ADVANTAGES OF USING MULTIPLE REGRESSION FOR DISCRETE FOURIER ANALYSIS[1]

Ricardo Joseph Simeoni and Peter Michael Mills

School of Physiotherapy and Exercise Science, Griffith University, Queensland 9726

e-mail: r.simeoni@mailbox.gu.edu.au

Abstract

This study presents a Discrete Fourier Analysis method that treats the Fourier expansion of a waveform as a long multivariable equation. A multiple regression (MR) procedure is then used to evaluate the Fourier expansion coefficients.

The MR method of Discrete Fourier Analysis allows for flexibility in the selection of the constituent frequencies, frequency resolution and maximum frequency of computed frequency-domain spectra and does not require uniformly spaced time-domain data. The MR method also offers an alternative means of teaching Fourier Analysis to students from non-traditional mathematics backgrounds, particularly if the students are familiar with multivariable equations and MR. A student group focused on in the study are Exercise Science students who generally have difficulty understanding the Fourier Analysis theory required for topics within biomechanics, neurology and motor control.

MR frequency-domain spectra closely match those calculated using a standard Fast Fourier Transform (FFT) and display less spectral leakage than FFT spectra. The FFT algorithm is computationally more efficient.

Introduction

The Fourier transformation, $F(\omega)$, of a time-varying waveform, $f(t)$, of duration, T, into the frequency-domain is usually expressed in terms of the Fourier integral:

$$F(\omega) = \frac{2}{T}\int_0^T f(t)e^{-i\omega t}dt, \tag{1}$$

or its discrete counterpart, the Discrete Fourier Transform (DFT):

$$F(\omega) = \frac{2}{N}\sum_{j=1}^{N} f(t_j)e^{-i\omega_n t_j} = \frac{2}{N}\sum_{j=1}^{N}\Big(f(t_j)cos(\omega_n t_j) + if(t_j)sin(\omega_n t_j)\Big) \quad (n = 1, \ldots N/2), \tag{2}$$

where $F(\omega_n)$ is the n^{th} harmonic, ω_n is the angular frequency of the n^{th} harmonic, N is the number of discrete time-domain measurements of $f(t)$ and t is time.

For existing computational algorithms such as the Fast Fourier Transform (FFT), the frequency resolution, $\Delta(\omega)$, of the calculated frequency-domain spectrum is dependent on N and the sampling frequency, ω_s, such that $\Delta\omega = \omega_s/N$. The frequency–domain spectrum is calculated for all ω_n up to a maximum of $\omega_{\max} = \omega_s/2$.

[1]Presented in modified form at the MATLAB User Conference, VIC, November 2000

The dependence of $\Delta\omega$ and $\omega_{\max}$ on N and ω_s may restrict the option for their adjustment. For example, the frequency-domain spectrum of an electroencephalogram (EEG) biosignal sampled at 125.0 Hz for 4.0 seconds (that is, $N = 500$) will have Δ_ω and $\omega_{\max}$ values of 0.25 and 62.50 Hz respectively. If one wished to double Δ_ω, then N would need to be halved, resulting in the omission of some data. The example also shows that since $\omega_{\max}$ is dependent on ω_s, which in a clinical situation may be set, $\omega_{\max}$ may not match the actual maximum frequency present in the analysed waveform. For the given example, $\omega_{\max}$ exceeds the maximum frequency (approximately 40 Hz) of many EEG biosignals. (Note that a prefix of $2p$ is implied for all stated angular frequency values.)

The use of large-scale multiple regression (MR) to calculate $F(\omega_n)$ theoretically allows for comparatively more flexibility in the selection of constituent $\omega_n, \Delta\omega$ and $\omega_{\max}$ and does not require uniformly spaced time-domain data. Although some curve-fitting software packages (for example, MATLAB curve-fitting tool box) fit trigonometric expansions to data points, large-scale MR for the calculation of high-resolution Fourier spectra is not typically an option.

