



OPTIMIZACIÓN DEL MÉTODO DE MÁXIMA ENTROPÍA PARA PREDICCIÓN DE SERIES DE TIEMPO

Eduardo Gómez Ramírez, Gabriel Palafox Suárez y Octavio Bautista Zenil*

Laboratorio del Centro de Investigación, *Laboratorio de Cómputo de Ingeniería
Universidad La Salle

RESUMEN

La Transformada Rápida de Fourier (FFT) "tradicionalmente" es un algoritmo que se ha aplicado en una gran cantidad de áreas y del cual se puede contar con bastante literatura en cuanto a algoritmos, programas, etc. No existe algún programa o paquete comercial de Procesamiento Digital de Señales (DSP) que no contenga la FFT como subrutina, función o instrucción. Aunque esta transformada es muy utilizada tiene ciertas limitaciones como son: errores de pseudointerferencia, sensibilidad al ruido, etc. Para estos casos, existen otros algoritmos alternativos a Fourier. En este trabajo se analiza el método de Máxima Entropía como herramienta en el análisis de espectro.

ABSTRACT

Traditionally the Fast Fourier Transform is an algorithm that has been applied in many areas and we can found enough literature for algorithms, programs, etc. There is no program or commercial package of Digital Signal Processing (DSP) that doesn't contain the FFT as a subroutine, function or instruction. Although this transform is used in many situations it has certain limitations like: leakage, picket fence, noise sensibility etc. In these cases, other alternative algorithms exists. In the current studies we analyze the Maximum Entropy Method like a tool for the spectra analysis.

INTRODUCCIÓN

Desde que Fourier publicó su trabajo en 1822 en su libro *Teoría Analítica del Calor*, la transformada de Fourier se ha utilizado en una gran cantidad de áreas. A pesar de la gran potencialidad de esta herramienta, el problema se presentaba cuando se requería aplicar a una gran cantidad de datos. Debido a esto, hubo un gran desarrollo en nuevos algoritmos que aplicaran la transformada discreta de Fourier. Uno de los más importantes fue elaborado en 1965, por James W. Cooley y John W. Tukey. El trabajo de ambos dio lugar a un algoritmo conocido como: Transformada Rápida de Fourier (TRF). Aunque esta aportación redujo considerablemente el número de operaciones de N^2 a $N \log_2 N$, la TRF no es el único algoritmo para procesamiento de una serie de tiempo y en varios casos no es el óptimo. Existen otros algoritmos con un gran desarrollo teórico que empiezan desde los trabajos de Heissenberg, Schrödinger, Von Neumann, Wiener, etc. (1,2).

Estos trabajos en la actualidad son poco utilizados, pero en varios casos tienen varias ventajas sobre el algoritmo de Fourier. La TRF tiene algunos inconvenientes tales como errores de pseudointerferencia y su gran sensibilidad a niveles de ruido en la señal de entrada (3-6). En el área de predicción se utilizan una gran cantidad de herramientas (7,8), dependiendo de la naturaleza y nivel de ruido que contengan los datos. Un algoritmo que se ha utilizado con mejores resultados que Fourier para señales ruidosas es el método de Máxima Entropía.

En este trabajo se analiza el método de Máxima Entropía para señales con diferentes niveles de ruido, se demuestra que existe un número mínimo de nodos en función del error de predicción para la estimación del espectro de potencia, y su adaptación al uso de llamadas a procedimientos remotos (RPC, Remote Procedure Call).



MÉTODO DE MÁXIMA ENTROPÍA (MEM)

El método de Máxima Entropía (MEM) originalmente fue desarrollado para el procesamiento de datos Geofísicos (9). Este método fue presentado por primera vez en el trabajo "Maximum Entropy Spectral Analysis" y fue desarrollado por John Burg en 1967.

El método de Máxima Entropía está basado en escoger que espectro corresponde al más aleatorio o a la más impredecible serie de tiempo cuya autocovarianza coincida con el conjunto de valores dado. Este método se conoce también con el nombre de *Modelo de todos polos* o *Modelo Autoregresivo (AR)* (10).

