

# An automatic denoising method for NMR spectroscopy based on low-rank Hankel model

Tianyu Qiu, Wenjing Liao, Yihui Huang, Jinyu Wu, Di Guo, Dongbao Liu, Xin Wang, Jian-Feng Cai, Bingwen Hu, and Xiaobo Qu\*

**Abstract**—Nuclear magnetic resonance (NMR) spectroscopy, whose time domain data is modeled as the sum of damped exponential signals, has become an indispensable tool in various scenarios, such as biomedicine, biology, and chemistry. NMR spectroscopy signals, however, are usually corrupted by Gaussian noise in practice, raising difficulties in sequential analysis and quantification. The low-rank Hankel property of exponential signals plays an important role in the denoising issue, but selecting an appropriate parameter still remains a problem. In this work, we explore the effect of the regularization parameter of a convex optimization denoising method based on low-rank Hankel matrices for exponential signals corrupted by Gaussian noise. An accurate estimate on the spectral norm of weighted Hankel matrices is provided as a guidance to set the regularization parameter. The bound can be efficiently calculated since it only depends on the standard deviation of the noise and a constant. Aided by the bound, one can easily obtain an auto-setting regularization parameter to produce promising denoised results. Our results on synthetic and realistic NMR spectroscopy data demonstrate a superior denoising performance of the proposed approach over typical Cadzow and the state-of-the-art QR decomposition methods, especially in the low signal-to-noise ratio regime.

**Index Terms**—spectral denoising, NMR spectroscopy, Hankel matrix, signal reconstruction, automatic parameter.

## I. INTRODUCTION

NUCLEAR magnetic resonance (NMR) spectroscopy has grown into an essential tool for biomedical studies [1], such as the structure determination [2], metabolic analysis [3], and medical diagnosis [4]. However, NMR spectroscopy signals are often corrupted by noise during acquisition and/or transmission. The noise problem turns out to be severe in the low Signal-to-Noise Ratio (SNR) regime [5], [6]. Therefore,

This work was supported in part by National Natural Science Foundation of China (62122064, 61971361, 61871341, 61811530021, 61672335), National Key R&D Program of China (2017YFC0108703), Science and Technology Program of Xiamen (3502Z20183053), Health-Education Joint Research Project of Fujian Province (2019-WJ-31), and Xiamen University Nanqiang Outstanding Talents Program. Asterisk indicates corresponding author (Email: quxiaobo@xmu.edu.cn)

Tianyu Qiu, Yihui Huang, Jinyu Wu, Dongbao Liu, Xin Wang and Xiaobo Qu are with Biomedical Intelligent Cloud R&D Center, Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, Xiamen University, Xiamen, China.

Wenjing Liao is with School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA.

Di Guo is with School of Computer and Information Engineering, Fujian Provincial University Key Laboratory of Internet of Things Application Technology, Xiamen University of Technology, Xiamen, China.

Jian-Feng Cai is with Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China.

Bingwen Hu is with Department of Physics, East China Normal University, Shanghai, China

there is a strong demand for denoising signals, particularly in the low SNR regime.

Gaussian noise is commonly encountered in NMR spectroscopy denoising applications [7]–[10]. One of the most effective and widely adopted approaches to suppress Gaussian noise is to average multiple signal acquisitions. However, the multiple acquisitions are not always available or too costly in real applications. For this reason, effective denoising of the signals with a limited number of scans is favorable.

Numerous efforts have been made to denoise NMR spectroscopy signals. Among them, exploiting the exponential characteristic of NMR spectroscopy signals has been grown into a powerful tool [11]–[18]. Such low-rank properties were also utilized in NMR spectroscopy reconstruction [9], [14], [19]–[21], NMR spectroscopic imaging [22]–[24], and magnetic resonance imaging [25] [26], [27]. The Cadzow enhancement approach is popular in spectra denoising with the exploitation of the low-rank property of exponentials [11]–[13]. Compared with some typical denoising methods, such as the smoothing approach [28], wavelet thresholding [29], [30], Maximum entropy [31], and covariance matrix [32], [33], Cadzow method is more theoretically adopted to the denoising of all NMR spectroscopy signals. However, it is a challenging task to choose a proper number  $R$  of exponential components in practical applications, unless a priori information is given. Efforts have been made to estimate  $R$ , such as the indicator function [34] and the significance level function [35], but the estimation of  $R$  may not be satisfactory enough to yield good results [36]. Another denoising method called random QR denoising method (rQRd) is based on an approximate low-rank decomposition, and accelerates the computation by avoiding the Singular Value Decomposition (SVD) in the Cadzow method [7]. It is, however, also based on an estimation of the rank  $R$ .

This low-rank Hankel property also can be exploited in an unconstrained convex optimization method for the reconstruction issue [14], [37]. The method, known as Low-Rank Hankel Matrix reconstruction method (LRHM), also can be used for denoising, and one may receive a good result. The regularization parameter  $\lambda$  plays an important role in the results. As an example, Fig. 1 shows the denoised results with different  $\lambda$ . If  $\lambda$  is too large, the majority of the noise remains since the effect of the nuclear norm minimization is ignorable; if  $\lambda$  is too small, the spectral peaks are seriously distorted. Unfortunately, the choice of  $\lambda$  is still based on users' experience. Exploring the effect and the proper choice of  $\lambda$  is still of great demand and challenging.

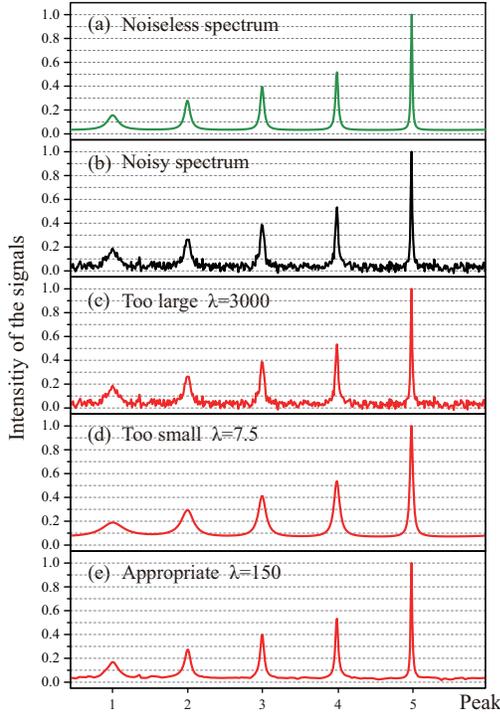


Fig. 1. A denoising example of LRHM with different choices of  $\lambda$ . (a) The noiseless spectrum. (b) The noisy spectrum with Gaussian noise ( $\sigma = 0.02$ ). (c)-(e) The denoised results with  $\lambda = 3000, 7.5$ , and  $150$ , respectively. Note: Without explicit illustration,  $\lambda$  is in the data consistency term in this paper. The model of LRHM is written as  $\min_{\mathbf{x} \in \mathbb{C}^{2N+1}} \|\mathcal{R}\mathbf{x}\|_* + \frac{\lambda}{2} \|\mathbf{y} - \mathbf{x}\|_2^2$ .

In this paper, we explore the effect of the regularization parameter, and show that a good  $\lambda$  can be automatically chosen according to the spectral norm of a weighted Hankel matrix, which is estimated by random matrix theory as a guideline for the selection of a proper  $\lambda$ . One only needs to estimate the standard deviation of the noise, which also can be set automatically, to calculate this proper  $\lambda$ . Numerical experiments on both synthetic and real NMR spectroscopy data show that noise can be effectively removed when the parameter is chosen according to our analysis.

The rest of the paper is organized as follows. Section II briefly reviews the signal model of NMR spectroscopy signals and LRHM in the denoising issue. Section III is devoted to analyzing the selection of  $\lambda$  and estimating the spectral norm of weighted Hankel matrices. Section IV contains numerical results on synthetic and real NMR spectroscopy data. Section V discusses the robustness to the estimate on the noise standard deviation, the effect of phase, the comparison with other methods, and the Denoising on  $^{13}\text{C}$  Solid-state NMR spectroscopy. Finally, we conclude and discuss future works in Section VI.

Notations used in the paper are introduced below. We denote vectors through bold lowercase letters and matrices through bold uppercase letters. The entry in vectors and matrices is denoted by a normal letter with a subscript which stands for its location. For example,  $x_n$  denotes the  $n^{\text{th}}$  element of  $\mathbf{x}$ , and  $X_{m,n}$  denotes the  $(m,n)^{\text{th}}$  entry of  $\mathbf{X}$ . For any vector  $\mathbf{x}$ ,  $\|\mathbf{x}\|_2$  represents the  $l_2$  norm. For any matrix  $\mathbf{X}$ ,  $\|\mathbf{X}\|_*$  and  $\|\mathbf{X}\|_2$

denote the nuclear norm and the spectral norm, respectively. The Hadamard product is denoted by  $\circ$ . We use superscript  $T$  and  $H$  to denote the transpose and the conjugate transpose of  $\mathbf{x}$  and  $\mathbf{X}$ . Most of operators are denoted by calligraphic letters.

## II. CONNECTION TO PRIOR WORK

In the time domain, NMR spectroscopy signal, referred to as Free Induction Decay (FID), can be expressed as the sum of  $R$  exponentials:

$$x_0(t_n) = \sum_{r=1}^R a_r e^{(j2\pi f_r - \tau_r)t_n}, \quad n = 0, \dots, 2N \quad (1)$$

where  $a_r$  denotes the signal amplitude,  $f_r$  is the central frequency, and  $\tau_r$  is the decay factor. When the number of peaks is small enough, usually  $R \leq 0.1(N+1)$ , the Hankel matrix can be treated as “low-rank” [14], [18]. In this work, we focus on the denoising of NMR spectroscopy with this property.

In practice, observations are often contaminated by noise and one receives  $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$ , where  $\mathbf{x}_0 = [x_0(t_0) \ x_0(t_1) \ \dots \ x_0(t_{2N})]^T$  is a noiseless signal and  $\mathbf{z} \in \mathbb{C}^{2N+1}$  is a random vector whose real and imaginary parts are i.i.d Gaussian with mean 0 and variance  $\sigma^2$ .

Exponential signals can be transformed into Hankel matrices with a Vandermonde decomposition. Given  $\mathbf{x}_0$ , one forms the square Hankel matrix

$$\mathcal{R}\mathbf{x}_0 = \begin{bmatrix} x_0(t_0) & x_0(t_1) & \dots & x_0(t_N) \\ x_0(t_1) & x_0(t_2) & \dots & x_0(t_{N+1}) \\ \vdots & \vdots & \ddots & \vdots \\ x_0(t_N) & x_0(t_{N+1}) & \dots & x_0(t_{2N}) \end{bmatrix},$$

where  $\mathcal{R} : \mathbb{C}^{2N+1} \rightarrow \mathbb{C}^{(N+1) \times (N+1)}$  is the operator transforming a vector to the square Hankel matrix. It is well known that  $\text{Rank}(\mathcal{R}\mathbf{x}_0) \leq R$  [38], [39].

The denoising method we explore is based on the low-rank property of  $\mathcal{R}\mathbf{x}_0$  [14], and called Convex Hankel l0-Rank matrix approximation for Denoising exponential signals (CHORD), where one solves the following optimization problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{2N+1}} \|\mathcal{R}\mathbf{x}\|_* + \frac{\lambda}{2} \|\mathbf{y} - \mathbf{x}\|_2^2, \quad (2)$$

where  $\lambda$  denotes the regularization parameter,  $\hat{\mathbf{x}}$  denotes the minimizer. The nuclear norm  $\|\cdot\|_*$  is a surrogate for the rank [40].

Alternating Direction Method of Multipliers (ADMM) [41] is a typical iterative algorithm, which can be used to solve (2). All details have been presented in the supplementary material (Section I).

The optimization problem in (2) involves a single regularization parameter  $\lambda$ , and the denoised result crucially depends on the choice of  $\lambda$ . Therefore, setting an appropriate  $\lambda$  is a crucial issue in this denoising method. This paper provides an automatic estimate on the proper choice of  $\lambda$ , and validations by experimental results.

### III. AN AUTOMATIC ESTIMATE OF THE REGULARIZATION PARAMETER $\lambda$

This section provides an estimate of the proper  $\lambda$  through establishing a relation between  $\lambda$  and the spectral norm of weighted Hankel matrices.

As  $\hat{\mathbf{x}}$  is the minimizer of (2), according to the subgradient of the nuclear norm [40], [42]–[44], the subgradient of (2) is derived as

$$\lambda(\mathbf{y} - \hat{\mathbf{x}}) = \mathcal{R}^* \left( \hat{\mathbf{U}} \hat{\mathbf{V}}^H + \hat{\mathbf{W}} \right), \quad (3)$$

where matrices  $\hat{\mathbf{U}}, \hat{\mathbf{V}} \in \mathbb{C}^{(N+1) \times (N+1)}$  are from the SVD of  $\mathcal{R}\hat{\mathbf{x}}$  such that  $\mathcal{R}\hat{\mathbf{x}} = \hat{\mathbf{U}} \hat{\mathbf{\Sigma}} \hat{\mathbf{V}}^H$ , and  $\hat{\mathbf{W}} \in \mathbb{C}^{(N+1) \times (N+1)}$  satisfies  $\hat{\mathbf{U}}^H \hat{\mathbf{W}} = \mathbf{0}$ ,  $\hat{\mathbf{W}} \hat{\mathbf{V}} = \mathbf{0}$ , and  $\|\hat{\mathbf{W}}\|_2 \leq 1$ .  $\hat{\mathbf{x}}$  should satisfies (3) for a matrix  $\hat{\mathbf{W}}$  that has the aforementioned properties.  $\mathcal{R}^* : \mathbb{C}^{(N+1) \times (N+1)} \rightarrow \mathbb{C}^{2N+1}$  is an operator transforming a matrix into vector via summing each anti-diagonal.

Denote the vector  $\mathbf{w}$  is the weights defined as  $\mathbf{w} = [1 \ 2 \ \dots \ N+1 \ \dots \ 2 \ 1]^T \in \mathbb{R}^{2N+1}$  and the symbol  $\circ$  stands for Hadamard product.

Since  $\lambda$  comes from the sub-gradient of (2), the optimal  $\lambda$  definitely satisfies (3). To obtain a specific  $\lambda$  value, it is necessary to know  $\hat{\mathbf{U}}, \hat{\mathbf{V}}$ , and  $\hat{\mathbf{W}}$ . However, according to the definition of the sub-gradient of the nuclear norm, the matrix  $\hat{\mathbf{W}}$  cannot be obtained directly, we have to use the inequality scaling and numerical experiments to obtain a proper  $\lambda$ .

Because  $\lambda \mathcal{R} \frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}})$  is an approximation of  $\hat{\mathbf{U}} \hat{\mathbf{V}}^H + \hat{\mathbf{W}}$ ,  $\|\lambda \mathcal{R} \frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}})\|_F \leq \|\hat{\mathbf{U}} \hat{\mathbf{V}}^H + \hat{\mathbf{W}}\|_F$  [11]. This relationship approximately holds in the spectral norm (See the details in the Supplement Section II). Therefore, the proper  $\lambda$  is chosen as below

$$\lambda \leq \frac{1}{\|\mathbf{Z} + \tilde{\mathbf{X}}\|_2} \leq \frac{1}{\left( \|\mathbf{Z}\|_2 - \|\tilde{\mathbf{X}}\|_2 \right)}, \quad (4)$$

where  $\mathbf{Z} = (\mathcal{R} \frac{1}{\mathbf{w}}) \circ \mathcal{R}\mathbf{z}$  denotes a weighted Hankel matrix such that

$$\mathbf{Z} = \left( \mathcal{R} \frac{1}{\mathbf{w}} \right) \circ \mathcal{R}\mathbf{z} = \begin{pmatrix} z_1 & \frac{z_2}{2} & \dots & \frac{z_{N+1}}{N+1} \\ \frac{z_2}{2} & \frac{z_3}{3} & \dots & \frac{z_{N+2}}{N} \\ \vdots & \vdots & \dots & \vdots \\ \frac{z_{N+1}}{N+1} & \frac{z_{N+2}}{N} & \dots & z_{2N+1} \end{pmatrix}, \quad (5)$$

and  $\tilde{\mathbf{X}}$  denotes  $\tilde{\mathbf{X}} = (\mathcal{R} \frac{1}{\mathbf{w}}) \circ \mathcal{R}(\mathbf{x}_0 - \hat{\mathbf{x}})$ .

In order to explore the relationship among the spectral norm of weighted Hankel matrices, the noise level, and the size of the matrix, we did sufficient Monte Carlo trials on synthetic data and Gaussian noise. Results in Fig. 2 and Fig. 3 show that the empirical means of  $\|\mathbf{Z}\|_2$  and  $\|\tilde{\mathbf{X}}\|_2$  are almost independent of  $N$ . Furthermore, these empirical means increase as the increasing of the standard deviation  $\sigma$  of the noise.

