J Biomol NMR
DOI 10.1007/s10858-015-9923-x

ARTICLE

Efficient and generalized processing of multidimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms

Shangjin Sun[1] · Michelle Gill[1] · Yifei Li[1] · Mitchell Huang[1] · R. Andrew Byrd[1]

Received: 14 November 2014 / Accepted: 17 March 2015

Abstract The advantages of non-uniform sampling (NUS) in offering time savings and resolution enhancement in NMR experiments have been increasingly recognized. The possibility of sensitivity gain by NUS has also been demonstrated. Application of NUS to multidimensional NMR experiments requires the selection of a sampling scheme and a reconstruction scheme to generate uniformly sampled time domain data. In this report, an efficient reconstruction scheme is presented and used to evaluate a range of regularization algorithms that collectively yield a generalized solution to processing NUS data in multidimensional NMR experiments. We compare l1-norm (L1), iterative re-weighted l1-norm (IRL1), and Gaussian smoothed l0-norm (Gaussian-SL0) regularization for processing multidimensional NUS NMR data. Based on the reconstruction of different multidimensional NUS NMR data sets, L1 is demonstrated to be a fast and accurate reconstruction method for both quantitative, high dynamic range applications (e.g. NOESY) and for all J-coupled correlation experiments. Compared to L1, both IRL1 and Gaussian-SL0 are shown to produce slightly higher quality reconstructions with improved linearity in peak intensities, albeit with a computational cost. Finally, a generalized processing system, NESTA-NMR, is described that utilizes a fast and accurate first-order gradient descent algorithm (NESTA) recently developed in the compressed sensing field. NESTA-NMR incorporates L1, IRL1, and Gaussian-SL0 regularization. NESTA-NMR is demonstrated to provide an efficient, streamlined approach to handling all types of multidimensional NMR data using proteins ranging in size from 8 to 32 kDa.

Keywords Non-uniform sampling · Multidimensional NMR data processing · Compressed sensing · NESTA · NUS · gp78 · ASAP1

Introduction

Multidimensional NMR experiments are powerful techniques for extracting structural and dynamic information about proteins and other macromolecules. However, these experiments can be very time consuming on high field NMR spectrometers, where the necessity of covering larger spectral widths leads to shorter dwell times, which in turn make it difficult to achieve acquisition times sufficient to afford the necessary resolution. Furthermore, conventional (uniform) sampling schemes require the acquisition of data at equally spaced time points in all indirect dimensions to enable discrete Fourier transform (DFT) for data processing and analysis. For these reasons, uniform sampling is limiting in three-dimensional (3D) experiments, and the impact is exacerbated in four-dimensional (4D) experiments. In contrast, non-uniform sampling (NUS) allows a significant portion of the data points on the sampling grid to be omitted, thus enabling acquisition at long evolution times, where closely spaced signals can be resolved, without increasing the experimental acquisition time (Hoch

Shangjin Sun and Michelle Gill have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-015-9923-x) contains supplementary material, which is available to authorized users.

✉ R. Andrew Byrd
byrdra@mail.nih.gov

[1] Structural Biophysics Laboratory, National Cancer Institute, Frederick, MD 21702, USA

Published online: 26 March 2015

et al. 2014; Szantay 2008). Recently, several studies demonstrated that sensitivity gains can also be achieved with NUS methods (Hyberts et al. 2012c; Paramasivam et al. 2012; Rovnyak et al. 2011). These advantages make NUS an attractive method for a broad range of NMR experiments. Additionally, recent updates to spectrometer software make NUS acquisition a simple runtime option. The wide range of NMR experiments of differing dynamic range necessitates an efficient, general processing scheme, which seamlessly functions on a variety of computational platforms, ranging from laptops to computational clusters, and accommodates all experiment types in order to facilitate common adaptation of NUS methods.

The application of NUS for NMR experiments requires two additional steps compared to conventional data acquisition and processing: (1) design of a NUS schedule (for the purposes of this report, only "on-grid" sampling will be considered); and (2) reconstruction of the unsampled data points with a suitable algorithm to create a uniformly sampled grid. Ideally, such an algorithm would utilize both information available from sampled data points and assumptions about the properties of the NMR signals.

For the first step, it is desirable to use sampling schedules that produce spectra with high fidelity (resonance frequencies and intensity) relative to a uniformly sampled counterpart while using the minimal acquisition time possible. The complex relationship of these requirements is an ongoing area of research (Hyberts et al. 2012a; Maciejewski et al. 2012). Nevertheless, it has been demonstrated recently that one can obtain high quality multidimensional spectra even if an "aggressive" undersampling NUS schedule (e.g. <1 % of sampled points) has been used (Hyberts et al. 2012b), though we do not specifically advocate such a degree of undersampling. This phenomenon can be explained by the theory of compressed sensing (CS) (Candès et al. 2006a, b; Candès and Tao 2005, 2006; Donoho 2006) that proves for a suitably sparse sample, a high quality reconstruction of the original signal is guaranteed when the following condition has been met:

$$n > cK\log(N/K) \quad (1)$$

where N is the full length of the vector, K is the number of non-zero values (peaks) in the vector, c is a small constant whose value is often empirically determined, and n is the number of points sampled. However, as noted previously (Hoch and Stern 1996; Kazimierczuk and Orekhov 2011; Mayzel et al. 2014), NMR data are not strictly sparse, as the term is defined by the CS equation. Nevertheless, as is implied by the equation, the principle driver of the required number of points (n) is the number of anticipated peaks (K), which we use as a guideline (with some additional percentage to account for error) for determining how many experimental points to collect.

The sampled points are dispersed over a grid of larger size, causing the sparsity of the data acquisition and the resolution of the final spectrum to increase without degrading quality. Since the number of NMR signals is generally independent of spectrum dimensionality, the number of sampled points required by NUS will not increase—at least not in a geometric fashion—when higher dimensional experiments are employed. Rather, the number of sampling points required is related to the number of spectral components and, hence, is correlated with the experiment type and size of the molecule under study. Therefore, dramatic under-sampling can be applied in high-dimensional NMR experiments, especially in 4D experiments. The design of sampling schemes remains an intense area of research (Aoto et al. 2014; Hyberts et al. 2012a; Maciejewski et al. 2012) and will not be discussed further in this report.

In the second step, reconstruction of the uniformly sampled data begins with creation of a grid where the sampled data points occupy their expected position in the final, reconstructed matrix and unsampled points are set to zero. Each vector of the matrix is then processed with an algorithm that minimizes a target function, as has been elegantly described for the maximum entropy (Hoch and Stern 1996) and multidimensional decomposition (MDD) (Orekhov et al. 2003; Orekhov and Jaravine 2011) methods. There are two important considerations in this process. First, the reconstructed data needs to be consistent with the experimental data at the sampled points. Consistency can be achieved by including a "data consistency term" (e.g. the sum of the square of the deviations) in the target function to penalize deviations of reconstructed data points from originally sampled ones or by keeping the experimentally sampled points unchanged during the procedure. Second, the reconstructed data can be modeled mathematically to include a property (or term in the target function) that, when minimized, represents the best reconstruction of the missing points onto the full data grid. This term is typically referred to as a "regularization term". Approaches such as these have been shown to enable a convergent path to the optimal reconstruction (Hoch and Stern 1996). This modeling procedure can utilize different properties of the signals, thus yielding different regularization terms and associated target functions. The choice of a specific regularization term leads to different NMR data processing methods, such as maximum entropy reconstruction (Hoch et al. 2014; Hoch and Stern 1996), MDD (Orekhov et al. 2003; Orekhov and Jaravine 2011), l1-norm minimization (Bostock et al. 2012; Hyberts et al. 2007, 2009; Stern et al. 2007) and iterative re-weighted least squares (IRLSs) (Kazimierczuk and Orekhov 2011). For a given model, various numerical optimization or minimization techniques are utilized to minimize the

target function and reconstruct the unsampled data points. The choice of a particular regularization term and the numerical algorithm used to minimize the target function constitute the core of a NUS NMR data processing method.

In this report, we consider several models of signal properties and regularization terms that can be used for processing NUS NMR data. Generally, NMR signals are sparse in the frequency domain of multidimensional experiments, which results in the majority of the data points being close to zero—i.e. they fall within the Gaussian distributed noise which is characterized by a standard deviation σ. This is especially true for the indirect dimensions of multidimensional NMR experiments, where the data corresponding to a single frequency along the directly detected dimension contains relatively few signals compared to a large number of data points. Within the constraint of data-consistency (vide supra), one needs to minimize the total number of non-zero points, hereafter referred to as the l0-norm, in the frequency domain. Extensive research in the CS-field has demonstrated that l0-norm minimization by an exhaustive combinatorial search is generally impractical for large scale problems, such as multidimensional NMR spectra. For example, a 3D NMR experiment with two NUS dimensions typically requires reconstruction of ~6000 (e.g. 64 × 96) hypercomplex points for every discrete frequency along the direct dimension (~512–1024 points). For such a task, a combinatorial search is impractical with current computational power. Instead, many regularization terms that exploit the "sparse" property of signals have been proposed. We consider the application of three of these terms to NUS NMR data and present a highly efficient optimization procedure (NESTA), which yields a general and complete processing package (NESTA-NMR) capable of handling both low-to-medium dynamic range experiments (such as triple-resonance assignment experiments) and high-dynamic range, quantitative experiments (such as NOESY, J-modulated dipolar coupling measurements, or relaxation spectra).

Regularization terms and reconstruction algorithms

In the context of NMR data processing, the l1-norm (L1, the sum of absolute values) regularization has been demonstrated to be suitable for high dynamic range (NOESY-type) NUS NMR data (Bostock et al. 2012; Hyberts et al. 2007, 2009; Kazimierczuk and Orekhov 2011). For complex data, L1 has been defined previously (Hyberts et al. 2007, 2009; Kim et al. 2007; Wright et al. 2009) as

$$\|f\|_{l1} = \sum |f_k| \tag{2}$$

where

$$|f_k| = \sqrt{f_{k,rr}^2 + f_{k,ri}^2 + f_{k,ir}^2 + f_{k,ii}^2} \tag{3}$$

and

$$|f_k| = \sqrt{f_{k,rrr}^2 + f_{k,rri}^2 + f_{k,rir}^2 + f_{k,rii}^2 + f_{k,irr}^2 + f_{k,iri}^2 + f_{k,iir}^2 + f_{k,iii}^2} \tag{4}$$

for frequency domain data that contain two and three NUS dimensions, respectively. Hence, processing NUS data can be reduced to a numerical minimization of L1 as defined above. In addition to L1, we consider several regularization terms based on the sparse assumption that has been used in the CS field. Candès et al. have demonstrated that using an iteratively re-weighted l1-norm (IRL1) as the regularization term can further improve data reconstruction from incomplete measurements (Candès et al. 2008). IRL1 is defined as

$$\|f_{ir}\|_{l1} = \sum \omega_k |f_k| \tag{5}$$

where

$$\omega_k^{i+1} = 1/\left(|f_k|^i + \varepsilon\right) \tag{6}$$

except for the first iteration, where $\omega_k = 1.0$.

The point-wise weight ω_k^{i+1} at the $i + 1$th iteration is calculated from $|f_k|^i$, which has the same definition as described above for the ith iteration. The parameter ε is set to a small positive value (e.g. 0.1) to avoid division by zero. This approach bears some conceptual similarity to the IRLS regularization method (Kazimierczuk and Orekhov 2011). Another derived regularization term that meets the sparse assumption is the Gaussian smoothed l0-norm (Gaussian-SL0) (Mohimani et al. 2009; Trzasko et al. 2007), defined as

$$\|f\|_{sl0} = \sum \left(1 - e^{-0.5|f_k|^2/\sigma^2}\right) \tag{7}$$

when σ is very small relative to the signal amplitude $|f_k|$ (vide supra), Gaussian-SL0 is a good approximation of the l0-norm as the term $1 - e^{-0.5|f_k|^2/\sigma^2}$ rapidly approaches either 1 or 0.

