ELSEVIER

Progress in Nuclear Magnetic Resonance Spectroscopy 34 (1999) 257–299

PROGRESS IN NUCLEAR
MAGNETIC RESONANCE
SPECTROSCOPY

Linear prediction spectral analysis of NMR data

P. Koehl[a,b]

[a]*Department of Structural Biology, Fairchild Building, Stanford, CA 94305, USA*
[b]*UPR 9003 du CNRS, Boulevard Sebastien Brant, 67400 Illkirch-Strasbourg, France*

Received 10 November 1998

Contents

PII: S0079-6565(99)00002-3

Keywords: Fourier transformation; Linear prediction; Autoregression methods; Extended Prony method

1. Introduction

Nuclear Magnetic Resonance (NMR) signals were originally derived from continuous wave (CW) spectroscopy. CW spectra were recorded by slowly and continuously scanning the frequency axis, yielding directly all intensities related to the resonance frequencies contained in the sample of interest [1]. The same information however can be obtained by recording the evolution of a combination of these intensities with respect to time, after proper preparation, followed by proper deconvolution of the individual intensities, usually based on Fourier transformation. The resulting experimental times are considerably reduced, and the main advantage is a dramatic increase in signal-to-noise ratio. A simple explanation for this can be derived from the following argument. If a combination of two properties P1 and P2 is recorded as a single value, two data points D1 and D2 at least are required to retrieve P1 and P2. Assuming that the deconvolution proceeds through a linear combination of the data points D1 and D2, the intensities of the ''signal'' associated with P1 and P2 follow this combination, while error propagation theory tells us that the ''noise'' is related only to the square root of the combination of errors on D1 and D2 (assuming the latter to be independent). The development of signal processing techniques, the introduction of fast techniques to calculate digital Fourier transform such as the Cooley–Tukey algorithm [2], as well as the development of the computer technology transformed this essential finding of Ernst and Anderson [3,4] into a major breakthrough for NMR spectroscopy [5]. Since then, NMR has experienced other major changes. An account of these transformations can be found in the Nobel lecture of Ernst [6], or in more detail in the recent Encyclopedia of NMR [7]. Spectrometers with magnets of 14.1 and 18.8 T (600 and 800 MHz for ^{1}H spectroscopy) are now quite common for high resolution liquid-state NMR, and manufacturers have prototypes and projects for much higher fields. The electronic components have drastically improved, heading towards a fully digital system. A great deal of effort has been put into developing new experiments [8], which are now multidimensional [5,9–11], and usually involve several spin types [12–15]; the samples themselves are chemically or biologically modified in order to enrich them in normally rare observable spins, such as ^{15}N and ^{13}C [16–19], at a reasonable price in the case of biological samples, as a result of the progress of biotechnologies. All these changes are now standard. NMR was at first concerned with interpreting spectra in terms of chemical structures, then with species identification and dynamics studies. The range of applications of NMR techniques has also increased

significantly, including in vivo studies [20–24], as well as three-dimensional structure determination and dynamics studies of macromolecules, using high resolution NMR [25–31]. In contrast to all these changes, after 30 years, Fourier Transform remains the standard procedure for NMR signal processing, though new signal processing techniques have emerged during the same period of time (for recent reviews, see [32,33]).

It is important to note at this point that signal processing in NMR may have very different purposes, depending on the applications. A high resolution NMR spectroscopist who studies a large biomolecule in solution usually has little a priori knowledge regarding the spectroscopic properties of the molecule. He is first confronted with the problem of correctly identifying in the NMR spectra a large number of chemically-shifted nuclei, as well as defining the corresponding underlying chemical structure. The presence of a large number of components requires that sophisticated experimental techniques be applied, such as multidimensional spectroscopy. The corresponding experiments are usually costly in time. Signal processing in these cases aims at maximizing the extraction of information from these experiments, mainly for the purpose of spectral assignment. In the case of in vivo experiments, the chemical species under study are known, as well as their NMR spectra. For these applications, signal processing is geared toward quantification in the context of poor signal-to-noise ratio. Missing data points are a common problem in some applications of solid state NMR, and signal processing is then expected to help recover information from these incomplete signals. Fourier transformation is the main common denominator in all these forms of signal processing. We will see in the following that it has its limitations.

Fourier analysis provides a mathematical model in which any function, and in particular any signal, is modeled as a sum of sine and cosine waves. The coefficients of these waves which provide the best fit to the original function are stored in the so-called spectrum, i.e. the Fourier Transform of the signal (see for example [34]). The FT is bijective and has an inverse (iFT), hence the same information can be retrieved in both the signal (i.e. time-domain information, referred to as the FID for free induction decay) and the spectrum (frequency-domain information). FT operates on continuous functions; an equivalent of FT exists for discrete signals, the Discrete Fourier Transform, or DFT. Analysis of a given discrete finite (N values) signal $\{x_n\}$ by means of DFT assumes only that the signal is sampled with constant time intervals and that it is zero outside of the observation window, which is equivalent to saying that the signal has been multiplied by the boxcar function ($b_n = 1$ for $n = 0,\ldots,N-1$ and 0 otherwise). Solutions have been proposed to alleviate the first condition, primarily in the case of a small number of missing data points within a large signal (see for example [35]). Missing data points at the beginning of the time series are more difficult to handle, resulting in severe distortions of the baseline of the spectrum. The consequence of the second condition (i.e. the multiplication by the boxcar function) is a spectrum distorted by a convolution with a Sine Cardinal function ($\mathrm{Sinc}(x) = \sin(x)/x$), i.e. the Fourier Transform of $\{b_n\}$ shows ''wiggles'' near each of its peaks. Attempts to attenuate this problem usually proceed through apodization, of the signal, i.e. multiplication of $\{x_n\}$ by a proper function such as to set the signal to 0 at the last experimental point. Apodization can either provide improvements in spectral resolution by splitting overlapping peaks, or improvements in the signal-to-noise ratio by recovering weak signals lost in the noise. Both enhancements cannot be achieved simultaneously. Also, performance is ultimately limited by the Nyquist theorem, which sets the ''dwell time'', i.e. the sampling time interval, according to the spectral range of interest. These limits prevent a quantitative analysis of crowded spectra which show severe peak overlaps. In these cases, alternative approaches have been proposed, which make use of additional information, such as models of the underlying physical signal and noise in $\{x_n\}$. These methods include wavelet transforms [36–40], Non-Linear Least Squares (NLLSQ) fits in the frequency domain [41], Padé–Laplace analysis [42–44], maximum entropy methods [45–52], Bayesian and maximum likelihood techniques [53–61], genetic algorithms [62], and linear prediction [63–66]. Several reviews of the latter method have been published (see for example [41,67–69]). In this review, we will focus on the recent areas of development related to this

technique, covering both methodological aspects with emphasis on matrix computation, as well as applications.

This review is organized as follows. First, the concept of linear prediction is explained through a simple example. Then, the various numerical techniques available for solving the linear prediction equations are described, with emphasis on how they handle noise in the experimental signal. Qualitative and quantitative issues for linear prediction will be covered in detail. The final section covers all recent applications of linear prediction in the analysis of NMR spectra.

The distinction between numerical analysis and engineering has become less well defined in recent years, and this statement can even be considered an euphemism in the case of signal processing. In most papers dealing with linear prediction, you are likely to see equations involving matrices covering most of the printed space. This review is designed to be comprehensive in describing the new procedures that are now applied to linear prediction, therefore will not elude the technical and mathematical aspects; however attempts are made to provide a ''practical'' explanation to the problems as much as possible. I do hope that after reading this review, the basic concepts behind the terms ''linear prediction'' will be clearer. To help the reader going through the mathematical formalisms, a small glossary of the main methods borrowed from applied mathematics is provided at the end of the manuscript.

2. The concept of linear prediction

2.1. The need for alternative solutions to Fourier transform

Classical treatment of NMR signals proceeds through a discrete Fourier transform (DFT) of the data points $\{a_k\}$ of a free induction decay (FID), sampled at constant time interval Δ:

$$A_i = \sum_{k=0}^{N-1} a_k \exp\left(-\frac{j2\pi ik}{N}\right) \quad (1)$$

where j is the complex number whose square is -1, and N the total number of points. $\{A_i\}$ is a discrete representation of the spectrum of the time domain signal $\{a_k\}$, calculated at all frequencies:

$$f_i = \frac{i}{N\Delta}, \qquad i = -\frac{N}{2}, \ldots, \frac{N}{2} - 1 \quad (2)$$

for a complex signal.

Despite its popularity, it is important to bear in mind that the application of DFT imposes certain restrictions. Firstly, all the elements of the signal must have been measured at fixed time intervals Δ, and included in the calculation. This can be detrimental in certain conditions: should the first data points be highly corrupted by noise, the spectrum will be accordingly affected, usually showing severe baseline problems. Ignoring these data points would usually not solve this problem, and in fact would induce severe phase problems. A similar problem exists at the other end of the time series. Secondly, the derived spectrum depends on N, the total number of data points collected (see Eq. (1)): for example, the digital resolution in the spectrum is inversely proportional to N. Thirdly, if the data vector is non-zero at the end of the acquisition, its Fourier transform exhibits artifacts (i.e. convolution by the Fourier transform of the Boxcar function, defined by $B(n\Delta) = 1$ for $0 \leq n < N$ and 0 otherwise). While this is usually not a problem in 1D NMR experiments, in ND experiments the number of data points in the reconstructed dimensions is usually kept as low as possible in order to reduce the length of the experiment; special care for the corresponding truncated FID is then needed. Eq. (1) also makes no assumption on the model that best describes the data: this can be seen as an advantage, as it makes DFT universal in its applications, but also as a weakness, as it does not allow inclusion of prior information on the signal in the spectral analysis. Eq. (1) is applied to the FID as is, which can be seen as a combination of signal and noise (this combination is usually assumed to be a simple sum), and DFT does not provide a tool to distinguish these two components. All these difficulties have set up the grounds for the search of alternate spectral analysis methods, Linear Prediction being one of them. I would not want to conclude this section by leaving the impression that Fourier transform should be completely replaced: though it has its limitations, some of them exposed earlier, DFT is still a very valuable technique, routinely used by all NMR spectroscopists. Interestingly, DFT even proves useful as a

numerical technique to considerably speed up the methods developed to replace it (an example is provided in the Appendix A).

2.2. A simple example

Let us consider a complex causal signal $S(t)$ containing a single damped exponential:

$$S(t) = A\mathrm{e}^{-Rt}\mathrm{e}^{-j2\pi f_0 t} \tag{3}$$

where A is a complex amplitude, R the damping coefficient, and f_0 the frequency of the sinusoid.

Sampling $S(t)$ at constant time interval Δ yields the time series $\{S_k\}$:

$$S_k = S(k\Delta) = A\mathrm{e}^{-Rk\Delta}\mathrm{e}^{-\mathrm{j}2\pi f_0 k\Delta} = AZ^k \tag{4}$$

where Z is a constant with respect to k, and equal to:

$$Z = \mathrm{e}^{-(R+\mathrm{j}2\pi f_0)\Delta} \tag{5}$$

Using basic properties of exponential functions, we directly derive from Eq. (4) that:

$$S_k = ZS_{k-1} \qquad ,\forall k > 0 \tag{6}$$

i.e. the data points in the time series are linearly related. In this simple case, knowledge of two consecutive data points S_i and S_{i+1} is enough to fully characterize the time series, by first computing $Z = S_{i+1}/S_i$. This has several major consequences.

1. All data points preceding S_i or following S_{i+1} can be derived by successive applications of Eq. (6). This is the concept of backward and forward linear prediction, respectively.
2. If N data points have been collected, the procedure aforementioned would make it possible to build a prolonged signal with much more data points than N. This would be useful if N is small and the time series is meant to be analyzed by Fourier transform, in which case artifacts related to truncation could be removed.
3. In fact, there is no need to prolong the time series S as its spectrum can be directly derived based solely on Eq. (6).The Fourier transform $X(f)$ of the time series S is defined by:

$$X(f) = \sum_{k=0}^{+\infty} S_k \exp\left[-\mathrm{j}2\pi f k\Delta\right] \tag{7}$$

By summing Eq. (6) multiplied by $\exp\left[-\mathrm{j}2\pi f k\Delta\right]$ for all $k > 0$, we get:

$$\sum_{k=1}^{+\infty} S_k \exp\left[-\mathrm{j}2\pi f k\Delta\right] = Z\sum_{k=1}^{+\infty} S_{k-1}\exp\left[-\mathrm{j}2\pi f k\Delta\right]. \tag{8}$$

Substituting Eq. (5) in Eq. (6), we obtain, after rearrangement:

$$X(f) - S_0 = Z\exp\left[-\mathrm{j}2\pi f\Delta\right]X(f). \tag{9}$$

Thus the Fourier transform of S is:

$$X(f) = \frac{A}{1 - Z\exp\left[-\mathrm{j}2\pi f\Delta\right]} \tag{10}$$

4. If finally we use the fact that the time series S is a discrete representation of the model function given by Eq. (3), the spectrum itself becomes useless: the damping coefficient R and the frequency f_0 can be directly derived from the module and phase of Z:

$$R = -\frac{1}{\Delta} ln|Z| \tag{11}$$

and

$$f_0 = -\frac{1}{2\pi\Delta}\arg(Z) \qquad (2\pi) \tag{12}$$

and the amplitude A is subsequently derived from Eq. (4), yielding all elements of the signal (the ambiguity in the definition of the frequency is a consequence of the discretization of the signal; the equivalent in the Fourier analysis is the periodicity of the continuous or discrete spectrum).

2.3. Linear prediction: autoregression, or functional modeling?

The relevant point to be deduced from the simple example described earlier is that if the time series under study satisfies a relationship similar to Eq. (6), then, even if it is short, it may contain enough information to allow a correct spectral analysis. The problem is to recover this information. When the only knowledge we rely on is Eq. (6), this leads to the concept of autoregressive modeling (AR) [70]. If in addition the underlying model for the time series is known (such as Eq.(3) for the example given earlier), then the problem becomes deterministic, in which

case more restrictive but more robust techniques for extracting information on the signal exist; these techniques belong to functional modeling. Linear prediction is in fact a generic term which describes both approaches [67,70].

Autoregressive time series have properties which greatly facilitate their analyses; in particular, they can be extrapolated in the ''future'' or in the ''past'' (i.e. after the last data point, or before the first data point, respectively). Such properties are potentially very useful for the analysis of NMR signals. In the case of high-dimensional NMR for example, experimental time considerations limit the number of acquired data points in the reconstructed dimensions; direct application of Fourier transform on this data induces distortions, which can be reduced by apodization prior to DFT, but at the cost of loss of resolution.

In the following section, we will briefly define autoregressive processes, as well as the common methods to study them. We will then show that under certain assumptions, an NMR signal satisfies the requirements of AR processes, but its analysis requires specific techniques, which will be reviewed in detail.

Functional modeling is described in detail in Section 4.

3. Autoregressive methods

3.1. Autoregressive processes

Let us consider a discrete time series as a noisy representation of a deterministic phenomenon. The total acquisition lasts $N\Delta$, where N is the number of data points and Δ the constant time interval between two points. Assuming that this window of observation provides a quasi-complete image of this phenomenon, it is quite natural to express any data point a_m not measured as a function of M known experimental points:

$$a_m = f(a_{N-1},\ldots,a_{N-M}) \qquad \text{for } m > N-1 \tag{13}$$

where M is a parameter at this stage.

The simplest possible expression for the function $f(a_{N-1},\ldots,a_{N-M})$ is linear; we could then predict a_N based on the relationship:

$$a_N = \sum_{m=1}^{M} b_m a_{N-m} + e_N \tag{14}$$

where M is the order of the linear estimator, the b are coefficients, and e_N is noise. More generally, we will assume:

$$a_k = \sum_{m=1}^{M} b_m a_{k-m} + e_k \tag{15}$$

for all k (i.e. Eq. (15) applies on known data points as well as outside of the observation window); e_k is the k-th element of a zero-mean, white noise vector. Time series that comply with Eq. (15) are referred to as autoregressive processes [70]; sinusoids are good examples of underlying models for these types of time series.

3.2. Spectrum of an autoregressive time series

One of the major interests of AR analysis is that it allows direct calculation of the spectrum of the time series, bypassing the Fourier analysis. Following the method used in the simple example described earlier, we multiply Eq. (15) by $\exp[-j2\pi fk\Delta]$ and sum over all k:

$$\begin{aligned}\sum_{k=-\infty}^{+\infty} a_k \exp[-j2\pi fk\Delta] &= \sum_{k=-\infty}^{+\infty}\sum_{m=1}^{M} b_m a_{k-m}\exp[-j2\pi fk\Delta] \\ &\quad + \sum_{k=-\infty}^{+\infty} e_k \exp[-j2\pi fk\Delta]\end{aligned} \tag{16}$$

(the summation extends to all k, positive and negative, as this approach is not restricted to causal time series). By definition, the left side of Eq. (16) corresponds to $A(f)$, the Fourier transform of the time series $\{a_k\}$, while the last term of the right term is $E(f)$, the Fourier transform of the noise sequence $\{e_k\}$. After

rearrangement, we obtain:

$$A(f) = \sum_{m=1}^{M} b_m \exp[-j2\pi fm\Delta] \sum_{k=-\infty}^{+\infty} a_{k-m} \exp[-j2\pi f(k-m)\Delta] + E(f) \quad (17)$$

Eq. (17) can be rewritten as:

$$A(f) = \sum_{m=1}^{M} b_m \exp[-j2\pi fm\Delta] A(f) + E(f) \quad (18)$$

which yields, after setting $b_0 = -1$,

$$A(f) = -\frac{E(f)}{\sum_{m=0}^{M} b_m \exp[-j2\pi fm\Delta]} \quad (19)$$

As $\{e_k\}$ is supposed to be a white noise with zero mean, its Fourier transform $E(f)$ is a constant, in which case the Fourier transform of the time signal is directly derived from the knowledge of the prediction coefficients $\{b_m\}$.

A word of caution is in order here: Eq. (19) would directly provide the spectrum of the signal under study, were it not from the fact that the summation of the Fourier series in Eq. (16) extends from $-\infty$ to $+\infty$. Experimental signals are causal however, and finite in size. While this is not a problem when a power spectrum is sufficient, deriving phased spectra requires more precaution. Solutions to this problem have been described by Tang and Norris [71–73] in the form of the LPZ transform, and by Ni and Scheraga [74].

3.3. Computing the autoregressive coefficients: the autocorrelation methods

It is important to notice that in the AR master equation e_k is completely included in the model: Eq. (15) cannot be considered as the summation of signal and noise, as e_k depends on the previous sequence $\{e_0,\ldots,e_{N-1}\}$. A crude but reasonable statement would be that in AR, the existence of observation noise is neglected, and the white ''noise'' sequence e_k is a measure of the validity of the prediction model.

The method used to solve for the coefficients b_m is called regression analysis or alternatively autoregressive (AR) model, as the summation extends over the ''past'' of the time series. Two different approaches for solving for the b_m have been proposed [67]: the first is based on the autocorrelation coefficients of the data points, while the second directly analyzes the data points themselves.

If we assume that the time series to be analyzed is stationary with zero mean, estimates of the autocorrelation coefficients are given by:

$$\hat{R}_n = \frac{1}{N-n} \sum_{k=1}^{N-n-1} a_{k+n} a^*_k \quad (20)$$

where * stands for the complex conjugate.

$\hat{R}_n$ can also be seen as an estimate of the true autocorrelation coefficients R_n, which are equal to the expectation value $E[a_{k+n}a^*_k]$, or equivalently $E[a_k a^*_{k-n}]$. The latter can be obtained by multiplying Eq. (15) by a^*_{k-n}, and taking the expectation:

$$E[a_k a^*_{k-n}] = E\left[-\sum_{m=1}^{M} b_m a_{k-m} a^*_{k-n}\right] + E[e_k a^*_{k-n}] \quad (21)$$

As the data points a_{k-n} are uncorrelated with the white ''noise'' e_k for $n > 0$, we obtain:

$$R_n = -\sum_{m=1}^{M} b_m R_{n-m} \quad (22)$$

and, for $n = 0$,

$$R_0 = -\sum_{m=1}^{M} b_m R_{-m} + \sigma^2 \quad (23)$$

where $R_{-m} = R^*_m$ and $\sigma^2 = E[e_k e^*_k]$, i.e. the variance of the white ''noise'' sequence $\{e_k\}$. Combining Eq. (22) and Eq. (23), and setting $b_0 = 1$, we obtain [70]:

$$\sum_{m=0}^{M} b_m R_{n-m} = \begin{cases} \sigma^2 & n = 0 \\ 0 & n > 0. \end{cases} \quad (24)$$

Eq. (24) corresponds to the Yule–Walker equations, which show that the linear coefficients b_m can be derived from the autocorrelation coefficients.

The Yule–Walker equations can be rewritten in

matrix form as:

$$\begin{bmatrix} R_0 & R_{-1} & . & . & . & R_{-M} \\ R_1 & R_0 & . & . & . & R_{1-M} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ R_M & R_{M-1} & . & . & . & R_0 \end{bmatrix} \begin{bmatrix} 1 \\ b_1 \\ . \\ . \\ . \\ b_M \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} \tag{25}$$

where the matrix **R** on the left hand side is called the autocorrelation matrix, which has an hermitian Toeplitz structure (see Glossary). The unknowns in the system of Eq. (25) are the linear coefficients $\{b\}$ and σ^2 which can be retrieved either by classical matrix operations, or by using the fast Levinson and Durbin algorithm [75,76]. The latter takes in account the Toeplitz structure of the matrix.