In applications, it is expected that signal details can be preserved as much as possible, thus we propose to select the regularization parameter as

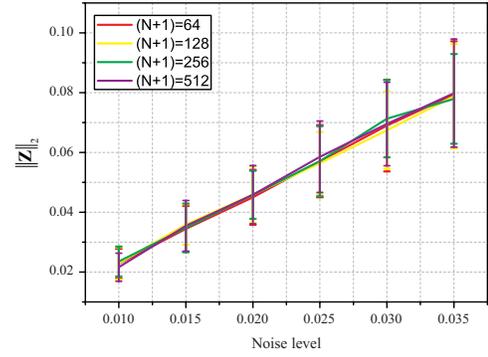


Fig. 2. The relation between  $\|\mathbf{Z}\|_2$  and the standard deviation  $\sigma$  of the Gaussian noise  $\mathbf{z}$  in 100 Monte Carlo trials. The Matrix  $\mathbf{Z}$  is of size  $(N+1) \times (N+1)$  with  $(N+1) = 64, 128, 256, 512$ , respectively. The curve represents the mean of  $\|\mathbf{Z}\|_2$  in 100 trails versus  $\sigma$ , and the standard deviation of  $\|\mathbf{Z}\|_2$  in 100 trails is indicated by the vertical bar.

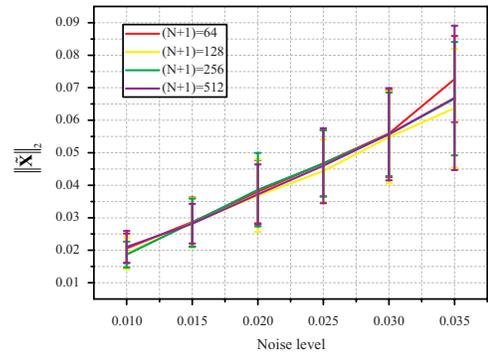


Fig. 3. The relation between  $\|\tilde{\mathbf{X}}\|_2$  and the standard deviation  $\sigma$  of the Gaussian noise  $\mathbf{z}$  in 50 Monte Carlo trials. The Matrix  $\tilde{\mathbf{X}}$  is of size  $(N+1) \times (N+1)$  with  $(N+1) = 64, 128, 256, 512$ , respectively. The curve represents the mean of  $\|\tilde{\mathbf{X}}\|_2$  in 50 trails versus  $\sigma$ , and the standard deviation of  $\|\tilde{\mathbf{X}}\|_2$  in 50 trails is indicated by the vertical bar. Note:  $\mathbf{x}_0$  are damped exponential signals with random  $a_r, f_r$  and  $\tau_r$ .  $\hat{\mathbf{x}}$  is obtained from CHORD. The parameter  $\lambda$  is chosen such that the Normalized Root-Mean-Square Error (NRMSE) is minimized.

$$\lambda^* = \frac{1}{\left| \mathbb{E} \|\mathbf{Z}\|_2 - \mathbb{E} \|\tilde{\mathbf{X}}\|_2 \right|}, \quad (6)$$

where the symbol  $\mathbb{E}$  denotes the expectation.

In order to provide a proper choice of  $\lambda$ , we estimate an upper and lower bound of  $\mathbb{E} \|\mathbf{Z}\|_2$ . With respect to  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$ , we provide an empirical value based on sufficient numerical experiments on synthetic data.

#### A. The bounds of $\mathbb{E} \|\mathbf{Z}\|_2$

Actually, for estimating bounds of the spectral norm of Hankel matrices given by random vectors, numerical achievements have been made [45]–[48]. In this subsection, we focus on estimating bounds of the spectral norm of weighted Hankel matrices. Theorem 1 and 2 provide a lower and upper bounds of  $\mathbb{E} \|\mathbf{Z}\|_2$ , respectively. All details of proofs and the asymptotic analysis have been presented in the supplementary material (Sections III–V).

**Theorem 1.** Suppose the real and imaginary parts of the entries in  $\mathbf{z} \in \mathbb{C}^{2N+1}$  are i.i.d. Gaussian random variables with mean 0 and variance  $\sigma^2$ . Define  $R_N$  and  $Q_N$  such that

$$R_N^2 = \sum_{k=0}^{2N} |d_k|^2 \text{ and } Q_N^4 = \sum_{k=0}^{2N} |d_k|^4, \quad (7)$$

$$\text{where } d_k = \begin{cases} \frac{2}{(k+1)(k+2)} \sum_{m=0}^k \frac{1}{m+1}, & 0 \leq k \leq N \\ \frac{2}{(2N-k+1)(k+2)} \sum_{m=k}^{2N} \frac{1}{m-N+1}, & N < k \leq 2N \end{cases}.$$

Then there exists a constant  $C$  such that the matrix  $\mathbf{Z}$  defined in (4) satisfies

$$\mathbb{E} \|\mathbf{Z}\|_2 \geq \sigma \frac{C(N+1)}{2N+1} \sqrt{R_N^2 \left(1 + \log \frac{R_N^4}{Q_N^4}\right)}. \quad (8)$$

**Theorem 2.** Suppose the real and imaginary parts of the entries in  $\mathbf{z} \in \mathbb{C}^{2N+1}$  are i.i.d. Gaussian random variables with mean 0 and variance  $\sigma^2$ . Then

$$\mathbb{E} \|\mathbf{Z}\|_2 \leq \sigma \sqrt{2C_{\mathbf{w}} \log(2N+2)}, \quad (9)$$

where  $C_{\mathbf{w}} = \max(\sum_{k=0}^N w_k^{-2}, \sum_{k=1}^{N+1} w_k^{-2}, \dots, \sum_{k=N}^{2N} w_k^{-2})$  with the vector  $\mathbf{w}$  defined in (5).

Two theorems above provide the following upper and lower bounds of  $\mathbb{E} \|\mathbf{Z}\|_2$ :

$$\sigma \frac{C(N+1)}{2N+1} \sqrt{R_N^2 \left(1 + \log \frac{R_N^4}{Q_N^4}\right)} \leq \mathbb{E} \|\mathbf{Z}\|_2 \leq \sigma \sqrt{2C_{\mathbf{w}} \log(2N+2)}. \quad (10)$$

The upper bound scales as  $\sigma \sqrt{\log N}$ , while the lower bound depends on  $R_N$  and  $Q_N$ . When  $N$  is large enough, the upper bound and the lower bound only differ by a factor of  $\sqrt{\log N}$ .

Choosing the lower bound of  $\mathbb{E} \|\mathbf{Z}\|_2$  tends to obtain a relatively large  $\lambda$ , which is beneficial to preserve more signal details. Therefore, we suggest to choose  $\mathbb{E} \|\mathbf{Z}\|_2$  as

$$\mathbb{E} \|\mathbf{Z}\|_2 = \frac{C(N+1)}{(2N+1)} \sqrt{R_N^2 \left(1 + \log \frac{R_N^4}{Q_N^4}\right)} \sigma. \quad (11)$$

We next find the empirical constant  $C$  through repetitive experiments on synthetic data. According to Theorem 1,  $C > 0$  is a constant, which is independent of the signal length and the standard deviation  $\sigma$ .

The slope values of lines in Fig. 2 are estimated by the least square method. According to the conclusion in (9),  $C$  values are obtained via dividing slopes by  $\frac{(N+1)}{(2N+1)} \sqrt{R_N^2 \left(1 + \log \frac{R_N^4}{Q_N^4}\right)}$ . Finally, after averaging four  $C$  values, we suggest that  $C = 2.9$  for denoising. Results in Fig. 4 confirms that the conclusion in Theorem 1 is well capable of estimating  $\mathbb{E} \|\mathbf{Z}\|_2$ .

### B. The empirical $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$

This subsection is devoted to an empirical estimate of  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$ . We perform experiments with different  $N$ ,  $\sigma$ ,  $\lambda$ ,

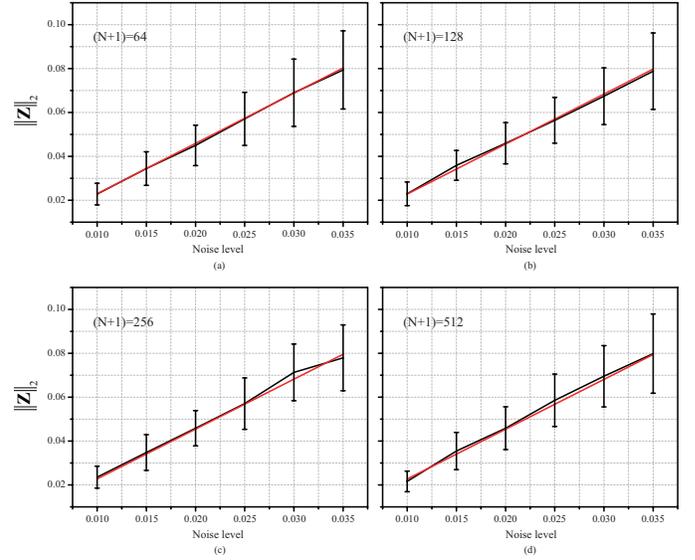


Fig. 4.  $\mathbb{E} \|\mathbf{Z}\|_2$  and the lower bound with the suggested  $C$  under different matrix sizes. The Matrix  $\mathbf{Z}$  is of size  $(N+1) \times (N+1)$  with  $(N+1) = 64$ (a), 128(b), 256(c), and 512(d), respectively. The vertical axis denotes the value of  $\|\mathbf{Z}\|_2$  and the horizontal axis denotes the standard deviation of Gaussian noise. The black curves stand for the empirical mean of  $\|\mathbf{Z}\|_2$  in Fig. 2. Red lines denote the lower bounds with  $C = 2.9$ . The error bars denote the standard deviation of  $\|\mathbf{Z}\|_2$  values in 100 Monte Carlo trials.

signals and noises in order to determine a proper empirical estimate value.

Before evaluating the denoising performance, we first introduce two objective criteria, Normalized Root-Mean-Square Error (NRMSE) [49] and Mean Absolute Error (MAE) [50].

$$\text{NRMSE} = \frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2}{\|\mathbf{x}_0\|_2}, \quad (12)$$

where  $\hat{\mathbf{x}}$  and  $\mathbf{x}_0$  are the denoised signal and the noiseless signal respectively.

$$\text{MAE} = \frac{\|\mathbf{f}_s - \mathbf{f}_0\|_1}{2N+1}, \quad (13)$$

where  $\mathbf{f}_s \in \mathbb{R}^{(2N+1) \times 1}$  and  $\mathbf{f}_0 \in \mathbb{R}^{(2N+1) \times 1}$  denote the real part of the noisy spectrum and the noiseless spectrum, respectively.

We generate a synthetic data set, including 90 random damping complex exponential signals with  $2N+1 = 255$ , 511, and 1023 respectively, and repeat 100 Monte Carlo trials to incorporate the randomness of Gaussian noise. Each signal in the data set has  $3R+1$  parameters, including  $R$ ,  $a_r$ ,  $f_r$  and  $\tau_r$ , where  $r = 1, 2, \dots, R$ . The number of exponential components is  $R = 4 + M_r$ , where  $M_r$  denotes a pseudo-random scalar integer of range  $[1, 9]$ . The amplitude  $a_r$  is uniformly sampled from  $(0, 10)$ . Each frequency  $f_r$  is uniformly sampled from  $(0, 1)$ . The damping factor is  $\tau_r = 5 + 60m_r$ , where  $m_r$  is uniformly sampled from  $(0, 1)$ .

Then, we use a series of  $\lambda$  to denoise signals in the data set above, find the optimal solution  $\hat{\mathbf{x}}$  corresponded to the lowest error, NRMSE, and calculate  $\|\tilde{\mathbf{X}}\|_2$ . 9 signals with different data lengths are randomly selected and the corresponding  $\|\tilde{\mathbf{X}}\|_2$  are presented in Fig. 5.

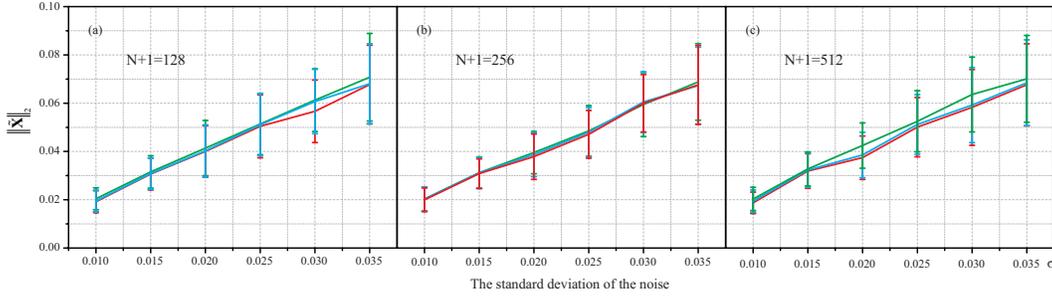


Fig. 5. The relation between  $\|\tilde{\mathbf{X}}\|_2$  given by random synthetic signals and noise levels. (a)-(c) show the average spectral norm of  $\tilde{\mathbf{X}}$  which is of size  $(N+1) \times (N+1)$  with  $(N+1) = 128$ (a), 256(b), 512(c), respectively. The vertical error bars represent the standard deviation of  $\|\tilde{\mathbf{X}}\|_2$  values in 100 Monte Carlo trials. Note: For each sub-plot, the green, blue, and red lines denote  $\|\tilde{\mathbf{X}}\|_2$  given by different random signals.

Results in Fig. 5 indicate that the standard deviation of noise is the main factor that determines  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$ . Moreover, this empirical mean of  $\|\tilde{\mathbf{X}}\|_2$  seems to be linear to the standard deviation of the noise. Additionally, the spectral parameters and the distribution they satisfy also slightly affects  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$  (See the details in the supplementary material Section IX). We estimate the slope on Matlab platform (2017b) and suggest  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2 = 1.94\sigma$  for denoising.

#### IV. NUMERICAL EXPERIMENTS

In this section, we evaluate the performance of CHORD with the suggested  $\lambda$  on the synthetic data and a realistic NMR spectroscopy data set. No preprocessing, including the phase correction and apodisation, was applied to the synthetic data. For the data acquired from real NMR spectrometer, we merely truncated the FID signal and only used the first 1000 points. For visualization, real-part spectra are presented and imaginary-part spectra are discarded, but all the processing are on complex data. In addition, for avoiding the bias, the proposed method also have been tested on random simulated data, whose detailed results are presented in the supplementary material (Section X).

The typical method, Cadzow [7], [13], and the state-of-the-art method, rQRd [7] are compared with our proposed method. For Cadzow, its key parameter is the rank of this Hankel matrix. For rQRd, its primary parameter is the number of the matrix  $\mathbf{Q}$ 's column, denoted as  $\text{rank}_Q$ , in QR decomposition. For the rest of the manuscript, without explicit illustration, the main parameters in Cadzow and rQRd are chosen to be the ones yielding the lowest reconstruction error, NRMSE.

##### A. Denoising of synthetic complex data

We generated a synthetic exponential complex data with five peaks (presented in Fig. 1(a)). In the following, the synthetic data indicates the signal in Fig. 1(a). The denoising performance of three methods is tested through recovering the signal from complex Gaussian noise with different standard deviation ( $\sigma = 0.01, 0.02, 0.03, 0.04, 0.05$ , and  $0.06$ , respectively). 100 Monte Carlo trials are done to avoid the randomness of noise.

In practice, we do not know in advance the standard deviation of the noise that corrupts the signal of interest. Here, we truncate signal from the end of noisy FID to estimate the standard deviation of the noise to mimic the real cases. The truncated length is verified by Kolmogorov-Smirnov (KS) test [51]. Also, we compare the denoising performances of CHORD given the known standard deviation and the estimated standard deviation. For clarity, we name the CHORD using the known standard deviation  $\text{CHORD}_{\text{Prior}}$  and the CHORD using the estimated standard deviation  $\text{CHORD}_{\text{Esti}}$ , respectively.

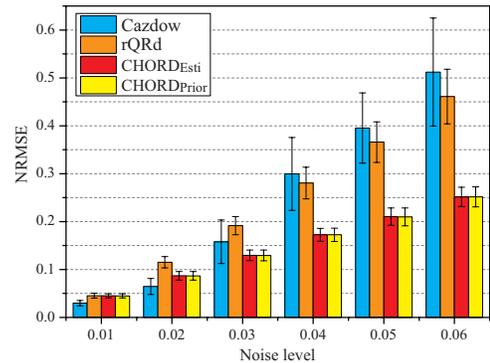


Fig. 6. The reconstruction error, NRMSE, for synthetic data (Fig. 1(a)) under different noise levels.  $\text{CHORD}_{\text{Esti}}$  and  $\text{CHORD}_{\text{Prior}}$  denote denoised results of CHORD with estimated standard deviation and the known standard deviation, respectively. Cadzow and rQRd present the optimal (minimal NRMSE) denoised results, respectively. The height of columns shows the average of the NRMSEs over 100 trials. The vertical bar comes from the randomness of noise.

Fig. 6 shows the denoising performance under different noise levels. Under relatively weak noise ( $\sigma \leq 0.02$ ), Cadzow achieves the lowest NRMSE compared to other approaches. Under relatively high noise ( $\sigma \geq 0.05$ ), however, the NRMSE of Cadzow increase faster than that of rQRd and, particularly, CHORD, implying Cadzow is not robust to relatively high noise levels. The proposed method produces the lowest NRMSE when the noise is higher than 0.03 and produces smallest variances. Furthermore, the results of  $\text{CHORD}_{\text{Esti}}$  are very close to that of  $\text{CHORD}_{\text{Prior}}$ , showing the feasibility of CHORD. In the following, without explicit illustration, the mentioned CHORD is  $\text{CHORD}_{\text{Esti}}$ .