Minimization of target functions that include the regularization terms described above (L1, IRL1, Gaussian-SL0, and many others) is an active research topic in the CS field. CS techniques have also been widely applied in signal and image processing, for instance in MRI (Lustig et al. 2007). The development of such algorithms provides opportunities to leverage these accomplishments for the purpose of processing NUS NMR data. Recently, the NESTA algorithm was introduced by Becker et al. as a first

order method for fast and accurate signal recovery or image reconstruction (Becker et al. 2011). It was demonstrated that NESTA, which implements Nestrov's ideas (2005), can rapidly and accurately recover noisy compressed signals with very large dynamic range (~60 dB power). The method incorporates: (1) the coupling of smoothing techniques with gradient methods for optimizing non-smooth functions; and (2) first-order methods with very rapid convergence rates. The complete mathematical description and rigorous analysis of the NESTA algorithm, as well as a comparison between NESTA and several state-of-the-art L1 minimization algorithms, can be found elsewhere (Becker et al. 2011). The salient features of NESTA and our adaptation of the algorithm for processing hypercomplex NMR data are briefly described here. First, the NMR data is processed in the direct dimension, and a region of interest (e.g. the left half of an H^N detected spectra) is extracted. This data is then divided along each frequency domain point in the direct dimension. For each of these slices, the corresponding data is shuffled according to the sampling schedule and the remaining (unsampled) points are zero-filled. This slice is two- and three-dimensional for 3D and 4D NMR experiments, respectively. Second, based on the definition of L1 for hypercomplex NMR data (vide supra), the point-wise gradient is computed. Sampled points are retained while unsampled points are updated according to the gradient and its previous value. The process of computing the gradient and updating the unsampled data points is repeated until convergence is reached. The NMR data can then be processed using standard methods as if it were uniformly sampled. For ILR1, an additional outer loop is introduced for weighting of the l1-norm. Inside this additional loop, the re-weighted l1-norm is minimized using slightly different gradient calculations, and then point-wise weights are updated. Gaussian-SL0 is a smooth, non-convex function, and, in our experience, such functions gain little acceleration from the NESTA algorithm. Furthermore, the minimization procedure for Gaussian-SL0 has a tendency to become trapped in local minima; hence, a procedure described by Mohimani et al. was adopted to improve the performance of Gaussian-SL0 (Mohimani et al. 2009). More detailed description and a flow chart for the data processing procedure are provided in the Supplementary Information.

Materials and methods

Experimental data were collected on three different proteins: (1) a 1 mM sample of the 8 kDa CUE domain containing residues 453–504 from human gp78 (Liu et al. 2012); (2) a 330 μM sample of the 15 kDa PH domain of ASAP1, which contains residues 339–451 (Luo et al. 2008); and (3) a 400 μM sample of the 32 kDa two domain construct (ZA) of ASAP1 containing residues 441–724 (Luo et al. 2008). Isotope labeling was performed by expressing and purifying the proteins from *E. coli* using standard techniques to produce either uniform ^{13}C, ^{15}N-labeled protein, uniform ^{2}H, ^{13}C, ^{15}N-labeled protein (DCN), or uniform ^{2}H, ^{13}C, ^{15}N, $^{13}C^{1}H_3$-methyl (Ileδ1, Leu, Val) labeled protein (DCN-ILV) or ^{2}H, ^{15}N, $^{13}C^{1}H_3$-methyl (Ileδ1, Leu, Val) labeled protein ($DC_{methyl}N$-ILV).

NESTA-NMR has been used to process a wide range of 3D and 4D NMR experiments that were collected on these three samples, and the salient information of all of these experiments is listed in Supplemental Table 1. Data discussed explicitly in the manuscript consist of the following four data sets:

1. A 4D methyl–methyl HMQC-NOESY-HMQC experiment (4D CC-NOESY) utilizing mixed constant-time evolution (Ying et al. 2007) was recorded on 1 mM DCN-ILV gp78 CUE using a Bruker Avance 900 MHz instrument running TopSpin 2 with cryoprobe at 298 K. The standard pulse sequence was modified to store all of the hypercomplex pairs adjacent to each other with quadrature modulations preceding time modulation and the delays in the indirect dimensions calculated according to a NUS sampling schedule. In order to compare reconstructions with those of different programs, the sampling schedule was produced by an in-house Python script according to the algorithm described by Mobli et al. (2010) which was additionally modified to ensure every index for a given dimension contained at least one sampling point. Sampling consisted of 7200 NUS points taken on a 48 ^{13}C × 32 ^{1}H × 48 ^{13}C grid with a sampling density of 9.8 %. In this report, the number of points of an indirect dimension is described in complex points—i.e. real and imaginary data are counted as one point. The maximum evolution times in the indirect dimensions were 11.5 ms for both ^{13}C dimensions and 34.1 ms for the indirect ^{1}H dimension. Spectral widths were 4098 Hz for both ^{13}C dimensions and 909 Hz for the indirect ^{1}H dimension. Each FID was recorded with 4 scans, and the NOE mixing period was 150 ms.
2. A variable (non-constant) time 4D methyl–methyl HMQC-NOESY-HMQC (Diercks et al. 1999) experiment was acquired on a 400 μM sample of $DC_{methyl}N$-ILV ZA on a Bruker Avance III 600 MHz instrument with cryoprobe at 298 K using TopSpin 3.2. The sampling schedule was designed with ScheduleTool, which is distributed

with RNMRTK (Hoch and Stern 1996), and consisted of 12,000 NUS points taken on a 48 $^{13}C \times 64$ $^{1}H \times 48$ ^{13}C grid with a sampling density of 8.1 %. The maximum evolution times in the indirect dimensions were 12.2 ms for both ^{13}C dimensions and 19.0 ms for the indirect ^{1}H dimension. Spectral widths were 3922 Hz for both ^{13}C dimensions and 3360 Hz for the indirect ^{1}H dimension. Each FID was recorded with 4 scans and the NOE mixing period was 200 ms.

3&4. Two 3D ^{15}N-edited NOESY-HSQC experiments were acquired on a 330 μM ^{15}N-labeled PH domain on a Bruker Avance III 600 MHz instrument with cryoprobe at 298 K using the Topspin 3.2 library pulse sequence *nosesyhsqcf3gp193d* (Sklenar et al. 1993). One data set was collected with uniform sampling (36 $^{13}C \times 180$ ^{1}H) and serves as the reference. The other was collected with 1620 NUS points (25 % sampling density) on a 36×180 grid. The sampling schedule was designed with ScheduleTool. For both experiments, the maximum evolution times in the indirect dimensions were 18.5 ms for ^{15}N and 25 ms for ^{1}H. Spectral widths were 1945 Hz for ^{15}N and 7194 Hz for ^{1}H. The NOE mixing period was 60 ms. Each FID contained 8 and 32 scans for the uniformly sampled and non-uniformly sampled data, respectively.

Data reconstruction was performed using in-house C programs for both the NESTA algorithm and alternative algorithms used for comparison. This was done to enable direct comparison of convergence rates since package-specific implementations may affect computing efficiency. Thus, all the algorithms utilized the same libraries and were compiled on the same computer. Mixed-radix FFT and IFFT routines from the GNU Scientific Library (GSL) (Galassi et al. 2009) capable of processing complex vectors of any length (not restricted to powers of 2) were used to construct multidimensional subroutines to transform hypercomplex data. Direct comparison of algorithms rather than a specific software package is enabled because the algorithms utilize the same libraries and the analysis of computational efficiency is measured by the number of iterations required to reach convergence.

The processing package NESTA-NMR was developed to apply NESTA minimization to 2D, 3D, and 4D NMR data. Data described in this manuscript were processed on a desktop computer running Centos 6 with a 2.13 GHz Intel Xeon processor containing 4 hyperthreaded cores (8 threads) or a Mac Pro with a 3.5 GHz Intel Xeon processor containing 6 hyperthreaded cores (12 threads). The software can also be run on a cluster to access even more threads; however, this is not generally necessary given the relatively short computational times of NESTA-NMR, even for 4D data (vide infra). After reconstructing the unsampled data points and merging them with experimentally sampled data, the indirect dimensions were processed with NMRPipe (Delaglio et al. 1995) using standard FFT methods for transformation and visualized using Sparky (Goddard and Kneller).

Results

NESTA-NMR: a general purpose NUS processing engine

We implemented the NESTA algorithm in the software package NESTA-NMR for processing multidimensional (2D, 3D, and 4D) NMR data using a variety of regularization terms: L1, IRL1, and Gaussian-SL0. For practical purposes, it is advantageous to process the direct dimension of NUS NMR data first and extract only the region of interest along this dimension. The appropriate regularization method can then be performed on the data corresponding to each point in the direct dimension. Reconstruction of each of these smaller data sets is completely independent of the others. This separation enables a simple parallel computing paradigm and, additionally, alleviates memory issues associated with higher dimensional NMR data files. A similar approach has been adopted by MddNMR (Orekhov et al. 2003; Orekhov and Jaravine 2011) and hmsIST (Hyberts et al. 2012b) for processing NUS NMR data. NESTA-NMR supports multithreading and this feature has been implemented using the C standard library. Thus, no additional software installation or scripts are required to enable this feature. Data reconstruction by NESTA-NMR can be performed in parallel for various computational environments (e.g. laptops, desktops, computing clusters, etc.).

The general flow of data processing is equivalent for 2D, 3D, and 4D data. The data are first converted into NMRPipe format. Customized NMRPipe macros for Bruker and Agilent data are included with NESTA-NMR that implement the Rance-Kay protocol (Cavanagh et al. 1991; Kay et al. 1992; Palmer et al. 1991, 1992) for frequency discrimination on NUS data. Other frequency discrimination protocols (States, States-TPPI) do not require additional processing steps. Using NMRpipe, the direct dimension is processed, which includes apodization, Fourier transformation, phasing, and extraction of the region of interest. Reconstruction is then performed using NESTA-NMR (see Supplemental Information for more details on the program), which returns the reconstructed data in NMRPipe format. NESTA-NMR requires only the data and the sampling scheme, in the same

format as is used by the instrument software for data collection (see Supplemental Information for more details). NESTA-NMR requires only a single command to execute, which can either be embedded inside the NMRPipe script or run from the command line. NMRPipe is then used for processing the indirect dimensions of the reconstructed data.

The core package of NESTA-NMR is modular and enabled simple comparison of other minimization algorithms. Separate routines were written that implement L1 and IRL1, in addition to the Gaussian-SL0 algorithm. The distributed version of NESTA-NMR contains all three of these algorithms.

NESTA is faster than IST algorithms when using L1 regularization

The most time-consuming operations in the reconstruction of NUS data are multidimensional fast Fourier transforms (FFTs) and inverse FFTs (IFFTs), collectively referred to as FFTs hereafter. Multidimensional FFTs are constructed from a series of one-dimensional FFTs, each of which has a computational cost $\mathcal{O}$ ($N\ logN$), where N is the number of complex points. A 3D FFT operation for a hypercomplex cube with dimensions (m, n, q) is composed of $4nq + 4mq + 4mn$ FFTs. Suppose m, n, q each have the value 50, then a 3D FFT operation requires $\sim$120,000 FFT operations. The utilization of different FFT libraries in addition to variations in computer hardware will impact computation time. Hence, for an unbiased measure of efficiency, it is best to evaluate the number of iterations required for convergence by different algorithms.

Using experimental 4D CC-NOESY NMR data (data set 1), we systematically tested several l1-norm minimization algorithms. We compared the convergence rate of three l1-norm minimization algorithms in processing a single 3D slice (a cube) of data set 1 (see "Materials and methods" and Fig. 1): IST-S (Kazimierczuk and Orekhov 2011; Stern et al. 2007), IST-D (Drori 2007) and NESTA (Becker et al. 2011). Because two versions of IST were tested, the terms IST-S and IST-D are used to differentiate these algorithms by the last name of first author of the corresponding literature. To insure an unbiased comparison, these algorithms are implemented in the same software framework and utilize the same multidimensional FFT subroutines (described in the "Materials and methods" section). There are several different implementations of the popular iterative soft thresholding algorithm reported in the NMR literature (Bostock et al. 2012; Drori 2007; Kazimierczuk and Orekhov 2011; Stern et al. 2007). Hyberts et al. implemented the Drori IST (IST-D) algorithm in hmsIST and recommended updating the threshold with a scaling factor 0.98 to gradually scale down the threshold from a large value to small value (Hyberts et al. 2012b). We used this scaling factor (0.98) for both IST-S and IST-D to test the convergence rate. The NESTA algorithm incorporates a parameter, μ, which is the smoothing factor for the gradient (see Supplementary Information) and is similar to the threshold used in IST algorithms. This parameter is also gradually scaled down from a large initial value (90 % of the largest absolute value) to a value of 0.002 within 15–30 steps. For this reconstruction, 30 steps were used, which we have found to be suitable for virtually all data types. The scaling of μ only happens when convergence is reached, which is determined to have occurred when the difference between the current L1 and the average L1 of ten prior runs is smaller than a predefined value.

