Tina Memo No. 2007-002
Internal

# Padé Approximant: How to... and Why.

D.C.Williamson and N.A.Thacker

Last updated
21 / 03 / 2007


Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

# Padé Approximant: How to... and Why.

D.C.Williamson, N.A.Thacker 21/3/2007

## Abstract

*This document has been written, in response to requests for more detailed information from researchers in this area, as a supplement to previous publications [1]. We give details regarding the practical implementation of the Padé Approximant for the analysis of MR data (How to..). It is tempting to think that the use of the Padé transform is somehow intrinsically justified by the mathematical form of the algebra. This notion can be dispelled by making the effort to relate the process to conventional statistics, we do this here in the appendices (.. and Why). We believe that this information will help others to get the best results from their own implementations.*

## How to...

MR spectroscopy data is expected to be composed of a super-position of resonances at specific frequencies and we would like to determine these and their amplitudes. One approach is simply to look for peaks (usually Lorentzian) in the measured Fourier data. In the approach detailed here this is done by approximating the data with a rational polynomial, the Padé Approximant.

The discrete Fourier transform can be written as a polynomial in $x$, where $x = exp(-i\omega\tau)$, and $\tau$ is the dwell time

$$F(\omega) \;=\; \sum_{n=0}^{N} a_n x^n \qquad (1)$$

where the $a_n$ are the FID data. This series is a truncation of the infinite series which is equivalent to the full Fourier integral. The Padé Approximant is constructed to approximate the infinite series,

$$\sum_{n=0}^{N} a_n x^n \;=\; \frac{\sum_{p=0}^{P} b_p x^p}{\sum_{q=0}^{Q} c_q x^q} \qquad (2)$$

where $P + Q = N$. In general, provide $P + Q = N$, $P$ and $Q$ can take any value offering a range of possible PA, e.g., $P = 0$, $Q = N$ is the Linear Prediction polynomial. With respect to NMR spectroscopy, $P = Q = N/2$ has been provided the most promising results[1].

To determine the coefficients $b_p$ and $c_q$ both sides of equation 2 are multiplied by the denominator of the right hand side and then we equate terms of the same order in $x$.

This leads to two sets of equations:

$$\sum_{q=0}^{Q} c_q a_{p-q} \;=\; b_p \qquad for \qquad p \;=\; 0, ...P$$

and

$$\sum_{q=0}^{Q} c_q a_{p-q} \;=\; 0 \qquad for \qquad p \;=\; P+1, ...P+Q$$

$C_0$ can be set as 1, since it is simply a scaling factor in equation 2. This allows the rearrangement of the second set of equations into the following matrix form.

$$\begin{bmatrix} a_p & a_{p-1} & ... & a_{P-Q+1} \\ a_{p+1} & a_p & ... & a_{P-Q+2} \\ ... & ... & & ... \\ a_{P+Q-1} & a_{P+Q-2} & ... & a_P \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ ... \\ c_Q \end{bmatrix} = - \begin{bmatrix} a_{P+1} \\ a_{P+2} \\ ... \\ a_{P+Q} \end{bmatrix} \qquad (3)$$

[1] Though it is tempting to interpret this as some indication of intrinsic suitability of the approximation, it may be a function of the degree to which the method approximates a valid Likelihood estimate. An indication of how this might arise is provided below.

Writing this a $DC = -A$, this matrix equation is solved by inverting matrix D and rearranging the equation $C = -D^{-1}A$. This approach is common in many data analysis systems and would probably be accepted without question by many on the basis of familiarity. However, such analytic approaches to estimation are not generally statistical valid as they take no account of the noise in the measured $a_n$ data. This will be discussed further below, for more details regarding this general issue see [4]. Once the values of $c_q$ have been determined, the coefficients $b_p$ can then be found directly from the first set of equations.

There is a slight complication concerning the NMR data. In general all the coefficients are complex numbers. The matrix inversion and multiplication can all be carried out as written, provided a full suite of complex algebra is available. There is, however, a more straight forward method. If we consider each coefficient $(a_j, b_j, c_j)$ to be a complex number, then we can separate the real and imaginary parts as

$$a_j \;=\; (a_j^R, a_j^i) \quad b_j \;=\; (b_j^R, b_j^i) \quad c_j \;=\; (c_j^R, c_j^i)$$

and define the following matrices

$$D^r \;=\; \begin{bmatrix} a_P^r & a_{P-1}^r & ... & a_{P-Q+1}^r \\ a_{P+1}^r & a_P^r & ... & a_{P-Q+2}^r \\ ... & ... & & ... \\ a_{P+Q-1}^r & a_{P+Q-2}^r & ... & a_P^r \end{bmatrix} \quad D^i \;=\; \begin{bmatrix} a_P^i & a_{P-1}^i & ... & a_{P-Q+1}^i \\ a_{P+1}^i & a_P^i & ... & a_{P-Q+2}^i \\ ... & ... & & ... \\ a_{P+Q-1}^i & a_{P+Q-2}^i & ... & a_P^i \end{bmatrix}$$

