Fast NMR Spectroscopy

DOI: 10.1002/ange.201100370

Accelerated NMR Spectroscopy by Using Compressed Sensing**

*Krzysztof Kazimierczuk and Vladislav Yu. Orekhov**

There is increased interest in high-resolution, fast multidimensional NMR spectroscopy for studying molecular structure, interactions, and dynamics. The distinct feature of the contemporary NMR spectroscopy, namely the possibility to observe hundreds of atoms in complex macromolecules simultaneously, finds its foundation in the invention of multidimensional experiments in the mid 1970s.[1] However, the ultimate resolution obtained in these experiments comes at the high price of the long data collection times needed to systematically sample the large multidimensional data sets. The number of measured data points increases polynomialy with desired spectral resolution and exponentially with a number of dimensions.[2] The problem of lengthy sampling often compromises or even prohibits many applications of the multidimensional spectroscopy in chemistry and molecular biology. Fortunately, the field of fast NMR spectroscopy offers a number of solutions.[3–9] A common approach is to replace the time-consuming systematic sampling of the signal on the fine Nyquist grid by the random non-uniform sampling (NUS).[4] For many years, however, NUS was associated with the inherent loss of the spectrum quality, such as the presence of spectral artefacts and false peaks.

Recently, Candès et al.[10] formulated a new NUS theorem, which states that for most of the practical cases, a significantly smaller number of data points in comparison to the size the full Nyquist grid is sufficient for obtaining the exact reconstruction of the spectrum. The theorem evoked the rapidly growing group of signal processing methods, referred to as the compressed sensing (CS) or compressive sampling. A number of CS applications has been recently demonstrated in various fields of science and technology, including the striking results obtained for fast magnetic resonance imaging (MRI).[11] Herein, we demonstrate CS as an effective tool for obtaining high-quality spectra from the NUS data and present the first experimental examples of compressed sensing in NMR spectroscopy (CS-NMR).

According to the classical Nyquist–Shannon sampling theorem, sampling at the constant rate, which is equal or larger than the spectral bandwidth, is the necessary condition for the exact reconstruction of the spectrum. This theorem is based on an implicit pessimistic assumption that every point in the spectrum carries important information. However, in the most of the practical cases, including the NMR spectroscopy, the peaks occupy only a small fraction of the spectrum, while the rest is the baseline. In other words, we say that the spectrum is sparse.

NMR signal sampled in the time domain and represented as a vector **s** is associated with a frequency-domain spectrum **S** in the following way:

$$\mathbf{F}\,\mathbf{S} = \mathbf{s} \tag{1}$$

where **F** stands for the inverse of discrete Fourier transform. Equation (1) can be defined for one or multiple dimensions. If the signal is sampled regularly on the Nyquist grid, the number of unknown points in the spectrum **S** is equal to the number of elements in the time-domain signal **s**, and Equation (1) has a unique solution. In the NUS strategy, measurements are performed only for a small fraction of the randomly selected points from the grid. This saves much experimental time, but the system of Equations (1) becomes underdetermined and thus has the infinite number of solutions. To choose the right solution and obtain the good spectrum, additional assumptions need to be introduced. The processing methods that have been developed over last three decades for dealing with the NUS signal differ primarily in the kind of assumptions and are more or less successful depending on its relevance for the particular type of spectrum and the level of sparseness. Thus, the simplest approach is to choose the solution that features the minimal power (or l_2 norm). As this is equivalent to assuming that signal is equal to zero in all the grid points that were omitted in the experiment, the spectrum can be obtained by the direct discrete Fourier transform.[6] Another traditional approach looks for the solution having maximum entropy.[7] The SIFT method[8] offers the third alternative, where the number of unknowns in the linear system of Equation (1) is significantly reduced by setting some points of spectrum to be equal to zero. Finally, an alternative to the Fourier basis set can be used that contains fewer basis functions. Thus, MDD describes the spectrum by the small number of tensor products of one-dimensional vectors.[9] Unfortunately, none of above methods was mathematically proven to give the exact, that is, artifact-free reconstruction of the spectrum.