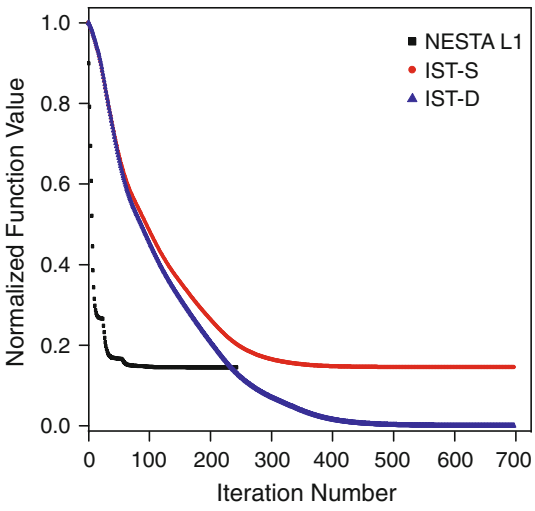


Fig. 1 Comparison of convergence rates for l1-norm minimization algorithms using iterative soft thresholding (IST) or NESTA (*black*) algorithms. Two IST algorithms are shown: IST-S (*red*) (Kazimierczuk and Orekhov 2011; Stern et al. 2007) and IST-D (*blue*) (Drori 2007). The algorithms were used to reconstruct a 4D NUS CC-NOESY spectrum of 1 mM DCN-ILV gp78 CUE domain. Function values at every iteration are normalized against the initial function values before optimization. Both NESTA L1 and IST-S preserve sampled points during optimization, which explains their convergence to similar function values. IST-D does not preserve sampled data, and its different convergence value is a reflection of this fact, rather than of the relative accuracy or quality of its reconstruction relative to NESTA L1 and IST-S

The two IST algorithms converge after approximately 400–500 iterations for each 3D hypercomplex cube, and the NESTA algorithm reaches convergence in <100–150 total iterations. While the actual number of iterations used by these algorithms may vary with the data being processed, NMR experiment type, parameters chosen for reconstruction, and/or the convergence criteria, we consistently find a similar ratio of performance between NESTA, IST-S and IST-D. It is worth noting that, in both IST-S and our implementation of NESTA, all of the sampled points are kept unchanged during the course of optimization. This variation in the treatment of data consistency explains the different minima reached by the

algorithms in Fig. 1. The method utilized by NESTA-NMR and IST-S implicitly ensures complete consistency with the experimental data and avoids both the necessity of measuring the noise level and of adjusting the weight of the regularization terms relative to the data consistency term. However, as noted by Stern et al. (2007), fine spectral features present in reconstructions performed by methods that retain all experimental data require further analysis to be deemed statistically significant. Algorithms that use unconstrained optimization and Bayesian procedures to minimize both data consistency and regularization terms, such as Maximum Entropy (Hoch and Stern 1996), do not suffer from the aforementioned artifacts (Stern et al. 2007) but generally require the setting of parameters associated with noise and data consistency.

We have also benchmarked computation times with NESTA-NMR. Two-dimensional data sets are processed with virtually no additional time compared to the processing of uniformly sampled data, while 3D data sets are reconstructed in $\sim$1–5 min. Four dimensional reconstruction of data set 1 (909 direct points in a 48 $\times$ 32 $\times$ 48 grid) using 30 iterations of μ for NESTA L1 requires 0.8 h on a Mac Pro with 12 threads and 1.2 h on a Linux desktop with 8 threads. For data set 2, the reconstruction of 849 direct points in a 48 $\times$ 64 $\times$ 48 grid requires 1.6 and 2.7 h on the respective hardware. The difference in reconstruction times is likely attributable to the greater number of peaks present in data set 2 (acquired on the 32 kDa ZA construct) relative to data set 1 (acquired on the 8 kDa CUE domain).

Preservation of spectra fidelity with NESTA L1 reconstruction

The ability of the NESTA algorithm to reconstruct weak NOESY peaks in a 4D CC-NOESY spectrum was examined for a small protein domain, CUE, from the ubiquitin E3 ligase gp78 (Liu et al. 2012). The sample was ^{2}H, ^{13}C, ^{15}N-labeled with the exception that Ileδ1, Leu, Val methyl groups were protonated (^{1}H). This labeling scheme, referred to as ILV-labeling, has become popular for obtaining methyl–methyl NOE constraints to generate moderate resolution structures (Tugarinov and Kay 2003). The narrow distribution of proton signals of Ile-δ1, Leu, Val methyl groups of the gp78 CUE enables use of a relatively narrow spectral width (see the "Materials and methods" section) in the ^{1}H dimensions. Limited chemical shift dispersion for the Ile, Leu, and Val methyl signals in the ^{13}C dimensions requires the use of NUS to enable digital resolution sufficient for unambiguous assignment. Use of NUS sampling (7200 NUS points on a 48 $\times$ 32 $\times$ 48 grid with 9.8 % sampling density) required a total acquisition time of 88 h (3.7 days), whereas sampling the equivalent uniform grid would require 37.5 days if utilizing the same four step phase cycle.

Reconstruction of the experimental data using several different l1-norm minimization algorithms (IST-S, IST-D and NESTA L1) provided very similar results (Fig. 2). A weak long-range NOE, corresponding to V34-γ1:L28-δ2 with a Cm–Cm distance of 5.8 Å, is $\sim$1/1000 the intensity of the diagonal peak and is easily detected and quantified. Data processed with the NESTA L1 algorithm are equivalent or superior to that processed with IST-S or IST-D, based on higher intensities and cross peak-to-diagonal intensity ratio for weak peaks. These results demonstrate the suitability of NUS methods, and of NESTA L1 in particular, for acquiring and processing high dynamic range NOESY NMR data. Additionally, cross peaks observed in the small spectral region shown in Fig. 2 are consistent with the CUE structure (Fig. 3) previously determined by NMR (Liu et al. 2012).

To test the ability of the NESTA algorithm to quantitatively reconstruct weak NOESY peaks, we assessed the accuracy of peak intensities in a 3D NUS (25 % sparsity) ^{15}N-edited NOESY-HSQC spectrum (reconstructed with NESTA L1) compared to a reference spectrum collected with uniform sampling (Fig. 4). The normalized intensities of the cross peaks in the NUS spectrum demonstrate excellent agreement when compared to the same peaks in the uniformly sampled spectrum (Fig. 5a). To ensure variations in peak intensity most accurately reflect differences relating to reconstruction with NESTA-NMR, NUS intensities in Fig. 5 were taken from a spectrum created by resampling the uniformly sampled spectrum using the same sampling schedule as the NUS spectrum depicted in Fig. 4. The critical issue for spectrum fidelity in NOESY experiments is the accurate representation of weak cross peaks correlating to long-range distances. A correlation plot of these weak cross peaks (Fig. 5b and highlighted in the boxed area of Fig. 5a) demonstrates that the intensities are accurately reconstructed with NESTA L1. The slightly reduced slope indicates some non-linearity in the weakest peaks; however, this deviation is quite small and would not represent any error in the distance constraint assigned to these NOEs in the typical structure calculation protocols.

Comparison of IRL1 and Gaussian-SL0 regularization terms

In addition to L1, two other regularization methods (IRL1 and Gaussian-SL0) were compared by processing the same 4D CC-NOESY (Fig. 6) and 3D ^{15}N NOESY-HSQC spectra (Fig. 7). All three spectra are equivalent in peak representation. Interestingly, in the spectra reconstructed by IRL1 with 5 external re-weighting iterations (IRL1-W5) and Gaussian-SL0, the cross peaks generally have slightly

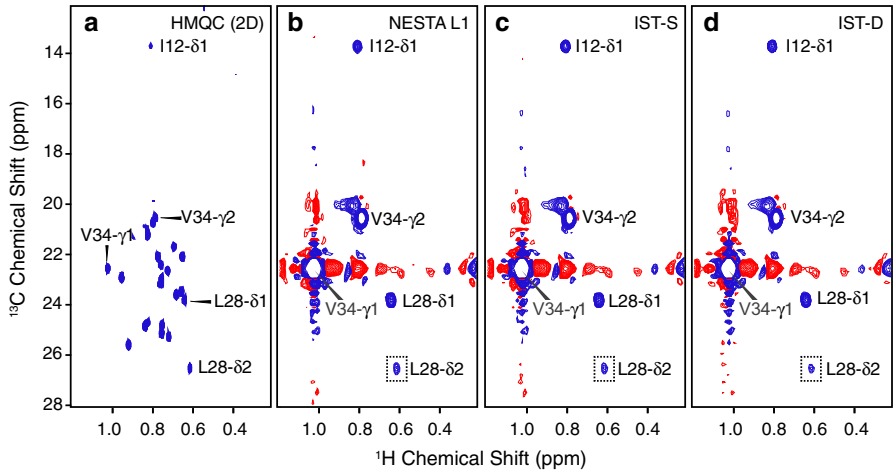


Fig. 2 Comparison of regions of 4D NUS CC-NOESY spectra of 1 mM DCN-ILV gp78 CUE domain to (**a**) a 2D ^{1}H–^{13}C HMQC collected with uniform sampling on the same sample and processed with standard methods. The 4D CC-NOESY was reconstructed with different l1-norm minimization algorithms: **b** NESTA L1, **c** IST-S, and **d** IST-D. The slices (**b**–**d**) correspond to the frequency of Val 34-γ1 (*gray*). A weak cross peak (*dashed box*) corresponding to an NOE between methyl groups Val 34-γ1 and Leu 28-δ2 has ~1/1000 of the intensity of the diagonal peak corresponding to Val 34-γ1. Slices from the reconstructed 4D spectra are plotted using the same contour level. Spectral acquisition parameters are given in the "Materials and methods" section

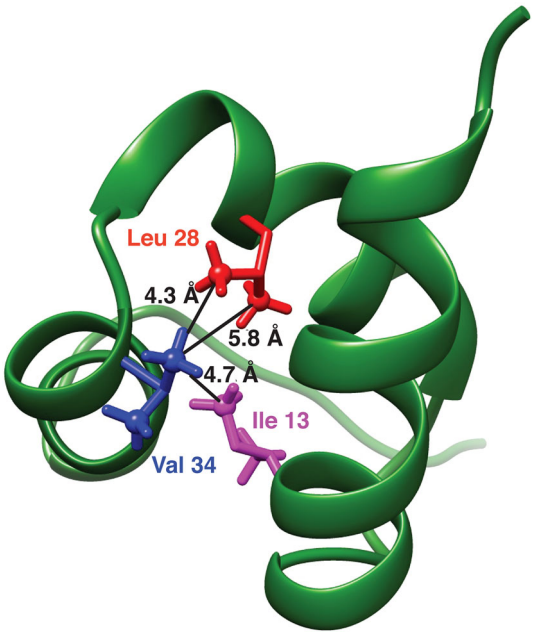


Fig. 3 Structure of human gp78 CUE from PDB 2LVN (Liu et al. 2012). Distances corresponding to the long-range NOEs observed in Fig. 2 are indicated. Methyl groups of Ile 13 (*magenta*), Leu 28 (*red*), and Val 34 (*blue*) are shown. Distances for Ile 13-δ1 to Val 34-γ1 (4.7 Å), L28-δ1 to Val 34-γ1 (4.3 Å), and L28-δ2 to Val 34-γ1 (5.8 Å) are indicated with *black lines*

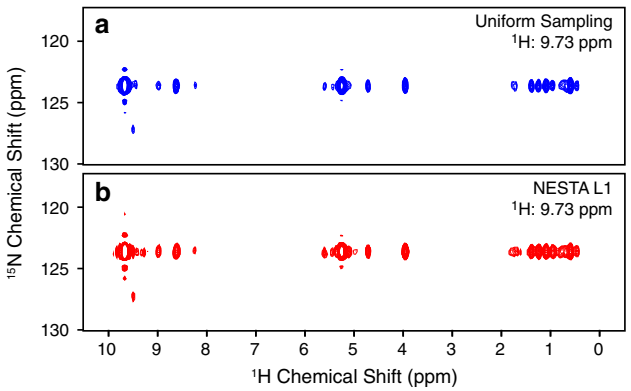


Fig. 4 Slices of 3D ^{15}N-edited NOESY-HSQC spectra of 330 μM ^{15}N-labeled PH domain obtained with (**a**, *blue*) uniform sampling and processed using standard methods or (**b**, *red*) with NUS and reconstructed with NESTA L1. Spectral acquisition parameters are given in the "Materials and methods" section

higher intensities than those in the spectrum reconstructed with L1. The crosspeak-to-diagonal intensity ratios for L28-δ2/V34γ1 (the weakest cross peak in this slice) are 1.16×10^{-3}, 1.32×10^{-3}, and 1.54×10^{-3} for L1, IRL1-W5 and Gaussian-SL0, respectively (Fig. 6b–d). This result suggests that IRL1 and Gaussian-SL0 are better at preserving the intensities of weak peaks. However, the spectrum processed with Gaussian-SL0 regularization has many residual artifacts (Fig. 6d), indicating it is best used in conjunction with a reconstruction from one of the two l1-norm regularizations.

