Denoising applied to spectroscopies-part I: concept and limits

Guillaume Laurent, William Woelffel, Virgile Barret-Vivin, Emmanuelle Gouillart, Christian Bonhomme

Denoising applied to spectroscopies – part I: concept and limits

Guillaume LAURENT (1*), William WOELFFEL (2), Virgile BARRET-VIVIN (1), Emmanuelle GOUILLART (2), Christian BONHOMME (1)

(1) Sorbonne Université, Collège de France, CNRS, Laboratoire de Chimie de la Matière Condensée de Paris, LCMCP, F-75005, Paris, France.

(2) Saint-Gobain, CNRS, Surface du Verre et Interfaces, SVI, F-93300, Aubervilliers, France.

(1) Case Courrier 174, 4 place Jussieu, 75252 Paris Cedex 05, France.

(2) 39 quai Lucien Lefranc, BP 135, 93303 Aubervilliers Cedex, France.

(*) Corresponding author

guillaume.laurent@sorbonne-universite.fr, orcid.org/0000-0002-8127-5326

william.woelffel@saint-gobain.com

vigile_bv@msn.com

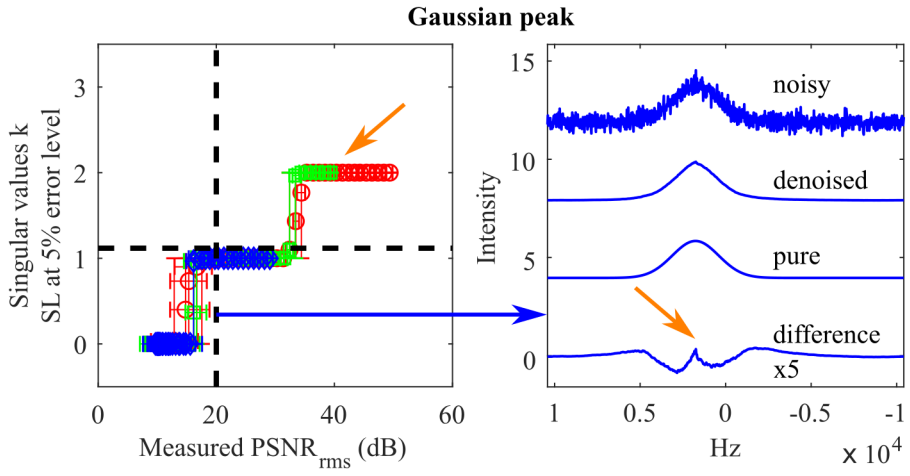
emmanuelle.gouillart@saint-gobain.com

christian.bonhomme@sorbonne-universite.fr, orcid.org/0000-0003-0802-6961

Denoising applied to spectroscopies – part I: concept and limits

Some spectroscopies are intrinsically poorly sensitive, such as Nuclear Magnetic Resonance (NMR) and Raman spectroscopy. This drawback can be overcome by using Singular Value Decomposition (SVD) and low-rank approximation to denoise spectra and consequently increase sensitivity. However SVD limits have not been deeply investigated until now. We applied SVD to NMR and Raman spectra and showed that best results were obtained with a square data set in time domain. Automatic thresholding was applied using Malinowski's indicators. 6×7380 noisy spectra with 41 signal-to-noise ratios were compared to their non-noisy counterparts, highlighting that SVD induces a systematic error for Gaussian peaks but faithfully reproduces shape of Lorentzian peaks, thus allowing quantification. Used carefully, SVD can decrease experimental time by a factor of 2.3 for spectroscopies. This study may help scientists to apply SVD to denoise spectra in a more efficient way, without falling into pitfalls.

Keywords: spectroscopy, sensitivity, signal processing, Cadzow denoising, Singular Value Decomposition (SVD)



Graphical abstract

A. Introduction

Spectroscopic techniques are of outmost importance in the field of materials analysis. Among them, Nuclear Magnetic Resonance (NMR) (1) and Raman (2) spectroscopies are very powerful local probes of the chemical structure. Especially, they allow to analyse liquid-state, solid-state, or even gas-state samples. While NMR informs about chemical and magnetic environment of atomic nuclei, Raman spectroscopy provides vibrational and rotational information on chemical entities. Unfortunately, these techniques suffer from a major drawback, namely their intrinsic low sensitivity. Although focused on NMR and Raman approaches, this work can be easily extended to other spectroscopies.

In the case of NMR, only one nucleus over 10^5 is detected under usual conditions (3). This is due to the low population difference between nuclear spin energy levels, which results from Boltzmann equilibrium. Many factors influence NMR sensitivity: magnetic field strength, sample volume, sample temperature, electronics temperature, radio-frequency coil quality factor and coil filling factor (4). When studying a solid-state sample, situation gets worse due to spectral line broadening, which results either from environment distribution or from relaxation (5). Indeed, NMR relevant anisotropic interactions, as chemical shift anisotropy, dipolar coupling and quadrupolar coupling, are no longer averaged to zero by fast and isotropic molecular motions, leading to spectra spreading over hundreds of ppm or a few megahertz (6). During the last decades, numerous technical progress has allowed to increase sensitivity of solid-state NMR: Magic Angle Spinning (MAS) up to 110 kHz (7–9), Cross Polarisation (CP) (10), high performance heteronuclear decoupling (11) and NMR magnetic field strength increase (12). Additionally, new very sensitive techniques were

developed, for instance micro-coils and microresonators adapted to MAS (13) or Dynamic Nuclear Polarisation (DNP) (14–16), with which a gain of up to 320 per unit of time was achieved (17).

In the case of Raman spectroscopy, only one photon over 10^6 is detected (18), due to its low scattering cross-section of $\approx 10^{-30}\,cm^2$ per molecule, to be compared with $\approx 10^{-20}\,cm^2$ for infrared absorption spectroscopy and $\approx 10^{-16}\,cm^2$ for fluorescence spectroscopy (19). Furthermore, some samples are fluorescent, thus hiding their Raman spectrum, or can be locally damaged by the laser light during analysis (20). A major advancement in Raman analysis has been achieved with the discovery of Surface-Enhanced Raman Scattering (SERS) (21, 22) with an enhancement factor of 10^4-10^6 (23). However, this effect applies only to transition metals (24) and strongly depends on surface roughness (25). To circumvent these limitations, Raman equipment has been hyphenated with an atomic force microscope or a tunnel effect microscope, leading to Tip-Enhanced Raman Spectroscopy (TERS) (26, 27). Another important progress has been achieved through the use of non-linear light sources and picosecond lasers (28, 29). Thanks to such improvements, it is now possible to study artworks (30), to map a surface (31) and even to follow cure kinetics of an epoxy resin (32).

In parallel with these instrumental and methodological developments, mathematical and computer tools have become increasingly widespread in the field of data processing (33). In addition to Fourier transform revolution (34, 35), other treatments have emerged in NMR spectroscopy and Magnetic Resonance Imaging (MRI). As examples, one may cite Hadamard transform (36), compressed sensing (37), non-uniform sampling (38), or quantitative signal reconstruction from multiple echoes (39). In Raman spectroscopy, pre-processing became of paramount importance to obtain

quantitative measurements (40): it is now recommended to suppress fluorescence background (41), to correct cosmic ray spikes (42) and to normalise spectra (43).

Furthermore, a very important mathematical tool family concerns noise reduction (44). Indeed, the above mentioned sensitivity-enhanced spectroscopic equipments are costly and are not accessible to all laboratories, hence the need to denoise. Moreover, even with hardware and methodological improvements, spectra can still be noisy, especially when studying amorphous materials (45, 46), for which distribution of bond lengths and bond angles broadens signals and reduce sensitivity. In order to decrease experimental time or in case of unstable samples, signal processing is mandatory to get a reasonable Signal-to-Noise Ratio (SNR). The easiest way to perform noise reduction is smoothing. In NMR, apodisation, especially exponential multiplication, is used prior to Fourier transform (47). In Raman spectroscopy, a polynomial algorithm named Savitzky–Golay is preferred (48). Other options to reduce noise are maximum entropy (49), covariance matrix (50), Wiener's estimation (51), wavelet transform (52), uncoiled random QR denoising (53), and the method initially proposed by Tufts *et al.* (54) and generalised by Cadzow (55).

Singular Value Decomposition (SVD) is an important part of Cadzow's denoising algorithm and its related low-rank approximation (56). An history of SVD can be found in (57). It has been discovered independently by Beltrami in 1873 (58), by Jordan in 1874 (59) and rediscovered by Lanczos in 1958 (60). It is preferred over eingenvalue decomposition, which is less precise, as demonstrated by Laüchli (61). Though quite old, Cadzow's procedure is currently a research domain of vivid interest (62), especially using sparse data, *i.e.* partially empty matrix (63). SVD is widely used in several domains including acoustics (64), geophysics (65), air quality (66), electrocardiograms (67), image compression (68), video surveillance (69), MRI (70),

data mining (71) and even on Facebook (72). Nevertheless, SVD is still not so commonly used in spectroscopies like NMR (73–76) and Raman spectroscopy(77–80), despite its use could significantly reduce experimental time and be of particular interest for the scientific community.

Recently, Man *et al.* developed a new SVD application for NMR (81). It was programmed under Java and two versions are currently available: one for processors (82) and the other one for Nvidia graphic cards (83) using CUDA (84). Indeed, graphic cards allow very efficient parallel computations. However, the limits of this approach are not clear: (i) which matrix shape should be preferred? (ii) what is the minimal experimental SNR? (iii) Are denoised spectra quantitative? (iv) is it suitable for other spectroscopies?

Following a previous communication (85), we tried to address these questions. This work is divided into two parts. In this first contribution (I), we focus on SVD concept and limits. Experimental details will be provided in Section B. Theoretical background on SVD and low-rank approximation concepts is developed in Section C.1. Hankel and Toeplitz matrices are explored in Section C.2. SNR definitions are given in Section C.3. Section D.1 is devoted to experimental results by applying SVD to solid-state NMR and Raman spectroscopies. The influence of both matrix shape and thresholding are studied in Sections D.2 and D.3, respectively. Time and frequency denoising are compared in Section D.4. The minimum SNR needed to have accurate results is investigated in Section D.5. The impact of SVD on peak shape is considered in Section D.6. Finally, denoising on a real NMR spectrum is analysed in Section D.7.

In a second part (II) (86) we will benchmark SVD using Java, Matlab and Python, on various processors and nvidia graphic cards ranging over 10 and 6 years,

respectively. We will try to optimise algorithms, software libraries and hardware capabilities to achieve the fastest possible denoising computation.

B. Materials and methods

B.1. Synthesis of the 50 / 50 MTEOS / TEOS sample

This sample is representative of typical materials obtained by sol-gel chemistry (87–89). This soft chemistry synthetic approach is a suitable route to design hybrid materials that contain both organic (methyltriethoxysilane, MTEOS, T species) and inorganic functions (tetraethylorthosilicate, TEOS, Q species), combining for instance hydrophobicity and high mechanical stability (90). T and Q stand for the number of oxygen on each silicon, namely Tri (3) and Quadri (4), for MTEOS and TEOS, respectively. The letter is associated with a superscript indicating the number of condensed Si-O-Si bridges. It is important to quantify the ratio T / Q and the condensation degree, mainly by ^{29}Si MAS NMR, in order to properly characterise such hybrid materials. This nucleus suffers from a low natural abundance of 4.7 % and an intermediate resonating frequency at 1 / 5th of ^{1}H one, both lowering SNR. It is thus current to average noise over one night or one weekend for a single spectrum. Using denoising is an interesting approach to decrease acquisition time.

Every chemical was used as received with no further purification. The solution was prepared by adding 10.18 g of MTEOS (98 %, Alfa Aesar; M = 178.30 g.mol^{-1}, 57.1 mmol) and 11.89 g of TEOS (>99 %, Aldrich; M = 208.33 g.mol^{-1}, 57.1 mmol), 29 mL of milliQ water and 50 mg of a 37 % w/w aqueous solution of hydrochloric acid (HCl, VWR; M = 36.46 g.mol^{-1}).

The solution was stirred at room temperature for at least one hour at 500 rpm to ensure hydrolysis of the precursors. Controlled condensation occurred during spray

drying of the sample, performed using a mini spray dryer B-290 (BUCHI) fitted with an atomiser (nozzle tip diameter = 0.7 mm) and a peristaltic pump. The temperatures at the inlet and outlet of the spray dryer were fixed at 220 °C and within the range of 95-120 °C, respectively. Polydisperse spherical particles of hybrid organic / inorganic amorphous silica (characteristic size: 1-10 µm) were obtained.