The usage of Fourier Analysis by students from non-traditional mathematics backgrounds is increasing. However, a significant number of these students have difficulty understanding the concepts and theory of Fourier Analysis and how Equations (1) and (2) have the capacity to transform a waveform into the frequency-domain. The theory of the FFT, involving the construction of the Fourier Matrix, may appear even more abstract.

Multivariable equations in contrast are common to the field of Exercise Science (for example, body surface area as a function of height and mass; $\dot{V}O_{2\,\max}$ as a function of speed, age and speed$\times$age; percent body fat as a function of various anthropometric measurements). Accordingly, our Bachelor of Exercise Science students perceive MR to be clinically relevant and learn to generate multivariable equations from clinical data using MATLAB and EXCEL software. Since Fourier Analysis can be explained as a large-scale MR, such an explanation may assist in the learning process of students from clinical backgrounds who have an appreciation and understanding of MR.

The study therefore presents a MR-based DFT method that allows for flexibility in the choice of constituent $\omega_n, \Delta\omega$ and $\omega_{\max}$; does not require uniformly spaced time-domain data; and offers a supplementary approach for teaching Fourier Analysis to students from non-traditional mathematics backgrounds.

MR Fourier Analysis Theory

The expansion coefficients, $b_{1\ldots m}$, of a multivariable equation of the form

$$y = b_0 + b_1 x + b_2 x_2 + b_3 x_3 + \ldots b_m x_m, \qquad (3)$$

are commonly found by establishing the matrix equation

$$\overbrace{\begin{pmatrix} 1 & {}^1x_1 & {}^1x_2 & \dots & {}^1x_m \\ 1 & {}^2x_1 & {}^2x_2 & \dots & {}^2x_m \\ 1 & {}^3x_1 & {}^3x_2 & \dots & {}^3x_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & {}^Nx_1 & {}^Nx_2 & \dots & {}^Nx_m \end{pmatrix}}^{\mathbf{X}} \times \overbrace{\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}^{\mathbf{B}} = \overbrace{\begin{pmatrix} {}^1y \\ {}^2y \\ {}^3y \\ \vdots \\ {}^Ny \end{pmatrix}}^{\mathbf{Y}} \tag{4}$$

Note that with MATLAB, **B** is conveniently calculated using the backslash operator (namely $\mathbf{B} = \mathbf{X}\backslash\mathbf{Y}$), which performs a least squares fit rather than a simultaneous solution, and therefore **X** does not have to be square. Without access to the backslash operator, **B** may be calculated by establishing the more conventional MR matrix equation

$$\begin{pmatrix} N & \sum_i {}^ix_1 & \sum_i {}^ix_2 & \dots & \sum_i {}^ix_m \\ \sum_i {}^ix_1 & \sum_i {}^ix_1^2 & \sum_i {}^ix_1^i x_2 & \dots & \sum_i {}^ix_1\,{}^ix_m \\ \sum_i {}^ix_2 & \sum_i {}^ix_2\,{}^ix_1 & \sum_i {}^ix_2^2 & \dots & \sum_i {}^ix_2\,{}^ix_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_i {}^ix_m & \sum_i {}^ix_m\,{}^ix_1 & \sum_i {}^ix_m\,{}^ix_2 & \dots & \sum_i {}^ix_m^2 \end{pmatrix} \times \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} \sum_i {}^iy \\ \sum_i {}^ix_1\,{}^iy \\ \sum_i {}^ix_2\,{}^iy \\ \vdots \\ \sum_i {}^ix_m\,{}^iy \end{pmatrix} \tag{5}$$

where the index i represents the i^{th} set of measurements of $x_{1...m}$ and y.

