Journal of Magnetic Resonance 197 (2009) 63–70

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Optimized sampling patterns for multidimensional T_2 experiments

Christopher Kumar Anand [a,*], Alex D. Bain [b], Anuroop Sharma [b]

[a] Department of Computing and Software, McMaster University, 1280 Main Street West, ITB-202, Hamilton, Ontario, Canada L8S 4K1
[b] Department of Chemistry McMaster University, Hamilton, Canada

ARTICLE INFO

Article history:
Received 4 September 2008
Revised 30 November 2008
Available online 13 December 2008

Keywords:
Optimized sampling
Nonlinear sampling
Sparse k-space sampling
Nonuniform Fourier transform
SDP
Semidefinite optimization

ABSTRACT

Non-uniform sampling in multidimensional NMR shows great promise to significantly decrease experimental acquisition times, especially for relaxation experiments for which peak locations are already known. In this paper we present a method for optimizing the non-uniform sampling points such that the noise amplification and numerical instabilities are minimized. In particular, the minimum singular value of the Moore-Penrose pseudo-inverse is maximized using sequential semi-definite programming, thereby minimizing the worst-case errors. We test this method numerically on a set of assignment data from the proteins ubiquitin (in both folded and unfolded states) and RIα (119–244), a cAMP-binding regulatory subunit of protein kinase A (PKA). This test indicates that optimizing more than doubles the efficiency over random selection of points, and the efficiency increases as we go to higher dimensions.

1. Introduction

With modern NMR methods, we are able to assign all the carbon, nitrogen and hydrogen resonances in large proteins—thousands of nuclei in all. The main tool in this work is multidimensional NMR, in which overlapping peaks are resolved by going to higher dimensions [1–3]. For instance, amide protons are sorted according to the chemical shift of the amide nitrogen, then further sorted by shifts of the α-carbon, the carbonyl carbon, etc. The aim of the multidimensional experiments is to resolve all the signals, so there is a unique peak in the spectrum for each group of nuclei. Once the structure is established, dynamics can be measured by embedding a relaxation experiment into the multidimensional pulse sequence and following the intensity of the peak as a function of delay times.

If all signals are clearly resolved, then a full assignment can be made in a straightforward fashion. Overlaps and ambiguities may have to be resolved by going to higher dimensions. However, if the spectra are acquired and reconstructed using the standard Fourier methods, then the time required for the experiment increases exponentially with the number of dimensions. At the moment, three and four dimensions represent a practical limit, with experiments that can last several days.

There are two different problems here. One is the resolution and assignment of all of the unknown peaks in a new sample. The other is the study of dynamics in a protein for which the assignments are known. The same Fourier transform methods can be used for both applications, but there are more efficient ways of obtaining the information.

For the first problem, the assignment problem, there has been considerable recent interest in reduced dimensionality methods, which sample the signal in the time domain differently [4–10,29]. Depending on the complexity of the spectrum, this may lead to dramatically shorter acquisition times. For systems with closely-spaced peaks, many samples will be needed to resolve them, and the number approaches that of standard Fourier methods. There is no general approach that will work successfully in all cases, since we do not know *a priori* how crowded the spectrum will be.

We offer a solution to the second problem, in which the frequencies of the peaks are known, and we wish to monitor their intensities as a function of some delay time. Dynamics is typically probed in this way by measurements of T_1 and T_2 [2] with the inversion-recovery and CPMG experiments, respectively. If we use the same methods as used for the structural assignment, this multiplies the spectrometer time needed by the number of delay times used. If we only want to measure their intensities when we already know the positions of the peaks, we can exploit this information. Non-uniform sampling in the time domain can save significant amounts of time.

In a relaxation experiment, there are two types of sampling times: one associated with the relaxation experiment, and one with the resolution of the different frequencies. Each of these can be optimized:

(1) When an experiment (a single FID or multidimensional experiment) is repeated with different delays inserted for the purposes of measuring decay. The number and timing of such delays should be optimized.

* Corresponding author.
E-mail address: anandc@mcmaster.ca (C.K. Anand).

1090-7807/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2008.12.005

(2) Within a single multidimensional experiment, delay times are inserted for the purpose of creating phase variation in indirect dimensions. The number of FIDs and the delays for each of those FIDs should be optimized.

The first problem was studied in [11–13]. The second problem has been addressed without using numerical optimization of sampling times by all of the reduced dimensionality methods mentioned above. In particular, [14] introduces the Cramér-Rao lower bound as a measure of experiment design and uses it to compare linear and nonlinear sampling for small numbers of peaks. Similarly, 2D NOESY spectra were used to extend the results of a 3D NOESY–NOESY experiment [15] by the appropriate use of 2D mixing times.

In this paper we propose to use numerical optimization for this second problem, to give us a "best" choice of sampling points. The novelty in this paper is the evaluation of a trust-region, sequential semi-definite optimization approach on assignment data taken from real proteins, and numerical simulations designed to quantify the sensitivity of the designs to expected machine variations and measurement errors.

In the future, it may be possible to combine these two design problems and design a single experiment with delays used for both purposes, but we have not attempted it in this paper.

2. Theory

There are two fundamentally different optimization problems involved in designing protein NMR experiments. In the problem of determining peak frequencies, which we do not address in this paper, the peak frequencies are not known. In this case, full spectra are reconstructed. Faster experiments undersample the time domain causing artifacts in the spectra. The *point spread function* (psf) is the standard measure of the undersampling artifact. Informally, it is the spectrum one would observe if the true spectrum were a single discrete point, or a delta function for continuous transforms. The psf isolates the effects of poor sampling from all other effects. For example, in projection reconstruction, [16], undersampling causes star-like artifacts. This is assuming a simple reconstruction, but we also seek to reduce the artifacts, because they can obscure neighbouring peaks. See [7] for an example of an optimized undersampling pattern.

