A Subspace Time-Domain Algorithm for Automated NMR Spectral Normalization

Philippe Lemmerling,*,[1] Leentje Vanhamme,* Rocco Romano,† and Sabine Van Huffel*

*Katholieke Universiteit Leuven, Department of Electrical Engineering, SCD-SISTA, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium; and †Istituto Nazionale per la Fisica della Materia, Unità di Napoli, c/o Dipartimento di Scienze Fisiche, Università degli Studi Napoli Federico II, Complesso Universitario di Monte S. Angelo, via Cinthia, 80126 Naples, Italy

E-mail: philippe.lemmerling@esat.kuleuven.ac.be, leentje.vanhamme@esat.kuleuven.ac.be, rocco.romano@na.infn.it, sabine.vanhuffel@esat.kuleuven.ac.be

Recently, two methods have been proposed for quantitatively comparing NMR spectra of control and treated samples, in order to examine the possible occurring variations in cell metabolism and/or structure in response to numerous physical, chemical, and biological agents. These methods are the maximum superposition normalization algorithm (MaSNAl) and the minimum rank normalization algorithm (MiRaNAl). In this paper a new subspace-based time-domain normalization algorithm, denoted by SuTdNAl (subspace time-domain normalization algorithm), is presented. By the determination of the intersection of the column spaces of two Hankel matrices, the common signal poles and further on the components having proportionally varying amplitudes are detected. The method has the advantage that it is computationally less intensive than the MaSNAl and the MiRaNAl. Furthermore, no approximate estimate of the normalization factor is required. The algorithm was tested by Monte Carlo simulations on a set of simulation signals. It was shown that the SuTdNAl has a statistical performance similar to that of the MiRaNAl, which itself is an improvement over the MaSNAl. Furthermore, two samples of known contents are compared with the MiRaNAl, the SuTdNAl, and an older method using a standard. Finally, the SuTdNAl is tested on a realistic simulation example derived from an *in vitro* measurement on cells.   © 2002 Elsevier Science (USA)

Key Words: NMR; algorithm; normalization; subspace methods.

1. INTRODUCTION

NMR spectroscopy is increasingly being used, both *in vitro* and *in vivo*, to examine variations induced by physical, chemical, or biological agents which have acted on a sample or on a patient. In this type of studies, in order to obtain relative quantitative information, a comparison between signal intensities of control samples and treated or exposed samples is often conducted. The problem that arises in such comparisons is that the concentrations of cells in control and treated samples will be different, thus leading to a change of spectral intensities of metabolites even when the concentration of those metabolites on the cellular level has not changed.

To overcome this problem various methods have been developed. An example is the use of reference compounds. For example, in *in vitro* studies, synthetic, such as 3-trimethylsilyl[2,2,3,3-d_4]propionate (*8, 2*), or naturally occurring, such as glucose in ^{13}C spectra (*11*), reference compounds have been used. In addition, in both *in vivo* and *in vitro* studies, enzymatic determination of the concentration of a metabolite present in the samples, as well as the assumption that the concentration of a particular metabolite (e.g., the alanine methyl doublet or the creatine singlet) does not change, has also been utilized (*3, 1*). In particular, a common method of normalization of spectra is to divide each spectral intensity by either the sum of intensities or the square root of the sum of squares of intensities of metabolites whose concentrations are supposed not to change (*6*).

Recently two new methods, which are not based on the addition of external substances of known concentration, on the indirect measurement of internal substances, or on an assumption about the concentration of particular substances, have been proposed: the maximum superposition normalization algorithm (MaSNAl) (*9*) and the minimum rank normalization algorithm (MiRaNAl) (*10*). In this paper a new algorithm, the subspace-based time-domain normalization algorithm (SuTdNAl), is proposed for quantitatively comparing NMR spectra. It is based on the determination of the intersection of the column spaces of two Hankel matrices. The next section introduces the problem setting, followed by a description of the algorithm. The algorithm is tested by Monte Carlo simulations applied to a simulation example. It is demonstrated that, like the MiRaNAl, the SuTdNAl is able to determine the normalization factor with very low bias, both in spectra with a low signal-to-noise ratio and in cases in which great changes are induced by the treatment (degree of diversity >50%). Furthermore it is demonstrated that the variance of the estimated normalization factors obtained with the SuTdNAl and the MiRaNAl are similar. The advantage of the new algorithm over the MiRaNAl is the increased computational

---

[1] To whom correspondence should be addressed. Fax: +32-16-321970.

efficiency and the fact that no initial estimate of the normalization factor is required. Furthermore, two samples of known contents are compared with MiRaNAl, with SuTdNAl, and with an older method using a standard. Finally, the SuTdNAl is tested on a realistic simulation example derived from an *in vitro* measurement on cells.

In the remainder of the paper we will adopt a Matlab-like notation for vectors and matrices:

- $A(p, q)$: the entry in the qth column of the pth row of A.
- $A(p, :)$: the pth row of A.
- $A(:, q)$: the qth column of A.
- $A(p : q, r : s)$: the $(q - p + 1) \times (s - r + 1)$ submatrix of A containing the entries that belong to rows p to q *and* to columns r to s.
- $b(p)$: the entry on the pth row of column vector b.
- $b(p : q)$: the $(q - p + 1) \times 1$ subvector of b containing the entries of row p till row q.
- $b(q : -1 : p)$: this vector is equal to the previous one but with the elements in reversed order.

Furthermore $i = \sqrt{-1}$.

2. RESULTS AND DISCUSSION

2.1. The Algorithm

A time-domain NMR-free induction decay (FID) experiment can be modeled as a sum of complex damped exponentials,

$$x(n) = \sum_{s=1}^{s=S} A(s)e^{i\tilde{\phi}(s)}e^{(-\alpha(s)+i2\pi v(s))t(n)} + e(n),$$

$$n = 0, \ldots, N - 1, \quad [1]$$

where S is the number of complex damped exponentials, $A \in \mathbb{R}^{S \times 1}$ contains the amplitudes, $\alpha \in \mathbb{R}^{S \times 1}$ contains the damping factors, $v \in \mathbb{R}^{S \times 1}$ contains the frequencies (in Hz) and $\tilde{\phi} \in \mathbb{R}^{S \times 1}, \tilde{\phi}(s) = \phi(s) + \phi_0, s = 1, \ldots, S$, where $\phi(s)$ is defined as the phase of the sth component and ϕ_0 is the zero-order phase. Furthermore, $e(n)$ is complex white Gaussian noise. The number of complex data points is represented by N. The discretely sampled time steps are $t(n) = (n + \eta)\Delta t$, with $t(0) = \eta \Delta t$ the begin time or dead time of the spectrometer and Δt the sampling interval. The phases $\phi(s)$, required only under particular experimental conditions, can usually be set equal to zero, as will be assumed in the remainder of the paper. Another way of writing Eq. [1] is as

$$x(n) = \sum_{s=1}^{S} c(s)z(s)^{(n+\eta)}, \quad n = 0, \ldots, N - 1,$$

$$[2]$$

with $c(s) \equiv A(s)e^{i\tilde{\phi}(s)}$ and $z(s) \equiv e^{(-\alpha(s)+i2\pi v(s))\Delta t}$,

where $c(s), s = 1, \ldots, S$, are the so-called complex amplitudes and $z(s), s = 1, \ldots, S$, are the so-called signal poles. Now we consider 2 FIDs $x_1(n)$ and $x_2(n)$ with

$$x_1(n) = \sum_{s=1}^{s=S_1} A_1(s)e^{i(\tilde{\phi}_1(s))}e^{(-\alpha_1(s)+i2\pi v_1(s))t(n)}$$

$$[3]$$

$$x_2(n) = \sum_{s=1}^{s=S_2} A_2(s)e^{i(\tilde{\phi}_2(s))}e^{(-\alpha_2(s)+i2\pi v_2(s))t(n)},$$

$$[4]$$

where the different quantities are defined in a way similar to that in Eq. [1].