The discrete Fourier expansion of a waveform in the time-domain takes the standard form

$$f(t) = a_0 + \sum_{n=1}^{N/2} c_n sin(\omega_n t + \theta_n) = a_0 + \sum_{n=1}^{N/2} (a_n cos(\omega_n t) + b_n sin(\omega_n t)), \tag{6}$$

where $c_n = \sqrt{a_n^2 + b_n^2}$ and $\theta_n = tan^{-1}(b_n/a_n)$. . Equation (6) can be considered as a multivariable equation:

$$\overbrace{f}^{y}(t) = \overbrace{a_0}^{b_0} + \overbrace{a_1}^{b_1}\overbrace{cos(\omega_1 t)}^{x_1} + \overbrace{b_1}^{b_2}\overbrace{sin(\omega_1 t)}^{x_2} + \overbrace{a_2}^{b_3}\overbrace{cos(\omega_2 t)}^{x_3} + \overbrace{b_2}^{b_4}\overbrace{sin(\omega_2 t)}^{x_4} + \dots,$$

with the trigonometric terms representing the previous ${}^ix_{1...m}$ variables and $a_{0...N/2}$, $b_{1...N/2}$ representing the expansion coefficients to be determined. The corresponding MR matrix representation of Equation (6) is thus, in the manner of Equation (4), given by

$$\overbrace{\begin{pmatrix} 1 & \dots & cos(\omega_i t_1) & sin(\omega_i t_1) & \dots & sin(\omega_{N/2} t_1) \\ 1 & \dots & cos(\omega_i t_2) & sin(\omega_i t_2) & \dots & sin(\omega_{N/2} t_2) \\ 1 & \dots & cos(\omega_i t_3) & sin(\omega_i t_3) & \dots & sin(\omega_{N/2} t_3) \\ 1 & \dots & cos(\omega_i t_4) & sin(\omega_i t_4) & \dots & sin(\omega_{N/2} t_4) \\ 1 & \dots & cos(\omega_i t_5) & sin(\omega_i t_5) & \dots & sin(\omega_{N/2} t_5) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \dots & cos(\omega_i t_N) & sin(\omega_i t_N) & \dots & sin(\omega_{N/2} t_N) \end{pmatrix}}^{\mathbf{D}} \times \overbrace{\begin{pmatrix} a_0 \\ a_1 \\ b_1 \\ a_2 \\ b_2 \\ \vdots \\ b_{N/2} \end{pmatrix}}^{\mathbf{E}} = \overbrace{\begin{pmatrix} f(t_1) \\ f(t_2) \\ f(t_3) \\ f(t^4) \\ f(t_5) \\ \vdots \\ f(t_N) \end{pmatrix}}^{\mathbf{F}} \tag{7}$$

The MATLAB source code required to establish and solve Equation (7) for a discretely measured waveform, $f(t)$, may be obtained from the authors. Note that if certain harmonics are known to be insignificant, then the number of columns in **D** (and rows in **E**) does not have to extend to $N+1$ as would be representative of a conventional DFT.

Hence, specific harmonics may be omitted from Equation (7) but with all N data points still being used within the MR, since the number of rows in **D** remains unchanged following any such omission. Additionally, $t_{1,\dots,N}$ in Equation (7) can be non-uniformly spaced and so the waveform to be analysed may be sampled at irregular time intervals. The default and minimum $\Delta\omega$ for the MR source code is set to $1/(t_N - t_1)$.

MR Frequency-Domain Spectra

Figure 1 compares the frequency-domain spectra calculated by MR and a standard FFT for MATLAB's time-domain sunspot activity data ($N = 288$), which displays the annual variation in sunspot activity with time since the year 1700.

