Integral Least Squares and Orthogonal Functions

CFS-245

June 26, 1997

David Dunthorn
www.c-f-systems.com

Note

This document predates Precise Signal Component and was not given a CFS number at the time. In reviewing material related to Precise Signal Component I find that it gives a very good development of the method of complex least squares, which is used in Precise Signal Component and rarely (if ever) found in the general literature. It also gets usefully into some areas of orthogonal and nearly orthogonal functions as I was thinking of them at that time. This line of thought led directly to Precise Signal Component. There is a brief excursion into functions of the form exp($-i\, f_0\, \beta^k\, t$) which I was exploring at the time and which is not relevant to Precise Signal Component, but I have decided to leave the document as it was and assign it a current CFS number.

This document is viewable only. I believe this will be adequate for people who do not intend to study it. Please contact me through our web site if you need a printable version. I am aware that the no-print can be defeated, but again I ask that you contact me instead. I really need to know if and how people are finding these documents useful, and this seems one of the few ways I have to encourage feedback.

Integral Least Squares and Orthogonal Functions

We wish to approximate the complex function $y(t)$ over $t_0 \leq t \leq t_f$ using a linear sum of other complex functions, $x_j(t),\ j = 1...n,$ so that $y = \sum\limits_{j=0}^{n} a_j x_j$. To accomplish this, we require that

$$\int_{t_0}^{t_f} w\left(y - \sum_{j=0}^{n} a_j x_j\right)\overline{\left(y - \sum_{j=0}^{n} a_j x_j\right)}\,dt$$

be a minimum, where w is a weighting factor, applied so that the relative importance of deviations at different t can be taken into account. The overline indicates the complex conjugate. With complex variables, the "least squares" concept must become "least magnitude," and the magnitude of x is $x\overline{x}$. This becomes

$$\int_{t_0}^{t_f} w\left(y - \sum_{j=0}^{n} a_j x_j\right)\left(\overline{y} - \sum_{j=0}^{n} \overline{a_j}\ \overline{x_j}\right)dt \qquad (1)$$

Although it is not necessary to do so, it may be a little less confusing at this point to write out the coefficient $a_j = a_{R_j} + i a_{I_j}$, where $i = \sqrt{-1}$ and a_{R_j}, a_{I_j} are the components of a_j:

$$\int_{t_0}^{t_f} w\left(y - \sum_{j=0}^{n} a_{R_j} x_j - i\sum_{j=0}^{n} a_{I_j} x_j\right)\left(\overline{y} - \sum_{j=0}^{n} a_{R_j}\ \overline{x_j} + i\sum_{j=0}^{n} a_{I_j}\ \overline{x_j}\right)dt$$

The object is to minimize this integral magnitude with respect to the choice of the constant multipliers $a_j = a_{R_j} + i a_{I_j}$. To do this we set $\frac{\partial(\text{Eqn 1})}{\partial a_{R_\mathrm{k}}} = \frac{\partial(\text{Eqn 1})}{\partial a_{I_\mathrm{k}}} = 0$ and evaluate for a_k.

$$\frac{\partial(\text{Eqn 1})}{\partial a_{R_\mathrm{k}}} = \int_{t_0}^{t_f} w\Bigg\{ - \left(y - \sum_{j=0}^{n} a_{R_j} x_j - i\sum_{j=0}^{n} a_{I_j} x_j\right)\overline{x_k}$$

$$- \left(\overline{y} - \sum_{j=0}^{n} a_{R_j}\ \overline{x_j} + i\sum_{j=0}^{n} a_{I_j}\ \overline{x_j}\right)x_k\Bigg\}dt = 0$$

now, expand the expression:

$$\int_{t_0}^{t_f} w\Bigg\{ -y\overline{x_k} + \sum_{j=0}^{n} a_{Rj} x_j\overline{x_k} + i\sum_{j=0}^{n} a_{Ij} x_j\overline{x_k}$$

$$-\overline{y}x_k + \sum_{j=0}^{n} a_{Rj}\ \overline{x_j}x_k - i\sum_{j=0}^{n} a_{Ij}\ \overline{x_j}x_k \Bigg\} dt = 0$$

$$\int_{t_0}^{t_f} w\Bigg\{ -(y\overline{x_k} + \overline{y}x_k) + \sum_{j=0}^{n} a_{Rj}(x_j\overline{x_k} + \overline{x_j}x_k)$$

$$+\, i\sum_{j=0}^{n} a_{Ij}(x_j\overline{x_k} -\ \overline{x_j}x_k) \Bigg\} dt = 0$$