We evaluate the effect of parameters selection of the tested

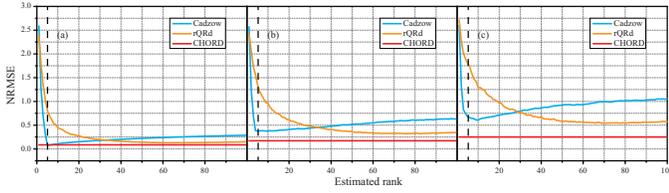


Fig. 7. The average NRMSE of denoised results of the synthetic data (in Fig. 1(a)) with different estimated ranks over 50 Mont Carlo trials. (a)-(c) denote the average NRMSE of denoised results with  $\sigma = 0.02, 0.04,$  and  $0.06,$  respectively. The black dash lines stand for the exact rank of the synthetic data (rank=5). Note: For rQRd, the estimated rank stands for rank $_Q$ .

approaches in Fig. 7. For Cadzow, when the noise is weak (Fig. 7(a)), an accurate estimate leads to a good result. But as the noise gets stronger, the optimal estimated rank (in terms of NRMSE) may be not equivalent to the actual rank (Fig. 7(c)), meaning that if the noise level is strong enough, an accurate estimated rank will not significantly improve denoised results. Compared with Cadzow, rQRd owns a more flexible parameter setting, but the average NRMSE of its denoised results is always higher than that of CHORD under large noise.

Fig. 8 presents the representative denoised results of the synthetic signal corrupted by strong noise. Typical denoised spectra of Cadzow and rQRd with three different parameters selection are presented. Cadzow tends to remove small peaks if using a much smaller estimated rank (see Fig.8(i)). And if the estimated rank is close to or larger than the real rank, Cadzow spectra introduce spectral distortions and distinct artifacts (see Fig.8(e) and (m)). For rQRd, a small rank $_Q$  leads to a smooth spectrum but with missed or weakened low-intensity peaks (see Fig.8(k)), while larger parameters introduce strong noise (see Fig.8(g) and (o)). For the CHORD, it provides a relatively reasonable denoised result using the suggested  $\lambda$  and the estimated noise level.

### B. Denoising of NMR spectroscopy data

NMR spectroscopy, as a non-invasive technology, has been widely utilized in the study of chemistry, biology, and medicine, such as the diagnosis of diseases [22]. One of the reasons that limits the widespread of this technology is its relatively low SNR. Therefore, CHORD is evaluated on the denoising of a real NMR spectroscopy data. We acquired the signal with high SNR as the reference and added the Gaussian noise retrospectively.

In applications, the unit of chemical shift is usually expressed in part per million (ppm) instead of the Hz, avoiding the ambiguity when spectrometers are at different magnet strengths. The definition of chemical shift is given by

$$\text{chemical shift(ppm)} = \frac{f_{test} - f_{ref}}{f_{spec}} \times 10^6, \quad (14)$$

where  $f_{test}$  denotes the resonance frequency of the sample,  $f_{ref}$  the absolute resonance frequency of a standard compound measured in the same magnetic field, and  $f_{spec}$  the frequency of the magnetic field strength of spectrometers.

The real data is a 1D  $^1\text{H}$  NMR spectrum that was acquired at 298 K on a Varian 500 MHz (11.7T) magnetic resonance sys-

tem (Agilent Technologies, Santa Clara, CA, USA) equipped with a 5 mm indirect detection probe. A  $8.3\mu\text{s}$  single pulse sequence was used, and 64 scans were acquired. The total acquisition time took 286.7s. The concentration of creatine, choline, magnesium citrate and calcium citrate are 0.03g/mL, 0.03g/mL, 0.06g/mL, and 0.06g/mL, respectively.

The denoised results of the metabolic spectrum are presented in Fig. 9, which supports the conclusion made on the synthetic data. Under a relatively strong noise level ( $\sigma = 0.035$ ), Cadzow smooths the spectrum, which, on the one side, offers a nice noise denoised results, on the other side, however, leads to the missing of some peaks (such as the peaks at 6.8 ppm). rQRd provides a spectrum with obvious noise (orange lines in Fig. 9(c)), and weakens low-intensity peaks (such as the peaks at 6.8 ppm). CHORD is capable of effectively removing noise and keeping more details of peaks (see Fig. 9(c)). For the high SNR scenario, all the three methods produce nice and comparable denoised results (see Fig. 9(a)).

Experiments on synthetic complex exponential and realistic NMR spectroscopy data demonstrate that CHORD with the auto-setting parameter achieves more robust and accurate results compared with Cadzow and rQRd method.

## V. DISCUSSIONS

### A. The estimate of noise

We estimate the noise level by calculating the standard deviation of signals truncated from the end of signals on Matlab platform (2017b). To ensure that the truncated signals satisfy the Gaussian distribution, KS test is introduced into the noise estimate. Details of  $p$  value test and the flowchart have been shown in the Supplement (Section VI).

The  $p$  values of different truncated lengths under various noise levels are tested on the simulated data (Fig. 10). 100 Monte Carlo trials have been done to avoid the randomness of noise.

The results in Fig. 10 illustrate that the reduction of  $p$  values is caused by the truncation of FID. In the areas without noise, such as the last 200 points in Fig. 10(a), the average of  $p$  values approximates 0.8. As the truncated length increases, a part of the signal is treated as noise, resulting in the apparent descent of  $p$  values. Additionally, a low noise level is beneficial to distinguish the noise and the ground truth, which is reflected in the earlier decrease of  $p$  values. Furthermore,  $p$  values fluctuate at 0.8 for truncated signals without FID. Therefore, we selected 0.8 as the threshold value.

### B. The effect of phase on $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$

Since the effect of phase can be alleviated by the correction in the preprocessing, in the above, we temporarily omitted the phase during the estimate of  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  to simplify the problem. This subsection is devoted to discuss the empirical value of  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  under two common situations where the signal  $\mathbf{x}_0$  contains the zero-order and relative phase.

Both the zero-order phase  $\theta$  and the relative phase  $\theta_r$  are uniformly random and sampled from  $(0, 1)$ . These phases are added for the experiments as:

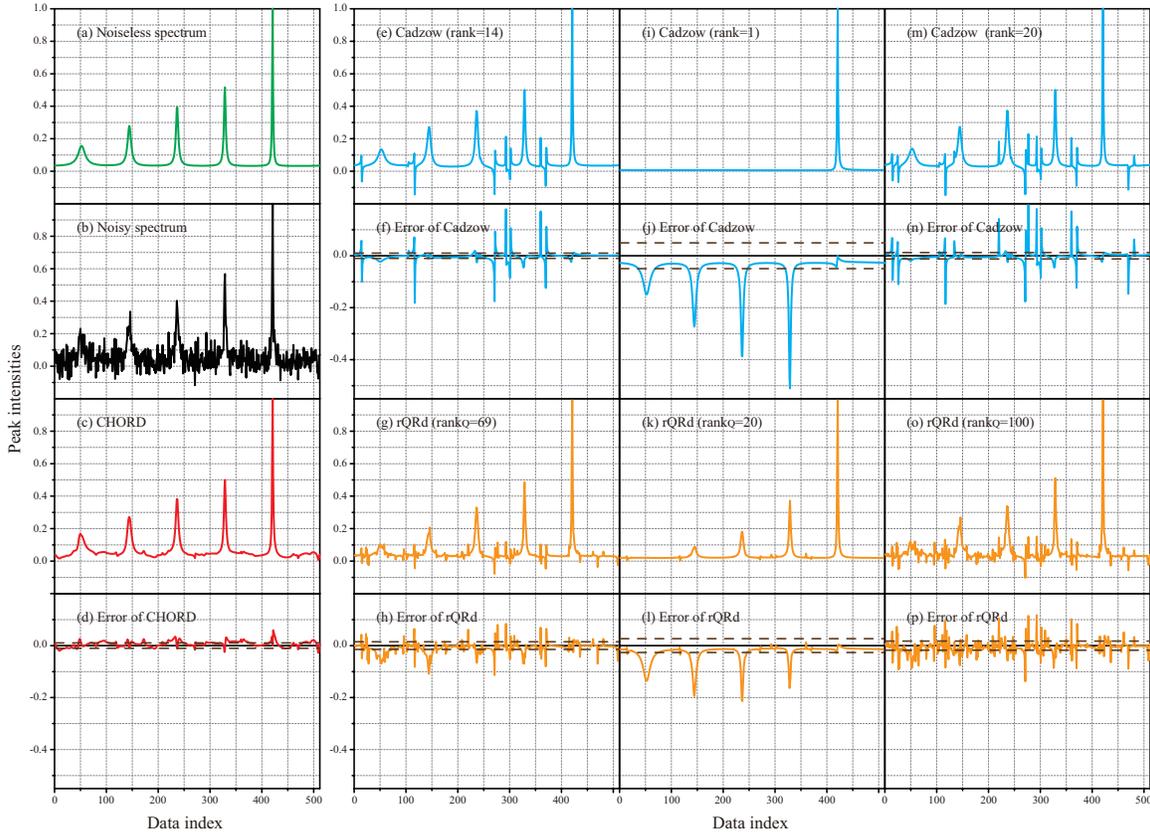


Fig. 8. The denoised results by typical methods. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.04$ ) respectively. (c)-(d) are the denoised spectra and errors of CHORD with the suggested parameter. (e)-(f), (i)-(j), and (m)-(n) show denoised spectra and errors of Cadzow with three different estimated ranks (optimal in terms of NRMSE, small, and large). (g)-(h), (k)-(l), and (o)-(p) are denoised results of rQRd with three different estimated ranks (optimal in terms of NRMSE, small, and large). The brown long dash lines denote the MAEs of the denoised spectra.

(Zero-order phase)

$$x_0(t_n) = e^{j2\pi\theta} \sum_{r=1}^R a_r e^{(j2\pi f_r - \tau_r)t_n}, n = 0, \dots, 2N, \quad (15)$$

and

(Relative phase)

$$x_0(t_n) = \sum_{r=1}^R a_r e^{j2\pi\theta_r} e^{(j2\pi f_r - \tau_r)t_n}, n = 0, \dots, 2N. \quad (16)$$

According to results of the Monte Carlo simulations (Fig. 11 and Fig. 13), it is observed that  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  tends to be proportional to the noise level. Utilizing the same technique in Section III-B, the empirical relation between  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  and the standard deviation of the noise can be acquired. Compared with the suggested value, the zero-order and relative phase slightly reduce  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$ . But this change causes little impact on the denoised spectra (Fig. 12 and 14).

The slopes of lines corresponding to all synthetic data with the zero-order phase were estimated by the least square method, and the averaged slope is 1.84. Compared with the suggested value (1.94), the relative error is 5%. For synthetic data with the relative phase, the same technique is utilized to

estimate the slope, and the average slope is 1.87. Compared with the suggested value, the relative error is 4%. In the supplementary material (Section VII), signals with 256 and 1024 data points are shown to complement the discussion.

### C. Comparison with other representative methods

This subsection respectively provides the comparison between the proposed method and two representative methods, a sparse regularization-based method, Compressed Sensing (CS) [52], and adaptive regularization parameters selection method with Discrepancy Principle (DP) [53]. We verify these methods on the denoising of the synthetic and the experimental NMR spectra. The synthetic denoised spectra and NRMSE are shown in the supplementary material (Section VIII).

1) *Comparison with CS*: The CS assumes the sparsity of NMR spectroscopy in the frequency domain. For the denoised spectra, CHORD better removes noise and retains peak details (such as the peak at 6.8 ppm in Fig. 15) than CS.

2) *Comparison with DP-based method*: The Discrepancy Principle (DP) tries to find an optimal regularization parameter so that the norm of denoising error is equal to that of the noise [53]. DP has been used in Tikhonov regularization [53], Total Variation (TV) [54], and Low-Rank reconstruction [55].

We implemented a DP-based method (See Fig. 16) to select a  $\lambda^*$  to satisfy

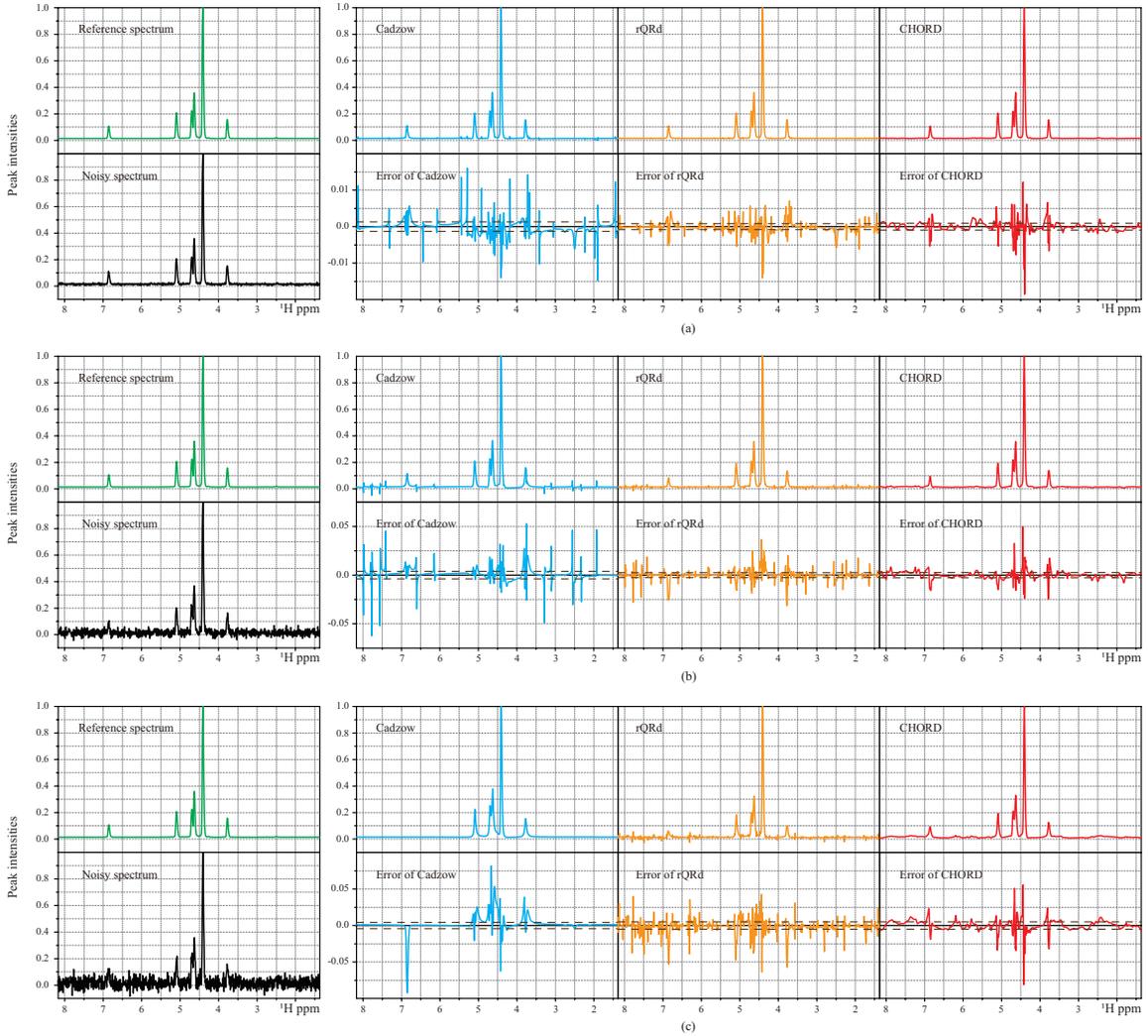


Fig. 9. Denoised results of realistic metabolite spectrum with  $\sigma = 0.035$  (a),  $0.020$  (b), and  $0.005$  (c), respectively. The green lines denote the reference spectrum with high SNR. The black lines indicate noisy spectra. The blue, orange and red line are denoised results of Cadzow, rQRd and CHORD, respectively. Note: The results of Cadzow and rQRd that enable the lowest NRMSE are presented here. The brown long dash lines denote the MAEs of the denoised spectra.

$$\|\hat{\mathbf{x}}(\lambda^*) - \mathbf{y}\|_2 = C\sigma, \quad (17)$$

where  $\sigma$  stands for the standard deviation of real/imaginary part of noise.  $C$  is a constant.

Spectra in Figs. 17 show that both the proposed method and DP can remove the noise well while the former obtains slightly lower NRMSE than the latter (See the supplementary material). However, the convergence of DP has not been proven in LRHM, which is worthy of further investigation in the future.

#### D. The denoising on $^{13}\text{C}$ Solid-state NMR spectroscopy

To evaluate the denoising performance under a more realistic scenario, we tested the performance of the proposed method (Fig. 18) on experimental  $^{13}\text{C}$  solid-state NMR spectra which were acquired with varying levels of average. Solid-state NMR spectroscopy has grown into a versatile tool to

analyze materials in the solid state but is limited by relatively low SNR due to its poor sensitivity [36].