The major assumption when solving the Yule–Walker equations is that the true autocorrelation coefficients R_n can be safely replaced by the estimates $\hat{R}_n$ given by Eq. (20). While for large N and small n this can be seen as reasonable (i.e. the sum will contain enough terms to average out the experimental noise), $\hat{R}_n$ will be a poor estimate for large n and more generally for small N, as well as for signals with poor signal-to-noise ratio and for non stationary signals. Several methods have been proposed to try to alleviate these problems.

1. Burg proposed a modification of the Levinson–Durbin algorithm with built-in construction of the estimates of the autocorrelation coefficients [77,78]. This method works directly on the data matrix, and makes no assumptions about elements outside the observation range, from 0 to $N-1$ (this method is sometimes referred to as a maximum entropy technique, though it is a true LP technique). Though this method is an improvement compared to solving directly the Yule–Walker equations, problems have been reported, such as line splitting (appearance of two lines where one is expected) [79], and frequency shifting [80,81].
2. For transient time series, Cadzow proposed to modify the number of terms included in the estimation of R_n based on its row position in the correlation matrix [82]; this has the drawback of destroying the Toeplitz structure of this matrix, making the procedure inefficient.
3. Fedrigo et al. [83] proposed a modification of Eq. (20) in which the number of terms included in the summation becomes independent of n:

$$\hat{R}_n = \frac{1}{K} \sum_{k=0}^{K-1} a_{k+n} a^*_k \tag{26}$$

When this modification is used, the Toeplitz structure of the matrix is maintained. This method, which is based on a modified correlation matrix, was shown on a few examples to perform better than a method working directly on the raw data, though it is not clear from their paper which method was used for the latter.

3.4. Noise-free NMR time series are autoregressive processes

The NMR signal is a transient process recorded as a finite time series $\{A_n\}$ which can be modeled by a function $\hat{A}_n$, consisting of K exponentially damped sinusoids characterized by signal frequencies f_k, amplitudes c_k, damping factors R_k and phase φ_k, defined by:

$$\hat{A}_n = \sum_{k=1}^{K} c_k \mathrm{e}^{\mathrm{j}\varphi_k} e^{(-R_k + j2\pi f_k)n\Delta} \tag{27}$$

where Δ is the constant sampling time interval. As a first step, noise is not considered. Any signal which can be described by Eq. (27) verifies the AR equation exactly, with an order parameter $M = K$. This property was introduced two centuries ago by the Baron de Prony [63], and was recently re-introduced in signal processing of NMR data by Kay and Marple [84]. A brief description of why Eq. (27) verifies the AR equations, and how it can be used to analyze NMR signal follows. Let us define:

$$Z_k = \mathrm{e}^{(-R_k - \mathrm{i}2\pi f_k)\Delta} \tag{28}$$

and P the complex polynomial function of order K, whose K roots are the Z_k:

$$P(z) = \prod_{k=1}^{K} (z - z_k) \tag{29}$$

After expansion, P can be rewritten as:

$$P(z) = z^K - \sum_{m=1}^{K} b_m z^{K-m} \tag{30}$$

where b_m are coefficients.

As by definition of P, $P(Z_k) = 0$ for all k, the following relation:

$$Z_k^K = \sum_{m=1}^{K} b_m Z_k^{K-m} \tag{31}$$

are true for all k in[1,K]. More generally,

$$Z_k^n = \sum_{m=1}^{K} b_m Z_k^{n-m}, \qquad n \geq K \tag{32}$$

Combining Eqs. (27), (28) and (32), we obtain:

$$\hat{A}_n = \sum_{k=1}^{K} c_k e^{j\varphi_k} \left(\sum_{m=1}^{K} b_m Z_k^{n-m} \right) \tag{33}$$

from which we derive:

$$\hat{A}_n = \sum_{m=1}^{K} b_m \hat{A}_{n-m} \tag{34}$$

for all n larger than K.

Eqs. (27)–(34) provide a justification for the fact that an NMR signal follows an AR model [Eq. (34) should be directly compared to Eq. (15)]. The main difference lies in the fact that there is no need to include the white noise sequence e_n, as the linear relation (34) is exact. It should be emphasized again at this point that $\{e_n\}$ in the AR model is a measure of the validity of the model, and does not account for measurement noise; the latter has not yet been included here. It is also interesting to notice that the polynomial introduced in Eq. (30) corresponds in fact to the denominator of the spectrum of the time series, as determined using the AR model (Eq. (19)). Though the $\{Z_k\}$ values are truly roots of the polynomial P, they can be seen as ''poles'' of the spectrum of the time series. For this reason, they are usually referred to as poles, and K, the number of sinusoids to be modeled, is sometimes referred to as the number of poles included in the calculation.

The theoretical time series described by Eq. (27) can be prolonged by an arbitrary number of points using Eq. (34): performing this extrapolation is defined as ''forward linear prediction''.

The same reasoning which prevails for Eqs. (28)–(34) can be repeated to define backward linear prediction, in which case:

$$\hat{A}_n = \sum_{m=1}^{K} c_m \hat{A}_{n+m} \tag{35}$$

In Eq. (35), $\{c\}$ are the backward linear predictions.

3.5. The experimental NMR signal is really an ARMA process

Under the assumption that the true NMR signals are exponentially damped sinusoids, i.e. that (26) describes the NMR free induction decay (FID) in the absence of noise, the data points of an experimental FID in the presence of noise can be modeled by

$$A_n = A(n\Delta) = \hat{A}_n + w_n, n = 0, 1 \ldots\ldots N-1 \tag{36}$$

where N is the total number of data points. The noise sequence w_n is supposed to be normally distributed, with zero mean and a standard deviation σ, and the real and imaginary part of w_n are supposed to be independent. After substitution of Eq. (36) into Eq. (34), we obtain:

$$A_n - w_n = \sum_{k=1}^{K} b_k (A_{n-k} - w_{n-k}) \tag{37}$$

hence:

$$A_n = \sum_{k=1}^{K} b_k A_{n-k} - \sum_{k=1}^{K} b_k w_{n-k} + w_n \tag{38}$$

A true least-squares solution vector $\{b\}$ of Eq. (38) would minimize the variance of w defined as

$$\sum_{k=1}^{N} w_k^* w_k \tag{39}$$

This would lead however to a set of nonlinear equations.

Based on Eq. (38), A_n is in fact an autoregressive-moving average process (ARMA), with identical autoregression and moving average parameters [70]. The estimation of ARMA parameters is complex, and remains an active area of research [85–93]. A widely used procedure is to exploit the fact that an ARMA

process can be expressed as an AR process of infinite order [70]. Eq. (38) is then rewritten as:

$$A_n = \sum_{k=1}^{P} b_k A_{n-k} + \varepsilon_n \tag{40}$$

in which the noise is now modeled by ϵ_n. The order of the sum has been modified from K, the true number of underlying sinusoids, to P, a parameter referred to as the order of the LP procedure, with $P > K$. It should be mentioned that K is usually not known, in which case increasing the LP order to P has the further advantage of ensuring the determination of all complex exponentials in the signal.

As a first approximation, it is therefore appropriate to analyze an NMR signal as an AR process. The first step of such a study is to find the coefficients of the linear relations between the data points. Under the ergodic condition these coefficients can be derived from the autocorrelation matrix **R**, using the Yule–Walker equations [see Eq. (25)]. Fast algorithms are usually applied to this problem such as the Levinson–Durbin procedure, which uses the special Toeplitz structure of the matrix **R**. NMR signals however are not stationary, and discarding this fact leads to poor results when applying AR (for a review see [41,67]). Solutions to this problem have been proposed, such as the Burg algorithm described earlier, which is directly applied on the data matrix [77,78]. It was shown however that even this procedure has problems, such as a line splitting when a single line is expected [79] and frequency shifting [80,81]. The most promising alternative techniques are least-squares methods, which make no assumptions on data points outside the window of observation, as well as on the ergodicity of the signal.

3.6. Computing the autoregressive coefficients: the least squares approach

3.6.1. Linear least-squares procedures

Let us consider a classical linear system of equation:

$$\mathbf{Ax} = \mathbf{b} + \mathbf{n} \tag{41}$$

where **n** is a noise sequence affecting the observation vector **b**, supposed to be white and of zero mean. A least-squares solution of Eq. (41) is obtained by minimizing the variance of **n**, or equivalently the difference squared between the model **Ax** and the observation vector **b**:

$$\boldsymbol{x}_{\mathrm{LSQ}} = \arg\min_x (\mathbf{n}^*\mathbf{n}) = \arg\min_x \left(\|\mathbf{Ax} - \mathbf{b}\|^2 \right) \tag{42}$$

where $\mathbf{n}^*$ is the hermitian transpose of **n**, and $\|\cdot\|$ the L2 norm [94].

This procedure can be directly applied to Eq. (40): under the assumption that $\boldsymbol{\epsilon}$ is a white noise process, with zero mean, a minimum variance estimate of **b**, $\mathbf{b}_{\mathrm{LSQ}}$, is obtained by minimizing $\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}$, based on Eq. (42). Several methods have been proposed to solve this problem, and we will review some of them. Let us first rewrite Eq. (40) in matrix form as:

$$\begin{bmatrix} A_{P-1} & A_{P-2} & . & . & A_0 & \\ A_P & A_{P-1} & . & . & A_1 & \\ . & . & . & . & . & \\ . & . & . & . & . & \\ A_{N-2} & A_{N-3} & . & . & A_{N-1} & -P \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ b_P \end{bmatrix}$$

$$= \begin{bmatrix} A_P \\ A_{P+1} \\ . \\ . \\ A_{N-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_P \\ \varepsilon_{P+1} \\ . \\ . \\ \varepsilon_{N-1} \end{bmatrix} \tag{43}$$

where N is the total number of data points in the NMR signal, and P the number of poles; we have seen earlier that P should be as large as possible. From Eq. (41) however, we also observe that the largest possible value for P is $N/2$.

Notice that in Eq. (43), the prediction remains on the available data, i.e. for example Eq.(40) is not applied for $n < P$. This is different from the autoregressive methods based on the correlation matrix described, which use all possible equations, filling up with zeros when data points are ''missing''. If the larger system which includes these incomplete equations had be considered instead, its normal equations would have led back to the Yule–Walker equations, with all the problems mentioned earlier. The reduced system of equations described by Eq. (43) is known as the covariance method of linear prediction.

Eq. (43) can be restated in compact form

$$\mathbf{Tb} = \mathbf{a} + \varepsilon \tag{44}$$

where **T** and **a** are a matrix and a vector containing signal data points, respectively.

T is highly structured, in the sense that all elements on any sub-diagonal of **T** are equal: **T** is a Toeplitz matrix (see Glossary). It should be noted that by simple manipulation, Eq. (43) can be rewritten as:

$$\begin{bmatrix} A_0 & A_1 & . & . & A_{P-1} \\ A_1 & A_2 & . & . & A_P \\ . & . & . & . & . \\ . & . & . & . & . \\ A_{N-1-P} & A_{N-P} & . & . & A_{N-2} \end{bmatrix} \begin{bmatrix} b_P \\ b_1 \\ . \\ . \\ b_1 \end{bmatrix} = \begin{bmatrix} A_P \\ A_{P+1} \\ . \\ . \\ A_{N-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_P \\ \varepsilon_{P+1} \\ . \\ . \\ \varepsilon_{N-1} \end{bmatrix} \tag{45}$$

or in compact form as:

$$\mathbf{Hb} = \mathbf{a} + \varepsilon \tag{46}$$

where **H** is a matrix in which all elements in a given anti-diagonal are equal: **H** is a Hankel matrix. Both forms of the system have been used in NMR. As Toeplitz and Hankel matrices have very similar properties, we will focus only on the Toeplitz representation.

The dimension of **T** is $(N - P) \times P$, hence the direct inverse of **T** cannot be computed. The most common approach is to form the normal equations [94]:

$$\mathbf{T}^*\mathbf{Tb} \approx \mathbf{T}^*\mathbf{a} \tag{47}$$

where $\approx$ indicates that the noise is not considered. In the specific case of linear prediction in which the matrix T is Toeplitz, building the normal equations is fast. Thus, let us call $\mathbf{C} = \mathbf{T}^*\mathbf{T}$, then

$$C(i,j) = \sum_{k=1}^{N-P} E(k,i)^*E(k,j) \quad \text{for} \quad (i,j) \in [1,P]^2 \tag{48}$$

where

$$E(k,i) = A_{P+1+i-j} \tag{49}$$

from the definition of the signal matrix. **C** is hermitian; in addition, if we consider $\mathbf{C}(i+1, j+1)$, we obtain:

$$C(i+1,j+1) = \sum_{k=1}^{N-P} A_{P+1+k-i-1}{}^*A_{P+1+k-j-1} = \sum_{k=0}^{N-P-1} A_{P+1+k-i}{}^*A_{P+1+k-j} \tag{50}$$

hence

$$C(i+1,j+1) = C(i,j) + A_{P+1-j}{}^*A_{P+1-i} - A_{P+1+N-j}{}^*A_{P+1+N-i} \tag{51}$$

The whole correlation matrix **C** can then be constructed at low computing cost from the knowledge of its first column and successive use of Eq. (51). In addition, the first column of **C** is obtained as the product of the Toeplitz matrix $\mathbf{T}^*$ with the vector corresponding to the first column of **T**, and the computational cost for that product can be reduced from N^2 in general to $N\log(N)$, using FFT (see Appendix A). The only drawback of this procedure is that $\mathbf{C} = \mathbf{T}^*\mathbf{T}$ does not have a particular structure anymore.

The least squares solution of (42) is derived as:

$$\mathbf{b}_{\mathrm{LSQ}} = (\mathbf{T}^*\mathbf{T})^{-1}\mathbf{T}^*\mathbf{a} \tag{52}$$

Eq. (50) can be simplified by first taking the Cholesky decomposition of **T**:

$$\mathbf{T}^*\mathbf{T} = \mathbf{S}^*\mathbf{S} \tag{53}$$

where **S** is an upper triangular matrix. This method is usually preferred to a LU decomposition of the same matrix, as it is faster and allows direct identification of the pseudo-rank of the system. It has been applied to LP in NMR by Gesmar and Led [95].

Even though the normal equations can be formed fast, other numerically more stable approaches have been proposed, either based on a Householder QR decomposition of $\mathbf{T}(\mathbf{T} = \mathbf{QR}$, where **R** is upper triangular, and **Q** orthogonal), or by first calculating the pseudo-inverse of **T**, using its singular value decomposition (SVD). The two latter approaches are

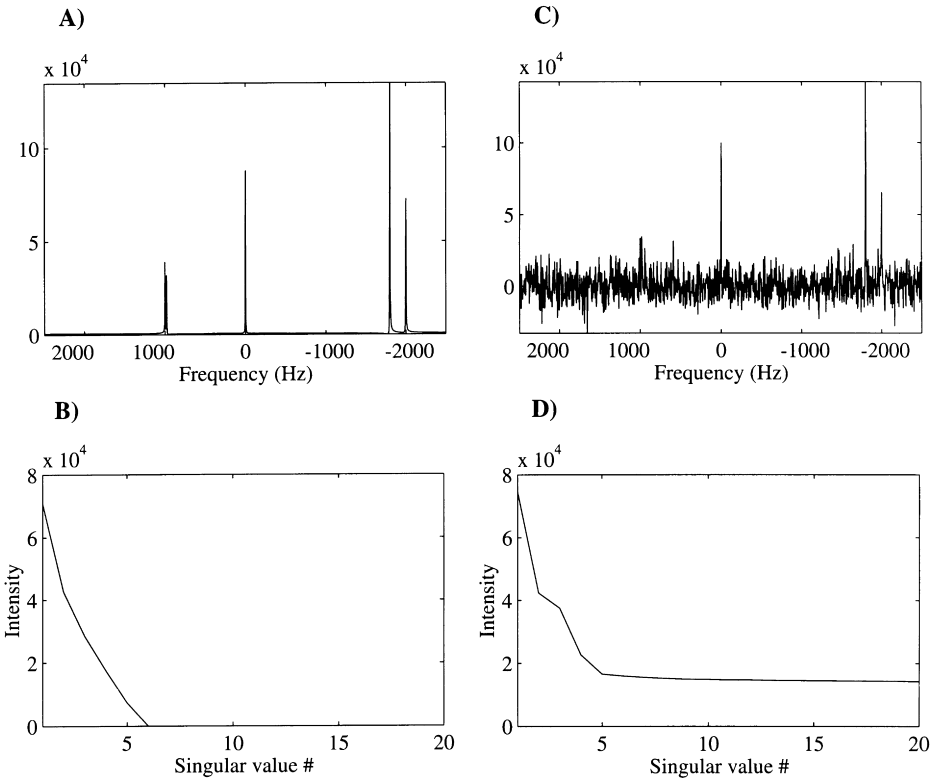

Fig. 1. Singular value decomposition is a rank-revealing procedure, which can be used to identify the number of components in a signal. To test this property, a synthetic signal was designed including 5 exponential functions, and computed over 1024 complex data points, with a dwell time of 200 μs (SW = 5000 Hz). The real part of the FT of this signal in the absence of noise is shown in (A) while (C) shows the same spectrum after corruption of the signal by a white noise sequence (SNR = 5 dB; the SNR is defined as 10 $\log(C_{\max}^2/\sigma^2)$, where $C_{\max}$ is the maximum amplitude in the signal, and σ the standard deviation of the noise sequence). The spectrum shows two peaks very close to each other, at 980 and 1000 Hz, respectively. The forward LP coefficients were computed using SVD, including all 1024 data points, and assuming 100 poles. The 20 first singular values for the signals corresponding to (A) and (C) are shown in (B) and (D), respectively. In the absence of noise, only 5 singular values are non-zero, corresponding to the 5 exponential functions of the signal (C). In the presence of noise however, all singular values are non-zero. Keeping all singular values before the plateau value would retain only 4 components in the signal.

referred to as LPQRD [96] and LPSVD [66], respectively, for obvious reasons.

3.6.2. Reducing the contribution of noise

There is no physical reason for ϵ in Eq. (40) to be a white noise vector with zero mean. In fact, we know that:

$$\varepsilon_n = -\sum_{k=1}^{P} b_k w_{n-k} + w_n \tag{54}$$

hence LSQ estimates of **b** based on Eq. (40) can be biased. A further complication occurs as K, the true number of sinusoids, is usually unknown. A first approximation to minimize both problems is to increase the number of forward prediction coefficients to P larger than K, with the hope that this will also increase the resolution in the frequency domain. Increasing the number of prediction coefficients has been shown to be useful without complete success in removing the bias (see for example [95,97]). Problems occur in the analysis of short NMR time series, in which case the value of P is limited, and in the analysis of FIDs with poor signal-to-noise ratios. In the latter case, there is the cumulative effect of the increase of observation noise, and the increase in the limitation of the model, which tries to fit an AR model to the noise in the signal. This can affect the

computation of the linear prediction coefficients in two ways: firstly, the data matrix can become numerically unstable, resulting in poor solutions for the linear system to be solved, and secondly the AR model itself might not apply anymore. Solutions to these problems would involve modification of the matrix itself to prevent numerical problems, and pre-treatment of the data in the signal themselves, in order to reduce the contribution of noise. This has led to the application for LP of two numerical procedures, namely regularization and preprocessing, which are presented in detail in the following.

3.6.2.1. Regularization We will briefly describe the concept of regularization in the case of SVD-based least-squares methods [98,99]. A general overview of regularization can be found in [100,101]; an illustration is provided in Fig. 1.

Let us go back to the problem of solving Eq. (46), which can be restated in terms of a χ^2:

$$\mathbf{b}_{\mathrm{LSQ}} = \min_b \chi^2 = \min_b \left(\|\mathbf{Tb} - \mathbf{a}\|^2 \right) \tag{55}$$

As we have seen before, several methods exist to solve the corresponding system; we will focus here on the SVD approach. The singular value decomposition of the data matrix **T** of size $(N - P)*P$ is defined as:

$$\mathbf{T} = \mathbf{U}\mathbf{S}^t\mathbf{V} \tag{56}$$

where **U** and **V** are orthogonal matrices containing the singular vectors, and **S** the diagonal matrix containing the singular values. In the absence of noise, the rank of **T** is exactly *K*, the total number of exponentials in the signal, and only *K* singular values are non-zero (an example is given in Fig. 1(b)). The least-squares solution of Eq. (56) is then defined as:

$$\mathbf{b}_{\mathrm{LSQ}} = \mathbf{V}\mathbf{S}'^t\mathbf{U}\mathbf{a} \tag{57}$$

where $\mathbf{S}'$ is the pseudo-inverse of **S**:

$$S(i,i) = \begin{cases} \dfrac{1}{S(i,i)}, & S(i,i) \neq 0 \\ 0, & S(i,i) = 0 \end{cases}. \tag{58}$$

In the presence of noise however, all singular values are usually greater than zero (Fig. 1(d)). Direct application of Eq. (58) can lead to a distorted solution: a small value of $S(i,i)$, which could probably be accounted for by noise, will yield an unrealistic large pseudo-inverse element, resulting in a least squares solution which is too large. This is used in fact to characterize the ''instability'' or ill-conditioning of the matrix, by defining its condition number as (see also the Glossary):

$$\mathrm{cond}(\mathbf{T}) = \frac{S(1,1)}{S(P,P)} \tag{59}$$

i.e. the ratio of the largest to the smallest singular value. The condition number is always larger than 1, and becomes infinite when the matrix is rank deficient. A small value for cond(**T**) indicates that **T** is full rank, and numerically stable, while large values for cond(**T**) indicates instability. In the latter case in fact, the matrix is probably not full rank, and it is only for numerical reasons that it appears as such.

