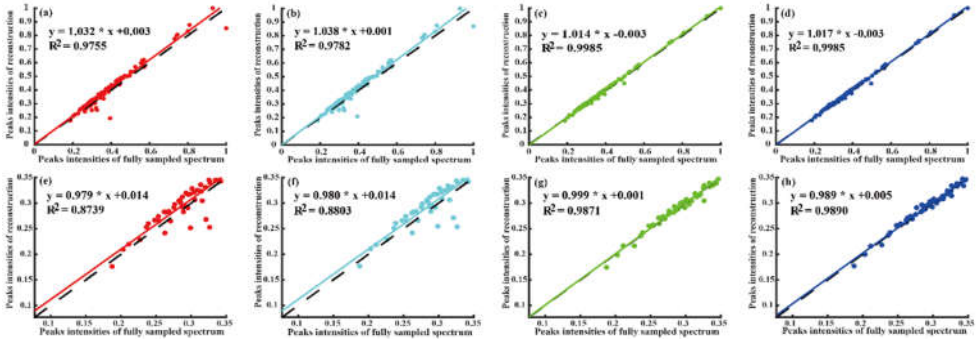

Figure 7. Peak intensity correlation between fully sampled spectrum and reconstructed spectra on the 2D HSQC of GB1. (**a**–**d**) are correlation plots of all peaks using LRHM, LRHMF, SLS, and SLSMF, respectively. (**e**–**h**) are correlation plots of low-intensity peaks using LRHM, SLS, and SLSMF, respectively.

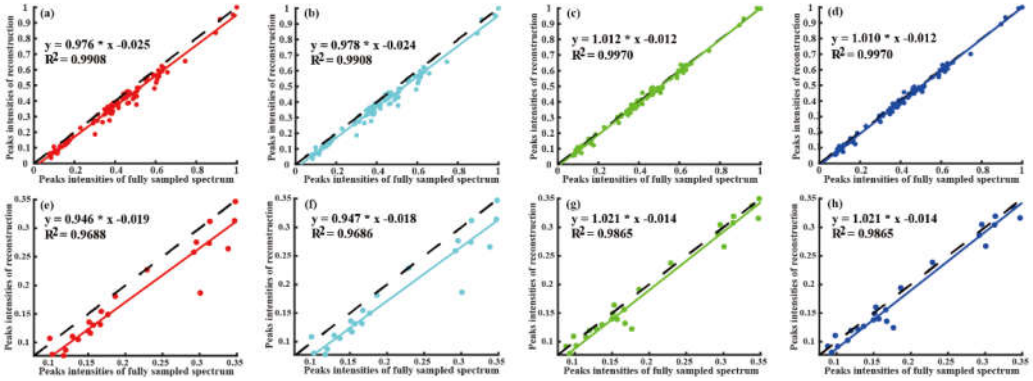

Figure 8. Peak intensity correlation between fully sampled spectrum and reconstructed spectra on the 2D ^{1}H-^{15}N best-TROSY spectrum of ubiquitin. (**a**–**d**) are correlation plots of all peaks using LRHM, LRHMF, SLS, and SLSMF, respectively. (**e**–**h**) are correlation plots of low-intensity peaks using LRHM, SLS, and SLSMF, respectively.

