ChemComm

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: X. Qu, T. Qiu, D. Guo, H. Lu, J. Ying, M. Shen, B. Hu, V. Orekhov and Z. Chen, *Chem. Commun.*, 2018, DOI: 10.1039/C8CC06132G.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemcomm

View Article Online
DOI: 10.1039/C8CC06132G

ChemComm

COMMUNICATION

High-fidelity Spectroscopy Reconstruction in Accelerated NMR

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/ chemcomm

Xiaobo Qu*[a], Tianyu Qiu[a], Di Guo[b], Hengfa Lu[a], Jiaxi Ying[a], Ming Shen[c], Bingwen Hu[c], Vladislav Orekhov[d], and Zhong Chen[a]

Abstract: Non-uniform sampling significantly accelerates the data acquisition time in NMR spectroscopy, but spectra must be reconstructed with proper methods. A high-fidelity reconstruction method is proposed to preserve low-intensity spectral peaks and provide stable reconstruction under different sampling trials.

High-resolution NMR spectroscopy plays an important role in modern biochemical analysis, such as characterizing complex protein structures[1], studying short-lived molecular systems [2] and monitoring real time chemical reactions [3, 4]. The data acquisition time of NMR, however, increases rapidly with spectral resolution and dimensionality. To reduce the measurement time, Non-Uniform Sampling (NUS) acquires fewer data and incorporates computational models to reconstruct the spectra[5-11].

A fundamental feature of NUS-NMR is the trade-off between the spectra quality and acceleration factor. In order to obtain a good spectrum, one needs to use both the optimal sampling scheme, e.g. Poisson-gap sampling [12] or spatiotemporally encoded ultrafast (STEU) NMR [8, 13], and an optimal spectra reconstruction algorithms. The later often differ in their prior assumptions on the spectrum and/or time-domain signal. Examples of priors are maximum entropy[9], spectrum sparsity in CLEAN[7] and compressed sensing [8, 14-19], spectral line-shape estimation in SMLIE [11], and tensor structures in MDD [6] or Hankel tensors[20].

A recent progress was a low rank (LR) approach[10] that models the time-domain NMR signal, also called free induction decay (FID), as a sum of a few decaying sine waves (or exponentials) [20-23]. From this model, it follows that the FID can be arranged into a Hankel matrix whose rank is equal to the number of spectral peaks. Assuming that the number of peaks is much smaller than the number of FID data points, the spectrum can be faithfully reconstructed by enforcing low rank of the Hankel matrix [10, 20, 22]. It was shown that the low rank approach can outperform the state-of-the-art compressed sensing algorithm on preserving low intensity broad peaks, and thus increases effective sensitivity in the reconstructed spectra [10].

However, at sufficiently high acceleration factor, the LR fails to reproduce high quality spectrum and the question arises if a better algorithm can be designed to reduce the sampling requirement? Figure 1 shows an obvious peak distortion in the LR spectrum, when only 10% data (acceleration factor is 10) are sampled in synthetic signal. An intuitive explanation of LR failure is that its objective function[10], nuclear norm defined as the sum of singular values [24], is a relatively poor approximation of the rank function, which is the count of the spectral peaks. As an example shown in Figure 1(a), the true rank function allows all the nonzero singular values of a matrix have equal contributions, but the nuclear norm used in the LR [10] treats the singular values differently . As a result, the LR has the tendency to discard or weaken small singular values [25] when minimizing the nuclear norm. This means that low intensity peaks may be lost or compromised in the reconstruction since a small singular value usually corresponds to a low intensity peak.