The experimental  $^{13}\text{C}$  solid-state NMR data is a decoupled static CSA spectrum with the sample of glycine and was acquired on a Bruker 600MHz spectrometer (14.1T) equipped with an AVANCE-III console. A commercial Bruker HX double-resonance MAS probe with a 4 mm outer diameter rotor was used in the static for the experiments.

Compared with the reference spectrum (average of 200 scans), CHORD with the automatic parameter effectively remove noise and preserve the details of the spectrum (Red lines in Fig. 18(c) and (d)), saving more than 80% acquisition time. For Cadzow, its optimal denoised result (in terms of NRMSE) over-smooths the spectrum and generates some fake peaks (black arrows in Fig. 18(e)). Reducing the estimate of the rank suppresses the fake peak, but causes a more serious loss of signal details. A large rank keeps more details but leads to more fake peaks. For rQRd, its optimal result remains too much residue, a smaller estimate of rank results in a smoother

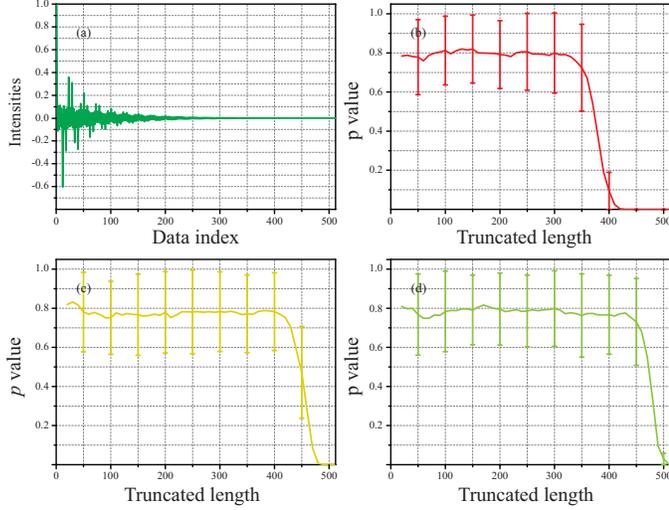


Fig. 10. The  $p$  values with different truncated lengths under various noise levels. (a) denotes the synthetic FID signal in Fig. 8(a). (b)-(d) are  $p$  values of (a) under noise with  $\sigma = 0.01, 0.03,$  and  $0.05,$  respectively.

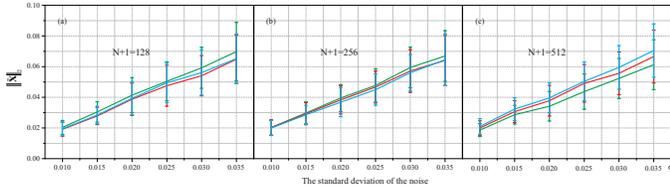


Fig. 11. The relation between  $\|\tilde{\mathbf{X}}\|_2$  given by random synthetic signals with zero-order phases and noise levels. (a)-(c) show the average spectral norm of  $\tilde{\mathbf{X}}$  which is of size  $(N+1) \times (N+1)$  with  $(N+1) = 128, 256, 512,$  respectively. The vertical error bars come from 100 Monte Carlo trials. Note: For each sub-plot, the green, blue, and red lines denote  $\|\tilde{\mathbf{X}}\|_2$  given by random signals with different spectral parameters.

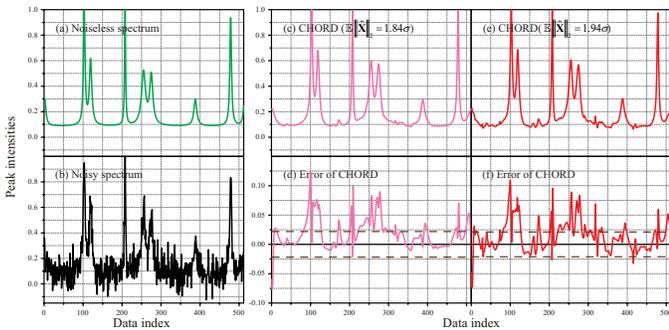


Fig. 12. The comparison of 512-point denoised spectra with the zero-order phase under the new and the suggested  $\mathbb{E}\|\tilde{\mathbf{X}}\|_2$  values. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.02$ ). (c) and (e) are the denoised spectra of CHORD with the new value ( $1.84\sigma$ ) and the suggested value ( $1.94\sigma$ ), respectively. (d) and (f) stand for the denoising errors of spectra that correspond to (c) and (e), respectively. The brown long dash lines denote the MAEs of the denoised spectra.

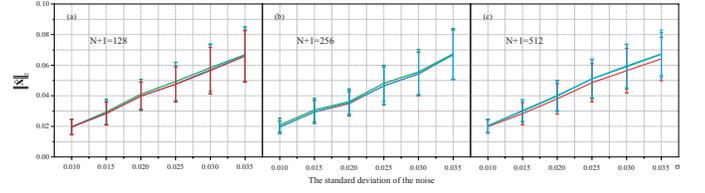


Fig. 13. The relation between  $\|\tilde{\mathbf{X}}\|_2$  given by random synthetic signals with the relative phase. (a)-(c) show the average spectral norm of  $\tilde{\mathbf{X}}$  which is of size  $(N+1) \times (N+1)$  with  $(N+1) = 128, 256, 512,$  respectively. The vertical error bars come from 100 Monte Carlo trials. Note: For each sub-plot, the green, blue, and red lines denote  $\|\tilde{\mathbf{X}}\|_2$  given by random signals with different spectral parameters.

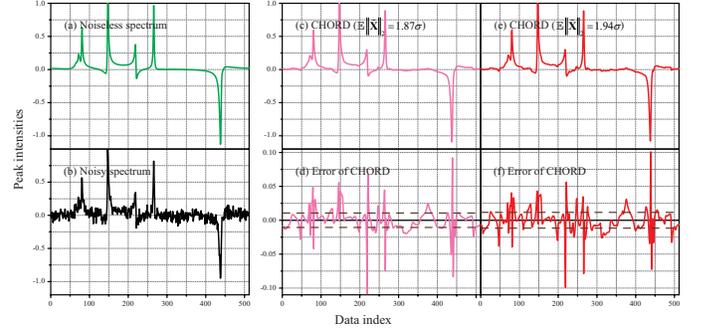


Fig. 14. The comparison of 512-point denoised spectra with the relative phase under the new and the suggested  $\mathbb{E}\|\tilde{\mathbf{X}}\|_2$  values. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.03$ ). (c) and (e) are the denoised spectra of CHORD with the new value ( $1.87\sigma$ ) and the suggested value ( $1.94\sigma$ ), respectively. (d) and (f) stand for the denoising errors of spectra that correspond to (c) and (e), respectively. The brown long dash lines denote the MAEs of the denoised spectra.

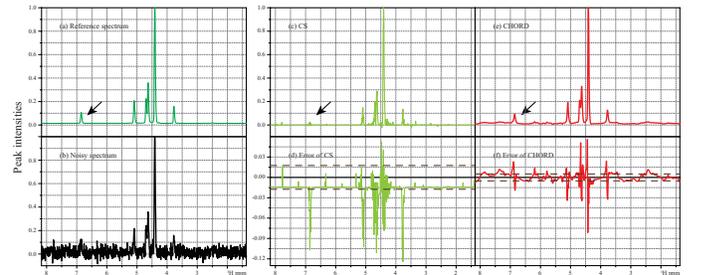


Fig. 15. The denoising of realistic metabolite spectrum by CS and CHORD. (a) denotes the reference spectrum with high SNR. (b) denotes the noisy spectrum ( $\sigma = 0.035$ ). (c) and (e) are the denoised results by CS and CHORD, respectively. (d) and (f) stand for the denoising errors of two methods, respectively. The brown long dash lines denote the MAEs of the denoised spectra.

spectrum with the loss of signal details.

## VI. CONCLUSION

Based on CHORD, a denoising method based on low-rank Hankel property of complex exponential signals, we attempt to figure out the bound of the regularization parameter, determine the empirical optimal constant, and estimate the standard derivation of the noise so that the users are able to apply CHORD with an auto-setting parameter. Experiments on synthetic complex exponential and realistic NMR spectroscopy data demonstrate that CHORD with the auto-setting parameter

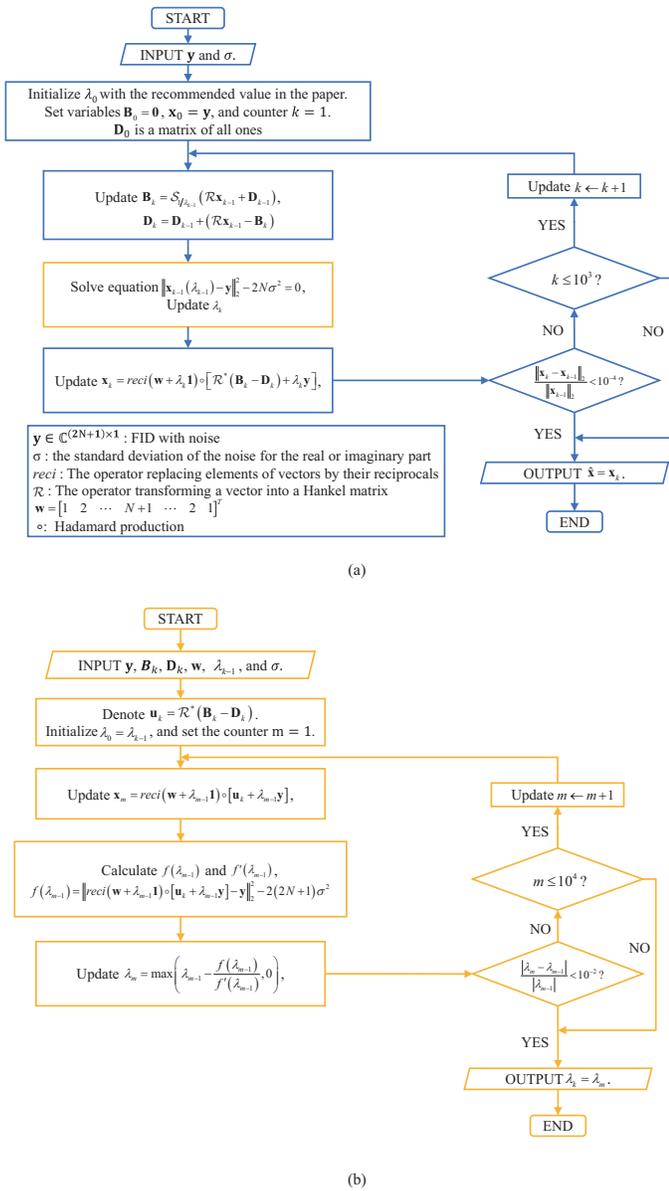


Fig. 16. The flowchart of CHORD-DP. (a) The flowchart of the whole algorithm. (b) The details of the orange box in (a).

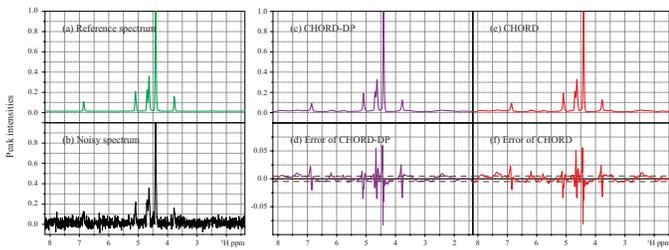


Fig. 17. Denoised results of realistic metabolite spectrum. (a) denotes the reference spectrum with high SNR. (b) denotes the noisy spectrum ( $\sigma = 0.035$ ). (c) and (e) are the denoised results by CHORD with parameter estimated by DP and the suggested parameter. (d) and (f) stand for the denoising error of two methods. The brown long dash lines denote the MAEs of the denoised spectra.

achieves more robust and accurate results compared with Cadzow and rQRd method.

In this paper, we did not discuss the effect of  $\mathbf{x}_0$  at great length and had not provided a theoretical estimate of  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$ . For the future work, it is worthwhile to explore an accurate estimate of  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$ . Moreover, we are also interested in exploring the probability distribution of the spectral norm, and extending the 1-D model in (2) to higher dimensional signals since their acquisition costs relatively more time in applications. In addition, non-exponential signals, such as Gaussian signals, are very common in applications [36]. Denoising Gaussian signals would be very different. How to denoise this type of signals is important and worthy to explore it in the future.

#### ACKNOWLEDGMENT

The authors would like to thank Hengfa Lu for polishing writing and Zhangren Tu for preparing part of code for comparison. The authors thanks You-Wei Wen, Haolin Zhan and Xinlin Zhang for helpful discussions. In addition, the authors are grateful to editors and reviewers for valuable suggestions.

#### REFERENCES

- [1] A. Wegemann, C. Staat, J. Rapp, A. Heidsieck, A. Haase, and B. Gleich, "A portable NMR spectrometer with a probe head combining RF and DC capabilities to generate pulsed-field gradients," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 8628–8636, 2020.
- [2] K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hiroaki, and M. Shirakawa, "High-resolution multi-dimensional NMR spectroscopy of proteins in human cells," *Nature*, vol. 458, no. 7234, pp. 106–109, 2009.
- [3] O. Beckonert, M. Coen, H. C. Keun, Y. Wang, T. M. D. Ebbels, E. Holmes, J. C. Lindon, and J. K. Nicholson, "High-resolution magic-angle-spinning nmr spectroscopy for metabolic profiling of intact tissues," *Nature Protocols*, vol. 5, no. 6, pp. 1019–1032, 2010.
- [4] M. C. Preul, Z. Caramanos, D. L. Collins, J.-G. Villemure, R. Leblanc, A. Olivier, R. Pokrupa, and D. L. Arnold, "Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy," *Nature Medicine*, vol. 2, no. 3, pp. 323–325, 1996.
- [5] P. P. Man, C. Bonhomme, and F. Babonneau, "Denoising NMR time-domain signal by singular-value decomposition accelerated by graphics processing units," *Solid State Nuclear Magnetic Resonance*, vol. 61, pp. 28–34, 2014.
- [6] F. Lam, Y. Li, and X. Peng, "Constrained magnetic resonance spectroscopic imaging by learning nonlinear low-dimensional models," *IEEE Transactions on Medical Imaging*, vol. 39, no. 3, pp. 545–555, 2020.
- [7] L. Chiron, M. A. V. Agthoven, B. Kieffer, C. Rolando, and M.-A. Delsuc, "Efficient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry," *Proceedings of the National Academy of Sciences*, vol. 111, no. 4, pp. 1385–1390, 2014.
- [8] F. Lam and Z.-P. Liang, "A subspace approach to high-resolution spectroscopic imaging," *Magnetic Resonance in Medicine*, vol. 71, no. 4, pp. 1349–1357, 2014.
- [9] J. Ying, J.-F. Cai, D. Guo, G. Tang, Z. Chen, and X. Qu, "Vandermonde factorization of Hankel matrix for complex exponential signal recovery—Application in fast NMR spectroscopy," *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5520–5533, 2018.
- [10] Y. Lu, S. Joshi, and J. M. Morris, "Noise reduction for NMR FID signals via Gabor expansion," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 6, pp. 512–528, 1997.
- [11] J. A. Cadzow, "Signal enhancement—A composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.
- [12] Y.-Y. Lin and L.-P. Hwang, "NMR signal enhancement based on matrix property mappings," *Journal of Magnetic Resonance, Series A*, vol. 103, no. 1, pp. 109–114, 1993.

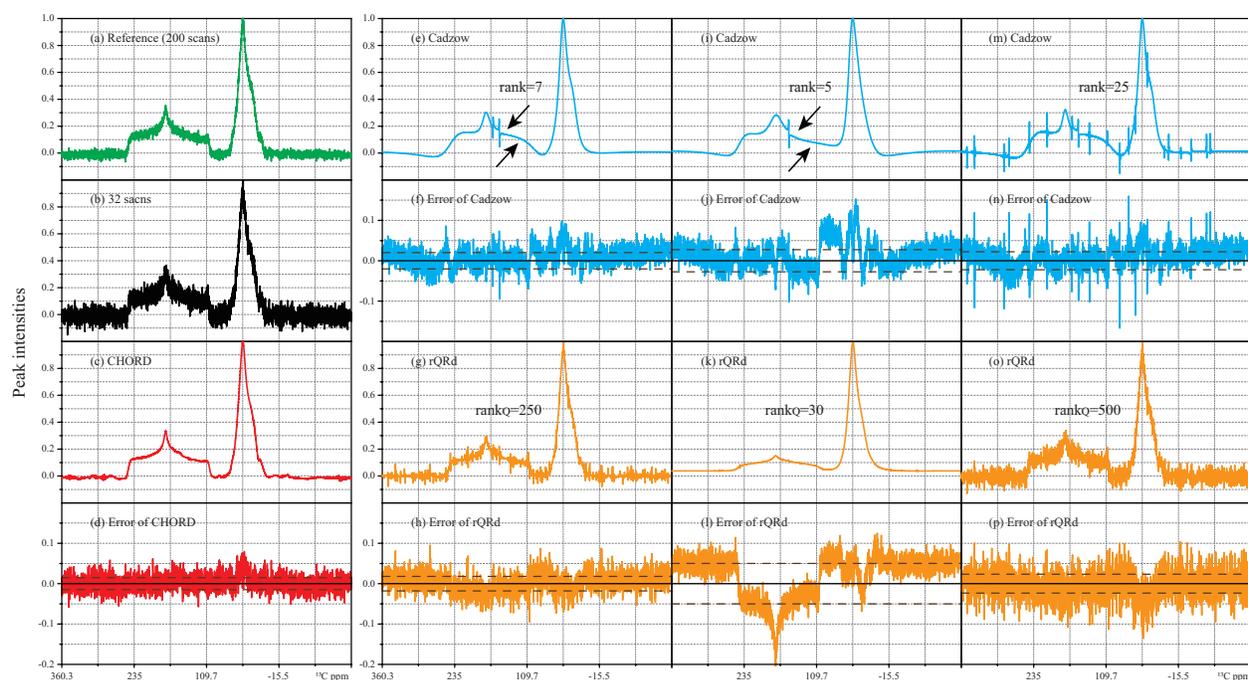


Fig. 18. The denoising of a  $^{13}\text{C}$  solid-state NMR spectra. (a) and (b) denote the reference spectrum (200 scans) and the observation (32 scans), respectively. (c)-(d) are the denoised spectra and the error of CHORD with the suggested parameter. (e)-(f), (i)-(j), and (m)-(n) show denoised spectra and their errors of Cadzow with three different estimated ranks (optimal in terms of NRMSE, small, and large). (g)-(h), (k)-(l), and (o)-(p) are denoised results of rQRd with three different estimated ranks (optimal in terms of NRMSE, small, and large). The brown long dash lines denote the MAEs of the denoised spectra.