What basically happens in the SuTdNAl is a time-domain fit of the FIDs x_1 and x_2 using the proposed model function of complex damped exponentials. A naive implementation would be to determine first the parameters of the model function for FID x_1 and then for FID x_2. The result would be two sets of parameters: $A_1(s), \alpha_1(s), v_1(s), \tilde{\phi}_1(s), s = 1, \ldots, S_1$, and $A_2(s), \alpha_2(s), v_2(s), \tilde{\phi}_2(s), s = 1, \ldots, S_2$. The next step in the procedure would then be looking for common signal poles, i.e., components for which $-\alpha_1(p) + i2\pi v_1(p) \equiv z_1(p) = z_2(q) \equiv -\alpha_2(q) + i2\pi v_2(q)$ for some p and q. The reason for looking for common signal poles is obvious: only signal poles representing metabolites that are present in signal $x_1(n)$ (obtained from the control sample) as well as in signal $x_2(n)$ (obtained from the treated sample) can give rise to proportional changes in their corresponding spectral intensities, i.e., amplitudes $A_1(s), A_2(s)$. In a typical experiment a majority of the spectral intensities of the S_c common signal poles scale with a factor R (the latter of course corresponds to the normalization factor that will be computed), whereas the signal poles whose corresponding spectral intensities do not scale with this factor R belong to metabolites whose concentration on the cellular level has been changed by a particular physical, chemical, or biological agent. It is clear that the above described straightforward approach works perfectly for *noiseless* signals x_1 and x_2. However, in real-life experiments signals x_1 and x_2 are contaminated by complex white Gaussian noise e_1 and e_2 and only x_{1n} and x_{2n} are observed:

$$x_{1n}(n) = x_1(n) + e_1(n) \text{ and } x_{2n}(n) = x_2(n) + e_2(n).$$

Furthermore, the involved signal poles can lie very close to each other. In such circumstances it will be extremely difficult—if not impossible—to detect common signal poles, i.e., determine which signal pole of x_{1n} corresponds to which signal pole of x_{2n}. A more advanced technique is called for in the noisy case and will be described in the next paragraph.

To explain the SuTdNAl we start again from the noiseless case. The algorithm starts by stacking x_1 and x_2 in $L \times M$ Hankel matrices H_1 and H_2,

$$H_1 \equiv \begin{bmatrix} x_1(0) & x_1(1) & \ldots & x_1(M-1) \\ x_1(1) & x_1(2) & \ldots & x_1(M) \\ \vdots & \vdots & \vdots & \vdots \\ x_1(L-1) & x_1(L) & \ldots & x_1(N-1) \end{bmatrix}$$

and

$$H_2 \equiv \begin{bmatrix} x_2(0) & x_2(1) & \dots & x_2(M-1) \\ x_2(1) & x_2(2) & \dots & x_2(M) \\ \vdots & \vdots & \vdots & \vdots \\ x_2(L-1) & x_2(L) & \dots & x_2(N-1) \end{bmatrix},$$

with M chosen greater than S_1 and S_2 and subject to the constraint $N = M + L - 1$. Before we proceed with the explanation of the procedure to find the common signal poles, it is useful to clarify the link that exists between the representation (see Eq. [2]) of a signal consisting of a sum of complex damped exponentials, such as x_1, and the Hankel matrix built from that signal, namely H_1. The relationship is expressed by the Vandermonde decomposition of the Hankel matrix:

$$H_1 = O_1 D_1 C_1^T \qquad [5]$$

with

$$O_1 \equiv \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1(1) & z_1(2) & \dots & z_1(S_1) \\ \vdots & \vdots & \vdots & \vdots \\ z_1(1)^{L-1} & z_1(2)^{L-1} & \dots & z_1(S_1)^{L-1} \end{bmatrix},$$

$$D_1 \equiv \begin{bmatrix} c_1(1) & 0 & \dots & 0 \\ 0 & c_1(2) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & c_1(S_1) \end{bmatrix}$$

and

$$C_1 \equiv \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1(1) & z_1(2) & \dots & z_1(S_1) \\ \vdots & \vdots & \vdots & \vdots \\ z_1(1)^{M-1} & z_1(2)^{M-1} & \dots & z_1(S_1)^{M-1} \end{bmatrix}.$$

From the latter equations, it follows that the column space of the matrix H_1 is determined by the vectors $v_{1k} \equiv [1 \ z_1(k) \ z_1(k)^2 \dots z_1(k)^{L-1}]^T$, $k = 1, \dots, S_1$. Therefore if the signals x_1 and x_2 have S_c signal poles in common, the corresponding vectors v_{1j}, $j = 1, \dots, S_c$ form a basis for the intersection of the column spaces of H_1 and H_2.[2] Therefore, to find the common signal poles first the intersection of the column spaces of H_1 and H_2 is determined and next the signal poles can be extracted from this common subspace. Finding the intersection of

---

[2] It can easily be proven that no vector other than the signal vectors coming from common signal poles lie in the common subspace. The proof relies entirely on the fact that a matrix like O_1 (see Eq. [5]) cannot be rank-deficient if the signal poles are distinct.

the column spaces of H_1 and H_2 can easily be done as follows,

$$H_1 X = H_2 Y,$$

$$[H_1 \ H_2] \begin{bmatrix} X \\ -Y \end{bmatrix} = 0,$$

and next use the Singular Value Decomposition (SVD) of $[H_1 \ H_2] = U_{ex} S_{ex} V_{ex}^H$ to find X and Y:

$$X = V_{ex}(1 : M, S_1 + S_2 - S_c + 1 : 2M)$$
$$Y = -V_{ex}(M + 1 : 2M, S_1 + S_2 - S_c + 1 : 2M). \qquad [6]$$

Note that the column space of the matrix $H_c \equiv H_1 X = H_2 Y$ is the intersection of the column spaces of H_1 and H_2.

Due to the shift-invariant structure of the matrix H_c it is now straightforward (see Appendix A) to determine the common signal poles $z_1(k) = z_2(k)$, $k = 1, \dots, S_c$. Note that the previous operation is represented by the function "*common_sp*" in the algorithmic outline of the SuTdNAl at the end of this section (i.e., at line 2).

