NMR Spectroscopy

DOI: 10.1002/anie.201409291

Accelerated NMR Spectroscopy with Low-Rank Reconstruction**

Xiaobo Qu, Maxim Mayzel, Jian-Feng Cai, Zhong Chen, and Vladislav Orekhov**

Abstract: *Accelerated multi-dimensional NMR spectroscopy is a prerequisite for high-throughput applications, studying short-lived molecular systems and monitoring chemical reactions in real time. Non-uniform sampling is a common approach to reduce the measurement time. Here, a new method for high-quality spectra reconstruction from non-uniformly sampled data is introduced, which is based on recent developments in the field of signal processing theory and uses the so far unexploited general property of the NMR signal, its low rank. Using experimental and simulated data, we demonstrate that the low-rank reconstruction is a viable alternative to the current state-of-the-art technique compressed sensing. In particular, the low-rank approach is good in preserving of low-intensity broad peaks, and thus increases the effective sensitivity in the reconstructed spectra.*

Achieving high spectral resolution and sensitivity while keeping a minimal measurement time is of primary importance in many NMR applications such as studies of short-lived molecular systems, in-cell NMR experiments,[1] characterizing intermediates of chemical reactions in real time,[2] high-throughput and metabolomic[3] applications. The duration of a multi-dimensional NMR experiment is proportional to the number of measured data points and increases rapidly with spectral resolution and dimensionality. The non-uniform sampling (NUS)[4] approach offers a general solution for a dramatic reduction in measurement time.

The reconstruction of a spectrum from a non-uniformly sampled signal is impossible without introducing additional constraints or assumptions on the signal in the time or frequency domains. Apart from computational issues such as convergence and stability in respect to noise and spectral artefacts, success of an algorithm in reconstructing a high-quality spectrum depends largely on the correctness of the used constraints that are derived from known NMR signal properties. For example, the compressed sensing (CS) approach[5] exploits a reasonable notion that an NMR spectrum is sparse,[6] that is, that only a few frequencies give rise to true peaks while the rest of the spectral space contains only baseline noise. Despite the evident success of CS,[3,6] it was recently noted that sparseness of the NMR spectrum is a crude approximation and the quality of the CS processed spectrum obtained from NUS data can be significantly improved by additional signal pre-processing steps such as zero filling or virtual echo.[7] In particular, spectra with broad lines deviate from the sparseness assumption made in CS and line shape distortions and even suppression of signals might be observed. Hence, the question remains, can an algorithm, based on alternative general signal property assumptions, provide NUS spectra reconstructions of even higher quality and from fewer acquired data points?

In this work, we introduce low-rank (LR) spectrum reconstruction that is inspired by recent developments in the field of low-rank matrix completion with many remarkable demonstrations in medical imaging,[8] computer vision,[9] and other applications. It has been proven that a low-rank matrix can be recovered faithfully from the limited number of its elements.[10] Unlike CS, which seeks for a spectrum with the fewest non-zero spectral intensities, the LR approach reconstructs a spectrum with the least number of spectral peaks. The latter assumes that the time-domain NMR signal can be approximated by a sum of a few decaying sine waves (or exponentials). This assumption has been applied to signal processing in NMR spectroscopy for decades, for example, in linear prediction,[11] filter-diagonalization,[12] recursive multi-dimensional decomposition,[13] and other algorithms. Yet, to the best of our knowledge, this signal property was never used as a constraint in non-parametric algorithms for the reconstruction of NUS spectra.

Let vector $\mathbf{x}$ be the complete NMR signal that is called the free induction decay (FID) and the operator $\mathbf{R}$ converts it into a Hankel matrix[11] $\mathbf{X} = \mathbf{Rx}$. The low rank of the FID means that the rank of its Hankel matrix $\mathbf{X}$ is low, that is, the number of non-zero singular values of $\mathbf{X}$ is small. As it is illustrated in Figure S1 in the Supporting Information, the FID rank is equal to the number of exponentials in $\mathbf{x}$.[11] Notably, the rank is independent of the line width of the peaks in the spectrum. When the FID is non-uniformly sampled, the matrix $\mathbf{X}$ contains missing entries. Thus, the task of the spectrum reconstruction from the NUS data is equivalent to recovering the matrix $\mathbf{X}$ and can be formulated as a low-rank matrix completion problem [Eq. (1)]:[10]

[*] Prof. X. Qu, Prof. Z. Chen
Department of Electronic Science
Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance
State Key Laboratory of Physical Chemistry of Solid Surfaces
Xiamen University, P.O. Box 979, Xiamen 361005 (China)
E-mail: quxiaobo@xmu.edu.cn
Dr. M. Mayzel, Prof. V. Orekhov
Swedish NMR Centre, University of Gothenburg
P.O. Box 465, Gothenburg 40530 (Sweden)
E-mail: vladislav.orekhov@nmr.gu.se
Prof. J. F. Cai
Department of Mathematics, University of Iowa
Iowa City, IA 52242 (USA)