If the true rank *K* of **T** is known in advance, a simple solution is to set $S(i,i) = 0$ for all *i* larger than *K*. This procedure is referred to as discrete regularization, and has been applied in the case of LPSVD to NMR signals [65]. It is worth noticing that when *K* is known, the regularization scheme described here is such that only the first *K* singular values of **T** are needed. If only a small number of singular values are to be computed, alternative schemes to classical SVD exist, such as the Lanczos algorithm which has been used for analyzing NMR signals [102,103]. The basic Lanczos algorithm [104,105] is known however to be very sensitive to round-off errors [103], and a numerically stable implementation requires great care. The method is still the subject of much research [106–109].

Small singular values correspond to noise. Inversely, the largest singular value usually corresponds to the predominant component in the signal, i.e. usually the solvent. This has led Zhu et al. [110] to propose a post-acquisition solvent processing technique, in which they first build the data matrix, calculate its SVD, set the first singular value to zero, and then reconstruct a signal which is analyzed by FT. It should be mentioned that such an application may lead to distortion in the amplitudes of the signal, as there is no reason to expect that the first singular value will refer to the water resonance only.

If the noise level is low and *K* is not known in advance, it can in principle be estimated from the spectrum of the singular values, as the singular values

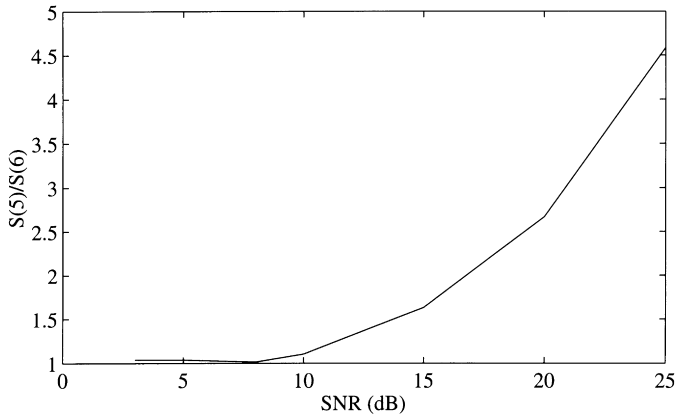

Fig. 2. SVD based LP prediction relies on the fact that the signal-related singular values can be distinguished from those only related to noise. The signal described in Fig. 1 contains only 5 components, hence a significant drop should appear between the values $S(5)$ and $S(6)$ of the 5th and 6th singular value of the corresponding data matrix. The ratio $S(5)/S(6)$ is plotted versus the SNR of the synthetic signal considered (see legend of Fig. 1 for details). For low SNR (< 10 dB), $S(5)$ and $S(6)$ are almost identical, and either the singular value spectrum is truncated after the 4th singular value, in which case one component is lost, or a few noise-related singular values have to be included in the computations.

related to noise will be much smaller than the singular values associated to signal. In most realistics case however, the spectrum of singular values will be continuous, and the cutoff value from which the effective rank can be derived is difficult to identify (see Fig. 2). A large condition number corresponds to an ill-conditioned matrix; the corresponding linear system has a solution with an unrealistically large norm. Intuitively, a reasonable solution to that problem would be to modify Eq. (55) such as to retain the minimization of χ^2 and at the same time to maintain the norm of the solution vector minimal. This is the principle of the Tikhonov regularization technique [111] in which the modified equation is:

$$\mathbf{b}_{\text{reg}} = \min_b \left(\|\mathbf{Tb} - \mathbf{a}\|^2 + \lambda^2 \|\mathbf{b}\|^2 \right) \tag{61}$$

where λ is a Lagrange multiplier and $\mathbf{b}_{\text{reg}}$ is the

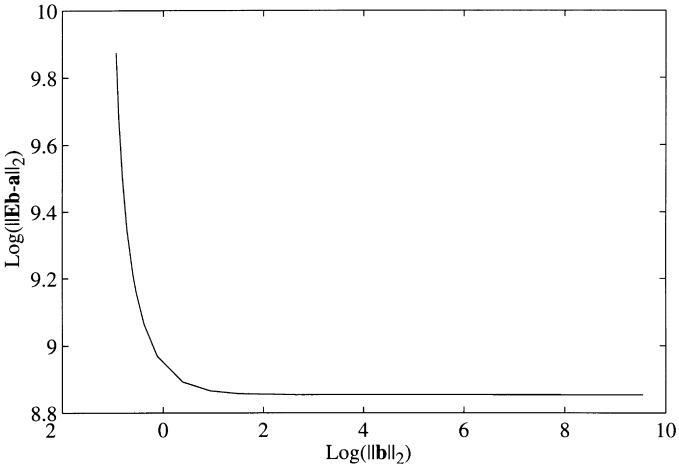

Fig. 3. The LP coefficients are obtained as solution of the linear system $\mathbf{Eb} = \mathbf{a}$, where $\mathbf{E}$ and $\mathbf{b}$ are the data matrix and observation vector, respectively, and b the vector of coefficients (see text for details). LP coefficients have been computed for the signal described in Fig. 1, for various noise sequences with SNR varying from 1 to 50 dB. The logarithm of the Euclidian norm of the residual $\mathbf{Eb} - \mathbf{a}$ is plotted versus the norm of the coefficient vector, displaying a characteristic L-shape curve.

regularized solution vector. The solution to Eq. (61) is given by:

$$\mathbf{b}_{\text{reg}} = \mathbf{V}\mathbf{S}''\mathbf{U}^*\mathbf{a} \tag{62}$$

where the regularized pseudo-inverse S'' is defined by:

$$S''(i,i) = \frac{S(i,i)}{S(i,i)^2 + \lambda^2}. \tag{63}$$

In its application to linear prediction of NMR signals, this technique has been referred to as continuous regularization [112], as it can be regarded as a continuous truncation of the singular values. For small values of λ only small singular values will be affected, while for large values of λ, most singular values will be affected. Choosing λ is a crucial step in this procedure and different strategies for estimating its optimal value have been proposed: for review, see [100]. One method is based on the L-curve described in Fig. 3, which identifies λ_{opt} as the parameter corresponding to the ''corner'' of this curve [113]. The discrepancy method is an alternative approach which defines λ_{opt} as the largest value λ for which $\|\mathbf{a} - \mathbf{T}\mathbf{b}_l\|^2 = \sigma^2$, where σ^2 is the variance of the noise which contaminates the signal (σ^2 can be estimated from a region of the experimental data which does not contain signal but noise only). The latter method has been applied to NMR data [112,114–116].

Regularization basically works by modifying the least-squares equation to be solved such as by restricting the modifications of the unknown vectors to reasonable values. Besides the Tikhonov regularization scheme, several methods have been proposed, which are beyond the scope of this review. It is worth mentioning that there is a maximum entropy regularization scheme [100,101], in which Eq. (55) is replaced by searching the minimum of:

$$f(\mathbf{b}) = \|\mathbf{T}\mathbf{b} - \mathbf{a}\|^2 + \lambda \sum_i b_i \ln(\omega_i b_i) \tag{64}$$

where ω_i are weight factors. The second nonlinear term can be seen as an entropy term, hence the name of this regularization technique.

3.6.2.2. Preprocessing Signal processing tries to extract as much information as possible from the experimental signal: the quality of the results will inherently depend on the quality of the data. As described earlier on the specific case of getting the linear prediction coefficients, one of the main aspects of signal processing is to find ways of dealing with the noise that pollutes the signal. The experiment itself is optimized to reduce the noise level. In addition, the noise has certain characteristics which can be used to distinguish it from the true signal. In the presence of white noise (i.e. independent random process with zero mean and constant variance) for example, least-squares methods can be readily used (even if the noise is not purely white, least-squares methods are a reasonable approximation). Separation of noise and signal can sometimes be derived in the process of analyzing the signal: as described earlier in the case of SVD, the first K singular values, where K is the true rank of the system, contains information about the signal, while the remaining singular values only relate to noise. This information can even be used prior to any analytical analysis of the signal, in procedures referred to as preprocessing.

Preprocessing is quite common in NMR. A well known example is the elimination of the water resonance in high resolution liquid NMR: if the spectrum is centered on the water resonance, and there are no signals of interest in a region adjacent to the peak from water, the latter can be reduced drastically using a low frequency digital filter [117–122].

Preprocessing can also be applied to reduce the noise prior to spectral analysis: a promising technique in that respect is the Cadzow regularization scheme [123]. A signal containing exactly K components will yield a data matrix $\mathbf{T}$ of rank K, in the absence of noise. In the presence of noise however, the same matrix will have a much larger numerical rank. The idea is then to compute a modified data matrix $\mathbf{T}_K$ of rank K, such that:

$$\|\mathbf{T} - \mathbf{T}_K\| \text{ is minimum.} \tag{65}$$

As mentioned before, SVD is one possible solution to that problem, based on the observation that small singular values computed from the signal matrix $\mathbf{T}$ only contain noise. The methodology is simple [114,115,124]: build the data matrix $\mathbf{T}$ of size $(N - P) \times P$ and perform a singular value decomposition on $\mathbf{T}$. Assuming that the rank K is known, all singular values from $K + 1$ to P are set to zero, as well as the corresponding singular vectors in $\mathbf{U}$ and $\mathbf{V}$, and a

''corrected'' data matrix is computed accordingly. The original data matrix has a special structure (Hankel or Toeplitz, depending on how the elements are organized) while this ''processed'' matrix has probably lost this structure and a correcting scheme is required: usually diagonal (for Toeplitz matrix) or anti-diagonal elements (for Hankel matrices) are averaged. A ''processed'' signal is derived from this matrix, and the procedure is iterated until the ''processed'' signal does not statistically differ from the original signal (i.e. $\|\mathbf{A} - \mathbf{A}_p\|^2 \leq \sigma^2$, where $\mathbf{A}$ is the original signal, $\mathbf{A}_p$ the processed signal, and σ^2 the variance of the noise contained in $\mathbf{A}$). A possible drawback of this method is that it requires knowledge of the true rank of the system. Automatic procedures have been designed to circumvent this problem [114,116].

SVD-based preprocessing is sensitive to the size $(N - P) \times P$ of the signal matrix with respect to K, its true rank, and this effect has been shown to be more important during the first iterations of the procedure [125]. This has led to the design of an efficient preprocessing technique, referred to as IEP (iterative enhancement procedure) [125], in which the number of column P of the data matrix is set to $K + 1$ during the first iteration. This number P is then increased to $2K$ during the second iteration, and to $N/2$ during all subsequent iterations, where N is the total number of data points in the signal.

The procedure can be further improved by setting the ''processed'' signal to be a mixture of the true and reconstructed signals [116]: the first points of the FID usually contain mainly signal, with a good signal-to-noise ratio, and are consequently maintained as such in the ''processed'' signal, while the data points at the end of the FID contain more noise, and are therefore replaced by ''reconstructed'' data points.

3.7. Total least-squares approaches

The least-squares problem [94] can be schematically written as follows:

(Model variables) $*$ (parameter) = (observations) + (error)

where (error) is considered to be Gaussian white noise on the observations. Eq. (46) is only an approximation of this scheme, as both $\mathbf{a}$ and $\mathbf{T}$ are constructed from the noise-corrupted data. Total least squares (TLS) provides a means of handling the latter situation. This procedure can be briefly described as follows: let us go back to the original Eq. (43) for forward linear prediction, which can be rewritten in matrix form as:

$$\mathbf{a} - \mathbf{w} = (\mathbf{T} - \mathbf{D})\mathbf{b} \tag{66}$$

$\mathbf{w}$ and $\mathbf{D}$ are the perturbation or error vector and matrix, respectively, such that $w(i) = w_i$ and $D(i,j) = w_{i+j}$, and $\mathbf{a}$ and $\mathbf{T}$ are the signal vector and matrix defined earlier.

Eq. (66) can be rewritten in augmented form as:

$$([\mathbf{a}, \mathbf{T}] - [\mathbf{w}, \mathbf{D}])\begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} = 0 \tag{67}$$

where $[\mathbf{a},\mathbf{T}]$ has dimension $(N - P) \times (P + 1)$ (with P the number of forward coefficients). For simplicity, Eq. (67) can be rewritten as

$$(\mathbf{C} + \mathbf{F})\mathbf{y} = 0 \tag{68}$$

where $\mathbf{C} = [\mathbf{a},\mathbf{T}]$ contains the signal, $\mathbf{F} = [\mathbf{w},\mathbf{D}]$ is the perturbation matrix and $\mathbf{y} = [\ -\ 1/\mathbf{b}]$ contains the coefficients. Solving the TLS problem amounts to finding a perturbation matrix $\mathbf{F}$ such that $\mathbf{C} + \mathbf{F}$ is rank deficient. If a minimum norm $\mathbf{F}$ can be determined, any $\mathbf{y}$ vector satisfying Eq. (68) is a solution. The TLS solution for y is chosen to be of minimum norm, from which b_{TLS}, the TLS estimate of $\mathbf{b}$, is derived. A complete description of TLS is given by Golub and Van Loan [105], and it was first applied to linear prediction on NMR signals by Tirendi and Martin [126]. Though this procedure significantly reduces the bias on the parameters estimated by LP, it was shown to still provide biased estimates of the true underlying parameters [97,127].

TLS is still the subject of much research. Recent developments relevant to LP include the following:

1. Extraction of the rank of the augmented signal matrix using TLS [128]. This is of potential interest in the case considered here, as it could lead to a knowledge of K, the true number of sinusoids.
2. Modification of the TLS algorithm in order to maintain the special structure of the data matrix [129].
3. Application of TLS for regularization [130].

3.8. Fast linear prediction techniques.

All the numerical methods described before (Cholesky, QR, SVD) for obtaining the linear prediction coefficients do not use the special structure of the matrix, and are computationally demanding. If M is the number of equations, and P the LP order, computation of the normal equations requires $MP + N^2/2$ multiplications, using the Toeplitz structure of the data matrix, and the Cholesky decomposition itself requires $P^3/3$ multiplications. The computational effort for QR and SVD is in the order MP^2. As M is larger than P, it results that all three methods have a complexity at least equals to P^3. Their demand for computer memory also represents a serious problem.

In the section dealing with autoregressive models, it was shown that fast algorithms which exploit the Toeplitz structure of the matrix exist. The ones we have described so far require however specific properties of the data matrix: the Durbin–Levinson algorithm only works on squared matrices and relies on the knowledge of good estimates of the autocorrelation function while the Burg algorithm is appropriate for stationary signals only. Toeplitz matrices appear in a large variety of problems, in which these conditions usually do not apply. There has been considerable interest in the numerical analysis community in looking at this type of matrix, resulting in the development of a significant number of ''new'' fast algorithms (see for examples [131–136]). One of these methods originally proposed by Cybenko [132] has been applied to linear prediction of NMR signals by Gesmar and Hansen [137]. In brief, the method computes an ''inverse orthogonal'' factorization of the Toeplitz matrix $\mathbf{T}$ of size $M \times P$, in the sense that instead of computing $\mathbf{T} = \mathbf{QR}$, with $\mathbf{Q}$ having orthonormal columns and $\mathbf{R}$ upper trianglular, it computes two matrices $\mathbf{Q}$ and $\mathbf{R}$ of size $M \times P$ and $P \times P$, respectively, for which

$$\mathbf{TR} = \mathbf{Q} \tag{69}$$

There is a direct implication of this for solving least squares problems, in that instead of solving a triangular system, this method requires only that a triangular matrix is multiplied by a vector, which is less prone to numerical instability. In Cybenko's algorithm, the matrices $\mathbf{Q}$ and $\mathbf{R}$ are updated recursively, until the desired prediction order L is reached. Based on Eq. (69), the solution of a linear system $\mathbf{Ta} = \mathbf{b}$ is found as:

$$\mathbf{a} = \mathbf{RQ}^*\mathbf{b} \tag{70}$$

Gesmar and Hansen [137] have shown that the matrix multiplications required in Eq. (70) can be avoided in the special case of linear prediction. In LP problems, the right hand side vector which we rewrite as t_1 is such that the augmented matrix $[\mathbf{t}_1\mathbf{T}]$ of size $M \times (P + 1)$ retains a Toeplitz structure. The ''inverse factorization'' of $[\mathbf{bT}]$ gives:

$$[\mathbf{t}_1\mathbf{T}][\mathbf{R}'\mathbf{r}] = [\mathbf{Q}'\mathbf{q}] \tag{71}$$

where $\mathbf{R}'$ and $\mathbf{Q}'$ differ from $\mathbf{Q}$ and $\mathbf{R}$ in Eq. (69). The part of Eq. (71) involving r and q can be expanded as:

$$r_1\mathbf{t}_1 + \sum_{j=2}^{P+1} r_j\mathbf{t}_j = \mathbf{q} \tag{72}$$

where $\mathbf{T} = (\mathbf{t}_2,\ldots,\mathbf{t}_{P+1})$, and reorganized as:

$$\mathbf{t}_{P+1} = -\sum_{i=1}^{P} \frac{\mathbf{r}_i}{\mathbf{r}_{P+1}}\mathbf{t}_i + \frac{1}{r_{P+1}}\mathbf{q} \tag{73}$$

As $[\mathbf{Q}'\mathbf{q}]$ is orthonormal, $\mathbf{q}$ is orthogonal to all P columns of $\mathbf{Q}'$. We also know that $\mathbf{Q}'$ and the first P column of $[\mathbf{t}_1\mathbf{T}]$ spans the same vector space, therefore $\mathbf{q}$ is orthogonal to all vectors $(\mathbf{t}_1,\ldots,\mathbf{t}_P)$. From this, and from Eq. (73), we derive that $\mathbf{q}/\mathbf{r}_{1+P}$ is the minimum distance between the vector $\mathbf{t}_{1+P}$ and the hyperspace span by $(\mathbf{t}_1,\ldots,\mathbf{t}_P)$, consequently the vector

$$\mathbf{c}_{min} = \left(-\frac{r_1}{r_{1+P}}, -\frac{r_2}{r_{1+P}}, \ldots, -\frac{r_P}{r_{1+P}}\right) \tag{74}$$

is the minimum norm solution to the least square problem $(\mathbf{t}_1,\ldots,\mathbf{t}_P)\mathbf{c} = \mathbf{t}_{P+1}$, hence corresponds to the forward linear prediction.

The method presented earlier is fast (i.e. requires approximately $9MP + (27/2)P^2$ operations), and requires a small amount of computer memory (approximately $5(M + P)$): both are true advantages for any type of high-resolution LP calculation, in which case the LP order L may be in the order of 1000 or more. Based on the approach presented earlier, Gesmar and Hansen have shown that it is even feasible to determine more than 10 000 coefficients [137].

Among the other fast procedures available to solve a Toeplitz linear system, it is worth mentioning the

direct QR factorization procedure developed by Nagy, which can be modified to handle Tikhonov regularization with very little computing overhead [138]. We have recently adapted this method to linear prediction and its application to NMR spectroscopy (P. Koehl, manuscript in preparation).

3.9. Conclusions on AR methods

If a time series has the convenient property of belonging to the class of functions that obey Eq. (13), then AR offers a good alternative to Fourier transform. AR solves some of the limitations of FT: for example, AR can work on short time series, which are difficult to analyze by means of FT because of the truncation.

Without consideration of noise, NMR signals belong to these classes of functions while in the presence of noise, AR is only a working approximation. Applications of AR to NMR requires then some precaution.

1. A strict AR model introduces the notion of noise related to the applicability of the model itself, and does not handle the observation noise. We have seen that an experimental NMR signal should be considered as an ARMA process instead, or as an AR process of infinite order. In practice, what is usually proposed is to increase the order of the model (M in Eq. (13)), but there is a limitation to this increase, because of the finite size of the data set; this method will also fail for low signal-to-noise ratios. Further attempts to reduce the effect of this approximation involve pre-processing of the signal, using techniques originally described by Cadzow [123]. Noise in the signal can also lead to ill-conditioned linear systems. Tikhonov Regularization [111] is one solution to this problem and has been described in great detail in the case of SVD. Total least-squares [105] is an alternative to least squares which directly compensates for the approximation that the noise sequence is uncorrelated.
2. Numerical techniques specific to AR (such as those described in Section 3.2) are usually inadequate in NMR because of the transient nature of the signals. They have been replaced by linear least squares procedures, such as Cholesky decomposition, QR factorization or SVD. The experimental noise and the size of the system to be solved are two concerns which have been studied in detail.
3. The order of the autoregressive procedure becomes a complete part of the model, and finding its correct value is a crucial issue, for which no universal solution exists. As this problem is important also in model-based quantitative LP methods, it will be discussed later.

It should be mentioned that linear prediction techniques have been used to study a signal with a high number of components, inducing systems of very high order. Solving such systems by computer programs is not easy, because of memory and computer time requirements. These systems have special structures however (usually Toeplitz or Hankel), and this has led to the introduction of fast LP techniques [137].

Within these limitations, AR has proven a useful alternative to zero-filling and a powerful technique for completing or replacing the first data points of a signal. This will be discussed in detail in the Application section.

AR is a general technique, which does not take into account the underlying mathematical model describing the signal. By including this information, it is possible to extract even more information from the time domain signal, to the point that the spectral analysis is no longer required. This is discussed in the next section.

4. Model-based linear prediction methods

4.1. The extended Prony method

As described in Eqs. (27) and (36), the NMR signal is a transient process recorded as a finite time series $\{A_n\}$ which can be modeled by a sum of K exponentially damped sinusoids characterized by signal frequencies f_k, amplitudes c_k, damping factors R_k and phase φ_k defined by:

$$A_n = \sum_{k=1}^{K} c_k e^{j\varphi_k} e^{(-R_k + j2\pi f_k)n\Delta} + w_n \tag{75}$$

where Δ is the constant sampling time interval and $\{w\}$ is a noise sequence, supposed to be white.