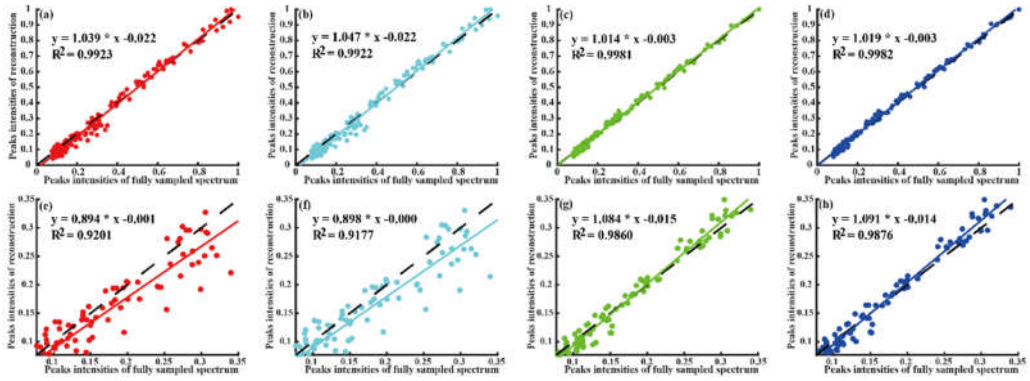

Figure 9. Peak intensity correlation between fully sampled spectrum and reconstructed spectra of the solid NMR. (**a**–**d**) are correlation plots of all peaks using LRHM, LRHMF, SLS, and SLSMF, respectively. (**e**–**h**) are correlation plots of low-intensity peak using LRHM, SLS, and SLSMF, respectively.

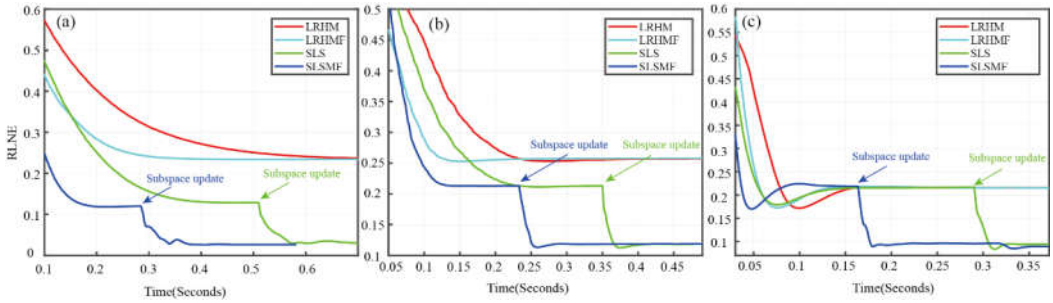

Figure 10. The RLNEs versus times of different approaches. (**a**–**c**) are the RLNE curves of the 1D experiments marked with a dotted line in Figures 3**a**, 4**a,** and 5**a**, respectively. Note: The point indicated by the arrow in the Figure is to update the outer subspace.

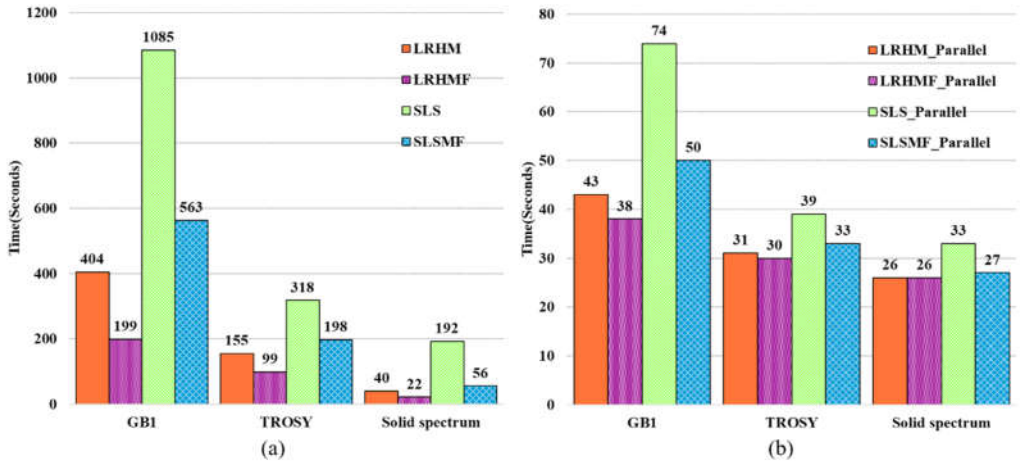

Figure 11. The computation time of spectra reconstruction on a computational server using the LRHM, LRHMF, SLS, and SLSMF methods. (**a**,**b**) computation time of spectra reconstruction without and with parallel computing, respectively. Note: The switching on and off parallel pool takes an average of 24 s.

