NMR Solvent Peak Suppression by Piecewise Polynomial Truncated Singular Value Decomposition Methods

Daesung Kim, Hyekyoung Lee, Youngdo Won, Dai-Gyoung Kim,[*] Youngwoo Lee,[‡] and Hoshik Won[*]

Department of Chemistry, Hanyang University, Seoul 133-791, Korea
[*]Department of Applied Mathematics, Hanyang University, Ansan 425-791, Korea
[‡]R & D Center, Samsung Engineering Co., Ltd, Yongin 449-844, Korea
Received March 26, 2003

A new modified singular value decomposition method. piecewise polynomial truncated SVD (PPTSVD), which was originally developed to identify discontinuity of the earth's radial density function, has been used for large solvent peak suppression and noise elimination in nuclear magnetic resonance (NMR) signal processing. PPTSVD consists of two algorithms of truncated SVD (TSVD) and L_1 problems. In TSVD, some unwanted large solvent peaks and noise are suppressed with a certain soft threshold value. whereas signal and noise in raw data are resolved and eliminated in L_1 problems. These two algorithms were systematically programmed to produce high quality of NMR spectra. including a better solvent peak suppression with good spectral line shapes and better noise suppression with a higher signal to noise ratio value up to 27% spectral enhancement, which is applicable to multidimensional NMR data processing.

Key Words : Nuclear magnetic resonance. Piecewise polynomial truncated singular value decomposition method. Peak suppression

Introduction

Most NMR spectroscopists are interested primarily in very tiny resonance peaks giving molecular information. although distinguishing real signal is not feasible due to the serious overlap with noise. Efforts to obtain the exact information of NMR signals are easily disturbed by typical noise caused by various reasons. such as imperfections in instrumentation and the condition of samples during the NMR data collection. Many useful techniques enhancing the signal to noise ratio by eliminating typical noise or unwanted huge solvent peaks have been developed over the last few decades. Window functions are used for their simplicity and powerful abilities of noise reduction. Although many window functions have been employed successfully in signal processing. there are serious limitations to noise reduction because a decrease in spectral resolution always occurs during the noise elimination process. On the other hand. other advanced methods have been developed to overcome this limitation. For example. some successful methods. including linear prediction (LP).[1,2] maximum entropy reconstruction (MaxEnt).[3,4] and maximum likelihood method (MLM).[5] have been widely used as an advanced tool of NMR signal processing.

As one of the advanced methods of signal enhancement. singular value decomposition (SVD) has been widely used for some time to obtain linear prediction (LP) coefficients[1] in the area of NMR signal processing. Because stable LP coefficients are easily predicted. conventional SVD methods were mainly used in solvent peak suppression. SVD became an important mathematical tool in identifying the pure signal

from the noise-contaminated signals. The resulting eigenvalues of SVD calculations provide direct information of signal intensities. Recently. noise elimination methods[6] and large solvent peak suppression methods have been developed by utilizing basic properties of SVD.[7,8]

Hansen developed a piecewise polynomial truncated singular value decomposition (PPTSVD) method and applied it to identification of the discontinuity of the earth's radial density function.[9] Hansen reported that PPTSVD can provide better results from signal processing compare with conventional SVD-based methods. The discontinuity obviously was identified from the earth's unclear original radial density signal. This PPTSVD method was applied to the advanced NMR data processing in obtaining a clear signal from the noise-contaminated signal and in the suppression of large solvent peaks in the present study. In addition. basic theories of SVD and PPTSVD are described here to simplify the mathematical method for developing algorithms.