The second problem is the determination of relaxation rates. In proteins, differing relaxation rates along the backbone can be used to infer function, and even change as the protein undergoes conformal changes, so determining relaxation rates as quickly as possible is desired. In this case, however, the peak frequencies are known, so the appearance of the spectra is much less important. The only new information at this point is the peak heights or volumes, and specifically how they change as a function of relaxation delay time.

In a conventional quadrature experiment, the peak amplitudes (areas) are each different linear combinations of the measured signal. For example, if there is only one peak in the spectrum, then its amplitude is proportional to the strength of the first data point. The relative signal to noise ratio varies with e^{-t/T_2^*}, where t is the time of the sample, and T_2^* is the observed decay rate.

If there are two peaks, a and b, in a one dimensional experiment, with Larmor frequencies ω_a and ω_b and relaxation times $T_{2,a}^*$ and $T_{2,b}^*$, then the signals are mixed in the time domain. Given samples at times t_1 and t_2, the samples are

$$\begin{aligned} h(t_1) &= f(\omega_a)e^{i\omega_a t_1}e^{-t_1/T_{2,a}^*} + f(\omega_b)e^{i\omega_b t_1}e^{-t_1/T_{2,b}^*} \\ h(t_2) &= f(\omega_a)e^{i\omega_a t_2}e^{-t_2/T_{2,a}^*} + f(\omega_b)e^{i\omega_b t_2}e^{-t_2/T_{2,b}^*}. \end{aligned} \tag{1}$$

We can rewrite this to emphasize the linearity as

$$\begin{pmatrix} h(t_1) \\ h(t_2) \end{pmatrix} = \begin{pmatrix} e^{i\omega_a t_1}e^{-t_1/T_{2,a}^*} & e^{i\omega_b t_1}e^{-t_1/T_{2,b}^*} \\ e^{i\omega_a t_2}e^{-t_2/T_{2,a}^*} & e^{i\omega_b t_2}e^{-t_2/T_{2,b}^*} \end{pmatrix} \begin{pmatrix} f(\omega_a) \\ f(\omega_b) \end{pmatrix}, \tag{2}$$

and for future reference, we define S to be the signal generation matrix:

$$= S \begin{pmatrix} f(\omega_a) \\ f(\omega_b) \end{pmatrix}.$$

A priori, S depends on the values of T_2, but since T_2s vary in a limited range, it is reasonable to design an experiment with approximate T_2 values. Numerical tests presented in this paper show that the optimal designs are insensitive to this parameter.

For a real system, the matrix S may not be invertible, so it may not be possible to determine the peak amplitudes from two data points. Even if it is possible, the signal to noise ratio depends on the singular values of this transformation. Recall that the singular values of a matrix S are defined to be the eigenvalues of S^*S where * means conjugate transpose. This is defined whether S is square or rectangular, and, in fact, we expect to have more samples than peaks, so the signal generation matrix in 2 will in general have more rows than columns.

The worst case in experiment design is easy to understand. With a bad choice of t_1, t_2, the matrix in 2 will be singular, caused by linear dependence of the rows, and we can never recover both peak heights. In this case one of the singular values is 0.

To get a bound on the best case, note that the real exponentials associated with T_2 decay are always less than one, but for short times will be close to one. For a rough estimate, we ignore the decay. This results in

$$S^*S = \begin{pmatrix} 2 & e^{i(\omega_b-\omega_a)t_1} + e^{i(\omega_b-\omega_a)t_2} \\ e^{i(\omega_a-\omega_b)t_1} + e^{i(\omega_a-\omega_b)t_2} & 2 \end{pmatrix}, \tag{3}$$

expanding into real and imaginary parts, we see the symmetry:

$$= \begin{pmatrix} 2 & x+iy \\ x-iy & 2 \end{pmatrix},$$

and with some algebra we see that the product of the eigenvalues is $4-x^2-y^2$. The maximum 4 is realizable by choosing t_1 and t_2 such that the off-diagonal elements are zero, in which case the eigenvalues are two. This result can be interpreted in terms of statistics. The matrix S^*S proportional to the inverse of the covariance matrix, which is diagonal if the peak estimates are not correlated, and hence give the maximum information.

In general, the maximum eigenvalue is bounded by the number of samples, hence our objective, the minimum eigenvalue, is also bounded by the number of samples. Let n be the number of samples, m the number of residues, and r the dimension of the experiment. The matrix S^*S has size $m \times m$, while the value of the matrix elements, and the computational cost of constructing them are bounded by n, the number of samples.

The exact optimum depends on the T_2s, and cannot be calculated analytically, so we report our efficiency in terms of how close a sampling pattern gets to this theoretical upper bound without relaxation, which is the same as n, the number of samples. Efficiency is a relative measure of accuracy. Accuracy, defined as the inverse of the expected error bars on reported peak amplitudes, can always be increased by collecting more samples, whether through averaging or collecting data at different delay times. Efficiency is the ratio of expected accuracy divided by the number of samples collected.

For two peaks, it is possible to select good sampling times by inspection, and invert the transformation by hand. For many peaks, and more samples than peaks, numerical methods are required, based on the Moore-Penrose pseudo-inverse. We seek to maximize

the signal to noise in this process. To define it and see the effect on noise, write the signal generation equation as

$$\begin{pmatrix} h(t_1) \\ \vdots \\ h(t_n) \end{pmatrix} = S \begin{pmatrix} f(\omega_1) \\ \vdots \\ f(\omega_m) \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{4}$$

where S is the complex $n \times m$ matrix

$$S_{i,j} = e^{\sqrt{-1}\langle t_i, \omega_j \rangle},$$

and the vector of ϵs represents the measurement noise.