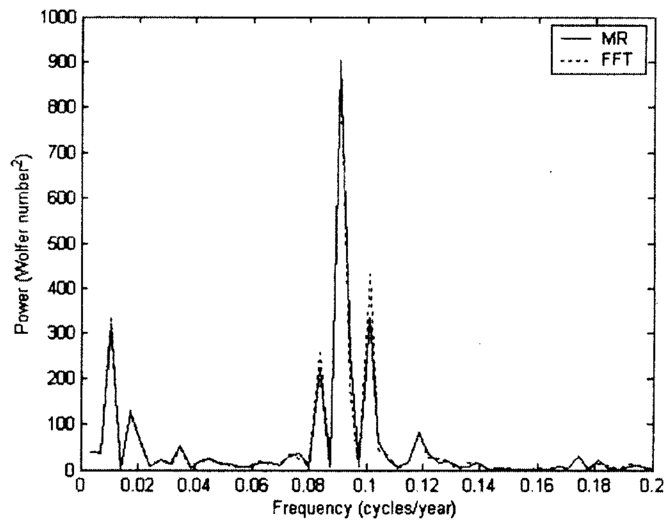


Figure 1 *Comparison of MR and FFT frequency-domain spectra for MATLAB's sunspot activity data.*

A second comparison between MR and a FFT is made for a synthesized waveform ($N = 288$) that consists of three harmonics ($n = 2, 5$ and 7) with respective powers of $8, 50$ and 98units^2 : $f(t) = \sqrt{8}sin(\omega_2 + \pi/4) + \sqrt{50}sin(\omega_5 + \pi/4) + \sqrt{98}sin(\omega_7 + \pi/4)$. Power by standard definition is the value of c_n^2, where c_n is defined by Equation (6).

Table 1 compares theoretical power with power calculated by MR and a FFT for the synthesized waveform for all harmonics from 1 to 10. (Note that since the scale of MR Fourier spectra is such that at any discrete point in time the sum of all frequency components add to the value of $f(t)$, which differs from the FFT scale, the FFT powers in Figure 1 and Table 1 are rescaled to allow appropriate comparison.)

Table 1

Harmonic Number	Power ($units^2$)		
	MR	FFT	Theoretical
1	9.51×10^{-20}	8.68×10^{-3}	-
2	8.00	8.04	8.00
3	4.53×10^{-21}	1.72×10^{-2}	-
4	3.99×10^{-20}	3.99×10^{-2}	-
5	50.0	51.7	50.0
6	3.36×10^{-20}	3.47×10^{-2}	-
7	98.0	98.0	98.0
8	1.21×10^{-21}	7.33×10^{-2}	-
9	2.73×10^{-20}	1.94×10^{-2}	-
10	2.17×10^{-21}	8.95×10^{-3}	-

Comparison of the theoretical power with the power calculated by MR and a FFT for the first ten harmonics of the synthesized waveform for $N = 288$.

MR and FFT methods give similar spectra. Table 1 also shows that spectral leakage (discussed in detail by Harris [1]) contributions by non-constituent harmonics are up to 10^{17} times higher for the FFT, though spectral leakage is still small for the FFT.

Figure 2 (a) shows MR frequency-domain spectra for vastus medialis quadriceps electromyogram (EMG) data ($\omega_s = 1000$ Hz, $N = 500$) for $\Delta\omega = 2$ and 3 Hz respectively. The solid curve in Figure 2 (a) is also representative of a FFT to within the accuracy indicated by Figure 1 and Table 1. The dashed curve in Figure 2 (a) is not representative of a FFT since a FFT would use $N = 333$ to achieve $\Delta\omega = 3$ Hz. Hence, Figure 2 (a) demonstrates that for MR, frequency resolution may be decreased without changing ω_s or N which is not the case for a FFT.

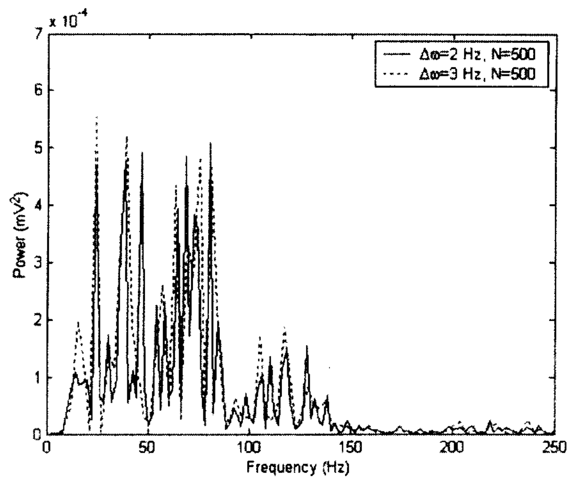


Figure 2 (a) *MR frequency-domain spectra for vastus medialis quadriceps EMG data for $\omega_s = 1000$ Hz, $N = 500$ and $\Delta\omega = 2$ and 3 Hz. The solid curve is also representative of a spectrum that would be obtained with a FFT.*