Theoretical Background

Singular Value Decomposition (SVD). The SVD matrix A for a given time domain NMR data can be represented by $m \times n$ dimensional elements as in equation (1).

$$A = U\Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T \tag{1}$$

where U and V are the square matrices carrying $m \times m$ and $n \times n$ dimensional size. respectively. These matrices basically have an orthonormal property. *i.e.*. $U^T U = V^T V = I_n$. and are related to the phase of the matrix A. The upper capital letter T refers to the transverse. and sigma. $\Sigma = diag(\sigma_1, \sigma_2, \sigma_3, \cdots, \sigma_n)$. is a diagonal matrix whose diagonal elements are σ_i.
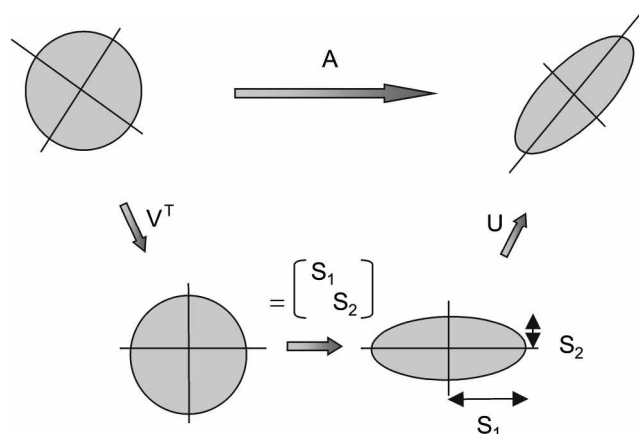
Figure 1. A graphical presentation of PPTSVD algorithm shows how matrices carrying unwanted peaks are operated. A resulting matrix of the operation of matrix A can be divided into three matrix operations, a phase correction (V^T), an adjustment of spectral magnitude (Σ), and finally an reorientation to the original phase (U).

The singular values implying the magnitude of the elements of matrix A satisfy the following condition $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$. The phase of the original NMR data can be changed by the operation of matrixes U and V. In addition, proper adjustment of the matrix Σ enables the suppression of unwanted signals as shown in Figure 1. An arbitrary phase of the NMR signal can be changed with V^t matrix, and, subsequently, spectral magnitude can be adjusted with Σ matrix. The final reorientation to the original phase can be achieved with the matrix U operation. In other words, the operational components of A can be manipulated by three decomposed matrix operations. To suppress large unwanted signals and small noises, the matrix Σ has to be properly modified.

Truncated Singular Value Decomposition (TSVD). In this method, one simply ignores all the small singular values that originated from noise and approximates a rank-n full matrix A_n by a rank k matrix in which only the largest k singular values are retained.

$$A_k = \Sigma u_i \delta_i v_i^T \quad \text{where } k < n \tag{2}$$

In this way, matrix A_n is replaced by A_k, which has a well defined null space of dimension n-k spanned by the right singular value vectors, $v_{k-1}, \ldots, v_n$. The original linear system of equation (2) is then replaced by the following problem set of equation (3), where b is ideal noise-free data obtained at the minimized point. The resulting TSVD solution of equation (3) is given by x_k and equation (4).

$$\min \|x\|_2 \text{ is subjected to } \|A_x x - b\|_2 = \min \tag{3}$$

$$x_k = \sum_{i=1}^{k} \frac{u_i^T b}{\delta_i} v_i \tag{4}$$

Piecewise Polynomial Truncated Singular Value Decomposition (PPTSVD). As in TSVD, the PPTSVD method

solves the problem[9,10] of equation (5), where L is a discrete approximation to a derivative operator that is represented as a matrix notation (6).

$$\min \|Lx\|_1 \text{ is subjected to } \|A_x x - b\|_2 = \min \tag{5}$$

$$L_1 = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \tag{6}$$

The solution of equation (5) is obtained by solving the equation (7), giving values of w:

$$\min \|(LV_k)w - Lx_k\| \tag{7}$$

The resulting PPTSVD solution $(\tilde{x}_k)$ consists of the TSVD solution, x_k and a modification factor, $-V_k w_k$ term as illustrated in equation (8).

$$\tilde{x}_{L,k} = x_k - V_k w_k \tag{8}$$

A self-consistent value for the solution of equation (8) is accomplished at that point; an extra norm, $\|\tilde{x}_{L,k}\|_2$, equals to the square root of the value of TSVD term and value of w as in equation (9).