To solve for the effect of noise on the amplitude estimates, multiply both sides of 4 by the adjoint, $S^* = \overline{S}^T$, and solve for f.

$$S^*S \begin{pmatrix} f(\omega_1) \\ \vdots \\ f(\omega_m) \end{pmatrix} = S^* \begin{pmatrix} h(t_1) \\ \vdots \\ h(t_n) \end{pmatrix} - S^* \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Assuming that sufficiently many samples have been collected to be able to reconstruct the peak amplitudes, S^*S is invertible, so we can write this as

$$\begin{pmatrix} f(\omega_1) \\ \vdots \\ f(\omega_m) \end{pmatrix} = (S^*S)^{-1}S^* \begin{pmatrix} h(t_1) \\ \vdots \\ h(t_n) \end{pmatrix} - (S^*S)^{-1}S^* \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \tag{5}$$

where the first term on the right hand side is the estimate, and the second term is the apparent noise. The linear system used for the estimation $(S^*S)^{-1}S^*$ is known as the Moore-Penrose pseudo-inverse. This shows the role of S in amplifying the noise. The factor S^*S is the Fisher information matrix of design theory, and, as noted above, proportional to the inverse of the covariance matrix, and the key to formulating an optimization problem. We can maximize the signal to noise ratio in different ways, depending on whether we are concerned with the experiment overall, or with individual peaks. If we need to measure each individual peak independent of the other peaks, we are concerned with the worst-case signal to noise, which corresponds to the minimum eigenvalue of S^*S.

We can weigh the efficiency of existing sampling patterns using this measure. In statistics this is called E-optimality, [17]. In the limit when the number of samples is very large and well distributed, the eigenvalues are all the same and E-optimality is equivalent to a more commonly used statistical measure, D-optimality, which considers the product of the eigenvalues, rather than the minimum. Geometrically, D-optimality minimizes the volume of the uncertainty ellipsoid (higher dimensional error bar), whereas E-optimality minimizes the radius. We believe that for the T_2 experiments there is enough signal to noise that we can reduce experiment times to the point where we are far from the high-sampling regime and E- and D-optimality provide different solutions. In this case, minimizing the worst case expected error (including apparent false correlations) is the right criterion.

Example: Dense Sampling. The easiest case to understand is dense sampling, in which the spectrum is reconstructed using a FFT. Since all the eignenvalues of the Fourier transform have norm 1, the noise level in the reconstructed spectrum is constant. So all peaks with equal widths will be effected by the same measurement noise. Other sampling patterns can be analysed numerically, and we report below that they all have this property if they include sufficiently many samples.

2.1. Efficiency

Numerically, we define efficiency as the minimum eigenvalue of S^*S divided by the number of samples. With this definition, the efficiency is proportional to sampling time requirements. It would have been equally natural to report efficiency in terms of ratios of standard deviations, which can be obtained by taking the square root of the reported efficiencies.

For the Fourier Transform applied to dense sampling, and with peak positions corresponding to all grid points, all eigenvalues of S^*S are equal to the number of samples, which is equal to the number of frequency grid points. So the peak efficiency (without relaxation) is achieved.

This definition of efficiency does not take into account extra precision in peak amplitude obtained by using several points in the spectrum to calculate an area, since this will depend on operator choices.

2.2. Optimization problem

[18] introduced a method of optimizing a *set* of steady-state MRI experiments with respect to expected noise. The same eigenvalue maximization approach using a semi-definite constraint can be adapted to the present problem. The present problem and the problem in [18] are both part of a general optimization problem of NMR experiments. In the previous work, we considered the effect of pulse sequence parameters on the generation of a steady state, and ignored the effect of sampling in frequency space (referred to as k-space in Magnetic Resonance Imaging), and the effect of changing repeat times. In this paper, we ignore the pulse sequence design issues, and focus on the sampling in the indirect frequency dimensions.

Maximizing the minimum eigenvalue of S^*S can be formulated as a semidefinite programming problem: given parameters $\omega_i \in R^r$,

$$\min_{\{t_i\}} \quad -\tau \tag{6}$$

$$\text{subject to} \quad S^*S - \tau I \succeq 0. \tag{7}$$

The inequality $\succeq$ in 7 is a semi-definite (matrix) inequality, see [19]. The variable τ which is equal to the minimal eigenvalue of S^*S, was introduced to put the problem in the standard form. The effect of this inequality is to require the eigenvalues in 7 to be non-negative, which is equivalent to bounding the eigenvalues of S^*S below by τ. The practical importance of formulating the problem in this way is that there are a growing number of open-source or academically available solvers for semi-definite problems, [20]. Available solvers cannot handle problems in quite the level of generality required for our work, but the problem may be decomposed using a trust-region method.

A trust-region method is a standard iterative approach to nonlinear, possibly non-convex optimization, see [21]. At each iteration, the function to be optimized is approximated by a simpler function. The approximation is then minimized within a "trust region", after which the original function is evaluated. If this results in no improvement, one shrinks the trust region, and tries again. If the improvement is better than a threshold value, one expands the trust region for the next iteration. For details of this formulation, see the technical report [22].

Without further insight into practical NMR experiments, this problem would not be solvable by current-generation software. Fortunately, all NMR experiments have a directly acquired dimension, in which sampling is dense relative to sampling in other dimensions, and Fourier methods are cheap.

2.3. One densely sampled direction

In real NMR experiments, there is always a directly acquired dimension. We have assumed that dense FIDs are acquired for hydrogen, but this is not required. Using this structure, the dimension of the subproblems can be significantly reduced. Samples are

organized in dense lines parallel to the directly sampled direction, so only the first sample point, $(0, t_{j,2}, t_{j,3}, \ldots)$ in each FID needs to be modelled. The other points in the FID are $(1, t_{j,2}, t_{j,3}, \ldots), (2, t_{j,2}, t_{j,3}, \ldots), \ldots$ in the appropriate units.

It follows that the matrix S^*S becomes

$$(S^*S)_{ij} = \left(\sum_{s=0}^{n_0-1} e^{\sqrt{-1}\cdot s\cdot\delta\cdot(\omega_{j,0}-\omega_{i,0})}\right)\left(\sum_{l=1}^{n/n_0} e^{\sqrt{-1}\langle t_l,\omega_j-\omega_i\rangle}\right) \tag{8}$$

where n_0 is the number of samples per FID, and δ is the inter-sample spacing. This reduces the cost of constructing each element of S^*S from rn to $(r-1)n/n_0$.

