Sparse Sampling Techniques

DOI: 10.1002/anie.201402049

Accelerating Diffusion-Ordered NMR Spectroscopy by Joint Sparse Sampling of Diffusion and Time Dimensions**

*Mateusz Urbańczyk, Wiktor Koźmiński, and Krzysztof Kazimierczuk**

Abstract: *Diffusion-ordered multidimensional NMR spectroscopy is a valuable technique for the analysis of complex chemical mixtures. However, this method is very time-consuming because of the costly sampling of a multidimensional signal. Various sparse sampling techniques have been proposed to accelerate such measurements, but they have always been limited to frequency dimensions of NMR spectra. It is now revealed how sparse sampling can be extended to diffusion dimensions.*

The phenomenon of nuclear magnetic resonance (NMR) underlies one of the most important techniques of chemical analysis, NMR spectroscopy. In an NMR experiment, nuclei are polarized by an external magnetic field and, after excitation with radiofrequency (RF) pulses, return to the equilibrium state emitting a free induction decay (FID) signal. This signal is usually weak because the nuclear magnetic moment is very small, and therefore, the experiment has to be a masterstroke of signal excitation, detection, and processing. Furthermore, the changes in resonance frequency that are caused by molecular structure and dynamics (chemical shift) are in the order of parts per million (ppm) and often difficult to resolve.

The latter problem can be circumvented by introducing additional spectral dimensions. For example, in a two-dimensional NMR experiment, the signal is a function of two time variables, each encoding one type of nucleus (^{1}H, ^{13}C, ^{15}N etc.), and its components correspond to pairs of nuclei that are coupled by a physical interaction:

$$s(t_1,t_2)=\sum_{i=1}^{N} s_i(t_1)\otimes s_i(t_2) \tag{1}$$

where $si(t_1)$ and $si(t_2)$ are resonance signals from the ith pair of nuclei. The spectrum of a signal that is emitted by N pairs of nuclei consists of N correlation peaks, each characterized by a pair of frequencies. The vector of spectral points **S** is obtained from the vector of the FID samples **s** by solving the following system of equations:

$$\mathbf{FS}=\mathbf{s} \tag{2}$$

where **F** is the inverse Fourier transform $M\times M$ matrix with $F_{lm}=\exp\left(\frac{i2\pi lm}{M}\right)$ or the unique solution, **s** and **S** have to be of equal size. Unfortunately, the sampling of indirect dimensions (t_1 in this case) is time-consuming. Usually, to obtain two-dimensional **S** of good resolution, data has to be collected for tens of minutes or more. This motivated several groups to develop sophisticated mathematical methods that allow the reconstruction of **S** from a relatively small number of samples.[1] Many of these approaches are based on random or non-uniform sampling (NUS).[2] In such cases, **S** is found by minimizing the penalty function involving the term $\boldsymbol{\Theta}(\mathbf{S})$, which represents additional assumptions that are needed to solve the underdetermined variant of Equation (2):

$$\min_{\mathbf{s}}\|\mathbf{FS}-s\|_{\ell_2}^2+\tau\Theta(S) \tag{3}$$

where τ is the Lagrange coefficient, which keeps the measured data and the assumption in accordance. Among these approaches, compressed sensing (CS) has been shown to be most effective for signals with sparse (i.e., "mostly empty") **S**.[3] This method, which was successfully employed to NMR spectroscopy,[1a,b] assumes that $\boldsymbol{\Theta}(\mathbf{S})=|\,\mathrm{S}\,|_{\ell_1}$.