$$C^r \;=\; \begin{bmatrix} c_1^r \\ c_2^r \\ ... \\ c_Q^r \end{bmatrix} \quad C^i \;=\; \begin{bmatrix} c_1^i \\ c_2^i \\ ... \\ c_Q^i \end{bmatrix} \quad A^r \;=\; \begin{bmatrix} a_1^r \\ a_2^r \\ ... \\ a_Q^r \end{bmatrix} \quad A^i \;=\; \begin{bmatrix} a_1^i \\ a_2^i \\ ... \\ a_Q^i \end{bmatrix}$$

Using this set of matrices we redefine

$$D \;=\; \begin{bmatrix} D^r & D^i \\ -D^i & D^r \end{bmatrix} \quad C \;=\; \begin{bmatrix} C^r \\ C_i \end{bmatrix} \quad A \;=\; \begin{bmatrix} A^r \\ A_i \end{bmatrix}$$

giving the same matrix equation, $DC = -A$ but using the new matrices. The solution is identical, $C = -D^{-1}A$.

The inversion of matrix D is not trivial, and as stated above is potentially statistically invalid. However, we can regain some statistical validity if the inversion is carried out using singular value decomposition (SVD). This approach can be considered as a hyperplane fit to the data on the assumption of uniform independent noise on the linear constraint equations [2] (See Appendix A).

Any SVD or numerical inversion routine will require something called a condition number. The condition number controls the accuracy with which the inversion is performed. It can be used to control the amount of noise which is retained in the spectrum. A large condition number reduces the amount of noise (and possible removes genuine small peaks) from the spectrum, while a very small condition number retains all the data plus all the noise and will produce an identical spectrum to the Fourier transform (at an arbitrary spectral resolution) but minus any Gibbs ringing due to truncation of the Fourier series. Unfortunately changing the condition number does not provide a systematic method for noise reduction, it is something that you need to play with to get a feel for how it affects the spectrum.

There is a systematic method which is based upon our statistical interpretation of SVD , but this relies on having access to the results of the SVD and being able to use them to generate the matrix inverse manually, as opposed to passing the matrices to a "black box" subroutine for matrix inversion (Appendix B).

The complex frequency of signal components can be identified as poles in the PA. These are most easily identified as the roots of the denominator polynomial. The most robust method for finding complex roots of complex polynomial is the Laguerre method, which is detailed in Numerical Recipes [3] but most maths packages will have an implementation. If $x_i$ represents the set of roots, then the frequencies can be determined from the definition of $x$ and are

$$\omega_i \;=\; i\; ln(x_i)/\tau$$

Given the frequency, the associated amplitude (peak area) can be obtained from

$$d_i \;=\; \frac{P(x_i)}{x_i Q'(x_i)} \quad where \quad P(x) \;=\; \sum_{p=0}^{P} b_p x^p \quad and \quad Q(x) \;=\; \sum_{q=0}^{Q} c_q x^q \quad and \quad Q'(x) \;=\; \frac{dQ(x)}{dx}$$

## Monte Carlo:

The Padé will provide a set of frequencies and amplitudes for a given data set however it does not tell us anything about the error on these estimates, nor does it tell us about the stability of each peak. As the Padé approach is only an approximation to a Likelihood estimation, it is possible for the Padé to produce spurious peaks due to specific patterns of noise. It is also likely to model peaks that deviate substantially from a Lorentzian lineshape (macromolecule baseline signals) as an arbitrary sum of Lorentzians. However, we can regain a quantitative understanding of the process sufficient for a scientific interpretation of the major resonance processes by assessing the stability of the solution. This can be done using a general Monte Carlo method.

From a single data set, we produce, say 100, synthetic data sets by adding Gaussian noise at the level estimated from the original peak. Each of these spectra are analysed using the PA. A 2D histogram of frequencies and amplitudes is constructed which allows us to identify which peaks are stable with respect to the noise and which are not. This also provides a mean and standard deviation for the peak amplitudes. There is more detail about the Monte Carlo in the paper [1].

## Appendix A: The Assumption of Uniform Independent Noise.

The linear constraints embodied by equation (3) can be considered as a difference between a predicted and observed data value in a $Q$ dimensional space. To determine an estimate of the required model parameters $C$, we should not just treat this as an exercise in algebra as this would not be consistent with a scientific anlaysis of data[2]. Instead we need to apply quantitative statistical methods and optimise the likelihood that these two points are consistent with the measurement noise on all $a_n$ (which can safely be assumed to be uniform independant and ranom Gaussian with standard deviation of $\sigma_{noise}$ for MR spectroscopy data ). In this case, as both sides of (3) contain measurements this is a potentially complicated problem.