Once the common poles are calculated, the remaining poles in x_1 and x_2 are determined. Note that at this point, the situation corresponds to the case in which we have some a priori known signal poles (in this case these are the now available common signal poles) and we have to determine the remaining poles (5, 4). As shown in (5, 4) and briefly summarized in Appendix B, the remaining signal poles can easily be determined, by projecting the row spaces of H_1 and H_2 onto the orthogonal complement of the subspace spanned by the common signal poles and their powers. This projection allows the computation of the remaining signal poles $z_1(k)$, $k = S_c + 1, \dots, S_1$ of x_1 (note that the latter operation is represented by the function "*diff_sp*" in the algorithmic outline of the SuTdNAl at the end of this section (i.e., at line 3)) and the remaining signal poles $z_2(k)$, $k = S_c + 1, \dots, S_2$ of x_2 (note that the latter operation is represented by the function "*diff_sp*" in the algorithmic outline of the SuTdNAl at the end of this section (i.e., at line 4)). Using the determined signal poles and the model function Eq. [3], it is straightforward to find the amplitudes $A_1(k)$, $k = 1, \dots, S_1$ and phases $\tilde{\phi}_1(k)$, $k = 1, \dots, S_1$ of x_1, simply by solving the following set of equations:

$$L_1 c_1 = x_1 \qquad [7]$$

with

$$L_1 \equiv \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1(1) & z_1(2) & \dots & z_1(S_1) \\ \vdots & \vdots & \vdots & \vdots \\ z_1(1)^{N-1} & z_1(2)^{N-1} & \dots & z_1(S_1)^{N-1} \end{bmatrix}.$$

Note that the latter operation is represented by the function

"ca" in the algorithmic outline of the SuTdNAl at the end of this section (i.e., at line 5). In a similar way, the amplitudes $A_2(k)$, $k = 1, \ldots, S_2$, and phases $\tilde{\phi}_2(i)$, $k = 1, \ldots, S_2$, of x_2 are found (note that the latter operation is represented by the function "ca" in the algorithmic outline of the SuTdNAl at the end of this section (i.e., at line 6)).

Let us now consider what changes when the above described subspace procedure is applied to the observed *noisy* signals x_{1n} and x_{2n} instead of to the noiseless signals x_1 and x_2. Obviously Eq. [5] is no longer satisfied exactly, for the simple reason that H_1 no longer has rank S_1. Nevertheless it is still possible to determine an *estimate* of the common subspace and thus an *estimate* of the common signal poles. To be more specific, in Appendix A, the noisy signals lead to an inconsistent set of equations [14] that will be solved in a Total Least Squares (TLS) sense (*13*). As in the noiseless case, the determination of the "uncommon" signal poles leads to a system of equations described by Eq. [20]. However, in the noisy case the latter system of equations is inconsistent and will be solved in a TLS sense. Finally, the set of equations [7]—used for the determination of the complex amplitudes—becomes inconsistent in the noisy case and will be solved in a Least Squares (LS) sense.

In order to understand the following procedure, it should be reminded that the S_c components with common signal poles can be divided into two sets. As mentioned at the beginning of this section, the spectral intensities (i.e., the amplitudes in A_1 and A_2) of the first set, which contains S_p components, will present proportional changes due to different cell concentrations while the spectral intensities of the second set, which contains $S_c - S_p$ components, will present nonproportional changes, which can most likely be attributed to the effects of a particular agent. If the signal intensities of the S_c common signal poles are considered, the vector $r \equiv A_2(1 : S_c, 1)./A_1(1 : S_c, 1)$, where "$./$" represents the pointwise division of two vectors, contains the ratios of the signal intensities. If the measurements would be noiseless, the normalization factor would occur several times as an entry of the vector r. However, if noisy measurements are considered, all entries of r will be different. Therefore the entries of r will be divided into classes, where membership of a class is determined by the closeness to the other members. The class with the largest number of members will then determine the normalization factor by taking the average of the different ratios of the members. Therefore, in order to find the correct normalization factor we proceed in the following way (note that the following operation is represented by the function "*findfactor*" in the algorithmic outline of the SuTdNAl at the end of this section (i.e., at line 7)). Set $r \equiv A_2(1 : S_c, 1)./A_1(1 : S_c, 1)$, where "$./$" represents the pointwise division of two vectors. Next, the first entry $r(1)$ is assigned to class 1, Next, we check if $r(2)$ lies close to $r(1)$ by evaluating $\|r(2) - r(1)\|/r(1) < 0.1$.[3] If this

is true, the second entry $r(2)$ is assigned to class 1 as well. If $r(2)$ is not close to $r(1)$, a new class, class 2 is created and the second entry of r is assigned to the latter class, ... In the end (i.e., when all entries of r have been considered) the class with the largest number of members determines the normalization factor, simply by taking the mean of the elements $r(i)$ that belong to the largest class. The entire procedure can be described in the following schematic way [SuTdNAl(S_1, S_2, S_c)]:

1. **begin**
2. $(\alpha_{1,2}(s), \nu_{1,2}(s), s = 1, \ldots, S_c)$:=
   common_sp(x_1, x_2);
   /*compute the common signal poles of x_1 and x_2*/
3. $(\alpha_1(s), \nu_1(s), s = S_c + 1, \ldots, S_1)$:=
   diff_sp($x_1, S_1, \alpha_1(s), \nu_1(s), s = 1, \ldots, S_c$)
   /*compute the remaining signal poles of x_1*/
4. $(\alpha_2(s), \nu_2(s), s = S_c + 1, \ldots, S_2)$:=
   diff_sp($x_2, S_2, \alpha_2(s), \nu_2(s), s = 1, \ldots, S_c$)
   /*compute the remaining signal poles of x_2*/
5. $(A_1(s), \tilde{\phi}_1(s), s = 1, \ldots, S_1)$:=
   ca($x_1, \alpha_1(s), \nu_1(s), s = 1, \ldots, S_1$)
   /*compute the complex amplitudes of x_1*/
6. $(A_2(s), \tilde{\phi}_2(s), s = 1, \ldots, S_2)$:=
   ca($x_2, \alpha_2(s), \nu_2(s), s = 1, \ldots, S_2$)
   /*compute the complex amplitudes of x_2*/
7. R_{est} :=findfactor($A_1(s), A_2(s), s = 1, \ldots, S_c$)
   /*determine the normalization factor*/
8. **end**.

2.2. Experimental Results

In this section the performance of the SuTdNAl is analyzed. The SuTdNAl calculates an estimate R_{est} for the normalization factor. As a performance measure of SuTdNAl the absolute percent bias (bias) and the variance (var) are calculated,

$$\text{bias}(R_{est}) \equiv 100|E[R_{est}] - R|/R$$

$$\text{var}(R_{est}) \equiv E[(R_{est} - E[R_{est}])^2],$$

where $E[.]$ represents the expectation operator.

In order to compare similar spectra we define two properties of spectra: the degree of diversity (Dod) and the signal to noise ratio (SNR). The degree of diversity of two spectra is defined as the percent ratio between the number of points that are not superimposed and the total number of spectral points, when the two spectra are correctly normalized. In practice, for these simulation examples, the degree of diversity is calculated as follows. First the FIDs are zero-padded to 32,768 points and then Fourier transformed. Next the noise level of x_{1n} is estimated by calculating the standard deviation σ_1 of the last 1001 points of f_{1n}, the real part of the Fourier transform of the zero-padded FID x_{1n}. In a similar way σ_2 is determined on the basis of f_{2n}, the real part of the Fourier transform of x_{2n}. Next the signal d is

---

[3] Note that 0.1 is a user-defined tolerance.

calculated:

$$d = |x_{2n} - Rx_{1n}|.$$

The percentage of points for which $d > \sigma_1 + \sigma_2$ is then the degree of diversity (Dod).