El espectro de Potencia¹ de una función se puede expresar de la siguiente forma (10):

$$P(f) = \frac{a_0}{\left| 1 + \sum_{k=1}^m a_k \cdot z^k \right|^2} \quad \text{Ec. 1}$$

donde: m = es el número de polos

$$z = e^{j2\pi f}$$

$$a_0 \dots a_k \in \mathfrak{R}$$

Para expresar la Ec. 1 de otra forma se utilizará el teorema de Wiener-Khinchin, que puede ser enunciado como:

"La Transformada de Fourier de la autocorrelación de una función es igual al espectro de potencia"

De esta manera el espectro de potencia se puede expresar como:

$$P(f) = \frac{a_0}{\left| 1 + \sum_{k=1}^m a_k \cdot z^k \right|^2} = \sum_{j=-m}^m \Phi_j \cdot z^j \quad \text{Ec. 2}$$

donde: Φ_j es la autocorrelación de la función muestreada.

$$\Phi_j = \Phi_{-j} = \frac{1}{N+1-j} \sum_{i=0}^{N-j} c_i c_{i+j} \quad \text{Ec. 3}$$

¹ Es la magnitud del espectro de Fourier elevado al cuadrado

donde: $j = 0, 1, 2, \dots, N$

c_i representa el i -ésimo dato de $f(x)$

Los coeficientes a_0, a_1, \dots, a_k y la autocorrelación Φ_j , cumplen con la siguiente relación (Matriz simétrica de Toeplitz):

$$\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_m \\ \Phi_2 & \Phi_3 & \dots & \Phi_{m-1} \\ \Phi_3 & \Phi_4 & \dots & \Phi_{m-2} \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_m & \Phi_{m-1} & \dots & \Phi_0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} a_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{Ec. 4}$$

de donde por el método de Burg y Andersen (10) pueden ser obtenidos los valores de los coeficientes.

RELACIÓN ENTRE EL ERROR DE PREDICCIÓN Y EL NÚMERO DE POLOS

Como se puede observar de las ecuaciones anteriores el número de polos define la exactitud de la aproximación del espectro de potencia con respecto al de Fourier. Pero de aquí surgen varias preguntas: ¿Existe un valor mínimo de polos? ¿Depende el número de polos del nivel de ruido? Para contestar estas preguntas se utilizará la siguiente consideración:

Primero supóngase que se tiene el modelo estático de una planta

$$y_t = f(x_t) + \omega_t \quad \text{Ec. 5}$$

donde: $f(x)$ es una función

ω_t es una variable aleatoria

La variable aleatoria ω representa el nivel de ruido en un instante t con la siguiente estadística:

$$E\{\omega_t\} = 0 \quad E\{\omega_t \omega_s\} = \sigma_t^2 \delta_{ts} \quad \text{Ec. 5.1}$$

Para el modelo de esta planta se pueden demostrar (11) los resultados siguientes:

$$D_n^N = E\{S_n^N\} + 2 \frac{N}{n} - 1 \quad \text{Ec. 6}$$



donde el valor

$$D_n^N \triangleq \frac{1}{n} \sum_{t=1}^n E \left\{ \left(f(x_t) - \hat{f}^N(x_t | \hat{c}_n^N) \right)^2 \right\} \quad \text{Ec. 7}$$

caracteriza la precisión cuadrática media de la aproximación del modelo de la planta después de n mediciones del par entrada-salida (x_t, y_t) , $t=1, \dots, n$. A esta función se le denomina *error de predicción o de estimación*.

El modelo estimado de la planta (Ec. 5) es el siguiente:

$$\hat{y}_t^N = \hat{f}^N(x_t | \hat{c}_n^N) = \sum_{i=1}^N c_i^N \phi_i^N(x_t) \quad \text{Ec. 8}$$

donde: $\phi_i^N(x_t)$ representa una función determinada y c_i^N el peso de esta función en la salida del modelo, N es el orden del modelo, es decir, el número de términos del estimado de la función $f(x)$.

donde:

$$\hat{c}_n^N = \hat{c}_n^N \triangleq \arg \min_{c^N} E \{ S_n^N \} \quad \text{Ec. 9}$$

$$S_n^N \triangleq \frac{1}{n} \sum_{t=1}^n \left(y_t - \hat{y}_t^N \right)^2$$

A la Ec. 7 se le conoce como *criterio de Mallows-Akaike* (12,13). Es importante destacar de la Ec. 7 y de la Figura 1 los siguientes puntos:

- La curva "exponencial" corresponde a la precisión de la aproximación $\tilde{S}_{y,n}^N$, es decir sino existiera ruido en la función $f(x)$, a mayor número de términos mejor será la aproximación.