If the studied sample is a mixture of several compounds, the spectral resolution can be improved by introducing diffusion dimensions.[4] This approach, known as diffusion-ordered spectroscopy (DOSY), allows the generation of a signal that exponentially decays with a rate that is proportional to the diffusion of the molecule containing nuclei i:

$$s(t_1,t_2,g)=\sum_{i=1}^{N} s_i(t_1)\otimes s_i(t_2)\otimes\alpha_i(g) \tag{4}$$

where $\alpha_i(g)$ is a decaying function that is connected to the distribution of the diffusion coefficients $A(D)$ through a Laplace transform, and g is the magnetic field gradient that is used to sample the decay. To determine $A(D)$, a Laplace transform has to be employed, which, unlike the Fourier transform (FT),

[*] M. Urbańczyk, Prof. Dr. W. Koźmiński
Faculty of Chemistry, Biological and Chemical Research Centre
University of Warsaw
Żwirki i Wigury 101, Warsaw, 02-089 (Poland)
Dr. K. Kazimierczuk
Centre of New Technologies, University of Warsaw
Żwirki iWigury 93, Warsaw, 02-089 (Poland)
E-mail: k.kazimierczuk@cent.uw.edu.pl

[**] We thank the Foundation for Polish Science for support within the TEAM Programme and the Polish National Science Centre for support (HARMONIA grant DEC-2013/08/M/ST4/00975 and SONATA BIS 2 grant DEC-2012/07/E/ST4/01386). This study was carried out at the Biological and Chemical Research Centre, University of Warsaw, established within a project co-financed by the European Union through the European Regional Development Fund under the Operational Programme Innovative Economy (2007–2013). Numerical computations were performed at the Interdisciplinary Centre for Computational and Mathematical Modelling (ICM), University of Warsaw (G52-15).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201402049.

is numerically unstable and needs to be regularized. Possible regularization methods include the maximum entropy method[5] or CONTIN.[6] Recently, we have proposed the ℓ_1-norm regularization algorithm ITAMeD for the processing of diffusion data:[7]

$$\min_{\mathbf{A}} \|\mathbf{LA} - \alpha\|_{\ell_2}^2 + \tau\|\mathbf{A}\|_{\ell_1} \quad (5)$$

where **L** is the Laplace transform matrix. This kind of regularization enforces the sparsity of the solution to the transform.[8]

Interestingly, the inverse Laplace transform has always been applied to DOSY data in a "sequential" manner, that is, after FT in all time dimensions. This required full (and thus time-consuming) sampling by varying the magnetic field gradient for each point of all time domains. However, the striking similarity of Equations (5) and (3) suggests that Fourier and Laplace transforms can be combined, and the signal can be sparsely sampled in a joint t_1/g domain (Figure 1). The minimized function becomes:

$$\min_{\mathbf{Q}} \|\mathbf{PQ} - \mathbf{q}\|_{\ell_2}^2 + \tau\|\mathbf{Q}\|_{\ell_1} \quad (6)$$

where **P** is the combined "Fourier–Laplace transform":

$$\mathbf{P} = \mathbf{F} \otimes \mathbf{L} \quad (7)$$

and **q** is the vector of the sparsely sampled indirect part of the three-dimensional signal from Equation (4); **Q** is its spectrum.

The robustness of this method to noise and undersampling has recently been studied by simulations.[9] Herein, we exploit the possibility of an experimental application of this approach, exemplified for the HSQC–DOSY technique.

We demonstrate the result of 3D ITAMeD processing on the 3D CT–HSQC–iDOSY[10] spectra of two chemical mixtures: a four-component mixture with different concentrations of citrate, L-alanine, 2,4,6-trimethylaniline (TMA), and taurine and a two-component equimolar mixture of rutin and quercetin, which show a strong degeneracy of ^{1}H chemical shifts. Cross sections and the projection over the ^{13}C dimension for the first sample are shown in Figure 2. Both the ^{1}H and ^{13}C frequency coordinates are reproduced up to the precision that is defined by digital resolution. The reconstructed diffusion coefficient values differ only slightly between peaks for the same compound.

The peak volumes in the reconstructed spectrum are in acceptable accordance with the reference ^{1}H NMR spectrum (Table 1). Only the volume of the L-alanine peak was significantly overestimated.

