Supplementary Information

Appendix A

Algorithm

The proposed EnhanCed Low Rank (ECLR) model is defined as follows:

$$\min_{\mathbf{x}} \sum_{q=1}^{Q} \phi(\sigma_q(\mathcal{R}\mathbf{x}); a) + \frac{\lambda}{2}\|\mathbf{y} - \mathcal{U}\mathbf{x}\|_2^2, \tag{A1}$$

where $\mathcal{U}$ is an operator of NUS schedule, $\mathbf{y}$ is the acquired NUS data, $\mathcal{R}$ denotes an operator that transforms $\mathbf{x}$ into a Hankel matrix $\mathcal{R}\mathbf{x}$, Q stands for total number of singular values of the Hankel matrix, and λ trades the low-rank approximation with the consistency between the reconstructed signal $\mathbf{x}$ and the acquired NUS data $\mathbf{y}$.

Since a larger a leads to better approximation of rank, i.e. the number of spectral peaks, two loops of iterations are designed in the algorithm. In the outer loop, we progressively increase the values of a and the output of Eq. (A1) with a smaller a will be taken as the initial solution of Eq. (A1) with a larger a. In the inner loop, we solve Eq. (A1) for a fix a according to the following way.

By introducing a variable $\mathbf{Z}$ and a parameter β, Eq. (A1) is rewritten as below

$$\min_{\mathbf{x},\mathbf{Z}} \sum_{q=1}^{Q} \phi(\sigma_q(\mathbf{Z}); a) + \frac{\beta}{2}\|\mathcal{R}\mathbf{x} - \mathbf{Z}\|_F^2 + \frac{\lambda}{2}\|\mathbf{y} - \mathcal{U}\mathbf{x}\|_2^2, \tag{A2}$$

which is alternatively solved by two sub-problems:

For a fixed $\mathbf{Z}$, solve

$$\min_{\mathbf{x}} \frac{\beta}{2}\|\mathcal{R}\mathbf{x} - \mathbf{Z}\|_F^2 + \frac{\lambda}{2}\|\mathbf{y} - \mathcal{U}\mathbf{x}\|_2^2, \tag{A3}$$

whose solution is

$$\mathbf{x} = \left(\beta\mathcal{R}^*\mathcal{R} + \lambda\mathcal{U}^*\mathcal{U}\right)^{-1}\left(\beta\mathcal{R}^*\mathbf{Z} + \lambda\mathcal{U}^*\mathbf{y}\right), \tag{A4}$$

where * means the adjoint operator.

For a fixed $\mathbf{x}$, solve

$$\min_{\mathbf{Z}} \sum_{q=1}^{Q} \phi(\sigma_q(\mathbf{Z}); a) + \frac{\beta}{2}\|\mathcal{R}\mathbf{x} - \mathbf{Z}\|_2^2, \tag{A5}$$

whose solution is

$$\mathbf{Z} = \mathbf{U}\Theta\left(\mathbf{\Sigma}; \beta, a\right)\mathbf{V}^H, \tag{A6}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the left and right unitary matrix in the singular value decomposition of the Hankel matrix $\mathcal{R}\mathbf{x}$=$\mathbf{U\Sigma V}^H$, and the particular thresholding function on the singular value stored in the diagonal entry of $\mathbf{\Sigma}$ is defined as

$$\Theta\left(\mathbf{\Sigma}; \beta, a\right) = \min\left\{\mathbf{\Sigma}, \max\left\{\left(\mathbf{\Sigma} - 2a/\beta\right)\Big/\left(1 - 2a^2/\beta\right), 0\right\}\right\}. \tag{A7}$$

The whole algorithm is summarized in the following flowchart and pseudo code.