To further explore the differences in peak intensity obtained with the three algorithms, the linearity of the peaks in a NUS 3D ^{15}N-edited NOESY-HSQC spectrum reconstructed with NESTA-NMR using Gaussian-SL0 regularization was compared to the uniformly sampled spectrum (Fig. 7a). NUS peak intensities were determined in a fashion analogous to those of Fig. 5. The intensities of the

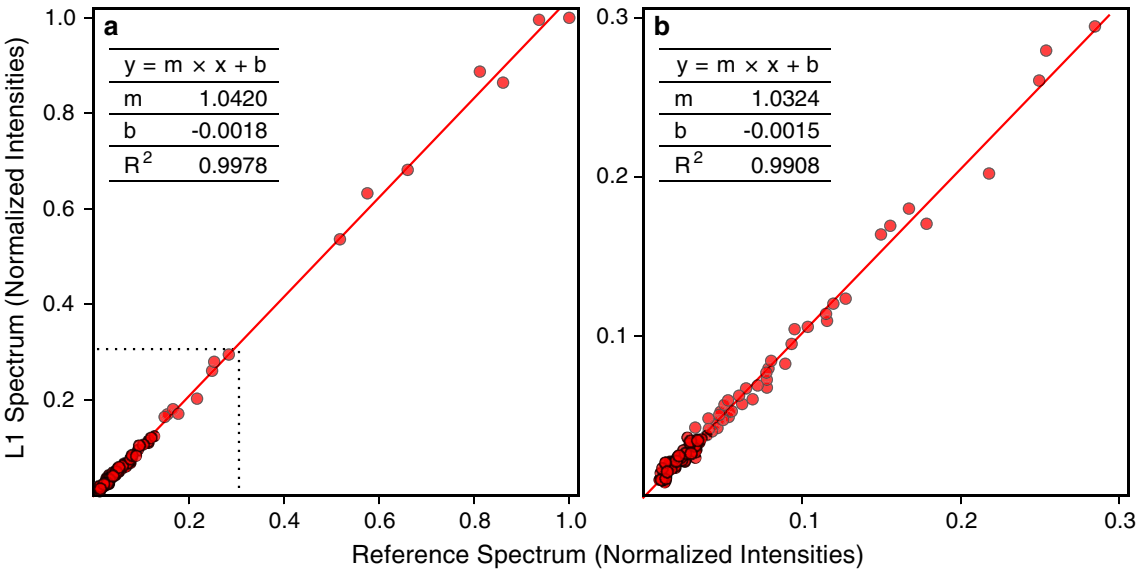


Fig. 5 Correlation of peak intensities between uniformly sampled and NUS ^{15}N NOESY-HSQC spectra processed using standard methods or reconstructed using the NESTA algorithm and L1 regularization. Both spectra were processed with NMRPipe and peak analysis was performed with SPARKY. Correlation plots of peak intensities corresponding to NOEs for seven residues are shown as **a** a total of 96 peaks including both diagonal and cross peaks, and **b** 89 weak cross peaks corresponding to the *boxed region* of **a**. The *red line* and *table* in each *panel* show the results of a linear regression performed on the respective peaks. To ensure the correlations depicted in **a** and **b** reflect only errors in reconstruction, the NUS peak intensities are from a spectrum created by resampling the uniformly sampled spectrum. However, the agreement between the two independently acquired data sets is also excellent ($R^2 = 0.9933$ for all peaks)

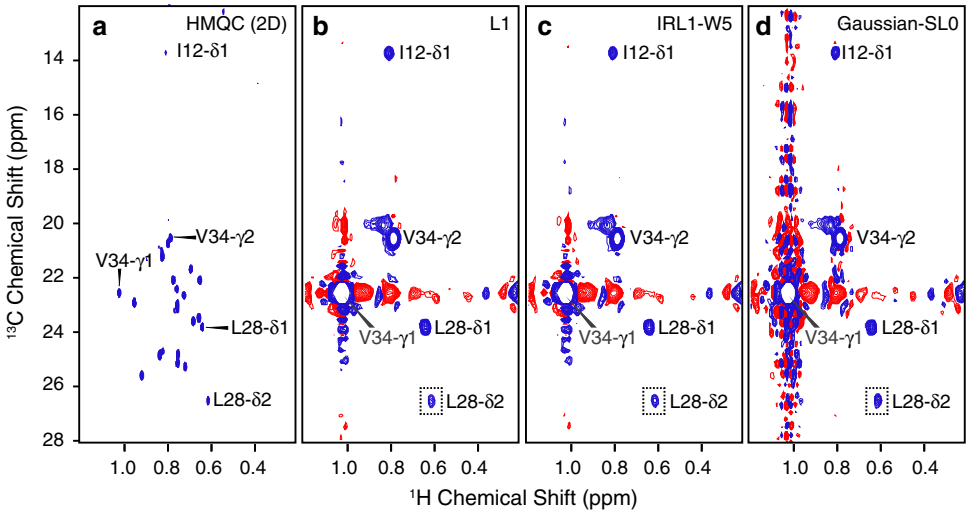


Fig. 6 Regions of a 4D CC-NOESY spectra collected with NUS of 1 mM DCN-ILV gp78 CUE domain compared to **a** a 2D ^{1}H–^{13}C HMQC collected with uniform sampling on the same sample and processed with standard methods. The 4D CC-NOESY was reconstructed with **b** L1 regularization, **c** five rounds of IRL1 regularization, or **d** Gaussian-SL0 regularization. The slices (**b**–**d**) correspond to the frequency of Val 34-γ1 (*gray*). Spectra reconstructed with different regularization terms are consistent with the uniformly sampled NMR data with regards to expected cross peaks (Das et al. 2009). Slices from the reconstructed 4D spectra are plotted using the same contour level. The *dashed box* is described in the caption to Fig. 2

weak peaks in the Gaussian-SL0 spectrum are plotted against those from the uniformly sampled spectrum (Fig. 7b), and a slope (1.0319) is observed that is similar to that from L1 regularization (1.0324, Fig. 5b). However, the slope derived from all peaks (1.0292, Fig. 7a) is slightly closer to unity than that of L1 (1.0420, Fig. 5b). This indicates Gaussian-SL0 preserves the intensities of peaks more accurately than L1 in certain situations. The same

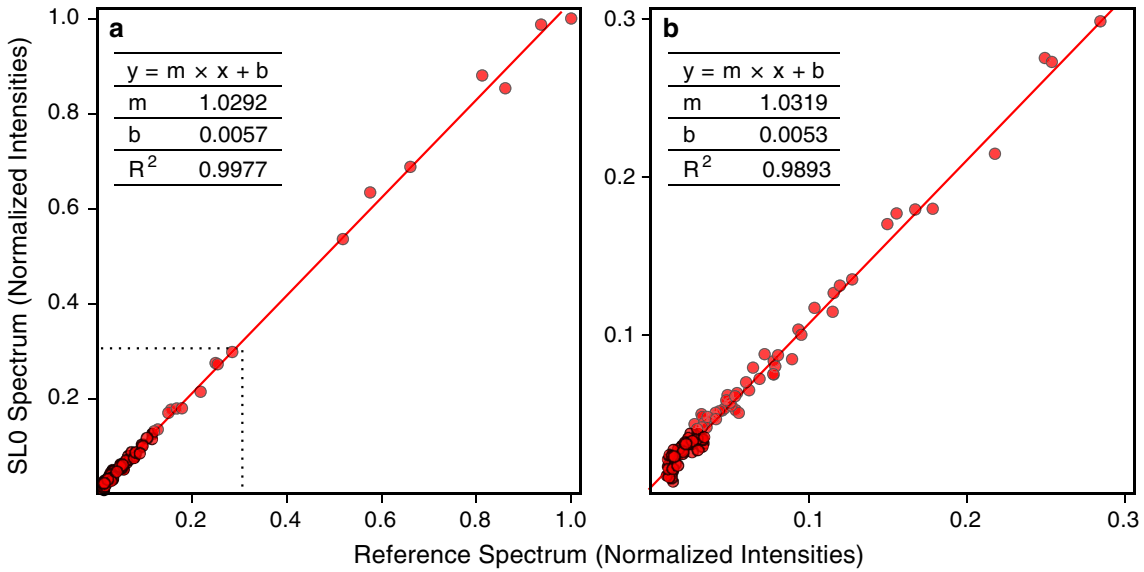


Fig. 7 Correlation of peak intensities between uniformly sampled and NUS ^{15}N NOESY-HSQC spectra processed using standard methods or reconstructed using the NESTA algorithm and Gaussian-SL0 regularization. Correlation plots of peak intensities corresponding to NOEs from seven residues are shown as **a** a total of 96 peaks including both diagonal and cross peaks, and **b** 89 weak cross peaks corresponding to the *boxed region* of **a**. The *red line* and *table* in each *panel* show the results of a linear regression performed on the respective peaks. The NUS peak intensities are derived as described in the caption to Fig. 5. The R^2 between peak intensities from two independently acquired data sets is 0.9937 (data not shown)

trend can be found for IRL1 regularization relative to uniformly sampled data (data not shown). Additionally, the intercepts show very small deviations from zero (Figs. 5; 7), indicating that constraints derived from weak peaks in NUS-reconstructed NOESY spectra can confidently be used in structure calculations.

Generality of NESTA L1 reconstruction and NESTA-NMR

The application of NUS to the breadth of NMR experiments in biomolecular studies is greatly enhanced if a common protocol can be used for all spectral types and molecular systems. We demonstrate this capability for NESTA-NMR by collecting data ranging from a 4D CC-NOESY spectrum of ILV-labeled D/C/N ZA domain of ASAP1 (32 kDa) (Fig. 8a) to the full range of triple resonance experiments used for backbone and sidechain assignment in proteins ranging from 8 to 32 kDa (Supplemental Table S1). The computational efficiency and robustness of NESTA-NMR allows the same protocol to be applied to this broad range of experiments, greatly simplifying usage. Of particular interest is the 4D HCCH(CO)NH TOCSY experiment (Fig. 8b). This spectrum was collected on the 15 kDa PH domain of ASAP1 with a sparsity of 1 %; nevertheless, the reconstructed data provide excellent resolution and spectral integrity. This data enabled assignment of sidechain ^{1}H and ^{13}C resonances in this protein.

Discussion

Four-dimensional NMR experiments are powerful tools for the determination of structural information. The separation of signals along additional dimension(s) greatly reduces degeneracy in resonance assignment and quantitation. However, the primary limitation of the routine use of 4D experiments is the excessive experimental time required to acquire data with sufficient digital resolution. The incorporation of NUS greatly reduces the necessary acquisition time and enhances resolution and sensitivity, making 4D NMR experiments a more attractive choice. Several pioneering studies (Bostock et al. 2012; Hyberts et al. 2007, 2009, 2012b; Kazimierczuk and Orekhov 2011) have demonstrated that L1 minimization is particularly suitable for processing NUS data that has a high dynamic range. We find that efficient handling of NUS 4D data enables one to quickly establish distance restraints and calculate three-dimensional structures by acquiring 4D CC-NOESY experiments and other related multidimensional experiments, such as 4D CN-NOESY and 4D HCCCONH-TOCSY. However, existing implementations of L1 minimization can be computationally intensive,