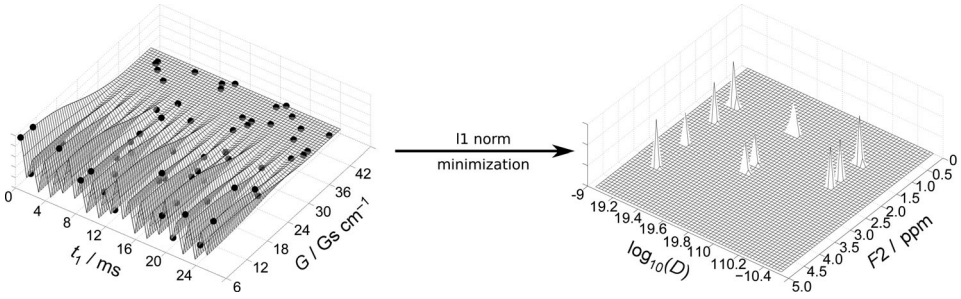

Figure 1. Joint sparse sampling of the time and gradient domains in diffusion-ordered NMR spectroscopy.

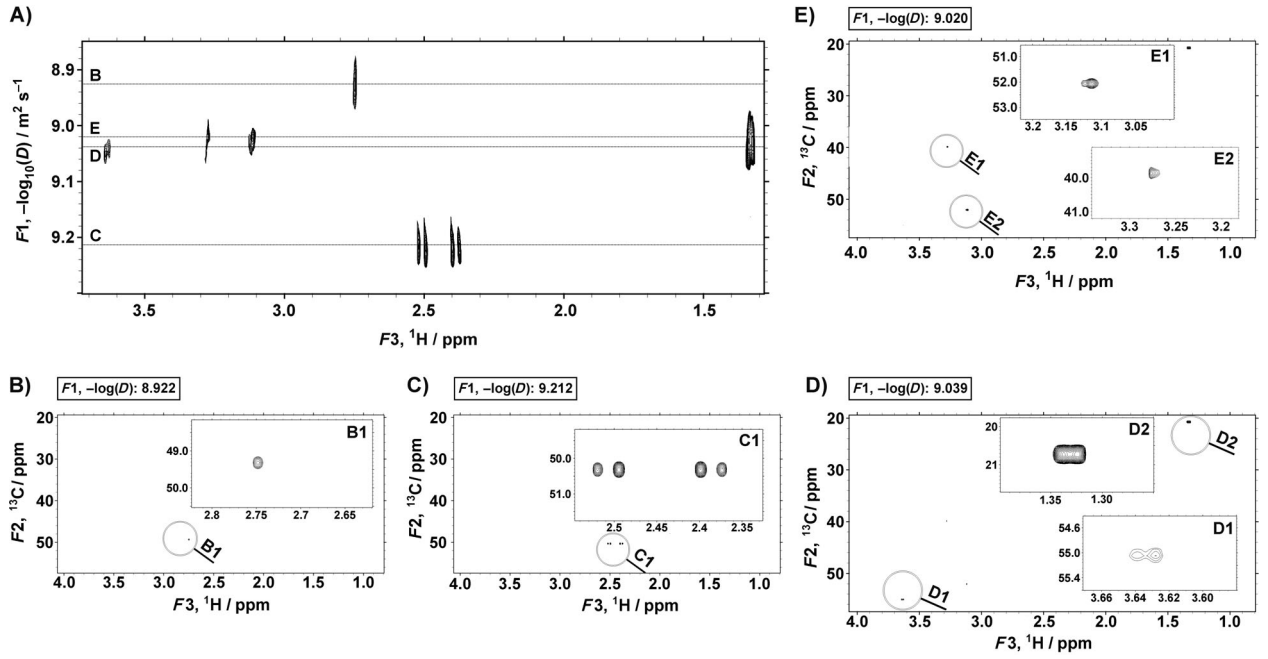

Figure 2. 3D HSQC–DOSY spectrum of a mixture of L-alanine, TMA, taurine, and citrate. A) Projection over the ^{13}C dimension. B–E) Cross sections for the diffusion coefficients of TMA (B), citrate (C), L-alanine (D), and taurine (E). The insets contain magnified images of certain peaks.