B.2. Solid-state NMR experiments

^{29}Si solid-state NMR experiments were performed on a Bruker Avance III spectrometer operating at 300.29 MHz for ^{1}H and 59.65 MHz for ^{29}Si with 4 mm zirconia rotors spun at 14 kHz (MAS broadband dual probe). Unless otherwise stated, CP was used with a contact time of 5 ms, a relaxation delay of 1 s, NS = 2048 scans. Low-power ^{1}H SPINAL-64 decoupling (ν_{1H} = 2.4 kHz) (91) was checked to be sufficient and was used to protect the probe as the total acquisition time of 197 ms was too long for high-power decoupling. During acquisition, 4096 complex points were acquired with 24 Carr-Purcell-Meiboom-Gill (CPMG) echoes (39) and a full echo delay of 8 ms.

SVD was applied on Free Induction Decay (FID, time domain) after removal of the first 68 points corresponding to oversampled digitalisation. Zero-filling to 16384 complex points and cosine multiplication were applied after SVD. This apodisation limits both signal truncation and broadening effects. One may note that SVD was not directly applied to spectra (SPC, frequency domain) because zero-filling increases matrix size and thus computation time.

B.3. Simulation of kinetics studied by Raman spectroscopy under Python

2000 Raman spectra of 2000 points each were calculated *in silico* using four Gaussian lines at 450, 510, 750 and 900 cm^{-1}, respectively. Full Widths at Half Maximum

(FWHM) ranged from 118 to 212 cm^{-1}. Such high FWHM are typical of amorphous materials like glasses (45). In order to reflect a kinetic evolution, the amplitude of peaks at 510 and 750 cm^{-1} were linearly decreased across the series while amplitude of peaks at 450 and 900 cm^{-1} were linearly increased. This series of spectra may mimic ageing of a material for instance. Homoscedastic white Gaussian noise was added on each spectrum. SVD was applied using Principal Component Analysis (PCA) function from Python Scikit-learn package (92). Computation took only a few seconds under Python Anaconda 3.5. The source code is available in file Figure_I.4a.py of (93).

B.4. Simulation of NMR spectra with known noise under Matlab

NMR complex FID were simulated *in silico* under Matlab (The MathWorks, Inc., Natick, MA, USA) with a complex exponential at the expected frequency ν and either an exponential decay (Equation 1) or a Gaussian decay (Equation 2), leading to a Lorentzian or a Gaussian peak on spectra after Fourier transform, respectively. While the former is typical of a relaxation-driven shape, the latter highlights a distribution of chemical environments (94) or more complex relaxation phenomena, *e.g.,* strong dipoar coupling. To obtain a Gaussian peak with the same FWHM as a Lorentzian peak, the shape is defined according to Equation 3.

$$y = e^{i2\pi\nu t} \cdot \mathrm{e}^{\frac{-t}{T_2}} \tag{1}$$

$$y = e^{i2\pi\nu t} \cdot \mathrm{e}^{\frac{-t^2}{2\sigma^2}} \tag{2}$$

$$\sigma = T_2\sqrt{2\ln(2)} \tag{3}$$

7380 NMR FID were simulated, grouped as follows:

- 2 shapes for decay: exponential and Gaussian;
- 3 T_2 values of 10, 1.0 and 0.10 ms corresponding to *narrow* (32 Hz), *intermediate* (320 Hz) and *broad* peaks (3200 Hz), respectively, which are typical values obtained by ^{13}C, ^{29}Si or ^{31}P solid-state NMR, for various types of crystalline or amorphous materials;
- 41 levels of homoscedastic white Gaussian noise ranging from -20 dB to +20 dB;
- 30 random noise patterns at the same noise level.

Additionally, each data set was repeated 6 times with different processing parameters:

- 2 with truncation or not, at 5 T_2, time above which signal is almost no longer existent;
- 3 Significance Level (SL, see Section D.3) for SVD automatic thresholding at error level of 5, 7.5 or 10 %.

Each FID was composed of 1024 points for a duration of 49 ms. SVD was applied before zero-filling (if truncation was applied) and Fourier transform. As peak was not at the middle of the spectrum, *signal region* was defined as the 512 points centred at peak frequency. *Noise region* corresponded to the other 512 points. Baseline zero-order offset was preliminary corrected by subtracting the mean value of noise region. This step was essential to avoid spectrum aliasing due to Fourier transform, especially for broad peaks. The source codes of SVD automatic thresholding and FID simulations are available in files sfa.m and Figure_I.7_I.8_I.S2.m of (93), respectively. Computation of the full set of $6 \times 7380 = 44280$ spectra took 30 minutes with an overclocked Intel Core i5 4670K @ 4.4 GHz processor with Matlab R2016b.

C. Theoretical background

In this section, SVD and low-rank approximation are first developed. Hankel and Toeplitz matrices are then presented. Finally, SNR is defined.

C.1. SVD and low-rank approximation

SVD is a mathematical tool used to decompose a matrix X with m rows and n columns, whatever its size or shape, into the product of three other matrices U, Σ and V^T (Equation 4). This is illustrated in Figure 1 by orange hatched rectangles. U and V are unitary square matrices, of size m × m and n × n, respectively. If complex numbers are used, V^T, the transpose of matrix V, is replaced by V*, its conjugate transpose. SVD can indifferently be applied on real or complex matrices, the only difference being a double computation time for complex matrices (see part (II) of this work (86)). The central matrix Σ has the same shape as the original matrix X. Nevertheless, it has values only on its main diagonal (green rectangle), sorted by amplitude. These diagonal entries are called singular values and are the non-negative square roots of the eigenvalues of X^TX or XX^T (95). One can notice that the more elongated is matrix X, the less singular values it has.

$$X = U \cdot \Sigma \cdot V^T \tag{4}$$

$$X \approx X_k = U_k \Sigma_k V_k^T \tag{5}$$

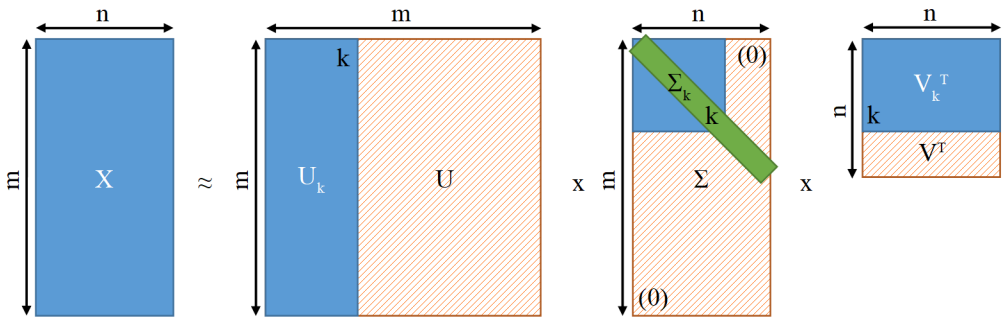


Figure 1: Singular value decomposition (orange hatched rectangles) and low-rank approximation (blue filled rectangles). The k singular values are sorted by size (in green).

Using low-rank k, matrix X can be approximated to X_k according to Equation 5, where U_k, Σ_k and V_k^T are the matrices U, Σ and V^T truncated at k values, represented as blue filled rectangles in Figure 1. This process is really useful to compress large data sets (68). In the case of noise-containing data, true signals correspond to low-k values while noise signals are related to high-k values. Thus, selecting the correct k-limit allows to keep all true signals while rejecting noise, including t_1-noise (73). Additionally, when a baseline distortion is present, its intense signal correspond to the first singular value which may be removed (96). The following section describes how to apply SVD on one-dimensional (1D) data.

C.2. Hankel and Toeplitz matrices

As stated above, SVD can only be applied to matrices. However, 1D data form only a (complex) row or column but not a matrix. In such a case, a transformation step is required. This can be performed thanks to Hankel matrix, or similarly to Toeplitz matrix. The former is defined by its first row and its last column. All anti-diagonal values are filled identically to the first ones (Figure 2). Toeplitz matrix is defined by its first column and its first row, all diagonal values being identical (97). Circulant matrices

are a special case of Hankel or Toeplitz matrices, where every row of the matrix is a cyclic shift of the row above (98). However, in general case, these are semi-circulant matrices, with one value being replaced from one row to the next one.

Hankel and Toeplitz matrices are not necessarily square, and are the vertical reflection of each other. From a programming point of view, extracting anti-diagonals needs a matrix vertical reflection. Although this step is computationally inexpensive, Toeplitz matrices were preferred for the sake of simplicity when using Python and Matlab. The Java application developed by Man *et al.* (81–83) used a Hankel matrix. It should be noted that after denoising, (anti-)diagonals values are no longer identical and averaging is needed as highlighted by Hansen and Jensen (64). They indeed stated that 'simply extract (and transpose) an arbitrary row of the matrix […] lacks a solid theoretical justification'.

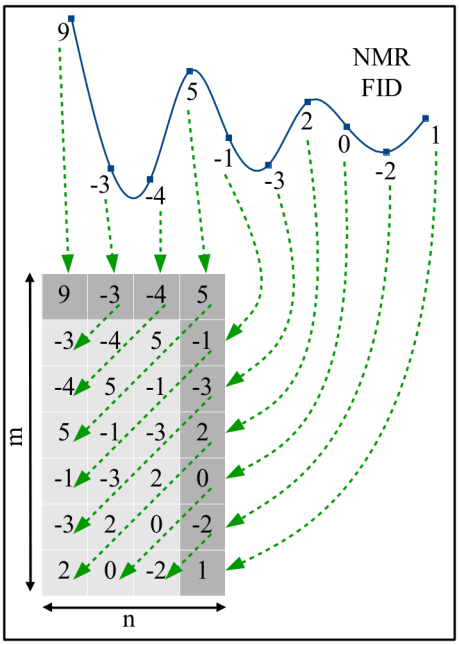


Figure 2: Hankel matrix applied to a NMR FID. Each point of the FID defines either one point of the first row or of the last column of the Hankel matrix. All anti-diagonal points are filled identically to the first point.

C.3. Signal-to-noise ratio

Two definitions of SNR are used. The first one corresponds to the mathematical formula (Equation 6), used in electronics, where y_i are the individual values while σ_{signal} and σ_{noise} are the standard deviations for signal region and noise region, respectively. Normalisation over (n-1) points is preferred to avoid a bias in standard deviation (99). This formula is valid only when signal can be measured without any noise, so-called *pure* signal. However, the only observable parameter on a real noisy signal is the signal-plus-noise-to-noise ratio (SNNR) (100) defined by replacing σ_{signal} by $\sigma_{signal+noise}$ in Equation 6. SNR can then be deducted from SNNR, following Equation 7.

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}} = \frac{\left(\frac{1}{n-1}\sum_i (y_i)^2\right)_{signal}}{\left(\frac{1}{n-1}\sum_i (y_i)^2\right)_{noise}} \quad (6)$$

$$SNR = \frac{\sigma^2_{signal+noise} - \sigma^2_{noise}}{\sigma^2_{noise}} = SNNR - 1 \quad (7)$$

The other possible definition is the analytical chemistry formula (Equation 8), where PSNR is the SNR based on peak amplitude (100). Signal height (H_{signal}) is measured from maximum of peak to mean of noise, whereas noise is measured on a region of 20 times the signal FWHM (101). Additionally, variants exist, depending on the way noise is defined, either as maximum noise (h_{noise_max}), mean noise (h_{noise_mean}) or Root Mean Square (RMS) noise (h_{noise_rms}) (102). Following this nomenclature, SNR used in NMR (102) and Raman spectroscopy (51) should rather be called $PSNR_{rms}$ (Equation 9). While SNR is related to the area of the studied peak, PSNR is related to its

height, leading to different results. They can be expressed in decibels (Equation 10), which is more convenient to explore a wide variation range.

$$PSNR_{max} = \frac{H_{signal}}{h_{noise\_max}} = \frac{H_{signal}}{h_{noise\_peak\_peak}/2} = \frac{2\,max(y_i)_{signal}}{(max(y_i) - min(y_i))_{noise}} \quad (8)$$

$$PSNR_{rms} = \frac{H_{signal}}{h_{noise\_rms}} = \frac{H_{signal}}{\sigma_{noise}} = \frac{max(y_i)_{signal}}{\sqrt{\left(\frac{1}{n-1}\sum_i (y_i)^2\right)_{noise}}} \quad (9)$$

$$SNR^{dB} = 10\log_{10}(SNR) \quad , \quad PSNR^{dB}_{rms} = 10\log_{10}(PSNR^2_{rms}) \quad (10)$$

Additionally, Currie defined a Critical Level (L_c), a Detection Limit (L_d) and a Quantitative Limit (L_q), at 1.64, 3.29 and 10 σ_{noise}, respectively (103). The corresponding levels, SNR and PSNR measurements are drown in Figure 3.