The resulting optimization problem is still too large for biologically-interesting proteins, but we have more structure to exploit. The first factor in 8 approaches zero on off-diagonal elements ($i \neq j$) as the peak frequencies in the direct dimension become separated, because this term is the dot product of Fourier basis functions. These functions are orthogonal as continuous functions, and nearly orthogonal when the corresponding peaks are visually separated in the corresponding spectrum. This allows us to decompose the matrix S^*S into diagonal blocks, corresponding to regions of the spectrum with overlap, ignoring very small components on the off diagonal. In terms of multidimensional spectra, we would not expect to have trouble separating peaks if they are already widely separated in the 1D hydrogen spectrum.

We exploit the fact that not all peaks overlap in the directly acquired spectrum by dividing the peaks into groups, such that overlap only occurs within groups. If each group can be resolved separately, then we can resolve all peaks. The groups are indexed by the 'fat' hyperplanes they occupy. Let P be the set of such planes. Since the directly acquired dimension is densely sampled, it is expected that the FIDs optimized for one plane p provide information about all other planes, further improving efficiency.

Another way of thinking of this is that since the directly acquired dimension is densely sampled, it would be possible to perform the Fourier transform in this dimension and consider each frequency in this dimension separately, as if it were possible to do a number of $(r-1)$ dimensional experiments each involving only a single set of peaks which overlap in the direct dimension.

This reduces the number of variables in the problem from rn to $(r-1)n/n_0$, and reduces the size of the most difficult element of the optimization problem, the semidefinite constraint, from $m \times m$ to $m_p \times m_p$, where n_0 is the number of samples per FID, r is the number of dimensions, and m_p is the number of peaks assigned to plane $p \in P$, the set of 'fat' hyperplanes. Of course, instead of one problem, one must solve a separate problem for each plane. The much smaller problems can be solved by existing software.

We will use the union of the sets of points as our experimental design. Solving different problems for each plane does not take into account that a single set of FIDs are acquired. By optimising sets of sample times for each plane, we are underestimating the efficiency of the sampling by not taking into account the sampling times acquired for other planes. Practically, this is only important if the number of samples grows large, which we have not observed in our numerical problems.

In the current implementation, we use the frequency in the directly-acquired direction to cluster the peaks. Hierarchical clustering is a general method which starts a cluster with a single peak and iteratively adds peaks which overlap any peak in the cluster until no more peaks need to be added. Hierarchical clustering works for the benchmark problem, but it is not optimal for very dense spectra in which such clusters are so big that most pairs of peaks within a "fat" plane do not themselves overlap. How to optimally cluster the peaks is an open problem for dense spectra, but workable solutions are available. Our current best solution uses two steps: first, cluster peaks with substantial overlap, and in a second step add points from neighbouring clusters when those points somewhat overlap points in the cluster. This approach is applicable to all spectra, including spectra without any gaps, but using it increases the amount of computation by assigning some peaks to more than one fat plane. Since any set of mutually overlapping peaks is contained in some fat plane, the reported efficiency will be a lower bound on the true efficiency.

To optimize sampling in such a clustered plane of peaks, we have tried two approaches: (1) budgeting a fixed number of samples and optimizing the sampling times; and (2) setting a target efficiency, and a floor on the number of sampling points, optimizing an initial set of sampling times corresponding to the floor, and until the target is met, adding new sampling points one at a time and reoptimizing the set of sample times. The second method could require an arbitrary number of samples, and is, not surprisingly, much slower, but always achieves the required efficiency with a minimum of samples. In this paper, only the first method is used. For a numerical comparison, see [22].

2.4. Sentinel points

Using prior information about the model in optimizing the solution of an inverse problem introduces the risk that the prior information may become invalid. The more efficient the optimized solution, the greater the danger. One way of detecting model failures in this problem is to add sentinel points: ω_is not corresponding to defined peaks. If the model is valid, the estimated signals at these points will be pure noise. Depending on the possible sources of failure, the co-ordinates of the sentinel points can be chosen to detect such problems. To test for machine failures, arbitrary points can be used. In the numerical examples in this paper, two points were added to each of the clustered planes. The projections of the sentinel points onto the hyperplane orthogonal to the directly acquired direction were the same for all planes, and the points were chosen to be well within the convex hull of the projected points, but well separated from actual peak values.

3. Numerical results

In this section we use numerical experiments to explore the sensitivity of the proposed method to variations in parameters, and the applicability of the current implementation to realistic proteins. There are two questions: (1) can we optimize for realistic proteins repeatedly, and (2) how sensitive are the designs to changes in parameters (primarily peak frequencies, but also relaxation times). We also address the effect of total sampling time, which is a user-controllable parameter, on the sensitivity of the design.

The group of Dr. Giuseppe Melacini provided us with assignment tables of protein RIα (119–244), a cAMP-binding regulatory subunit of protein kinase A (PKA) [23–26]. We also optimized ubiquitin, since published data are available in folded and denatured versions, which allows us to evaluate the effect of overlap. We used the RIα data to investigate the sensitivity of this method to different constraints, design choices and parameter uncertainty. Having extracted good design settings from these experiments, we tried to optimize sample times for folded and denatured ubiquitin.

To test our optimization approach, all algorithms were implemented in C, calling the library CSDP 2.3. Resonance frequencies from an assignment table for the protein RIα (119–244) were used. Not all frequencies for all residues were given, and these residues were ignored, leaving 113 points. Some frequencies could not be determined because of overlap, these can usually be resolved in higher dimensions. If the peaks are not of interest, one or more peaks can be added to estimate the overlapped peak so that their

signals do not contaminate other peak measurements, but this is not required for benchmarking the optimization algorithm.

Since reaching high levels of efficiency with very sparse sampling will make higher dimensional experiments practical, we repeated most of the numerical tests with multiple dimensions, using the seventeen PKA residues with six recorded resonance frequencies (corresponding to hydrogen, nitrogen, and four carbons).