Table 1: Peak positions from the presented 3D HSQC–DOSY measurements and their normalized integrals. The reference integrals were taken from the ^{1}H NMR spectrum.

	D [$\mu m^2 s^{-1}$]	$F2$ [ppm]	$F3$ [ppm]	Integral	Ref.
Citrate	601	50.314	2.520	2.57	2.40
	605	50.315	2.495		
	605	50.313	2.399	2.71	2.46
	607	50.315	2.374		
TMA	1180	49.317	2.748	1	1
Taurine	942	39.874	3.279	2.04	2.16
	940	52.054	3.115	1.69	2.17
L-Alanine	900	55.037	3.636	1.67	1.33
	906	20.773	1.330	9.08	6.53

Furthermore, we examined the influence of undersampling on the quality of the spectrum (see the Supporting Information, Tables S3–S6). For low sampling levels, the diffusion coefficients were disturbed. However, even for only 50 points, the reconstruction separated the compound with the smallest diffusion coefficient from the one with the largest diffusion coefficient. Therefore, it seems that a large reduction in experimental time is possible if the components of the mixture differ significantly in their diffusion rates.

The proposed 3D method preserves all of the features of the ITAMeD algorithm that are associated with sparsity-enforcing ℓ_1-norm regularization, that is, it works best for monodisperse samples.[7] It should be noted, however, that the assumption of sparsity of reconstructed dimensions is even more relevant in three dimensions. As shown recently,[7] the one-dimensional ITAMeD algorithm is efficient in resolving multiexponential decays (overlapping peaks in the frequency dimension). The same feature was confirmed for the approach discussed here.[9] Yet, the peak dispersion is larger in two frequency dimensions than in the 1D case, and therefore, multiexponential decay rarely occurs in 3D HSQC–DOSY spectroscopy.

A comparison of the proposed method with the approach that is based on conventional sampling and monoexponential fit processing in the diffusion dimension is shown in Figure 3. This comparison is made for the sample with only two components (quercetin and rutin), but these components are characterized by a strong degeneracy of ^{1}H chemical shifts.[10b] As for the previous sample, the number of NUS points that corresponds to two regular HSQC spectra (256) turned out to be sufficient for a good reconstruction. Interestingly, even five regular HSQC spectra do not provide a precision in the diffusion dimensions that would be comparable with the ITAMeD result. For example, the deviations of the diffusion coefficients for the rutin peaks in the conventional measurement were up to 27 % (35 % for two HSQC spectra), whereas for the proposed method, they did not exceed 13 %. This can be explained by the fact that NUS provides much more