$$\|\tilde{x}_{L,k}\|_2 = (\|x_k\|_2^2 + \|w_k\|_2^2)^{1/2} \tag{9}$$

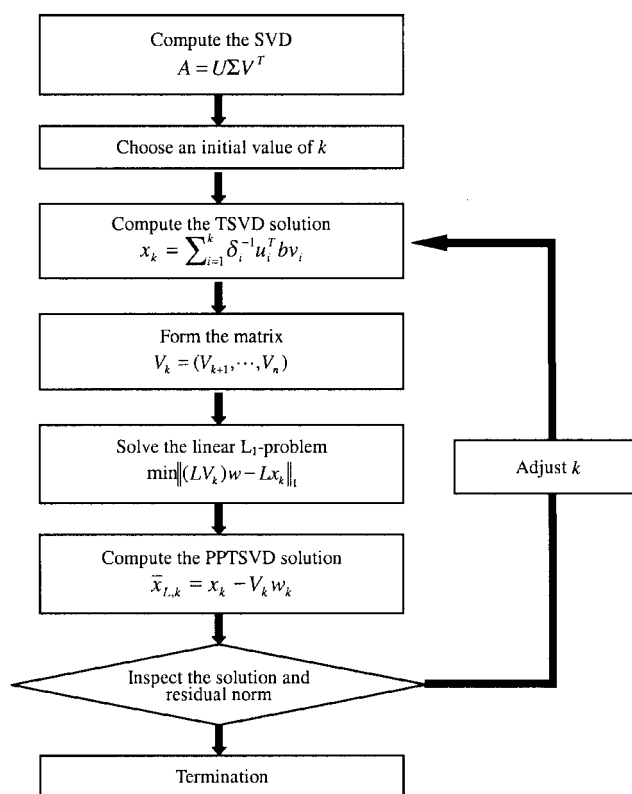The important mathematical expressions necessary for the

Figure 2. A flow diagram of PPTSVD calculations applying to the advanced NMR signal processing.

coding and application of PPTSVD method to NMR data processing are summarized as a flow chart diagram in Figure 2.

In conventional SVD methods, the first step of computation is to build a Toeplitz matrix carrying the size of $m \times (n - m + 1)$ dimension, where m is an integer rank from the frequency domain linear NMR data set having n data points. The diagonal elements of Σ matrix are manipulated after the execution of the SVD calculation. First few points are must be replaced with zero to suppress the large solvent peak. Toeplitz matrix having solvent peaks suppressed is then reconstructed by a matrix product operation. Noise suppression is accomplished by assuming that the last few points close to zero, and by calculating the average value and standard deviation of signals. The other parts of eigenvalues are modified by the calculated standard deviation value to obtain a noise free signal. The noise-eliminated NMR signals can be obtained by filling zero to the noise region and, subsequently, by reconstructing the matrix.

In PPTSVD methods, the algorithms solve TSVD and L_1-problem to determine the optimal solution of the equation. Figure 2 illustrates the detail procedure of stepwise calculation methods. The first step is to solve TSVD calculation. The second step is to find the solution of equation (7) that gives the modification factor with the results of the first step. The third step is to calculate the exact solution with the modification factor obtained from the previous step.

Experimental Section

10 mM zinc binding leutenizing hormone releasing hormone (Zn-LHRH) dissolved in DMSO-d6/H$_2$O mixed solvent was used as the NMR sample. The large HOD water peak at off resonance region was used to test current NMR data processing. NMR data of complex 2048 data points were recorded with a Varian UNITY plus 500 MHz NMR spectrometer. The first free induction decay (FID) of 2D NOESY data was used to test 1D version of algorithm. FID was converted to HyNMR (in house program) format and double precision text format. All of source coding works were accomplished under the IRIX C/C++ compiler of an SGI octane workstation.