However, we can use the method of error propagation to determine the structure of the correlation matrix describing expected errors on the difference between the model and data. For example for the system of measurements $(u, v, w, x, y, z)$ equation (3) becomes

$$\begin{bmatrix} w & v & u \\ x & w & v \\ y & x & w \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \Delta$$

where $\Delta$ is the vector residual. This has an expected covariance of

$$\Sigma = \sigma^2_{noise} \begin{bmatrix} a^2+b^2+c^2+1 & cb+ab+a & ac+b \\ cb+ab+a & a^2+b^2+c^2+1 & cb+ab+a \\ ac+b & cb+ab+a & a^2+b^2+c^2+1 \end{bmatrix}$$

Strictly we should estimate the parameters $C$ by minimising the Mahalanobis distance in the data space on the basis of this covariance (often called the Variational Method).

$$-\, log\ L \ = \ \Delta^T \Sigma^{-1} \Delta$$

This is a circular problem as we need $C$ to compute $C$, and it would generally require an iterative solution. However, in general the structure of the problem tells us that the diagonal terms of $\Sigma$ will always be 1 plus the sum of squared $C$ elements. The off diagonal terms will be sums of products of $C$ elements which decrease in number away from the diagonal. Therefore, to a first approximation the covariance matrix is diagonal which is equivalent to an assumption of uniform independent noise on the linear constraints, which is in turn consistent with the use of SVD. The solution of (3) as the basis for data analysis is only valid provided this assumption produces a negligible change from the Likelihood solution. This approximation will improve for increasing $Q$ and performance (statistical stability) of the method is expected to be a function of the value of $P + Q = N$. It is reasonable to speculate that the observed enhanced behaviour for $P = Q = N/2$ is at least partly a consequence of this varying approximation to Likelihood.

[2]This very common error is often pointed out politely at conferences in the form of the question; "How does this method perform in the presence of noise?". Equally however, attempts to address these issues in a journal publication by deriving the method from statistical principles such as Likelihood, will often be attacked by reviewers on the basis of "lack of novely". Sadly, as with our publication [1], it is easier to get material published if we simply describe and evaluate the new technique and leave statistical interpretation to the informed reader.

## Appendix B: SVD noise reduction.

Given a matrix D, generally m x n in dimensions, SVD allows us to factorise the matrix as follows:

$$D \;=\; UWV^{*}$$

Here $U$ and $V$ are unitary $m$ by $m$ and $n$ by $n$ matrices (the * denotes the conjugate transpose) and $W$ is a diagonal $m$ by $n$ matrix. If we describe $U$ and $V$ as some kind of orthogonal components and $W$ is the weighting of each component. It is the $W$ matrix that is important.

When inverting the matrix

$$D^{-1} \;=\; (UWV^{*}) \;=\; (V^{*})^{-1}W^{-1}U^{-1} \;=\; VW^{-1}U^{-1}$$

By definition the inverse of a unitary matrix is equal to it's transpose, allowing a straightforward inversion of $U$ and $V^{*}$ and the inverse of $W$ can be written as:

$$D \;=\; \begin{bmatrix} 1/w_1 & 0 & ... & 0 \\ 0 & 1/w_2 & & ... \\ ... & & ... & 0 \\ 0 & ... & 0 & 1/w_n \end{bmatrix}$$

where $w_i$ is an element of the $W$.

In order to eliminate the noise in the inversion all terms in the inverse of $W$ are set to zero for

$$1/w_i \;=\; 0 \qquad for \quad i > j \quad s.t. \quad \sum_{i=j}^{n} 1/w_i \;=\; \sigma_{noise}$$

where (noise is an estimate of the noise on the data. In practice this means estimating the noise from the tail of the FID or a part of the spectrum with no signal and sum the elements in the inverse of $W$ from bottom right diagonally upwards until the sum equal the noise. At this point all elements which contributed to the sum are set equal to zero and the modified matrix is used to generate the inverse of $D$.

# References


[1] D.C.Williamson and N.A.Thacker, Assessment of the Padé Approximant as a Method for Quantifying $^{1}$H Magnetic Resonance Spectroscopic Data Tina memo 2003-008, www.tina-vision.net, 2003.

[2] M.E.Tipping and C.M.Bishop, Probabilistic Principle Component Analysis, Jou. Royal Stat. Soc., B, 6, 3, 611-622, 1999.

[3] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C, The Art of Scientific Computing", Cambridge University Press, second edition, New York, 1992.

[4] A. Nayak, E. Trucco and N.A. Thacker, Comparing the Performance of Least-Squares Estimators: when is GTLS Better than LS? Tina memo 2005-011, www.tina-vision.net, 2005.