Furthermore, the SNR is calculated as follows:

$$\text{SNR} = \min(\max(f_{1n})/\sigma_1, \max(f_{2n})/\sigma_2).$$

In the first section we compare the statistical performance of the SuTdNAl to the statistical performance of the MiRaNAl. The following section elaborates on the computational efficiency of the SuTdNAl compared to the MiRaNAl. Since it is hard to estimate the necessary model orders, we investigate in a separate subsection the dependence of the result of the algorithm on the chosen model orders. Next, the new method is tested by comparing spectra of two samples of known contents using the SuTdNAl, the MiRaNAl and an older method using a standard. The latter example allows to determine how SuTdNAl performs on measured signals under realistic noise conditions. The example is rather simple in the sense that it contains only one metabolite of which the intensities change proportionally. Considering an experiment in which more than just one of the metabolites has proportionally changing amplitudes would complicate the assessment of the performance of the algorithm since part of the bad performance might be due to the fact that it is very hard to make all the ratios experimentally exactly the same. Therefore the last subsection describes a simulation example derived from an *in vitro* measurement on cells.

2.2.1. Comparison of Statistical Performance

Since the MiRaNAl statistically outperforms the MaSNAl only the former algorithm is involved in this comparison. For several simulation signals, Monte Carlo simulations with different noise levels consisting of 30 runs each, were performed, yielding spectra with different SNRs and Dods. In Figs. 1 and 2 the results are summarized for spectra with Dod $\approx 59\%$. Figure 1 shows the absolute percent bias of the results obtained by the MiRaNAl[4] (circles) and the SuTdNAl (triangles), whereas Fig. 2 shows the variance of the results obtained by the MiRaNAl (circles) and the SuTdNAl (triangles). In terms of absolute percent bias, the MiRaNAl and the SuTdNAl have a similar performance. Furthermore, the variance of the estimates obtained with the SuTdNAl is similar to the variance of the estimates obtained with the MiRaNAl.

2.2.2. Comparison of Computational Performance

A first difference observed upon comparing the computational efficiency of the MiRaNAl and the SuTdNAl is the fact that the
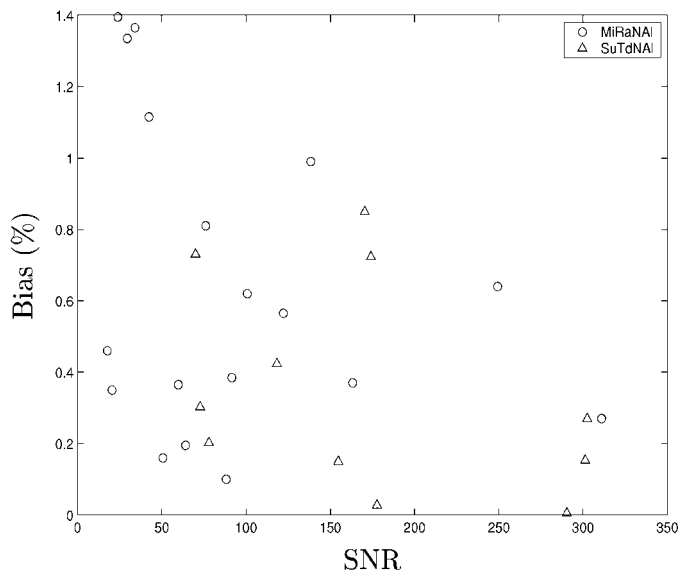

FIG. 1. This figure illustrates the absolute percent bias versus the SNR (mean degree of diversity $= (59 \pm 3)\%$).

MiRaNAl is an iterative method as opposed to the SuTdNAl. This is a consequence of the fact that the MiRaNAl determines the optimal normalization factor R by calculating the singular value spectra of several $\lceil N/2 \rceil \times (N - \lceil N/2 \rceil + 1)$ matrices (where $\lceil . \rceil$ rounds to the nearest integer towards $+\infty$), i.e., matrices obtained for different values of the normalization factor R. The normalization factor R that results in the matrix with the lowest numerical rank is the solution obtained by the MiRaNAl.
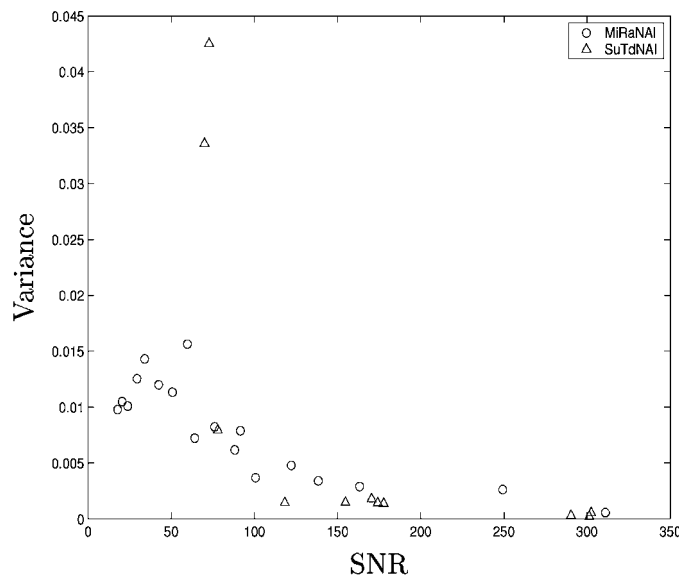

FIG. 2. This figure illustrates the variance versus the SNR (mean degree of diversity $= (59 \pm 3)\%$).

---

[4] These results correspond to the results shown in Fig. 5 of (*10*).

TABLE 1
Computationally Most Important Parts

Matrix size $(L = \lceil N/2 \rceil, M = N - L + 1)$	Formula	Line no. in SuTdNAl
$L \times 2M$	Eq. [6]	2
$L \times (2M - S_1 - S_2 + S_c)$	Eq. [11]	2
$L \times (M - S_c)$	Eq. [18]	3
$L \times (M - S_c)$	Eq. [18]	4

The latter implies that

1. A good starting value for R has to be available and
2. Depending on the required accuracy of R, many SVD calculations are necessary.

As far as the SuTdNAl is concerned, most of the computational effort is related to the calculation of four SVDs. In Table 1 an overview is given of these SVDs, indicating the size of the involved matrices, the corresponding formula in this paper as well as the corresponding line in the algorithm description of the SuTdNAl in Section 2.