The most common approach to derive all elements of Eq. (75) is Fourier analysis: the DFT of the signal

described by Eq. (75) is given by:

$$F_i = \sum_{k=1}^{K} \frac{c_k e^{j\varphi_k}}{R_k + j2\pi(f - f_k)} + W_i \quad (76)$$

where $\{W\}$ is the DFT of the noise sequence $\{w\}$; $\{W\}$ is a white noise process in the spectral domain [41]. The first term on the right side of Eq. (76) is a sum of complex Lorentzian functions. In the simple case in which all φ_k are zero, the real part of the spectrum described by Eq. (76) will appear as the superposition of pure absorption Lorentzians, centered at the characteristic frequencies f_k, of amplitude c_k/R_k, linewidth at half height R_k and surface c_k. In the more general case in which the phase parameters are non-zero, the spectrum is first ''phased'' prior to any quantitative analysis, i.e. a correction factor of the form $\exp(-j\phi_i)$ is applied to each spectral point F_I, which restores the pure absorption appearance for the real part of the spectrum. A general scheme for extracting information such as intensities from NMR signals has been accordingly defined, which includes apodization of the FID (to remove artifacts because of finite size samples as well as to improve the SNR), spectral analysis through Fourier transformation, phasing and baseline correction of the frequency-domain spectra (baseline offset in the spectral domain usually results from corruption in the first data point of the FID), and quantification through peak-picking and numerical integration. Limitations of this procedure appear however for signals with low signal-to-noise ratio, as well as in crowded or complicated spectra which show severe overlaps. In these cases, we have to look for alternative solutions.

Direct fitting of Eq. (75) to noisy experimental data leads to a difficult nonlinear least-squares problem, which requires that K be known, as well as good initial values for all spectral parameters [41]. Methods to solve this nonlinear problem have been mainly implemented as a final refinement procedure, after initial guesses for all or some of the parameters have been derived using another technique (which could be FT and visual inspection of the spectrum, when all peaks can be clearly identified) [139–143]. Fitting in the spectral domain has the same problems.

Maximum entropy [46] is a non-parametric approach whose aim is to enhance the signal content in a reconstructed spectrum. As such, it does not directly provide quantitative information on the signal, though this is an area in development now [52]. It should be noted here that there have been some ambiguities in the literature about the use of the term ''maximum entropy''. The Burg algorithm is sometimes referred to as maximum entropy, though it is an AR procedure. There is also a maximum entropy regularization procedure, though to my knowledge, it has not been applied to linear prediction. The maximum entropy technique mentioned here performs a nonlinear least-squares fit of a model signal to the experimental spectrum, under the constraint that the ''entropy'' of the signal should be minimal. Daniell and Hore provide a nice explanation of the significance of this entropy term [49].

Linear prediction in its model-based definition has attracted considerable interest as a quantitative alternative to FT analysis. It includes the following three steps:

1. Firstly, the NMR signal is considered to be an AR process, and the linear coefficients $\{b\}$ are computed, using one of the techniques presented earlier.
2. Once the $\{b\}$ are known, the frequencies f_k and the relaxation rates R_k of the signal can be calculated from the complex roots Z_k of the characteristic polynomial defined by:

$$P(z) = \sum_{m=0}^{P} b_m z^{P-m} \quad (77)$$

   where P is the order chosen for the AR process, and $b_0 = 1$. Eq. (77) is directly derived from Eq. (30), while the relation between the roots and the frequencies and damping factors is described by Eq. (28).
3. Finally, the complex amplitudes $c_k e^{j\varphi_k}$ are determined from Eq. (75) in the modified form

$$A_n = \sum_{k=1}^{P} c_k e^{j\varphi_k} z_k^n + w_n \quad (78)$$

   by a least squares calculation.

It is worth noting here that the sum in Eq. (75) has been extended from K, the true number of exponentials, to P, the presumed order of the AR process.

Procedures for deriving the AR coefficients have been described at length in the preceding sections.

We will therefore concentrate here on the root finding procedure required in step 2, as well as on the LSQ calculation of step 3.

4.2. Extracting the frequencies and decay rates

4.2.1. Roots of the characteristic polynomial

Considering that at this stage the linear coefficients $\{b\}$ have been computed, the next step is to calculate the roots of the polynomial given by Eq. (77), from which the frequencies and amplitudes of the signal components can be derived based on Eq. (78). Several methods exist to find the roots of a polynomial, including an algorithm proposed by Steiglitz and Dickinson [144] which has been used for LP by Barkhuisen et al. [65,66] and by Tang et al. [65,66,96] as well as the algorithm proposed by Svejgard [145] which is used by Gesmar and co-workers [41,95]. The latter procedure performs linearly with the order of the polynomial and is robust, making it a good option for high order LP calculations. Another approach [146] is to transform the root finding step into an eigenvalue problem. Any polynomial equation of the form:

$$P(x) = \sum_{i=0}^{N} a_i X^i = 0 \tag{79}$$

can be recast into the problem of finding the eigenvalues of the $(N + 1) \times (N + 1)$ matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & . & . & -a_0 \\ 1 & 0 & . & . & -a_1 \\ 0 & 1 & 0 & . & -a_2 \\ . & . & . & . & . \\ 0 & 0 & . & 1 & -a_N \end{bmatrix} \tag{80}$$

as $P(x)$ is the characteristic polynomial of $\mathbf{A}$, i.e. $P(x) = \det(\mathbf{A} - x\mathbf{I})$ where $\mathbf{I}$ is the identity matrix of size $(N + 1) \times (N + 1)$. It should be noticed however that $\mathbf{A}$ has no special structure, in which case the eigenvalue problem cannot be simplified. In the case of high polynomial order, it also suffers from high requirements in computer memory usage.

Polynomial rooting can also be avoided using the state–space models: this will be described later.

All procedures aforementioned rely on the fact that the linear prediction order P is much larger than the true number of exponential functions contained in the signal. Increasing P however results in a large number of extraneous roots for the polynomial. The problem of distinguishing these roots from true signal-related roots has been discussed by many authors, with major contributions from Kumaresan and Tufts [147,148]. Briefly, in the case of forward linear prediction, all roots fall within the unit circle in the complex plane. However, in the case of backward linear prediction, signal-related roots will fall outside the unit circle, while extraneous roots will fall both inside and outside the unit circle: all roots falling inside the circle can usually be discarded.

The distinction of signal roots from noise roots based on their relative position with respect to the unit circle in the complex plane becomes difficult for signals with low signal-to-noise ratios: even in the case of backward prediction, signal roots can move inside the unit circle, while noise roots move outside. Delsuc et al. [149] have tried to overcome this problem using a method suggested by Porat and Friedlander [150]. Basically the method relies on the fact that the same signal roots should be found in forward LP and in backward LP after reflection inside the unit circle, while noise roots should not be related to each other. The method calculates the pairwise distances between roots obtained from the two types of LP, and signal roots are defined as those for which the distance is minimum. An estimate of the exact number of signal roots is needed.

Zhu and Bax [151] proposed an alternative way to use the information contained in the backward and forward LP coefficients. As mentioned earlier, the roots derived from forward coefficients and those derived from backward prediction should be identical, after reflection within the unit circle of the latter. These sets of roots can be used to recalculate two sets of LP coefficients. In the presence of noise, these two new sets of coefficients will be different, and Zhu and Bax [151] proposed to reduce the contribution of the random errors induced by noise by simply averaging the two sets of coefficients. New roots are then computed based on the average coefficients.

NMR studies of macromolecules in solution generate signals with a high number of exponential functions. We have seen earlier that LP techniques require that the number of poles defined in the calculation be

much larger than the true number of peaks. The combination of these two conditions requires that very high degree polynomial equation be solved in order to get the characteristic frequencies and relaxation rates of the nuclei of the macromolecule under study (Gesmar and Hansen [137] have published a case with a polynomial order of 16 000). The large exponents in the polynomial require special care in order to avoid computing overflows. Fortunately, in the case of high resolution NMR, the roots of the polynomial fall very close to the unit circle in the complex plane. A stable strategy to solve high order polynomial equations of the form $P(z) = 0$ was described by Svejgard [145], in which roots are found iteratively by searching for minima of $|P(z)|$, directing the search away when an overflow occurs. When a root is found, the polynomial is deflated, and the procedure is repeated up to the last root. With the introduction of fast techniques to solve for the coefficients, solving for the roots of the characteristic equation has become the time limiting step in quantitative NMR.

4.2.2. Avoiding polynomial rooting: The state–space models

Barkhuijsen et al. [152] proposed an alternative method to directly extract the poles of the signal which makes use of the Hankel structure of the matrix as well as of the exponential nature of the signal. We will give a brief description of this method.

The NMR signal is modeled by a sum of damped exponential functions, as described by Eq. (26). This equation can be rewritten in matrix form:

$$\begin{bmatrix} A_1 \\ A_2 \\ . \\ . \\ A_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & . & . & 1 \\ Z_1 & Z_2 & . & . & Z_K \\ . & . & . & . & . \\ . & . & . & . & . \\ Z_1^{N-1} & Z_2^{N-1} & . & . & Z_K^{N-1} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ . \\ . \\ c_K \end{bmatrix} \tag{81}$$

where the Z_i are the poles, defined by Eq. (27), and c the amplitude. Eq. (69) is a good illustration that the problem of extracting the components of a signal can be broken down into first finding its poles, and then the complex amplitudes.

Eq. (81) can be applied to each column of the complete data matrix (in Hankel from) from which we derive:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ Z_1 & Z_2 & . & . & Z_P \\ . & . & . & . & . \\ . & . & . & . & . \\ Z_1^{N-P} & Z_2^{N-P} & . & . & Z_P^{N-P} \end{bmatrix} \times \begin{bmatrix} c_1 & c_1 Z_1 & . & . & c_1 Z_1^{P-1} \\ c_2 & c_2 Z_2 & . & . & c_2 Z_2^{P-1} \\ . & . & . & . & . \\ . & . & . & . & . \\ c_P & c_P Z_P & . & . & c_P Z_P^{P-1} \end{bmatrix} \tag{82}$$

or

$$\mathbf{H} = \mathbf{ZC} \tag{83}$$

H is the Hankel data matrix, **Z** is the matrix containing the pole, and **C** a mixed matrix containing both poles and amplitudes. The aim is to propose a method to extract **Z**, based on **H**, and then to extract the individual Z_i, based on the matrix **Z**.

Another decomposition of **H** can be obtained from its singular value decomposition:

$$\mathbf{H} = \mathbf{UDV}^* = \mathbf{UV}' \tag{84}$$

where $\mathbf{V}'$ combines **D** and $\mathbf{V}^*$.

Replacing Eq. (83) in (84), we obtain:

$$\mathbf{Z} = \mathbf{UVC}^{-1} = \mathbf{UQ} \tag{85}$$

where $\mathbf{Q} = \mathbf{V}'\mathbf{C}^{-1}$.

Z has the special form of a Vandermonde matrix. In that respect, it has the general property that any two subsets of the same size of its rows share the same subspace. For example, if we call $\mathbf{Z}_l$ and $\mathbf{Z}_f$ the two sub-matrices obtained by removing the last and first row, respectively, it is easy to see that they verify:

$$\mathbf{Z}_l = \mathbf{Z}_f \Gamma \tag{86}$$

where Γ is the diagonal matrix defined by:

$$\Gamma = \mathrm{Diag}\{\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_P\}.$$

If we define in the same way $\mathbf{U}_l$ and $\mathbf{U}_f$ as the sub-matrices of **U** in Eq. (85) in which the last and first

rows are missing, respectively, Eq. (85) yields

$$\mathbf{Z}_l = \mathbf{U}_l\mathbf{Q} \qquad \text{and} \qquad \mathbf{Z}_f = \mathbf{U}_f\mathbf{Q} \tag{88}$$

Combining Eq. (86) and (88), we obtain:

$$\mathbf{U}_f = \mathbf{U}_l(\mathbf{Q}\Gamma\mathbf{Q}^{-1}) \tag{89}$$

which provides a solution to the problem of finding Γ.

1. Find matrix $\mathbf{M}$ such that $\mathbf{U}_f = \mathbf{U}_l\mathbf{M}$; a least squares solution to this problem is obtained by first forming the normal equations:

$$\mathbf{M} = (\mathbf{U}_l^* U_l)^{-1} U_l^* U_f \tag{90}$$

2. Diagonalize $\mathbf{M}$, whose eigenvalues are the poles we are trying to estimate.

The complete procedure includes the following steps: build the Hankel signal matrix $\mathbf{H}$, get its singular value decomposition from which we derive $\mathbf{U}$, build $\mathbf{U}_f$ and $\mathbf{U}_l$, find the matrix $\mathbf{M}$ based on Eq. (90) and diagonalize $\mathbf{M}$. Compared to all methods presented before, this method eliminates the polynomial routing steps required to get the poles, and makes linear prediction a procedure only involving linear algebra. The computationally intensive step of this procedure is the singular value decomposition of $\mathbf{H}$. For that reason, Mutzenhardt et al. [153] have proposed replacing this step by first building the correlation matrix $\mathbf{H}^*\mathbf{H}$, and then replacing $\mathbf{U}$ mentioned earlier by $\mathbf{U}'$, the matrix containing the eigenvectors of $\mathbf{H}^*\mathbf{H}$. Though the latter is theoretically completely equivalent to performing a singular value decomposition in which the left singular values only are computed, it benefits numerically from the fact that the correlation matrix $\mathbf{H}^*\mathbf{H}$ can be computed analytically.

This approach, initially proposed by Barkhuijsen et al. [152], was recently modified by Van Huffel et al. [154] who proposed to compute the matrix $\mathbf{M}$ (Eq. (90)) by Total Least Squares instead of Least Squares.

In practice, after SVD of the Hankel data matrix has been performed, only the first L singular values and the corresponding L first vectors of $\mathbf{U}$ are retained in order to decrease the impact of noise, based on the regularization concept (see Section 3.6.2.1). Ober and Ward [155] have found however that this truncation may lead to severe phase distortion in the corresponding LP spectrum, even though the Fourier transform of the raw signal leads to a well phased, noise corrupted spectrum. To solve this problem, they derived a modification of the method based on the theory of stochastic realization, which always provides correctly phased spectrum (obviously, their method only works when the spectrum is expected to be phased).

Rotation invariance of the signal matrix has been studied in detail for general signal processing, yielding methods such as MUSIC [156] and ESPRIT [157]. Both techniques are widely studied.

4.3. Amplitudes and phases of the signals

Having found the poles Z of the signal by the rooting procedure described earlier, the complex amplitudes of the exponential functions (which contains both the real amplitudes and the phases) are derived by another least-squares calculation corresponding to Eq. (78). In the case where some roots have been identified to be only noise-related, this second set of equations has a lower order. All methods described earlier for solving linear equations can be applied. It should be noted that the matrix form corresponding to Eq. (78):

$$\begin{bmatrix} 1 & 1 & . & . & 1 \\ Z_1 & Z_2 & . & . & Z_L \\ Z_1^2 & Z_2^2 & . & . & Z_L^2 \\ . & . & . & . & . \\ Z_1^{N-1} & Z_2^{N-1} & . & . & Z_L^{N-1} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ . \\ . \\ C_L \end{bmatrix} = \begin{bmatrix} A_0 \\ A_1 \\ . \\ . \\ A_{N-1} \end{bmatrix} \tag{91}$$

in which L is the effective number of poles, after elimination of the noise-related ones, and N the total number of data points, involves a Vandermonde matrix $\mathbf{V}$. In this case the normal equations are reasonably efficient as $\mathbf{E} = \mathbf{V}^*\mathbf{V}$ can be calculated analytically:

$$E(i,j) = \sum_{k=1}^{N} Z_i^{*k-1} Z_j^{k-1} = \frac{1 - \left(Z_i^* Z_j\right)^N}{1 - \left(Z_i^* Z_j\right)} \tag{92}$$

thus reducing significantly the number of operations, as well as the condition number of $\mathbf{E}$. This is the method proposed by Gesmar and Led [41,95].

Noise or glitches in the first data points of the experimental NMR signal yield to an offset in the spectrum after Fourier transformation as well as to baseline distortion, which makes least squares analysis of the spectrum difficult. The first points however cannot be removed prior to Fourier analysis, or large phase problems occur. In the LP procedure however, the first points can be omitted both for finding the LP coefficients, and for the calculation of the amplitudes described earlier.

4.4. Conclusions on model-based analysis of NMR signals

The NMR FID can be modeled as the sum of complex exponential functions. Finding the characteristic elements of the signal components by nonlinear least-squares minimization is a hard numerical problem, which requires knowledge of good initial conditions which are not readily available in the case of complex signals. Under reasonable assumptions however, the digital experimental signal sampled at constant time interval can be defined as an AR process. The problem of solving for the amplitudes and phases of the exponential functions can then be solved by a method known as the extended Prony method which consists of three basic steps. Firstly, the AR coefficients are determined, which are used to define a characteristic polynomial. The second step involves finding the roots of this polynomial, from which the frequencies and damping factors of the components of the NMR signal are derived. These roots are also used in the final procedure which evaluates the complex amplitudes of the signal using a least squares approach.

Several methods have been proposed to find the roots of the AR polynomial. One of the major issues dealt with by these methods is the problem of potential overflow as a result of the high orders involved. Fortunately, all signal roots fall very close to the unit circle in the complex plane. Solving for the roots of the characteristic polynomial is the only nonlinear step in the whole procedure. It can be recast into an eigenvalue problem using the companion matrix approach, or even completely bypassed using the state–space approach proposed by Barkhuijsen et al. [152] which combines the two approaches.

The final step involves solving a linear system to extract the complex amplitudes of the exponential functions which describe the signal. Several linear least-squares techniques are available to solve this problem; it is worth noting that the normal equations are a reasonable solution as they can be built analytically, instead of involving the costly product of two matrices (both in time and precision).

5. Linear prediction applied to ND NMR experiments

The development of NMR for solving solution structures of macromolecules has greatly benefited from the successive introductions of multi-dimensional NMR techniques (2D in the 1970s, 3D and even higher since then) [5,9,10,12,158]. Dimensions in an ND NMR experiments are not equivalent: in the detected dimension, long acquisition times can be used without adding to the total time needed for recording the spectrum, while in the indirectly detected dimensions, acquisition time is kept small, and usually shorter than the macroscopic relaxation rate T_2. Limitations are mainly related to the fact that there are constraints of time as well as storage for recording these types of experiments. As signals acquired for long t_i values contribute less to the signal-to-noise ratio, where t_i corresponds to one of the reconstructed dimension, the total t_i time is usually kept shorter than T_2.

Truncation of a time-domain signal is a problem in the case of a Fourier analysis, and zero-filling followed by severe apodization only offers a partial solution. In the case of 1D experiments, we have shown that linear prediction offers an alternative solution, using the autoregressive properties of the signal; it should therefore also provide an effective improvement of ND experiments. LP is a time-consuming procedure though, and simplifications and/or shortcuts are usually needed to keep the computation for a ND experiment tractable.

5.1. Extracting amplitudes from a ND spectrum

The simpler solution is to extract as much information as possible prior to analysing the ND experiment. This was originally proposed by Manassen et al. [159] for a 2D experiment: they first determined the resonance frequencies and the decay rates of the spins of

the sample under study from a well resolved 1D experiment, and then calculated the amplitude of the 2D signals by a least-squares calculation. The same procedure was rediscovered 10 years later by Muller [160].

It should be mentioned that this method would not be reliable in the case of complicated spectra showing severe signal overlap in 1D experiments (for which higher dimensional experiments are usually required to resolve the overlaps).

5.2. Hybrid spectral analysis for ND spectra

The most common applications of linear prediction to 2D experiments uses the fact that multidimensional Fourier transform is separable, i.e. each dimension can be treated independently [34]. In fact, there is no need to use the same technique for spectral analysis in each dimension: for example, DFT could be applied to the detected dimension, as it usually contains enough data points, while a high-resolution modeling method could be used afterwards for any of the reconstructed dimensions. This leads to the concept of hybrid spectral analysis. All techniques presented before for 1D signal are then adequate, leading to three different approaches:

1. After DFT has been applied to the detected dimension, each FID in the reconstructed dimension is extended by linear prediction prior to Fourier transform: this minimizes the need for strong apodization, and should therefore improve both sensitivity and resolution. Such techniques have been applied to 2D, 3D and 4D data matrices [32,33,41,68,161]. It should be mentioned that in 3D and 4D, the truncation problem is unavoidable, to the extent that in some of the indirectly detected dimensions, only 8 or 16 time increments are used. These cases highlight another problem, i.e. the fact that the number of correlations present in a given dimension can exceed the number of data points available. Zhu and Bax [162] have proposed an elaborate procedure to account for this problem: in the case of a 2D experiment for which $t2$ and $t1$ are the detected and indirectly detected dimensions, respectively, they use a sequence composed of Fourier transformation in $t2$, followed by extrapolation by linear prediction of all FID in $t1$, then FT in $t1$, inverse Fourier transform in $t2$, extrapolation by linear prediction in $t2$, and finally FT in $t2$.
2. The first approach described before only uses the autoregressive properties of the signal to extrapolate the FID, and then applies DFT. We have seen in the preceding sections that spectral analysis can be performed based solely on the fact that the signal is an AR process: this has been used by Hoch [163] for a 2D COSY spectrum in a technique that can be referred to as ''hybrid 2D spectral estimator'', in which 1D DFT in the detected domain $t2$ is followed by spectral estimation in the $t1$ or indirectly detected dimension. He used Burg's algorithm to calculate the AR coefficients from which he derived a power spectrum based on Eq. (19). The final hybrid spectrum showed higher resolution compared to a 2D spectrum generated by FT in the two dimensions on the same amount of data.
3. Linear prediction can also be used quantitatively to derive directly all components of the signal. Schusseim and Cowburn [164] have used this property for the hybrid spectral analysis of a 2D signal, by first applying DFT in the $t2$ dimension, followed by quantitative LP analysis in $t1$. Application of the extended Prony method yields lists of signal parameters for each column of the data matrix, and these lists can be scanned to filter for true signals and to eliminate noise.