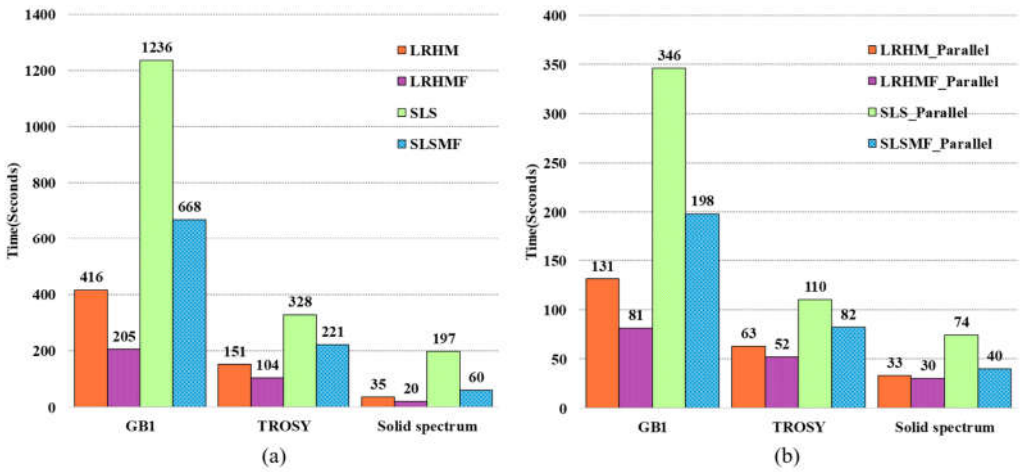

Figure 12. The computation time of spectra reconstruction on a personal computer. (**a**,**b**) computation time of spectra reconstruction without and with parallel computing, respectively. Note: All reconstruction algorithms were performed using MATLAB 2017b (Mathworks Inc., Natick, MA, USA) on a personal computer with 1.80 GHz 4 cores CPU and 8 GB RAM. The switching on and off parallel pool takes an average of 24 s.

Table 5. Performance comparison of various methods.

Method \ Quality	Peak Intensity Correlation		High-Fidelity Reconstruction Peaks	Reconstruction Time (Seconds)		Fast Reconstruction
	Low Intensity Peaks	All Peaks		Computing Server	Personal Computer	
LRHM	0.9209	0.9826	No	116.5	138.2	No
LRHMF	0.9222	0.9870	No	69.0	82.0	Yes
SLS	0.9865	0.9979	Yes	290.2	381.8	No
SLSMF	0.9877	0.9979	Yes	154.5	211.5	Neutral

Note: R^2 denotes the square of the Pearson correlation coefficient. The correlation of low-intensity peaks and all peaks are averaged on all spectra in Figures 7–9. The reconstruction time are averaged

on all spectra in Figures 11 and 12. The Pearson correlation coefficients of low-intensity peaks and all peaks are greater than 0.95 and 0.99, respectively, indicating high-fidelity reconstruction.

5. Conclusions

In this paper, a self-learning subspace matrix factorization (SLSMF) method is introduced to speed up the faithful reconstruction of highly-accelerated NMR spectroscopy. Comparisons with the state-of-the-art self-learning subspace method and the low-rank method imply that the new approach remarkably reduces computation time without sacrificing spectrum quality.

Author Contributions: Experiments and simulations, Z.T. and J.Z.; methodology, D.G.; writing—original draft preparation, Z.T., H.L. and D.G.; project administration, D.G.; conceptualization, supervision, funding acquisition, D.G.; manuscript preparation, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (61871341 and 61672335), and the Science and Technology Program of Xiamen (3502Z20183053).