- La curva 2 (el término $2N/n$) corresponde a la influencia de los ruidos para diferentes valores de N (mientras mayor sea el número de entradas que contengan ruido mayor será la distorsión en la medición de la salida de cada nodo).

- El valor óptimo del número de entradas $(N_n^y)^*$ corresponde al valor mínimo de la función $D_{y,n}^N$.

Lo anterior puede ser relacionado con nuestro espectro suponiendo que $f(x)$ es el espectro de la señal sin ruido obtenido por FFT ($P(f)$). Por lo que el espectro con ruido es análogo a la función "y" (Ec. 5), y puede ser representado de la siguiente forma:

$$PR(f) = P(f) + \omega_t, \quad t=1,2,\dots,n \quad \text{Ec. 5'}$$

donde $\omega_t \in R$ es una variable aleatoria que representa la diferencia que existe entre el espectro sin ruido y el espectro de la señal con ruido.

Considerando lo anterior se pueden demostrar los resultados siguientes:

$$D_n^N = E \{ S_n^N \} + 2 \frac{N}{n} - 1 \quad \text{Ec. 6'}$$

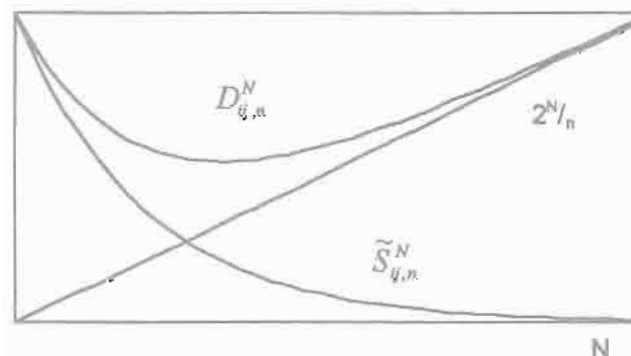


Figura 1. Error de predicción $D_{y,n}^N$.



donde el valor

$$D_n^N \triangleq \frac{1}{n} \sum_{i=1}^n E \left\{ \left(P(x_i) - \hat{P}^N(x_i | \hat{c}_n^N) \right)^2 \right\} \quad \text{Ec. 7'}$$

caracteriza la precisión cuadrática media de la aproximación del espectro. Esta ecuación representa el *error de estimación de nuestro espectro*.

El espectro estimado:

$$\hat{P}(f) = \sum_{j=-m}^m \Phi_j * Z^j \quad \text{Ec. 8'}$$

y:

$$\hat{c}_n^N = c_n^N \triangleq \underset{c_n^N}{\text{argmin}} E \{ S_n^N \} \quad \text{Ec. 9'}$$

$$S_n^N \triangleq \frac{1}{n} \sum_{f=1}^n (PR(f) - \hat{P}(f))^2$$

Para poder analizar el comportamiento del valor esperado de $D_{ij,n}^N$ se requiere de una gran cantidad de operaciones, por esta razón para esta parte del proyecto se realizó un programa en lenguaje C. De esta forma, se procesó y demostró el algoritmo en una *workstation* utilizando el procedimiento que se describe brevemente en la siguiente sección.

ADECUACIÓN DEL PROGRAMA AL USO DE LLAMADAS A PROCEDIMIENTOS REMOTOS

El programa de MEM fue separado en dos partes siguiendo el paradigma Cliente/Servidor. La parte del "cliente" es procesado en una *workstation* SUN SPARC/classic del Laboratorio de Cómputo de Ingeniería en la Universidad La Salle y lo que corresponde al "servidor" se procesó en una *workstation* Silicon Graphics Indigo con 2 procesadores del Laboratorio de Visualización en la UNAM.

La comunicación entre ambos procesos se llevó a cabo mediante el uso del mecanismo de llamadas a procedimientos remotos, el cual establece los protocolos e interfases necesarias para el desarrollo de aplicaciones distribuidas. En este caso se usó el ambiente de cómputo

distribuido ONC+, desarrollado por SUN Microsystems, en especial, su compilador de protocolos RPCGEN. Así, el cliente es la interfase para el usuario recibiendo los parámetros de entrada del usuario, de manera transparente a éste se comunica con el servidor en la UNAM, le manda su petición, el servidor la procesa y le entrega los resultados al cliente. La comunicación entre los dos sistemas de cómputo se realizó a través de Internet utilizando un enlace RDI de 64 Kb/seg.