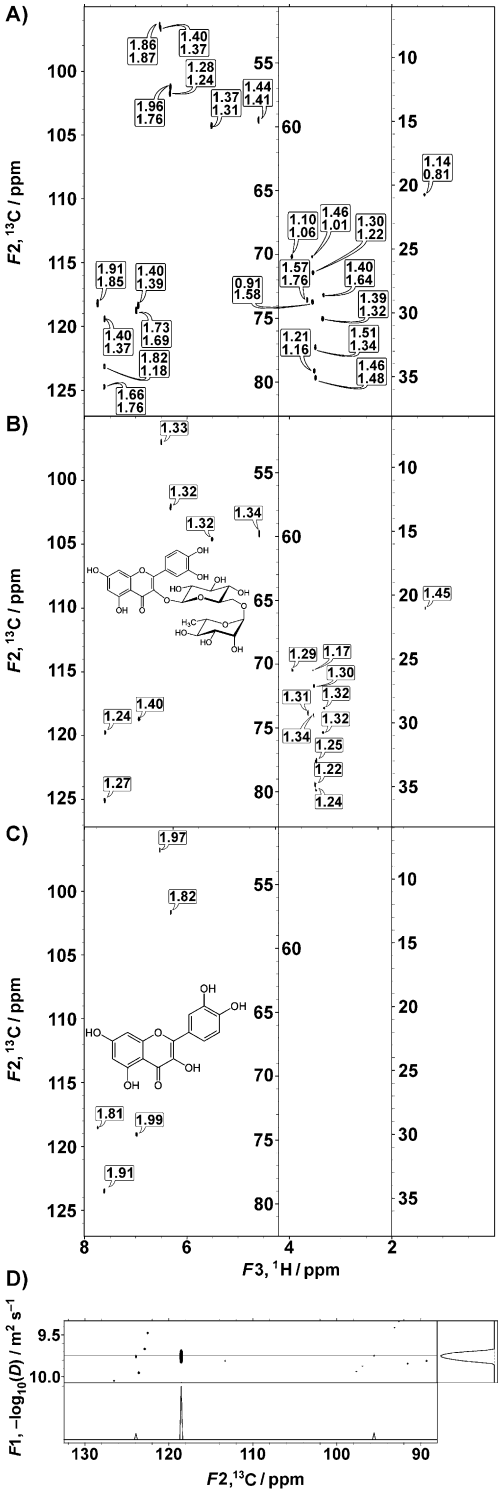

Figure 3. 3D HSQC–DOSY spectrum of the mixture of rutin and quercetin. A) 2D HSQC spectrum obtained with a 3D HSQC–DOSY pulse sequence for the lowest of five diffusion-encoding gradient levels. Markers show the diffusion coefficients obtained by a monoexponential fit (in 10^{-10} $m^2 s^{-1}$). Upper and lower values correspond to the fit using five and two (first and third out of the original five) HSQC spectra, respectively. B) Projection from the ITAMeD reconstruction for diffusion coefficients from 1.07×10^{-10} to 1.45×10^{-10} $m^2 s^{-1}$; markers show the peak coordinates in the diffusion dimension. C) Projection from the ITAMeD reconstruction for diffusion coefficients from 1.76×10^{-10} to 2.04×10^{-10} $m^2 s^{-1}$; markers show the peak coordinates in the diffusion dimension. D) Cross section from the ITAMeD reconstruction for the peak at $F3 = 7.64$ ppm showing artifacts obtained by this method.

extensive sampling of the diffusion domain, that is, much more gradient values are used. Furthermore, the effect of noise is different in ITAMeD than in monoexponential fit processing. With the former method, the noise manifests itself mainly in the presence of artifacts (Figure 3D), whereas with the latter method, the value of the fit (the diffusion coefficient) is affected. This is due to the fact that ITAMeD is able to process the multicomponent (multiexponential) function and, to some extent, separates noise as additional components. Additionally, by ITAMeD processing we were able to increase the resolution (decrease the linewidths) in the ^{13}C dimension by double extrapolation inside the NUS reconstruction algorithm. The diffusion coefficients were higher than those reported by McLachlan and co-workers[10b] because of a lower sample concentration and higher temperature.

In our opinion, many other 3D DOSY techniques may benefit from the application of joint NUS in the t_1/g space. However, according to the CS theory, the number of sampling points that is required for a proper reconstruction is proportional to the number of peaks in a spectrum.[3] Therefore, techniques that provide less sparse spectra (e.g., 3D TOCSY–DOSY) will require more sampling points.

Finally, it is worth mentioning that other regularization methods, such as MaxEnt[5] or CONTIN,[6] could also be applied to NUS data in the time-gradient domain. However, the ITAMeD method exploits the principle of compressed sensing in the frequency dimensions and thus provides superior results when the spectrum is sparse (which is usually the case in liquid-state NMR spectroscopy).

We propose that the developed method can significantly accelerate diffusion-ordered NMR spectroscopy. It is based on a sparsity constraint, which allows the efficient reconstruction of a 3D spectrum in an experimental time frame that is typical for 2D measurements. In practice, this will provide an at least three-fold acceleration of the experiment and more precise results. We believe that this method will open new avenues in a variety of high-dimensional diffusion NMR techniques.