In the CS approach, it is assumed that the best among the solutions fulfilling Equation (1) is the sparsest one, that is, containing the highest number of zeros. This corresponds to the minimal l_0 norm of the spectrum. Unfortunately, finding such solution is the NP-hard task, and typically unachievable in a reasonable computational time. Here the power of the CS theory comes into play, which states that the sparsest solution

[*] K. Kazimierczuk, V. Yu. Orekhov
Swedish NMR Centre, University of Gothenburg
Box 465, 40530 Göteborg (Sweden)
Fax: (+46) 31-786-3886
E-mail: vladislav.orekhov@nmr.gu.se
K. Kazimierczuk
Faculty of Chemistry, University of Warsaw
Pasteura 1, 02093 Warsaw (Poland)

[**] The Swedish Research Council (grant 2008-4299), Foundation for Polish Science (KOLUMB stipend) for supporting K.K., Bruker BioSpin, and EC Bio-NMR project no. 261863 are acknowledged.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201100370.

can be almost always found by the minimization of penalty function involving the l_p norm of the spectrum[12] with $p=1$ [Equation (2)]:

$$\|\mathbf{FS}-\mathbf{s}\|_{l_2}^2+\lambda\|\mathbf{S}\|_{l_p}^p \quad (2)$$

The penalty function is convex and thus has only one (global) minimum. Moreover, the reconstruction of the spectrum from NUS signal is exact (that is, the same as from the full dataset) with overwhelming probability, and no other reconstruction method can in general perform better than CS.[13] We should note, however, that strictly speaking, neither the NMR signal nor the accompanying Gaussian noise are sparse in the frequency domain. Nevertheless, the theory predicts that CS can be also applied to approximately sparse and/or noisy signals (for a comprehensive review, see Ref. [14]). Notably, the l_1 norm penalty function was recently successfully employed in the reconstruction of NMR spectra by the forward maximum entropy method.[5] The penalty functions [Eq. (2)] with the norm l_p, where p less than 1 are non-convex, that is, may have more than one local minimum. However, in real applications the solution is at least as good as the one obtained for the l_1 norm. Besides, spectral reconstructions using a norm with $0<p<1$ may display better convergence[15] and require fewer measurements in comparison with the l_1 norm.

Finally, the matrix **F** in Equations (1) and (2) can be easily defined to span only a fraction of the multidimensional spectrum, and it is possible to apply the CS approach to reconstruct regions of high-dimensional spectra when signal positions in some of dimensions are known.[16] This would correspond to assuming, in a similar fashion to the SIFT method,[8] that there are no peaks beyond the selected regions.

Herein we present results obtained using two different CS algorithms: iterative soft thresholding (IST)[17] and iterative re-weighted least squares (IRLS)[18] (for details of the algorithms, see the Supporting Information). The IST algorithm has been demonstrated on simulated NUS datasets[19] in combination with the wavelet transform and for spectrum reconstruction from a spatially encoded signal.[20] It was also used for suppressing *sinc*-"wiggles" in the spectra of truncated signals.[21] IST is reminiscent of various algorithms for cleaning aliasing artefacts in the NMR spectra obtained from sparse data.[22] Contrary to the IST, which is equivalent to l_1 norm minimization, the IRLS algorithm allows also the minimization of norms less than 1 (later referred to as the $l_{p\to 0}$ norm).[23] When IRLS algorithm is used for the l_1 norm penalty function, the convergence and results are similar to IST, albeit with longer computational times.