- [13] J. Gillard, "Cadzow's basic algorithm, alternating projections and singular spectrum analysis," *Statistics and Its Interface*, vol. 3, no. 3, pp. 335–343, 2010.
- [14] X. Qu, M. Mayzel, J.-F. Cai, Z. Chen, and V. Orekhov, "Accelerated NMR spectroscopy with low-rank reconstruction," *Angewandte Chemie International Edition*, vol. 54, no. 3, pp. 852–854, 2015.
- [15] Y. Huang, H. Zhan, X. You, Y. Yang, C. Li, S. Cai, and Z. Chen, "A pure shift-based NMR method for transverse relaxation measurements on complex samples," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 1, pp. 201–211, 2020.
- [16] H. Li, Y. Yang, H. Zhan, X. Lin, and Z. Chen, "Periodic artifact suppression for pure shift NMR spectroscopy," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.
- [17] D. Chen, Z. Wang, D. Guo, V. Orekhov, and X. Qu, "Review and prospect: Deep learning in nuclear magnetic resonance spectroscopy," *Chemistry - A European Journal*, vol. 26, pp. 10391–10401, 2020.
- [18] T. Qiu, Z. Wang, H. Liu, D. Guo, and X. Qu, "Review and prospect: NMR spectroscopy denoising and reconstruction with low-rank Hankel matrices and tensors," *Magnetic Resonance in Chemistry*, vol. 59, pp. 324–345, 2021.
- [19] X. Qu, Y. Huang, H. Lu, T. Qiu, D. Guo, T. Agback, V. Orekhov, and Z. Chen, "Accelerated nuclear magnetic resonance spectroscopy with deep learning," *Angewandte Chemie International Edition*, vol. 59, pp. 10297–10300, 2020.
- [20] H. Lu, X. Zhang, T. Qiu, J. Yang, J. Ying, D. Guo, Z. Chen, and X. Qu, "Low rank enhanced matrix recovery of hybrid time and frequency data in fast magnetic resonance spectroscopy," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 4, pp. 809–820, 2018.
- [21] J. Ying, H. Lu, Q. Wei, J.-F. Cai, D. Guo, J. Wu, Z. Chen, and X. Qu, "Hankel matrix nuclear norm regularized tensor completion for n-dimensional exponential signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3702–3717, 2017.
- [22] H. M. Nguyen, X. Peng, M. N. Do, and Z. P. Liang, "Denoising MR spectroscopic imaging data with low-rank approximations," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 78–89, 2013.
- [23] A. Santos-Díaz and M. D. Noseworthy, "Comparison of compressed sensing reconstruction algorithms for  $^3\text{P}$  magnetic resonance spectroscopic imaging," *Magnetic Resonance Imaging*, vol. 59, pp. 88–96, 2019.
- [24] P. Cao, P. J. Shin, I. Park, C. Najac, I. Marco-Rius, D. B. Vigneron, S. J. Nelson, S. M. Ronen, and P. E. Z. Larson, "Accelerated high-bandwidth MR spectroscopic imaging using compressed sensing," *Magnetic Resonance in Medicine*, vol. 76, no. 2, pp. 369–379, 2016.
- [25] J. P. Haldar, "Low-rank modeling of local k-space neighborhoods (LO-RAKS) for constrained MRI," *IEEE Transactions on Medical Imaging*, vol. 33, pp. 668–681, 2014.
- [26] X. Zhang, D. Guo, Y. Huang, Y. Chen, L. Wang, F. Huang, Q. Xu, and X. Qu, "Image reconstruction with low-rankness and self-consistency of k-space data in parallel MRI," *Medical Image Analysis*, vol. 63, p. 101687, 2020.
- [27] Y. Huang, X. Zhang, H. Guo, H. Chen, D. Guo, F. Huang, Q. Xu, and X. Qu, "Phase-constrained reconstruction of high-resolution multi-shot diffusion weighted image," *Journal of Magnetic Resonance*, vol. 312, p. 106690, 2020.
- [28] J. C. Lindon and A. G. Ferrige, "Digitisation and data processing in Fourier transform NMR," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 14, no. 1, pp. 27–66, 1980.
- [29] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [30] D. Barache, J.-P. Antoine, and J.-M. Dereppe, "The continuous wavelet transform, an analysis tool for NMR spectroscopy," *Journal of Magnetic Resonance*, vol. 128, no. 1, pp. 1–11, 1997.
- [31] D. L. Donoho, I. M. Johnstone, A. S. Stern, and J. C. Hoch, "Does the maximum entropy method improve sensitivity?" *Proceedings of the National Academy of Sciences*, vol. 87, no. 13, pp. 5066–5068, 1990.
- [32] K. Takeda, "Solid-state covariance NMR spectroscopy," *Annual Reports on NMR Spectroscopy*, vol. 84, pp. 77–113, 2015.
- [33] C. Kaiser, J. J. Lopez, W. Bermel, and C. Glaubitz, "Dual transformation of homonuclear solid-state NMR spectra—an option to decrease measuring time," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1768, no. 12, pp. 3107–3115, 2007.
- [34] E. R. Malinowski, "Determination of the number of factors and the experimental error in a data matrix," *Analytical Chemistry*, vol. 49, no. 4, pp. 612–617, 1977.
- [35] —, "Abstract factor analysis of data with multiple sources of error and a modified Faber-Kowalski f-test," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 13, no. 2, pp. 69–81, 1999.
- [36] G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, and C. Bonhomme, "Denoising applied to spectroscopies—part I: concept and limits," *Applied Spectroscopy Reviews*, vol. 1–29, 2019.
- [37] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured

- matrix completion," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6576–6601, 2014.
- [38] J. C. Hoch and A. S. Stern, *NMR Data Processing*. Wiley-Liss New York, 1996.
- [39] P. Koehl, "Linear prediction spectral analysis of NMR data," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 34, no. 3-4, pp. 257–299, 2009.
- [40] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [41] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [42] G. A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra and Its Applications*, vol. 170, pp. 33–45, 1992.
- [43] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003.
- [44] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [45] M. W. Meckes, "On the spectral norm of a random Toeplitz matrix," *Electronic Communications in Probability*, vol. 12, pp. 315–325, 2007.
- [46] V. V. Nekrutkin, "Remark on the norm of random hankel matrices," *Vestnik St. Petersburg University: Mathematics*, vol. 46, pp. 189–192, 2013.
- [47] J. A. Tropp, "An introduction to matrix concentration inequalities," *Foundations and Trends® in Machine Learning*, vol. 8, no. 1-2, pp. 1–230, 2015.
- [48] W. Liao, "MUSIC for multidimensional spectral estimation: stability and super-resolution," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6395–6406, 2015.
- [49] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, pp. 1247–1250, 2014.
- [50] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, pp. 79–82, 2005.
- [51] J. F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, pp. 68–78, 1951.
- [52] K. Kazimierczuk and O. Orekhov, "Accelerated NMR spectroscopy by using compressed sensing," *Angewandte Chemie International Edition*, vol. 50, no. 24, pp. 5556–5559, 2011.
- [53] H. Engl, "Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates," *Journal of Optimization Theory Applications*, vol. 52, no. 2, pp. 209–215, 1987.
- [54] Y.-W. Wen and R. H. Chan, "Parameter selection for total-variation-based image restoration using discrepancy principle," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1770–1781, 2011.
- [55] B. Li, L. Jin, C. Leng, C. Lu, and J. Zhang, "A smooth approximation algorithm of rank-regularized optimization problem and its applications," in *International Conference on Image and Graphics*, 2015, pp. 470–479.

# Supplementary

Tianyu Qiu, Wenjing Liao, Yihui Huang, Jinyu Wu, Di Guo, Dongbao Liu, Xin Wang, Jian-Feng Cai, Bingwen Hu, and Xiaobo Qu \*

## I. THE SOLVER OF CHORD

The model of CHORD is expressed as bellow

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{2N+1}} \|\mathcal{R}\mathbf{x}\|_* + \frac{\lambda}{2} \|\mathbf{y} - \mathbf{x}\|_2^2. \quad (\text{S1})$$

Alternating Direction Method of Multipliers (ADMM) [1] is a typical iterative algorithm, which can be used to solve (S1). After introducing two variables  $\mathbf{B}, \mathbf{D} \in \mathbb{C}^{(N+1) \times (N+1)}$ , we can reformulate (S1) as follows:

$$\max_{\mathbf{D}} \min_{\mathbf{x}, \mathbf{B}} \|\mathbf{B}\|_* + \frac{\beta}{2} \|\mathcal{R}\mathbf{x} - \mathbf{B}\|_F^2 + \langle \mathcal{R}\mathbf{x} - \mathbf{B}, \mathbf{D} \rangle + \frac{\lambda}{2} \|\mathbf{y} - \mathbf{x}\|_2^2. \quad (\text{S2})$$

(S2) is solved with the following iterative scheme:

$$\begin{cases} \mathbf{x}_{k+1} = (\beta\mathcal{R}^*\mathcal{R} + \lambda\mathbf{1})^{-1} [\beta\mathcal{R}^*(\mathbf{B}_k - \mathbf{D}_k/\beta) + \lambda\mathbf{z} + \lambda\mathbf{x}_0] \\ \mathbf{B}_{k+1} = \mathcal{S}_{1/\beta}(\mathcal{R}\mathbf{x}_{k+1} + \mathbf{D}_k/\beta) \\ \mathbf{D}_{k+1} = \mathbf{D}_k + \tau(\mathcal{R}\mathbf{x}_{k+1} - \mathbf{B}_{k+1}) \end{cases} \quad (\text{S3})$$

where  $\mathbf{1} \in \mathbb{R}^{(N+1) \times 1}$  denotes a vector whose elements are 1. The subscript  $k$  denotes results in the  $k^{\text{th}}$  iteration,  $\mathcal{R}^* : \mathbb{C}^{(N+1) \times (N+1)} \rightarrow \mathbb{C}^{2N+1}$  is the adjoint operator of  $\mathcal{R}$ , which transforms a Hankel matrix into a vector through summing each anti-diagonal.

Let  $\mathbf{X} \in \mathbb{C}^{(N+1) \times (N+1)}$  be with the SVD  $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ , where  $\mathbf{\Sigma} = \text{diag}(\{\sigma_r\}_{r=1}^R)$ . The singular thresholding operator which applies in matrix  $\mathbf{X}$  is  $\mathcal{S}_{1/\beta}(\mathbf{X}) = \mathbf{U}\text{diag}(\{\sigma_r - 1/\beta\}_+) \mathbf{V}^H$ , where  $t_+ = \max(0, t)$  [2]. In ADMM,  $\beta$  and  $\tau$  are two parameters and we set  $\beta = 1$  and  $\tau = 1$ .

## II. THE DETAILED DERIVATION FROM (3) TO (4)

Since (3) in the main text is derived from the sub-gradient of the model, the optimal  $\lambda$  definitely satisfies (3). To obtain a specific  $\lambda$  value, it is necessary to know  $\hat{\mathbf{U}}, \hat{\mathbf{V}}$ , and  $\hat{\mathbf{W}}$ .

This work was supported in part by National Natural Science Foundation of China (62122064, 61971361, 61871341, 61811530021, 61672335), National Key R&D Program of China (2017YFC0108703), Science and Technology Program of Xiamen (3502Z20183053), and Xiamen University Nanqiang Outstanding Talents Program. Asterisk indicates corresponding author (Email: quxiaobo@xmu.edu.cn)

Tianyu Qiu, Yihui Huang, Jinyu Wu, Dongbao Liu, Xin Wang and Xiaobo Qu are with with Biomedical Intelligent Cloud R&D Center, Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, Xiamen University, Xiamen, China.

Wenjing Liao is with School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA.

Di Guo is with School of Computer and Information Engineering, Fujian Provincial University Key Laboratory of Internet of Things Application Technology, Xiamen University of Technology, Xiamen, China.

Jian-Feng Cai is with Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China.

Bingwen Hu is with Department of Physics, East China Normal University, Shanghai, China

However, according to the definition of the sub-gradient of the nuclear norm, the matrix  $\hat{\mathbf{W}}$  cannot be obtained directly, so we have to use the inequality scaling and numerical experiments to obtain a proper  $\lambda$ .

A. *The proof of  $\|\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}\|_2 = 1$*

We rewrite (3) as

$$\lambda(\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}}) = \mathcal{R}^*(\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}). \quad (\text{S4})$$

Since  $\hat{\mathbf{U}}$  and  $\hat{\mathbf{V}}$  unitary matrices, according to the definition of sub-gradient of the nuclear norm,

$$\|\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}\|_2 = \|\hat{\mathbf{U}}\hat{\mathbf{V}}^H\hat{\mathbf{V}} + \hat{\mathbf{W}}\hat{\mathbf{V}}\|_2 = \|\hat{\mathbf{U}}\|_2 = 1. \quad (\text{S5})$$

B. *The relationship between  $\lambda\|\mathcal{R}[\frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}})]\|_2$  and  $\|\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}\|_2$*

In this subsection, we analyze  $\lambda\|\mathcal{R}[\frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}})]\|_2$  under two situations where  $\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}$  is a Hankel matrix and  $\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}$  is not a Hankel matrix.

When  $\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}$  is a Hankel matrix,

$$\lambda\mathcal{R}\left[\frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}})\right] = \hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}, \quad (\text{S6})$$

where  $\mathbf{w}$  denotes a weight defined as  $\mathbf{w} = [1 \ 2 \ \dots \ N+1 \ \dots \ 2 \ 1]^T \in \mathbb{R}^{2N+1}$ .

Thus,

$$\lambda\left\|\mathcal{R}\left[\frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}})\right]\right\|_2 = 1. \quad (\text{S7})$$

When  $\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}$  is not a Hankel matrix,  $\lambda\mathcal{R}[\frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}})]$  is a Hankel approximation of  $\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}$  satisfies that [3]

$$\lambda\left\|\mathcal{R}\left[\frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}})\right]\right\|_F < \|\hat{\mathbf{U}}\hat{\mathbf{V}}^H + \hat{\mathbf{W}}\|_F. \quad (\text{S8})$$

Here we do numerical experiments on complex random matrices to illustrate that the relationship in (S8) approximately holds in the spectral norm.

In the experiments, we tested two types of random matrices: the real and imaginary part of entries in  $\mathbf{X}$  satisfy 1) the standard normal distribution and 2) the uniform distribution. The matrix  $\mathbf{X}$  is of size  $M \times M$  with  $M = 10, 20, 30, \dots$ , and 1000, respectively, and we repeated 100 Monte Carlo trials to incorporate the randomness of matrices. The results of  $\|\mathbf{X}\|_2 - \|\mathcal{R}(\frac{1}{\mathbf{w}} \circ \mathcal{R}^*\mathbf{X})\|_2$  were recorded and presented.

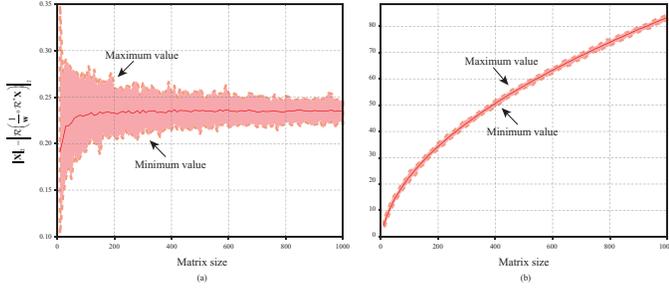


Fig. S1. Empirical test of  $\|\mathbf{X}\|_2 - \|\mathcal{R}(\frac{1}{\mathbf{w}} \circ \mathcal{R}^* \mathbf{X})\|_2$ . The real and imaginary part of entries in  $\mathbf{X}$  satisfy the standard normal distribution (a) and the uniform distribution (b). The red lines denote the average of  $\|\mathbf{X}\|_2 - \|\mathcal{R}(\frac{1}{\mathbf{w}} \circ \mathcal{R}^* \mathbf{X})\|_2$  in 100 trials. The pink areas stand for the distribution. The red dash lines are the maximum and minimum values of  $\|\mathbf{X}\|_2 - \|\mathcal{R}(\frac{1}{\mathbf{w}} \circ \mathcal{R}^* \mathbf{X})\|_2$  in Monte Carlo trials, respectively.