Several identities are useful at this point: $\overline{a}b + a\overline{b} = 2Re(\ \overline{a}b) = 2Re(a\overline{b})$; $\overline{a}b - a\overline{b} = 2iIm(\ \overline{a}b) = -2iIm(a\overline{b})$; $a\overline{b} - \overline{a}b = -2iIm(\ \overline{a}b) = 2iIm(a\overline{b})$. Thus:

$$\int_{t_0}^{t_f} w\Bigg\{ -2Re(y\overline{x_k}) + 2\sum_{j=0}^{n} a_{Rj}Re(x_j\overline{x_k})$$

$$+\, 2i^2\sum_{j=0}^{n} a_{Ij}Im(x_j\overline{x_k}) \Bigg\} dt = 0$$

$$\int_{t_0}^{t_f} w\Bigg(-2Re(y\overline{x_k}) + 2\sum_{j=0}^{n} a_{Rj}Re(x_j\overline{x_k}) - 2\sum_{j=0}^{n} a_{Ij}Im(x_j\overline{x_k}) \Bigg) dt = 0 \tag{2}$$

Note that all the terms in this equation are real (not complex). Now the same process is applied, setting the $a_{I\mathrm{k}}$ derivative equal to zero:

$$\frac{\partial(\text{Eqn 1})}{\partial a_{I\mathrm{k}}} = \int_{t_0}^{t_f} w\Bigg\{ +\, i\Bigg(y - \sum_{j=0}^{n} a_{Rj}x_j - i\sum_{j=0}^{n} a_{Ij}x_j \Bigg)\overline{x_k}$$

$$-\, i\Bigg(\overline{y} - \sum_{j=0}^{n} a_{Rj}\ \overline{x_j} + i\sum_{j=0}^{n} a_{Ij}\ \overline{x_j} \Bigg)x_k \Bigg\} dt = 0$$

$$\begin{aligned}\int_{t_0}^{t_f} wi\Bigg\{ &+ y\overline{x_k} - \sum_{j=0}^{n} a_{Rj}x_j\overline{x_k} - i\sum_{j=0}^{n} a_{Ij}x_j\overline{x_k} \\ &\qquad - \overline{y}x_k + \sum_{j=0}^{n} a_{Rj}\ \overline{x_j}x_k - i\sum_{j=0}^{n} a_{Ij}\ \overline{x_j}x_k \Bigg\}dt = 0\end{aligned}$$

$$\begin{aligned}\int_{t_0}^{t_f} wi\Bigg\{ &+ (y\overline{x_k} - \overline{y}x_k) - \sum_{j=0}^{n} a_{Rj}(x_j\overline{x_k} - \ \overline{x_j}x_k) \\ &\qquad - i\sum_{j=0}^{n} a_{Ij}(x_j\overline{x_k} + \ \overline{x_j}x_k) \Bigg\}dt = 0\end{aligned}$$

$$\begin{aligned}\int_{t_0}^{t_f} wi\Bigg\{ &+ 2iIm(y\overline{x_k}) - 2i\sum_{j=0}^{n} a_{Rj}Im(x_j\overline{x_k}) \\ &\qquad - 2i\sum_{j=0}^{n} a_{Ij}Re(x_j\overline{x_k}) \Bigg\}dt = 0\end{aligned}$$

$$\int_{t_0}^{t_f} w\Bigg(- 2Im(y\overline{x_k}) + 2\sum_{j=0}^{n} a_{Rj}Im(x_j\overline{x_k}) + 2\sum_{j=0}^{n} a_{Ij}Re(x_j\overline{x_k})\Bigg)dt = 0 \tag{3}$$