2.2.3. Influence of Model Orders on Performance of SuTdNAl

As can be deduced from the algorithm outline in one of the previous subsections, the model orders S_1, S_2 and S_c have to be provided by the user. Furthermore, with closely spaced signal poles it is extremely difficult to determine these numbers exactly from the noisy measurements x_{1n} and x_{2n}. Therefore the influence of the choice of (S_1, S_2, S_c) on the performance of the SuTdNAl is studied in this section, using two simulated ^1H NMR FIDs starting from the model function in Eq. [3], Eq. [4], $N = 128$, $\Delta t = 0.000125$, $\eta_1 = 0.00030$, $\eta_2 = 0.00035$, $\phi_{01} = 0.54$, $\phi_{02} = 0.35$, and the parameters in Table 2. For each noise level, 30 independent Gaussian noise realizations with standard deviation σ_n are constructed for signal x_{1n} and x_{2n}. As can be derived from Table 2, the normalization factor $R = 2$.

The results are summarized in Table 3. It is clear that overestimating the model orders by even more than 20% still yields low absolute percent biases, while the variance does not change considerably.

2.2.4. Comparison of Two Measured Spectra

In order to test the SuTdNAl experimentally, two samples of known contents were compared. Both samples contained thyrotropin-releasing factor (THR, Calbiochem, MW 362.4), deuterated methanol (CD_3OD 99.96%, Cambridge Isotope Laboratories), and sodium trimethylsilyl[2,2,3,3-d_4]propionate (TSP, 10 μmol/ml). The first sample (sample A) consisted of 4.5 mg of THR, 400 μl of CD_3OD, and 10 μl of TSP, while the second sample (sample B) consisted of 4.5 mg of THR, 600 μl of CD_3OD, and 10 μl of TSP.

Five ^1H NMR spectra were obtained, for each sample, using a Bruker DPX digital spectrometer operating at 300 MHz. The spectra were accumulated with a 90° flip angle pulse and 64 transients of 8K data points corresponding to a spectral window of $\pm$2097.3 Hz. The magnitude spectra of sample A and B are shown in Fig. 3.

A normalization algorithm allows, after normalization of the spectra, to obtain directly the correct percent differences of the non proportional signals. For example, let us suppose we want to obtain the percent difference ($\triangle_{CD_3OD}(\%)$) of CD_3OD relative to THR. Using the traditional method, i.e., measuring the concentrations of both THR and CD_3OD, in both spectra with respect to TSP, the results in Table 4 were obtained (10, 9). If the percent difference ($\triangle_{CD_3OD}(\%)$) of CD_3OD relative to THR of the sample A spectrum with respect to the sample B spectrum is to be computed, the following equation can be used,

$$\triangle_{CD_3OD}(\%) = 100 \frac{CD_3OD(A) - \frac{CD_3OD(B)}{THR(B)}THR(A)}{\frac{CD_3OD(B)}{THR(B)}THR(A)} \quad [8]$$

where $CD_3OD(A)$ is the CD_3OD concentration in spectrum A and the other symbols have similar meaning. In fact, $\triangle_{CD_3OD}(\%)$ is the difference in CD_3OD relative to THR concentration. The CD_3OD concentrations are proportional to those obtained by considering the CD_3OD residual peaks due to residual protons (7); thus, the concentrations obtained by the CD_3OD residual peaks in Table 4 can be used directly in Eq. [8]. By Eq. [8], using the concentrations in Table 4, obtained by the traditional TSP standard quantification (traditional method), a $\triangle_{CD_3OD}(\%) = (-39.4 \pm 8.7)\%$ percent difference

TABLE 2
FID1 and FID2 Parameters

i	$A_1(i)$	$A_2(i)$	$\alpha_1(i)$	$\alpha_2(i)$	$v_1(i)$	$v_2(i)$
1	17	34	369	369	−3500	−3500
2	10	20	327	327	−3450	−3450
3	40	80	177	177	−3335	−3335
4	36	72	913	913	−3150	−3150
5	30	60	123	123	−3100	−3100
6	15	30	651	651	−3000	−3000
7	20	40	236	236	−2900	−2900
8	12	24	950	950	−2750	−2750
9	25	50	139	139	−2000	−2000
10	33	66	210	210	−1500	−1500
11	35	70	178	178	1600	1600
12	15	30	555	555	1800	1800
13	23	46	710	710	2100	2100
14	35	70	275	275	2400	2400
15	22	44	345	345	2700	2700
16	35	70	455	455	2850	2850
17	39	78	700	700	3000	3000
18	27	54	825	825	3020	3020
19	35	15	765	865	3100	2900
20	14	50	455	555	3500	3400

<div align="center">

TABLE 3
Influence of Model Orders S_1, S_2, S_c (Exact Model Orders are $S_1 = 20$, $S_2 = 20$, $S_c = 18$)

</div>

	$(S_1, S_2, S_c)=$	(20,20,18)	(22, 22, 18)	(24, 24, 22)	(24, 24, 20)	(26, 26, 20)	(26, 26, 24)
$\sigma_n = 0.6$, $SNR = 419.57$, $Dod = 68.49\%$	bias	0.21%	0.22%	0.09%	0.58%	0.45%	0.07%
	variance	3.91e-04	2.91e-04	4.60e-04	8.60e-04	6.12e-04	5.32e-04
$\sigma_n = 0.8$, $SNR = 261.59$, $Dod = 59.27\%$	bias	0.11%	0.33%	0.20%	0.55%	0.33%	0.25%
	variance	2.62e-04	8.28e-04	9.75e-04	7.43e-04	5.27e-04	1.10e-03
$\sigma_n = 1.2$, $SNR = 151.00$, $Dod = 53.80\%$	bias	0.18%	0.49%	0.67%	0.58%	0.79%	0.16%
	variance	7.10e-04	8.86e-04	2.34e-03	7.88e-04	2.08e-03	1.19e-03
$\sigma_n = 1.4$, $SNR = 169.09$, $Dod = 55.67\%$	bias	0.40%	0.36%	0.51%	0.34%	0.60%	0.39%
	variance	6.26e-04	1.20e-03	2.68e-03	1.01e-03	2.82e-03	1.62e-03
$\sigma_n = 1.6$, $SNR = 144.75$, $Dod = 52.09\%$	bias	0.13%	0.35%	0.34%	0.02%	0.05%	0.05%
	variance	8.76e-04	1.37e-03	1.69e-03	1.80e-03	1.21e-03	1.49e-03
$\sigma_n = 1.8$, $SNR = 108.17$, $Dod = 52.69\%$	bias	0.04%	0.23%	0.47%	0.46%	0.10%	0.03%
	variance	2.83e-03	2.52e-03	1.19e-03	1.41e-03	1.18e-03	1.24e-03
$\sigma_n = 2.2$, $SNR = 141.78$, $Dod = 46.96\%$	bias	0.75%	0.42%	0.11%	0.16%	0.39%	0.02%
	variance	1.63e-03	2.78e-03	3.26e-03	1.75e-03	2.19e-03	3.13e-03
$\sigma_n = 2.4$, $SNR = 98.03$, $Dod = 45.52\%$	bias	0.25%	0.25%	0.42%	0.15%	0.56%	0.42%
	variance	2.07e-03	2.97e-03	3.17e-03	3.84e-03	2.26e-03	1.92e-03
$\sigma_n = 2.6$, $SNR = 65.53$, $Dod = 51.07\%$	bias	0.70%	0.36%	0.15%	0.23%	0.13%	0.74%
	variance	3.49e-03	2.38e-03	4.73e-03	1.81e-03	3.79e-03	2.58e-03
$\sigma_n = 2.8$, $SNR = 70.23$, $Dod = 42.81\%$	bias	1.02%	0.24%	0.13%	0.74%	0.06%	0.40%
	variance	2.40e-03	3.69e-03	2.30e-03	4.75e-03	4.72e-03	5.06e-03
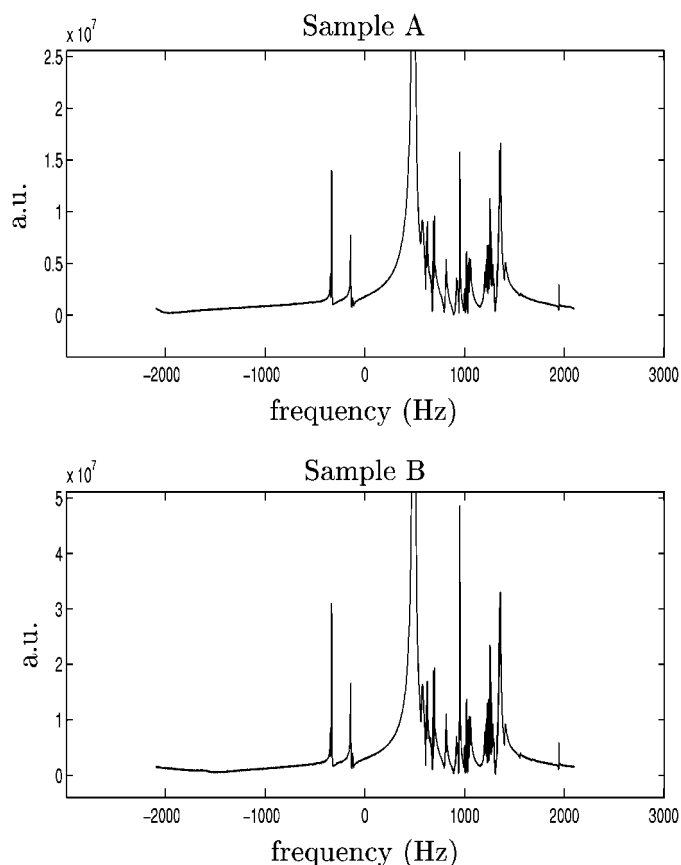
FIG. 3. This figure shows the magnitude spectra of sample A (top) and sample B (bottom).