In this work, we introduce a better approximation of the rank function [26] as shown in Figure 1(a), and a new reconstruction model is proposed to infer the NUS data. A rescaling function ϕ of the singular values is defined as

$$\phi(\sigma_q(\mathbf{X});a)=\begin{cases}|\sigma_q|-\frac{a}{2}\sigma_q^{\,2}, & |\sigma_q|\le\frac{1}{a}\\ \frac{1}{2a}, & |\sigma_q|\ge\frac{1}{a}\end{cases}, \qquad (1)$$

where σ_q denotes the q^{th} singular value of a matrix $\mathbf{X}$ and a is a parameter that controls the closeness to the rank. As shown in Figure 1(a), as a gets larger, ϕ better approaches the rank and consequently the number of peaks. By incorporating the Hankel matrix of FID $\mathbf{x}$, an ultimate reconstruction model is

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

[a] Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, Xiamen University, Xiamen 361005, China.
[b] School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China.
[c] Shanghai Key Laboratory of Magnetic Resonance, Department of Physics, East China Normal University, Shanghai 200062, China.
[d] Department of Chemistry and Molecular Biology and Swedish NMR Centre, University of Gothenburg, Box 465, Gothenburg 40530, Sweden.
*Corresponding author: quxiaobo@xmu.edu.cn
†Electronic Supplementary Information (ESI) available. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/C8CC06132G

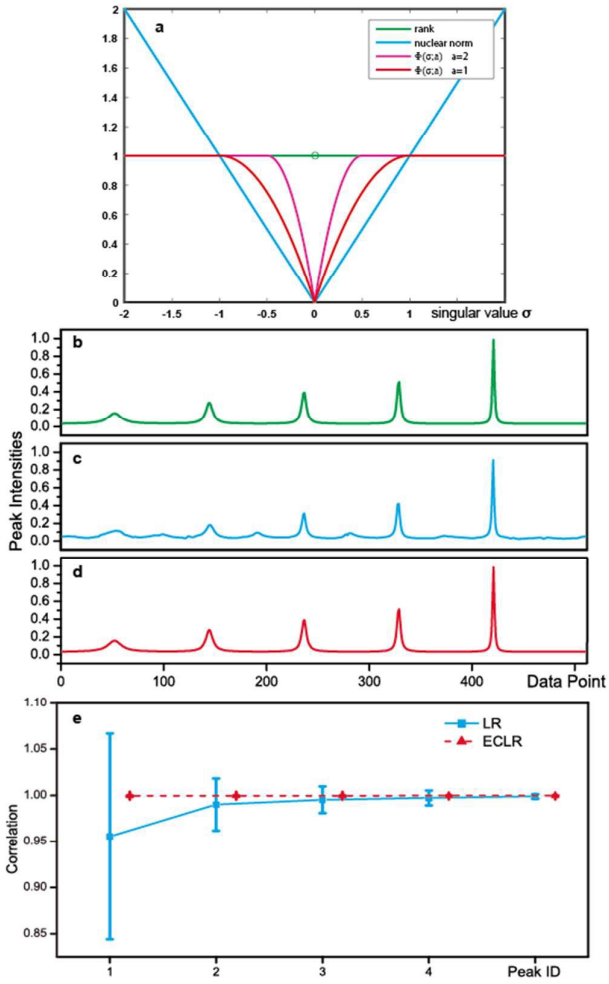

Fig. 1 Functions approximating the matrix rank and the reconstructions of the synthetic spectrum containing five peaks with different line widths. (a) blue, and red curves represent approximation of the matrix rank (green) with nuclear norm and the new function (Eq. 1), respectively; (b) the reference fully sampled spectrum, (c) and (d) are the LR and ECLR reconstructions, respectively, obtained from 10% NUS, (e) correlation analysis of spectral intensities in small regions, which are around the peaks (3 times the line width) between the full reference spectrum and the LR (c) and ECLR (d) reconstructions. The error bars are the standard deviations of the correlations over 100 NUS resampling trials.

where U is an operator of NUS schedule, $\mathbf{y}$ is the acquired NUS data, R denotes an operator that transforms $\mathbf{x}$ into a Hankel matrix $\mathrm{R}\mathbf{x}$, Q stands for total number of singular values of the Hankel matrix, and λ trades low rank approximation of the reconstruction $\mathbf{x}$ with consistency to acquired NUS data $\mathbf{y}$. In implementation, we start from a small a and then increases its value to achieve better spectroscopy reconstruction. The whole algorithm is summarized in supplementary Appendix A. In the following description, this new approach is called EnhanCed Low Rank (ECLR) method.