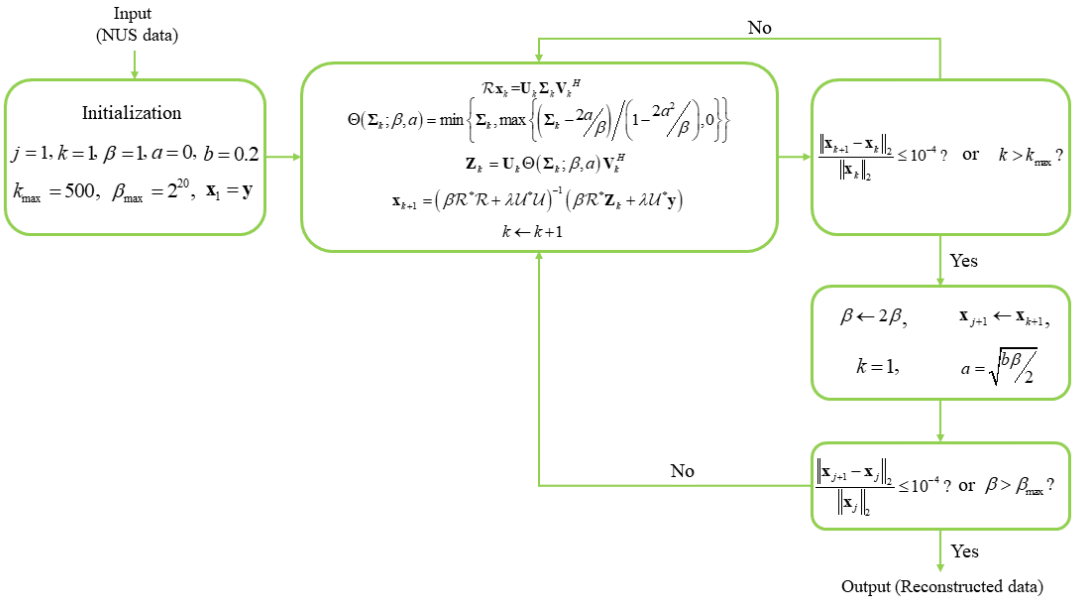

Fig. S1. Flowchart of the proposed ECLR method.

Pseudo code of the algorithm

Input: $\mathbf{y}$, λ, $\mathcal{R}$, $\mathcal{U}$; **Output**: $\hat{\mathbf{x}}$

Initialization: $j=1$, $k=1$, $\beta=1$, $a=0$, $b=0.2$, $k_{\max}=500$, $\beta_{\max}=2^{20}$, $\mathbf{x}_1=\mathbf{y}$

1) While $\beta_j<\beta_{\max}$ and $\dfrac{\|\mathbf{x}_j-\mathbf{x}_{j-1}\|_2}{\|\mathbf{x}_{j-1}\|_2}>10^{-4}$, **do**

2) **While** $k<k_{\max}$ and $\dfrac{\|\mathbf{x}_k-\mathbf{x}_{k-1}\|_2}{\|\mathbf{x}_{k-1}\|_2}>10^{-4}$, **do**

3) $\mathcal{R}\mathbf{x}_k=\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^H$;

4) $\mathbf{Z}_k=\mathbf{U}_k\Theta(\mathbf{\Sigma}_k;\beta,a)\mathbf{V}_k^H$;

5) $\mathbf{x}_{k+1}=\left(\beta\mathcal{R}^*\mathcal{R}+\lambda\mathcal{U}^*\mathcal{U}\right)^{-1}\left(\beta\mathcal{R}^*\mathbf{Z}_k+\lambda\mathcal{U}^*\mathbf{y}\right)$;

6) $k\leftarrow k+1$;

7) **End while**

8) $k=1$, $\beta\leftarrow 2\beta$, $a=\sqrt{b\beta/2}$, $\mathbf{x}_{j+1}\leftarrow\mathbf{x}_{k+1}$;

9) **End while**

Output: $\hat{\mathbf{x}}=\mathbf{x}_{j+1}$

Appendix B

Experiment for the 2D ^{1}H-^{15}N HSQC spectrum

300 μM ^{15}N-^{13}C labeled sample of cytosolic CD79b in 20 mM sodium phosphate buffer, pH 6.7 was prepared as described previously [1]. Fully sampled 2D ^{1}H-^{15}N HSQC with 256 complex points in the ^{15}N dimension (143.5 ms acquisition time) was acquired at 55 °C on 800 MHz Bruker AVANCE III HD spectrometer equipped with 3 mm CPTCI cryoprobe. The directly detected dimension of the amide region of the full reference 2D spectrum (8.75-7.85 ^{1}H ppm) was processed using NMRPipe software [2] and imported in MATLAB for consecutive reconstruction by the ECLR and LR [3] methods. The 25% sparse non-uniform Poisson-gap sampling table was produced according to reference [4].

Experiment for the 2D ^{1}H-^{1}H double-quantum solid NMR spectrum

The solid-state NMR spectrum was acquired from an imidazole sample at room temperature on a 21.1 T Bruker AVANCE-III spectrometer. The sample was spinning at 60 kHz inside a 1.3 mm probe. The 2D ^{1}H-^{1}H double-quantum spectrum was recorded using a symmetry-based R1225 pulse sequence [5] with a recycle delay of 5 s. Data were recorded by co-adding 16 transients with 300 t_1 increments. The number of fully sampled data points in the indirect dimension is 150. The 15% sparse non-uniform Poisson-gap sampling table that was produced according to reference [4].

Appendix C

Low rank reconstruction of STEU 2D COSY spectrum

Experimental Section

Fully sampled STEU 2D ^{1}H-^{1}H COSY [6,7] (Correlation Spectroscopy) spectrum of a special oil with 800 complex points in the direct dimension was acquired on a Varian 500 MHz NMR System (Agilent Technologies, Santa Clara, CA, USA) equipped with a 5-mm indirect detection probe. The oil sample exists in intrahepatic fat of liver, quantification of its spectrum serves as an indispensable tool in clinical diagnose [8]. In this STEU COSY spectrum, signals are uniformly sampled with the acquisition parameters as follows: encoding gradients = 3.91 G/cm, decoding gradients = 48.8 G/cm, compensative gradient = −7.81 G/cm, duration of each decoding gradient lobe = 220 us, duration of the chirp pulse = 12 ms, duration of compensative gradient = 0.5 ms, and number of alternating gradient pairs = 150. The WURST profile with the sweep frequency range of 30 KHz in 6 ms.