The lower dimensional experiments used the first resonances in the order listed.

3.1. Repeatability

In the first experiment we consider a three indirect dimensional (i.e. 4D) experiment, and took the nine peaks in the first clustered plane. In Fig. 1 the numerical optimization method consistently converges to 100% efficiency, and the efficiency as a function of the number of sample times is much less variable than the results of simple greedy optimization.

In Fig. 1 the red (+) efficiencies are the result of initially selecting eight random sets of nine sampling times and plotting the calculated minimum eigenvalue, and at each subsequent step trying eight new random sampling points and keeping the one which results in the largest minimum eigenvalue. In addition to this simple greedy algorithm, the green (x) points plot the efficiencies which initially resulted from applying the trust region method to three sets of nine, and on subsequent iterations, taking the best result after adding eight new random sampling times to each set one at a time and re-running the trust-region method on the whole set of points.

What is most important here is not the approach to 100% efficiency, nor the slowing improvement of greedy optimization, but the large gap between the two methods for very small numbers of sample points. For very sparse sampling, numerical optimization enables significant reduction in sampling with much lower loss of efficiency, and therefore higher signal to noise.

We use the 17 peaks with all six resonance frequencies as our problem for all the numerical experiments, unless stated otherwise.

3.2. Effect of dimension, initial sampling pattern

To test the efficiency of existing sampling patterns and the sensitivity of the optimized pattern to the initial distribution of sample times, we randomly chose sets of 40 initial sampling times from uniform and exponential distributions in the range [0 ms 20 ms], and also chose random radii to use in radial sampling patterns, but the results for radial sampling are not shown since they were indistinguishable from uniform distributions. Efficiency was measured and then optimized against the same set of random T_2s in the range of 200 ms for the 17 peak set. The exponentially distributed initial points were generated using a decay parameter (λ) of 1/200 ms. Although numerically the exponentially distributed points were better than random, see Fig. 2, the difference is not significant relative to the effect of numerical optimization.

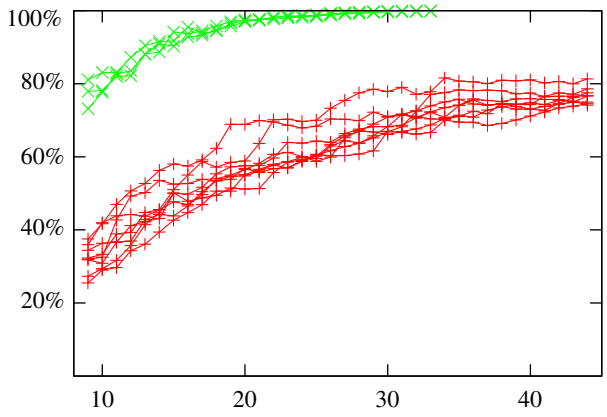

Fig. 1. Numerical optimization shows consistent results with different starting sets, and is much more efficient for very small numbers of samples. Shown are efficiency for infinite relaxation times. Sample points were added one by one, and are plotted on the horizontal axis, starting with 9 samples (also the number of peaks). The efficiency is plotted on the vertical axis.

3.3. Sensitivity to peak frequencies

We evaluate the sensitivity of the optimized sampling trajectories to peak frequencies similarly.

(1) Peak frequencies, and optimized sampling times from the last test were taken.
(2) The peak frequencies were randomly perturbed by normally distributed differences, generated for each of 10 different standard deviations.
(3) The efficiency for each set of perturbed values was calculated, and plotted as a function of the standard deviation of the perturbation used (Fig. 3).

Fig. 3 shows that optimized sampling is sensitive to variations in peak frequencies. For these methods to work it is important that assigned frequencies are accurate to within 10 Hz.

A densely sampled experiment is optimal for all discrete frequencies generated by the FFT. At some point, adding additional sample points will reduce the sensitivity of the design to changes in peak frequencies. Unfortunately, small changes (up to doubling) the number of sampled times has little effect on the sensitivity to peak locations. To determine this, we measured the efficiency of an optimal design for frequencies obtained by perturbing the originals using random deviations with standard deviation 10 Hz, for designs using 40, 50, 60 and 70 sets of delay times. In Fig. 4, we see that this has no significant effect.

3.4. Sensitivity to maximum delay time

In conventional experiments, the maximum delay time determines resolution, and is chosen with this in mind, but the maximum delay time will also effect the sensitivity to errors in peak frequencies, because in our model frequencies are always multiplied by delay times. To test this sensitivity,

(1) Different sets of sampling times were generated, uniformly distributed over the ranges [0 s,2.5 s],[0 s,5 s],[0 s,10 s] and [0 s,20 s].
(2) Each set of sampling times was optimized.
(3) The peak frequencies were perturbed by normally distributed differences of magnitude 80 Hz.
(4) The efficiency of the original sampling pattern with perturbed values were calculated, and plotted as a function of the maximum delay time in Fig. 5.

Note that the maximum sampling range was the initial range, and that the optimization was not constrained to keep points in this range. Nevertheless, more than 90% of them remained within the initial range.