Figure 1 shows a comparison between a synthetic fully sampled spectrum composed of five peaks with different widths and reconstructions obtained using the LR and the proposed ECLR. Low intensity peaks (peaks #1 and #2 in Figure 1(c)), which are seriously distorted by LR, are reconstructed much better by ECLR (Figure 1(d)). This improvement has been further confirmed by the higher correlation of low intensity spectral peaks (Figure 1(e)). Besides, correlations of other relatively stronger peaks (peak #3, #4, #5 in Figure 1(e)) have also been increased with ECLR, implying that the new approach can simultaneously improve the reconstruction of all peaks. What is more, much lower error bars achieved by ECLR indicate that the new method is more robust to different NUS trials, leading to more stable reconstruction. These observations imply that the ECLR has advantages on preserving low intensity peaks and providing more stable reconstruction.

Figure 2 shows reconstructed 2D ^{1}H-^{15}N HSQC spectrum of the intrinsically disordered cytosolic domain of human CD79b protein from B-cell receptor. When only very limited data are available in NUS (25% of full data in this case), LR produced artefact peaks (peak D in Figure 2(b)), weakened or lost low intensity peaks (blue dashed lines in Figure 2(e) and (f)). The reconstruction by ECLR is significantly better. Particularly, weak peaks are restored (red lines in Figure 2(e) and (f)). The limitation of LR on recovering low intensity peaks was also observed in another experiment of the 2D ^{1}H-^{1}H double-quantum solid NMR spectrum (Figure 3(b) and (c)). Peak intensity regression evaluation (Figure 3(d) and (e)) on the solid NMR spectra further demonstrated that the spectrum obtained by ECLR was more consistent to the reference spectrum than that obtained by LR. Thus, ECLR provides more faithful reconstruction and can achieve higher practical sensitivity in fast NMR with NUS.

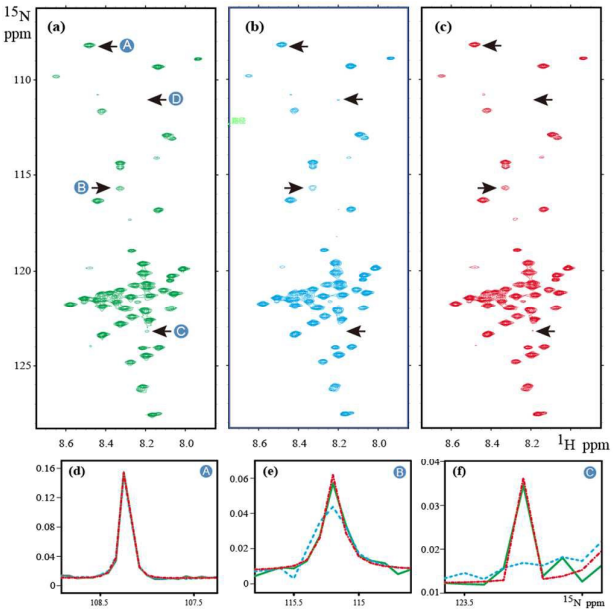

Fig. 2. 2D ^{1}H-^{15}N HSQC spectrum of the cytosolic domain of CD79b protein from B-cell receptor. (a) the fully sampled reference spectrum, (b) and (c) are the reconstructions using LR and ECLR from 25% NUS data, (d)-(f) are zoom out 1D ^{15}N traces, and green, blue, and red lines represent reference, the LR and ECLR spectra, respectively.