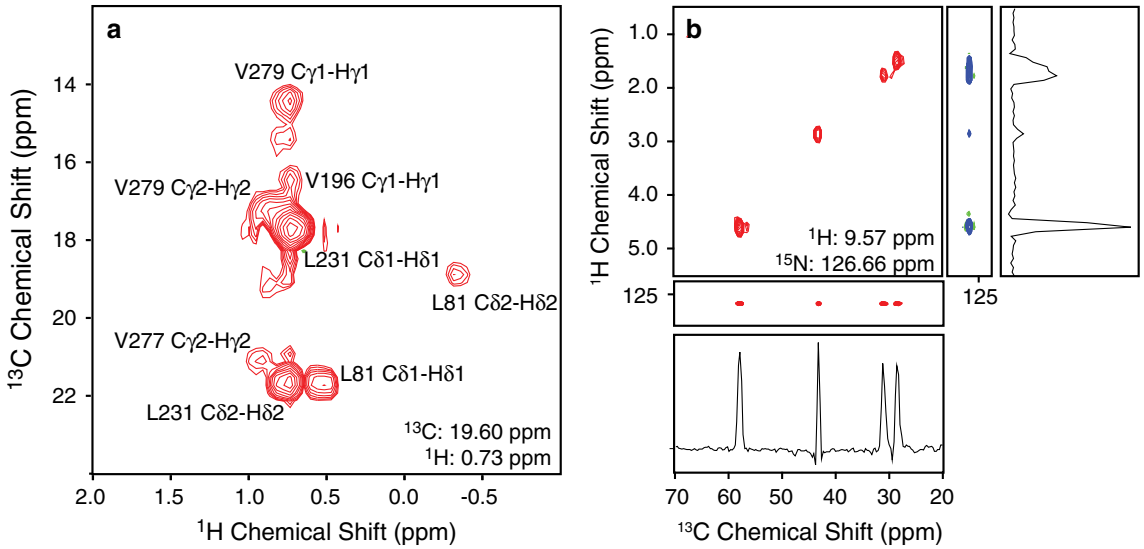


Fig. 8 NESTA-NMR is applicable to many NMR experiments as shown by **a** a plane corresponding to the chemical shift of the Leu231-δ1 methyl group from a 4D CC-NOESY experiment performed on the 32 kDa ILV-DCN-ZA domain of ASAP1 (400 μM), and **b** a plane corresponding to the side chain of a single amino acid residue in the 15 kDa C/N labeled PH domain of ASAP1 (330 μM). The *boxed contour* plot and trace on the right side are from a 3D H(CC)(CO)NH experiment, and the boxed contour plot and trace on the bottom are from a 3D (H)CC(CO)NH experiment. See "Materials and methods" and Supplemental Table S1 for further details

especially when a high degree of accuracy is needed for high dynamic range experiments. This problem is exacerbated for 4D data because reconstruction requires about one million FFT operations (assuming each indirect dimension has about 50 complex points) for each point in the direct dimension. Since the computational cost for a single iteration is very large, a high-efficiency algorithm must converge in the smallest number of iterations possible, thus minimizing the total number of FFT operations. After a systematic comparison of several state-of-the-art L1 algorithms previously described for processing NUS NMR data, we find that incorporation of NESTA, a first order gradient descent algorithm recently developed in the field of CS, enables L1 minimization to converge with the least number of iterations ($\leq$200) and at a low computational cost per iteration. In addition, consistency with original experimental data is implicitly realized by preserving the sampled data points during reconstruction. This also improves the robustness of the reconstruction by eliminating both the need to estimate noise and to choose parameter(s) that accurately gauge the level of data consistency. Generally, a 3D NUS cube can be reconstructed within several minutes, enabling one to process 4D NUS NMR data with a laptop or workstation computer within a few hours.

Our comparison of different reconstruction algorithms using high-dynamic range NOESY data demonstrated that Gaussian-SL0 preserves linearity better than L1 methods. However, as a non-convex function, Gaussian-SL0 is inherently a less stable function to minimize. Furthermore, Gaussian-SL0 cannot benefit from the acceleration schemes in the NESTA algorithm, which is designed for minimization of convex functions. For this reason, Gaussian-SL0 generally requires a greater number of steps for convergence, thus reducing the computing efficiency relative to that of the NESTA algorithm with L1 regularization. This hinders the application of Gaussian-SL0 in cases where fast processing speeds are needed. Therefore, we recommend using Gaussian-SL0 for smaller regions of interest in tandem with L1 regularization to process the whole dataset. Additionally, although IRL1 is also a non-convex function, it is implemented as an iterative minimization of a weighted L1 where the weights are updated after each NESTA run. IRL1 provides the increased fidelity of Gaussian-SL0, while only incurring moderate additional computational time (generally $\sim$3–5 times longer than L1). Consequently, NESTA L1 minimization provides the most efficient and general algorithm for NUS reconstruction, and either Gaussian-SL0 or IRL1 can be utilized as desired for a more detailed examination of spectral features.

The discussion in this report emphasizes the application of the regularization terms L1, IRL1, and Gaussian-SL0 and the NESTA algorithm in the context of processing NOESY NMR data that have a high dynamic range; however, we also demonstrate that L1 can form the basis of a generalized approach for a broad range of NMR experiments such as HCCCONH-TOCSY, CN-NOESY, and

backbone assignment experiments. We recommend using the NESTA algorithm and L1 regularization for general purpose NUS NMR data processing. In the case of NOESY NMR data and other situations where more accurate estimation of peak intensity is required, IRL1 or Gaussian-SL0 can be used. The software package NESTA-NMR implements L1, IRL1, and Gaussian-SL0 for reconstruction of 2D, 3D, and 4D NUS NMR data that can be subsequently processed with NMRPipe. Thus, NESTA-NMR provides a generalized, efficient solution that integrates with popular workflows for the application of NUS to biomolecular NMR studies.

Software availability

NESTA-NMR binary executables for Mac and Linux, an NMRPipe macro for Rance-Kay frequency discrimination, installation instructions, documentation, and sample data are available on-line at http://nestanmr.com.

Acknowledgments We thank Dr. Aleksandras Gutmanas (EBI, Hinxton, UK), Dr. Jinfa Ying (NIDDK, NIH, Bethesda, MD) and Mr. William Hanisch for useful discussions. This work was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research.

References

Aoto PC, Fenwick RB, Kroon GJA, Wright PE (2014) Accurate scoring of non-uniform sampling schemes for quantitative NMR. J Magn Reson 246:31–35. doi:10.1016/j.jmr.2014.06.020

Becker S, Bobin J, Candès EJ (2011) NESTA: a fast and accurate first-order method for sparse recovery. SIAM J Imaging Sci 4:1–39. doi:10.1137/090756855

Bostock MJ, Holland DJ, Nietlispach D (2012) Compressed sensing reconstruction of undersampled 3D NOESY spectra: application to large membrane proteins. J Biomol NMR 54:15–32. doi:10.1007/s10858-012-9643-4

Candès EJ, Tao T (2005) Decoding by linear programming. IEEE Trans Inf Theory 51:4203–4215

Candès EJ, Tao T (2006) Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Trans Inf Theory 52:5406–5425. doi:10.1109/Tit.2006.885507

Candès EJ, Romberg J, Tao T (2006a) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory 52:489–509. doi:10.1109/Tit.2005.862083

Candès EJ, Romberg JK, Tao T (2006b) Stable signal recovery from incomplete and inaccurate measurements. Commun Pure Appl Math 59:1207–1223. doi:10.1002/Cpa.20124

Candès EJ, Wakin MB, Boyd SP (2008) Enhancing sparsity by reweighted l(1) minimization. J Fourier Anal Appl 14:877–905. doi:10.1007/s00041-008-9045-x

Cavanagh J, Palmer AG, Wright PE, Rance M (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear relay spectroscopy. J Magn Reson (1969) 91:429–436. doi:10.1016/0022-2364(91)90209-C

Das R et al (2009) Allosteric activation of E2-RING finger-mediated ubiquitylation by a structurally defined specific E2-binding region of gp78. Mol Cell 34:674–685. doi:10.1016/j.molcel.2009.05.010

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Diercks T, Coles M, Kessler H (1999) An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. J Biomol NMR 15:177–180. doi:10.1023/A:1008367912535

Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52:1289–1306

Drori I (2007) Fast l(1) minimization by iterative thresholding for multidimensional NMR Spectroscopy. EURASIP J Adv Signal Process. doi:10.1155/2007/20248

Galassi M et al (2009) GNU scientific library reference manual, 3rd edn. Network Theory, Bristol

Goddard TD, Kneller DG Sparky 3. University of California, San Francisco. https://www.cgl.ucsf.edu/home/sparky/

Hoch JC, Stern AS (1996) NMR data processing. Wiley-Liss, New York

Hoch JC, Maciejewski MW, Mobli M, Schuyler AD, Stern AS (2014) Nonuniform sampling and maximum entropy reconstruction in multidimensional NMR. Acc Chem Res 47:708–717. doi:10.1021/ar400244v

Hyberts SG et al (2007) Ultrahigh-resolution (1)H-(13)C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. J Am Chem Soc 129:5108–5116. doi:10.1021/ja068541x

Hyberts SG, Frueh DP, Arthanari H, Wagner G (2009) FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation. J Biomol NMR 45:283–294. doi:10.1007/s10858-009-9368-1

Hyberts SG, Arthanari H, Wagner G (2012a) Applications of non-uniform sampling and processing. Top Curr Chem 316:125–148. doi:10.1007/128_2011_187

Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G (2012b) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling. J Biomol NMR 52:315–327. doi:10.1007/s10858-012-9611-z

Hyberts SG, Robson SA, Wagner G (2012c) Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. J Biomol NMR 55:167–178. doi:10.1007/s10858-012-9698-2

Kay L, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114:10663–10665. doi:10.1021/ja00052a088

Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. Angew Chem Int Ed Engl 50:5556–5559. doi:10.1002/anie.201100370

Kim S-J, Koh K, Lustig M, Boyd S, Gorinevsky D (2007) An interior-point method for large-scale L1-regularized least squares. IEEE J Sel Top Signal Process 1:606–617. doi:10.1109/jstsp.2007.910971

Liu S et al (2012) Promiscuous interactions of gp78 E3 ligase CUE domain with polyubiquitin chains. Structure 20:2138–2150. doi:10.1016/j.str.2012.09.020

Luo R, Miller Jenkins LM, Randazzo PA, Gruschus J (2008) Dynamic interaction between Arf GAP and PH domains of ASAP1 in the regulation of GAP activity. Cell Signal 20:1968–1977. doi:10.1016/j.cellsig.2008.07.007

Lustig M, Donoho D, Pauly JM (2007) Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn Reson Med 58:1182–1195. doi:10.1002/mrm.21391

Maciejewski MW, Mobli M, Schuyler AD, Stern AS, Hoch JC (2012) Data sampling in multidimensional NMR: fundamentals and strategies. Top Curr Chem 316:49–77. doi:10.1007/128_2011_185

Mayzel M, Kazimierczuk K, Orekhov VY (2014) Causality principle in reconstruction of sparse NMR spectra. Chem Commun 50:8947–8950. doi:10.1039/C4CC03047H

Mobli M, Stern AS, Bermel W, King GF, Hoch JC (2010) A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment. J Magn Reson 204:160–164. doi:10.1016/j.jmr.2010.02.012

Mohimani H, Babaie-Zadeh M, Jutten C (2009) A fast approach for overcomplete sparse decomposition based on smoothed l(0) norm. IEEE Trans Signal Process 57:289–301. doi:10.1109/Tsp.2008.2007606

Nesterov Y (2005) Smooth minimization of non-smooth functions. Math Program 103:127–152. doi:10.1007/s10107-004-0552-5

Orekhov VY, Jaravine VA (2011) Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. Prog Nucl Magn Reson Spectrosc 59:271–292. doi:10.1016/j.pnmrs.2011.02.002

Orekhov VY, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. J Biomol NMR 27:165–173

Palmer AG, Cavanagh J, Wright PE, Rance M (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. J Magn Reson (1969) 93:151–170. doi:10.1016/0022-2364(91)90036-S

Palmer AG, Cavanagh J, Byrd RA, Rance M (1992) Sensitivity improvement in three-dimensional heteronuclear correlation NMR spectroscopy. J Magn Reson (1969) 96:416–424. doi:10.1016/0022-2364(92)90097-Q

Paramasivam S et al (2012) Enhanced sensitivity by nonuniform sampling enables multidimensional MAS NMR spectroscopy of protein assemblies. J Phys Chem B 116:7416–7427. doi:10.1021/jp3032786

Rovnyak D, Sarcone M, Jiang Z (2011) Sensitivity enhancement for maximally resolved two-dimensional NMR by nonuniform sampling. Magn Reson Chem. doi:10.1002/mrc.2775

Sklenar V, Piotto M, Leppik R, Saudek V (1993) Gradient-tailored water suppression for 1H–15N HSQC experiments optimized to retain full sensitivity. J Magn Reson Ser A 102:241–245. doi:10.1006/jmra.1993.1098

Stern AS, Donoho DL, Hoch JC (2007) NMR data processing using iterative thresholding and minimum l(1)-norm reconstruction. J Magn Reson 188:295–300. doi:10.1016/j.jmr.2007.07.008

Szantay C (2008) NMR and the uncertainty principle: how to and how not to interpret homogeneous line broadening and pulse nonselectivity. III. Uncertainty? Concepts Magn Reson A 32A:302–325. doi:10.1002/cmr.a.20116

Trzasko J, Manduca A, Borisch E (2007) Sparse MRI reconstruction via multiscale L0-continuation. Paper presented at the Proceedings of the 2007 IEEE/SP 14th workshop on statistical signal processing

Tugarinov V, Kay LE (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc 125:13868–13878. doi:10.1021/ja030345s

Wright SJ, Nowak RD, Figueiredo MAT (2009) Sparse reconstruction by separable approximation. IEEE Trans Signal Process 57:2479–2493. doi:10.1109/tsp.2009.2016892

Ying J, Chill JH, Louis JM, Bax A (2007) Mixed-time parallel evolution in multiple quantum NMR experiments: sensitivity and resolution enhancement in heteronuclear NMR. J Biomol NMR 37:195–204. doi:10.1007/s10858-006-9120-z

Efficient and generalized processing of multidimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms

Supplemental Information

Shangjin Sun[1], Michelle Gill[1], Yifei Li, Mitchell Huang and R. Andrew Byrd[*]

Structural Biophysics Laboratory
National Cancer Institute, Frederick, MD 21702

[1] These authors contributed equally to this work.