Again, this equation is all in real terms (not complex). Thus Equations 2 and 3 may be combined into one complex equation by multiplying Equation 3 by i and adding the two together. Note that $a_jx_j\overline{x_k} = (a_{Rj} + a_{Ij}i)(Re(x_j\overline{x_k}) + Im(x_j\overline{x_k})i)$, which is exactly the form of the four terms in equations 2 and 3. Thus we have:

$$2\int_{t_0}^{t_f} w\Bigg(- y\overline{x_k} + \sum_{j=0}^{n} a_jx_j\overline{x_k}\Bigg)dt = 0$$

or

$$\sum_{j=0}^{n} \int_{t_0}^{t_f} w a_j x_j \overline{x_k} dt = \int_{t_0}^{t_f} w y \overline{x_k} dt \quad \text{j=0,...,n} \tag{4}$$

$$\sum_{j=0}^{n} a_j \int_{t_0}^{t_f} w x_j \overline{x_k} dt = \int_{t_0}^{t_f} w y \overline{x_k} dt \quad \text{j=0,...,n} \tag{5}$$

There are $n + 1$ equations of this form. They can be expressed in matrix form as:

$$\mathbb{X}\boldsymbol{a} = \boldsymbol{q}$$

Where $X_{j,k} = \int_{t_0}^{t_f} w x_j \overline{x_k} dt,\ a_k = a_k,\ q_k = \int_{t_0}^{t_f} w y \overline{x_k} dt.$ The a_k's can be determined by inverting the $\mathbb{X}$ matrix:

$$\boldsymbol{a} = \mathbb{X}^{-1}\boldsymbol{q}$$

The derivation of this general form for linear least squares is very similar to the derivation of the coefficients for a series developed from orthogonal functions, such as the Fourier series. The primary difference is that for a series of orthogonal functions, $\mathbb{X}$ is always a diagonal matrix (for orthonormal functions it is the identity matrix). The off-diagonal terms of $\mathbb{X}$ are interdependencies of the x functions. Since by definition the members of a set of orthogonal functions are linearly independent, there are no interdependencies and all off-diagonal terms are zero. Thus for orthogonal functions each a is given directly by the corresponding q integral and any a (i. e. frequency component) can be calculated without having to calculate the others. This feature; the fact that no $\mathbb{X}$ matrix need be calculated or inverted and further, that any one a can be calculated independently, have given orthogonal functions a central role in practical applications, particularly the Fourier series in frequency spectrum analysis.

For the orthogonal functions, the $\boldsymbol{q}$ vector easily can be seen to be a set of resonators. Each sum effectively beats the object function $y(t)$ against one of the $x_k(t)$'s to determine how well they resonate with one another. With the Fourier series, the collection of various resonances become the frequency spectrum of the object signal and it is easy to see that the amplitude of the spectral response at each frequency is precisely the response of the resonator for that frequency.

Note that the $\boldsymbol{q}$ vector for the general form of linear least squares is precisely the same as for a set of orthogonal functions. Again each sum effectively beats the object function $y(t)$ against one of the x$_k(t)$'s to determine how well they resonate with one another. Again each q_k is a resonator for the function for which the corresponding a_k is the amplitude of that "spectral component." The difference is that for general least squares $\mathbb{X}$ is not the identity matrix and thus each amplitude a_k will not equate exactly with the corresponding resonator q_k, but will have various corrections based upon the responses of other resonators. However, the device still produces a spectrum, and the matrix

$\mathbb{X}^{-1}$adjusts the raw resonator data to produce spectral components which also reflect the interactions.

In the case of interest, we will be using functions $e^{-if_0\beta^k t}$ which are very similar to the Fourier components e^{-ijt}. The difference is that products of pairs of the Fourier functions integrate conveniently to zero over the convenient interval $[-\pi, \pi]$ while the β^k form does not have this property. However, the β^k form does generally integrate to produce small off-diagonal elements and furthermore, the disparity between the sizes of the diagonal and off-diagonal elements tends to grow larger as the sample time increases so that the $\mathbb{X}$ matrix tends to approach the identity matrix . Unlike the orthogonal form, the coefficients associated with each frequency can be estimated even for short sample times, with the accuracy and/or confidence increasing as the sample time increases. A mechanism similar to this may explain how human hearing apparently can distinguish frequencies with shorter sample times than Fourier analysis requires. The concept that Fourier series analysis can exactly determine the component of frequency f in an arbitrary signal is not true in general, and will work only if the spectrum is defined to consist of just frequencies related as $1, 2, ..., n$.