Results in Fig. S1 show that the relationship in (S8) approximately holds in the spectral norm. Therefore,

$$\lambda \left\| \mathcal{R} \left[ \frac{1}{\mathbf{w}} \circ (\mathbf{x}_0 + \mathbf{z} - \hat{\mathbf{x}}) \right] \right\|_2 < \left\| \hat{\mathbf{U}} \hat{\mathbf{V}}^H + \hat{\mathbf{W}} \right\|_2 = 1. \quad (\text{S9})$$

### III. THE PROOF OF THEOREM 1

The proof of Theorem 1 are based on the following lemmas:

**Lemma 1.** *Suppose the real and imaginary parts of the entries in  $\mathbf{z} \in \mathbb{C}^{2N+1}$  are i.i.d. Gaussian random variables with mean 0 and variance  $\sigma^2$ . Define  $\{d_k\}_{k=0}^{2N}$  as Theorem 1, and then*

$$\|\mathbf{Z}\|_2 \geq \frac{(N+1)\sigma}{2N+1} \sup_{0 \leq \omega \leq 1} \left| \sum_{k=0}^{2N} d_k p_k e^{i2\pi k \omega} \right|, \quad (\text{S10})$$

where  $p_k \sim \mathcal{N}(0, 1)$ ,  $k = 0, 1, 2, \dots, 2N$ .

*Proof.* For any vectors  $\mathbf{a}, \mathbf{b} \in \mathbb{C}^{N+1}$ ,

$$|(\mathbf{b}, \mathbf{Z}\mathbf{a})| = |\mathbf{b}^H \mathbf{Z}\mathbf{a}| \leq \|\mathbf{b}\|_2 \|\mathbf{a}\|_2 \|\mathbf{Z}\|_2.$$

We use the technique in [4] to derive the lower bound of  $\mathbb{E}\|\mathbf{Z}\|_2$  by choosing proper vectors  $\mathbf{a}$  and  $\mathbf{b}$ . Let  $a_{k_1} = \frac{1}{k_1+1} e^{i2\pi k_1 \omega}$  and  $b_{k_2} = \frac{1}{k_2+1} e^{-i2\pi k_2 \omega}$ , where  $\omega \in [0, 1]$  and  $k_1, k_2 = 0, \dots, N$ . Then

$$\begin{aligned} \|\mathbf{Z}\|_2 &\geq \frac{1}{C_N} \sup_{0 \leq \omega \leq 1} \left| \sum_{k_1=0}^N \sum_{k_2=0}^N \frac{1}{k_1+1} \frac{1}{k_2+1} \frac{e^{i2\pi k_1 \omega} e^{i2\pi k_2 \omega} z_{k_1+k_2}}{w_{k_1+k_2}} \right| \\ &= \frac{1}{C_N} \sup_{0 \leq \omega \leq 1} \left| \sum_{k=0}^{2N} d_k z_k e^{i2\pi k \omega} \right| \\ &= \frac{\sigma}{C_N} \sup_{0 \leq \omega \leq 1} \left| \sum_{k=0}^{2N} d_k p_k e^{i2\pi k \omega} \right|, \end{aligned} \quad (\text{S11})$$

where  $C_N = \sum_{k_1=0}^N \frac{1}{(k_1+1)^2} \leq \frac{2N+1}{N+1}$ , which yields the conclusion in (S10).  $\square$

**Lemma 2.** *Let  $\{d_k\}_{k=0}^{2N}$  and  $\{p_k\}_{k=0}^{2N}$  be the sequences defined in Theorem 1 and Lemma 1. If  $R_N$  and  $Q_N$  are defined as*

$$R_N^2 = \sum_{k=0}^{2N} |d_k|^2 \text{ and } Q_N^4 = \sum_{k=0}^{2N} |d_k|^4,$$

then there exists a constant  $C$  such that

$$\mathbb{E} \left( \sup_{0 \leq \omega \leq 1} \left| \sum_{k=0}^{2N} d_k p_k e^{i2\pi k \omega} \right| \right) \geq C \sqrt{R_N^2} \left( 1 + \log \frac{R_N^4}{Q_N^4} \right), \quad (\text{S12})$$

Lemma 2 is a special case of Theorem in [5]. Details of the proof is shown as below.

Here is the proof of Lemma 2.

**Proposition 1.** [5] *For every  $M < \infty$  there exists a constant  $C(M) > 0$  such that, whenever  $\{\psi_k\}_{k=0}^{2N}$  is a system of functions in an  $L_2(\mu)$ -space satisfying*

- (1°)  $\|\psi_k\|_{L_2(\mu)} = 1$  and  $\|\psi_k\|_{L_3(\mu)} \leq M$ , for all  $0 \leq k \leq 2N$ ,
  - (2°)  $\left\| \sum_{k=0}^{2N} d_k \psi_k \right\|_{L_2(\mu)} \leq M \sqrt{\sum_{k=0}^{2N} |d_k|^2}$ , for all  $0 \leq k \leq 2N$ ,
  - and  $\{p_k\}_{k=0}^{2N}$  are independent random variables over a probability space  $(T, \mathcal{T}, \tau)$  with
  - (3°)  $\mathbb{E}(p_k) = 0$ ,  $\mathbb{E}|p_k|^2 = 1$ , and  $\sqrt[3]{\mathbb{E}|p_k|^3} \leq M$ , for all  $0 \leq k \leq 2N$ ,
- then, for any choice of the coefficients of  $\{d_k\}_{k=0}^{2N}$ , we have

$$\mathbb{E} \left\| \sum_{k=0}^{2N} d_k p_k \psi_k \right\|_{L_\infty(\mu)} \geq C \sqrt{\sum_{k=0}^{2N} |d_k|^2} \sqrt{1 + \log \frac{(\sum_{k=0}^{2N} |d_k|^2)^2}{\sum_{k=0}^{2N} |d_k|^4}}.$$

*Proof.* According to (S11),  $\psi_k = e^{i2\pi k \omega}$ . It is obvious that for all  $0 \leq k \leq 2N$ ,

$$\|e^{i2\pi k \omega}\|_{L_2(\mu)} = 1 \text{ and } \|e^{i2\pi k \omega}\|_{L_3(\mu)} = 1. \quad (\text{S13})$$

According to the triangle inequality,

$$\left\| \sum_{k=0}^{2N} d_k \psi_k \right\|_{L_2(\mu)} \geq \sqrt{\int_0^1 \sum_{k=0}^{2N} |d_k e^{i2\pi k \omega}|^2 d\omega} = \sqrt{\sum_{k=0}^{2N} |d_k|^2}. \quad (\text{S14})$$

$p_k$  is a random variable which satisfies normal distribution, thus

$$\mathbb{E}(p_k) = 0, \mathbb{E}|p_k|^2 = 1 \text{ and } \sqrt[3]{\mathbb{E}|p_k|^3} = 0. \quad (\text{S15})$$

Combining (S13), (S14) and (S15) yields Lemma 2.  $\square$

Combining Lemma 1 and Lemma 2 results in Theorem 1.

### IV. THE PROOF OF THEOREM 2

*Proof.* We express  $\mathbf{Z}$  as a sum of independent matrices such that

$$\mathbf{Z} = \sum_{k=0}^{2N} \frac{1}{w_k} z_k \mathbf{B}_k, \quad (\text{S16})$$

where  $\mathbf{B}_k \in \mathbb{R}^{(N+1) \times (N+1)}$  has one on the  $(k+1)^{th}$  skew diagonal and all other entries are 0. For example,

$$\mathbf{B}_1 = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

According to [6],

$$\mathbb{E} \|\mathbf{Z}\|_2 \leq \sqrt{2\nu^2(\mathbf{Z}) \log(2N+2)}, \quad (\text{S17})$$

where  $\nu^2(\mathbf{Z}) = \max\{\|\mathbb{E}(\mathbf{Z}^H \mathbf{Z})\|_2, \|\mathbb{E}(\mathbf{Z} \mathbf{Z}^H)\|_2\}$ . The parameter  $\nu^2(\mathbf{Z})$  can be calculated as follows

$$\begin{aligned} \|\mathbb{E}(\mathbf{Z}^H \mathbf{Z})\|_2 &= \|\mathbb{E}(\mathbf{Z} \mathbf{Z}^H)\|_2 \\ &= \left\| \mathbb{E} \left( \left( \sum_{k=0}^{2N} \frac{z_k}{w_k} \mathbf{B}_k \right) \left( \sum_{m=0}^{2N} \frac{z_m}{w_m} \mathbf{B}_m \right)^H \right) \right\|_2 \\ &= \left\| \mathbb{E} \left( \sum_{k=0}^{2N} \frac{|z_k|^2}{w_k^2} \mathbf{B}_k \mathbf{B}_k^H \right) \right\|_2. \end{aligned} \quad (\text{S18})$$

Denote the diagonal matrix  $\mathbf{C}_k = \mathbf{B}_k \mathbf{B}_k^H \in \mathbb{R}^{(N+1) \times (N+1)}$ ,  $k = 0, 1, \dots, 2N$ . When  $0 \leq k \leq N$ , the first  $k+1$  diagonal entries of  $\mathbf{C}_k$  are one, and others are zero. When  $N+1 \leq k \leq 2N$ , the last  $2N+1-k$  diagonal entries of  $\mathbf{C}_k$  are one, and others are zero. For example,

$$\mathbf{C}_1 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \text{ and } \mathbf{C}_{N+1} = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

Substituting (S18) into the definition of  $\nu^2(\mathbf{Z})$  in (S17) results in

$$\nu^2(\mathbf{Z}) = \left\| \mathbb{E} \left( \sum_{k=0}^{2N} \frac{|z_k|^2}{w_k^2} \mathbf{C}_k \right) \right\|_2 = \sigma^2 C_w. \quad (\text{S19})$$

Finally we combine (S19) and (S17) to obtain Theorem 2.  $\square$

## V. ASYMPTOTIC ANALYSIS OF THE ESTIMATES OF BOUNDS

**Theorem 1.** Let  $R_N$  and  $Q_N$  be defined as Theorem 1. Then there exists a constant  $C_L > 0$  such that

$$\lim_{N \rightarrow +\infty} \frac{(N+1)C}{2N+1} \sqrt{R_N^2 \left(1 + \log \frac{R_N^4}{Q_N^4}\right)} = C_L. \quad (\text{S20})$$

The proof of Theorem 1 is based on two lemmas below which study the asymptotic of  $R_N^2$  and  $Q_N^4$  as  $N \rightarrow \infty$ .

**Lemma 3.** Let  $R_N$  be defined as  $R_N^2 = \sum_{k=0}^{2N} |d_k|^2$ . Then there exists a constant  $C_R > 0$  such that

$$\lim_{N \rightarrow +\infty} R_N^2 = C_R. \quad (\text{S21})$$

**Lemma 4.** Let  $Q_N$  be defined as  $Q_N^4 = \sum_{k=0}^{2N} |d_k|^4$ . Then there exists a constant  $C_Q > 0$  such that

$$\lim_{N \rightarrow +\infty} Q_N^4 = C_Q. \quad (\text{S22})$$

Lemma 3 and Lemma 4 are proved as following.

Here is the proof of Lemma 3.

*Proof.* According to the definition,  $R_N^2$  is expressed as

$$\begin{aligned} R_N^2 &= 4 \underbrace{\sum_{k=0}^N \left( \frac{1}{(k+1)(k+2)} \sum_{m=0}^k \frac{1}{m+1} \right)^2}_{R_N^{(1)}} \\ &+ 4 \underbrace{\sum_{k=N+1}^{2N} \left( \frac{1}{(2N-k+1)(k+2)} \sum_{m=k}^{2N} \frac{1}{m-N+1} \right)^2}_{R_N^{(2)}}. \end{aligned} \quad (\text{S23})$$

The sequence  $R_N^{(1)}$  is positive, and increases as  $N$  increases. It is straight forward

$$1 \leq \sum_{m=0}^k \frac{1}{m+1} \leq k+1. \quad (\text{S24})$$

Let  $F_1(N) = \int_0^N \frac{1}{(l+1)^2(l+2)^2} dl$ . The sequence  $R_N^{(1)}$  satisfies the following upper bound and lower bound

$$\begin{aligned} R_N^{(1)} &\geq \sum_{k=0}^N \frac{1}{(k+1)^2(k+2)^2} > F_1(N) \\ &= \frac{3}{2} - 2 \log 2 - \left( \frac{1}{N+1} + \frac{1}{N+2} \right) \\ &+ 2 \log \left( 1 + \frac{1}{N+1} \right), \end{aligned} \quad (\text{S25})$$

and

$$R_N^{(1)} \leq \sum_{k=0}^N \frac{1}{(k+2)^2} < 1 - \frac{1}{N+2}. \quad (\text{S26})$$

Combining (S25) and (S26) gives rise to

$$\frac{3}{2} - 2 \log 2 < \lim_{N \rightarrow +\infty} R_N^{(1)} < 1. \quad (\text{S27})$$

Since  $R_N^{(1)}$  increases as  $N$  increases,

$$\lim_{N \rightarrow +\infty} R_N^{(1)} = C_{R_1}, \quad (\text{S28})$$

where  $\frac{3}{2} - 2 \log 2 < C_{R_1} < 1$ .

The sequence  $R_N^{(2)}$  can be rewritten as

$$R_N^{(2)} = \sum_{k=0}^{N-1} \left( \frac{1}{(N-k)(k+N+3)} \sum_{m=0}^{N-1-k} \frac{1}{m+k+2} \right)^2. \quad (\text{S29})$$

For  $\left( \sum_{m=0}^{N-1-k} \frac{1}{m+k+2} \right)^2$  where  $k = 0, 1, \dots, N-1$ , it is obvious that

$$\left( \sum_{m=0}^{N-1-k} \frac{1}{m+k+2} \right)^2 \leq \left( \sum_{m=0}^{N-1} \frac{1}{m+2} \right)^2.$$

According to Cauchy-Buniakowsky-Schwarz inequality,

$$\begin{aligned} \left( \sum_{m=0}^{N-1} \frac{1}{m+2} \right)^2 &\leq \sum_{m=0}^{N-1} \left( \frac{1}{m+2} \right)^2 \sum_{m=0}^{N-1} 1 \\ &< N \left( 1 - \frac{1}{N+1} \right) \\ &< N. \end{aligned} \quad (\text{S30})$$

We have

$$0 < \sum_{m=0}^{N-1-k} \frac{1}{m+k+2} < \sqrt{N}. \quad (\text{S31})$$

Denote  $g(k) = \frac{N}{(N-k)^2(k+N+3)^2}$  where  $k = 0, 1, 2, \dots, N-1$ .

The limit of  $\sum_{k=0}^{N-1} g(k)$  can be calculated as

$$\begin{aligned} &\sum_{k=0}^{N-1} \frac{N}{(N-k)^2(k+N+3)^2} \\ &= \sum_{k=0}^{N-1} \frac{2N}{(2N+3)^3(N-k)} + \frac{2N}{(2N+3)^3(k+N+3)} \\ &+ \sum_{k=0}^{N-1} \frac{N}{(2N+3)^2(N-k)^2} + \frac{N}{(2N+3)^2(k+N+3)^2} \\ &< \frac{4N^2}{(2N+3)^3} + \frac{2N-1}{(2N+3)^3} + \frac{N^2}{(N+1)(N+2)(2N+3)^2} \end{aligned} \quad (\text{S32})$$

thus

$$\lim_{N \rightarrow +\infty} \sum_{k=0}^{N-1} g(k) = 0. \quad (\text{S33})$$

Since  $R_N^{(2)} \geq 0$  and it satisfies  $R_N^{(2)} < \sum_{k=0}^{N-1} g(k)$ ,

$$\lim_{N \rightarrow +\infty} R_N^{(2)} = 0. \quad (\text{S34})$$

Combining (S27) and (S33) results in Lemma 3.  $\square$

Lemma 4 is proved is below.

*Proof.* According to the definition,  $Q_N^4$  is given by

$$\begin{aligned} Q_N^4 &= 16 \underbrace{\sum_{k=0}^N \left( \frac{1}{(k+1)(k+2)} \sum_{m=0}^k \frac{1}{m+1} \right)^4}_{Q_N^{(1)}} \\ &+ 16 \underbrace{\sum_{k=N+1}^{2N} \left( \frac{1}{(2N-k+1)(k+2)} \sum_{m=k}^{2N} \frac{1}{m-N+1} \right)^4}_{Q_N^{(2)}}. \end{aligned} \quad (\text{S35})$$

The same technique as Lemma 3 to prove (S35)

$$\frac{111}{8} - 20 \log 2 \leq \lim_{N \rightarrow +\infty} Q_N^{(1)} < \frac{1}{3}, \quad (\text{S36})$$

and the details will not be shown here.