Address correspondence to R. Andrew Byrd: byrdra@mail.nih.gov

Input data format for NESTA-NMR

To perform reconstruction, NESTA-NMR requires only the data and a sampling schedule. The file containing the data must be organized so all of the real and imaginary (complex) pairs are adjacent to each other. For a four-dimensional data set whose indirect dimensions are (t_1, t_2, t_3), this is: (R,R,R), (I,R,R), (R,I,R), (I,I,R), (R,R,I), (I,R,I), (R,I,I), and (I,I,I), where R and I denote data that is real and imaginary, respectively, in the appropriate indirect dimension. Because a complex pair must be collected for each indirect dimension associated with a NUS point, the data file will contain 2^N times as many FIDs as NUS points, where N is the number of indirect dimensions. As of TopSpin 3.0, this is the default storage format for Bruker NUS experiments. NESTA-NMR has been tested thoroughly on Bruker data. It is expected to be compatible with Agilent data, where limited testing was performed with development versions of NESTA-NMR, provided the complex pairs are organized according to VnmrJ's "phase2,phase" or "phase3,phase2,phase" acquisition flags for three- and four-dimensional experiments, respectively. The first step of NESTA-NMR processing (see below) involves conversion to NMRPipe format, and data can be organized in this format on Agilent spectrometers based on the definition of the "array" parameter.

The sampling schedule, referred to as a nuslist by Bruker, is a text file whose rows correspond to the consecutively acquired NUS points. Each row contains the respective indices of the indirect dimension(s) as zero-indexed, space-delimited integers. A four-dimensional NUS experiment will have three integers in each row, for example. The initial ten lines of a three- and four-dimensional sampling schedule are shown in Fig. S1A and B, respectively.

Overview of NESTA-NMR processing steps

NESTA-NMR has been designed with a focus on computational speed, ease of use, and integration with existing workflows involving NMRPipe (Delaglio et al. 1995). The processing of non-uniformly sampled (NUS) data can be completed in four steps, three of which are essentially identical to those used to process uniformly sampled multidimensional NMR data. A graphical representation of this process is shown in Fig. S2. Briefly, these steps are: (1) convert NUS data to NMRPipe format; (2) process the direct dimension of the data using NMRPipe and perform Rance-Kay processing on the indirect dimensions if required; (3) reconstruct missing data using NESTA-NMR; (4) process the indirect dimensions using NMRPipe.

NESTA-NMR can process two-, three-, and four-dimensional data. It attempts to correctly determine input parameters wherever possible, but allows these values to be overridden with command line flags, which can be listed from the help function of the program (Fig. S3**)**. A detailed description of each of these parameters and further information about the steps in the data processing workflow (Fig. S2) are provided in the manual included with NESTA-NMR, which can be downloaded at http://nestanmr.com.

Rance-Kay processing of NUS data

NESTA supports the three most common protocols for frequency discrimination in the indirect dimension: States, States-TPPI, and Rance-Kay. Of these three methods, Rance-Kay is the only one that requires specific treatment prior to processing the direct dimension.

Since it is most convenient to handle such processing without the introduction of separate programs, NESTA includes macros, called NESTA_bruk_ranceN.M and NESTA_var_ranceN.M, that perform Rance-Kay processing on Bruker and Agilent data, respectively, within NMRPipe. These macros can perform Rance-Kay processing for any indirect dimension (Y, Z, or A) and on any dimension experiment (2D, 3D, 4D) that was collected with the real and imaginary pairs adjacent to each other in the Y-dimension (see above). If more than one dimension requires Rance-Kay processing, the macro can be called additional times. Fig. S4 contains the code for NESTA_bruk_ranceN.M. The code for the Agilent version is similar. Fig. S5 contains an example NMRPipe processing script that utilizes the macro. The NUS dimension of the converted matrix (Y, Z, or A) is specified by setting the parameter "nShuf" to 1, 2, or 3, respectively.

Parallel processing in NESTA-NMR

NESTA-NMR utilizes parallel processing methods during data reconstruction (step 3 of Fig. S2). For this task, NESTA invokes the POSIX threads (pthreads) C library. This library utilizes lower level functions than other parallelization libraries, such as OpenMP, meaning that more lines of code are required to implement parallelization with pthreads. However, pthreads affords more control over thread creation and management than higher level libraries. Perhaps most importantly, pthreads is part of the standard C library on POSIX-compliant operating systems (Linux, BSD, Mac OS X, etc.) and, as such, does not require the installation of additional libraries. The use of a native C library for parallel processing is desirable over implementations involving external scripts or shell functions due to ease of use and a potential increase in speed and reduction in memory.

Brief description of the NESTA algorithm

NESTA-NMR seeks to reconstruct non-uniformly sampled (NUS) NMR data by minimizing the $\ell 1$-norm (L1) of the sampled data in the frequency domain relative to the analogous points in the reconstructed data while simultaneously preserving the experimental data. The NESTA algorithm has been described extensively by Becker *et al.* (Becker et al. 2011). Our implementation is discussed briefly, with minor differences from the more general method of Becker and coworkers noted.

The goal of our implementation of NESTA can be described as follows:

$$\text{Minimize } \|F(x)\|_{l1}, \text{ subject to } Rx = b \qquad (1)$$

where $F(x)$ is implemented in NESTA-NMR as the Fourier transform of x. The desired solution is x, R is a sparse matrix of the sampled data, and b is a vector that contains experimental data at the appropriate locations and is zero elsewhere. For simplicity, the target function, $\|F(x)\|_{l1}$, is referred to as $f(x)$ henceforth in the text. The solution criteria consider only sampled data, thus allowing the remaining (reconstructed) data to adopt values that minimize the target function.

The version of NESTA described by Becker and coworkers includes an additional term, ϵ, in the above equation, which is an upper bound on the noise. Our implementation of NESTA sets this value to zero because allowing ϵ to be adjustable did not significantly improved the quality of the reconstruction in our tests and it has the additional benefit of eliminating the need to estimate noise.

To reduce computational complexity, NESTA computes a smoothed version of the target function, called $f_u(x)$, which is then used to calculate the gradient, $\Delta f(x)$. When the smoothing parameter, μ, is small, the minima—and, consequently, also the gradients—of the original and smoothed functions are very similar.

The gradient, $\Delta f(x)$, is computed as follows:

1. The current result of x at iteration k, called x_k, is Fourier transformed to produce h_k, which is then used to calculate L1. The value of $f(x)$ is used to monitor the decrease of L1 and determine if the current result meets the stop criterion. Becker, *et al.*, used $f_u(x)$ to monitor optimization progress rather than $f(x)$. For our implementation, we found using $f(x)$ improves computing efficiency without affecting the final results.

2. Then, h_k is divided by the larger of either its absolute value, $|h_k|$, or the smoothing parameter, μ. This is essential to ensure the step size of x_{k+1} is small after gradient calculation (see below). The value of μ, which corresponds to the degree of gradient smoothing, starts at 0.9 times the L1-norm of the zero-filled FID prior to optimization and is scaled down by 0.9 (0.81, 0.73, etc.) for successive optimization steps to a final value of 0.002. This scaling balances the rate of convergence, which is faster when μ is large and more important in early stages of optimization, with accuracy of the final solution, which occurs when μ is small and is more important in later optimization steps.

3. Finally, h_k is inverse Fourier transformed to return the gradient, $\Delta f(x)$.

After computing $\Delta f(x)$, two auxiliary vectors y_k and z_k are calculated. As described by Becker *et al.*, these vectors store the search steps and the results of previous iterations, respectively. The value for x_{k+1} is then calculated from a linear combination of y_k and z_k. Calculating x_{k+1} from y_k and z_k rather than directly from x_k significantly increases the convergence rate of the NESTA algorithm.

Convergence is achieved when the relative difference between the L1 for a given iteration compared to the mean L1 for ten prior iterations is smaller than the stop criterion, which is calculated as the inverse of L1 for the zero-filled FID prior to reconstruction. For some 4D spectra, this stop criterion proved to be too stringent in cases where the number of peaks is very large (e.g. NOESYs of large proteins), so the stop criterion is also scaled by the number of reconstructed (complex) points for 4Ds.

Computational efficiency of NESTA-NMR

This implementation of the NESTA algorithm is designed for efficiency with regards to both computational requirements and memory usage. Only two multidimensional Fourier transforms are required per iteration, and NESTA-NMR utilizes the highly optimized GNU Scientific Library (Galassi et al. 2009) for these steps. Point-wise calculations of absolute value are used for monitoring the value of the target function, *f(x)*, and for comparison to the smoothing parameter, μ. The computational cost per iteration is roughly equivalent to that of iterative soft thresholding (IST). However, as described in the main text, NESTA-NMR reaches a global minimum in fewer iterations, and, consequently, in less computational time, than IST. NESTA-NMR operates on linearized versions of matrices (vectors), which greatly reduces memory usage relative to that of true matrix operations. Lastly, each NESTA-NMR process reads only the data necessary for its respective calculation.

As is the case with signal processing in general, the computational time required by NESTA-NMR increases with data dimensionality and size, with the most important of these two being dimensionality. Performing a reconstruction on four-dimensional data requires optimization of multiple three-dimensional slices. This is more time consuming than the analogous calculation for three- or two-dimensional data, which involves optimization of only two- or one-dimensional slices, respectively. NESTA-NMR performs a separate optimization process for each of the directly detected points, so it is advantageous to limit the size of this dimension by extracting only the region necessary during processing of the direct dimension with NMRPipe (step 2 of Fig. S2).

Effects of sparsity and sampling schedule on NESTA-NMR reconstructions

To demonstrate that reconstruction by NESTA-NMR is not dependent on sampling schedule over a range of sparsity values, a uniformly sampled spectrum was downsampled at four different sparsity levels using multiple, randomly generated sampling schedules for each sparsity level. The spectrum was a 3D HNCO experiment collected on a uniformly-^{15}N, ^{13}C -labeled sample of 700 μM human ubiquitin using a Bruker Avance III 600 MHz instrument with cryoprobe at 298 K. The data set contained 6528 indirect points (51 ^{15}N × 128 ^{13}C) with spectral widths of 2189.2 and 2112.7 Hz and maximum evolution times of 23.3 ms and 60.6 ms for ^{15}N and ^{13}C, respectively. Each FID contained 8 scans.

One hundred randomly generated sampling schedules were created using Schedule Tool (Hoch and Stern 1996; Maciejewski et al.) for each of four different sparsity levels: 3000, 2000, 1000, and 400 points, corresponding to a 46.0%, 30.6%, 15.3%, and 6.1% sampling density, respectively. The uniformly sampled spectrum was downsampled to correspond to each of the 400 (total) sampling schedules, and then reconstruction was performed using NESTA-NMR with 30 iterations of regularization. The error from the final L1 of each reconstruction was used as a means to evaluate reconstruction quality. For each sparsity level, a representative experiment whose L1 error was closest to the mean of the group was processed using standard FFT methods with NMRPipe.

A histogram of the L1 errors for each sparsity level is shown in Fig S6, panels A, D, G, and J. As expected, the mean value of the L1 error (vertical red line) increases with sparsity. For the 30.6% sampling density (2000 points), the distribution is both very symmetric and Gaussian, as

demonstrated by the skew and kurtosis values, respectively. The distributions at the other three sampling densities are less regular, but that of the sparsest group (6.1%, 400 points) does not appear to be significantly more irregular (based on kurtosis and skew) than the the other sampling densities. A single ^{1}H-^{13}C plane taken at 136.0 ppm (^{15}N) from a fully processed reconstruction is shown with identical contouring for each sparsity level in Fig S6, panels B, E, H, and K. As has been observed previously (Orekhov et al. 2003), the presence of artifacts and/or noise gradually increases with sparsity. However, even the most sparsely sampled spectrum (400 pts, 6.1%) is quite usable when contouring is scaled based on relative sparsity, as shown in Fig S6, panels C, F, I, and L.