Results and Discussion

Toeplitz matrix was constructed with a rank of 100 for the complex 2048 NMR data points. The magnitude of spectrum was plotted as a function of rank and sums of diagonal Σ matrix elements in Figure 3 after PPTSVD calculation. The curve exhibits an exponential decay that can be classified into three regions. In the first region, large Σ matrix values are located in the front four points. A typical exponential decay appeared in the second region of 5 to 80 points, whereas the magnitudes of Σ matrix are close to zero in the third region of 81 to 100 points. The diagonal elements of Σ matrix correspond to the magnitude of each peak of the spectrum.

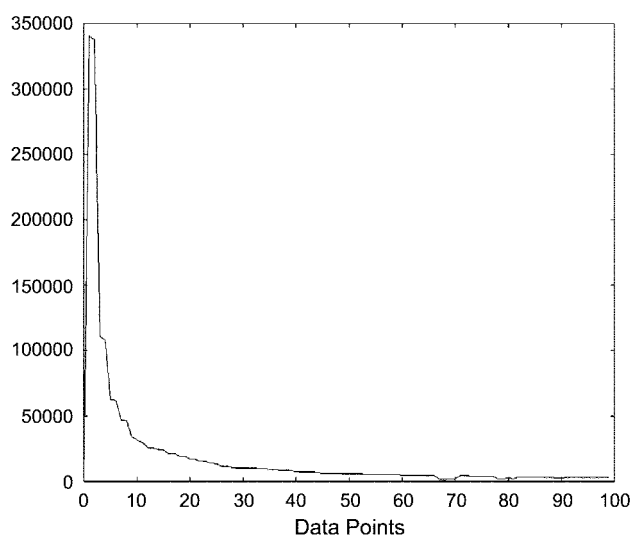The large four points in front represent a huge HOD water

Figure 3. The magnitude of NMR spectrum was plotted as function of rank and sums of diagonal Σ matrix elements. The large front four points correspond to a solvent peak whereas the range of 81 to 100 points corresponds to noises.

solvent peak, so these points were zero filled. The product of matrices U and V^T produced a spectrum with the water peak eliminated. The third region at the last 20 points close to zero can be interpreted as magnitude of noise. Noise was suppressed by applying these selected noise points of the third region to all the matrix elements. Ranks were adjusted to 80 points, except the noise components, to resolve the magnitude and to produce a better spectrum in the current PPTSVD.

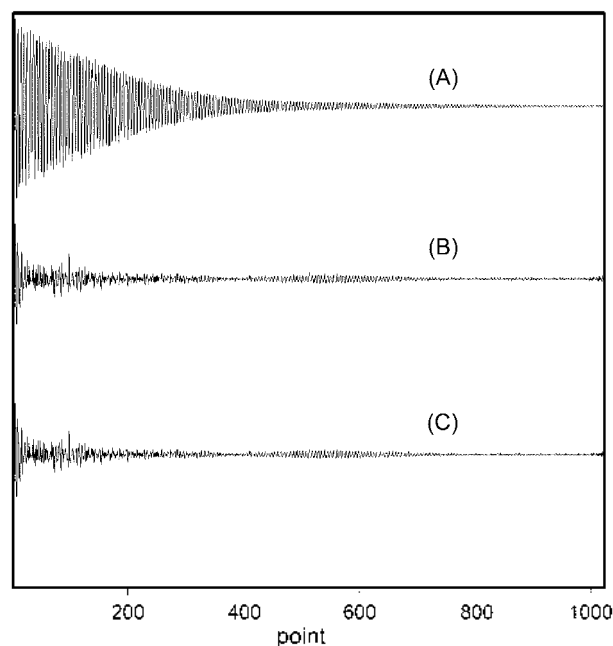Stack plots of the free induction decay curves correspond-