Hybrid spectral analysis does not have to include DFT: linear prediction could be applied in dimension 2, followed by nonlinear least-squares modeling in dimension 1. This strategy was applied to the quantitative determination of the exchange rates of slowly exchanging amide protons in proteins [165]. In this study, the exchange rates are obtained by fitting a model that includes exponential decay (i.e. NLSQ in a pseudo-dimension 1) to the parameters produced by a linear prediction analysis of a series of 1D FID recorded during the exchange of the amide protons with solvent (LP in dimension 2). LP analysis of each FID requires a large number of LP coefficients in order to assure a correct estimate of the spectral parameters and this calculation must be repeated for each FID. This was made possible using the fast LP procedure of Gesmar and Hansen [137].

5.3. Spectral analysis using sequential linear prediction

The hybrid procedure aforementioned can be extended one step further, in the sense that both dimensions can be analyzed by a high-resolution modeling technique. This was initially proposed by Barkhuijsen and coworkers for 2D electron spin data [65], and then modified by Gesmar and Led [166] to comply with NMR data. In the following, we describe the latter technique.

Let us consider a 2D experiment in which the data are complex in both dimensions. The corresponding 2D can be expressed as a sum of bi-complex signals under the three conditions that the experiment can be phased in absorption mode, there is not first-order phase correction in one of the dimension, and the noise contribution can be neglected [167]. Most of the phase-sensitive experiments satisfy these conditions, providing that the experimental conditions are defined such that the detected dimension requires very little first order correction. The general expression for an hypercomplex FID is:

$$x(\mathbf{t}_1,\mathbf{t}_2)=\sum_{k=1}^{K}\sum_{l=1}^{L}C_{kl}*\exp[\mathrm{i}\varphi_k^{(1)}]\times\exp[\mathrm{j}\varphi_l^{(2)}]\exp[-R_k^{(1)}t_1]\exp[\mathrm{i}2\pi f_k^{(1)}t_1]\times\exp[-R_l^{(2)}t_2]\exp[\mathrm{j}2\pi f_l^{(2)}t_2] \tag{93}$$

in which modulation occurs in the two dimensions defined by t_2 and t_1. C_{kl} is the real amplitude of the correlation between peak k and l, φ stands for the phase, R for the decay rate, and f for the frequency of resonance. The superscripts 1 and 2 indicate that the quantity is related with $t1$ and $t2$, respectively. For sake of clarity, we differentiate the number for complex representation in the two dimensions, and denote them as i and j. $x(t_1,t_2)$ corresponds in fact to two complex numbers or four real numbers which are usually referred to as rr, ir, ri and ii where r stands for real and i for imaginary.

When this signal is sampled at regularly spaced intervals in the two dimensions, it can be represented by an hypercomplex time series:

$$A_{mn}=\sum_{k=1}^{K}B_{mk}\exp[\mathrm{j}\varphi_k^{(2)}]\times\exp[-R_k^{(2)}n\Delta_2]\exp[\mathrm{j}2\pi f_k^{(2)}n\Delta_2]+w_{mn} \tag{94}$$

where

$$B_{mk}=\sum_{l=1}^{L}C_{kl}\exp[\mathrm{i}\varphi_l^{(1)}]\times\exp[-R_l^{(1)}m\Delta_1]\exp[\mathrm{i}2\pi f_l^{(1)}m\Delta_1] \tag{95}$$

$\Delta1$ and $\Delta2$ are the sampling interval in the $t1$ and $t2$ dimensions, respectively, and w the noise sequence. A and w have hypercomplex values which depend on i and j, while B is complex, and depends on i only. The concept of hypercomplex representation might be confusing at this stage and should be seen simply as a condensed mathematical notation. What happens is that in the case of a States experiment [168], two FIDS are recorded for each $t1$ time value, i.e. for each m, with a phaseshift between the two of $\pi/2$. A single compact model for these two FIDs is directly encoded into Eq. (94), which can be decomposed into standard complex notation as:

$$A^{r}_{(2m-1)n}=\sum_{k=1}^{K}\mathrm{real}(B_{mk})\exp[\mathrm{j}\varphi_k^{(2)}]\times\exp[-R_k^{(2)}n\Delta_2]\exp[\mathrm{j}2\pi f_k^{(2)}n\Delta_2]+w^{r}_{(2m-1)n} \tag{96}$$

and

$$A^{i}_{(2m)n}=\sum_{k=1}^{K}\mathrm{imag}(B_{mk})\exp[\mathrm{j}\varphi_k^{(2)}]\times\exp[-R_k^{(2)}n\Delta_2]\exp[\mathrm{j}2\pi f_k^{(2)}n\Delta_2]+w^{i}_{(2m)n} \tag{97}$$

in which $2m-1$ and $2m$ refer to the FID index, and A^{r}, A^{i}, w^{r}, w^{i} are standard complex numbers.

Eqs. (96) and (97) relate to the $t2$ or detected dimension, while Eq. (95) corresponds to the reconstructed dimension. This separation provides the basis for the sequential LP spectral analysis technique proposed by Gesmar and Led [166], which schematically proceeds as follows:

1. A model-based linear prediction procedure is applied on each FID, yielding the frequencies and decay rates in the $t2$ dimension, as well as the amplitudes defined as real(B) and imag(B).
2. For each frequency k in dimension 2, a complex FID is reconstructed in dimension 1, based on Eq. (95).
3. Each reconstructed FID is analyzed by quantitative linear prediction, yielding the frequencies and decay rates in dimension 1, as well as the amplitudes C of the correlation between the frequencies.

Performing LP on each FID of the 2D data set can be very costly in computer time; however this procedure can be greatly simplified. In Eqs. (96) and (97), the two exponential expressions which contain the decays rates and the frequencies are independent of m, consequently the linear coefficients are independent of m. The entire two-dimensional data set can therefore be included in the determination of the coefficients for the $t2$ dimension. Gesmar and Led [166] proposed to generate the matrix corresponding to the total set of normal equations by summing the normal equation matrices built from each FID. Having found the autoregressive coefficients based on this cumulative data matrix, the poles of the signal in the $t2$ dimension are derived as the root of the characteristic equation, from which the frequencies $f^{(2)}$ and decay rates $R^{(2)}$ are extracted.

As the poles are known at this point, the amplitudes B are evaluated using a LSQ procedure. This calculation must be carried out for each FID, though the inversion of the matrix required to solve the systems need only be performed once.

The simplification proposed earlier for the $t2$ dimension could be applied theoretically for the $t1$ dimension also. However, Gesmar and Led [166] preferred to perform the complete LP calculations independently for each reconstructed FID in dimension 1. The main reason behind this choice is that the number of available data points in the $t1$ dimension is usually small, which limits the choice for the linear prediction order; the latter however should be chosen to be larger than the expected number of resonance in the signal, in order to comply with the autoregressive model (see Section 3). In most cases, each resonance in the $t2$ dimension is only correlated with a small number of frequencies, and a small LP order is sufficient. Including all possible frequencies however would require a much larger number of data points.

Let us define $Z^{(1)}$ and $Z^{(2)}$ by:

$$Z_k^{(2)} = \exp\left[-R_k^{(2)}\Delta_2\right]\exp\left[\mathrm{j}2\pi f_k^{(2)}\Delta_2\right]$$
$$Z_l^{(1)} = \exp\left[-R_l^{(1)}\Delta_1\right]\exp\left[\mathrm{j}2\pi f_l^{(1)}\Delta_1\right] \tag{98}$$

and let us consider the special case where the signal is ''symmetric'' with respect to the two frequency axes, and the sample intervals for the two time axes are the same. Under these conditions, all $Z^{(1)}$ and $Z^{(2)}$ are the elements of the same list L, i.e. the same list of poles is expected in both dimensions. Under these restrictive conditions, there exists a set of AR equations satisfied by the signal in both dimensions. From these equations, a set of AR coefficients is derived, and used to build the characteristic polynomial whose roots correspond to L, the list of Z values defined earlier. The knowledge of these poles allows the calculation of the frequencies and corresponding decay rates of the signal, as well as the amplitude of the correlation, based on a least-squares calculation. This is the strategy proposed by Imanashi et al. [169], which can be seen as an extension of the method of Gesmar and Led [166].

The two major challenges in the estimation of 2D signal intensities are the handling of signal overlaps and the accurate definition of a base plane for the signals to be examined. While overlaps can be handled directly by both methods described earlier, the baseline may be more of a problem. This is an important issue for signals recorded on samples in H_2O solutions, in which cases the spectra show severe distortion of the baseline caused by the dispersive tails from the water resonance (''F2 ridges''). As shown by Holak et al. [170], the intensity I of the correlation between two frequencies f_1 and f_2 in a phase-sensitive 2D spectrum can be estimated from the product of the 1D intensities, $I_2(f_1)$ and $I_1(f_2)$, as determined from the row at $f = f_1$ and the column at $f = f_2$, respectively, and from the spectrum amplitude $S(f_1,f_2)$, using:

$$I = \frac{I_2(f_1)I_1(f_2)}{S(f_1,f_2)} \tag{99}$$

Within this framework, Gesmar et al. [171] designed an alternative approach based on sequential application of linear predictions in both dimensions of the 2D

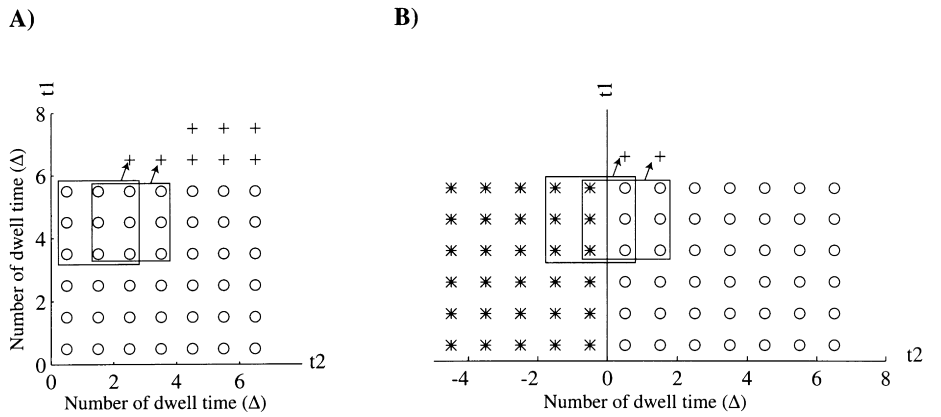

Fig. 4. Schematic representation of the LP procedure of Zhu and Bax [174] for extrapolation of 2D NMR signals. Extrapolation in dimension 1 defined by Eq. (101) is described. Experimentally determined data points are shown as circles (○), while extrapolated points are shown as (+). (A) The experimental signal is causal in both dimensions, restricting the extrapolation: for example, with the LP order considered [(*M*,*P*) = (3,3)], the data points at (0.5,6.5) and (1.5,6.5) cannot be computed. A solution to this problem was proposed by Zhu and Bax, in which the signal is first extrapolated in the negative time domain by reflection (data points shown as *), as shown in (B).

signal. In this method, the 1D intensities $I_2(f_1)$ and $I_1(f_2)$ are determined independently by linear prediction, followed by nonlinear least-squares analysis, after which $S(f_1,f_2)$ is calculated from the estimated spectral parameters. This method will fail in the presence of large baseline distortion, but can easily accommodate a baseline correction. Such a modification was suggested by Kristensen et al. [172] which proceeds as follow: the 1D intensity $I_2(f_1)$ is derived directly from the partially Fourier transformed dataset $S(t2,f_1)$, as originally proposed by Gesmar et al. [171]. To enable a linear prediction in the F1 dimension on a baseline corrected signal, the 2D data set $S(t1,t2)$ is Fourier transformed and phase corrected in both dimensions, then a baseline correction is performed in the F2 dimension, and finally an inverse Fourier transform reverts the F1 dimension to the time domain, to give the required data set $S(f_2,t1)$, from which an apparent intensity $I_1{}^*(f_2)$ is derived. At this point, it is important to realize that the baseline correction in the F2 dimension affects the estimated 1D intensity in the F1 dimension, and that $I_1(f_2) \neq I_1{}^*(f_2)$. This comes from that fact that the discrete Fourier transform of an exponentially damped sinusoid can be approximated as the sum of a Lorentzian and a pseudo-baseline [41]. By removing the baseline in the 2D spectrum, both the ridges and the pseudo-baseline for each column of the matrix have been removed. The latter can be accounted for when estimating the F1 1D signal intensity by setting:

$$I_1(f_2) = I_1^*(f_2) + Ib_2 \tag{100}$$

where I is the intensity of the correlation between f_1 and f_2, and b_2 the F2 pseudo-baseline level. The 2D intensity I is then derived by replacing Eq. (100) in Eq. (99), and solving for I.

5.4. Direct 2D LP analysis

Both the hybrid procedure as well as the sequential LP method described earlier make use of the separability of spectral analysis of multidimensional data. The concept of direct 2D LP spectral analysis of NMR spectra has been suggested originally by Marple [70] and by Kumerasan and Tufts [173], and exploited for the first time by Zhu and Bax [174]. The latter authors proposed to replace the general equation for 1D autoregression by two master equations in 2D, which are:

$$x(m,n) = \sum_{k=1}^{M} \sum_{l=0}^{P-1} c_{kl} x(m-k, n-l) \tag{101}$$

and

$$x(m,n) = \sum_{k=0}^{M-1} \sum_{l=1}^{P} d_{kl} x(m-k, n-l) \tag{102}$$

where x is the hypercomplex signal, M and P the linear prediction order in dimensions 1 and 2,

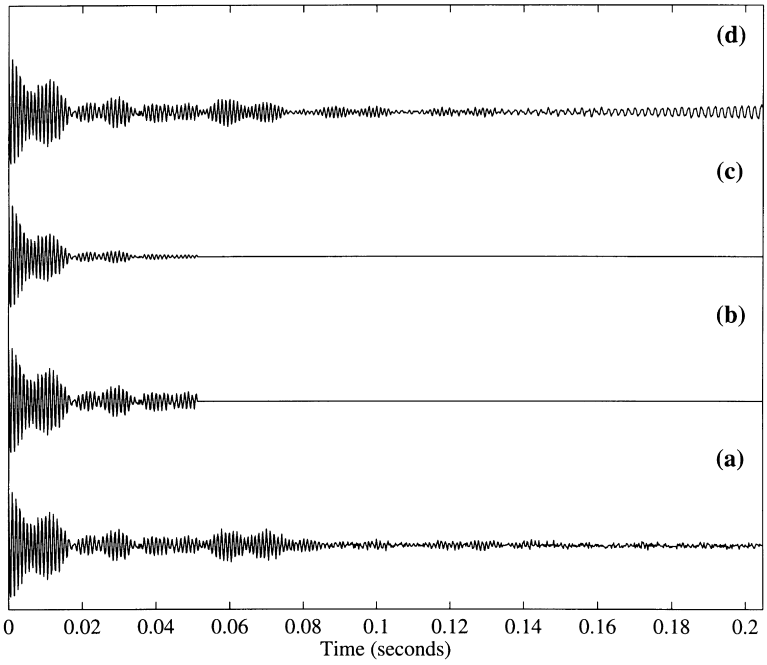

Fig. 5. A synthetic FID corresponding to a signal containing 4 components at frequencies − 1000, − 985, − 970 and − 900 Hz. All signals have the same decay rates (7 Hz), same amplitude (100), and same phase (0). The synthetic FID is sampled at 200 μs, and a white noise sequence is added, with a SNR of 30 dB. (a) real part of the complete FID over 1024 points. (b) real part of the FID truncated after 256 points. (c) same as in (b), after exponential apodization (lb = 10Hz). (d) the truncated FID shown in (b) is extrapolated by linear prediction up to original size of 1024 data points. The linear coefficients are computed based on a SVD of the data matrix, including all 256 available data points, with a LP order of 128.

respectively, and c_{kl} and d_{kl} two sets of linear coefficients. When using Eq. (101), the data matrix can be extended in dimension 1 only, while Eq. (102) allows extension in dimension 2, as illustrated in Fig. 4.

Based on Eq. (101), it is only possible to predict $x(m,n)$ for which $n > P$: for example, all data points $x(N1 + 1,n)$ cannot be predicted if $n < P$. To circumvent this limitation, Zhu and Bax [174] have proposed extending the data in the negative time-domain first, and then apply the 2D LP procedure, as shown schematically in Fig. 4. Extension into the negative time-domain is performed as follows. First, as the time-domain is short, it is possible to assume, as a first approximation, the decay rates to be the same for all frequency components, in which case the damping of the signal can be remove by multiplication by a single increasing exponential. After removal of the decreasing component, the signal can be extended into the negative time-domain by simple symmetry. To avoid discontinuity, the phase of the signal has to be adjusted experimentally to zero at time zero, and the first data point is set to exactly $\Delta/2$.

This method was found to work better than the hybrid FT-LP procedure based on the separability of the spectral analysis on 2D data that are severely truncated in both dimensions. In practice, this method is only of real use for the analysis of 3D or 4D data sets. The price in computing time of this method is however enormous. Zhu and Bax [174] mentioned in their paper that application of this procedure to a small 16×16 hypercomplex data matrix using a 12×12 hypercomplex prediction matrix required 20 min on a 10 Mflops computer. Considering that the procedure has to be applied twice for extension in both dimension, and repeated for all cross sections through a 3D and 4D data set, the total computational time required limits its application. To our knowledge, there has not yet been any follow-up to the paper of Zhu and Bax [174].

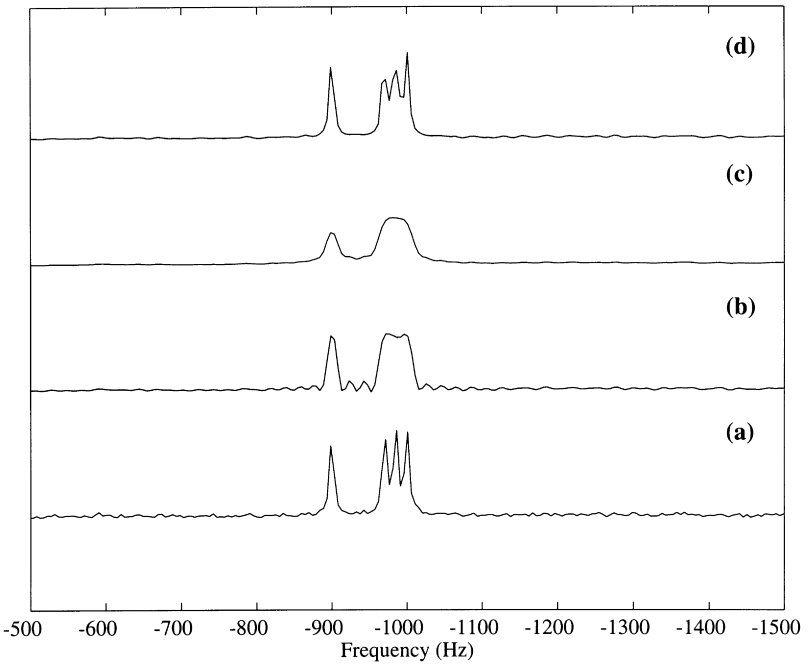

Fig. 6. (a), (b), (c) and (d) show the real part of the FT spectra corresponding to the signals shown in Figs.5(a),(b),(c)and(d), respectively. Truncation of the signal induces a loss of resolution, as well as the apparition of ''wiggles'' on the spectrum, as shown in (b). Apodization with an exponential function removes the wiggles as well as reduces the contribution of noise, at the cost however of a loss of resolution (see (c)). By extrapolation using linear prediction, the resolution is recovered.

6. Linear prediction analysis of NMR spectra

As we have seen before, the name ''linear prediction'' in the field of NMR spectroscopy covers two different types of applications:

1. The NMR signal can be modeled as an autoregressive process, and the AR coefficients can be used either to extrapolate the time domain signal, or to construct a spectrum;
2. A complete quantitative spectral analysis of the NMR signal is performed based on the AR coefficients and a mathematical model for the signal, using the so-called extended Prony method.

The next two sections cover examples of these two applications, as well as a discussion of their limitations.

6.1. Linear prediction can complement Fourier transformation

The NMR free induction decay signal is measured at N intervals of time to give the data vector $\{A_n\}$ $n = 0, \ldots, N-1$. Provided that the time interval is constant, this time series can be analyzed by application of the discrete Fourier transform. DFT provides as output a spectrum, from which information on the spin system can be derived, such as resonance frequencies, relaxation rates and population (intensities). However, DFT has strict requirements as well as limitations, and some of these can be alleviated by linear prediction.