El servidor es una aplicación de cómputo intensivo que puede realizar una gran cantidad de iteraciones; para 1,000 iteraciones ocupó casi 30 horas de tiempo efectivo de procesador. Este mismo procedimiento en una computadora 486/66 por lo menos tardaría 4 días.

El objetivo final es correr el servidor en la Supercomputadora Cray Y-MP 4/464 de la Dirección General de Servicios de Cómputo Académico en la UNAM, utilizando sus capacidades vectoriales y de paralelismo una vez que se autorice la apertura de un puerto RPC y que se implanten los mecanismos de seguridad necesarios para el acceso restringido a sólo un cliente.

EJEMPLO

Para ejemplificar el comportamiento de la Ec. 7, se utilizó la función representada en la gráfica de la Figura 2.

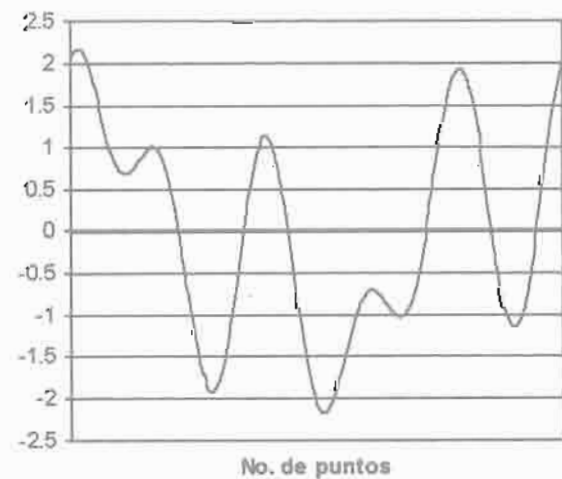


Figura 2. Función $f(x)$.

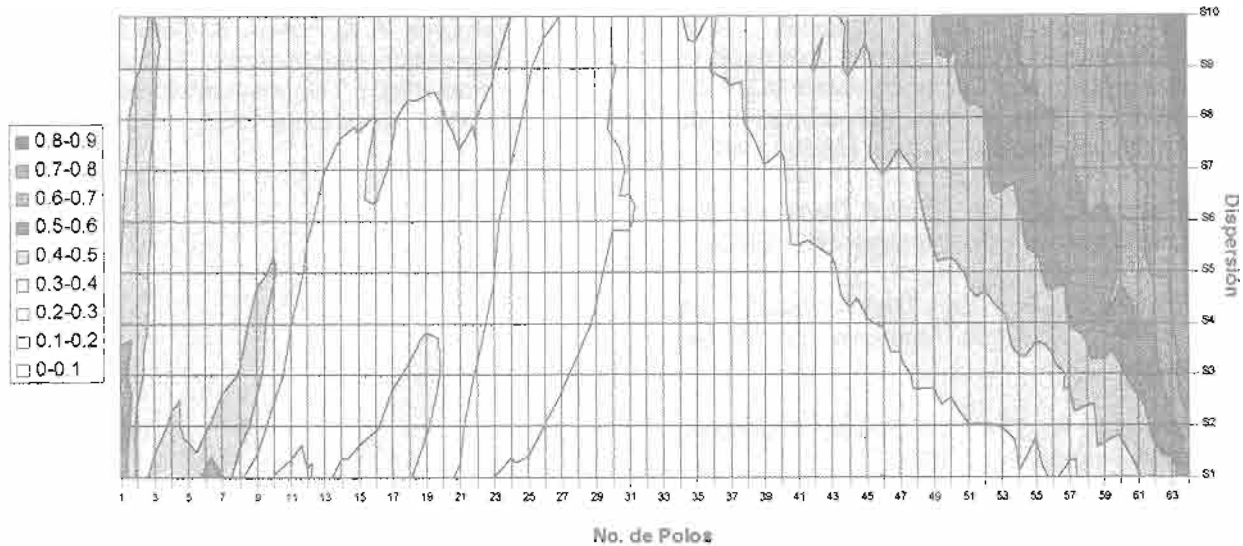


Figura 3. Error de Predicción $D_{y,n}^N$.