was obtained (*10, 9*), while, from the known quantities of the added substances and using the same above-cited equation, a $\triangle_{CD_3OD}(\%) = (-33.4 \pm 0.2)\%$ difference was expected. The same two spectra (A, B) were utilized to obtain the percent difference $\triangle_{CD_3OD}(\%)$ of CD_3OD relative to THR by using the SuTdNAl. In particular, spectra were normalized with the algorithm and a normalization constant $R = 1.89$ was found. After normalization, the $\triangle_{CD_3OD}(\%)$ of CD_3OD relative to THR between the spectrum of sample A, normalized with respect to the spectrum of sample B, was obtained directly by comparing the areas of the two spectral CD_3OD residual peaks. The value found for $\triangle_{CD_3OD}(\%)$ was $\triangle_{CD_3OD}(\%) = (-33.0 \pm 1.4)\%$. It is interesting to observe that using the MiRaNAl results in $\triangle_{CD_3OD}(\%) = (-32.5 \pm 1.6)\%$ (*10*). As can be seen, all three methods, the traditional method, the SuTdNAl and the MiRaNAl, yielded percent difference results which were consistent with the expected ones. However, the SuTdNAl and the MiRaNAl allowed good results to be obtained without the use of any standard and without quantifying all the spectral lines, but rather by comparing the signals of interest in the normalized spectra.

<div align="center">

TABLE 4
Results Obtained Using a Standard

</div>

Sample	THR (μmol/ml)	CD_3DO (μmol/ml)
A	38.7 ± 3.6	41.9 ± 3.9
B	20.2 ± 2.5	36.1 ± 4.6

2.2.5. Simulated in Vitro Cell Spectra

For this section an *in vitro* measurement of a cell spectrum was first acquired, followed by a quantitation with VARPRO (*12*). Next, the quantified parameters of the 58 components are used to reconstruct a first signal s_1 using a Lorentzian lineshape model. A second signal s_2 is composed of 58 components as well. The latter components are constructed as follows. A first set of 32 components is constructed using exactly the same parameters as the corresponding components of s_1 with the exception that the amplitudes are scaled by a factor $R = 2$ (these are the intensities that do change proportionally). A second set of 21 components is constructed using exactly the same parameters as the corresponding components of s_1 with the exception that the amplitudes differ randomly from their counterparts in s_1 (these are the intensities that do not change proportionally). A third set of components have amplitudes, dampings, and frequencies that differ randomly from their counterparts in s_1. The noiseless signals s_1 (thick line) and s_2 divided by $R = 2$ (thin line) are shown in Fig. 4. At several noise levels Monte Carlo simulations involving the SuTdNAl are performed. For each noise level, 30 independent Gaussian noise realizations with standard deviation σ_n are constructed for signal s_1 and s_2. As mentioned before the normalization factor $R = 2$. To get an idea of the different noise levels, Fig. 5 shows 4 instances of the real part of the spectra of the noisy signal s_1 with noise levels $\sigma_n = 190, 285, 380$, and 570 (from bottom to top).