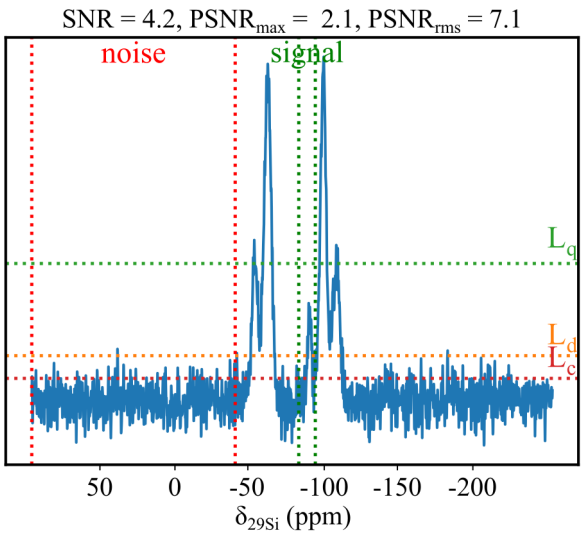


Figure 3: Measurement of Signal-to-Noise Ratio (SNR) and Peak SNR (PSNR) based on $noise_{max}$ and $noise_{rms}$ for the ^{29}Si MAS solid-state NMR spectrum of the 50 / 50 MTEOS / TEOS sample. Signal and noise regions are highlighted with dotted vertical green and red lines, respectively. Critical (L_c), detection (L_d) and quantitative (L_q) limits (103) are shown with red, orange and dotted horizontal green lines, respectively.

D. Results and discussion

In this section, SVD is applied to NMR and Raman spectra. We then focus on practical aspects of denoising, namely the impact of the matrix shape and the number of components used for thresholding. Time and frequency denoising are compared. The minimum experimental SNR needed for valid use of SVD and its impact on peak shape are thoroughly investigated. Finally, a limit case is evaluated.

D.1. Denoising of NMR and Raman spectra

Images are already matrices and SVD can directly be applied on them. Two-dimensional (2D) spectra can be treated similarly. However one-dimensional (1D) spectra are not directly suitable for SVD. If a series of spectra is available, one just need to stack the successive spectra to obtain a 2D data set. Figure 4a shows such a stack of spectra simulating a reaction kinetics studied by Raman spectroscopy, as described in Section B.3. A similar stack can be obtained when a surface is mapped to analyse species in a sample region (31). The spectrum consisted of four overlapped peaks with varying intensities. Without SVD, the two peaks at 510 and 900 cm^{-1} were difficult to detect due to the amount of noise (red arrows). However, after SVD, these components were identified as evidenced by green arrows. One should note the very low residual noise.

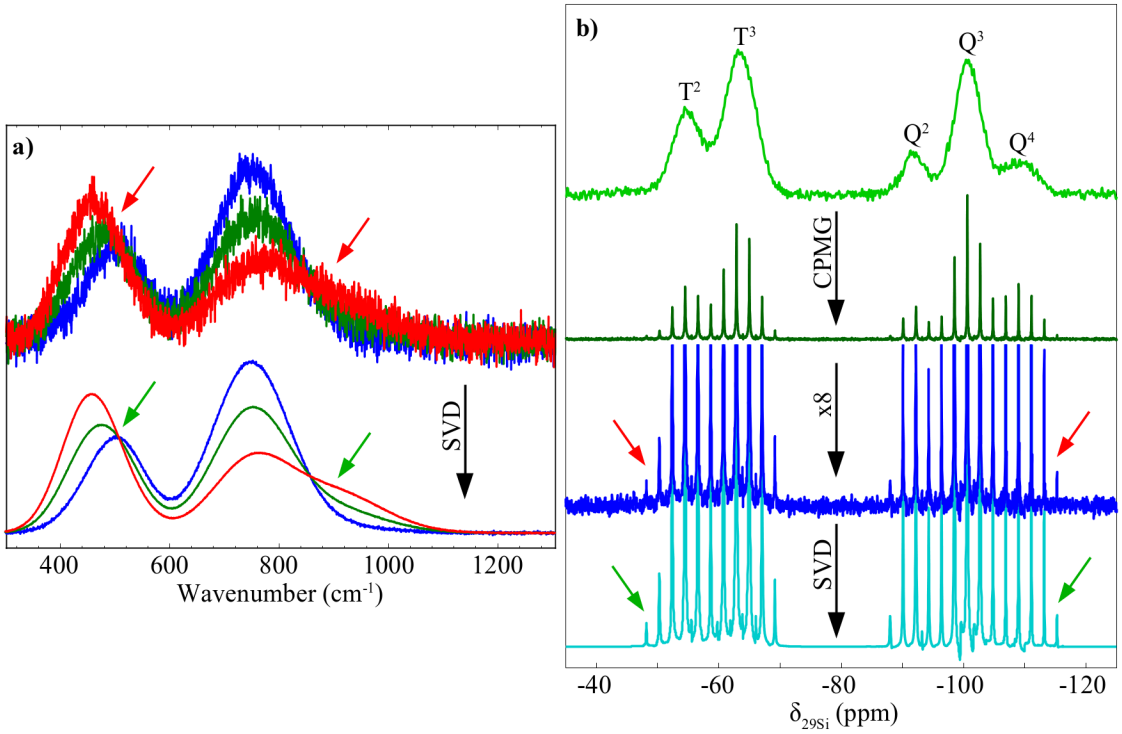


Figure 4: a) SVD applied to a set of spectra mimicking a kinetic reaction, as probed by Raman spectroscopy. Four overlapping bands are simulated on 2000 spectra with 2000 points each, resulting into a matrix of 2000×2000 points. Only three representative spectra are shown. b) ^{29}Si MAS solid-state NMR spectra of the 50 / 50 MTEOS / TEOS sample. From top to bottom: standard spectrum, CPMG spectrum, vertically zoomed CPMG spectrum, spectrum obtained after SVD applied on time domain signal of the above transformed to an Hankel matrix of 2015×2014 points. Red and green arrows show difficult to detect and enhanced signals, respectively.

It is also possible to use SVD on a 1D data set, by way of a Hankel or a Toeplitz matrix, as described in Section C.2. This feature was used in the context of ^{29}Si MAS solid-state NMR (Figure 4b) using the GPU Java application (83). The ^{29}Si MAS NMR spectrum of the 50 / 50 MTEOS / TEOS sample (see Section B.1) is shown in the top of the figure. It should be noted that this spectrum already displays a very good SNR. The signal has been enhanced using CP, allowing a non-quantitative spectrum to be acquired in 40 minutes. The second spectrum shows the same sample analysed using CPMG echoes (39), which led to numerous spikelets. The overall shape of the CPMG spectrum

is qualitatively similar to the above spectrum. Such an approach was proved to be a mean to increase SNR during acquisition step by discretising broad peaks (104). The original shape with improved SNR can be recovered by summing echoes, but nevertheless this leads to relaxation distortions. The third spectrum of Figure 4b is a vertical zoom to highlight noise level and much less intense spikelets marked by red arrows. By comparison with the same spectrum after SVD processing, such small peaks were highly enhanced and noise has disappeared. This ^{29}Si CPMG MAS NMR spectrum was used hereafter as a reference to evaluate the performances of the SVD process.

D.2. Matrix shape

It is sometimes argued that efficient denoising can be obtained using an iterative process on a rectangular matrix, with a number of columns higher than the number of signals of interest (105), *i.e.* roughly the number of peaks. The iteration consists in converting into a matrix, applying SVD, and reverting to a 1D set of denoised data. Due to (anti-)diagonals averaging (see Section C.2), the matrix at the beginning of the second iteration is not exactly the one at the end of the first iteration, which explains that multiple iterations give different results. However in our case, such a procedure led to some residual noise (Figure S1). Even with $m \times n = 3901 \times 128$ points and 10 iterations, noise was only marginally reduced. Nevertheless, this procedure has the advantage of a low computation time *per* iteration, as Hankel or Toeplitz matrix size is smaller for a rectangular shape than a square shape, when starting with the same number of points in 1D data set (see part (II) of this work (86)).

In a second step, Figure 5a depicts the efficiency of denoising on various matrix shapes, either rectangular ones ($m \times n = 3997 \times 32$ points) or square ones ($m \times n = 2015 \times 2014$ points). While noise was strongly present for elongated matrices,

it decreased when the number of columns increased and finally disappeared for a square matrix. Square matrices were used hereafter. More generally, our results suggest to tend to a square data matrix before applying SVD, as also recommended by Van Huffel *et al.* (106). For n = 512 and n = 2014, small peaks seem to be missing, highlighted by orange arrows. This feature is explained in next section.

D.3. Thresholding

Another parameter to be varied in order to optimise denoising is the number of singular values k (corresponding to signals) used for low-rank approximation (Figure 5b). In a first attempt, k was set to the number of peaks present, k = 22. However, k = 25 resulted in a better shape for the three more intense spikelets, corresponding to Q^3 and T^3 peaks (Figure 4b) and k = 47 was necessary to select all small peaks that were missing in previous section (green circles and orange arrows). Above this value, isolated artefacts were observed as shown for k = 50 and k = 75 (red ellipsis). They are usually narrow and out of phase, which make them easy to detect. It is thus necessary to carefully adjust k to discriminate signals from noise.

Another approach to manually select the correct number of singular values was to plot singular values in logarithmic scale (Figure 5c). Up to k = 25, singular value amplitude was strongly decreasing. Between k = 25 and k = 47 (dashed vertical lines), a slow slope was present and it was hard to distinguish the optimal value, because these singular values had a too low SNR. Above k = 47, the curve exhibited a plateau and finally a cliff for last indexes. The plateau and the cliff are characteristic of noise values (107).

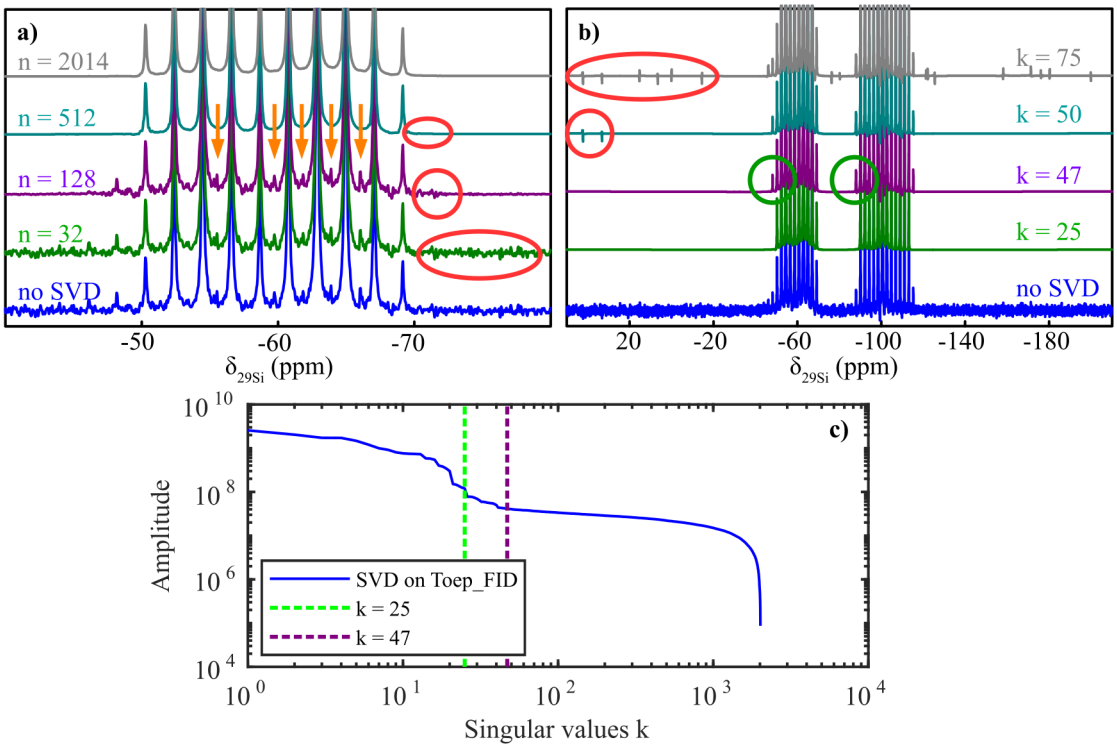


Figure 5: a) Influence of the number of columns (n) for Toeplitz FID (Toep_FID) matrix construction, with m + n = 4028 points and k = 22 singular values, for a ^{29}Si CPMG solid-state NMR spectrum of the 50 / 50 MTEOS / TEOS sample. b) influence of the number of singular values (k) and c) singular values in logarithmic scale for a matrix of 2015×2014 points (n = 2014). 22 major spikelets and 47 in total are present. Red and green circles show artefacts and enhanced small signals, respectively; orange arrows are minor spikelets visible at k = 47.