Acknowledgments: The authors are grateful to Luke Arbogast and Frank Delaglio for providing the 2D HSQC spectrum of GB1 and Bingwen Hu for providing the solid NMR. Thanks to Hengfa Lu and Xiaobo Qu for valuable suggestions on this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Reference

1. Cavalli, A.; Salvatella, X.; Dobson, C.M.; Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci.* **2007**, *104*, 9615–9620.
2. Tkáč, I.; Öz, G.; Adriany, G.; Uğurbil, K.; Gruetter, R. In vivo 1H NMR spectroscopy of the human brain at high magnetic fields: Metabolite quantification at 4t vs. 7t. *Magn. Reson. Med.* **2009**, *62*, 868–879.
3. Besghini, D.; Mauri, M.; Simonutti, R. Time domain NMR in polymer science: From the laboratory to the industry. *Appl. Sci.* **2019**, *9*, 1801.
4. James, K. *Understanding NMR Spectroscopy*; Wiley: Hoboken, NJ, USA, 2005.
5. Orekhov, V.Y.; Jaravine, V.A. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 271–292.
6. Lam, F.; Li, Y.; Guo, R.; Clifford, B.; Liang, Z.-P. Ultrafast magnetic resonance spectroscopic imaging using spice with learned subspaces. *Magn. Reson. Med.* **2020**, *83*, 377–390.
7. Barna, J.C.J.; Laue, E.D.; Mayger, M.R.; Skilling, J.; Worrall, S.J.P. Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *J. Magn. Reson.* **1987**, *73*, 69–77.
8. Mobli, M.; Hoch, J.C. Nonuniform sampling and non-fourier signal processing methods in multidimensional NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *83*, 21–41.
9. Hyberts, S.G.; Takeuchi, K.; Wagner, G. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J. Am. Chem. Soc.* **2010**, *132*, 2145–2147.
10. Kazimierczuk, K.; Stanek, J.; Zawadzka-Kazimierczuk, A.; Koźmiński, W. Random sampling in multidimensional NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 420–434.
11. Jiang, B.; Jiang, X.; Xiao, N.; Zhang, X.; Jiang, L.; Mao, X.; Liu, M. Gridding and fast fourier transformation on non-uniformly sparse sampled multidimensional NMR data. *J. Magn. Reson.* **2010**, *204*, 165–168.
12. Petrellis, N. Undersampling in orthogonal frequency division multiplexing telecommunication systems. *Appl. Sci.* **2014**, *4*, 79–98.
13. Donoho, D.L. Compressed sensing. *IEEE Trans. Inform. Theory* **2006**, *52*, 1289–1306.
14. Qu, X.; Guo, D.; Cao, X.; Cai, S.; Chen, Z. Reconstruction of self-sparse 2D NMR spectra from undersampled data in the indirect dimension. *Sensors* **2011**, *11*, 8888–8909.
15. Ying, J.; Lu, H.; Wei, Q.; Cai, J.-F.; Guo, D.; Wu, J.; Chen, Z.; Qu, X. Hankel matrix nuclear norm regularized tensor completion for $ n $-dimensional exponential signals. *IEEE Trans. Signal. Process.* **2017**, *65*, 3702–3717.