We demonstrate results of the IST and IRLS algorithms for three two-dimensional spectra: ^{1}H–^{15}N HSQC of a globular 14 kDa protein azurin[24] (1 mM, temperature 25 °C, 900 MHz Varian UNITY *Inova* with cryoprobe) and 2D NOESY and 2D DQF-COSY of human ubiquitin (1 mM, temperature 25 °C, 600 MHz Varian UNITY Inova). It should be emphasized that despite the modest computational costs, the low-dimensional spectra of complex molecules, such as proteins, are among the most difficult for the fast sampling methods. Thus, for two-dimensional spectra, the reduced dimensionality is not applicable and the NUS methods are rarely used because 2D spectra are not significantly sparse and the low absolute number of points leads to the poor statistics for the random sampling.[25]

Figure 1 a–c shows the comparison of HSQC spectrum obtained by the Fourier transform of a full dataset (120 complex points) with the CS reconstructions using 18.3 % of the indirect dimension samples, chosen randomly according

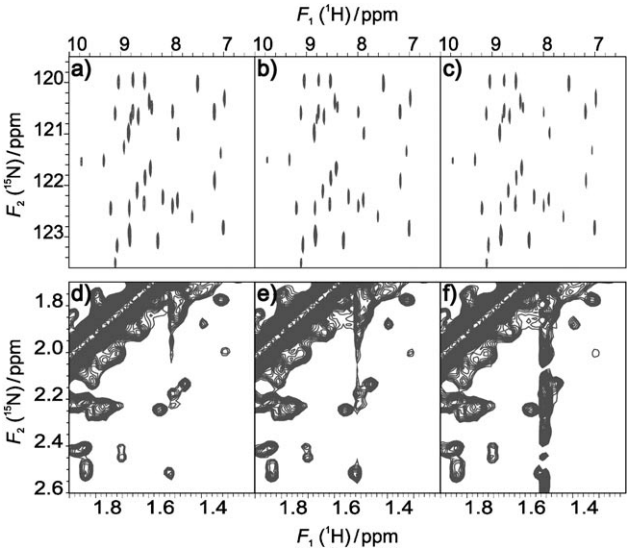

Figure 1. a–c) ^{15}N HSQC spectrum of azurin: a) Full spectrum (120 pts), b) IRLS reconstruction (22 pts, 10 iterations), c) IST reconstruction (22 pts, 900 iterations). d–f) 2D NOESY spectrum of ubiquitin: d) Full spectrum (512 pts), e) IRLS reconstruction (200 pts, 10 iterations), f) IST reconstruction (200 pts, 1500 iterations). Residual artifacts originating from a strong diagonal peak can be observed in panel (f). Exponentially decaying distributions of points were used in both NUS experiments.

to the exponentially biased probability distribution.[26] Reconstructions were performed using the $l_{p\to 0}$ norm (IRLS) and l_1 norm (IST) minimizations. Figure 1 d–f presents results for 2D NOESY with undersampling at the level of 39 % (other regions of the NOESY spectrum are shown in the Supporting Information, Figure S3). Figure 2 illustrates linearity of the peak intensities in the IRLS reconstruction of the NOESY spectrum. The number of iterations in IRLS was set to 10.

Results for IST show that the algorithm works well for spectra with good signal-to-noise ratios (S/N) and moderate dynamic range; for example, the HSQC depicted in Figure 1 a–c. In the more demanding situations, such as the NOESY-type spectra (Figure 1 d–f) with high dynamic range and many weak peaks, the convergence of the algorithm becomes slow. Although the IST finally converged to the artefact-free spectrum, it required about 5000 iterations, and thus significantly longer calculations than required for the IRLS. Similarity of the spectra obtained using l_1 (IST) and $l_{p\to 0}$ (IRLS) norms is in line with the CS theory.[12] The limited performance of the IST is in line with observations reported for some of the thresholding-based artefact-cleaning algo-