In order to select automatically the proper number of singular values, Malinowski developed an INDicator function (IND) and a Significance Level (SL) function (108), which are available under Matlab in file sfa.m of (93). While IND is based on the residual standard deviation, SL is a Fisher variance test giving the probability for a singular value to correspond to noise. In the former case, the minimum of IND reflects the number of singular values to select. In the latter case, the singular value is rejected if it has a probability of being noise higher than a desired level. The influence of this level will be investigated in Section D.5.b. When applying these

functions to the spectrum reported in Figure 5b, we found k = 53 with IND and k = 31, 36 and 39 with SL at 5, 7.5 and 10 % error level, respectively. As stated before, k = 47 was the manual optimum for the spectrum considered here, selecting all singular values attributed to signals while rejecting artefacts. While IND overestimated k and introduces artefacts, SL underestimated k, thus ignoring small signals. Nevertheless, SL was the tool of choice for automatic thresholding as the result was close from the expected one and did not display artefacts. It should be noted that singular value thresholding is also available (109) but it needed to adjust too many parameters and to sparsify data.

D.4. Time and frequency domains

In this section, spectra resulting from a same FID with SVD denoising applied either before (Figure 6, Toep_FID, in cyan) or after Fourier transform (Toep_SPC, in green) are compared. Both were simulated under Matlab and converted into a square Toeplitz matrix, according to Section D.2, before applying SVD. Additionally, multiple similar spectra with identical signal and random noise were stacked (mult_SPC, in red), to simulate a mapping of a homogeneous region in MRI or in Raman. For mult_SPC, we chose a number of spectra identical to the number of points per spectrum, again to get a square matrix. Mult_SPC was also useful to compare the influence of multiple sampling *vs.* a single sampling converted to a semi-circulant matrix. Singular values are presented in Figure 6a. By construction, the size of Toep_FID and Toep_SPC was half the one of mult_SPC, which explains that their maximum singular value indexes were lower. Singular value plot was very similar for Toep_FID and mult_SPC, but amplitude of the latter was higher. The difference between signals and noise singular values was also more prononced on mult_SPC. On the contrary, Toep_SPC had a very different signature, with 17 singular values corresponding to signals, instead of only one that was

expected.

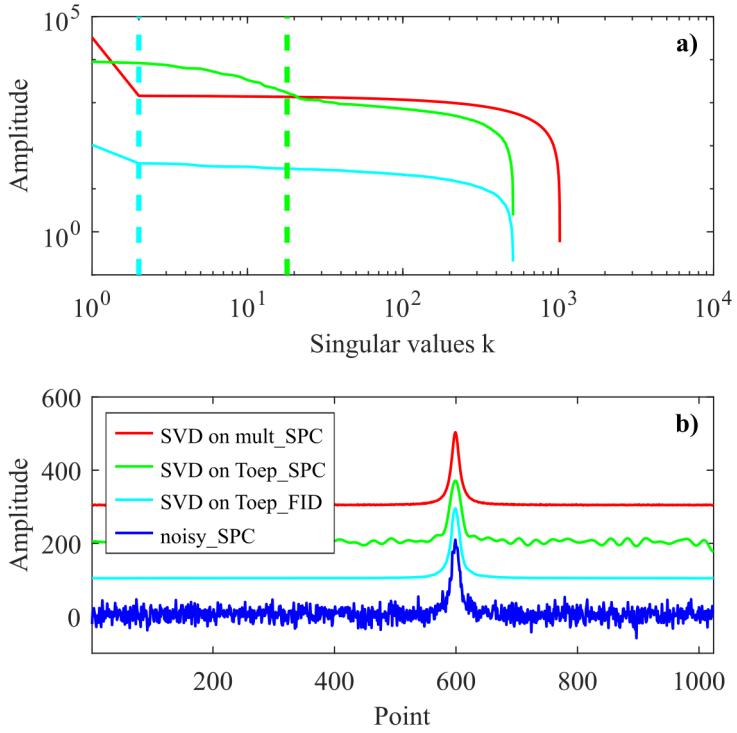


Figure 6: a) Singular values and b) spectra obtained with and without SVD. noisy_SPC: Fourier transform (SPC) of noisy FID simulated with 1024 points; Toep_FID and Toep_SPC: Toeplitz matrix of FID and SPC, respectively, with 513x512 points; mult_SPC: matrix of stacked multiple identical spectra with random noise, characterised bt 1024x1024 points. Dashed lines: corresponding Significance Level (SL) noise limit.

Noisy and denoised spectra are shown in Figure 6b. Interestingly, very different behaviours were observed. Satisfactory denoising was achieved using mult_SPC or Toep_FID. This result explains why SVD was efficient in the case of the afore mentioned Raman spectra mimicking a kinetic process (Figure 4a) and the NMR FID (Figure 4b). On the contrary, for Toep_SPC, noise –though reduced– was still present. This clearly denotes that in case of 1D frequency domain, a reverse Fourier transform before converting data into a Toeplitz matrix and applying SVD should be preferred (110).

D.5. Minimum signal-to-noise ratio

D.5.a. Comparison of SNR^{dB} and $PSNR_{rms}^{dB}$

An additional question is the sensitivity of SVD, that is to say the minimum SNR needed to get a proper signal detection. The corresponding indicators, SNR^{dB} and $PSNR_{rms}^{dB}$, were defined in Section C.3. 7380 noisy spectra were simulated with either Lorentzian or Gaussian shape and with a peak width of either 32 Hz, 320 Hz or 3200 Hz (see Section B.4). First, we compared in Figures 7a and 7b the SNR^{dB} determined using the pure signal and the separated desired noise (calculated SNR^{dB}) to the $PSNR_{rms}^{dB}$ measured on signal and noise regions of the noisy data (measured $PSNR_{rms}^{dB}$). Calculated SNR^{dB} was a simulating tool, close to the theoretical SNR^{dB}, whereas measured $PSNR_{rms}^{dB}$ reflected an experimental assessment, which was directly obtained on noisy spectra. Comparing them was a way to check how the noise itself could influence $PSNR_{rms}^{dB}$. For narrow peaks (in red) a linear relationship between SNR^{dB} and $PSNR_{rms}^{dB}$ was obtained. However, for intermediate and broad peaks (in green and blue, respectively), the evolution of SNR^{dB} with $PSNR_{rms}^{dB}$ displayed a steeper increase, with a vertical asymptote at $PSNR_{rms}^{dB} = 10$ dB (dashed black line). This value reflected an undetectable signal with $H_{signal} \leq h_{noise\_max}$, as $\sigma_{noise} = h_{noise\_max} / 3.3$ with a probability of 99.9 % for Gaussian noise. The increase resulted from spreading of peak area over a wider range. This implied a lower amplitude for broad peaks than for narrow peaks and consequently a lower $PSNR_{rms}^{dB}$ value.

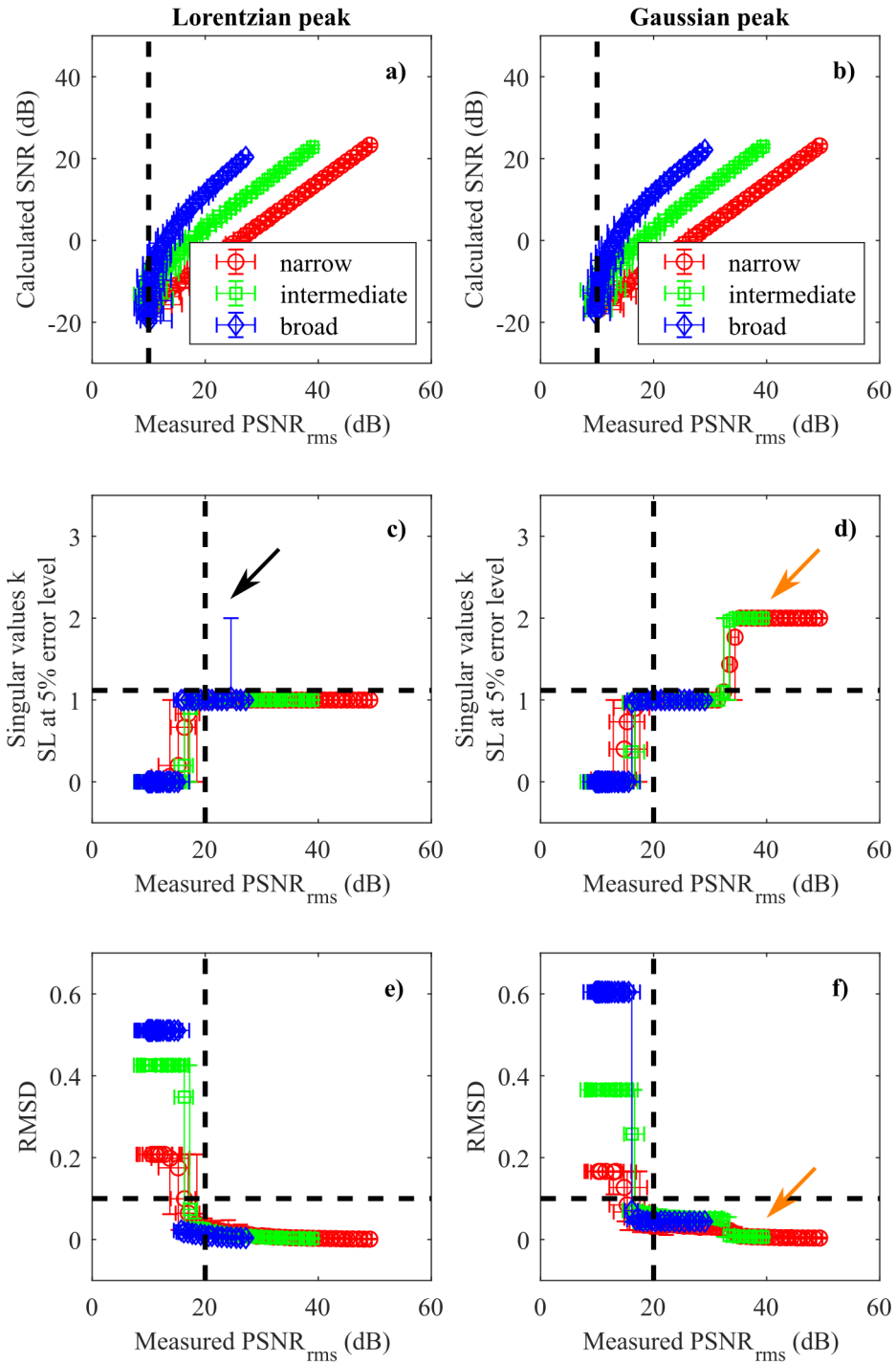


Figure 7: 7380 simulated spectra with known added homoscedastic white Gaussian noise. a), c) and e) Lorentzian peak; b), d) and f) Gaussian peak; a) and b) comparison of SNR definitions; c) and d) automatic thresholding with SL at 5% error level; e) and f) root mean square deviation of denoised spectra. Error bars correspond to the repetition of 30 simulated spectra with the same level of added noise. 41 levels of noise were used ranging from -20 dB to +20 dB. SVD was applied on time data. SNR and PSNR were obtained on frequency data. Dashed black lines represent characteristic values (see text

for more details); black arrow shows an artefact detection; orange arrows highlight the second singular value for Gaussian peaks. Data set is available in file Data_SL5.mat of (93).

The 7380 noisy spectra were then truncated to 5 T_2 and the same comparison was performed between calculated SNR^{dB} and measured $PSNR_{rms}^{dB}$ (Figures S2a and S2b). This truncation removed the vertical asymptote observed at low $PSNR_{rms}^{dB}$, as less points were defining noise, and intense noise peaks were less probable. This led to an artificial increase of SNR and to a vertical shift of $SNR^{dB} = f(PSNR_{rms}^{dB})$ evolution for intermediate and broad lines, pictured in green and blue, respectively, in Figures S2a and S2b. Moreover, a broader distribution of SNR values was obtained, especially for broad peaks (blue line). An additional feature was observed for broad Lorentzian peaks: a vertical asymptote at $PSNR_{rms}^{dB} = 37$ dB. This feature was a consequence of the spectral extension of the wings of Lorentzian peaks that were contributing to amplitude within noise region. On the contrary, Gaussian peak wings are much less intense and in this case, this vertical asymptote was not observed on $SNR^{dB} = f(PSNR_{rms}^{dB})$ evolution.

D.5.b. Automatic thresholding

SVD and Malnowski's SL automatic thresholding (see Section D.3) were applied to these simulated NMR FID corresponding to only one peak. When SL error level was set to 5 % and above $PSNR_{rms}^{dB} = 20$ dB this single peak was detected with k = 1 singular value, whatever peak width and shape (Figures 7c-d and S3a-b, dashed vertical black line). However, this was only an upper limit and many peaks were detected around $PSNR_{rms}^{dB} = 17$ dB, between the detection limit of 3.3 σ_{noise} and the quantification limit of 10 σ_{noise}, as defined by Currie (103). Thus, to be detected through Malinowski's algorithm, a signal has to be enough different from noise, *i.e.* between two to three

times higher than $noise_{max}$. Surprisingly, a second singular value was detected with k = 2 for Gaussian shapes above $PSNR_{rms}^{dB}$ = 36 dB, (Figures 7d and S3b). The amplitude of this second component was significant and improved the resulting shape of denoised spectrum. This second singular value will be explained in Section D.6.c. On the full data set of 7380 spectra, only one false detection was observed, as indicated by the black arrow in Figure 7c. Thus, the amount of artefacts was negligible. On the truncated FID (Figures S2c and S2d), this limit of $PSNR_{rms}^{dB}$ = 20 dB was not so abrupt, due to lack of accuracy on noise measurement. However, an advantage of truncation was a much faster computation for broad peaks, thanks to the smaller matrix used.