Primero se obtuvo el espectro de potencia por medio de FFT para esta señal y se comparó con el espectro obtenido por MEM de la función con distinto número de polos y diferente dispersión de ruido. La distribución del ruido fue normal y con media cero y la dispersión se varió en el rango de 0 a 1.

En la Figura 3, se muestra una gráfica que representa el error de predicción $D_{y,n}^N$ normalizado del espectro de potencia, considerando el número de polos contra la dispersión del ruido (eje s). Cada división del eje s corresponde a un incremento de 0.05 en la dispersión. Como se puede observar, existe un valle, representado como la zona más clara en la gráfica, que corresponde al número de polos mínimo para cada espectro. Este número de polos depende del número de puntos analizado, el nivel de ruido y la naturaleza de la función. Con este número "óptimo" se puede obtener el espectro de potencia y analizar cuáles son las tendencias de la señal y predecir su comportamiento futuro.

CONCLUSIONES Y COMENTARIOS

Como aproximación, comúnmente se hace que el número de polos corresponda a la mitad del número de puntos, pero este valor no es el más óptimo. En este trabajo se demostró que existe un número mínimo de polos para la obtención

del espectro por medio de MEM. Esta herramienta ofrece grandes ventajas para la recuperación de señales con altos niveles de ruido o que se encuentren "incompletas", como pueden ser los datos registrados de un experimento o por algún sistema de comunicaciones. El uso de RPC fue un alternativa muy importante para reducir el tiempo de cálculo que se requería para la demostración de la teoría de este trabajo.

AGRADECIMIENTOS

Se agradece la colaboración para este proyecto de manera especial al Ing. Marta Sánchez, Jefe de Supercómputo de la UNAM, al Dr. Enrique Daltabuit, Director de Cómputo para la investigación, al personal de visualización de la UNAM y al Laboratorio de Cómputo de Ingeniería de la Universidad La Salle.

REFERENCIAS

1. Robinson, E. A. "A Historical Perspective of Spectrum Estimation". *Proceedings of the IEEE*. Vol. 70 No. 90. pp. 885-907. September 1982.
2. Raggi, González M. G. "Fundamentos e implementación de la Transformada Discreta del Coseno para el análisis y



- Codificación de Señales en Comunicación". *III Congreso Internacional de Electrónica y Comunicaciones UDLA'92*. del 17 al 20 de febrero. Universidad de las Américas, Cholula, Puebla, México 1992.
3. Bracewell, R. *The Fourier Transform and its aplicaciones*, USA, McGraw-Hill, 1965.
 4. Bracewell, R. "La Transformación de Fourier", *Scientific American*, pp. 56-64, Agosto 1989.
 5. Brigham, E. *The Fast Fourier Transform*, USA, Prentice Hall, 1974.
 6. Gómez E. "Las Transformadas de Fourier, Hartley, Walsh, Haar y Máxima Entropía: comparación, uso e implementación en Comunicaciones y Electrónica". *Tercer Congreso Internacional de Electrónica y Comunicaciones*, Universidad de las Américas, Cholula, Puebla, México, 17-20 febrero 1992.
 7. Box G. & Jenkins G. *Time Series Analysis Forecasting and Control*. Holden-Day Series in Time Series Analysis. 1970.
 8. Kimoto T. & Asakawa K. Stock Market Prediction System with Modular Neural Networks. *International Conference of Neural Networks*. 1990.
 9. Kay, Steven M.; Marple, Stanley L. Jr. "Spectrum Analysis - A Modern Perspective". *Proceedings of the IEEE*. Vol. 69, No. 11, pp. 1380-1419. November 1981.
 10. Press, William H., Brian, Flannery P., Teukolsky, Saul A., Vetterling, William T. *Numerical Recipes*, USA, Cambridge University Press, 1986.
 11. Gómez E. & Poznyak A. "How to Select a Numbers of Nodes in Artificial Neural Networks". *CAM 94*. Cancún México. Septiembre 26-30 1994.
 12. Akaike H. Stochastic Theory of Minimal Realization. *IEEE Transactions on Automatic Control*. Vol. ac-19, No. 6. Diciembre 1974.
 13. Verulava Y. & Polyak B. *Selection of the Regression Model Order*. Translated from *Avtomatika i Telemekhanika*. No. 11, pp. 113-129. Noviembre 1988.