Fig. 5 clearly shows our sampling becomes more robust to variations in peak frequencies if initial sampling times are constrained to be short. Fig. 6, which plots the attained efficiency as a function

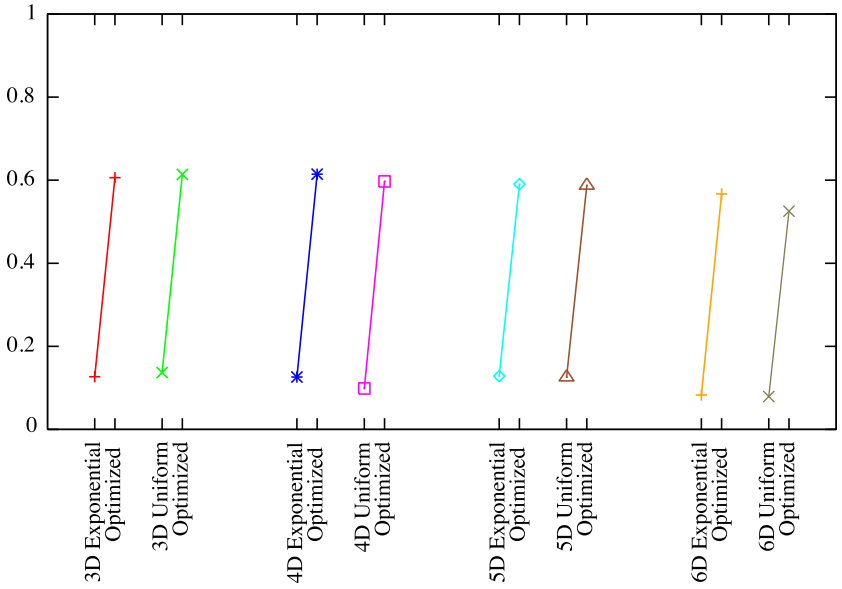

Fig. 2. Efficiency of optimized sampling is very slightly improved by starting with exponential sampling. The efficiency of the initial exponentially and uniformly distributed sampling patterns is connected to the efficiency after optimization, for different numbers of indirect dimensions. Each colour corresponds to the efficiency of one set of sampling times before and after trust-region optimization.

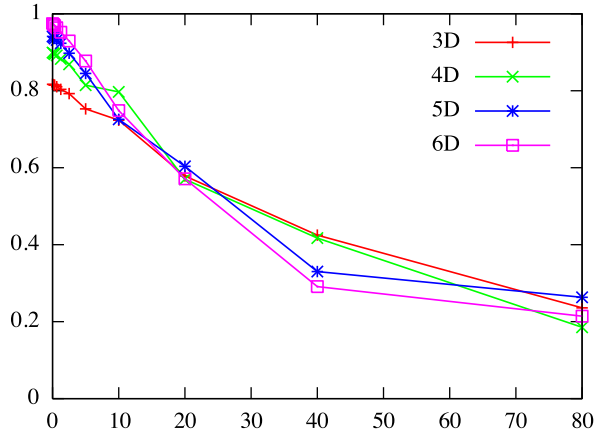

Fig. 3. Efficiency of optimized sampling is sensitive to variation in peak frequencies. The efficiency is plotted on the vertical as a fraction of the theoretical maximum is plotted against the standard deviation in Hz for each number of indirect dimensions.

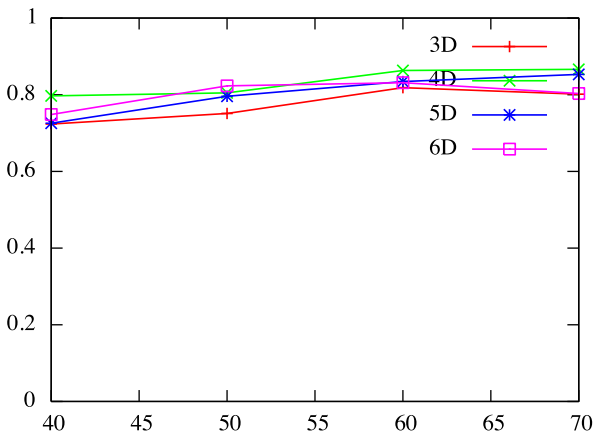

Fig. 4. Increasing the number of samples does not make our sampling more robust to variations in peak frequencies. The number of samples in the experiment is plotted on the horizontal axis.

of maximum delay time, shows that efficiency is not compromised with very short sampling ranges. Note that sampling times were constrained in the initial generation, but during optimization, some sampling times exceeded the initial constraint. Nevertheless, at least 90% of samples remained in the constrained interval, and no sample times exceeded the initial limit by a factor of two. Adding constraints degrades performance of the optimizer, and it was felt that there is no reason to make the maximum sample time a hard limit.

3.5. Sensitivity to T_2^*

We first consider the effect of relaxation on the design problem. The method we propose optimizes the sampling times before the relaxation times are known, so it is important that the results remain valid over a range of relaxation times. To do this we compare the reported sampling efficiency given nominal T_2s and in the case that these T_2s are inaccurate estimates, using the following procedure:

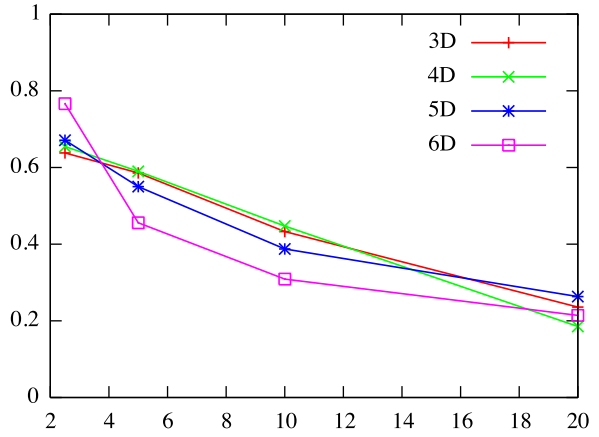

Fig. 5. Optimized sampling patterns are more robust when the totaling sampling range is shortened.

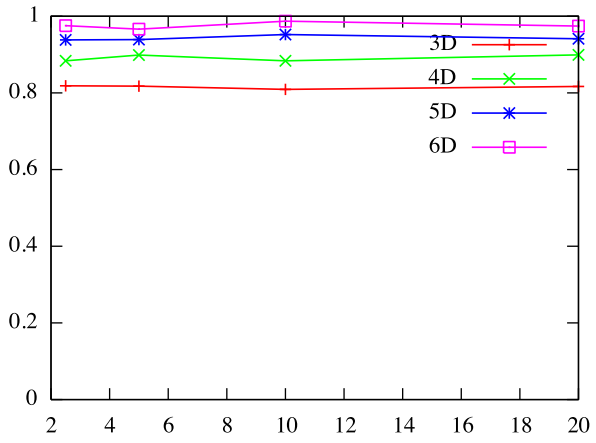

Fig. 6. Shortened sampling ranges do not compromise the efficiency of sampling, for experiments with at least three indirect dimensions. The efficiency, plotted on the vertical axis is nearly independent of the maximum delay time, plotted in ms on the horizontal axis.