The FID is acquired over a finite time interval, $D = N\Delta$. The corresponding truncation results in distortions in the signal, i.e. ''wiggles'' which appear on both side of each frequency component. The total acquisition time is also inversely proportional to the spectral resolution. Both conditions call for FIDs with a large number of data points covering most of the transient physical signal. This is usually not a problem for 1D NMR as data points can be collected without any cost; in the case of high dimensional NMR however, the number of measurements in any of the reconstructed dimensions should be kept as low as

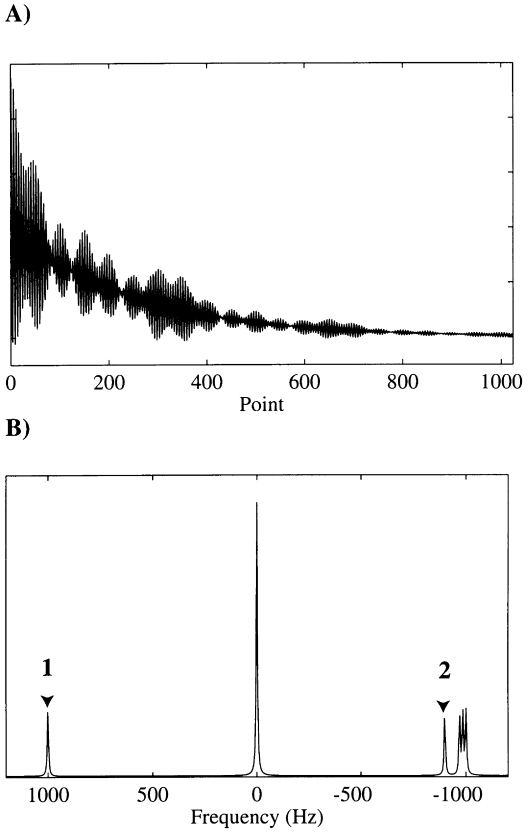

Fig. 7. Test case to study the influence of LP extrapolation on the quantification of NMR spectra. A synthetic complex signal $S(t)$ is generated, containing 6 exponential functions, sampled at 5000 Hz. The real part of the FID computed over 1024 points is shown in (A), and the corresponding spectrum in (B). Characteristics of peak 1 are: 1000 Hz, decay rate 7 Hz, amplitude 100 and phase 0. Peak 2 is located at − 970 Hz, and has the same decay rate, amplitude and phase as peak 1.

possible to reduce the length of the experiment. In natural abundance heteronuclear experiments, lack of sensitivity may also impose truncation, in the sense that only the first and most information-containing part of an FID is acquired to limit the contribution of noise. Setting the missing data points at the end of the signal to zero (i.e. ''zero-filling'') would artificially improve the resolution, but would not solve the problem of truncation. The latter usually requires that the signal is apodized, at the cost of lost of resolution. Zero-filling also introduces baseline problems in the spectrum, in the sense that the process of eliminating the pseudobaseline by multiplying the first data point by 1/2 no longer works. Linear prediction offers an alternative solution to zero filling as the autoregressive coefficients can be used to extrapolate the signal, hence limiting the effect of truncation, and improving both resolution and sensitivity. This is illustrated on synthetic data in Figs. 5 and 6. Truncations of experimental FIDs usually result from time contraints or sensitivity issues while recording the experiment, but it may also be inherent in the experimental procedure itself. The latter case occurs for example in variable angle correlation spectroscopy (VACSY) in solid state NMR. With conventional 2D VACSY it is not possible to obtain pure absorption mode spectra because of ''phase-twist'' artifacts inherent to the experiments. Pines and co-workers however have implemented linear prediction to extrapolate the missing data points in the 2 D FID [175]. Fourier transform of the processed FID yields an absoprtion-mode 2D VACSY spectra free of artifacts.

Extrapolation using forward linear prediction was first proposed by Tirendi and Martin [176] for the case of a NOESY spectrum of a small protein. Since then, this technique has become very popular for processing all types of homonuclear as well as heteronuclear multi-dimensional experiments in high resolution liquid NMR spectroscopy of macromolecules. It is consequently implemented as standard in all NMR processing programs. For example, approximately one-third of all papers, including high dimensional spectra, which have been published in the Journal of Biological NMR between 1996 and 1997 explicitly mention the use of linear prediction as a tool to extrapolate the signal (19 out of 62 in 1996, and 20 out of 55 in 1997). The popularity of the technique is also apparent from the fact that it is often referred to as ''standard linear prediction'', without further reference to any paper. The problem however of such a quote is the definition of ''standard''. While zero-filling and even Fourier transformation can be used through a ''black box'', linear prediction depends on a small list of parameters. When using a standard signal processing package, the NMR spectroscopist has no control on the numerical technique implemented to calculate the autoregressive coefficients; however, he does have the choice of values for the LP parameters. Two parameters are essential.

Firstly, the order L of the autoregressive process; this in fact corresponds to the number of LP

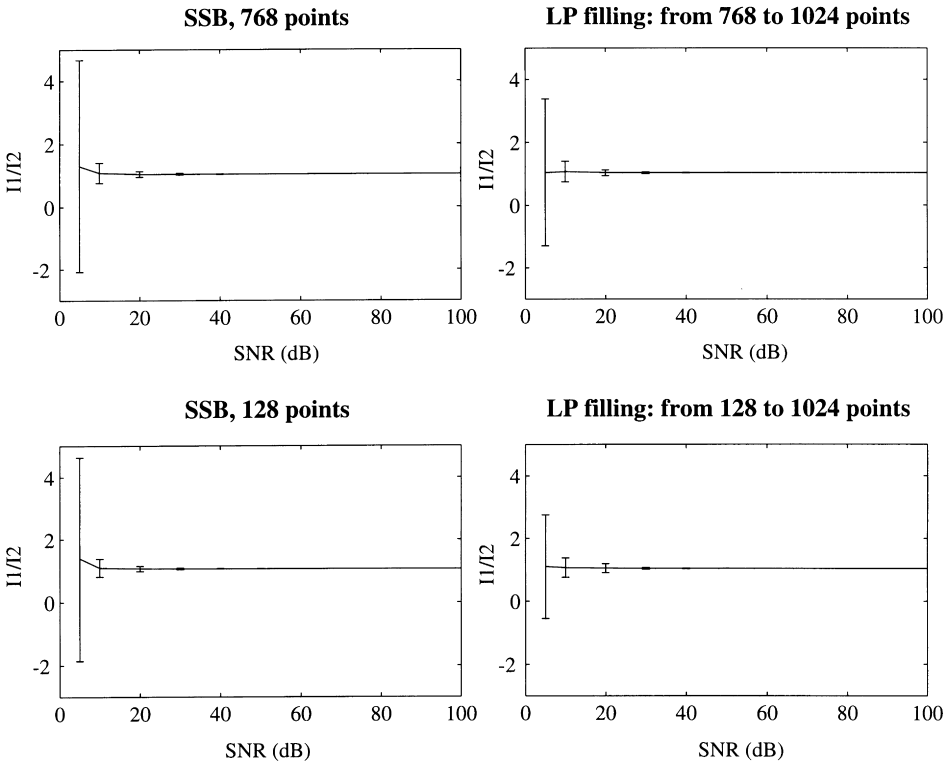

Fig. 8. Mean and standard deviation of the ratio $I1/I2$ of measured surface area for peak 1 and peak 2 defined in Fig. 7, computed over a set of 100 independent spectra, versus SNR in the synthetic FID. The latter is truncated after 128 or 768 data points, and either apodized using a square sine bell (ssb) window with phase 90, or extrapolated to a total of 1024 points using LP, followed by a smooth exponential apodization with an LB of 1. All FIDs are Fourier transformed over a total of 1024 points, and the surface areas $I1$ and $I2$ of peaks 1 and 2 in the spectra are computed by summing all intensities in a box of 30 points centered on the peak. Notice that preprocessing with LP always results in more accurate and more precise estimate of the ratio $I1/I2$.

coefficients. We have seen in Section 4 that the AR model is correct only in the limit of an infinite number of coefficients. While this is not practical, L should still be chosen to be larger than the number of exponentials in the signal, and at the same time to be small enough to improve the statistics in the process of calculating the L coefficients, by increasing the number of equations included in the linear system to be solved. Both $L = N/2$, which is the largest possible value for L, and $L = N/4$ have been used.

Secondly, the extent of extrapolation; this parameter is usually related to the desired resolution in the frequency domain. In most applications it is set to 2, i.e. to doubling the number of available data points. If the SNR of the signal is small, artifacts will appear in the reconstructed data point, such as the increase at the end of the FID, as described in Fig. 5(d). This increase can occur even when special care has been taken to remove or modify any roots of the characteristic AR polynomial which would correspond to signal with negative ''decay rates'', i.e. that would increase with time. It can be related to the presence of noise in the last data points of the signal, which are used to initiate the extrapolation. In general, the signal is apodized using a smooth window function after extrapolation by LP, in order to eliminate these artifacts.

Surprisingly however, extrapolation of the signal by linear prediction has less impact in the NMR structural biology community: out of 52 papers published in the Journal of Molecular Biology in 1996 which report the study of the structure and or the dynamic properties of a macromolecule, only 6 of them specifically refer to linear prediction (the figures for 1997

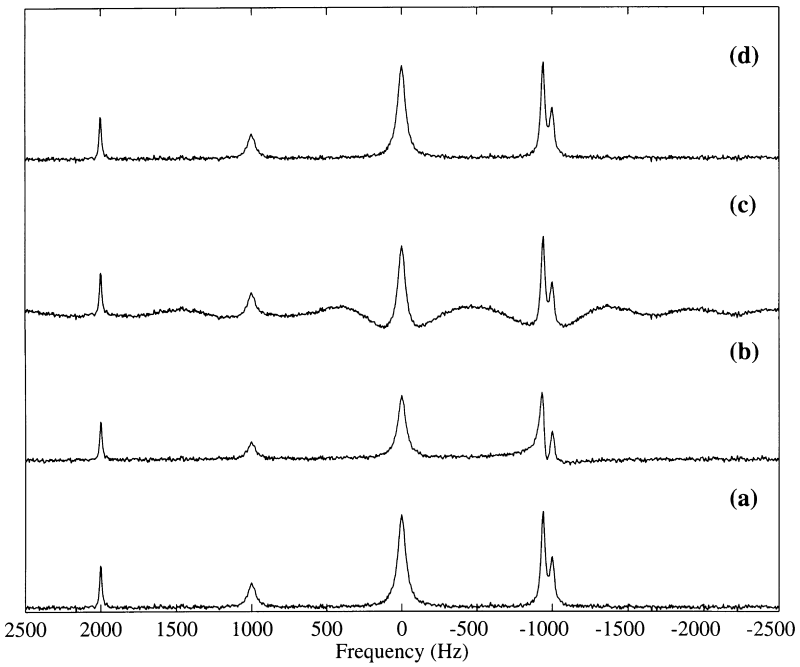

Fig. 9. (a) Real part of the spectrum of a synthetic signal containing 5 exponential functions (at frequencies − 1000, − 940, 0, 1000 and 2000 Hz; the corresponding decay rates are 30, 30, 60 and 15 Hz, amplitudes 100, 200, 400, 100, 50, and phases 0). The signal is sampled over 1024 points with a sampling frequency of 5000 Hz, and a white noise sequence is added (SNR = 30). (b) The first data points have been omitted, and the incomplete FID is transformed as if data point 11 corresponds to $t = 0$. The resulting spectrum shows distorsions in the amplitudes of the peaks (see for example peak at frequency − 1000 Hz). (c) Another solution is to maintain data point 11 at 10 dwell times, and set the missing data points at 0. This results in severe distortions of the baseline. (d) Backward extrapolation of the incomplete FID using linear prediction restores a distorsion-free signal.

are 9 out of 60). Reynolds and coworkers [177] reported a similar observation in the case of NMR studies of small organic molecules: out of the 100 papers published in Magnetic Resonance in Chemistry between the year 1995 and 1996 on structure determination of organic compounds by NMR, only 8 make reference to the use of linear prediction. One possible explanation for this lack of interest is the fear that extrapolation using LP might introduce distortions, preventing further quantification of the corresponding spectrum. In a very simple example described in detail in Figs. 7 and 8, we show that the use of LP does in fact improve quantitative analysis, compared to the use of basic zero-filling followed by severe apodization. The same conclusions were reached by Reynolds et al. [177], who showed that linear prediction used to extrapolate the reconstructed dimension can double the sensitivity or reduce the acquisition time by 4 for a 2D heteronuclear experiment such as an HSQC. By contrast however, they show that forward linear prediction in the acquired dimension is not recommended. A similar study was performed by Babcook et al. [178], in the case of the quantification of a NOESY spectrum for a nucleic acid. They have shown that the estimated interproton distances obtained from a relaxation matrix analysis of NOESY data are insensitive to the use of LP to extend relatively small data sets. Another reason for this lack of interest can simply be the fact that LP may not be needed in some cases. 2D experiments still represent the majority of spectra used for structure elucidation. Advances in chemistry and biology have made it possible to prepare highly concentrated samples, resulting in a major decrease in the NMR experimental time. The time saved in this way is used to acquire more data points in the indirectly detected dimension. Should the corresponding reconstructed FID cover most of the

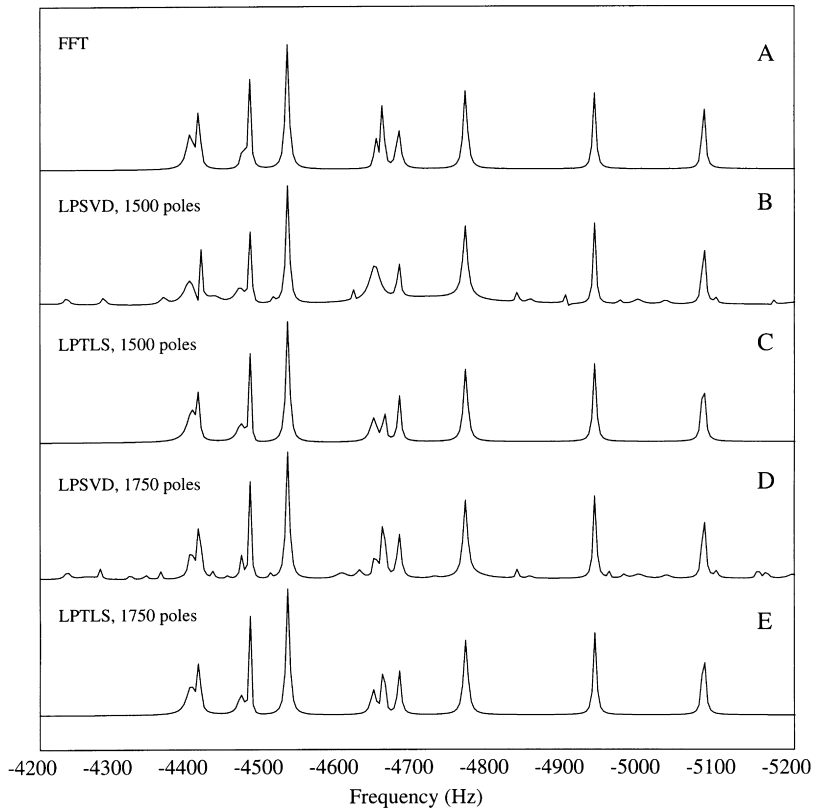

Fig. 10. (A) Selected region of the FFT NMR absorption spectrum of a test signal including 57 components, sampled at 32 μs intervals (SW = 31 250 Hz). The signal mimics the ^{13}C spectrum of an undecapeptide in which artefacts, such as peaks with small and large linewidths or small amplitudes, have been introduced to test the limits of linear prediction techniques. The synthetic FID was computed over 4096 complex data with a sweep width of 31 250 Hz. (B) and (C) show the same region of LP spectra resulting from an LPSVD calculation and an LPTLS calculation of order 1500, respectively. Uncorrelated noise sequences were added to the synthetic noise-free FID, such that the SNR is 10 dB. (D) and (E): same as in (B) and (C), with an LP order of 1750. In all LP calculations, poles with modules larger than 1 are discarded. Notice that at a given order, LPTLS performs better than LPSVD.

decay of the signal, extrapolation to avoid truncation-induced distortion becomes much less crucial.

When the first data points of an NMR signal are missing and ignored, or are present but badly corrupted by noise, the spectrum obtained by Fourier transformation shows severe phase problems as well as large ripples in the baseline. Setting these initial defective points to zero does not alleviate the problem, and even affects any possible quantification as the first data point (which is the sum of all intensities in the spectrum) has been modified. Application of backward linear prediction to reconstruct the missing data points removes the phase and the baseline problem. This is schematically described in Fig. 9, and was originally proposed by Marion and Bax [179]. Soft-pulse experiments such as soft COSY, soft NOESY and soft HMQC are examples in which this problem occurs: the length of the selective pulses causes considerable frequency-dependent dephasing, which may require first-order phase correction of the order of 1000°, which would lead to severe distortion. A simple solution to that problem is to extrapolate the experimental FID backward to zero, that is at the time at which all the exponential functions in the signal are in phase. Backward LP is an ideal method for this, and its application to a soft-HMBC experiment was demonstrated by Led et al. [180].

When data points are missing within the FID by

itself, DFT cannot be applied as such; alternative options include other numerical techniques such as Fourier transform of nonuniformly sampled data. LP does not usually perform well on such signals.

6.2. Complete spectral analysis using linear prediction

Linear prediction as defined in the extended Prony method provides a means of deriving all spectral parameters of a signal. The procedure involves three steps: deriving the autoregressive coefficients by solving a linear set of equations, using these coefficients to build a characteristic polynomial whose roots define the resonance frequencies and the decay rates of the components of the signal, and finally deriving the amplitude of the signals from the signal and the knowledge of these roots, using a least-squares calculation (all 3 parts have been reviewed in details in Sections 3 and 4). However, the success of this method has only be theoretical so far: while every spectroscopist would agree that DFT has limitations, and that it could be replaced completely by linear prediction, to our knowledge there has not yet been any application paper in the field of NMR spectroscopy in which DFT has not been used for spectral analysis. Nevertheless, the use of LP as a quantitative tool has attracted and continues to attract considerable interest. Most of the methodology papers on quantitative LP deal with one of the two following subjects: the choice of the numerical techniques for deriving the LP coefficients, and the accuracy and precision with which the characteristics of the spectral components can be derived. We will briefly review these two aspects.

Good estimates of the linear prediction coefficients are required in order to yield correct estimates of the signal components. This suggests, without any surprise, that the better the input data, the better the results, and that LP methods cannot be used to alleviate severe distortion of the time-domain signal, if quantification is expected. To derive the LP coefficients, a large linear system has to be solved; this system can be represented as:

(Data matrix)*(LP coefficients) = (Observation vector) + noise.

We have seen before that several numerical techniques have been applied to this problem. These methods can basically be divided into three categories:

1. Least-squares (LS) procedures which proceed through a decomposition of the data matrix, either by SVD, QR or Cholesky;
2. Total least square (TLS) procedures which incorporate the effect of noise into both the data matrix and the observation vector (both are derived from the experimental signal);
3. State space models which use the special structure of the data matrix (Hankel or Toeplitz), and its rotational invariance.

In the original paper which introduced the application of TLS for linear prediction on NMR signals, Tirendi and Martin [126] showed that TLS provides a substantial improvement for spectral estimation compared to standard least-squares procedures, when the SNR is low, or the data are truncated. Improvements are more apparent at low prediction orders. These results were later confirmed by Uike et al. [127] and Koehl et al. [97]. We will briefly describe the results of the latter study in order to illustrate specific aspects of LP calculations. The abilities of both LS and TLS methods to provide correct estimates of NMR signal parameters were tested on synthetic data containing different noise sequences. The LSQ method chosen is based on Cholesky decompositon of the normal equations. A Monte Carlo study was performed at various signal-to-noise ratios, in order to estimate the accuracy and precision with which the four signal parameters (frequency, decay rate, amplitude and phase) can be derived. The calculations were repeated for various prediction orders. Several points emerged from these computer experiments.

1. The TLS linear prediction method always performs better than the LSQ LP procedure, for the same number of backward coefficients, in agreement with Tirendi and Martin [126] and Uike et al. [127]. However it should be noted that the LSQ LP method is much faster than the TLS method, and was shown to yield similar results to TLS, provided that a larger number of backward coefficients are used (see Fig. 10). This is important if computing time is a problem.

2. Both LS and TLS based LP methods provide biased estimates of the four NMR parameters listed earlier, and this bias increases as the SNR decreases. The extended Prony method for Linear Prediction relies on the hypothesis that a finite AR model is a good approximation for describing the signal. We have seen in Section 3 that this approximation is exact in the case of a noise-free NMR signal. In the presence of noise however, the signal is better described by an ARMA process, for which an AR process is only an approximation. The bias described here indicates in fact that this approximation breaks down at low SNR.
3. Errors on all four spectral parameters (frequency, linewidth, amplitude and phase) increase when the SNR of the simulated signal decreases, and decrease when the number of backward LP coefficients is increased.
4. The frequency is the most reliable parameter determined by both LP methods. Relative errors on the linewidths, amplitudes and phases of the spectral components are comparatively larger, specifically at low SNR ratios and small number of backward LP coefficients, in which cases some signals have not been detected.

It is possible to improve the performance of the LSQ procedure with regularization of the data matrix. A regularized matrix can be derived from the data matrix by first performing an SVD, and then setting to zero all singular values lower than a given threshold, defined to distinguish signal from noise. Regularized LPSVD was still found however to be less efficient than LPTLS [127].

As a side note, it is worth mentioning that there are at least three types of application for NMR signal processing that manipulate the singular values of the data matrix. Firstly, the smaller singular values are set to zero in order to lower the background noise; this is the discrete regularization process. Secondly, the root-mean-square value obtained by combining the singular values defined as noise is removed from the retained, signal-related singular values. This bias compensation was first applied by Kumaresan et al. [64], and was shown to reduce $t1$ noise when applied within the Cadzow preprocessing procedure [181]. Thirdly, the largest singular value can be set to zero in order to minimize the impact of the largest component in the signal, such as water in solution NMR [110].