The sequence  $Q_N^{(2)}$  is restated as

$$Q_N^{(2)} = \sum_{k=0}^{N-1} \left( \frac{1}{(N-k)(k+N+3)} \sum_{m=0}^{N-1-k} \frac{1}{m+k+2} \right)^4. \quad (\text{S37})$$

According to (S29),

$$\begin{aligned} \left( \sum_{m=0}^{N-1-k} \frac{1}{m+k+2} \right)^4 &\leq \left( \sum_{m=0}^{N-1} \frac{1}{m+2} \right)^4 \\ &< N^2 \\ &< (k+N+3)^3. \end{aligned} \quad (\text{S38})$$

$$0 < \left( \sum_{m=0}^{N-1-k} \frac{1}{m+k+2} \right)^4 < (k+N+3)^3. \quad (\text{S39})$$

Define  $h(k) = \frac{1}{(N-k)^4(k+N+3)}$  where  $k = 0, 1, 2, \dots, N-1$ , then the limit of  $\sum_{k=0}^{N-1} \frac{1}{(N-k)^4(k+N+3)}$  can be calculated as

$$\begin{aligned} &\sum_{k=0}^{N-1} \frac{1}{(N-k)^4(k+N+3)} \\ &= \sum_{k=0}^{N-1} \frac{1}{(2N+3)^4(N-k)} + \sum_{k=0}^{N-1} \frac{1}{(2N+3)^3(N-k)^2} \\ &+ \sum_{k=0}^{N-1} \frac{1}{(2N+3)^2(N-k)^3} + \sum_{k=0}^{N-1} \frac{1}{(2N+3)(N-k)^4} \\ &+ \sum_{k=0}^{N-1} \frac{1}{(2N+3)^4(k+N+3)} \\ &< \frac{2N}{(2N+3)^4} + \frac{2N-1}{N(2N+3)^3} + \frac{2N-1}{N(2N+3)^2} + \frac{2N-1}{N(2N+3)}. \end{aligned} \quad (\text{S40})$$

Thus

$$\lim_{N \rightarrow +\infty} \sum_{k=0}^{N-1} h(k) = 0. \quad (\text{S41})$$

Since sequence  $Q_N^{(2)} \geq 0$  and it satisfies  $Q_N^{(2)} < \sum_{k=0}^{N-1} h(k)$ ,

$$\lim_{N \rightarrow +\infty} Q_N^{(2)} = 0. \quad (\text{S42})$$

Combining (S36) and (S42) yields Lemma 4.  $\square$

Combining Lemma 3 and Lemma 4 gives rise to

$$\begin{aligned} &\lim_{N \rightarrow +\infty} \frac{(N+1)C}{2N+1} \sqrt{R_N^2 \left( 1 + \log \frac{R_N^4}{Q_N^4} \right)} \\ &= C \sqrt{C_R \left( 1 + \log \frac{C_R^2}{C_Q} \right)} \lim_{N \rightarrow +\infty} \frac{(N+1)}{2N+1} \\ &= \frac{C}{2} \sqrt{C_R \left( 1 + \log \frac{C_R^2}{C_Q} \right)}, \end{aligned} \quad (\text{S43})$$

which gives rise to Theorem 1 with

$$C_L = \frac{C}{2} \sqrt{C_R \left( 1 + \log \frac{C_R^2}{C_Q} \right)}.$$

When  $N$  is large enough, the upper bound and the lower bound only differ by a factor of  $\sqrt{\log N}$ . According to the analysis above, we suggest to choose  $\mathbb{E} \|\mathbf{Z}\|_2$  as

$$\mathbb{E} \|\mathbf{Z}\|_2 = \frac{C(N+1)}{(2N+1)} \sqrt{R_N^2 \left(1 + \log \frac{R_N^4}{Q_N^4}\right)} \sigma. \quad (\text{S44})$$

## VI. THE KS TEST IN THE NOISE ESTIMATE

In the main text, we introduce KS test into the noise estimate to ensure that the truncated signals satisfy the Gaussian distribution. In order to avoid the bias, the  $p$  values with different truncated lengths under various noise levels are tested on three synthetic exponential signals with varying decays (See Figs. S2 and S3).

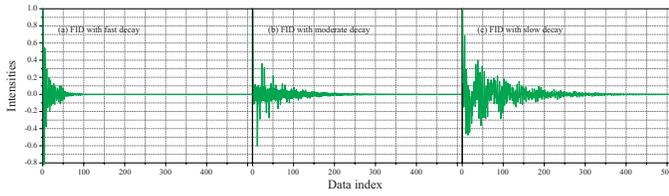


Fig. S2. Synthetic exponential signals used to test  $p$  values. (a)-(c) represents the real part of three FID signals with fast, moderate, and slow decay, respectively.

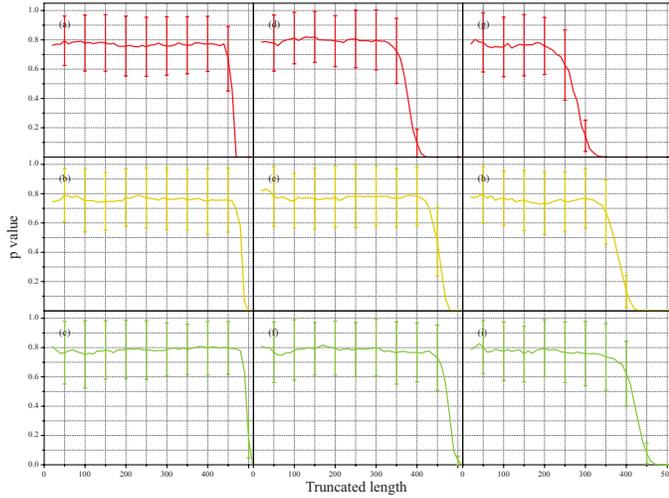


Fig. S3. The  $p$  values with different truncated lengths under various noise levels. (a)-(c) are  $p$  values of signals in Fig. S2(a) under noise with  $\sigma = 0.01, 0.03, \text{ and } 0.05$ . (d)-(f) are  $p$  values of signals in Fig. S2(b) under noise with  $\sigma = 0.01, 0.03, \text{ and } 0.05$ . (g)-(i) are  $p$  values of signals in Fig. S2(c) under noise with  $\sigma = 0.01, 0.03, \text{ and } 0.05$ . Note: The curves denote the average of  $p$  values. The bars represent the standard deviation and come from the randomness of noise.

Results in Figs. S2 and S3 confirm our conclusions in the main text. Therefore, we suggest  $p = 0.8$  as the threshold in the noise estimate.

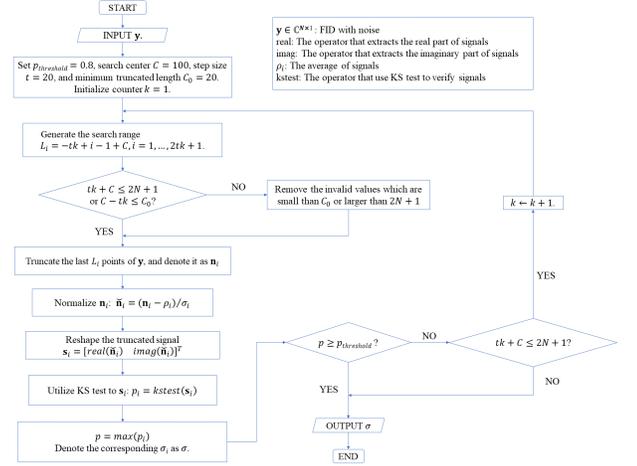


Fig. S4. The flowchart of the noise estimate based on KS test in the proposed method.

## VII. THE EFFECT OF PHASE

In Section V.B, the effect of the zero-order and relative phase has been discussed. In this section, we presented the denoised synthetic spectra with 256 and 1024 points to avoid the bias. Spectra and the errors in Figs. S5 to S8 confirm the conclusion (Figs. 12 and 13 in the main text) that, the slight change of  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$  does not greatly affect the denoising.

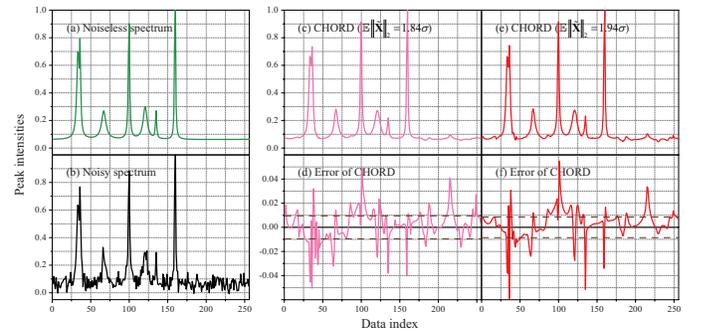


Fig. S5. The comparison of 256-point denoised spectra with the zero-order phase under the new and the suggested  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$  values. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.02$ ). (c) and (e) are the denoised results of CHORD with the new value ( $1.84\sigma$ ) and the suggested value ( $1.94\sigma$ ), respectively. (d) and (f) stand for the denoising errors of spectra that correspond to (c) and (e), respectively. The brown long dash lines denote the MAEs of the denoised spectra.

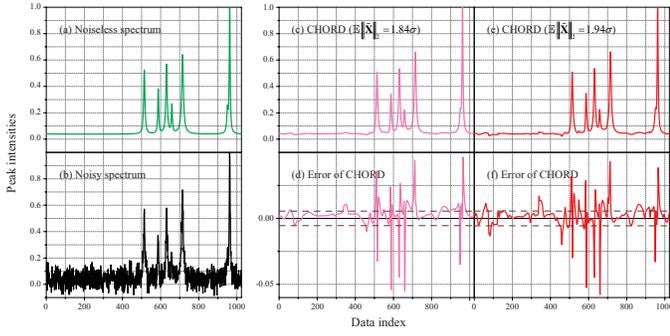


Fig. S6. The comparison of 1024-point denoised spectra with the zero-order phase under the new and the suggested  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$  values. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.02$ ). (c) and (e) are the denoised spectra of CHORD with the new value ( $1.84\sigma$ ) and the suggested value ( $1.94\sigma$ ), respectively. (d) and (f) stand for the denoising errors of spectra that correspond to (c) and (e), respectively. The brown long dash lines denote the MAEs of the denoised spectra.

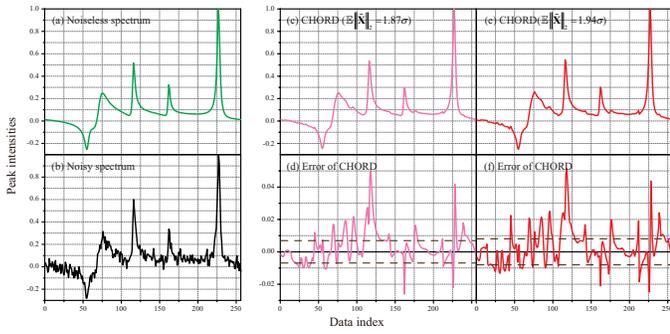


Fig. S7. The comparison of 256-point denoised spectra with the relative phase under the new and the suggested  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$  values. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.02$ ). (c) and (e) are the denoised spectra of CHORD with the new value ( $1.87\sigma$ ) and the suggested value ( $1.94\sigma$ ), respectively. (d) and (f) stand for the denoising errors of spectra that correspond to (c) and (e), respectively. The brown long dash lines denote the MAEs of the denoised spectra.

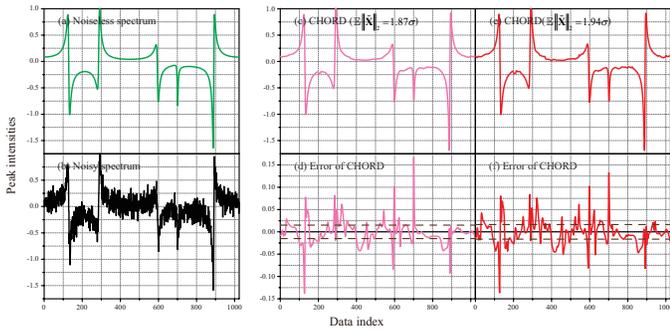


Fig. S8. The comparison of 1024-point denoised spectra with the relative phase under the new and the suggested  $\mathbb{E} \|\tilde{\mathbf{X}}\|_2$  values. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.02$ ). (c) and (e) are the denoised results of CHORD with the new value  $1.87\sigma$  and the suggested value  $1.94\sigma$ , respectively. The brown long dash lines denote the MAEs of the denoised spectra.

## VIII. COMPARISON WITH CS AND CHORD-DP

Due to the overlength of the main text, the denoised spectra and the errors of simulated signals are shown in this section.

### A. CS

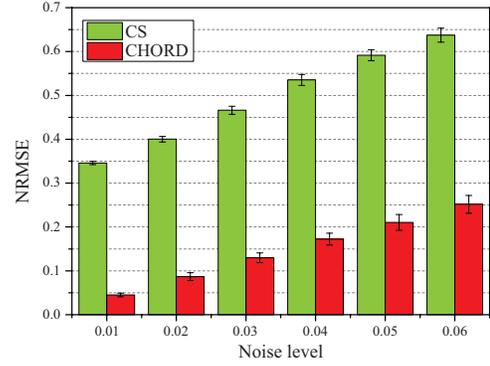


Fig. S9. The NRMSE of CS and CHORD for the synthetic data under different noise levels. The height of columns shows the average of the NRMSEs over 100 trials. The vertical bar comes from the randomness of noise. Note: The standard deviation of the noise is estimated, and the details of noise estimate has been explained in the main text Section V.A.

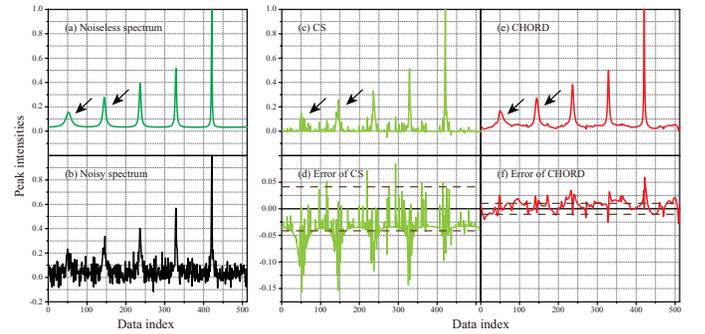


Fig. S10. Denoised results of synthetic data by CS and CHORD. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.04$ ), respectively. (c) and (e) are the denoised spectra of CS and CHORD. (d) and (f) stand for the denoising errors of two methods, respectively. The brown long dash lines denote the MAEs of the denoised spectra.

### B. CHORD-DP

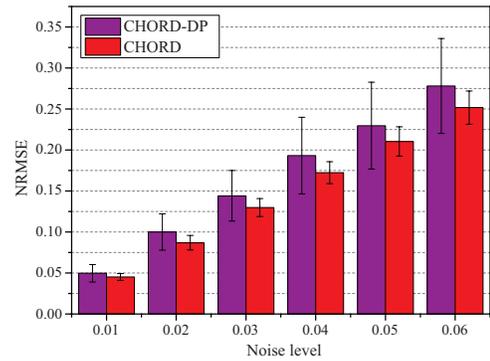


Fig. S11. The NRMSE of CHORD-DP for synthetic data under different noise levels. The height of columns shows the average of the NRMSEs over 100 trials. The vertical bar comes from the randomness of noise. Note: The standard deviation of the noise is estimated, and the details of noise estimate has been explained in the main text Section V.A.

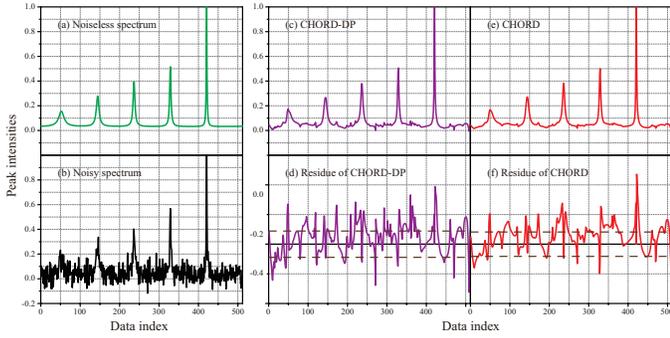


Fig. S12. The typical denoised spectra of CHORD-DP and CHORD. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.04$ ), respectively. (c) and (e) denote spectra of CHORD with parameter estimated by DP and the suggested parameter. (d) and (f) stand for the denoising errors of two methods, respectively. The brown long dash lines denote the MAEs of the denoised spectra.

### IX. OTHER FACTORS AFFECTING $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$

In this section, we force the amplitude and damping factor of exponentials to satisfy the Poisson and Beta distributions, respectively. The frequency is still uniformly sampled to maintain the maximum randomness.

Following the main text and other publication [7], we randomly generated 90 exponential signals with different lengths (255, 511, and 1023). Each signal has  $3R + 1$  parameters, including  $R$ ,  $a_r$ ,  $f_r$  and  $\tau_r$ , where  $r = 1, 2, \dots, R$ . The exponential components are  $R = 4 + M_r$ , where  $M_r$  denotes a pseudo-random scalar integer of range  $[1, 9]$ . Each frequency  $f_r$  is uniformly sampled from  $(0, 1)$ .

1) Poisson distribution: The amplitude  $a_r$  is i.i.d Poisson with mean 5 and variance 5. The damping coefficient is  $\tau_r = 5 + m_r$ , where  $m_r$  satisfies the Poisson distribution with mean 30 and variance 30.