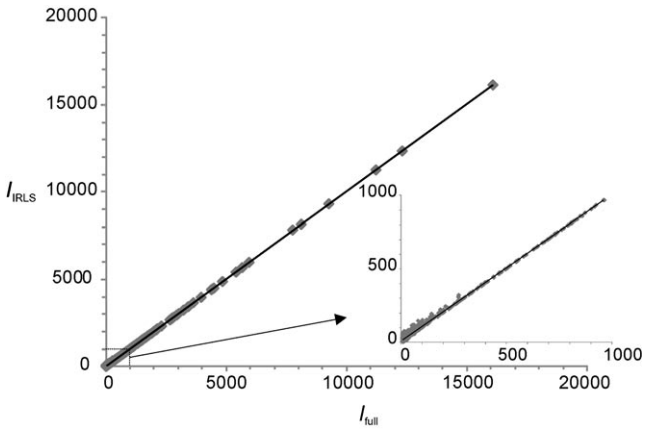

Figure 2. Accuracy of peak intensities in the reconstructed 2D NOESY spectrum of azurin. Intensities of peaks are shown in the IRLS-reconstructed spectrum (200 points out of 512) against the corresponding intensities in the full spectrum. The intensities are measured in the units of standard deviation of noise in the full spectrum. A complete manually verified list of peaks with S/N $\geq$ 8.0 in the full dataset were considered. Considering the shorter measurement time, this cut-off corresponds to S/N $\geq$ 5 in the reconstruction. The trend line was fitted with equation $I_{IRLS} = 1.0032\, I_{full} + 6.77$ with $R^2 = 0.9998$ (0.9963 for the inset).

rithms, which may require experiments with suppression of strong diagonal signals for successful processing of NOESY.[27]

The IRLS approach allows the accurate reconstruction of complex peak shapes. Figure 3 shows the example of the DQF COSY spectrum of ubiquitin. A multiplet pattern is reconstructed with high fidelity, which may be required, for example, for quantitative measurement of the scalar couplings.

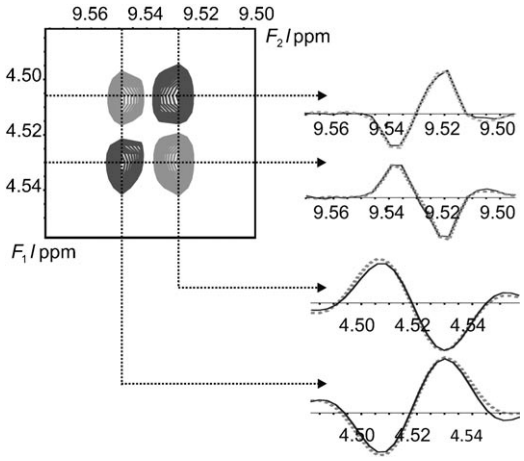

Figure 3. Fidelity of the lineshape reconstruction. The I13 HN-Hα peak from the IRLS-reconstructed DQF COSY spectrum of ubiquitin is shown. Positive contour levels are marked with dark gray, negative with light gray. 150 points (29.3%) were randomly selected from the full dataset (512 points). A sine-bell-weighted point distribution was used. The side panels show peak intersections (——) and a comparison with peak shape in the spectrum obtained with the full dataset (- - - -).

The field of compressed sensing, which recently exploded thanks to the pioneering theoretical works of Candès, Romberg, Tao, Donoho, and co-workers found successful applications in a multitude of signal processing applications.[12–14] With this work, we demonstrate that it is also a powerful tool for the NMR spectroscopy. The approach presented herein opens an avenue for a variety of CS applications in two- and multidimensional NMR spectroscopy.

Received: January 15, 2011
Published online: April 29, 2011

Keywords: data acquisition · NMR spectroscopy · non-uniform sampling · protein structures · structure elucidation