(1) 40 initial random sampling points were generated in [0 ms,20 ms].
(2) "True" T_2^* values were randomly generated in the range [150 ms,250 ms].
(3) the sampling times were optimized and the efficiency reported,
(4) the T_2^* values were randomly perturbed by normally distributed differences, generated for each of 10 different standard deviations
(5) the efficiency for each set of perturbed values was calculated
(6) the efficiency as a function of the standard deviation of the perturbation was plotted (Fig. 7).

This was repeated for experiments of dimension 3, 4, 5 and 6. The conclusion is that the optimal design is not sensitive to variation in T_2^*, so we are justified in optimizing sample times before the T_2^*s are available.

3.6. Full experiments

The full set of 113-RIα residues with measured H, N and C resonances was used to test the clustered hyperplane approach.

Applying the greedy random + trust algorithm to all planes (with 2 sentinel peaks), using a single set of samples with triple the cardinality of the plane, resulted in efficiencies shown in Table 1. Efficiency decreases as the cardinality of the plane increases. This computation took 3 h on a 2.6 GHz, 8 Dual Core Opteron server in a shared environment. To optimize NMR experiments which can run days on expensive spectrometers, this is more than justified.

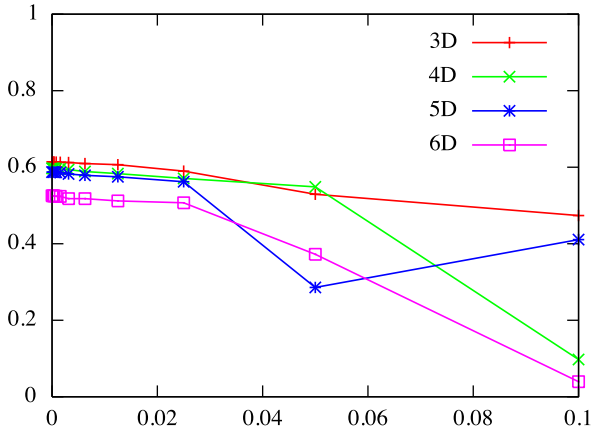

Fig. 7. Efficiency of optimized sampling is insensitive to some variation in T_2^*. The efficiency as a fraction of the theoretical.

When 512-point FIDS with μs sampling are collected for the union of these 360 *k*-space points, the sampling efficiency as measured by the minimal eigenvalue of *A* will be 88.2%. This number is higher than the efficiency for the 38-peak plane because (1) the clustered peaks are only partially overlapping, and (2) more than half the total points optimized for other planes also contribute.

As a final test, assignment tables for Folded and Denatured Ubiquitin were obtained from the BMRB database, and optimized HNCO experiments were designed using initial dwell times in the range [0 ms, 20 ms].

As expected, the folded version is much easier to optimise for since it has much less overlap among peaks. Table 2 shows the division into four sets of peaks for which sampling times were optimized and the net efficiency when all 185 sampling times are combined.

The denatured ubiquitin has a much more crowded spectrum, and conservatively clustering overlapping peaks together results in very large planes for which the trust-region optimization algorithm fails. Clustering only strongly overlapping peaks results in much smaller sets, but does not capture all overlapping pairs of peaks. To resolve this problem, peaks which partially overlap a peak in a neighbouring cluster were added to the neighbouring cluster one by one until every pair of overlapping peaks was contained in at least one set. The resulting overlapping clusters contained 33 duplicated peaks. The efficiencies for each clustered set was lower than for the folded case, but the net efficiency after combining all the delay times is higher because some of the peaks are duplicated. The denatured protein would require almost double the sampling time, but the expected signal to noise (assuming equal molar concentrations) would be higher than for the folded case Table 3.

3.7. Simulation

As a check on correctness of the reconstruction, and a verification of the reported noise estimates. Data corresponding to 113 peaks with additional noise was added, and the peak values were estimated by solving 5 using the conjugate gradient method. The observed variances matched the results of the optimization.

4. Conclusion

This paper proposes a general model for optimizing the delay times used in indirect dimensions. Numerical tests, based on a

Table 1
Efficiencies found by greedy random + trust method for all planes clustered from 3d RIα (119–244) peaks. The notation 7 + 2 means that 2 sentinel peaks were added to 7 peaks in the clustered plane.

Plane	Number of peaks	Number of samples	Efficiency
1	7 + 2	26	99%
2	38 + 2	119	72%
3	26 + 2	83	82%
4	15 + 2	50	87%
5	22 + 2	71	80%
6	2 + 2	11	100%
Overall	104 + 12	360	88%

Table 2
Folded Ubiquitin (BMRB 5387, [27]).

Plane	Number of peaks	Number of samples	Efficiency
1	15	40	77%
2	20	60	73%
3	20	60	77%
4	10	25	89%
Overall	65	185	85%

Table 3
Denatured Ubiquitin (BMRB 4375, [28]).

Plane	Number of peaks	Number of samples	Efficiency
1	10	25	78%
2	16	50	47%
3	17	50	74%
4	22	80	57%
5	22	80	27%
6	11	25	72%
Overall	65+33 repeats	320	75%

truly typical set of protein data, show great promise. They also show that for the relaxation experiment, in which peak positions are already known, none of the existing heuristics provide significantly better efficiency than randomly selected delay times when compared to optimized sampling. One of the surprises in the numerical simulations was the possibility that higher dimensions can be acquired at lower cost than lower dimensional experiments. At this point it is only a possibility, made possible by the fact that higher dimensional experiments can be optimized for lower numbers of samples than lower dimensional experiments. At the same time, the efficiency achieved may be lower at higher dimensions. Efficiency is a function of dimension, maximum delay time, and relaxation times. Ignoring relaxation, results in better efficiency in higher dimensions. With typical relaxation times and delay times, the efficiency is slightly reduced for higher dimensional sampling. Given the fine balance between these factors, it is likely that higher dimensions will win out for specimens with high levels of signal, capable of supporting very sparse sampling, but lower dimensional experiments will be preferred for specimens with very low concentrations, supporting low signal to noise ratios. We are currently pursuing experimental validation of these results.