2) Beta distribution: The amplitude  $a_r$  is i.i.d. Beta with mean 5 and variance 27.27. The damping coefficient is  $\tau_r = 5 + 60m_r$ , where  $m_r$  satisfies the Beta distribution with mean 0.5 and variance 0.27.

A grid search method is adopted to estimate the empirical distribution of  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$ . For each length, we randomly selected 3 signals (Green, blue, and red lines in Figs. S13 and S14) and used color lines to present the distribution of  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  (the average and the standard deviation).

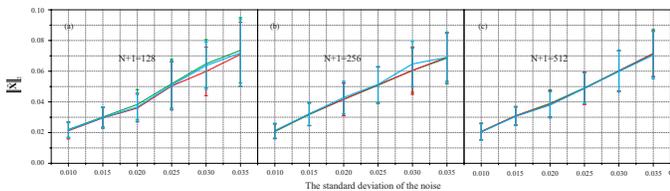


Fig. S13. The relation between  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  given by random synthetic signals with Poisson distribution and noise levels. (a)-(c) show the spectral norm of  $\tilde{\mathbf{X}}$  which is of size  $(N + 1) \times (N + 1)$  with  $(N + 1) = 128, 256, 512$ , respectively. The vertical error bars come from 50 Monte Carlo trials. For each sub-plot, the green, blue, and red lines denote  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  given by different random signals.

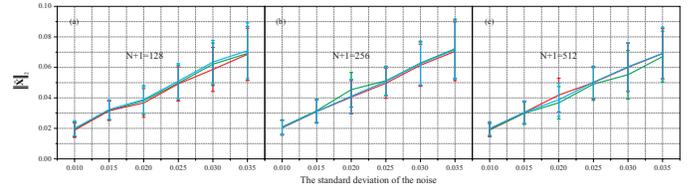


Fig. S14. The relation between  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  given by random synthetic signals with Beta distribution and noise levels. (a)-(c) show the spectral norm of  $\tilde{\mathbf{X}}$  which is of size  $(N + 1) \times (N + 1)$  with  $(N + 1) = 128, 256, 512$ , respectively. The vertical error bars come from 50 Monte Carlo trials. For each sub-plot, the green, blue, and red lines denote  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  given by different random signals.

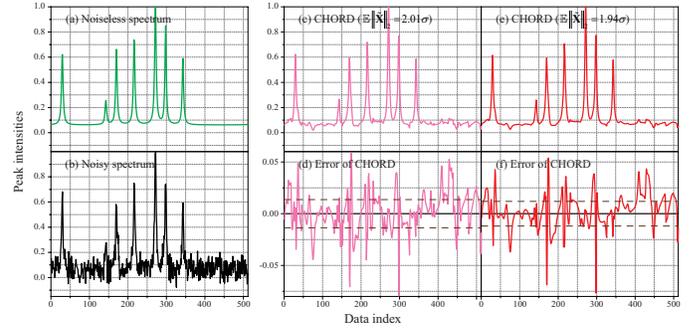


Fig. S15. The comparison of denoised spectra with Poisson distribution under the new and the suggested  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  values. (a) and (b) denote the 512-point synthetic signals without and with noise ( $\sigma = 0.02$ ). (c) and (e) are the denoised spectra obtained by CHORD with the new value ( $2.01\sigma$ ) and the suggested value ( $1.94\sigma$ ), respectively. (d) and (f) stand for the denoising errors of spectra that correspond to (c) and (e), respectively. Note: The brown long dash lines denote the MAEs of the denoised spectra.

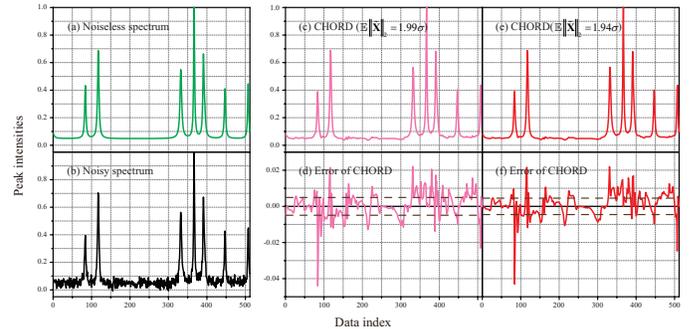


Fig. S16. The comparison of denoised spectra with Beta distribution under the new and the suggested  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  values. (a) and (b) denote the 512-point synthetic signals without and with noise ( $\sigma = 0.01$ ). (c) and (e) are the denoised spectra of CHORD with the new value ( $1.99\sigma$ ) and the suggested value ( $1.94\sigma$ ), respectively. (d) and (f) stand for the denoising errors of spectra that correspond to (c) and (e), respectively. Note: The brown long dash lines denote the MAEs of the denoised spectra.

For Poisson and Beta distributions, the averaged slopes (Figs S13 and S14), e.g. the ratio of  $\mathbb{E} \left\| \tilde{\mathbf{X}} \right\|_2$  over noise standard deviation, are 2.01 and 1.99, respectively. Compared with the suggested value 1.94 in the main text, the relative change in slope is small (4% for Poisson distribution and 3% for Beta distribution). This slight change did not greatly affect the denoising (Figs. S15 and S16).

Figs. S17 and S18 show the denoising error of the spectra at different noise levels. These two figures imply that there are two types of errors, noise fluctuation at the no-peak region and pseudo peaks at the locations of true peaks. Thus,  $\|\tilde{\mathbf{X}}\|_2$  can be affected by both noise and spectrum. Additionally, the error of denoising is increased with the increase of noise since the denoising becomes harder for heavier noise, although the noiseless spectrum does not change.

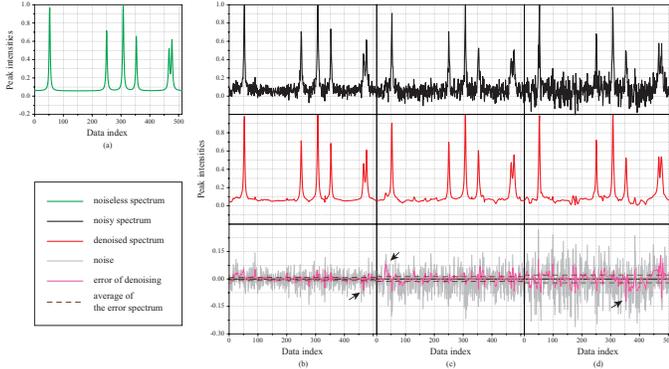


Fig. S17. The simulated noiseless signal with Poisson distribution, the denoised spectra and their error of CHORD denoising under three noise levels. (a) denotes the noiseless synthetic signal with 512 points. (b)-(d) denote the noisy, denoised, and error spectra under the noise standard deviation  $\sigma = 0.015, 0.025,$  and  $0.035,$  respectively. Note: The brown long dash lines denote the MAEs of the denoised spectra. The grey fluctuation is the pure noise.

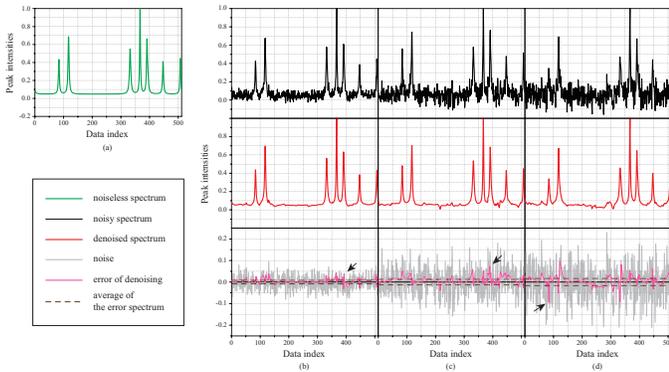


Fig. S18. The simulated noiseless signal with Beta distribution, the denoised spectra and their error of CHORD denoising under three noise levels. (a) denotes the noiseless synthetic signal with 512 points. (b)-(d) denote the noisy, denoised, and error spectra under the noise standard deviation  $\sigma = 0.015, 0.025,$  and  $0.035,$  respectively. Note: The brown long dash lines denote the MAEs of the denoised spectra. The grey fluctuation is the pure noise.

Combing the observations that a higher noise leads to greater error of denoising (Figs. S17 and S18), we can conclude that the standard deviation of noise is the main factor that determines  $\mathbb{E}\|\tilde{\mathbf{X}}\|_2$ . Thus, even with Poisson or Beta distributions, the suggested automatic denoising approach is still very valuable.

## X. THE DENOISING OF RANDOM SIMULATED DATA

We randomly generated three exponential signals with different lengths. Each signal has  $3R + 1$  parameters, including

$R, a_r, f_r$  and  $\tau_r$ , where  $r = 1, 2, \dots, R$ . The number of exponential components is  $R = 4 + M_r$ , where  $M_r$  denotes a pseudo-random scalar integer of range  $[1, 9]$ . The amplitude  $a_r$  is uniformly sampled from  $(0, 10)$ . Each frequency  $f_r$  is uniformly sampled from  $(0, 1)$ . The damping factor is  $\tau_r = 5 + 60m_r$ , where  $m_r$  is uniformly sampled from  $(0, 1)$ .

The denoised spectra in Figs. S19-S22 also confirm our original conclusions. With the increase of noise levels, CHORD provides lower NRMSE (Fig. S19). For the denoised spectra, Cadzow and rQRd lead to missing and weaken peaks that are in low intensity (Black arrows in Figs. S20-S22). CHORD preserves low intensity peaks better.

## REFERENCES

- [1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [2] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [3] J. A. Cadzow, "Signal enhancement-A composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.
- [4] I. Masri and A. Tonge, "Norm estimates for random multilinear Hankel forms," *Linear Algebra and Its Applications*, vol. 402, no. 1, pp. 255–262, 2005.
- [5] B. Kašin and L. Tzafriri, *Lower estimates for the supremum of some random processes, II*. Max-Planck-Institut für Mathematik, 1995.
- [6] J. A. Tropp, "An introduction to matrix concentration inequalities," *Foundations and Trends® in Machine Learning*, vol. 8, no. 1-2, pp. 1–230, 2015.
- [7] J. Ying, J.-F. Cai, D. Guo, G. Tang, Z. Chen, and X. Qu, "Vandermonde factorization of Hankel matrix for complex exponential signal recovery—Application in fast NMR spectroscopy," *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5520–5533, 2018.

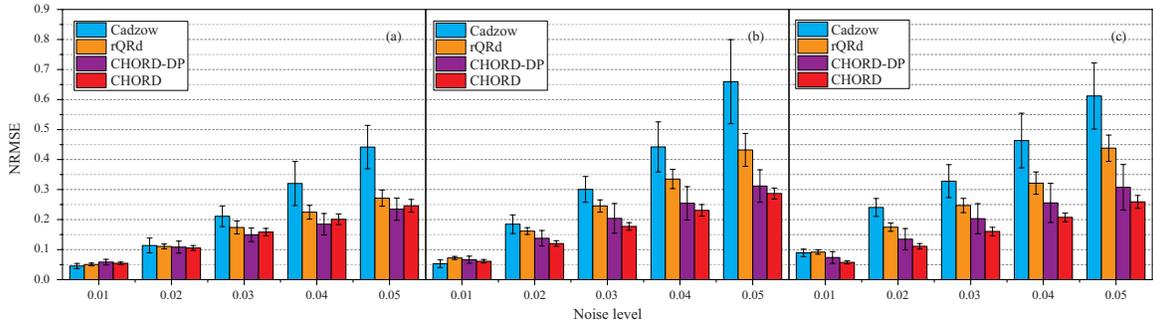


Fig. S19. The NRMSE for random synthetic data under different noise levels. (a)-(c) denote the NRMSE for synthetic data with 256, 512, and 1024 points, respectively. The height of columns shows the average of the NRMSEs over 50 trials. The vertical bar comes from the randomness of noise.

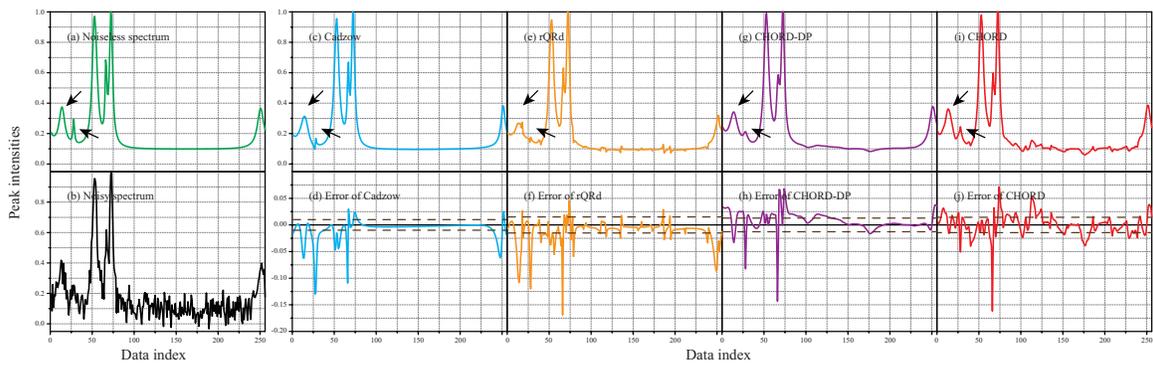


Fig. S20. The denoised spectra of random synthetic data with 256 points. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.02$ ). (c), (e), (g), and (i) are the denoised results of Cadzow, rQRd, CHORD-DP, and CHORD, respectively. (d), (f), (h), and (j) stand for the denoising error of four methods. Note: The results of Cadzow and rQRd that enable the lowest NRMSE are presented here. The brown long dash lines denote the MAEs of the denoised spectra.

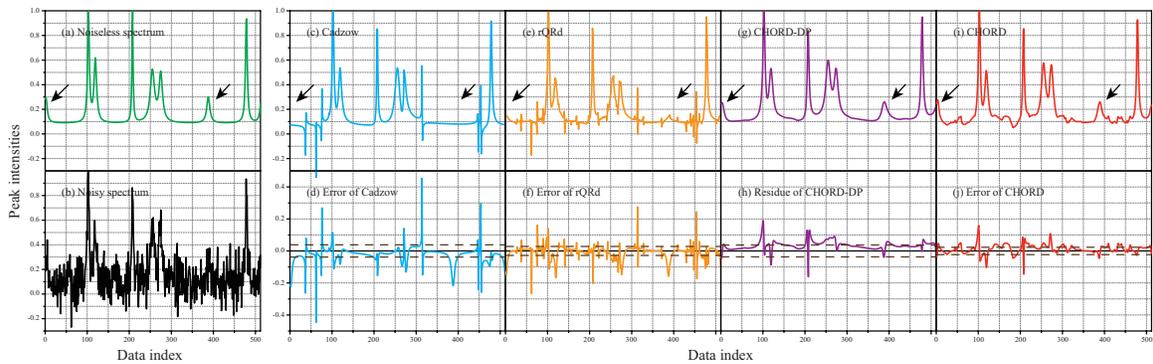


Fig. S21. The denoised spectra of random synthetic data with 512 points. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.03$ ). (c), (e), (g), and (i) are the denoised results of Cadzow, rQRd, CHORD-DP, and CHORD, respectively. (d), (f), (h), and (j) stand for the denoising error of four methods. Note: The results of Cadzow and rQRd that enable the lowest NRMSE are presented here. The brown long dash lines denote the MAEs of the denoised spectra.

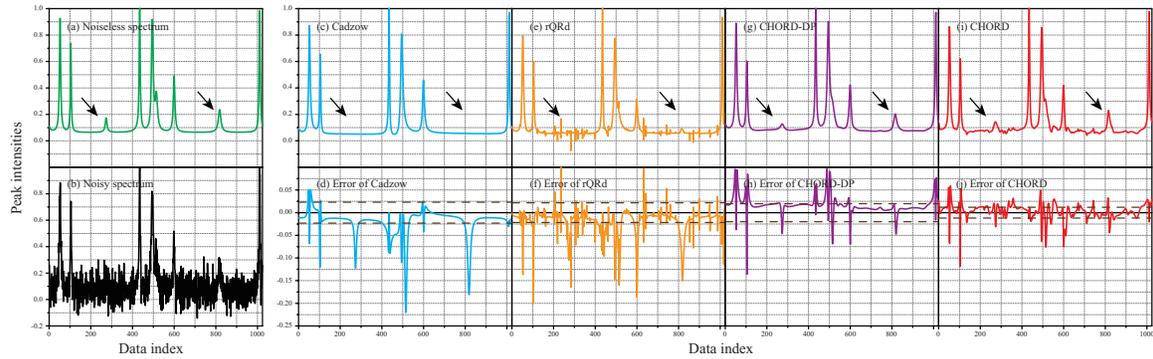


Fig. S22. The denoised spectra of random synthetic data with 1024 points. (a) and (b) denote the synthetic signals without and with noise ( $\sigma = 0.02$ ). (c), (e), (g), and (i) are the denoised results of Cadzow, rQRd, CHORD-DP, and CHORD, respectively. (d), (f), (h), and (j) stand for the denoising error of four methods. Note: The results of Cadzow and rQRd that enable the lowest NRMSE are presented here. The brown long dash lines denote the MAEs of the denoised spectra.