View Article Online
DOI: 10.1039/C8CC06132G

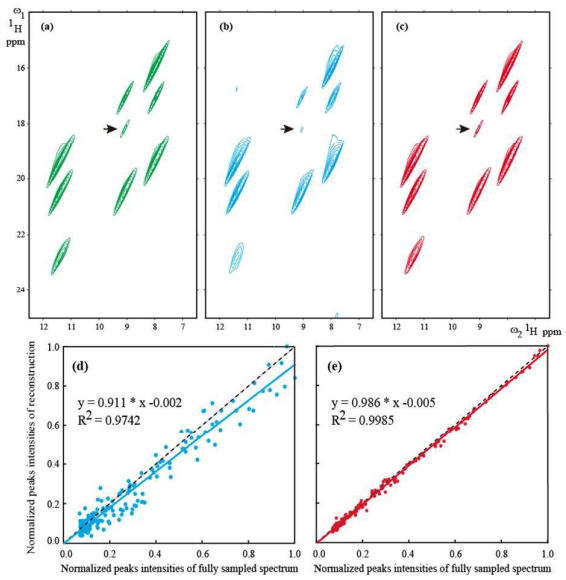

Fig. 3. Reconstruction of a solid NMR spectrum. (a) is the fully sampled reference spectrum, (b) and (c) are reconstructed spectra from 15% NUS data by LR and ECLR, respectively, (d) and (e) are the peak intensity correlations achieved by LR and ECLR, respectively. Note: The notation R^2 denotes Pearson linear correlation coefficient of fitted curve. The closer that the value of R^2 gets to 1, the stronger the correlation between reference and reconstructed spectra is.

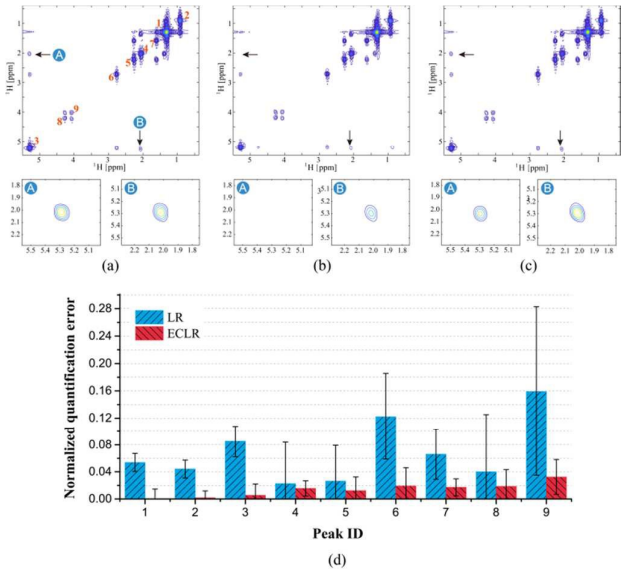

Fig. 4. Reconstruction of a STEU spectrum. (a) The full sampled STEU COSY spectrum, (b) and (c) are reconstructed spectra from 20% NUS data by LR and ECLR, respectively, (d) normalized quantification of the volume errors. Note: Peak #1~#9 are descending sorted following their volumes listed in Table S1 of the supplement. The error bars in (d) represent the deviation of the normalized quantification error from 10 different NUS trials.

Beyond the typical NUS that reduces the data acquisition time in indirect NMR dimensions, the STEU enables the ultrafast spectra acquisition within seconds [8]. The acquired signals, however, may need to be recovered when NUS is applied to reduce the instrumental gradient requirement [8]. Although compressed sensing has been applied to restore the missing data, it was recently found that the original LR can reconstruct the spectrum much better [27]. Then, extending ECLR into STEU NMR, will be valuable since weak signals may be reconstructed much better.

Figure 4 shows the reconstruction of STEU COSY spectrum obtained from a liver fat sample. The weak peaks haven been restored much better when using ECLR (Figure 4(c)) than LR (Figure 4(b)). To demonstrate more biomedical meanings of the new algorithm, we quantify volumes of the spectral cross-peak that reflect the ratio of different components in the intrahepatic fat, and thus carry important information for diagnostics of liver diseases [28]. Statistical analysis of the peak volumes is listed in the Table S1 of the supplement. The normalized quantification of the volume errors (Figure 4(d)) implies that ECLR achieves consistently higher accuracy for all cross-peaks, with the largest improvement for the low intensity signals. Thus, spectra reconstructed by ECLR provides much better data for defining the ratio of different components in the intrahepatic fat. Furthermore, the much lower error bars imply more stable reconstruction under different NUS sampling trials.