The results are summarized in Table 5. Note that Table 5 actually contains the results for $(S_1, S_2, S_3) = (58, 58, 53)$ and for $(S_1, S_2, S_3) = (100, 100, 100)$. The first triplet corresponds to the
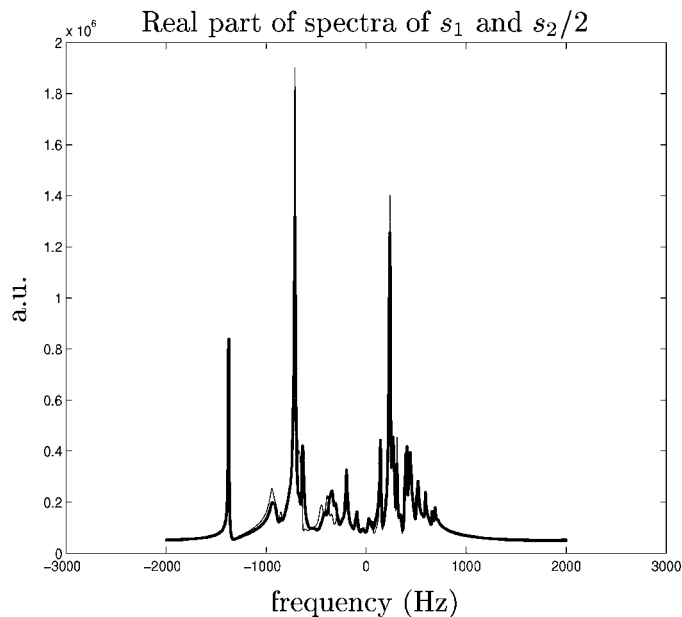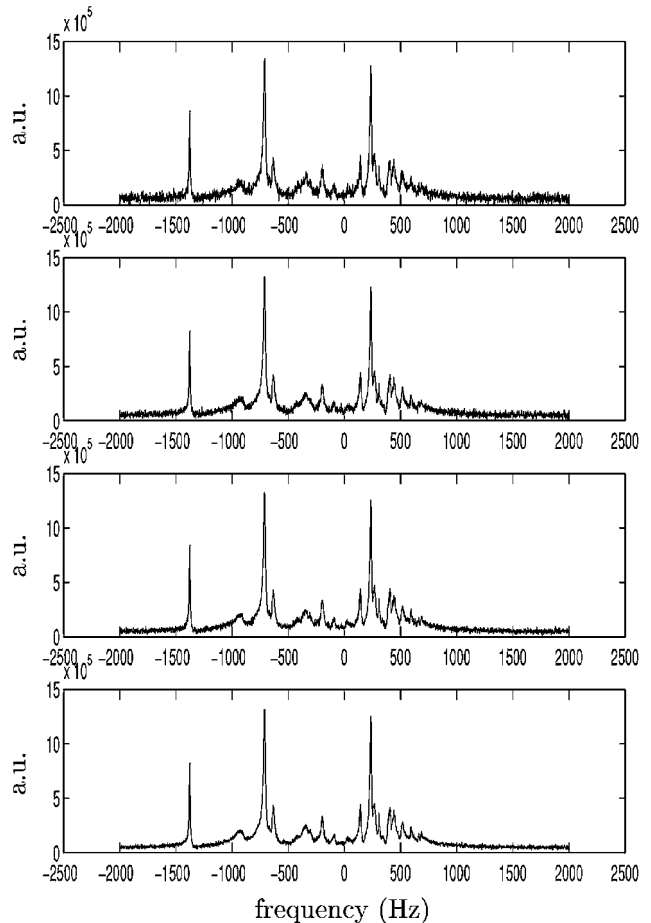

FIG. 5. This figure shows the real part of the spectrum of s_1 with additive noise of standard deviation $\sigma_n = 190, 285, 380, 570$ (from bottom to top).

correct model orders since the signal contains 53 common signal poles (of which only 32 have proportionally changing amplitudes). The second triplet corresponds to a large overestimation of the model orders. This case is included since it is often hard to obtain exact model order estimates starting from the noisy data. However, using prior knowledge and/or the SVD spectrum of the data matrices it is always possible to obtain a conservative

FIG. 4. This figure shows the real part of the spectra of s_1 (thick line) and of s_2 divided by $R = 2$ (thin line).

TABLE 5
Results for Simulated Cell Spectra (Exact Model Orders are $S_1 = 58, S_2 = 58, S_c = 53$)

	$(S_1, S_2, S_c)=$	$(58, 58, 53)$	$(100, 100, 100)$
$\sigma_n = 190, SNR = 156.83,$	bias	1.06%	0.95%
$Dod = 49.44\%$	variance	8.56e-04	2.78e-02
$\sigma_n = 285, SNR = 104.66,$	bias	0.73%	1.76%
$Dod = 47.78\%$	variance	2.36e-03	6.53e-02
$\sigma_n = 380, SNR = 78.08,$	bias	0.97%	3.7%
$Dod = 46.15\%$	variance	4.28e-03	9.08e-02
$\sigma_n = 570, SNR = 52.69,$	bias	4.02%	11.39%
$Dod = 44.18\%$	variance	3.04e-02	2.08e-01

upperbound of the model orders. The second triplet corresponds to such a conservative upper bound. It also illustrates a rule of thumb that can be used in case the model orders S_1, S_2, and S_c are difficult to determine: simply set them equal to one and the same overestimated value. The SuTdNAl then determines many common signal poles and no uncommon signal poles, yielding many poles that will model the noise. However, these "spurious" poles do not decrease the performance of the SuTdNAl too much since their amplitudes will in general not vary proportionally and thus they will not influence the result of the function "*findfactor*" (see the outline of the SuTdNAl, line 7). They do play a role in the function "*common_sp*" (see the outline of the SuTdNAl, lines 5 and 6) since they influence the amplitudes of the "true" signal poles, but as will be illustrated below, this influence remains small at normal noise levels.

It is clear that for noise levels that are typical for in vitro measurements (i.e., $\sigma_n = 190$) the SuTdNAl performs well both for the correct model orders and the overestimated model orders. When the noise standard deviation is doubled, the SuTdNAl still performs well for the correct model orders while the accuracy starts decreasing for the overestimated model orders. Only when the noise is tripled also the accuracy of SuTdNAl applied with the correct model orders starts decreasing.

3. CONCLUSION

In this paper a new subspace-based time-domain normalization algorithm, the SuTdNAl, is presented. It is shown that the SuTdNAl has a statistical performance similar to that of a previously presented normalization method: the MiRaNAl. However, the newly presented method is not iterative and thus is computationally less intensive. Furthermore, no initial guess of the normalization factor is required.

APPENDIX A

Determination of Common Signal Poles

In this appendix it is shown how the S_c common signal poles can be extracted from the column space of H_c, i.e., the common column space of H_1 and H_2, when both H_1 and H_2 contain respectively the noiseless signal x_1 (see Eq. [3]) and x_2 (see Eq. [4]).

If $z(1), z(2), \ldots, z(S_c)$ represent the common signal poles, it follows that

$$\mathcal{R}(H_c) = \mathcal{R}(O), \; O \in \mathbb{C}^{L \times S_c}$$

with

$$O = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z(1) & z(2) & \ldots & z(S_c) \\ \vdots & \vdots & \vdots & \vdots \\ z(1)^{L-1} & z(2)^{L-1} & \ldots & z(S_c)^{L-1} \end{bmatrix} \quad [9]$$

and thus

$$\exists B \in \mathbb{R}^{S_c \times (2M - S_1 - S_2 + S_c)} \quad \text{such that } H_c = O B. \quad [10]$$

Furthermore let

$$H_c = U S V^H \quad [11]$$

be the truncated SVD of H_c, with $U \in \mathbb{C}^{L \times S_c}$, $S \in \mathbb{C}^{S_c \times S_c}$, $V \in \mathbb{C}^{S_c \times (2M - S_1 - S_2 + S_c)}$. It then follows from Eq. [10] and Eq. [11] that

$$\exists \Xi \in \mathbb{C}^{S_c \times S_c} : U = O \Xi \text{ and } S V^H = \Xi^{-1} B, \quad [12]$$

where Ξ is a nonsingular matrix. From Eq. [9] it is clear that the column space of H_c is shift-invariant, i.e.,

$$\underline{O} Z = \bar{O}, \quad [13]$$

where $\underline{O}$ is the matrix O with its bottom row deleted, $\bar{O}$ is the matrix O with its top row deleted and Z is a diagonal matrix with $z(1), z(2), \ldots, z(S_c)$ on its diagonal. From Eq. [12] and Eq. [13] it then follows that

$$\underline{U} W = \bar{U}, \quad [14]$$

where $W \equiv \Xi^{-1} Z \Xi \in \mathbb{C}^{S_c \times S_c}$, $\underline{U}$ is the matrix U with its bottom row deleted and $\bar{U}$ is the matrix U with its top row deleted. Since by the definition of W, Z and W are related by a similarity transformation, the eigenvalues of Z and W are equal and thus the common signal poles of the noiseless signals x_1 and x_2 can be determined from the left singular vectors corresponding to the S_c largest singular values of H_c by solving [14] followed by the determination of the eigenvalues of the solution.