When SL error level was set to 7.5 or 10 %, the minimum SNR to get a peak detection was decreasing (Figures S3c-f). However, the number of false detections also increased noticeably, as indicated by the black arrow. In some rare cases, evidenced with lines being higher than the figure vertical limit (black arrow on Figure S3d), SL was unable to distinguish signal from noise, resulting in a noisy spectrum after SVD. While an error level of 5 % is really safe, a level of 7.5 % may be necessary to detect tiny peaks. A value of 10 % seems too high to avoid artefacts. Results are summarised in Table 1. SNR is not presented in this table as it is not a relevant parameter, that is too much depending on peak width and shape.

Table 1. PSNR needed for SVD with SL automatic thresholding depending on the desired error level.

SL error level	**$PSNR_{rms}^{dB}$**	**$PSNR_{rms}$**	**$PSNR_{max}$**	**Artefacts**
5 %	17	7.1	2.1	no
7.5 %	16	6.3	1.9	small
10 %	15	5.6	1.7	strong

D.5.c. Error measurement

The difference between denoised signal and simulated non-noisy signal (pure signal),

was measured using the Root Mean Square Deviation (RMSD), defined on Equation 11, where $y_i^{denoised}$ and y_i^{pure} are the individual values for denoised and pure SPC, respectively (Figures 7e and 7f). A high RMSD was obtained below $PSNR_{rms}^{dB}$ = 17 dB. As no peak was detected in this range, the obtained value corresponded to RMSD of pure signal compared to zero, which was higher for broad peaks, due to its wide spread range. Above this $PSNR_{rms}^{dB}$ value, RMSD displayed a steep decrease under 0.1 (dashed horizontal black line). These results emphasised the very good agreement between denoised and pure data. However, RMSD was higher for Gaussian than for Lorentzian peaks (Figure 7f). Above the second threshold of $PSNR_{rms}^{dB}$ = 36 dB, RMSD exhibited a further decrease downto the level obtained for Lorentzian peaks. This confirms the significance of the second singular value. A similar trend was observed on truncated data (Figures S2e and S2f).

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i^{denoised} - y_i^{pure} \right)^2}{n}} \qquad (11)$$

D.6. Quantification

D.6.a. Pure and denoised spectra

The next step that was investigated concerns the possibility to use denoised spectra for the sake of quantification. For each peak width and shape at $PSNR_{rms}^{dB}$ = 20 dB, *i.e.* at quantification limit, the spectrum with the worst RMSD is presented in Figure 8. These spectra were fitted with a Voigt function with error estimation implemented in Matlab (111). Amplitude, position, shape and width were automatically adjusted. Results are reported in Table 2 and 3 for Lorentzian and Gaussian peaks, respectively. A very large uncertainty occurred on fitting parameters derived from noisy spectra (top traces). For

denoised spectra (middle top traces), uncertainty decreased significantly, roughly by a factor of 10, but was still higher than for pure spectra (middle bottom traces). Difference between denoised and pure spectra is presented in bottom trace of Figure 8.

D.6.b. Lorentzian and Gaussian peaks

Surprisingly, despite a RMSD lower than 0.1, the area Percent Error (PE_{area} given by Equation 12) could be as high as 8.5 % and 42.6 % for Lorentzian and Gaussian peaks, respectively. While the former was acceptable at detection limit, the latter evidenced an overestimation. Although PE_{area} decreased after SVD denoising on Lorentzian peaks, it increased for Gaussian peaks. Moreover, difference spectrum on Gaussian peak exhibited a mix of narrow and wide components with opposite amplitudes (Figure 8f). Such a shape modification was not observed for Lorentzian peaks. Besides, the Gaussian / Lorentzian ratio was around 0.5 instead of 1.0 after denoising on Gaussian peaks (dark grey and light grey rows of Table 3). This result highlighted that SVD induced a change in peak shape from Gaussian peaks to more Lorentzian ones. Above the second threshold of $PSNR_{rms}^{dB} = 36$ dB, the shape was corrected thanks to the second singular value, giving a pure Gaussian peak after denoising (white rows of Table 3).

$$PE_{area}(\%) = \frac{Area^{denoised} - Area^{pure}}{Area^{pure}} \times 100 \qquad (12)$$

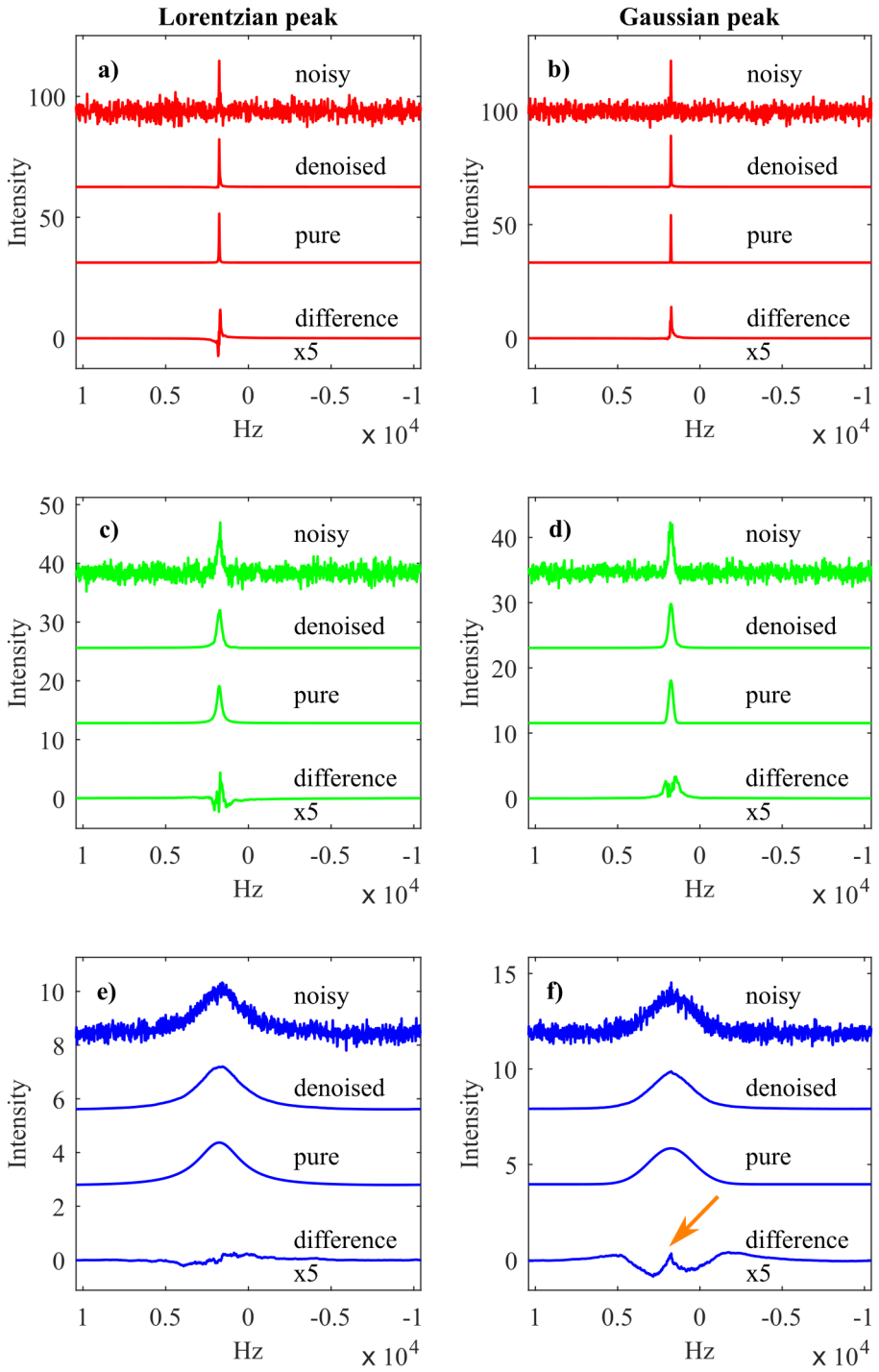


Figure 8: Spectra obtained in Figure 7 at $PSNR_{rms}^{dB}$ = 20 dB with worst RMSD for a), c) and e) Lorentzian peak; b), d) and f) Gaussian peak; a) and b) narrow peak in red; c) and d) intermediate peak in green; e) and f) broad peak in blue. On each sub-figure, top spectrum, middle top, middle bottom, bottom spectra correspond to noisy, denoised, pure spectra and difference between denoised and pure spectra, respectively. Orange arrow highlight the Gaussian peak distortion.

Table 2. Modelling of pure, noisy and denoised **Lorentzian peak** with narrow, intermediate and broad widths (in red, green and blue, respectively), for low, intermediate and high $PSNR_{rms}^{dB}$ (dark grey, light grey and white rows, respectively). a: Matlab file exchange 52321 (111); b: root mean square deviation to pure spectrum (Equation 11); c: Percent error on pure spectrum area (Equation 12).

				Modelling [a]					
Shape	**T_2 (ms)**	**Signal**	**RMSD** [b]	**Ampli-tude**	**Position (Hz)**	**G/L**	**Width (Hz)**	**Area**	**PE_{area} (%)** [c]
Lorentzian low $PSNR_{rms}^{dB}$	10	Noisy 19.6 dB	2.78	21 ± 5	1763 ± 6	0.0 ± 0.8	40 ± 10	1185	+ 17.3
		Denoised	0.05	19.7 ± 0.3	1760.3 ± 0.4	0.00 ± 6e-2	35 ± 1	1096	+ 8.5
		Pure	-	20.191 ± 3e-3	1761.87 ± 4e-3	9.2e-3 ± 7e-4	31.99 ± 1e-2	1010	-
	1	Noisy 19.5 dB	1.12	6.1 ± 0.8	1710 ± 20	0.3 ± 0.4	320 ± 60	2693	- 14.4
		Denoised	0.04	6.27 ± 7e-2	1729 ± 2	0.27 ± 4e-2	342 ± 6	2992	- 4.9
		Pure	-	6.299 ± 3e-3	1761.89 ± 6e-2	4e-3 ± 1e-3	318.5 ± 0.2	3145	-
	0.1	Noisy 20.2 dB	0.22	1.54 ± 6e-2	1520 ± 50	0.3 ± 0.2	3200 ± 200	6665	+ 2.5
		Denoised	0.02	1.52 ± 1e-2	1590 ± 10	0.35 ± 4e-2	3190 ± 40	6529	+ 0.4
		Pure	-	1.52 ± 1e-2	1762 ± 8	0.34 ± 3e-2	3160 ± 30	6502	-
Lorentzian interm $PSNR_{rms}^{dB}$	10	Noisy 30.1 dB	0.79	20 ± 1	1763 ± 2	0.1 ± 0.3	31 ± 4	952	- 5.7
		Denoised	0.01	20.56 ± 7e-2	1762.8 ± 0.1	0.00 ± 2e-2	30.8 ± 0.3	995	- 1.5
		Pure	-	20.191 ± 3e-3	1761.87 ± 4e-3	9.2e-3 ± 7e-4	31.99 ± 1e-2	1010	-
Lorentzian high $PSNR_{rms}^{dB}$	10	Noisy 36.2 dB	0.40	19.8 ± 0.6	1761.4 ± 0.8	0.1 ± 0.1	32 ± 2	963	- 4.7
		Denoised	0.01	19.90 ± 1e-2	1761.60 ± 2e-2	0.061 ± 3e-3	32.40 ± 4e-2	984	- 2.6
		Pure	-	20.191 ± 3e-3	1761.87 ± 4e-3	9.2e-3 ± 7e-4	31.99 ± 1e-2	1010	-

Table 3. Modelling of pure, noisy and denoised **Gaussian peak** with narrow, intermediate and broad widths (in red, green and blue, respectively), for low, intermediate and high $PSNR_{rms}^{dB}$ (dark grey, light grey and white rows, respectively). a: Matlab file exchange 52321 (111); b: root mean square deviation to pure spectrum (Equation 11); c: Percent error on pure spectrum area (Equation 12).