It is also apparent that the "magic" of Fourier series analysis is a direct result of the choice of a set of spectral frequencies which produce an orthogonal set of functions. That set of functions produces a spectrum which is additively spaced in frequency: $f_j = j$. Human hearing responds to frequency differences which are multiplicatively spaced in frequency: $f_k = f_0\,\beta^k$. Thus Fourier series analysis, no matter how attractive its properties, simply does not produce the correct response. It is clear that this is indeed the case when one realizes that many of the functions $e^{-if_0\beta^k t}$ can be made to equal a function e^{-ijt} with the same frequency and identical sample time. Thus the integral for that resonator will be identical for the Fourier analysis and the present analysis. Fourier analysis will give a precise amplitude of response for those frequency components, but that same amplitude will have to be adjusted to be correct for the multiplicative spectrum.

Consider the least squares relationship that was developed earlier:

$$\boldsymbol{a} = \mathbb{X}^{-1}\boldsymbol{q}$$

The expression for a_k can be written out as a sum:

$$a_k = \sum_{j=1}^{n} \overline{x}_{k,j}\, q_j$$

where $\overline{x}_{k,j}$ indicates the k, j^{th} element of the inverse matrix, $\mathbb{X}^{-1}$. Substituting for q_j

$$a_k = \sum_{j=1}^{n} \overline{x}_{k,j} \int_{t_0}^{t_f} wy\overline{x_j}dt$$

This can be rearranged:

$$a_k = \int_{t_0}^{t_f} wy \sum_{j=1}^{n} \overline{x}^{\urcorner}_{k,j}\, \overline{x_j} dt$$

At this point define $s_k = \sum_{j=1}^{n} \overline{x}^{\urcorner}_{k,j}\, \overline{x_j}$. Then

$$a_k = \int_{t_0}^{t_f} wy \mathrm{s}_k dt$$

Note that this summation has exactly the same form as $q_k = \int_{t_0}^{t_f} wy \overline{x_\mathrm{k}} dt$, but with s$_k$ in place of $\overline{x_\mathrm{k}}$. In the case of orthogonal functions, $a_k = q_k = \int_{t_0}^{t_f} wy \overline{x_\mathrm{k}} dt$. For general linear least squares, the resonator based on the function $x_k(t)$ is still $q_k = \int_{t_0}^{t_f} wy \overline{x_\mathrm{k}} dt$, but the amplitude response for $x_k(t)$ is $a_k = \int_{t_0}^{t_f} wy \mathrm{s}_k dt$, which has the form of a resonator for a different function $s_k(t)$. Thus we can call $s_k(t)$ the surrogate for $x_k(t)$. Once a sampling interval $t_0 \leq t \leq t_f$ is chosen, the set of surrogate functions $s_k(t)$ can be computed directly from the set of functions $x_k(t)$, independent of $y(t)$, and the individual spectral components, a_k can be independently computed from a single weighted sum of the data, just as is the case for orthogonal functions.

As well as the similarities, there are some important differences between general linear least squares and orthogonal function series. One of the reasons for the widespread use of Fourier series analysis is the availability of the Fast Fourier Transform (FFT), by which mechanism an entire spectrum of Fourier coefficients can be calculated in a number of operations proportional to $n\, ln\, n$ rather than proportional to n^2 as direct evaluation of the spectral sums would require. Such a mechanism is not available for general linear least squares. A second very important difference is that orthogonal function series and the Fourier series and the FFT in particular, are almost always computed from a number of sample points exactly matching the number of a_k coefficients. The only requirement for general linear least squares is that the number of sample points be greater than or equal to the number of coefficients or, since there are $n + 1$ coefficients, $m > n$. Since the use of surrogate functions allows each spectral amplitude to be expressed as an independent sum, in practical analysis of signals there is no reason that m, need be the same for all of the n functions. The same is true of the sampling rate. In multiplicative frequency spectrum analysis it may prove very useful to adjust both the effective sampling rate and the number of samples when treating a broad range of frequencies.