[**] The work was supported by the National Natural Science Foundation of China (61201045, 11375147); Fundamental Research Funds for the Central Universities (PRC) (2013SH002); The Swedish Research Council (grant 2011-5994); US National Science Foundation (DMS-1418737). X.Q. is grateful to the Wenner-Gren Foundation. We thank Dr. L. Isaksson for preparing the CD79b sample.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201409291.

$$\min_{\mathbf{x}} \|\mathbf{Rx}\|_* + \frac{\lambda}{2}\|\mathbf{y} - \mathbf{Ux}\|_2^2 \quad (1)$$

where **y** is the acquired NUS FID data, **U** is an operator of the NUS schedule, $\|\mathbf{Rx}\|_*$ is the nuclear norm[14] defined as a sum of matrix singular values, and λ trades the low-rank constraint with the consistency between the reconstructed signal **x** and the experimental data **y**. It is worth noting that the LR constraint in Equation (1) is imposed on the time-domain signal. Equation (1) can be efficiently solved by the alternating direction minimization method[15] (see the Supporting Information).

Figure 1 shows a comparison between a simulated fully sampled reference spectrum and its NUS reconstructions obtained using the CS and LR algorithms. The spectrum

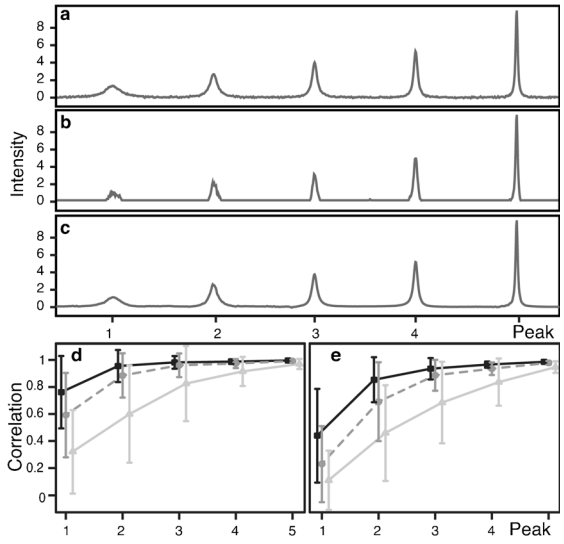

Figure 1. Reconstructions of the synthetic spectrum containing five peaks with different line widths. a) A fully sampled spectrum. b,c) The CS and LR reconstructions, respectively, obtained from 20% NUS. Correlation analysis of spectral intensities in small regions around the peaks (3 times the line width) between the full reference spectrum and the LR (d) and CS (e) reconstructions. The black, dashed gray, and solid gray lines in (d) and (e) connect results of the reconstructions using 20, 15, and 10% NUS, respectively. The error bars are the standard deviations of the correlations over 100 NUS resampling trials.

contains five peaks with the same integrals but different line widths. Both NUS processing methods successfully recover the narrowest peak to the right in Figure 1a–c. The broadest peak to the left is faithfully recovered by the LR approach but is seriously distorted by the CS. For the three middle peaks with moderate line width, the CS produces clearly visible line shape distortions as shrinkage of the peaks. Correlation analysis of the spectral intensities, shown in Figure 1d,e indicates that, for the NUS level in the range 10–20%, the two broadest peaks are recovered systematically better using the LR than by the CS. For the three remaining narrower peaks, the LR and CS provide comparable results. These observations imply that the LR, while performing similarly for the narrow peaks, outperforms CS when the peaks are relatively broad.

This effect can be explained by using the basic CS theorem, binding the number of properly reconstructed spectral points, which is essentially a measure of spectrum sparseness, with the sampling level.[5b] For broad peaks, more points contribute to each signal in the spectrum and thus more data points are needed to fulfil the condition for a successful CS reconstruction. On the other hand, the rank of the FID signal is independent on the line width, and thus the LR produces correct reconstruction of line shapes for both sharp and broad peaks.