APPENDIX B

Determination of Uncommon Signal Poles

In this appendix it is shown how—after having determined the common signal poles of the noiseless signals x_1 and x_2—the "uncommon" (i.e., the signal poles contained in x_1 that are different from the common signal poles) signal poles of x_1 (and in a similar way those of x_2) can be determined.

Assume that the signal poles of x_1 are ordered in such a way that $z(i), i = 1, \ldots, S_c$, are the S_c common signal poles. Equation [5] can be written as

$$H_1 = [O_1' \; O_1''] \begin{bmatrix} D_1' & 0 \\ 0 & D_1'' \end{bmatrix} [C_1' \; C_1'']^T = O_1' D_1' C_1'^T + O_1'' D_1'' C_1''^T, \quad [15]$$

where $O_1' \in \mathbb{C}^{L \times S_c}$, $O_1'' \in \mathbb{C}^{L \times (S_1 - S_c)}$, $D_1' \in \mathbb{C}^{S_c \times S_c}$, $D_1'' \in \mathbb{C}^{(S_1 - S_c) \times (S_1 - S_c)}$, $C_1' \in \mathbb{C}^{M \times S_c}$, $C_1'' \in \mathbb{C}^{M \times (S_1 - S_c)}$. Let the QR factorization of C_1' be represented as

$$C_1' = [Q_1 \ Q_2][R^T \ 0]^T,$$

with $Q_1 \in \mathbb{C}^{M \times S_c}$, $Q_2 \in \mathbb{C}^{M \times (M - S_c)}$, $R \in \mathbb{C}^{S_c \times S_c}$. Therefore it follows that

$$C_1'^T Q_2^* = R^T Q_1^T Q_2^* = R^T \left(Q_1^H Q_2 \right)^* = 0,$$

where $(\ )^*$ stands for the complex conjugate. It is thus clear that

$$H_1 Q_2^* = O_1'' D_1'' C_1''^T Q_2^*. \qquad [16]$$

From Eq. [16] it is clear that the column space of $H_1 Q_2^*$ is shift-invariant meaning that

$$\underline{O_1''} Z'' = \overline{O_1''}, \qquad [17]$$

where $\underline{O_1''}$ is the matrix O_1'' with its bottom row deleted, $\overline{O_1''}$ is the matrix O_1'' with its top row deleted and Z'' is a diagonal matrix with $z_1(i)$, $i = S_c + 1, \ldots, S_1$, i.e., the uncommon signal poles, on its diagonal. Let

$$H_1 Q_2^* = U S V^H \qquad [18]$$

be the truncated SVD of $H_1 Q_2^*$, with $U \in \mathbb{C}^{L \times (S_1 - S_c)}$, $S \in \mathbb{R}^{(S_1 - S_c) \times (S_1 - S_c)}$ and $V \in \mathbb{C}^{(M - S_c) \times (S_1 - S_c)}$. From Eq. [16] and Eq. [18] it follows that

$$\exists \Xi : U = O_1'' \Xi \quad \text{and} \quad SV^H = \Xi^{-1} D_1'' C_1''^T Q_2^* \qquad [19]$$

with $\Xi \in \mathbb{C}^{(S_1 - S_c) \times (S_1 - S_c)}$ a regular matrix. Combining Eq. [17] and Eq. [19] yields

$$\underline{U} W'' = \bar{U}, \qquad [20]$$

where $W'' \equiv \Xi^{-1} Z'' \Xi \in \mathbb{C}^{(S_1 - S_c) \times (S_1 - S_c)}$, $\underline{U}$ is the matrix U with its bottom row deleted, and $\bar{U}$ is the matrix U with its top row deleted. Since by the definition of W'', Z'' and W'' are related by a similarity transformation, the eigenvalues of Z'' and W'' are equal and thus the uncommon signal poles of the noiseless signals x_1 can be determined from the left singular vectors corresponding to the $S_1 - S_c$ largest singular values of $H_1 Q_2^*$ by solving Eq. [20] followed by the determination of the eigenvalues of the solution.

REFERENCES

1. H. Bachelard and R. Badar-Goffer, NMR spectroscopy in neurochemistry, *J. Neurochem.* **61**(2), 412–429 (1993).

2. S. J. Berners-Price, M. E. Sant, R. I. Christopherson, and P. W. Kuchel, ^{1}H and ^{31}P NMR and HPLC studies of mouse l1210 leukemia cell extracts: The effect of Au(l) and Cu(l) diphosphine complexes on the cell metabolism, *Magn. Reson. Med.* **18**, 142–158 (1991).

3. S. Cérdan, R. Parrilla, J. Santoro, and M. Rico, ^{1}H NMR detection of cerebral myo-inositol, *FEBS Lett.* **187**, 167–172 (1985).

4. H. Chen, S. Van Huffel, D. van Ormondt, and R. De Beer, Parameter estimation with prior knowledge of known signal poles for the quantification of NMR spectroscopy data in the time domain, *J. Magn. Reson. A*, 225–234 (1996).

5. H. Chen, S. Van Huffel, and J. Vandewalle, Improved methods for exponential parameter estimation in the presence of known poles and noise, *IEEE Trans. Signal Process.* **45**(5), 1390–1393 (1997).

6. Robin A. de Graaf, "In Vivo NMR Spectroscopy: Principles and Techniques," Wiley, New York, 1998.

7. A. E. Derome, "Modern NMR Techniques for Chemistry Research," Prentice–Hall, Englewood Cliffs, NJ, 1996.

8. C. Rémy, C. Arús, A. Ziegler, E. Sam Lai, A. Moreno, Y. Le Fur, and M. Décorps, *In vivo*, *ex vivo*, and *in vitro* one- and two-dimensional nuclear magnetic resonance spectroscopy of an intracerebral glioma in rat brain: Assignment of resonances, *J. Neurochem.* **62**, 166–179 (1994).

9. R. Romano, R. Lamanna, M. T. Santini, and P. L. Indovina, A new algorithm for NMR spectral normalization, *J. Magn. Reson.* **138**, 115–122 (1999).

10. R. Romano, M. T. Santini, and P. L. Indovina, A time-domain algorithm for NMR spectral normalization, *J. Magn. Reson.* **146**, (2000).

11. S. M. Ronen, E. Rushkin, and H. Degani, Lipid metabolism in large T47D human breast cancer spheroids: ^{31}P- and ^{13}C-NMR studies of choline and ethanolamine uptake, *Biochim. Biophys. Acta* **1138**, 203–212 (1992).

12. J. W. C. van der Veen, R. de Beer, P. R. Luyten, and D. van Ormondt, Accurate quantification of ^{31}P NMR signals using the variable projection method and prior knowledge, *Magn. Reson. Med.* **6**, 92–98 (1988).

13. S. Van Huffel and J. Vandewalle, "The Total Least Squares Problem: Computational Aspects and Analysis," Vol. 9, SIAM, Philadelphia, 1991.