Figure 2 (b) shows, by comparing MR spectra for $\Delta\omega = 4$ Hz calculated for $N = 250$ and 500, the effect of not using all available data to achieve a desired $\Delta\omega$. For $N = 250$, information is lost and the resulting spectrum is less accurate as shown. The dashed curve in Figure 2 (b) is representative of a FFT to within previously described accuracy.

A clinical example of where it is sometimes useful to decrease frequency resolution is seen in bispectral analysis, which is used to detect the degree of quadratic phase coupling between the harmonics of an EEG biosignal [2]. Decreasing resolution can make the detection of quadratic phase coupling between bands of frequencies more easy to detect.

Note that for MR, reducing $\omega_{\max}$ below the conventional value of $\omega_s/2$ and retaining all N will not improve accuracy but will reduce computation time.

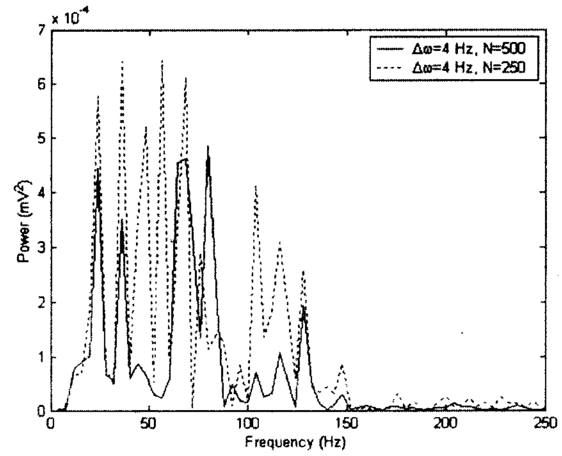


Figure 2 (b). *MR frequency-domain spectra for vastus medialis quadriceps EMG data for* $\omega_s = 1000$ *Hz,* $\Delta\omega = 4$ *Hz and* $N = 250$ *and* 500. *The dashed curve is also representative of a spectrum that would be obtained with a FFT.*

MR Advantages and Conclusion

Fourier Analysis through large-scale MR allows for flexibility in the selection of constituent $\omega_n, \Delta\omega$ and $\omega_{\max}$ and, if the number of data points exceed the minimum number required, then the additional data points will in some circumstances add to the accuracy of the MR. For a standard FFT, constituent ω_n, $\Delta\omega$ and $\omega_{\max}$ are more rigidly set which can result in unnecessary calculations or limit the number of data points used. Also, the MR method does not require uniformly spaced time-domain data.

The differences between MR and FFT spectra for the same $\Delta\omega$ and $\omega_{\max}$ are minimal. However, FFT spectra display up to 10^{17} times higher spectral leakage. As expected, the FFT which requires $N log_2 N$ calculations is computationally more efficient than MR Fourier Analysis which requires N^2 calculations. The difference in computation time does not translate to a significant disadvantage for many applications.

The use of large-scale MR offers a supplementary approach to teaching Fourier Analysis and this approach has been particularly useful for teaching students from non-traditional mathematics backgrounds who are at ease with multivariable equations and MR.

Acknowledgements

The Authors wish to acknowledge the use of MATLAB's sunspot activity data.

Thanks to Georgina Bignell, Bridget Jenner, Ryan McDonald, Jim Tran, Ben Weeks an Rebecca Wunsch for their valuable contributions as teaching team members within the subje Mathematics for Clinical Sciences at Griffith University Gold Coast Campus.