Figure 2 shows a NUS 2D ^{1}H-^{15}N HSQC spectrum of the intrinsically disordered cytosolic domain of human CD79b protein from the B-cell receptor. Similarity between the LR

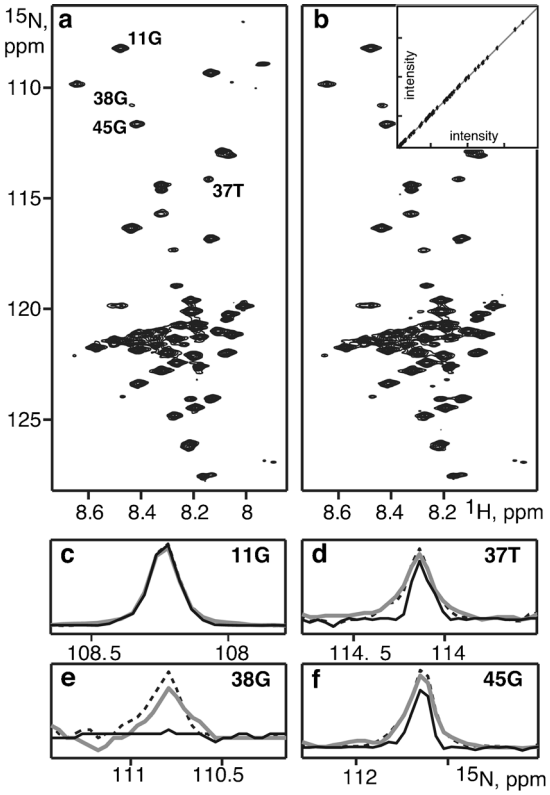

Figure 2. 2D 1H-15N HSQC spectrum of the cytosolic domain of CD79b. a) The LR reconstruction from 35% NUS data, b) the fully sampled reference spectrum. The inset shows correlation of the peak intensities between the reference and the LR spectra; the correlation coefficient equals to 0.99. c–f) Representative reconstructions for the 11G, 37T, 38G, and 45G amide group peaks, respectively; dashed, gray, and black lines show 1D ^{15}N traces through the peaks in the full reference, the LR and CS spectra, respectively.

reconstruction in Figure 2a and the fully sampled reference spectrum in Figure 2b illustrates the high quality of the LR reconstruction obtained from only 35% of the traditionally acquired spectrum. This qualitative observation corroborates with the faithful reproduction of the peak intensities shown in the inset of Figure 2b. Similar results are obtained for a 2D NOESY spectrum of ubiquitin (see the Supporting Information). The quality of the CS reconstruction obtained

from the same NUS HSQC data (not shown) is generally as good, with the majority of the peaks reproduced equally well by the CS and LR. This is illustrated for the amide group of Gly11 in Figure 2c. Nevertheless, several low-intensity peaks are notably compromised in the CS spectrum as shown in Figure 2d–f. While peaks for Thr37 and Gly45 show clear line shape shrinkage, the peak of Gly38 is completely lost. The opposite situation, when a true peak is present in the CS but is missing in the LR reconstruction, never occurred in our spectra. It should be also noted that the virtual-echo preprocessing used for all of the CS reconstructions in this work improves quality of the spectra but requires prior knowledge about the signal phase.[7] In general, when the phase is unknown, the virtual-echo cannot be used and the comparison between the CS and LR would be even more in favour of the LR method. The experimental results are fully consistent with the simulations shown in Figure 1 and lead us to the conclusion that the LR produces at least as good spectral reconstructions as the CS and often outperforms it for broadest and weakest peaks. Effective sensitivity of a spectrum reconstruction method from NUS data is defined as a possibility to detect weak peaks and discriminate them from eventual false signals.[16] Thus, the observed good reconstructions of the low-intensity peaks by the LR indicate high sensitivity of the new method.

We introduce the LR reconstruction as a new general technique for obtaining high-quality spectra from a small number of NUS data points. The method allows a significant reduction in measurement time, which is particularly useful for high throughput applications, studies of short lived systems, time-resolved experiments, and many other practical cases. We demonstrate the first NUS reconstruction algorithm using the low-rank property of the NMR time-domain signal. The LR and CS approaches are based on distinctly different assumptions and in future work we envisage design of an even more powerful NUS processing algorithm that combines the low rank and sparseness signal properties.

Received: September 19, 2014
Published online: November 11, 2014

Keywords: low-rank approximation · NMR spectroscopy · real-time measurements · structure elucidation

[1] a) D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Wälchli, B. O. Smith, M. Shirakawa, P. Güntert, Y. Ito, *Nature* **2009**, *458*, 102–105; b) F.-X. Theillet, H. M. Rose, S. Liokatis, A. Binolfi, R. Thongwichian, M. Stuiver, P. Selenko, *Nat. Protoc.* **2013**, *8*, 1416–1432.

[2] a) J. Balbach, V. Forge, W. S. Lau, N. A. van Nuland, K. Brew, C. M. Dobson, *Science* **1996**, *274*, 1161–1163; b) I. Landrieu, L. Lacosse, A. Leroy, J.-M. Wieruszeski, X. Trivelli, A. Sillen, N. Sibille, H. Schwalbe, K. Saxena, T. Langer, G. Lippens, *J. Am. Chem. Soc.* **2006**, *128*, 3575–3583; c) E. Rennella, T. Cutuil, P. Schanda, I. Ayala, V. Forge, B. Brutscher, *J. Am. Chem. Soc.* **2012**, *134*, 8066–8069; d) M. Mayzel, J. Rosenlow, L. Isaksson, V. Y. Orekhov, *J. Biomol. NMR* **2014**, *58*, 129–139.