We believe the practical applications of NUS involve the ability to collect data of a quality comparable to or better than that of a uniformly sampled counterpart in a reasonable amount of time and, thus, do not advocate aggressive undersampling as a common laboratory practice. Nevertheless, these results demonstrate that NESTA-NMR can faithfully reconstruct data collected with a practical range of sampling densities and that the quality of the reconstruction at a particular sparsity level is largely independent of sampling schedule.

Convergence of NESTA-L1 regularization

As has been noted previously (Stern et al. 2007), change in stepsize is not a reliable test of convergence for IST and associated methods, such as NESTA-NMR. To further test the convergence of NESTA-NMR, the reconstruction of a 1D data set produced by NESTA-NMR was then subject to further regularization with IST-S (Stern et al. 2007). These results were then compared to those produced by IST-S alone. The L1 was used to measure progress at each iteration. As noted (*vide supra*), NESTA-NMR and IST-S have the same functional, and thus (should) converge to the same, un-normalized value for a given data set. As before (see Fig 1), the NESTA-NMR regularization (Fig S7, black) converges in fewer iterations than IST-S (Figure S7, red) and to the same final L1. The L1-norms were not normalized to reinforce that the respective values are the exact same for NESTA-NMR and IST-S. Furthermore, when data reconstructed with NESTA-NMR were then subject to an additional 2000 iterations of IST-S regularization (Fig S7, green), L1 remained unchanged. This result is consistent with convergence of NESTA-NMR to a minimum similar to that of IST-S.

References

Becker S, Bobin J, Candès EJ (2011) NESTA: A fast and accurate first-order method for sparse recovery Siam Journal on Imaging Sciences 4:1-39 doi:10.1137/090756855

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes J Biomol NMR 6:277-293

Galassi M et al. (2009) GNU Scientific Library Reference Manual. 3rd ed. Network Theory,

Hoch JC, Stern AS (1996) NMR data processing. Wiley-Liss, New York

Maciejewski MW, Gorbatyuk V, Hoch JC ScheduleTool. University of Connecticut Health Center. http://sbtools.uchc.edu/nmr/sample_scheduler.

Orekhov VY, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. J Biomol NMR 27:165-173

Stern AS, Donoho DL, Hoch JC (2007) NMR data processing using iterative thresholding and minimum l(1)-norm reconstruction J Magn Reson 188:295-300 doi:10.1016/j.jmr.2007.07.008

Supplemental Table 1

Table S1. NESTA-NMR has been utilized for reconstruction on a vast range of NUS multidimensional NMR experiments and samples.

Experiment	Dimensions	% Sparsity	Acquisition Time		MW (kDa)	Isotope Labeling	Field Strength (MHz)
HNCA	3	26.0%	26	hrs	15	C/N	600
CBCA(CO)NH	3	29.6%	45.5	hrs	15	C/N	600
HNCO	3	25.0%	12.5	hrs	15	C/N	600
HN(CA)CB	3	25.0%	50.5	hrs	15	C/N	600
HCC(CO)NH_H	3	33.3%	43	hrs	15	C/N	600
HCC(CO)NH_C	3	32.5%	90	hrs	15	C/N	600
HCC(CO)NH	4	1.0%	57	hrs	15	C/N	600
N-edited NOESY	3	25.0%	70.5	hrs	15	15N	600
C-edited NOESY	3	23.1%	111	hrs	15	C/N	700
CN-NOESY	4	2.2%	108	hrs	15	C/N	600
CC(methyl)-NOESY	4	9.8%	88	hrs	8	C/N/D	900
CC(methyl)-NOESY	4	8.1%	6	days	32	ILV-D/C/N	600
HNCACB	3	26.0%	71.3	hrs	32	ILV-D/C/N	800
HN(CO)CACB	3	26.0%	70.4	hrs	32	ILV-D/C/N	800
(H)CC(CO)NH	3	24.0%	70.9	hrs	32	ILV-D/C/N	600
HCC(CO)NH	3	23.0%	81.8	hrs	32	ILV-D/C/N	600
HNCA	3	26.0%	43.6	hrs	32	C/N	600
HN(CO)CA	3	26.0%	47.7	hrs	32	C/N	600
(H)CCH-TOCSY	3	20.0%	92.7	hrs	32	C/N	700

Figure Captions

Fig. S1 The first ten lines of a sampling schedule for a (A) three- and (B) four-dimensional NUS spectrum

Fig. S2 Flowchart for processing NUS data with NESTA-NMR. The left column represents the data files present between each of the four steps. After the first step, multiple data files, indicated by gray rectangles, are present for the example data set. Each of the processing steps is listed in the right column

Fig. S3 The help function lists all of the flags accepted by NESTA-NMR and their default values

Fig. S4 NMRPipe macro NESTA_bruk_ranceN.M for performing Echo-AntiEcho processing on Bruker NUS data

Fig. S5 An example NMRPipe script that utilizes NESTA_bruk_ranceN.M for Rance-Kay processing in the second (nShuf = 2) NUS dimension. This is equivalent to performing Rance-Kay processing on the Z-dimension of the reconstructed data

Fig. S6 The reconstruction of HNCO data by NESTA-NMR using data derived from 100 randomly generated sampling schedules at four different sampling densities 46.0, 30.6, 15.3, and 6.1 %, which correspond to 3000, 2000, 1000, and 400 points, respectively. For each sampling density, the L1-norm error of the samples is plotted (panels A, D, G, and J) with the mean denoted (red line). A two-dimensional ^{1}H-^{13}C plane at 136.0 ppm (^{15}N) of a representative experiment for each group is shown with uniform (panels B, E, H, and K) and scaled (panels C, F, I, and L) contouring.

Fig. S7 The convergence of NESTA-NMR relative to IST-S was compared for a 1D data set. Both NESTA-NMR (black) and IST-S (red) reach the same final L1-norm value, but NESTA-NMR does so in fewer iterations. Furthermore, IST-S is not able to further improve the L1 value of data reconstructed by NESTA-NMR for 2000 iterations (green). To reduce overplotting, every fifth iteration of the NESTA-NMR regularization is shown. For IST-S, every second and 100^{th} iteration is shown for the regularization alone (red) and following NESTA-NMR (green), respectively.

Figure S1

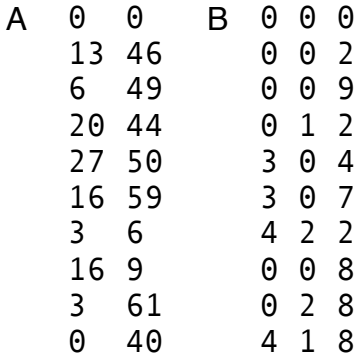

A 0 0
13 46
6 49
20 44
27 50
16 59
3 6
16 9
3 61
0 40
B 0 0 0
0 0 2
0 0 9
0 1 2
3 0 4
3 0 7
4 2 2
0 0 8
0 2 8
4 1 8

Figure S2

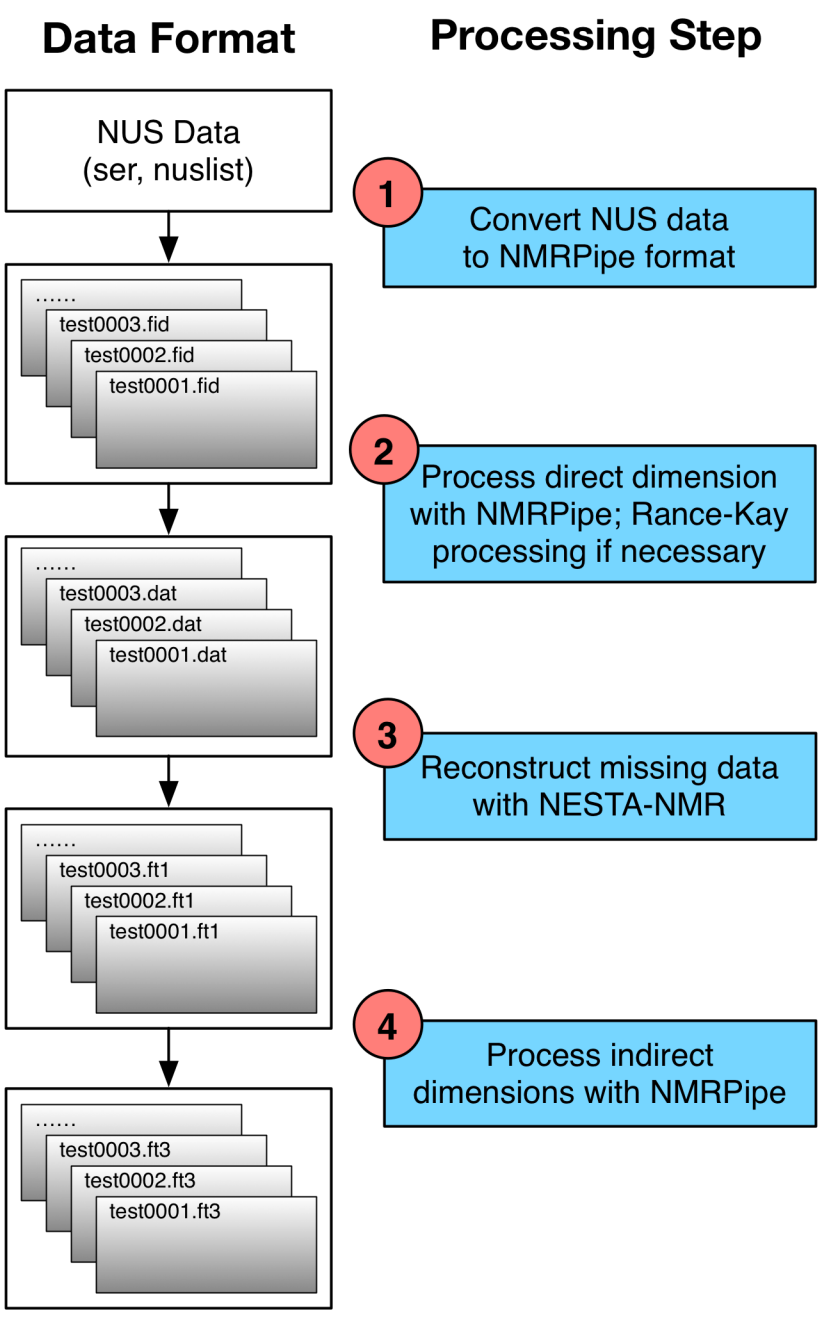

Data Format
Processing Step
NUS Data
(ser, nuslist)
1
Convert NUS data
to NMRPipe format
.....
test0003.fid
test0002.fid
test0001.fid
2
Process direct dimension
with NMRPipe; Rance-Kay
processing if necessary
.....
test0003.dat
test0002.dat
test0001.dat
3
Reconstruct missing data
with NESTA-NMR
.....
test0003.ft1
test0002.ft1
test0001.ft1
4
Process indirect
dimensions with NMRPipe
.....
test0003.ft3
test0002.ft3
test0001.ft3