We introduce a new approach for reconstruction of NMR spectra from very limited number of NUS data points. The method allows faithful and stable spectra reconstruction with the main improvement for low intensity peaks. Reconstruction on the spectrum of a liver fat sample clearly demonstrate that peaks with small volumes have been quantitatively measured much better. This approach is particularly valuable to fast decaying NMR signals, thus may be very useful for short living systems, time resolved experiments, and many other practical cases.

This work was partially supported by National Natural Science Foundation of China (61571380, 61871341, 61811530021, U1632274, 61672335 and 61601276), National Key R&D Program of China (2017YFC0108703), Natural Science Foundation of Fujian Province of China (2018J06018 and 2016J05205), Science and Technology Program of Xiamen (3502Z20183053), Fundamental Research Funds for the Central Universities (20720180056), the Swedish Research Council (2015-04614), STINT (Sweden) grant CH2017-7231. We thank Dr. L. Isaksson for preparing the CD79b sample.

Notes and references

1. M. W. Maciejewski, A. D. Schuyler and J. C. Hoch, in *Protein NMR: Methods and Protocols*, ed. R. Ghose, Springer New York, New York, 2018, pp. 341-352.
2. F.-X. Theillet, H. M. Rose, S. Liokatis, A. Binolfi, R. Thongwichian, M. Stuiver and P. Selenko, *Nat. Protoc.*, 2013, **8**, 1416-1432.
3. E. Rennella, T. Cutuil, P. Schanda, I. Ayala, V. Forge and B. Brutscher, *J. Am. Chem. Soc.*, 2012, **134**, 8066-8069.
4. M. Mayzel, J. Rosenlow, L. Isaksson and V. Y. Orekhov, *J. Biomol. NMR*, 2014, **58**, 129-139.
5. M. Billeter, *J. Biomol. NMR*, 2017, **68**, 65-66.
6. V. Jaravine, I. Ibraghimov and V. Y. Orekhov, *Nat. Methods*, 2006, **3**, 605-607.