[3] A. L. Guennec, P. Giraudeau, S. Caldarelli, *Anal. Chem.* **2014**, *86*, 5946–5954.

[4] a) J. C. J. Barna, E. D. Laue, M. R. Mayger, J. Skilling, S. J. P. Worrall, *J. Magn. Reson.* **1987**, *73*, 69–77; b) K. Kazimierczuk, J. Stanek, A. Zawadzka-Kazimierczuk, W. Kozminski, *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 420–434; c) S. G. Hyberts, K. Takeuchi, G. Wagner, *J. Am. Chem. Soc.* **2010**, *132*, 2145–2147; d) V. Y. Orekhov, V. A. Jaravine, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 271–292; e) J. C. Hoch, M. W. Maciejewski, M. Mobli, A. D. Schuyler, A. S. Stern, *Acc. Chem. Res.* **2014**, *47*, 708–717.

[5] a) D. L. Donoho, *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306; b) E. J. Candès, J. K. Romberg, T. Tao, *Commun. Pure Appl. Math.* **2006**, *59*, 1207–1223.

[6] a) K. Kazimierczuk, V. Y. Orekhov, *Angew. Chem. Int. Ed.* **2011**, *50*, 5556–5559; *Angew. Chem.* **2011**, *123*, 5670–5673; b) D. J. Holland, M. J. Bostock, L. F. Gladden, D. Nietlispach, *Angew. Chem. Int. Ed.* **2011**, *50*, 6548–6551; *Angew. Chem.* **2011**, *123*, 6678–6681; c) Y. Shrot, L. Frydman, *J. Magn. Reson.* **2011**, *209*, 352–358; d) X. B. Qu, D. Guo, X. Cao, S. H. Cai, Z. Chen, *Sensors* **2011**, *11*, 8888–8909; e) S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari, G. Wagner, *J. Biomol. NMR* **2012**, *52*, 315–327; f) E. C. Lin, S. J. Opella, *J. Magn. Reson.* **2013**, *237*, 40–48.

[7] M. Mayzel, K. Kazimierczuk, V. Y. Orekhov, *Chem. Commun.* **2014**, *50*, 8947–8950.

[8] a) B. Zhao, J. P. Haldar, A. G. Christodoulou, Z. P. Liang, *IEEE T. Med. Imag.* **2012**, *31*, 1809–1820; b) H. M. Nguyen, X. Peng, M. N. Do, Z. P. Liang, *IEEE Trans. Biomed. Eng.* **2013**, *60*, 78–89; c) P. J. Shin, P. E. Z. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, M. Lustig, *Magn. Reson. Medicine* **2013**, DOI: 10.1002/mrm.24997; d) J. F. Cai, X. Jia, H. Gao, S. B. Jiang, Z. Shen, H. Zhao, *IEEE T. Med. Imag.* **2014**, *33*, 1581–1591.

[9] a) E. J. Candès, X. Li, Y. Ma, J. Wright, *J. ACM* **2011**, *58*, 11; b) H. Ji, S. Huang, Z. Shen, Y. Xu, *SIAM J. Imag. Sci.* **2011**, *4*, 1122; c) W. Dong, G. Shi, X. Li, Y. Ma, F. Huang, *IEEE T. Image Processing* **2014**, *32*, 3618–3632.

[10] a) E. J. Candès, B. Recht, *Found. Comput. Math.* **2009**, *9*, 717–772; b) E. J. Candes, Y. Plan, *IEEE Proceedings* **2010**, *98*, 925–936.

[11] a) P. Koehl, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 257–299; b) J. C. Hoch, A. S. Stern, *NMR data processing*, Wiley-Liss, **1996**.

[12] V. A. Mandelshtam, *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 159–196.

[13] V. Jaravine, I. Ibraghimov, V. Y. Orekhov, *Nat. Methods* **2006**, *3*, 605–607.

[14] J. F. Cai, E. J. Candès, Z. W. Shen, *SIAM J. Optim.* **2010**, *20*, 1956–1982.

[15] A. Ganesh, L. Zhouchen, J. Wright, W. Leqin, C. Minming, M. Yi, in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2009 3rd IEEE International Workshop on*, 10.1109/CAMSAP.2009.5413299, **2009**, pp. 213–216.

[16] S. G. Hyberts, S. A. Robson, G. Wagner, *J. Biomol. NMR* **2013**, *55*, 167–178.