Acknowledgements

The authors acknowledge Giuseppe Melacini and Rahul Das who provided data for benchmarking. NSERC, CFI and OIT provided research support.

References

[1] R.R. Ernst, G. Bodenhausen, A. Wokaun, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, Oxford, 1987.
[2] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, N.J. Skelton, Protein NMR Spectroscopy, Principles and Practice, Academic Press, San Diego, 1996.
[3] E. Kupce, R. Freeman, Hyperdimensional NMR spectroscopy, J. Am. Chem. Soc. 128 (18) (2006) 6020–6021.
[4] E. Kupce, R. Freeman, Fast multi-dimensional NMR by minimal sampling, J. Magn. Reson. 191 (2008) 164–168.
[5] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise highdimensional NMR spectral information, J. Am. Chem. Soc. 125 (2003) 1385–1393.
[6] E. Kupce, R. Freeman, Fast multidimensional NMR: Radial sampling of evolution space, J. Magn. Reson. 173 (2005) 317–321.
[7] B.E. Coggins, P. Zhou, Sampling of the NMR time domain along concentric rings, J. Magn. Reson. 184 (2007) 207–221.
[8] P. Schmieder, A. Stern, G. Wagner, J. Hoch, Quantification of maximum-entropy spectrum reconstructions, J. Magn. Reson. 125 (1997) 332–339.
[9] J.C. Hoch, A.S. Stern, NMR Data Processing, Wiley, Liss, 1996.
[10] V.A. Jaravine, A.V. Zhuravleva, P. Permi, I. Ibraghimov, V.Y. Orekhov, Hyperdimensional NMR spectroscopy with nonlinear sampling, J. Am. Chem. Soc. 130 (12) (2008) 3927–3936.
[11] J.A. Jones, P. Hodgkinson, A.L. Barker, P.J. Hore, Optimal sampling strategies for the measurement of spin-spin relaxation times, J. Magn. Reson. B 113 (1996) 25–34.
[12] J.A. Jones, Optimal sampling strategies for the measurement of relaxation times in proteins, J. Magn. Reson. 126 (1997) 283–286.
[13] A. Bain, The choice of parameters in an NMR experiment: application to the inversion-recovery t1 experiment, J. Magn. Reson. 89 (1990) 153–160.
[14] R.J. Ober, Z. Lin, H. Ye, E.S. Ward, Achievable accuracy of parameter estimation for multidimensional NMR experiments, J. Magn. Reson. 157 (1) (2002) 1–16.
[15] J. Habazettl, A. Ross, H. Oschkinat, T.A. Holak, Secondary NOE pathways in 2d NOESY spectra of proteins estimated from homonuclear three-dimensional NOE–NOE nuclear magnetic resonance spectroscopy, J. Magn. Reson. (1969) 97 (3) (1992) 511–521.
[16] K. Nagayama, P. Bachmann, K. Wüthrich, R. Ernst, The use of cross-sections and of projections in two-dimensional NMR spectroscopy, J. Magn. Reson. 31 (1978) 133–148.
[17] F. Pukelsheim, Optimal Design of Experiments, Vol. 50 of Classics in Applied Mathematics, SIAM, 2006.
[18] C.K. Anand, R. Sotirov, T. Terlaky, Z. Zheng, Magnetic resonance tissue quantification using optimal bSSFP pulse-sequence design, Optimization and Engineering 8 (2) (2007) 215–238.
[19] H. Wolkowicz, R. Saigal, R. Saigal (Eds.), Handbook Of Semidefinite Programming: Theory, Algorithms, And Applications, Springer-Verlag, 2000.
[20] B. Borchers, CSDP, a C library for semidefinite programming, Optimization Methods Software 11 (1999) 613–623.
[21] A.R. Conn, N.I.M. Gould, P.L. Toint, Trust Region Methods, MPS—SIAM Series on Optimization, 2000.
[22] C.K. Anand, A. Sharma, Optimal fourier transform sampling patterns, Tech. rep., McMaster University (2007).
[23] R. Das, V. Esposito, M. Abu-Abed, S. Taylor, camp activation of PKA defines an ancient signaling mechanism, Proc. Natl. Acad. Sci. USA 104 (2007) 93–98.
[24] R. Das, M. Abu-Abed, G. Melacini, Mapping allostery through equilibrium perturbation NMR spectroscopy, J. Am. Chem. Soc. 128 (2006) 8406–8407.
[25] C. Kim, D. Vigil, G. Anand, S. Taylor, Structure and dynamics of PKA signaling proteins, Eur. J. Cell Biol. 85 (2006) 651–654.
[26] C. Kim, C. Chang, S. Saldanha, S. Taylor, PKA-1 holoenzyme structure reveals a mechanism for camp-dependent activations, Cell 130 (2007) 1032–1043.
[27] P. Flynn, M. Milton, C. Babu, A. Wand, A simple and effective NMR cell for studies of encapsulated proteins dissolved in low viscosity solvents, J. Biomol. NMR 23 (2002) 311–316.
[28] W. Peti, L. Smith, C. Redfield, H. Schwalbe, Chemical shifts in denatured proteins: resonance assignments for denatured ubiquitin and comparisons with other denatured proteins, J. Biomol. NMR 19 (2001) 153–165.
[29] K. Kazimierczuk, W. Koźmiński, I. Zhukov, Two-dimensional Fourier transform of arbitrarily sampled NMR data sets, J. Magn. Reson. 179 (2006) 323–328.