Both LP-LSQ and LP-TLS fail for low SNR because the AR approximation breaks down. An obvious solution to this problem is to repeat the experiment in order to improve the SNR of the signal. Though obvious, this solution is often not realistic as the original signal already comes from a highly optimized experiment. Another approach is to pre-process the signal in order to reduce the contribution of noise. One possibility is to use the Cadzow enhancement procedure [123], described in Section 3.6.2 which results in a processed signal that fits the noisy data in a least-squares sense. All of the aforementioned methods, including the Cadzow enhancement procedure [123], require the number of signal components as an input from the user. A method that automatically and reliably determines this number would prove very useful. In this context, the application of the concept of continuous regularization seem promising, as described by Kolber and Schaffel [112]. Successive applications of the combination of continuous regularization and SVD in order to derive the number of component in the signal, signal enhancement using the Cadzow procedure, and LP processing of the resulting processed signal using the state–space model based on TLS, has led to a fully automated signal processing procedure, referred to as LPSVD(CR) + IEP + HTLS, which was shown (on synthetic data) to perform better than previously used protocols [116].

A major challenge in quantitative signal processing is the handling of signal overlaps. With DFT techniques, discerning two lines separated by δ Hz requires the total acquisition time to be at least $1/\delta$ This presupposes that the NMR FID persists this long; if it does not, then the two lines cannot be resolved because the intrinsic linewidth caused by the sample is too large. Zero-filling artificially improves the spectral resolution by increasing the number of points, but does not solve this problem. One solution is to apply a line-sharpening filter to the signal prior to Fourier transform, such as a sine or cosine window function. With linear prediction techniques, the spectral resolution was found to be dependent on the linear prediction order, L. Increasing the prediction order of linear prediction requires an increase of the total number of data points. A solution is to increase the

total acquisition time at a given sampling rate. This does not work in the case of signals with short T_2 values however, as adding data points past the decay of the true signal results in a decrease of the signal-to-noise ratio. An alternative approach is oversampling [182]. Clearly, much more information can be gained on an analog signal at a higher sampling rate. Oversampling has been shown to improve the dynamic range of NMR spectra [182], to improve the resolution of proton NMR images [183] and to reduce baseline artefacts [184]. Koehl et al. [185] have shown that oversampling also increases the identification power of linear prediction, as well as the accuracy and precision of the LP-estimated peak separations.

6.3. A broader range of application for linear prediction

Linear prediction is in fact a general technique in signal processing, and has been applied in a large variety of applications, including for example speech recognition, telecommunication and image processing. Among all these applications, data compression, usually referred to as linear predictive coding, can be relevant to NMR. The principle is simple. Let us consider a signal $S(t)$, which has been sampled at constant time interval Δ, and digitalized, The corresponding time series is a collection of integer values, $\{S_n\}$. In the case of NMR data, this time series can be modeled by an autoregressive process, yielding

$$S_i = \sum_{k=1}^{N} b_k S_{i-k} + \varepsilon_i \tag{103}$$

in which ϵ_i is a measure of the error of the model, supposedly small. This is what makes this approach useful for data compression: the idea is to store the small residual errors rather than the raw data, which can be large. As data are usually stored as 32 bit integers, the residuals are defined as:

$$\delta_i = E\left[\sum_{k=1}^{N} b_k X_{i-k}\right] - X_i \tag{104}$$

The compression scheme for a 1D NMR signal is therefore performed as follows: for a given N, the autoregressive coefficients $\{b_k\}$ are computed and stored with the N first data values of the FID, then the δ_i values are computed for the rest of the FID, and stored efficiently. This technique is fully reversible and does not result in any loss of information; it has been tested for storing data from multidimensional experiments in solution-state NMR spectroscopy [186], as well as 2D images in magnetic resonance imaging (MRI) [187], with compression ratios varying from 1.7:1 to 3.8:1, depending on the SNR of the signal or image to be compressed.

All applications described earlier deal with complex signals modeled by a sum of complex exponential functions. In the case of NMR, it was shown that linear prediction also applies to 1D signals detected in sequential quadrature mode (or Redfield mode) [41,68], as well as to an FID in the F1 dimension of a 2 D experiment acquired in TPPI mode [188]. One step further is to apply linear prediction to real signals modeled by the sum of real exponential functions. Two such applications have been proposed, both concerned with extracting information from NMR relaxation curves [189,190]. Extracting parameters from a sum of real exponential functions is a difficult problem however, and I do not expect linear prediction to work much better than any other numerical technique.

7. Conclusions

NMR spectroscopy is becoming an ubiquitous, and sometimes crucial technique in physics, chemistry and biology. It is probably in the latter that progress has been most significant: the introduction of high dimensional NMR spectroscopy, the availability of high field NMR equipment as well as the major advances in the preparation of biological samples have turned high resolution solution-state NMR into a major tool in structural biology as well as in medical research. An interesting feature of the study of the structure of a macromolecule by NMR is the fact that the time spent in analyzing the information extracted from the NMR data is now larger than the time needed to collect the data by order of magnitudes. It is therefore important to optimize the experimental conditions in order to obtain the best possible raw data from a given sample, as well as to optimize the signal processing techniques used to analyze these raw data and extract the spectral parameters. DFT has been and remains the technique of choice for NMR

signal processing; we have tried to show in this review that the linear prediction technique is mature enough to improve the Fourier analysis, and in some circumstances even replace it completely.

Qualitative linear prediction is based solely on the hypothesis that the signal can be modeled as an autoregressive process. It allows backward as well as forward extrapolation of the time domain signal, prior to spectral analysis by DFT. Both operations yield to significant improvements of the NMR spectra, including attenuation of baseline, phase and intensity distortions, as well as improvement of resolution and sensitivity. Acquisition and treatment of multidimensional NMR spectra have greatly benefited from LP.

Quantitative linear prediction directly provides estimates of the frequencies, decay rates, amplitudes and phases of the components of the signal by further assuming a mathematical model for the data in the form of the sum of complex exponential functions. At the time of this review, the main developments concerning quantitative LP are related to in vivo NMR spectroscopy, in which case the number of resonances in the spectra are usually small. There has been great attention given to the automation of the LP procedure in these cases. Applications to more complex systems of spins have been limited, and mainly theoretical. It remains an active area of research though, as quantification remains a bottleneck in any high resolution NMR study.

It is clear, however, that the advent of faster computers as well as the progress of numerical methods have reduced the computing time limitations of linear prediction to a minimum, and this technique is now applied routinely for NMR signal processing, at least in its qualitative approach.

Acknowledgements

I gratefully thank Dan Spielman and Steve Lynch for their comments on the manuscript, as well as all members of the Michael Levitt laboratory at Stanford and the Jean-Francois Lefevre laboratory at Strasbourg for their support. Financial support from the Union Internationale Contre le Cancer (Geneva, Switzerland) is acknowledged. This work is dedicated to the memory of Helene Hietter-Koehl.

Appendix A. Circulant and Toeplitz matrix computations

An $N \times N$ matrix $\mathbf{C}$ is called a circulant matrix if its entries $C(i,j)$ satisfy $\mathbf{C}(i,j) = r[(i-j)\mathrm{mod}(N)]$ where r is the first row of $\mathbf{C}$. For example, a 3×3 circulant matrix has the form

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & c_2 \\ c_2 & c_0 & c_1 \\ c_1 & c_2 & c_0 \end{bmatrix}. \tag{A1}$$

An important property of circulant matrices is that they are all diagonalizable, and all have the same complete set of eigenvectors. Specifically the circulant matrix $\mathbf{C}$ can be written as

$$\mathbf{C} = \mathbf{F}^*\mathbf{D}\mathbf{F} \tag{A2}$$

where $\mathbf{F}$ is the unitary discrete Fourier transform matrix:

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 & . & 1 \\ 1 & \omega & \omega^2 & . & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & . & \omega^{2(N-1)} \\ . & . & . & . & . \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & . & \omega^{(N-1)(N-1)} \end{bmatrix} \tag{A3}$$

where $\omega = \exp[-(\mathrm{j}2\pi/N)]$, and $\mathbf{D}$ is the diagonal matrix containing the eigenvalues of $\mathbf{C}$, given by:

$$\mathrm{diag}(\mathbf{D}) = DFT(\mathbf{c}) = \sqrt{N}\mathbf{F}\mathbf{c} \tag{A4}$$

where $\mathbf{c}$ is the first column of $\mathbf{C}$ and DFT(.) denotes the discrete Fourier transform (DFT). As DFT can usually be performed using fast algorithms such as FFTs in $O(n \log n)$ operations, the eigenvalues of any circulant matrix can be computed amazingly fast. These properties are described in detail in the book ''Circulant Matrices'' by Davis [191].

Eqs (A2) and (A4) imply that the following matrix computations can be performed in $O(n \log n)$ operations.

1. The circulant matrix-vector multiplication $\mathbf{y} = \mathbf{C}\mathbf{x}$ is computed as

$$\mathbf{y} = \mathbf{F}^*\mathbf{D}\mathbf{F}\mathbf{x} \tag{A5}$$

therefore

$$\mathbf{y} = \mathrm{iFFT}\Big[\mathrm{FFT}(\mathbf{c})^{-1} \otimes \mathrm{FFT}(\mathbf{x})\Big] \tag{A6}$$

where iFFT is the inverse fast Fourier transform, and $\otimes$ is used to denote the component wise multiplication of vectors (or Hadamard matrix multiplication).

2. The inverse of a non singular circulant matrix is a circulant matrix given by:

$$\mathbf{C}^{-1} = \mathbf{F}^{*}\mathbf{D}^{-1}\mathbf{F}. \tag{A7}$$

3. Non-singular circulant systems of equations $\mathbf{Cx} = \mathbf{y}$ can be solved in $O(n \log n)$ operations by using both Eqs (A5) and A(6), yielding:

$$\mathbf{x} = \mathrm{iFFT}\Big[\mathrm{FFT}(\mathbf{c})^{-1} \otimes \mathrm{FFT}(\mathbf{y})\Big] \tag{A8}$$

where $\mathrm{FFT}(\mathbf{c})^{-1}$ denotes the component-wise inversion of the vector FFT(**c**).

An $N \times N$ matrix **T** is called a Toeplitz matrix if its entries $T(i,j)$ verify $T(i,j) = t(I - j)$. For example, a 3×3 Toeplitz matrix has the form

$$\mathbf{T} = \begin{bmatrix} t_0 & t_1 & t_2 \\ t_3 & t_0 & t_1 \\ t_4 & t_3 & t_0 \end{bmatrix}. \tag{A9}$$

Circulant matrices are particular Toeplitz matrices.

The matrix multiplication properties of circulant matrices can be extended to Toeplitz matrices, by observing that any Toeplitz matrix **T** can be embedded into a circulant matrix of twice its dimension. For example, the matrix **T** in Eq. (A9) can be embedded into the 6×6 circulant matrix

$$\mathbf{C} = \begin{bmatrix} t_0 & t_1 & t_2 & 0 & t_4 & t_3 \\ t_3 & t_0 & t_1 & t_2 & 0 & t_4 \\ t_4 & t_3 & t_0 & t_1 & t_2 & 0 \\ 0 & t_4 & t_3 & t_0 & t_1 & t_2 \\ t_2 & 0 & t_4 & t_3 & t_0 & t_1 \\ t_1 & t_2 & 0 & t_4 & t_3 & t_0 \end{bmatrix}. \tag{A10}$$

Notice the position of the zeros in **C**. Additional zeros can be added between 0 and t_4 on each line in order to set the size of **C** to be a power of 2.

Then $\mathbf{y} = \mathbf{Tx}$ is computed by using FFTs to form the circulant matrix vector product

$$\mathbf{y} = \mathbf{C}\begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} \tag{A11}$$

and extracting **y** as the first N component of $\mathbf{y}'$.

Appendix B. Glossary

B.1. Basic matrix definitions

Let us define $M_{m,n}(C)$ as the ring of matrices with complex entries of size $m \times n$, and $M_m(C)$ the sub-ring of $M_{m,n}(C)$ containing all square matrices of size m.

For any matrix $A \in M_{m,n}(\mathrm{C})$, we can define the three following matrices:

$\mathbf{A}^{\mathrm{T}}$, or transpose of **A**, defined as $A^{\mathrm{T}}(i,j) = A(j,i)$;

$\overline{A}$ or complex conjugate of **A**, defined as $\overline{A}(i,j) = \mathrm{conjg}(\mathbf{A}(i,j))$;

$\mathbf{A}^{*}$, or hermitian of A, defined as $\mathbf{A}^{*}(\mathrm{i,j}) = \mathrm{conjg}(\mathbf{A}(j,i))$.

A square matrix **A** is invertible if, and only if $\det(\mathbf{A}) \neq 0$, where det stands for determinant. In such case, **A** is non-singular.

A $m \times m$ Hermitian matrix **A** is positive definite if $\mathbf{x}^{*}\,\mathbf{Ax} > 0$, for all non-zero m-dimensional vector **x**. It can be shown that **A** is positive definite if and only if all its eigenvalues are positive.

B.1.1. Choleski decomposition of a matrix

Any positive definite matrix **A** in $M_{\mathrm{m}}(C)$ may be factored as

$\mathbf{A} = \mathbf{S}\,\mathbf{S}^{*}$

Where **S** is a lower triangular matrix having positive values on its diagonal. The decomposition is unique.

B.1.2. Circulant matrices

An $M \times M$ matrix **C** is called a circulant matrix if its entries $C(i,j)$ satisfy $C(i,j) = r[(i - j)\mathrm{mod}(N)]$ where **r** is the first row of **C**. That is, each column or row of **C** can be obtained by a circulant shift of its previous column or row, respectively. For example, a 4×4

circulant matrix has the form

$$\mathbf{C} = \begin{bmatrix} c_0 & c_3 & c_2 & c_1 \\ c_1 & c_0 & c_3 & c_2 \\ c_2 & c_1 & c_0 & c_3 \\ c_3 & c_2 & c_1 & c_0 \end{bmatrix}.$$

Basic properties of circulant matrices have been described in the Appendix A.

B.1.3. Condition number of a matrix

The condition number $\kappa(\mathbf{A})$ of a matrix $\mathbf{A}$ in $M_m(C)$ is defined as:

$$\kappa(\mathbf{A}) = \begin{cases} \|\mathbf{A}^{-1}\|\|A\| & A \text{ non-singular} \\ \infty & A \text{ singular} \end{cases}$$

where $\|\cdot\|$ stands for any matrix norm (see the following text). In the special case the spectral norm is chosen, the condition number of a squared, non-singular matrix $\mathbf{A}$ of size $m \times m$ is given by:

$$\kappa(\mathbf{A}) = \frac{S(1,1)}{S(m,m)}$$

where $S(1,1)$ and $S(m,m)$ are the two extreme singular values of $\mathbf{A}$.

The condition number is in fact a good measure of the ''conditioning'' of a matrix: if $\kappa(\mathbf{A})$ is close to 1, the matrix is said to be well-conditioned, and its inverse can be computed with acceptable precision; if $\kappa(\mathbf{A})$ is ∞, the matrix is rank deficient therefore non-invertible. If $\kappa(\mathbf{A})$ is large, the matrix $\mathbf{A}$ is said to be ''ill-conditioned'', and most numerical techniques used to solve a linear system of the form $\mathbf{Ax} = \mathbf{y}$ will not yield a correct estimate for $\mathbf{x}$. Solutions to this problem exist which usually involve preconditioning of the system (see for example Chan et al. [192] for the case of Toeplitz matrix), or regularization.

B.1.4. Hankel matrices

An $M \times M$ matrix $\mathbf{H}$ is called an Hankel matrix if its entries $H(i,j)$ satisfy $H(i,j) = h(i + j - 2)$ that is the entries $H(i,j)$ depend only on the sum of the two indices: all elements on a diagonal perpendicular to the main diagonal are constant. For example, a 3×3 Hankel matrix has the form

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_1 & h_2 & h_3 \\ h_2 & h_3 & h_4 \end{bmatrix}$$

B.1.5. Least squares (LS) and Total least-squares (TLS) solutions of a linear system

Given a matrix $\mathbf{A}$ in $M_{m,n}(C)$ and a vector $\mathbf{b}$ in $M_{m,1}(C)$, the least squares solution of the linear system of equations $\mathbf{Ax} = \mathbf{b}$ is the vector x_{LSQ} which solves

$$\min\|\mathbf{b} - \mathbf{Ax}\|_2$$

where $\|\cdot\|_2$ is the Euclidian norm.

The least squares viewpoint is that the observation vector $\mathbf{b}$ only is perturbed by noise:

$$\mathbf{Ax} = \mathbf{b} + \mathbf{e}$$

and the least squares solution is equivalent to determining the minimum perturbation $\mathbf{e}$ applied to $\mathbf{b}$ such that $\mathbf{b} + \mathbf{e}$ becomes a linear combination of the columns of $\mathbf{A}$.

In the TLS problem, both the matrix $\mathbf{A}$ and the observation vector $\mathbf{b}$ are considered to be perturbed by noise:

$$(\mathbf{A} + \mathbf{E})\mathbf{x} = \mathbf{b} + \mathbf{e}$$

which can be rewritten as:

$$([\mathbf{b}, \mathbf{A}] + [\mathbf{e}, \mathbf{E}])\begin{bmatrix} -1 \\ \mathbf{x} \end{bmatrix} = 0.$$

In the TLS problem, one attempts to minimize the perturbation matrix, $\|\mathbf{e},\mathbf{E}\|_{\mathrm{F}}$, subject to $(\mathbf{b} + \mathbf{e}) \in \mathrm{Range}(A + E)$.

B.1.6. Norm of a matrix

Given two square matrices $\mathbf{A}$ and $\mathbf{B}$ with complex entries of size $N \times N$, a matrix norm is a non-negative scalar which has the following properties:

1. $\|\mathbf{A}\| = 0$ if, and only if $\mathbf{A} = \mathbf{0}$, where $\mathbf{0}$ is the $N \times N$ null matrix;
2. $\|k\mathbf{A}\| = |k|\,\|\mathbf{A}\|$ for all k;
3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$;
4. $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|$.

The two most common matrix norms are:

1. the Frobenius or Euclidian or L2 norm of matrix is defined as:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}|A(i,j)|^2}$$

2. the spectral norm is defined as:

$$\rho(\mathbf{A}) = \max\{\sqrt{\lambda}, \qquad \lambda \text{ eigenvalues of } \mathbf{A}^*\mathbf{A}\}$$

B.1.7. Norm of a vector

Give a vector $\mathbf{x}$ of size $N \times 1$ with complex or real entries, the vector norm is a non-negative scalar $\|\mathbf{x}\|$ which has the following properties:

1. $\|\mathbf{x}\| = 0$ if, and only if $\mathbf{x} = \mathbf{0}$;
2. $\|k\mathbf{x}\| = |\mathrm{k}|\cdot\|\mathbf{x}\|$ for all k;
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The most common vector norm is the Euclidian or L2 norm, defined as:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{N}|x(i)|^2}$$

B.1.8. QR decomposition of a matrix

Every $m \times n$ complex matrix can be factored into the product of a matrix $\mathbf{Q}$ having orthonormal vectors for its columns, and an upper (right) triangular matrix $\mathbf{R}$:

$$\mathbf{A} = \mathbf{QR}.$$

B.1.9. Singular value decomposition (SVD) of a matrix

If $\mathbf{A} \in M_{m,n}(C)$ has rank k, then it may be written in the form:

$$\mathbf{A} = \mathbf{USV}^*$$

where $\mathbf{U} \in M_{m,m}(C)$ and $\mathbf{V} \in M_{n,n}(C)$ are unitary matrices (i.e. $\mathbf{UU}^* = \mathbf{I}_m$ and $\mathbf{VV}^* = \mathbf{I}_n$, where $\mathbf{I}_k$ is the $k \times k$ identity matrix). The matrix $\mathbf{S} \in M_{m,n}$ has entries $S(i,j) = 0$ for all $i \neq j$ and $S(1,1) \geq S(2,2)\ldots \geq S(k,k) > S(k+1,k+1) = \ldots S(q,q) = 0$, where $q = \min(m,n)$. The numbers $\{S(i,i)\}$ are the non negative square roots of the eigenvalues of $\mathbf{AA}^*$, and hence are uniquely defined. They are known as the singular values of $M_{m,n}(C)$. The columns of $\mathbf{U}$ are eigenvectors of $\mathbf{AA}^*$, and the column of $\mathbf{V}$ are eigenvectors of $\mathbf{A}^*\mathbf{A}$.

B.1.10. Toeplitz matrix

An $n \times n$ matrix $\mathbf{T}$ is called a Toeplitz matrix if its entries are constant down each of the diagonals. The entries $T(i,j)$ depend then only on the difference of the two indices, and are often written $T(i,j) = t(i-j)$. For example, a 3×3 Toeplitz matrix is usually represented as:

$$\mathbf{T} = \begin{bmatrix} t_0 & t_{-1} & t_{-2} \\ t_1 & t_0 & t_{-1} \\ t_2 & t_1 & t_0 \end{bmatrix}$$

B.1.11. Vandermonde matrix

An $n \times n$ matrix $\mathbf{V}$ is called a vandermonde matrix if its entries are given by:

$$\mathbf{V}(i,j) = x_j^{i-1}$$

where $\mathbf{x} = [x1,\ldots,x_n]$ is a n-dimensional vector.

For example, a 4×4 vandermonde matrix is represented by:

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ x_1^2 & x_2^2 & x_3^2 & x_4^2 \\ x_1^3 & x_2^3 & x_3^3 & x_4^3 \end{bmatrix}$$

The determinant of a vandermonde matrix is directly given by:

$$\det(\mathbf{V}) = \prod_{i,j=1\ i>j}^{n}(x_i - x_j).$$

References

[1] A. Abragam, The Principles of Nuclear Magnetism, Clarendon Press, Oxford, 1961.
[2] J.W. Cooley, J.W. Tukey, Mathematics of Computation 19 (1965) 297.