				Modelling [a]					
Shape	**σ (ms)**	**Signal**	**RMSD** [b]	**Ampli-tude**	**Position (Hz)**	**G/L**	**Width (Hz)**	**Area**	**PE_{area} (%)** [c]
Gaussian low $PSNR_{rms}^{dB}$	12	Noisy 19.9 dB	2.81	22 ± 4	1762 ± 4	1 ± 1	37 ± 9	927	+ 31.3
		Denoised	0.08	22.51 ± 8e-2	1761.00 ± 9e-2	0.42 ± 2e-2	34.2 ± 0.2	1007	+ 42.6
		Pure	-	20.827 ± 9e-5	1761.89 ± 9e-5	1.000 ± 3e-5	31.861 ± 2e-4	706	-
	1.2	Noisy 20.4 dB	0.88	6.7 ± 0.5	1790 ± 10	1.0 ± 0.5	330 ± 30	2329	+ 5.3
		Denoised	0.07	6.78 ± 2e-2	1757.3 ± 0.4	0.46 ± 1e-2	331 ± 1	2891	+30.8
		Pure	-	6.5193 ± 2e-5	1761.89 ± 5e-4	1.000 ± 2e-5	318.61 ± 1e-3	2211	-
	0.12	Noisy 20.5 dB	0.32	1.86 ± 8e-2	1810 ± 50	0.9 ± 0.3	3200 ± 100	6499	+ 1.9
		Denoised	0.05	1.839 ± 7e-3	1734 ± 4	0.62 ± 2e-2	2980 ± 10	6637	+ 4.1
		Pure	-	1.8800 ± 2e-5	1761.89 ± 1e-2	1.000 ± 8e-5	3186.0 ± 3e-2	6376	-
Gaussian intem $PSNR_{rms}^{dB}$	12	Noisy 29.7 dB	0.89	21 ± 1	1761 ± 2	0.2 ± 0.4	30 ± 4	899	+ 27.3
		Denoised	0.04	22.24 ± 4e-2	1761.80 ± 4e-2	0.38 ± 1e-2	30.77 ± 9e-2	913	+ 29.3
		Pure	-	20.827 ± 9e-5	1761.89 ± 9e-5	1.000 ± 3e-5	31.861 ± 2e-4	706	-
Gaussian high $PSNR_{rms}^{dB}$	12	Noisy 36.3 dB	0.40	21.4 ± 0.6	1761.7 ± 0.6	1.0 ± 0.2	31 ± 1	715	+ 1.3
		Denoised	0.02	21.32 ± 6e-2	1761.85 ± 6e-2	1.00 ± 2e-2	31.4 ± 0.1	713	+ 1.0
		Pure	-	20.827 ± 9e-5	1761.89 ± 9e-5	1.000 ± 3e-5	31.861 ± 2e-4	706	-

D.6.c. Real and extracted errors

The error on the measured area can origin from two sources: first the error introduced by the added noise, known as the *real error*, and second the error coming from the denoising itself, so-called the *extracted error* (108). An example of real error is presented in Figure S4a. Two successive measurements with NS = 120 scans gave a different amplitude for Q^2 peak at -91 ppm (red arrow). With a higher noise averaging at NS = 360 (not shown), amplitude ratios were similar to NS = 840. Thus, SNR at NS = 120 was too low and amplitude was tainted by error. A strong apodisation has been used here to artificially improve SNR (see Section D.7.a). Automatic thresholding was unable to correctly discriminate signals from noise and manual thresholding with k = 5 singular values was preferred. Nevertheless, the real error was kept after denoising (Figure S4b), which demonstrated that manual thresholding is a dangerous tool. Failure of automatic thresholding is thus an indication that SNR has to be improved.

The extracted error was especially present for Gaussian spectra, for which the Gaussian / Lorentzian ratio modification (see Section D.6.b) led to an increase of peak area. Indeed, SVD fits time decays with a sum of exponential (55). When fitting a Gaussian decay with a single exponential component, corresponding to one singular value at low $PSNR_{rms}^{dB}$, the peak area is correspondingly overestimated by 20 % (112). Our results were consistent with this value. Gaussian and exponential decays are very different, as Gaussian is flatter around its maximum. Above the second threshold, the Gaussian decay was fitted with two exponential decays, improving peak area value.

Unfortunately, when studying solid-state samples by using spectroscopic approaches, peaks are most of the time not Lorentzian. In such cases, SVD quantitative results are difficult to obtain. A workaround would be to model the resulting spectrum with pseudo-Voigt functions. For peaks with a Gaussian / Lorentzian ratio around 0.5,

dividing their area by 1.2 (20 %) should improve quantification. Taking this precaution into account for analysis of our data, PE_{area} was found to be similar between denoised spectra at $PSNR_{rms}^{dB}$ = 20 dB (dark grey rows in Table 2 and 3) (101) and noisy spectra at $PSNR_{rms}^{dB}$ = 30 dB, (light grey rows in Table 2 and 3). In solid-state NMR, another possibility to avoid Gaussian peak error relies on use of CPMG echoes (39) as in Figure 4b. This technique transforms a peak driven by chemical shift distribution (Gaussian shape, inhomogeneous interaction) into multiple narrow peaks driven by relaxation (Lorentzian shapes, homogeneous interaction) (113, 114), which are very suitable for SVD, being both sensitive and quantitative.

D.7. Limit case on a real NMR spectrum

D.7.a. Pre-processing

A pre-processing step called apodisation can be applied on FID before SVD and Fourier transform. The aim is first to reduce noise, and second to remove truncation artefacts leading to oscillations at peak foot, hence the name. In NMR, one can use for instance either exponential, cosine, or (shifted-)Gaussian decays. Their shape were compared in (115). While exponential is concave, cosine is convex and Gaussian is intermediate. In Figure S4c, we compared the influence of apodisation on initial noise. The resulting denoised SPC with automatic thresholding at an SL error level of 7.5 % are presented in Figure S4d. Without apodisation, SNR was too low to detect Q^2 peak at -91 ppm (red circle) and k = 4 singular values were found. With cosine apodisation, the correct number of peaks was obtained with k = 5 singular values (green ellipsis). With an exponential decay of 20 Hz, corresponding to the intrinsic SPC resolution, a similar result was obtained, but with k = 6 singular values, leading to small baseline distortions. Surprisingly, with an exponential decay of 50 Hz, a much higher number of k = 94

singular values was found, with almost no denoising (orange circle). An explanation was that apodisation changed the amplitude of noise values, especially at the end of the FID. By this way, noise became heteroscedastic, decreasing efficiency of SVD and Malinowski's criterion. When plotting singular values in logarithmic scale (Figure S4e), the slope moved from a plateau for cosine (purple curve) to a decay for exponential with 50 Hz (blue curve). In such a case, it is harder to discriminate signals from noise, as the slope is similar. Cosine is thus a good compromise before SVD as it decreases noise without changing too much singular values. An alternative would be to combine SVD and Savitzky–Golay smoothing filter (116), which process noise the same way all over the FID.

D.7.b. Denoising

On the sol-gel 50 / 50 MTEOS / TEOS sample, four hours and NS = 240 scans with cosine pre-processing, were needed to have a spectrum with a sufficient SNR to apply SVD and to detect Q^2 peak with automatic thresholding at SL error level of 10 % (not shown). If SL error level was limited to 7.5 %, six hours and NS = 360 scans were necessary (Figure 9 top trace). The corresponding denoised spectrum (middle top trace) was very close to the reference spectrum acquired in fourteen hours and NS = 840 (middle bottom trace), with $PSNR_{rms} = 9.7$, *i.e.* at quantification limit. Their difference (bottom trace) was comparable to noise. This was confirmed by peaks integration (Table 4) where very good agreement was obtained between the denoised spectra at NS = 240 or NS = 360 and the reference noisy spectrum at NS = 840. In particular, Q^2 / Q^3 ratio was very consistent. Time gain was thus between 2.3 to 3.5, depending on the SL error level allowed.

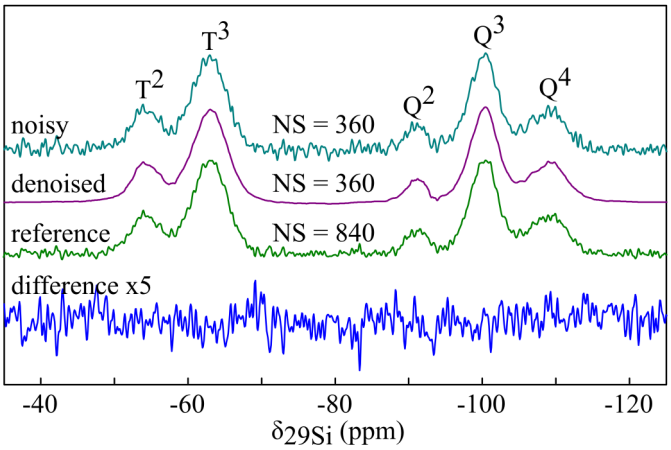


Figure 9: ^{29}Si MAS solid-state NMR spectrum of the 50 / 50 MTEOS / TEOS sample. From top to bottom: noisy spectrum at NS = 360 scans; denoised spectrum at NS = 360 scans with k = 5 singular values at SL error level 7.5 %; reference spectrum at NS = 840; difference between denoised and reference spectra. The spectrum is quantitative with an impulsion of 30° and a relaxation delay of 60 s.

Table 4. Peaks integration on noisy (grey rows) and denoised (white rows) spectra of the 50 / 50 MTEOS / TEOS sample for various number of scans (NS). The spectrum is quantitative with an impulsion of 30° and a relaxation delay of 60 s. Spectra were modelled using Dmfit (117).

NS	**Pre-proces-sing**	**PSNR rms**	**Svd thres-hold**	**% T^2**	**% T^3**	**% Q^2**	**% Q^3**	**% Q^4**	**Q^2 / Q^3**
120	cosine	3.9	no	11.5	38.8	3.3	29.1	17.2	0.11
			10 %	12.9	39.8	0	29.2	18.1	0
240	cosine	5.2	no	12.0	37.2	4.7	29.6	16.5	0.16
			10 %	12.5	36.3	5.5	28.8	16.8	0.19
360	cosine	7.1	no	12.3	36.5	4.8	30.2	16.2	0.16
			7.5 %	12.8	35.6	5.5	29.5	16.6	0.19
840	cosine	9.7	no	13.2	35.8	5.4	29.7	15.9	0.18
			7.5 %	13.7	35.2	5.8	29.1	16.1	0.20

E. Conclusion

Singular Value Decomposition is of crucial importance in many mathematical treatments involved in spectrocopies. In this first part (I), SVD with low-rank

approximation was successfully applied to denoise NMR and Raman spectra. This approach can easily be generalised to other spectroscopies. We have shown that a better denoising was obtained with square matrices and with SVD applied to time domain signal rather than to the corresponding frequency spectrum. Automatic thresholding was used thanks to Malinowski's Significant Level indicator and a 7.5 % error value was a good compromise between sensitivity and unwanted artefacts. 6×7380 SVD were carried out to compare pure, noisy and denoised spectra with SNR^{dB} ranging over 41 dB. Our results proved that this technique can detect signals as low as twice $noise_{max}$, *i.e.* with $PSNR_{max} = 2.0$ and $PSNR_{rms} = 6.6$, whatever the peak width. A systematic shape modification has been highlighted for Gaussian peaks with an overestimation of peak area by 20 %. This overestimation for Gaussian peaks is a major result as peak shape is often neglected when denoising, which can give misinterpreted data. A correction step is thus needed if Gaussian / Lorentzian ratio of denoised peak is around 0.5. When used carefully, SVD can lead to similar results between denoised spectra at $PSNR_{rms} = 6.6$ and noisy spectra at quantification limit ($PSNR_{rms} = 10$). As $PSNR_{rms}$ is increasing with the square root of time, this difference is equivalent to a considerable gain on acquisition time of 2.3, which is of paramount importance for low sensitivity experiments.

In a second part (II) (86), we will focus on the computation time needed for SVD treatment under Java, Matlab and Python, using both processors and graphic cards. We will check the influence of algorithms, especially the divide and conquer one, as well as the influence of single precision calculation will be investigated. Software libraries such as MKL (Intel Math Kernel Library) and hardware capabilities such as SSE3 (Streaming SIMD Extensions) (118) will be evaluated. All these optimisations will decrease computation time by a factor of 100.

F. Supplementary material

Figure S1: Influence of SVD iterations; Figure S2: SVD applied to 7380 simulated spectra with truncation; Figure S3: Influence of significance level for automatic thresholding; Figure S4: influence of number of scans and preprocessing; simulated data sets and source codes are available online in (93).

G. Acknowledgements

Sylvie Masse and Cédric Lorthioir are thanked for fruitful discussions.

H. References

1. Bonhomme, C., Gervais, C., and Laurencin, D. (2014) Recent NMR developments applied to organic–inorganic materials. *Prog. Nucl. Magn. Reson. Spectrosc.* 77: 1–48.