[1] a) R. R. Ernst, W. A. Anderson, *Rev. Sci. Instrum.* **1966**, *37*, 93; b) J. Jeener in *Proc. AMPERE International Summer School II*, Basko Polje, Yugoslavia, **1971**.
[2] V. Jaravine, I. Ibraghimov, V. Y. Orekhov, *Nat. Methods* **2006**, *3*, 605.
[3] a) B. E. Coggins, R. A. Venters, P. Zhou, *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 381; b) A. Tal, L. Frydman, *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 241; c) R. Bruschweiler, F. L. Zhang, *J. Chem. Phys.* **2004**, *120*, 5253; d) E. Kupce, R. Freeman, *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *52*, 22; e) V. A. Jaravine, A. V. Zhuravleva, P. Permi, I. Ibraghimov, V. Y. Orekhov, *J. Am. Chem. Soc.* **2008**, *130*, 3927.
[4] M. W. Maciejewski, H. Z. Qui, I. Rujan, M. Mobli, J. C. Hoch, *J. Magn. Reson.* **2009**, *199*, 88.
[5] a) S. G. Hyberts, D. P. Frueh, H. Arthanari, G. Wagner, *J. Biomol. NMR* **2009**, *45*, 283; b) S. G. Hyberts, K. Takeuchi, G. Wagner, *J. Am. Chem. Soc.* **2010**, *132*, 2145.
[6] K. Kazimierczuk, W. Kozminski, I. Zhukov, *J. Magn. Reson.* **2006**, *179*, 323.
[7] S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue, J. Staunton, *Nature* **1984**, *311*, 446.
[8] Y. Matsuki, M. T. Eddy, J. Herzfeld, *J. Am. Chem. Soc.* **2009**, *131*, 4648.
[9] V. Y. Orekhov, I. Ibraghimov, M. Billeter, *J. Biomol. NMR* **2003**, *27*, 165.
[10] E. J. Candès, J. Romberg, T. Tao, *IEEE Trans. Inf. Theory* **2006**, *52*, 489.
[11] a) S. Hu, M. Lustig, A. P. Chen, J. Crane, A. Kerr, D. A. C. Kelley, R. Hurd, J. Kurhanewicz, S. J. Nelson, J. M. Pauly, D. B. Vigneron, *J. Magn. Reson.* **2008**, *192*, 258; b) M. Lustig, D. Donoho, J. M. Pauly, *Magn. Reson. Med.* **2007**, *58*, 1182.
[12] D. L. Donoho, *Commun. Pure Appl. Math.* **2004**, *59*, 797.
[13] E. J. Candès, J. K. Romberg, T. Tao, *Commun. Pure Appl. Math.* **2006**, *59*, 1207.
[14] E. J. Candes, M. B. Wakin, *IEEE Sig. Proc. Mag.* **2008**, *25*, 21.
[15] R. Chartrand, W. Yin in *33rd International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), **2008**.
[16] K. Kazimierczuk, A. Zawadzka-Kazimierczuk, W. Kozminski, *J. Magn. Reson.* **2010**, *205*, 286.
[17] A. Papoulis, *IEEE Trans. Circuits Syst.* **1975**, *22*, 735.
[18] a) C. L. Lawson, *PhD Thesis*, UCLA **1961**; b) E. J. Candès, M. B. Wakin, S. P. Boyd, *J. Fourier Anal. Appl.* **2008**, *14*, 877.
[19] I. Drori, *EURASIP J. Adv. Sig. Pr.* **2007**.
[20] Y. Shrot, L. Frydman, *J. Magn. Reson.* **2011**, *209*, 352.
[21] A. S. Stern, D. L. Donoho, J. C. Hoch, *J. Magn. Reson.* **2007**, *188*, 295.
[22] B. E. Coggins, P. Zhou, *J. Biomol. NMR* **2008**, *42*, 225.

[23] I. Daubechies, R. Devore, M. Fornasier, C. S. Gunturk, *Commun. Pure Appl. Math.* **2010**, *63*, 1.
[24] B. G. Karlsson, T. Pascher, M. Nordling, R. H. A. Arvidsson, L. G. Lundberg, *FEBS Lett.* **1989**, *246*, 211.
[25] K. Kazimierczuk, A. Zawadzka, W. Kozminski, *J. Magn. Reson.* **2009**, *197*, 219.
[26] J. C. J. Barna, E. D. Laue, M. R. Mayger, J. Skilling, S. J. P. Worrall, *J. Magn. Reson.* **1987**, *73*, 69.
[27] J. W. Werner-Allen, B. E. Coggins, P. Zhou, *J. Magn. Reson.* **2010**, *204*, 173.