Figure S3

```
*******************************************************************************
*                                NESTA-NMR v1.0                               *
*     Shangjin Sun, Michelle Gill, Yifei Li, Mitchell Huang, and R. Andrew Byrd    *
*                         Structural Biophysics Laboratory                    *
*                            National Cancer Institute                        *
*                               Frederick, MD 21702                           *
*                                                                             *
*              Use of NESTA-NMR implies acceptance of the user license        *
*******************************************************************************

    Flag           Argument        Information
=============   ==============   ================================================

-f, --fids         TEMPLATE        Template for input data from NMRPipe
                                   DEFAULT: ./ft/test%04d.dat

-n, --nuslist      FILE            Path to the nuslist
                                   DEFAULT: ./nuslist

-d, --outdir       DIRECTORY       Output directory
                                   DEFAULT: ./nesta

-o, --outname      FILE            Basename for output files
                                   DEFAULT: test

-t, --threads      INTEGER         Number of regularization threads to use
                                   DEFAULT: Determine # CPUs otherwise 1

-m, --method       INTEGER         Regularization method: L1/IRL1 or Gaussian-SL0
                                   Set to 1 for L1/IRL1 or 0 for Gaussian-SL0
                                   DEFAULT: 1

-i, --iter         INTEGER         Maximum number of iterations for regularization
                                   For L1/IRL1 regularization, suggested range is 10-50
                                   For Gaussian-SL0, suggested range is >= 5000
                                   DEFAULT: 30 for L1/IRL1, 5000 for Gaussian-SL0

-r, --rwiter       INTEGER         Number of re-weighted iterations for IRL1 regularization
                                   Improved results, but longer computation time
                                   Suggested range is 2-10, no re-weighting done when <= 1
                                   DEFAULT: 1

-s, --scaling      FLOAT           Scaling value for Gaussian-SL0 optimization.
                                   Suggested range is 0.98-0.99.
                                   Iterations should increase if scaling is set to 0.99.
                                   DEFAULT: 0.98

-c, --cutoff       FLOAT           Cutoff value for Gaussian-SL0 optimization
                                   DEFAULT: 0.1

-a, --alt          CHAR(S)         Set TPPI sign alternation in appropriate dimensions
                                   Dimensions are a case insensitive string of characters:
                                   e.g. yza for all three indirect dimensions.
                                   DEFAULT: no sign alternation (for States or Rance-Kay)

-y, --sizeY        INTEGER         Increase number of points in Y-dimension
                                   DEFAULT: max value from nuslist

-z, --sizeZ        INTEGER         Increase number of points in Z-dimension
                                   DEFAULT: max value from nuslist

-l, --sizeA        INTEGER         Increase number of points in A-dimension
                                   DEFAULT: max value from nuslist

-b, --keepnesta                    Flag to preserve intermediate NESTA-NMR files
                                   DEFAULT: delete intermediate files

-v, --version                      Show version number and quit

-h, --help                         Show this help information and quit
```

Figure S4

```
/***/
/*  NESTA_bruk_ranceN.M: Echo-AntiEcho-mode gradient shuffling for NUS data
/*  converted as a pseudo 2D (real/imaginary and NUS points in Y-dimension)
/*  or pseudo 3D (real/imaginary in Y-dimension, NUS points in Z-dimension)
/*  This macro is for Bruker data.
/*
/*  by Michelle L. Gill, National Cancer Institute
/*
/*  USAGE:
/*  If Echo-AntiEcho is inner loop:
/*     nmrPipe -fn MAC -macro NESTA_bruk_ranceN.M -noRd -noWr -var nShuf 1
/*
/*  If Echo-AntiEcho is second loop (3D or 4D):
/*     nmrPipe -fn MAC -macro NESTA_bruk_ranceN.M -noRd -noWr -var nShuf 2
/*
/*  If Echo-AntiEcho is third loop (4D only):
/*     nmrPipe -fn MAC -macro NESTA_bruk_ranceN.M -noRd -noWr -var nShuf 3
/***/

shufLen = 2^(nShuf);

if ( sliceCode % shufLen )  {
    exit( 0 );
};

xSizeI = integer(xSize + 0.1);
dxSizeI = integer(xSize * 2.0 + 0.1);
dxySizeI = integer(xSize * shufLen * 2.0 + 0.1);

float inData[dxySizeI];
float tmpDataR[dxSizeI];
float tmpDataI[dxSizeI];
float tmpData[dxSizeI];
float half1[xSizeI];
float half2[xSizeI];

imagPos = integer(2^(nShuf - 1) * xSize * 2.0 + 0.1);
nLoops = integer(2^(nShuf - 1) + 0.1);

(void) dReadB(inUnit, inData, wordLen * dxySizeI);

for ( i = 0; i < nLoops; i++ ) {
    /* Starting position for the real data */
    xStartR = i * dxSizeI;
    /* Starting position for the imaginary data */
    xStartI = xStartR + imagPos;

    /* Extract the real and imaginary portions of the data */
    vvCopyOff(tmpDataR, inData, dxSizeI, 0, xStartR);
    vvCopyOff(tmpDataI, inData, dxSizeI, 0, xStartI);

    /* Preserve a copy of the real data for the addition step */
    vvCopy(tmpData, tmpDataR, dxSizeI);


    /* Bruker */
    /* Subtract the imaginary data from the real for the new real data */
    vvSub(tmpDataR, tmpDataI, dxSizeI);
    /* Then add the preserved copy of the real data to the imaginary data */
    vvAdd(tmpDataI, tmpData, dxSizeI);


    /* Invert the last half of the new imaginary data and swap halves */
    vvCopyOff(half1, tmpDataI, xSizeI, 0, 0);
    vvCopyOff(half2, tmpDataI, xSizeI, 0, xSizeI);
    vNeg(half2, xSizeI);
    vvCopyOff(tmpDataI, half2, xSizeI, 0, 0);
    vvCopyOff(tmpDataI, half1, xSizeI, xSizeI, 0);

    /* Copy the data back to the main array */
    vvCopyOff(inData, tmpDataR, dxSizeI, xStartR, 0);
    vvCopyOff(inData, tmpDataI, dxSizeI, xStartI, 0);
}

(void) dWrite(outUnit, inData, wordLen * dxySizeI);
```

Figure S5

```
#!/bin/csh

setenv NESTAHOME /usr/local/bin/NESTANMR/NMRPipe

xyz2pipe -in fid/test%06d.fid –verb
| nmrPipe -fn MAC -macro $NESTAHOME/NESTA_bruk_ranceN.M -noRd -noWr -var nShuf 2 \
| nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5                              \
| nmrPipe -fn ZF –auto                                                         \
| nmrPipe -fn FT -auto                                                         \
| nmrPipe -fn PS -p0 58.4 -p1 0.0 -di                                          \
| nmrPipe -fn EXT -x1 2.0ppm -xn -1.0ppm -sw                                   \
| pipe2xyz -out ./ft/test%06d.dat -x
```

Figure S6

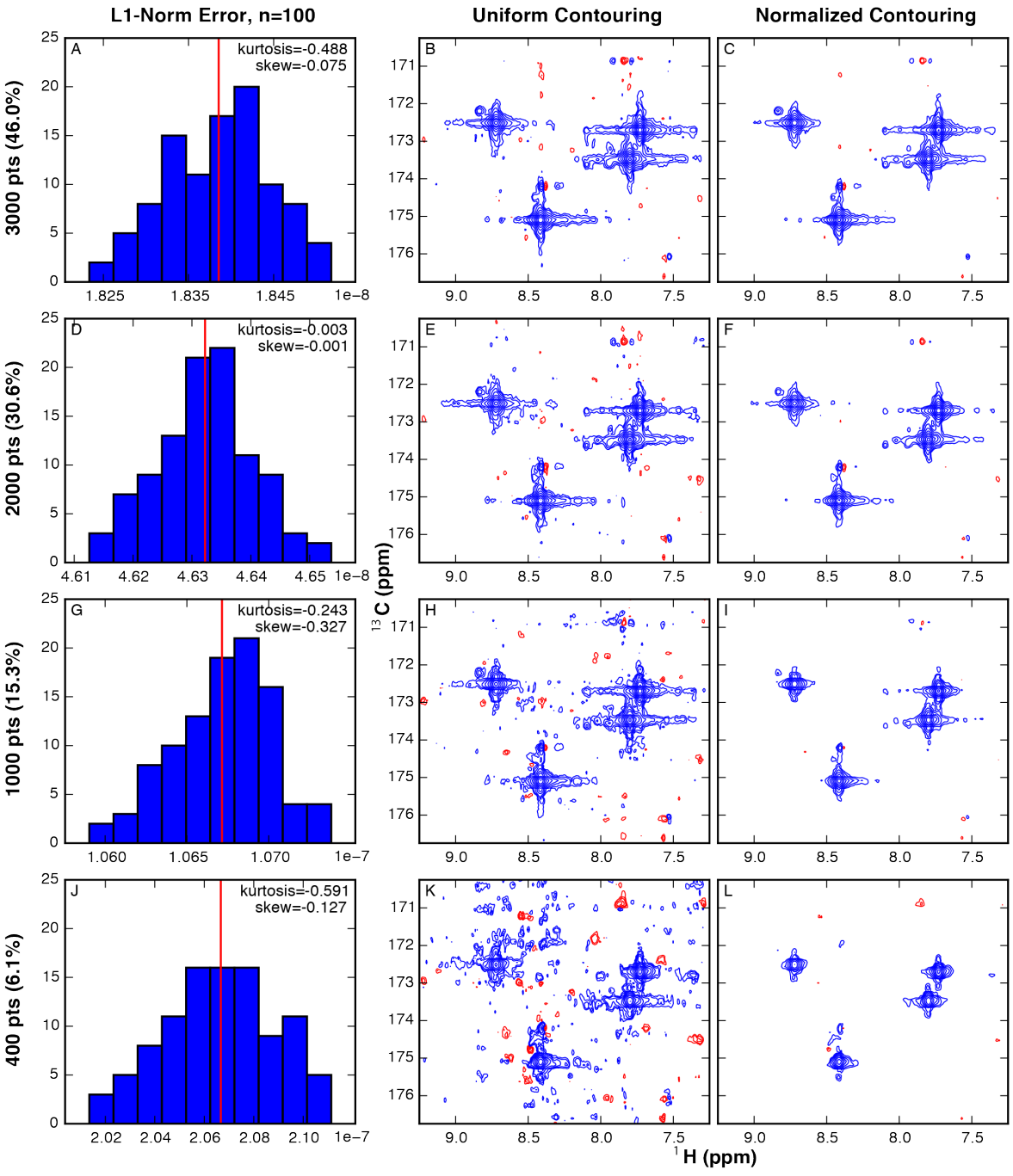
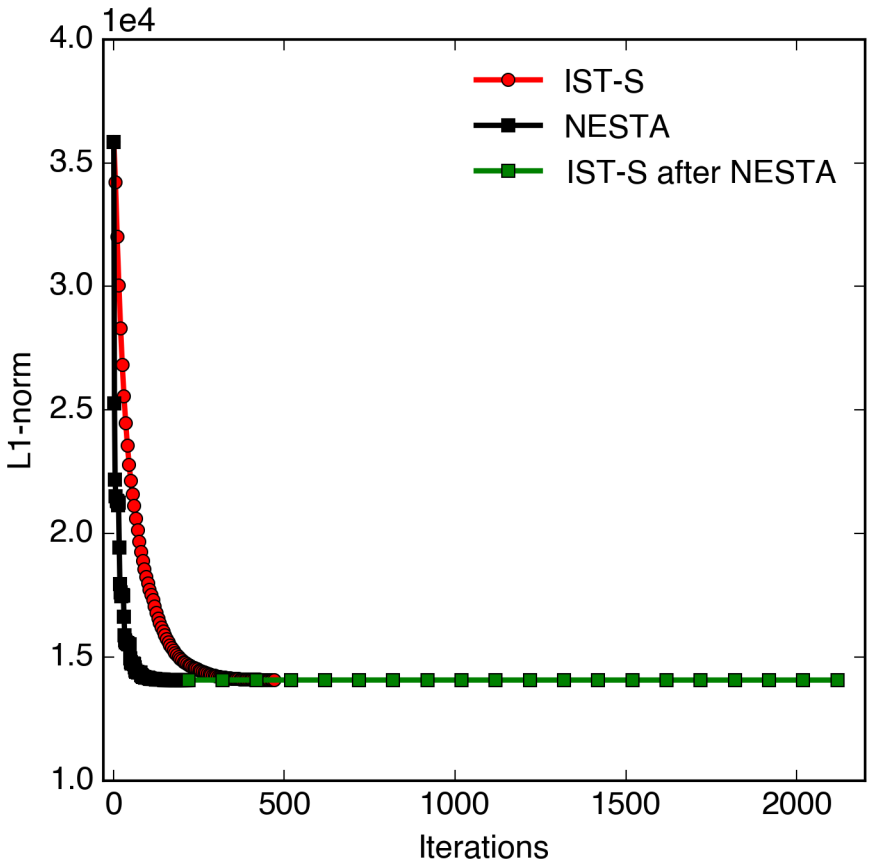

L1-Norm Error, n=100
Uniform Contouring
Normalized Contouring
3000 pts (46.0%)
2000 pts (30.6%)
1000 pts (15.3%)
400 pts (6.1%)
A
kurtosis=-0.488
skew=-0.075
B
C
D
kurtosis=-0.003
skew=-0.001
E
F
G
kurtosis=-0.243
skew=-0.327
H
I
J
kurtosis=-0.591
skew=-0.127
K
L
1e−8
1e−7
^{13}C (ppm)
^{1}H (ppm)

Figure S7


1e4
4.0
3.5
3.0
2.5
2.0
1.5
1.0
L1-norm
0
500
1000
1500
2000
Iterations
IST-S
NESTA
IST-S after NESTA