2. Lohumi, S., Kim, M. S., Qin, J., and Cho, B.-K. (2017) Raman imaging from microscopy to macroscopy: Quality and safety control of biological materials. *TrAC Trends Anal. Chem.* 93: 183–198.

3. Levitt, M. H. (2008) *Spin dynamics: basics of nuclear magnetic resonance*. 2nd ed. John Wiley & Sons Ltd: Chichester, England.

4. Hoult, D. I., and Richards, R. E. (1976) The signal-to-noise ratio of the nuclear magnetic resonance experiment. *J. Magn. Reson. 1969* 24 (1): 71–85.

5. Heitz, C., Laurent, G., Briard, R., and Barthel, E. (2006) Cross-condensation and particle growth in aqueous silane mixtures at low concentration. *J. Colloid Interface Sci.* 298 (1): 192–201.

6. Laws, D. D., Bitter, H.-M. L., and Jerschow, A. (2002) Solid-state NMR spectroscopic methods in chemistry. *Angew. Chem. Int. Ed.* 41 (17): 3096–3129.

7. Andrew, E. R., Bradbury, A., and Eades, R. G. (1958) Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature* 182 (4650): 1659.

8. Lowe, I. J. (1959) Free induction decays of rotating solids. *Phys. Rev. Lett.* 2 (7): 285–287.

9. Cala-De Paepe, D., Stanek, J., Jaudzems, K., Tars, K., Andreas, L. B., and Pintacuda, G. (2017) Is protein deuteration beneficial for proton detected solid-state NMR at and above 100 kHz magic-angle spinning? *Solid State Nucl. Magn. Reson.* 87: 126–136.

10. Pines, A., Gibby, M. G., and Waugh, J. S. (1972) Proton-enhanced nuclear induction spectroscopy. A method for high resolution NMR of dilute spins in solids. *J. Chem. Phys.* 56 (4): 1776–1777.

11. Mithu, V. S., Pratihar, S., Paul, S., and Madhu, P. K. (2012) Efficiency of heteronuclear dipolar decoupling schemes in solid-state NMR: investigation of effective transverse relaxation times. *J. Magn. Reson.* 220: 8–17.

12. Frydman, L. (2014) High magnetic field science and its application in the United States: a magnetic resonance perspective. *J. Magn. Reson.* 242: 256–264.

13. Lehmann-Horn, J. A., Jacquinot, J.-F., Ginefri, J. C., Bonhomme, C., and Sakellariou, D. (2016) Monolithic MACS micro resonators. *J. Magn. Reson.* 271: 46–51.

14. Hu, K.-N., Song, C., Yu, H., Swager, T. M., and Griffin, R. G. (2008) High-frequency dynamic nuclear polarization using biradicals: a multifrequency EPR lineshape analysis. *J. Chem. Phys.* 128 (5): 052302, 1–17.

15. Rossini, A. J., Zagdoun, A., Lelli, M., Lesage, A., Copéret, C., and Emsley, L. (2013) Dynamic Nuclear Polarization Surface Enhanced NMR Spectroscopy. *Acc. Chem. Res.* 46 (9): 1942–1951.

16. Yarava, J. R., Chaudhari, S. R., Rossini, A. J., Lesage, A., and Emsley, L. (2017) Solvent suppression in DNP enhanced solid state NMR. *J. Magn. Reson.* 277: 149–153.

17. Takahashi, H., Viverge, B., Lee, D., Rannou, P., and De Paëpe, G. (2013) Towards structure determination of self-assembled peptides using dynamic nuclear polarization enhanced solid-state NMR spectroscopy. *Angew. Chem. Int. Ed.* 52 (27): 6979–6982.

18. Gautam, R., Samuel, A., Sil, S., Chaturvedi, D., Dutta, A., Ariese, F., and Umapathy, S. (2015) Raman and mid-infrared spectroscopic imaging: applications and advancements. *Curr. Sci.* 108 (3): 341–356.

19. Hayazawa, N., Inouye, Y., Sekkat, Z., and Kawata, S. (2002) Near-field Raman imaging of organic molecules by an apertureless metallic probe scanning optical microscope. *J. Chem. Phys.* 117 (3): 1296–1301.

20. Vítek, P., Ali, E. M. A., Edwards, H. G. M., Jehlička, J., Cox, R., and Page, K. (2012) Evaluation of portable Raman spectrometer with 1064 nm excitation for geological and forensic applications. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* 86: 320–327.

21. Fleischmann, M., Hendra, P. J., and McQuillan, A. J. (1974) Raman spectra of pyridine adsorbed at a silver electrode. *Chem. Phys. Lett.* 26 (2): 163–166.

22. Sur, U. K., and Chowdhury, J. (2013) Surface-enhanced Raman scattering: overview of a versatile technique used in electrochemistry and nanoscience. *Curr. Sci.* 105 (7): 923–939.

23. Das, R. S., and Agrawal, Y. K. (2011) Raman spectroscopy: Recent advancements, techniques and applications. *Vib. Spectrosc.* 57 (2): 163–176.

24. Ren, B., Liu, G.-K., Lian, X.-B., Yang, Z.-L., and Tian, Z.-Q. (2007) Raman spectroscopy on transition metals. *Anal. Bioanal. Chem.* 388 (1): 29–45.

25. Laor, U., and Schatz, G. C. (1981) The role of surface roughness in surface enhanced Raman spectroscopy (SERS): the importance of multiple plasmon resonances. *Chem. Phys. Lett.* 82 (3): 566–570.

26. Stöckle, R. M., Suh, Y. D., Deckert, V., and Zenobi, R. (2000) Nanoscale chemical analysis by tip-enhanced Raman spectroscopy. *Chem. Phys. Lett.* 318 (1–3): 131–136.

27. Tian, Z.-Q., Ren, B., Li, J.-F., and Yang, Z.-L. (2007) Expanding generality of surface-enhanced Raman spectroscopy with borrowing SERS activity strategy. *Chem. Commun.* (34): 3514–3534.

28. Duncan, M. D., Reintjes, J., and Manuccia, T. J. (1982) Scanning coherent anti-Stokes Raman microscope. *Opt. Lett.* 7 (8): 350–352.

29. Kano, H., Segawa, H., Leproux, P., and Couderc, V. (2014) Linear and nonlinear Raman microspectroscopy: history, instrumentation, and applications. *Opt. Rev.* 21 (6): 752–761.

30. Perez-Rodriguez, J. L., Robador, M. D., Centeno, M. A., Siguenza, B., and Duran, A. (2014) Wall paintings studied using Raman spectroscopy: a comparative study between various assays of cross sections and external layers. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* 120: 602–609.

31. Krafft, C., Kirsch, M., Beleites, C., Schackert, G., and Salzer, R. (2007) Methodology for fiber-optic Raman mapping and FTIR imaging of metastases in mouse brains. *Anal. Bioanal. Chem.* 389 (4): 1133–1142.

32. Hardis, R., Jessop, J. L. P., Peters, F. E., and Kessler, M. R. (2013) Cure kinetics characterization and monitoring of an epoxy resin using DSC, Raman spectroscopy, and DEA. *Compos. PartA - Appl. Sci. Manuf.* 49: 100–108.

33. Byrne, C. L. (2013) Mathematics of signal processing: a first course. Available at: http://faculty.uml.edu/cbyrne/SP1text.pdf (accessed 29 March 2018).

34. Ernst, R. R., and Anderson, W. A. (1966) Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Rev. Sci. Instrum.* 37 (1): 93–102.

35. Freeman, R. (1993) The Fourier transform revolution in NMR spectroscopy. *Anal. Chem.* 65 (17): 743A-753A.

36. Kupče, E., Nishida, T., and Freeman, R. (2003) Hadamard NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* 42 (3–4): 95–122.

37. Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. (2008) Compressed sensing MRI. *IEEE Signal Process. Mag.* 25 (2): 72–82.

38. Mobli, M., and Hoch, J. C. (2014) Nonuniform sampling and non-Fourier signal processing methods in multidimensional NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* 83: 21–41.

39. Malfait, W. J., and Halter, W. E. (2008) Increased ^{29}Si NMR sensitivity in glasses with a Carr–Purcell–Meiboom–Gill echotrain. *J. Non-Cryst. Solids* 354 (34): 4107–4114.

40. Liland, K. H., Kohler, A., and Afseth, N. K. (2016) Model-based pre-processing in Raman spectroscopy of biological samples. *J. Raman Spectrosc.* 47 (6): 643–650.

41. Kandjani, A. E., Griffin, M. J., Ramanathan, R., Ippolito, S. J., Bhargava, S. K., and Bansal, V. (2013) A new paradigm for signal processing of Raman spectra using a smoothing free algorithm: Coupling continuous wavelet transform with signal removal method. *J. Raman Spectrosc.* 44 (4): 608–621.

42. Ryabchykov, O., Bocklitz, T., Ramoji, A., Neugebauer, U., Foerster, M., Kroegel, C., Bauer, M., Kiehntopf, M., and Popp, J. (2016) Automatization of spike correction in Raman spectra of biological samples. *Chemom. Intell. Lab. Syst.* 155: 1–6.

43. Schulze, H. G., Konorov, S. O., Okuda, K., Piret, J. M., Blades, M. W., and Turner, R. F. B. (2012) A volume-exclusion normalization procedure for quantitative Raman confocal microspectroscopy of immersed samples applied to human embryonic stem cells. *J. Raman Spectrosc.* 43 (3): 360–369.

44. Chatterjee, P., and Milanfar, P. (2010) Is denoising dead? *IEEE Trans. Image Process.* 19 (4): 895–911.

45. Perriot, A., Vandembroucq, D., Barthel, E., Martinez, V., Grosvalet, L., Martinet, C., and Champagnon, B. (2006) Raman microspectroscopic characterization of amorphous silica plastic behavior. *J. Am. Ceram. Soc.* 89 (2): 596–601.

46. Davis, M. C., Kaseman, D. C., Parvani, S. M., Sanders, K. J., Grandinetti, P. J., Massiot, D., and Florian, P. (2010) $Q^{(n)}$ species distribution in $K_2O{\cdot}2SiO_2$ glass by ^{29}Si magic angle flipping NMR. *J. Phys. Chem. A* 114 (17): 5503–5508.

47. Lindon, J. C., and Ferrige, A. G. (1980) Digitisation and data processing in Fourier transform NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* 14 (1): 27–66.

48. Člupek, M., Matějka, P., and Volka, K. (2007) Noise reduction in Raman spectra: finite impulse response filtration versus Savitzky–Golay smoothing. *J. Raman Spectrosc.* 38 (9): 1174–1179.

49. Donoho, D. L., Johnstone, I. M., Stern, A. S., and Hoch, J. C. (1990) Does the maximum entropy method improve sensitivity? *Proc. Natl. Acad. Sci.* 87 (13): 5066–5068.

50. Takeda, K. (2015) Solid-state covariance NMR spectroscopy. In *Annual Reports on NMR Spectroscopy*, Webb, G.A., Ed., Academic Press, 84, pp 77–113.

51. Chen, S., Lin, X., Yuen, C., Padmanabhan, S., Beuerman, R. W., and Liu, Q. (2014) Recovery of Raman spectra with low signal-to-noise ratio using Wiener estimation. *Opt. Express* 22 (10): 12102–12114.

52. Barache, D., Antoine, J.-P., and Dereppe, J.-M. (1997) The continuous wavelet transform, an analysis tool for NMR spectroscopy. *J. Magn. Reson.* 128 (1): 1–11.

53. Chiron, L., Agthoven, M. A. van, Kieffer, B., Rolando, C., and Delsuc, M.-A. (2014) Efficient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry. *Proc. Natl. Acad. Sci.* 111 (4): 1385–1390.

54. Tufts, D. W., Kumaresan, R., and Kirsteins, I. (1982) Data adaptive signal estimation by singular value decomposition of a data matrix. *Proc. IEEE* 70 (6): 684–685.

55. Cadzow, J. A. (1988) Signal enhancement – a composite property mapping algorithm. *IEEE Trans. Acoust. Speech Signal Process.* 36 (1): 49–62.

56. Tufts, D. W., and Shah, A. A. (1993) Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix. *IEEE Trans. Signal Process.* 41 (4): 1716–1721.

57. Stewart, G. W. (1993) On the early history of the singular value decomposition. *SIAM Rev.* 35 (4): 551–566.

58. Beltrami, E. (1873) Sulle funzioni bilineari. *G. Mat. Ad Uso Degli Stud. Delle Univ.* 11 (2): 98–106.

59. Jordan, C. (1874) Mémoire sur les formes bilinéaires. *J. Mathématiques Pures Appliquées* 19: 35–54.

60. Lanczos, C. (1958) Linear Systems in Self-Adjoint Form. *Am. Math. Mon.* 65 (9): 665–679.

61. Läuchli, P. (1961) Jordan-elimination und ausgleichung nach kleinsten quadraten. *Numer. Math.* 3 (1): 226–240.

62. Suykens, J. A. K. (2016) SVD revisited: A new variational principle, compatible feature maps and nonlinear extensions. *Appl. Comput. Harmon. Anal.* 40 (3): 600–609.