The fully sampled reference 2D spectrum (1.0-5.0 ^{1}H ppm) was imported in MATLAB for reconstruction. Retrospective NUS is applied to evaluate the reconstruction performance. The 20% sparse non-uniform gradient switching-based sampling pattern was produced according to STEU technology [9]. The pattern was generated with a pseudo random oscillating gradient used in pulse sequence, and the wavenumber k is obtained according to

$$k = \int_0^t \gamma_a G_a\left(t^{'}\right) dt^{'}, \tag{A8}$$

where t denotes the pulse duration, γ_a the gyromagnetic ratio and G_a the acquisition gradient. In the NUS implementation, t =152 ms, γ_a = 4.25 and G_a =8 G/cm. After obtaining the $k-t$ data trajectories, we arranged data points (off-grid) into a data matrix (800×100), of which sampled data point locations set as one while others set as zeros.

Spectrum Quantification

Accuracy of the cross-peak volume is adopted as the criteria for method evaluation because the resonance volumes reflect the ratio between different components in the intrahepatic fat, which earns importance in diagnosis of liver disease [8].

We assigned spectral peaks in STEU 2D COSY to protons according to reference work [8]. The peak resonance volumes integrated over the pixel intensities were normalized to the sum of volumes of all the peaks (total 100) in the spectrum. We carried out 10 Monte Carlo trials with different sampling patterns using LR methods, average resonance volumes for every peak and calculated the standard deviations. The results are presented in Table S1 in which the peaks are sorted by their volumes.

To further measure the quantification, a normalized quantification of volume error (NQVE) is defined as

$$\text{NQVE}=\frac{|v-\hat{v}|}{v}, \tag{A9}$$

where v and $\hat{v}$ denote the peak volume in reference and reconstructed spectra, respectively. The smaller NQVE is, the more accurate quantification is obtained from the reconstruction. The NQVE of peak #1~#9, with a descending order of their volumes listed in Table S1, are analyzed in Figure 3(d) of the paper. It shows that ECLR achieves consistently higher accuracy for all cross-peaks, with the largest improvement for the low intensity signals.

Table S1. QUANTITATIVE ANALYSIS OF STEU 2D SPECTRUM

Peak identity information			Peak volume		
ID	Chemical group	ppm	Reference	LR	ECLR
1	$-(CH_2)_n-$	1.30	37.69	39.72±0.50	37.69±0.55
2	$-(CH_2)_n-CH_3$	0.90	16.79	17.53±0.22	16.81±0.17
3	-CH = CH-	5.29	12.98	11.88±0.30	12.91±0.21
4	$-CH_2-CH = CH-CH_2-$	2.02	12.08	11.81±0.74	12.27±0.14
5	$-CO-CH_2-CH_2-$	2.24	8.16	7.94±0.43	8.26±0.16
6	$=CH-CH_2-CH =$	2.75	5.23	4.59±0.33	5.13±0.14
7	$-CO-CH_2-CH_2-$	1.60	4.78	4.46±0.18	4.70±0.06
8	$-CH_2$	4.30	1.19	1.14±0.10	1.17±0.03
9	-CH-	4.05	1.10	0.93±0.14	1.06±0.03

References

[1] L. Isaksson, M. Mayzel, M. Saline, A. Pedersen, J. Rosenlöw, B. Brutscher, B. G. Karlsson, V. Y. Orekhov, *PLoS ONE,* 2013, **8**, e62947.

[2] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR,* 1995, **6**, 277-293.

[3] X. Qu, M. Mayzel, J.-F. Cai, Z. Chen, V. Orekhov, *Angew. Chem. Int. Ed.,* 2015, **54**, 852-854.

[4] S. G. Hyberts, K. Takeuchi, G. Wagner, *J. Am. Chem. Soc.,* 2010, **132**, 2145-2147.

[5] M. Carravetta, M. Edén, O. G. Johannessen, H. Luthman, P. J. E. Verdegem, J. Lugtenburg, A. Sebald and M. H. Levitt, J Am Chem Soc, 2001, 123, 10628-10638.

[6] L. Frydman, T. Scherf, A. Lupulescu, *P. Natl. Acad. Sci.,* 2002, **99**, 15858-15862.

[7] Y. Shrot, L. Frydman, *J. Am. Chem. Soc.,* 2003, **125**, 11385-11396.

[8] G. Hamilton, T. Yokoo, M. Bydder, I. Cruite, M. E. Schroeder, C. B. Sirlin, M. S. Middleton, *NMR Biomed.,* 2011, **24**, 784-790.

[9] Y. Shrot, L. Frydman, *J. Magn. Reson.,* 2011, **209**, 352-358.