Figure 4. Free induction decay curves carrying a large water solvent peak (A), and after treatment of large peak suppression routine (B), and an enhanced NMR signal after the additional noise elimination procedure (C).
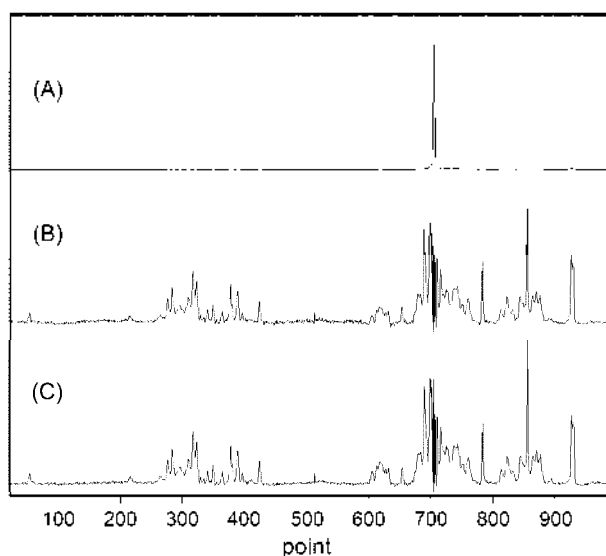
Figure 5. ¹H NMR spectrum of Zn-LHRH carrying a large water peak (A), and the spectrum after treatment of large peak suppression (B), and the spectrum after additional noise elimination procedure (C).

ing to three regions of Σ matrix elements are exhibited in Figure 4. Three free induction decay curves correspond to the raw NMR data (A) containing both huge solvent peak and noise peaks, the large solvent peaks suppressed (B), and the enhanced NMR signal after additional noise elimination (C), respectively. Stack plots of frequency domain NMR spectra after fourier transformation of the decay curves are illustrated in Figure 5. Solvent peak was completely suppressed with good spectral line shapes as in Figure 5(B), with an S/N ratio value of 165. In addition, good noise

suppression was accomplished as in Figure (C) with S/N ratio values of 210 exhibiting 27% of spectral enhancement. Conventional SVD gives approximately 15% of signal enhancement in noise reduction.[6-8]

SVD and solution of L_1-problems are in general time consuming procedures compared with other advanced processing methods as described in introduction. Solvent peak and noise suppression by PPTSVD took 445 seconds, compared with 370 seconds by conventional SVD. Although PPTSVD consumes 20% more time, it provides high quality NMR spectra, including better solvent peak suppression with good spectral line shapes and better noise suppression with higher S/N ratio value. Considering the results of our study, current methods may be applicable to multidimensional NMR data processing that needs better signal enhancement.

References

1. Domoulin. C. L.; Levy. G. C. *Bull. Magn. Reson.* **1984**, *6*, 47.
2. Barkhuijsen. H.; Beer. R. D.; Ormondt. D. V. *J. Magn. Reson.* **1986**, *67*, 371.
3. Schmieder. P.; Stern. A. A.; Wagner. G.; Hoch. J. C. *J. Magn. Reson.* **1997**, *125*, 332.
4. Hoch. J. C.; Stern. A. S. *NMR Data Processing.* Wiley-Liss Inc.; **1996**; p 102.
5. Umesh. S.; Donald. W. *IEEE Transactions on Signal Processing.* **1996**, *44*, 2245.
6. Fedrigo. M.; Fogolari. F.; Viglino. P.; Esposito. G. *J. Magn. Reson. Series B* **1996**, *113*, 160.
7. Zhu. G.; Smith, D.; Hua. Y. *J. Magn. Reson.* **1997**, *124*, 286.
8. Zhu. G.; Choy. W. Y.; Song. G.; Song, G.; Sanctuary. B. C. *J. Magn. Reson.* **1998**, *132*, 176.
9. Hansen. P. C. *Num. Lin. Alg. Appl.* **1996**, *3*, 513.
10. Hansen. P. C.; Sekii. T.; Shibahashi. H. *SIAM J. Sci. Stat. Comput.* **1992**, *13*, 1142.