63. Condat, L., and Hirabayashi, A. (2015) Cadzow denoising upgraded: a new projection method for the recovery of Dirac pulses from noisy linear measurements. *Sampl. Theory Signal Image Process.* 14 (1): 17–47.

64. Hansen, P. C., and Jensen, S. H. (2007) Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: survey and analysis. *EURASIP J. Adv. Signal Process.* 2007 (1): 092953, 1–24.

65. Chen, Y., Zhang, D., Jin, Z., Chen, X., Zu, S., Huang, W., and Gan, S. (2016) Simultaneous denoising and reconstruction of 5-D seismic data via damped rank-reduction method. *Geophys. J. Int.* 206 (3): 1695–1717.

66. Park, E. S., Henry, R. C., and Spiegelman, C. H. (1999) Determining the number of major pollution sources in multivariate air quality receptor models. (Technical report NRCSE-TRS No. 034) Available at: http://www.nrcse.washington.edu/pdf/trs34_receptor.pdf (accessed 29 March 2018).

67. Clifford, G. D. (2005) Singular value decomposition & independent component analysis for blind source separation. Available at: http://www.mit.edu/~gari/teaching/6.222j/ICASVDnotes.pdf (accessed 29 March 2018).

68. Aharon, M., Elad, M., and Bruckstein, A. (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54 (11): 4311–4322.

69. Baumann, M. (2010) *Real-time robust principal component analysis for video surveillance* (Master thesis), Institut für Kommunikationstechnik, Eidgenössische Technische Hochschule, Zürich, Switzerland.

70. Nguyen, H. M. (2011) *Towards high-resolution magnetic resonance spectroscopic imaging: spatiotemporal denoising and echo-time selection* (PhD dissertation), Electrical & Computer Eng, University of Illinois, Urbana-Champaign, USA.

71. Schulze, H. G., and Turner, R. F. B. (2015) Development and integration of block operations for data invariant automation of digital preprocessing and analysis of biological and biomedical Raman spectra. *Appl. Spectrosc.* 69 (6): 643–664.

72. Tulloch, A. (2014) Fast randomized SVD. Available at: https://research.fb.com/fast-randomized-svd/ (accessed 29 March 2018).

73. Brissac, C., Malliavin, T. E., and Delsuc, M. A. (1995) Use of the Cadzow procedure in 2D NMR for the reduction of t_1 noise. *J. Biomol. NMR* 6 (4): 361–365.

74. Stoica, P., Sandgren, N., Selén, Y., Vanhamme, L., and Van Huffel, S. (2003) Frequency-domain method based on the singular value decomposition for frequency-selective NMR spectroscopy. *J. Magn. Reson.* 165 (1): 80–88.

75. Nikel, O., Laurencin, D., Bonhomme, C., Sroga, G. E., Besdo, S., Lorenz, A., and Vashishth, D. (2012) Solid state NMR investigation of intact human bone quality: balancing issues and insight into the structure at the organic–mineral interface. *J. Phys. Chem. C* 116 (10): 6320–6331.

76. Zou, Y., and Xie, R. (2015) A novel method for NMR data compression. *Comput. Geosci.* 19 (2): 389–401.

77. Uzunbajakava, N., Lenferink, A., Kraan, Y., Volokhina, E., Vrensen, G., Greve, J., and Otto, C. (2003) Nonresonant confocal Raman imaging of DNA and protein distribution in apoptotic cells. *Biophys. J.* 84 (6): 3968–3981.

78. Palacký, J., Mojzeš, P., and Bok, J. (2011) SVD-based method for intensity normalization, background correction and solvent subtraction in Raman spectroscopy exploiting the properties of water stretching vibrations. *J. Raman Spectrosc.* 42 (7): 1528–1539.

79. Khmaladze, A., Jasensky, J., Price, E., Zhang, C., Boughton, A., Han, X., Seeley, E., Liu, X., Holl, M. M. B., and Chen, Z. (2014) Hyperspectral imaging and characterization of live cells by broadband Coherent Anti-Stokes Raman Scattering (CARS) microscopy with Singular Value Decomposition (SVD) analysis. *Appl. Spectrosc.* 68 (10): 1116–1122.

80. Caraher, M. C., Sophocleous, A., Beattie, J. R., O'Driscoll, O., Cummins, N. M., Brennan, O., O'Brien, F. J., Ralston, S. H., Bell, S. E. J., Towler, M., and Idris, A. I. (2018) Raman spectroscopy predicts the link between claw keratin and bone collagen structure in a rodent model of oestrogen deficiency. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1864 (2): 398–406.

81. Man, P. P., Bonhomme, C., and Babonneau, F. (2014) Denoising NMR time-domain signal by singular-value decomposition accelerated by graphics processing units. *Solid State Nucl. Magn. Reson.* 61–62: 28–34.

82. Man, P. P. (2012) 2012 Java application for FID denoising with SVD. Available at: http://pascal-man.com/navigation/faq-java-browser/SVD-Java-application2012.shtml (accessed 14 May 2018).

83. Man, P. P. (2012) Graphic processing unit (GPU), 2012 SVD Java 7 application for FID denoising. Available at: http://www.pascal-man.com/navigation/faq-java-browser/SVD-Java-application-GPU.shtml (accessed 14 May 2018).

84. Nickolls, J., Buck, I., Garland, M., and Skadron, K. (2008) Scalable parallel programming with CUDA. *Queue* 6 (2): 40–53.

85. Laurent, G. (2016) SVD Performances to denoise NMR and Raman spectra. Available at: https://hal.sorbonne-universite.fr/hal-01277387 (accessed 14 May 2018).

86. Laurent, G., Gilles, P.-A., Woelffel, W., Barret-Vivin, V., Gouillart, E., and Bonhomme, C. Denoising applied to spectroscopies – part II: decreasing computation time. *Appl. Spectrosc. Rev.* submitted.

87. Figueira, R. B., Fontinha, I. R., Silva, C. J. R., and Pereira, E. V. (2016) Hybrid Sol-Gel Coatings: Smart and Green Materials for Corrosion Mitigation. *Coatings* 6 (1): 12, 1–19.

88. Owens, G. J., Singh, R. K., Foroutan, F., Alqaysi, M., Han, C.-M., Mahapatra, C., Kim, H.-W., and Knowles, J. C. (2016) Sol–gel based materials for biomedical applications. *Prog. Mater. Sci.* 77: 1–79.

89. Ben Ahmed, N., Masse, S., Laurent, G., Piquemal, J.-Y., Yéprémian, C., Brayner, R., and Coradin, T. (2018) Optical microalgal biosensors for aqueous contaminants using organically doped silica as cellular hosts. *Anal. Bioanal. Chem.* 410 (4): 1205–1216.

90. Faustini, M., Nicole, L., Boissière, C., Innocenzi, P., Sanchez, C., and Grosso, D. (2010) Hydrophobic, antireflective, self-cleaning, and antifogging sol−gel coatings: an example of multifunctional nanostructured materials for photovoltaic cells. *Chem. Mater.* 22 (15): 4406–4413.

91. Fung, B. M., Khitrin, A. K., and Ermolaev, K. (2000) An improved broadband decoupling sequence for liquid crystals and solids. *J. Magn. Reson.* 142 (1): 97–101.

92. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12: 2825−2830.

93. Laurent, G., Gilles, P.-A., Woelffel, W., Barret-Vivin, V., Gouillart, E., and Bonhomme, C. (2018) Denoising applied to spectroscopies: parts I and II - programs and datas. Available at: http://doi.org/10.5281/zenodo.1406172 (accessed 30 August 2018).

94. Petrakis, L. (1967) Spectral line shapes: Gaussian and Lorentzian functions in magnetic resonance. *J. Chem. Educ.* 44 (8): 432–436.

95. Golub, G., and Kahan, W. (1965) Calculating the singular values and pseudo-inverse of a matrix. *J. Soc. Ind. Appl. Math. Ser. B Numer. Anal.* 2 (2): 205–224.

96. Brown, D. E., and Campbell, T. W. (1990) Enhancement of 2D NMR spectra using singular value decomposition. *J. Magn. Reson. 1969* 89 (2): 255–264.

97. Heinig, G., and Rost, K. (1984) Toeplitz and Hankel matrices. In *Algebraic methods for Toeplitz-like matrices and operators*, Birkhäuser Basel, pp 9–135.

98. Karner, H., Schneid, J., and Ueberhuber, C. W. (2003) Spectral decomposition of real circulant matrices. *Linear Algebra Its Appl.* 367: 301–311.

99. Ross, R. (1996) Bias and standard deviation due to Weibull parameter estimation for small data sets. *IEEE Trans. Dielectr. Electr. Insul.* 3 (1): 28–42.

100. Priyadarshani, N., Marsland, S., Castro, I., and Punchihewa, A. (2016) Birdsong denoising using wavelets. *PLoS ONE* 11 (1): e0146790, 1–26.

101. Shrivastava, A., and Gupta, V. B. (2011) Methods for the determination of limit of detection and limit of quantitation of the analytical methods. *Chron. Young Sci.* 2 (1): 21–25.

102. Hyberts, S. G., Robson, S. A., and Wagner, G. (2013) Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *J. Biomol. NMR* 55 (2): 167–178.

103. Currie, L. A. (1968) Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal. Chem.* 40 (3): 586–593.

104. Larsen, F. H., Skibsted, J., Jakobsen, H. J., and Nielsen, N. C. (2000) Solid-state QCPMG NMR of low-γ quadrupolar metal nuclei in natural abundance. *J. Am. Chem. Soc.* 122 (29): 7080–7086.

105. Hassani, H. (2007) Singular spectrum analysis: methodology and comparison. *J. Data Sci.* 5 (2): 239–257.

106. Van Huffel, S., Chen, H., Decanniere, C., and Vanhecke, P. (1994) Algorithm for time-domain NMR data fitting based on total least squares. *J. Magn. Reson. A* 110 (2): 228–237.

107. Faber, N. M., Buydens, L. M. C., and Kateman, G. (1994) Aspects of pseudorank estimation methods based on the eigenvalues of principal component analysis of random matrices. *Chemom. Intell. Lab. Syst.* 25 (2): 203–226.

108. Malinowski, E. R. (2002) *Factor analysis in chemistry*. 3rd ed. Wiley: Hoboken, NJ, USA.

109. Cai, J.-F., Candès, E. J., and Shen, Z. (2010) A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* 20 (4): 1956–1982.

110. Sanliturk, K. Y., and Cakar, O. (2005) Noise elimination from measured frequency response functions. *Mech. Syst. Signal Process.* 19 (3): 615–631.

111. Maxim. (2016) Peak fitting to either Voigt or LogNormal line shapes. Available at: http://fr.mathworks.com/matlabcentral/fileexchange/52321-peak-fitting-to-either-voigt-or-lognormal-line-shapes (accessed 29 March 2018).

112. Dong, Z., Dreher, W., Leibfritz, D., and Peterson, B. S. (2009) Challenges of using MR spectroscopy to detect neural progenitor cells in vivo. *Am. J. Neuroradiol.* 30 (6): 1096–1101.

113. Garroway, A. N. (1977) Homogeneous and inhomogeneous nuclear spin echoes in organic solids: Adamantane. *J. Magn. Reson. 1969* 28 (3): 365–371.

114. Lipton, A. S., Sears, J. A., and Ellis, P. D. (2001) A General Strategy for the NMR Observation of Half-Integer Quadrupolar Nuclei in Dilute Environments. *J. Magn. Reson.* 151 (1): 48–59.

115. Palmer, M. R., Suiter, C. L., Henry, G. E., Rovnyak, J., Hoch, J. C., Polenova, T., and Rovnyak, D. (2015) Sensitivity of Nonuniform Sampling NMR. *J. Phys. Chem. B* 119 (22): 6502–6515.

116. Hassanpour, H., Zehtabian, A., and Sadati, S. J. (2012) Time domain signal enhancement based on an optimized singular vector denoising algorithm. *Digit. Signal Process.* 22 (5): 786–794.

117. Massiot, D., Fayon, F., Capron, M., King, I., Le Calvé, S., Alonso, B., Durand, J.-O., Bujoli, B., Gan, Z., and Hoatson, G. (2002) Modelling one- and two-dimensional solid-state NMR spectra. *Magn. Reson. Chem.* 40 (1): 70–76.

118. Hofmann, J., Treibig, J., Hager, G., and Wellein, G. (2014) Comparing the performance of different x86 SIMD instruction sets for a medical imaging application on modern multi- and manycore chips. *Proceedings of the 2014 Workshop on Programming Models for SIMD/Vector Processing*, 57–64, ACM: New York, NY, USA.