Experimental Section

The CT–HSQC–iDOSY pulse sequence[10] was modified to employ States–TPPI quadrature and is described in detail in the Supporting Information. The first sample was a mixture of potassium citrate (85.2 mM), L-alanine (145.7 mM), TMA chloride (8.5 mM), and taurine (85 mM) dissolved in D_2O. The second sample was a mixture of rutin (40 mM) and quercetin (40 mM) in deuterated DMSO. Both mixtures have been used as DOSY standard samples before.[10b,11] The experiments were performed on an Agilent 600 MHz DDR2 NMR spectrometer with a Penta probe. All of the measurements were performed at 298 K. Signals were acquired using 400 (first sample) and 256 (second sample) pseudo-2D increments with delays/gradients set to random points from the 200×5000 (first sample) or 128×5000 (second sample) t_1/g grid with sixteen scans per point, which corresponded to 8 hours and 5 hours of experimental time, respectively. Regular sampling for the second sample was performed using 128 points in the ^{13}C dimension and five gradients, which corresponded to ca. 13 hours of experimental time. Spectra of the second sample were folded twice in the ^{13}C dimension, as described by McLachlan et al.[10b] For details of the sampling schedules, see the Supporting Information.

Received: February 4, 2014
Revised: March 5, 2014
Published online: April 24, 2014

Keywords: analytical methods · diffusion · diffusion-ordered spectroscopy · NMR spectroscopy

[1] a) D. J. Holland, M. J. Bostock, L. F. Gladden, D. Nietlispach, *Angew. Chem.* **2011**, *123*, 6678–6681; *Angew. Chem. Int. Ed.* **2011**, *50*, 6548–6551; b) K. Kazimierczuk, V. Y. Orekhov, Angew. Chem. **2011**, *123*, 5670–5673; *Angew. Chem. Int. Ed.* **2011**, *50*, 5556–5559; c) M. Mobli, J. C. Hoch, *Concepts Magn. Reson. Part A* **2008**, *32*, 436–448; d) Y. Matsuki, M. T. Eddy, J. Herzfeld, *J. Am. Chem. Soc.* **2009**, *131*, 4648–4656; e) V. Y. Orekhov, V. A. Jaravine, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 271–292; f) J. Stanek, W. Koźmiński, *J. Biomol. NMR* **2010**, *47*, 65–77; g) B. E. Coggins, R. A. Venters, P. Zhou, *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 381–419.

[2] K. Kazimierczuk, J. Stanek, A. Zawadzka-Kazimierczuk, W. Koźmiński, *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 420–434.

[3] E. Candès, M. Wakin, *IEEE Sign. Proc. Mag.* **2008**, *25*, 21–30.

[4] W. S. Price, *NMR Studies of Translational Motion: Principles and Applications*, Cambridge University Press, Leiden, **2009**.

[5] M. A. Delsuc, T. E. Malliavin, *Anal. Chem.* **1998**, *70*, 2146–2148.

[6] S. Provencher, *Comput. Phys. Commun.* **1982**, *27*, 229–242.

[7] M. Urbańczyk, D. Bernin, W. Koźminski, K. Kazimierczuk, *Anal. Chem.* **2013**, *85*, 1828–1833.

[8] D. L. Donoho, *Commun. Pure Appl. Math.* **2006**, *59*, 797–829.

[9] M. Urbańczyk, K. Kazimierczuk in *Signal Processing Symposium* (SPS), IEEE **2013**, pp. 1–6.

[10] a) B. Vitorge, D. Jeannerat, *Anal. Chem.* **2006**, *78*, 5601–5606; b) S. McLachlan, J. J. Richards, R. Bilia, G. Morris, *Magn. Reson. Chem.* **2009**, *47*, 1081–1085.

[11] L. M. Smith, A. D. Maher, O. Cloarec, M. Rantalainen, H. Tang, P. Elliott, J. Stamler, J. C. Lindon, E. Holmes, J. K. Nicholson, *Anal. Chem.* **2007**, *79*, 5682–5689.