View Article Online
DOI: 10.1039/C8CC06132G

7. B. E. Coggins and P. Zhou, *J. Biomol. NMR*, 2008, **42**, 225-239.
8. Y. Shrot and L. Frydman, *J. Magn. Reson.*, 2011, **209**, 352-358.
9. J. C. Hoch, M. W. Maciejewski, M. Mobli, A. D. Schuyler and A. S. Stern, *Accounts. Chem. Res.*, 2014, **47**, 708-717.
10. X. Qu, M. Mayzel, J.-F. Cai, Z. Chen and V. Orekhov, *Angew. Chem. Int. Edit.*, 2015, **54**, 852-854.
11. J. Ying, F. Delaglio, D. A. Torchia and A. Bax, *J. Biomol. NMR*, 2017, **68**, 101-118.
12. S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari and G. Wagner, *J. Biomol. NMR*, 2012, **52**, 315-327.
13. L. J. Lin, Z. L. Wei, Y. Q. Lin and Z. Chen, *Chem. Commun.*, 2015, **51**, 1234-1236.
14. X. Qu, X. Cao, D. Guo and Z. Chen, *Proc. of ISMRM*, Stockholm, Sweden, **2010**, 3371.
15. X. Qu, D. Guo, X. Cao, S. Cai and Z. Chen, *Sensors*, 2011, **11**, 8888-8909.
16. D. J. Holland, M. J. Bostock, L. F. Gladden and D. Nietlispach, *Angew. Chem. Int. Edit.*, 2011, **50**, 6548-6551.
17. K. Kazimierczuk and V. Y. Orekhov, *Angew. Chem. Int. Edit.*, 2011, **50**, 5556-5559.
18. M. Mayzel, K. Kazimierczuk and V. Y. Orekhov, *Chem. Commun.*, 2014, **50**, 8947-8950.
19. Y. Wu, C. D'Agostino, D. J. Holland and L. F. Gladden, *Chem. Commun.*, 2014, **50**, 14137-14140.
20. J. Ying, H. Lu, Q. Wei, J. F. Cai, D. Guo, J. Wu, Z. Chen and X. Qu, *IEEE T. Signal Proces.*, 2017, **65**, 3702-3717.
21. H. M. Nguyen, X. Peng, M. N. Do and Z.-P. Liang, *IEEE T. Bio.-Med. Eng.*, 2013, **60**, 78-89.
22. D. Guo, H. Lu and X. Qu, *IEEE Access*, 2017, **5**, 16033-16039.
23. D. Guo and X. Qu, *IEEE Access*, 2018, **6**, 4933-4940.
25. J. -F. Cai, E. Candès and Z. Shen, *SIAM J. Optimiz.*, 2010, **20**, 1956-1982.
25. S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, L. Zhang, *Int. J. Comput. Vision*, 2017, **121**, 183-208.
26. A. Parekh and I. W. Selesnick, *IEEE Signal Proc. Let.*, 2016, **23**, 493-497.
27. H. Lu, X. Zhang, T. Qiu, J. Yang, J. Ying, D. Guo, Z. Chen and X. Qu, *IEEE T. Bio.-Med. Eng.*, 2018, **65**, 809-820.
28. G. Hamilton, T. Yokoo, M. Bydder, I. Cruite, M. E. Schroeder, C. B. Sirlin and M. S. Middleton, *NMR Biomed.*, 2011, **24**, 784-790.

View Article Online
DOI: 10.1039/C8CC06132G

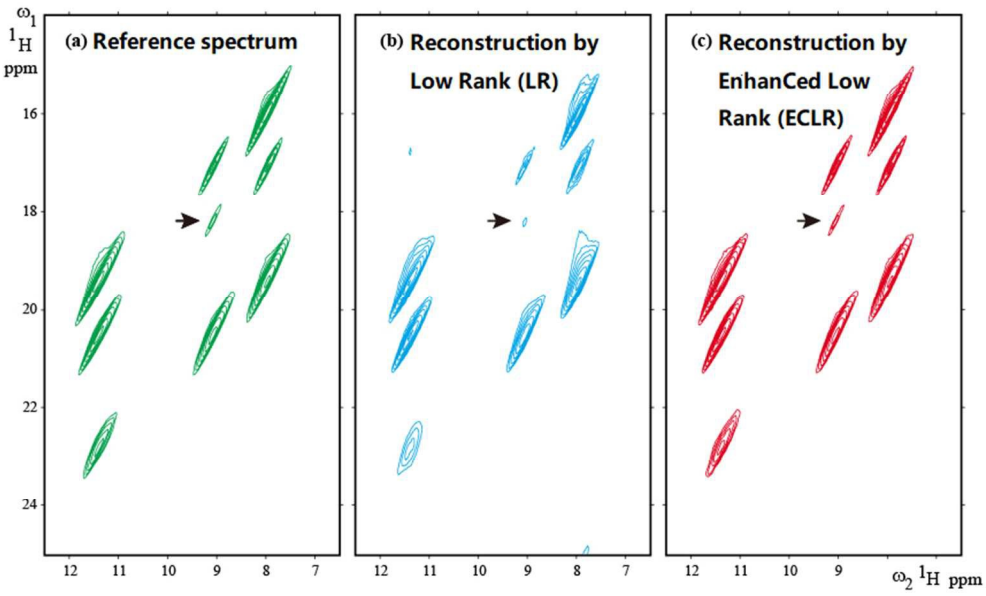

Figure: Table Of Contents Entry.
High-fidelity spectra, particularly low intensity peaks, are reconstructed for fast NMR with better rank approximation in EnhanCed Low Rank (ECLR).