References

[1] Harris F.J., On the use of windows for harmonic analysis with the discrete Fourier transforn *Proc. IEEE*, **66** (1978), 51-83.

[2] Ning T. and Bronzino J.D., Autoregressive and bispectral analysis techniques: EEG appl cations, *IEEE Engineering in Medicine and Biology*, March (1990), 47-50.

MathMedia

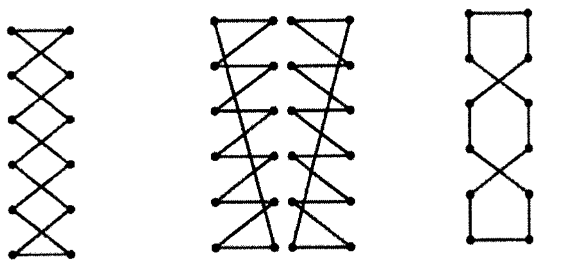
Deborah Smith, Science Writer, "It's a tie: two best shoelace knots named in race of 4 million *The Sydney Morning Herald*, Thursday December 5, 2002.

This item reports on the publication in *Nature* of the work of Monash University mathematici Dr Burkard Polster on the strongest and most efficient way of tying shoe laces into knots. Tl two traditionally favoured methods, viz. 'straight' lacing and 'criss-cross' lacing result in tl strongest fit. The 'bow-tie' method, which is rarely used, is the most efficient, in the sense th for a reasonably effective job, it uses the least amount of lace. The relative efficiency of granı knots and reef knots in completing the task is also investigated.

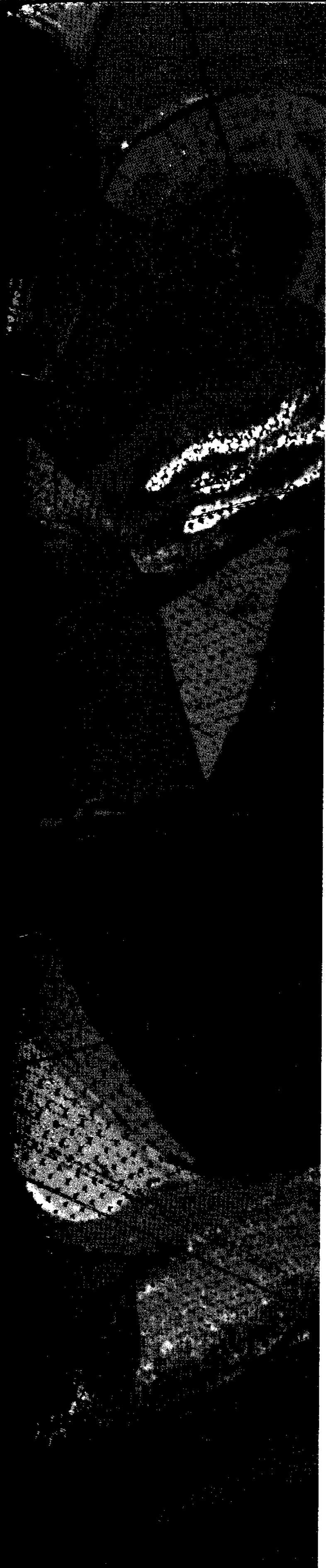
See also the Monash web-site:

http://www-pso.adm.monash.edu.au/news/ and follow the prompts. [Accessed on 10 Decen ber, 2002]. This is also the source of the diagrams reproduced below.

Noticed by a number of members, including David Pask and Kim Burgess, Admin. Assista AustMS.



1. Criss Cross *2. Straight* *3. Bow-tie*



THE AUSTRALIAN MATHEMATICAL SOCIETY

GAZETTE

Volume 30 Number 1 March 2003

ISSN 0311-0729
Registered by Australia Post
Publication No. QBQ3508