16. Lu, H.; Zhang, X.; Qiu, T.; Yang, J.; Ying, J.; Guo, D.; Chen, Z.; Qu, X. Low rank enhanced matrix recovery of hybrid time and frequency data in fast magnetic resonance spectroscopy. *IEEE Trans. Bio-Med. Eng.* **2018**, *65*, 809–820.
17. Lam, F.; Liang, Z.-P. A subspace approach to high-resolution spectroscopic imaging. *Magn. Reson. Med.* **2014**, *71*, 1349–1357.
18. Kazimierczuk, K.; Orekhov, V.Y. Accelerated NMR spectroscopy by using compressed sensing. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 5556–5559.
19. Holland, D.J.; Bostock, M.J.; Gladden, L.F.; Nietlispach, D. Fast multidimensional NMR spectroscopy using compressed sensing. *Angew. Chem. Int. Ed.* **2011**, *50*, 6548–6551.
20. Goowicz, D.; Kasprzak, P.; Kazimierczuk, K. Enhancing compression level for more efficient compressed sensing and other lessons from NMR spectroscopy. *Sensors* **2020**, *20*, 1325.
21. Qu, X.; Mayzel, M.; Cai, J.-F.; Chen, Z.; Orekhov, V. Accelerated NMR spectroscopy with low-rank reconstruction. *Angew. Chem. Int. Ed.* **2014**, *54*, 852–854, doi:10.1002/anie.201409291.
22. Guo, D.; Qu, X. Improved reconstruction of low intensity magnetic resonance spectroscopy with weighted low rank Hankel matrix completion. *IEEE Access* **2018**, *6*, 4933–4940.
23. Hu, Y.; Zhang, D.; Ye, J.; Li, X.; He, X. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans. Pattern Anal.* **2013**, *35*, 2117–2130.
24. Guo, D.; Tu, Z.; Lu, H.; Qiu, T.; Xiao, M.; Qu, X. Reconstruction of Highly Accelerated NMR Spectra with Self-learning Subspace. *Submitt. Anal. Chem.* **2020**.
25. Jennings, A.; Mckeown, J.J. *Matrix Computation*; John Wiley & Sons: Hoboken, NJ, USA, 1992.
26. Guo, D.; Lu, H.; Qu, X. A fast low rank Hankel matrix factorization reconstruction method for non-uniformly sampled magnetic resonance spectroscopy. *IEEE Access* **2017**, *5*, 16033–16039.
27. Zhang, X.; Guo, D.; Huang, Y.; Chen, Y.; Wang, L.; Huang, F.; Qu, X. Image reconstruction with low-rankness and self-consistency of k-space data in parallel MRI. *Med. Image Anal.* **2020**, *63*, 101687.
28. Hoch, J.C.; Stern, A.S. *NMR Data Processing*; Wiley-Liss: New York, NY, USA, 1996.
29. Chen, D.; Wang, Z.; Guo, D.; Orekhov, V.; Qu, X. Review and prospect: Deep learning in nuclear magnetic resonance spectroscopy. *Chem.-Eur. J.* **2020**, doi:10.1002/chem.202000246.
30. Qu, X.; Huang, Y.; Lu, H.; Qiu, T.; Guo, D.; Agback, T.; Orekhov, V.; Chen, Z. Accelerated nuclear magnetic resonance spectroscopy with deep learning. *Angew. Chem. Int. Ed.* **2019**, doi:10.1002/anie.201908162.
31. Kazimierczuk, K.; Kasprzak, P. Modified omp algorithm for exponentially decaying signals. *Sensors* **2014**, *15*, 234–247.
32. Wan, S.; Mak, M.; Kung, S. Transductive learning for multi-label protein subchloroplast localization prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *14*, 212–224.
33. Srebro, N. *Learning with Matrix Factorizations*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2004.
34. Signoretto, M.; Cevher, V.; Suykens, J.A.K. An SVD-free approach to a class of structured low rank matrix optimization problems with application to system identification. *Organometallics* **2013**, *12*, 4283–4285.
35. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122.
36. Vanessa, V.; Pablo, R.; Jean-Francois, M.; Raoul, V. NMR-Mpar: A fault-tolerance approach for multi-core and many-core processors. *Appl. Sci.* **2018**, *8*, 465.
37. Mayzel, M.; Kazimierczuk, K.; Orekhov, V.Y. The causality principle in the reconstruction of sparse NMR spectra. *Chem. Commun.* **2014**, *50*, 8947–8950.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).