[3] R.R. Ernst, W.A. Anderson, Rev. Sci. Instrum. 37 (1966) 93.
[4] R.R. Ernst, Adv. Mag. Reson. (1966) 1.
[5] R.R. Ernst, G. Bodenhausen, A. Wokaun, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, Oxford, 1987.
[6] R.R. Ernst, Angew. Chem. Int. Ed. Eng. 31 (1992) 805.
[7] D.M. Grant, P.K. Harris (Eds.), Encyclopedia of NMR, John Wiley, Chichester, 1996.
[8] A. Bax, S. Grzesiek, Acc. of Chem. Res. 26 (1993) 131.
[9] K. Schmidt-Rohr, H.W. Spiess, Multidimensional Solid-State NMR of Polymers, Academic Press, San Diego, 1994.
[10] F.J.M.V.d. Ven, Multidimensional NMR in Liquids: Basic Principles and Experimental Methods, VCH, New York, 1995.
[11] L.E. Kay, Prog. in Biophys. and Molec. Biol. 63 (1995) 277.
[12] A. Pardi, Methods in Enzymology 261 (1995) 350.
[13] A.M. Gronenborn, G.M. Clore, Critical Rev. In Biochem. and Molec. Biol. 30 (1995) 351.
[14] J.G. Pelton, D.E. Wemmer, Ann. Rev. of Phys. Chem. 46 (1995) 139.
[15] M.A. Keniry, Magn. Reson. in Chem. 34 (1996) 33.
[16] D.M. Lemaster, Prog. in NMR Spectrosc. 26 (1994) 371.
[17] R.T. Batey, J.L. Battiste, J.R. Williamson, Methods in Enzymology 261 (1995) 300.
[18] G.C. King, J.W. Harper, Z.J. Xi, Methods in Enzymology 261 (1995) 436.
[19] M. Kainosho, Nature Struct. Biol. 4 (1997) 858.
[20] F.A. Howe, R.J. Maxwell, D.E. Saunders, M.M. Brown, J.R. Griffiths, Magn. Reson. Quart. 9 (1993) 31.
[21] M. Maier, Brit. J. Psychiatry 167 (1995) 299.
[22] I.J. Cox, Prog. in Biophys. and Molec. Biol. 65 (1996) 45.
[23] J.P. Renou, Magn. Reson. in Chem. 35 (1997) S153.
[24] P. Bachert, Prog. in NMR Spectrosc. 33 (1998) 1.
[25] K. Wuthrich, NMR of Proteins and Nucleic Acids, Wiley, New York, 1986.
[26] G.M. Clore, A.M. Gronenborn, Nature Struct. Biol. 4 (1997) 849.
[27] S.J. Opella, Nature Struct. Biol. 4 (1997) 845.
[28] G. Wagner, Nature Struct. Biol. 4 (1997) 841.
[29] I.D. Campbell, A.K. Downing, Nature Struct. Biol. 5 (1998) 496.
[30] L.E. Kay, Nature Struct. Biol. 5 (1998) 513.
[31] K. Wuthrich, Nature Struct. Biol. 5 (1998) 492.
[32] D.N. Rutledge (Ed.), Signal Treatment and Signal Analysis in NMR, Elsevier Science, New York, 1996.
[33] J.C. Hoch, A.S. Stern, NMR Data Processing, Wiley–Liss, New York, 1996.
[34] E.O. Brigham, The Fast Fourier Transform and its Applications, Prentice Hall, Englewood Cliffs, NJ, 1988.
[35] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in FORTRAN: The Art of Scientific Computing, Cambridge University Press, Cambridge, 1996.
[36] J.B. Weaver, Y.S. Xu, D.M. Healy, J.R. Driscoll, Magn. Reson. in Med. 24 (1992) 275.
[37] A.R. Tate, D. Watson, S. Eglen, T.N. Arvanitis, E.L. Thomas, J.D. Bell, Magn. Reson. in Med. 35 (1996) 834.
[38] G. Neue, Solid State NMR 5 (1996) 305.
[39] D. Barache, J.P. Antoine, J.M. Dereppe, J. Magn. Res. 128 (1997) 1.
[40] W. Schempp, Acta Applicandae Mathematicae 48 (1997) 185.
[41] H. Gesmar, J.J. Led, F. Abildgaard, Prog. in NMR Spectrosc. 22 (1990) 255.
[42] C. Tellier, M. Guilloucharpin, D. Lebotlan, F. Pelissolo, Magn. Reson. in Chem. 29 (1991) 164.
[43] H.R. Halvorson, Methods in Enzymology 210 (1992) 54.
[44] S. Boudhabhay, B. Ancian, P. Levoir, R. Dubest, J. Aubard, Computers and Chemistry 16 (1992) 271.
[45] J.C. Hoch, J. Magn. Res. 64 (1985) 436.
[46] P.J. Hore, J. Magn. Res. 62 (1985) 561.
[47] B.C. Desimone, F. Deluca, B. Maraviglia, Magn. Reson. in Med. 8 (1988) 332.
[48] J.M. Dereppe, N. Jakus, J. Magn. Res. 79 (1988) 307.
[49] G.J. Daniell, P.J. Hore, J. Magn. Res. 84 (1989) 515.
[50] J.A. Jones, P.J. Hore, J. Magn. Res. 92 (1991) 276.
[51] P. Hodgkinson, H.R. Mott, P.C. Driscoll, J.A. Jones, P.J. Hore, J. Magn. Res. Series B 101 (1993) 218.
[52] P. Schmieder, A.S. Stern, G. Wagner, J.C. Hoch, J. Magn. Res. 125 (1997) 332.
[53] J.J. Kotyk, N.G. Hoffman, W.C. Hutton, G.L. Bretthorst, J.J.H. Ackerman, J. Magn. Res. 98 (1992) 483.
[54] J.J. Kotyk, N.G. Hoffman, W.C. Hutton, G.L. Bretthorst, J.J.H. Ackerman, J. Magn. Res. Series A 116 (1995) 1.
[55] G.W. Jeong, P.N. Borer, S.S. Wang, G.C. Levy, J. Magn. Res. Series A 103 (1993) 123.
[56] S. Wang, I. Pelczer, P.N. Borer, G.C. Levy, J. Magn. Res. Series A 108 (1994) 171.
[57] R.A. Chylla, J.L. Markley, J. Biomolec. NMR 5 (1995) 245.
[58] S. Umesh, D.W. Tufts, IEEE Trans. on Signal Processing 44 (1996) 2245.
[59] A. Gottvald, Int. J. Appl. Electromagnetics and Mechanics 8 (1997) 17.
[60] G. Zhu, Y.B. Hua, Chem. Phys. Lett. 264 (1997) 424.
[61] O.M. Weber, C.O. Duc, D. Meier, P. Boesiger, Magn. Res. in Med. 39 (1998) 723.
[62] G.J. Metzger, M. Patel, X.P. Hu, J. Magn. Res. Series B 110 (1996) 316.
[63] G.R.B. Prony, J. de L′Ecole Polytechnique, Paris 1 (1795) 24.
[64] R. Kumaresan, D.W. Tufts, IEEE Trans. On Acoustics Speech and Signal Processing 30 (1982) 833.
[65] H. Barkhuijsen, R. Debeer, W. Bovee, D. VanOrmondt, J. Magn. Res. 61 (1985) 465.
[66] H. Barkhuijsen, R. Debeer, W. Bovee, J.H.N. Creyghton, D. VanOrmondt, Magn. Reson. in Med. 2 (1985) 86.
[67] D.S. Stephenson, Prog. in NMR Spectrosc. 20 (1988) 515.
[68] J.J. Led, H. Gesmar, Chem. Rev. 91 (1991) 1413.
[69] J.J. Led, H. Gesmar, Methods in Enzymology 239 (1994) 318.
[70] S.L. Marple, Digital Spectral Analysis: With Applications, Prentice-Hall, Englewood Cliffs, NJ, 1987.
[71] J. Tang, J.R. Norris, J. Magn. Res. 69 (1986) 180.
[72] J. Tang, J.R. Norris, J. Chem. Phys 84 (1986) 5210.
[73] J. Tang, J.R. Norris, Chem. Phys. Lett. 131 (1986) 252.

[74] F. Ni, H.A. Scheraga, J. Magn. Res. 70 (1986) 506.
[75] N. Levinson, J. Math Physics 25 (1947) 261.
[76] J. Durbin, Rev. Int. Statist. I. 28 (1960) 233.
[77] J.P. Burg, in: Proceedings of the 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City, 1967.
[78] J.P. Burg, in: D.J. Childers (Ed.), A New Technique for Time Series Data, IEEE Press, New York, 1978.
[79] P.F. Fougere, E.J. Zawalick, H.R. Radoski, Phys. Earth and Planet. Interiors 12 (1976) 201.
[80] W.Y. Chen, G.R. Stegen, J. Geophys. Res. 79 (1974) 3019.
[81] D.N. Swingler, IEEE Trans. Acoustics Speech and Signal Processing 28 (1980) 257.
[82] J.A. Cadzow, Proc. IEEE 70 (1982) 907.
[83] M. Fedrigo, G. Esposito, S. Cattarinussi, P. Viglino, F. Fogolari, J. Magn. Reson. Series A 121 (1996) 97.
[84] S.M. Kay, S.L. Marple, Proc. IEEE 69 (1981) 1380.
[85] M. Chakraborty, S. Prasad, IEEE Trans. Signal Processing 41 (1993) 1692.
[86] G.M. Cordeiro, R. Klein, Stat. and Prob. Letters 19 (1994) 169.
[87] Y. Inouye, T. Umeda, Ieice Trans. Fund. Elec. Comm. and Comp. Sci. E77A (1994) 748.
[88] C.W. Therrien, C.H. Velasco, IEEE Trans. Signal Processing 43 (1995) 358.
[89] C.B. Xiao, X.D. Zhang, Y.D. Li, IEEE Trans. Signal Processing 44 (1996) 2900.
[90] J.K. Sabiti, Automatica 32 (1996) 235.
[91] K.H. Chon, R.J. Cohen, IEEE Trans. Biomed. Eng. 44 (1997) 168.
[92] M.H.A. Davis, W.X. Zheng, IEEE Trans. Automatic Control 42 (1997) 400.
[93] S. Vanbuuren, Psychometrika 62 (1997) 215.
[94] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.
[95] H. Gesmar, J.J. Led, J. Magn. Res. 76 (1988) 183.
[96] J. Tang, C.P. Lin, M.K. Bowman, J.R. Norris, J. Magn. Res. 62 (1985) 167.
[97] P. Koehl, C. Ling, J.F. Lefevre, J. Magn. Res. Series A 109 (1994) 32.
[98] P.C. Hansen, BIT 27 (1987) 534.
[99] P.C. Hansen, Siam J. Scient. Stat. Comp. 11 (1990) 503.
[100] M. Hanke, P.C. Hansen, Surv. Math. Ind. 3 (1993) 253.
[101] P.C. Hansen, Num. Algorithms 6 (1994) 1.
[102] G.L. Millhauser, A.A. Carter, D.J. Schneider, J.H. Freed, R.E. Oswald, J. Magn. Res. 82 (1989) 150.
[103] W.W.F. Pijnappel, A. Vandenboogaart, R. Debeer, D. VanOrmondt, J. Magn. Res 97 (1992) 122.
[104] J.K. Cullum, R.A. Willoughby, Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Birkhauser, Boston, 1985.
[105] G.H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, 1990.
[106] K.D. Ikramov, D.A. Kushnereva, Comp. Math. Math. Phys. 35 (1995) 1527.
[107] H.G. Yu, S.C. Smith, J. Comp. Phys. 143 (1998) 484.
[108] T.F. Chan, L. Depillis, H. Vandervorst, Num. Algorithms 17 (1998) 51.
[109] A. Bjorck, T. Elfving, Z. Strakos, Siam J. Matrix Analysis and Applications 19 (1998) 720.
[110] G. Zhu, D. Smith. Y.B. Hua, J. Magn. Res. 124 (1997) 286.
[111] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-posed Problems, Wiley, New York, 1977.
[112] W. Kolbel, H. Schafer, J. Magn. Res. 100 (1992) 598.
[113] P.C. Hansen, D.P. Oleary, Siam J. Sci. Comput. 14 (1993) 1487.
[114] A. Diop, W. Kolbel, D. Michel, A. Briguet, D. Graverondemilly, J. Magn. Res. Series B 103 (1994) 217.
[115] A. Diop, Y. Zaimwadghiri, A. Briguet, D. Graverondemilly, J. Magn. Res. Series B 105 (1994) 17.
[116] J. Totz, A. Vandenboogaart, S. Vanhuffel, D. Graverondemilly, I. Dologlou, R. Heidler, D. Michel, J. Magn. Res. 124 (1997) 400.
[117] D. Marion, M. Ikura, A. Bax, J. Magn. Res. 84 (1989) 425.
[118] M.S. Friedrichs, W.J. Metzler, L. Mueller, J. Magn. Res. 95 (1991) 178.
[119] L. Mitschang, C. Cieslar, T.A. Holak, H. Oschkinat, J. Magn. Res. 92 (1991) 208.
[120] M. Adler, G. Wagner, J. Magn. Res. 91 (1991) 450.
[121] Y. Kuroda, A. Wada, T. Yamazaki, K. Nagayama, J. Magn. Res. 88 (1990) 141.
[122] Y. Kuroda, A. Wada, T. Yamazaki, K. Nagayama, J. Magn. Res. 84 (1989) 604.
[123] J.A. Cadzow, IEEE Trans. Acoustics Speech and Signal Processing 36 (1988) 49.
[124] H. Chen, S. Vanhuffel, C. Decanniere, P. Vanhecke, J. Magn. Res. Series A 109 (1994) 46.
[125] I. Dologlou, S. Vanhuffel, D. Vanormondt, IEEE Trans. Signal Processing 45 (1997) 799.
[126] C.F. Tirendi, J.F. Martin, J. Magn. Res. 85 (1989) 162.
[127] M. Uike, T. Uchiyama, H. Minamitani, J. Magn. Res. 99 (1992) 363.
[128] T.F. Chan, P.C. Hansen, Siam J. Sci. Stat. Comput. 13 (1992) 727.
[129] S. Vanhuffel, H. Park, J.B. Rosen, IEEE Trans. Signal Processing 44 (1996) 2464.
[130] R.D. Fierro, G.H. Golub, P.C. Hansen, D.P. Oleary, Siam J. Sci.Comput. 18 (1997) 1223.
[131] H.S. Park, L. Elden, Numerische Mathematik 76 (1997) 383.
[132] G. Cybenko, Siam J. Sci. Stat. Comput. 8 (1987) 734.
[133] T. Kailath, J. Chun, Siam J. Matrix Analysis and Applications 15 (1994) 114.
[134] J.G. Nagy, Siam J. Sci. Comput. 14 (1993) 1174.
[135] C.P. Rialan, L.L. Scharf, IEEE Trans. Acoustics Speech and Signal Processing 36 (1988) 1740.
[136] D.R. Sweet, Numerische Mathematik 58 (1991) 613.
[137] H. Gesmar, P.C. Hansen, J. Magn. Res Series A 106 (1994) 236.
[138] J. Nagy, R.J. Plemmons, Linear Algebra and its Applications 172 (1992) 169.
[139] B. Noorbehesht, H. Lee, D.R. Enzmann, J. Magn. Res. 73 (1987) 423.
[140] A. Vandenboogaart, F.A. Howe, L.M. Rodrigues, M. Stubbs, NMR in Biomed. 8 (1995) 87.
[141] L. Vanhamme, A. Vandenboogaart, S. Vanhuffel, J. Magn. Res. 129 (1997) 35.

[142] C. Decanniere, P. Vanhecke, F. Vanstapel, H. Chen, S. Vanhuffel, C. Vandervoort, B. Vantongeren, D. Vanormondt, J. Magn. Res. Series B 105 (1994) 31.
[143] A. Vandenboogaart, Magn. Reson. in Chem. 35 (1997) S146.
[144] K. Steiglitz, B. Dickinson, IEEE Trans. Acoustics Speech and Signal Processing 30 (1982) 984.
[145] B. Svejgard, BIT 7 (1967) 240.
[146] P. Barone, L. Guidoni, E. Massaro, V. Viti, J. Magn. Res. 73 (1987) 23.
[147] R. Kumaresan, IEEE Trans. Acoustics Speech and Signal Processing 31 (1983) 217.
[148] R. Kumaresan, D.W. Tufts, L.L. Scharf, Proceedings of the IEEE 72 (1984) 230.
[149] M.A. Delsuc, F. Ni, G.C. Levy, J. Magn. Res. 73 (1987) 548.
[150] B. Porat, B. Friedlander, IEEE Trans. Acoustics Speech and Signal Processing 34 (1986) 1336.
[151] G. Zhu, A. Bax, J. Magn. Reson. 100 (1992) 202.
[152] H. Barkhuijsen, R. Debeer, D. VanOrmondt, J. Magn. Res. 73 (1987) 553.
[153] P. Mutzenhardt, P. Palmas, J. Brondeau, D. Canet, J. Magn. Res. Series A 104 (1993) 180.
[154] S. Vanhuffel, H. Chen, C. Decanniere, P. Vanhecke, J. Magn. Res. Series A 110 (1994) 228.
[155] R.J. Ober, E.S. Ward, J. Magn. Res. Series A 114 (1995) 120.
[156] R.O. Schmidt, IEEE Trans. Antennas and Propagation 34 (1986) 276.
[157] R. Roy, T. Kailath, IEEE Trans. Acoustics Speech and Signal Processing 37 (1989) 984.
[158] A.S. Edison, F. Abildgaard, W.M. Westler, E.S. Mooberry, J.L. Markley, Methods in Enzymology 239 (1994) 3.
[159] Y. Manassen, G. Navon, C.T.W. Moonen, J. Magn. Res. 72 (1987) 551.
[160] A. Muller, J. Magn. Res. Series A 114 (1995) 238.
[161] P. Guntert, V. Dotsch, G. Wider, K. Wuthrich, J. Biomolec. NMR 2 (1992) 619.
[162] G. Zhu, A. Bax, J. Magn. Res. 90 (1990) 405.
[163] J.C. Hoch, J. Magn. Res. 64 (1985) 436.
[164] A.E. Schussheim, D. Cowburn, J. Magn. Res. 71 (1987) 371.
[165] R. Moss, H. Gesmar, J.J. Led, J. Am. Chem. Soc. 116 (1994) 747.
[166] H. Gesmar, J.J. Led, J. Magn. Res. 83 (1989) 53.
[167] M.A. Delsuc, J. Magn. Res. 77 (1988) 119.
[168] D.J. States, R.A. Haberkorn, D.J. Ruben, J. Magn. Reson. 48 (1982) 286.
[169] Y. Imanishi, T. Matsuura, C. Yamasaki, T. Yamazaki, K. Ogura, M. Imanari, J. Magn. Reson. Series A 110 (1994) 175.
[170] T.A. Holak, J.N. Scarsdale, J.H. Prestegard, J. Magn. Res. 74 (1987) 546.
[171] H. Gesmar, P.F. Nielsen, J.J. Led, J. Magn. Res. Series B 103 (1994) 10.
[172] S.M. Kristensen, M.D. Sorensen, H. Gesmar, J.J. Led, J. Magn. Res. Series B 112 (1996) 193.
[173] R. Kumaresan, D.W. Tufts, Proc. IEEE 69 (1981) 1515.
[174] G.A. Zhu, A. Bax, J. Magn. Res. 98 (1992) 192.
[175] Y.K. Lee, R.L. Vold, G.L. Hoatson, Y.Y. Lin, A. Pines, J. Magn. Res. Series A 112 (1995) 112.
[176] C. F. Tirendi, J.F. Martin, J. Magn. Res. 81 (1989) 577.
[177] W.F. Reynolds, M. Yu, R.G. Enriquez, I. Leon, Magn. Res. in Chem. 35 (1997) 505.
[178] D.M. Babcook, P.V. Sahasrabudhe, W.H. Gmeiner, Magn. Res. in Chem. 34 (1996) 851.
[179] D. Marion, A. Bax, J. Magn. Res. 83 (1989) 205.
[180] J.J. Led, F. Abildgaard, H. Gesmar, J. Magn. Res. 93 (1991) 659.
[181] C. Brissac, T.E. Malliavin, M.A. Delsuc, J. Biomolec. NMR 6 (1995) 361.
[182] M.A. Delsuc, J.Y. Lallemand, J. Magn. Res. 69 (1986) 504.
[183] D.G. Cory, A.N. Garroway, J.B. Miller, J. Magn. Res. 87 (1990) 202.
[184] G. Wider, J. Magn. Res. 89 (1990) 406.
[185] P. Koehl, C. Ling, J.F. Lefevre, J. Chimie Phys. 91 (1994) 595.
[186] T.E. Malliavin, M.A. Delsuc, J.Y. Lallemand, J. Magn. Res. 93 (1991) 630.
[187] C.C. Li, M. Gokmen, A.D. Hirschman, Y. Wang, Comput. Med. Imag. Graph. 15 (1991) 277.
[188] D. Marion, K. Wuthrich, Biochem. Biophys. Res. Comm. 113 (1983) 967.
[189] T.E. Malliavin, M.A. Delsuc, J.Y. Lallemand, J. Biomolec. NMR 2 (1992) 349.
[190] Y.Y. Lin, N.H. Ge, L.P. Hwang, J. Magn. Reson. Series A 105 (1993) 65.
[191] P.J. Davis, Circulant Matrices, Wiley, New York, 1979.
[192] R.H. Chan, J.G. Nagy, R.J. Plemmons, SIAM J. Matrix Anal. Appl. 15 (1994) 80.