

# Análisis de Formas de Onda de Plasmas con Wavelets y Máquinas de Vectores Soporte

S. Dormido\*, J.M. de la Cruz\*\*, J. Vega\*\*\*, M. Santos\*\*, S. Dormido-Canto\*, J. Sánchez\*, R. Dormido-Canto\*, Gonzalo Farias\*\*

\* Dpto. Informática y Automática, UNED,  
28040 Madrid (SPAIN)

\*\* Dpto. Arquitectura de Computadores y Automática, Universidad Complutense,  
28040 Madrid (SPAIN)

\*\*\*Asociación EURATOM/CIEMAT para Fusión  
Avda. Complutense, 22, 28040 Madrid (SPAIN)

## RESUMEN

Cada experimento en fusión termonuclear genera cientos de señales. Para el análisis de estas señales es importante tener mecanismos automáticos de clasificación y recuperación de señales. En este trabajo se muestra como las técnicas de transformada de wavelets para la caracterización, la eliminación de ruido y la compresión de la información, y las máquinas de los vectores soporte para el reconocimiento de patrones, y la clasificación de la información, pueden ser utilizados para el fin señalado.

**Palabras claves:** Reconocimiento de patrones. Wavelets. Máquina de Vectores Soporte (MSV). Fusión termonuclear.

## 1. INTRODUCCIÓN

Los experimentos actuales realizados con plasma producen cientos de señales dinámicas en cada descarga. Estas señales provienen de muy diversos sensores y han sido obtenidas en distintas condiciones, por lo que habitualmente no presentan, con un análisis inmediato, una forma común [1]. Resulta de gran utilidad desarrollar métodos o aproximaciones que conduzcan, de forma rápida, a la recuperación de señales de interés, definidas por ciertas características, o que faciliten y aceleren el proceso de análisis de las mismas. En definitiva, se trata de encontrar patrones que permitan identificarlas y clasificarlas, y diseñar mecanismos computacionales automáticos de búsqueda.

En [2] se propone un método para encontrar secuencias temporales similares utilizando la Transformada Discreta de Fourier (TDF). En [3] los coeficientes de la serie de Fourier se utilizan como referencia para buscar fenómenos similares en grandes bases de datos. El método se ha utilizado en señales que varían lentamente. Sin embargo, la TDF no da buenos resultados cuando se utiliza con señales con variaciones rápidas.

En el análisis de Fourier la información temporal se pierde en la transformación al dominio de las frecuencias y no permite detectar características transitorias o no estacionarias. La transformada de Fourier de tiempo corto (TFTC) permite obtener características no estacionarias utilizando una ventana temporal para el análisis. Sin embargo, la precisión está determinada por la ventana usada en el análisis y es la misma para todas las frecuencias.

La transformada de wavelets (TW) es una alternativa a la TFTC. En esta transformada se usan ventanas pequeñas para las altas frecuencias y grandes para bajas frecuencias. Permite de

este modo analizar la señal en distintas escalas o niveles de resolución. Permite también poner de manifiesto la existencia de discontinuidades cuando se desplaza hacia escalas menores. Entre sus aplicaciones se encuentran la aproximación y compresión de señales, la reducción y eliminación de ruido, la representación de señales en múltiples escalas y la localización de patrones [4-5].

Las MVS y, de forma más general, los algoritmos de aprendizaje basados en núcleos, son un método muy efectivo para el reconocimiento de patrones [7-11].

En la sección siguiente 2 presentamos los aspectos de la transformada de wavelets que la hacen adecuada para ser usada en el análisis de señales de fusión termonuclear. Posteriormente presentamos el método MSV y la forma en que éste se implementa, y finalizamos en la sección 4 mostrando la aplicación de la TW y los MVS en una aplicación en la clasificación de distintas señales procedentes de experimentos con fusión termonuclear.

## 2. WAVELETS

Las wavelets son funciones matemáticas de cuadrado integrable, bien localizadas en los dominios temporal y de frecuencias con ciertas propiedades matemáticas; por ejemplo, la señal  $\psi(t)$  con

$$\int_{-\infty}^{\infty} |\psi(t)|^2 < \infty$$

y tal que

$$|\psi(t)| \leq c(1+|t|)^{-1-\varepsilon}, \quad |\Psi(w)| \leq c(1+|w|)^{-1-\varepsilon}, \quad \varepsilon > 0$$

donde  $\Psi(w)$  representa la transformada de Fourier de  $\psi(t)$ . A partir de esta señal, denominada *wavelet madre*, se forma una familia de funciones mediante las relaciones

$$\psi_{s,d}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-d}{s}\right), \quad s > 0, d \in R \quad (1)$$

Los parámetros  $s$  y  $d$  se denominan factor de escala y desplazamiento, respectivamente. La transformada continua de wavelet de una señal  $f(t)$  de cuadrado integrable,  $f(t) \in L_2(R)$ , se define como el producto escalar

$$T_w f(s,d) = \int_{-\infty}^{\infty} f(t) \psi^*\left(\frac{t-d}{s}\right) \quad (2)$$

donde el asterisco indica conjugado. En las aplicaciones prácticas el cálculo de la transformada se hará sobre una cierta rejilla de valores de  $s$  y  $d$ . La más utilizada es aquella en la que  $s$  es un múltiplo entero de 2 y  $d$  un desplazamiento entero del factor de escala

$$s=2^j, \quad d=k2^j, \quad j,k \in \mathbb{Z}$$

Tenemos así la familia de funciones

$$\left\{ \psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-k2^j}{2^j}\right) = 2^{-j/2} \psi(2^{-j}t-k) \right\}_{(j,k) \in \mathbb{Z}^2} \quad (3)$$

La transformada discreta de wavelet es ahora:

$$T_w f(j,k) = \int_{-\infty}^{\infty} f(t) 2^{-j/2} \psi^*(2^{-j}t-k) dt$$

Normalmente, la familia de wavelets se elige de modo que forman una base ortonormal completa del espacio  $L_2(\mathbb{R})$ , de modo que cualquier función de  $L_2(\mathbb{R})$  se puede poner en la forma

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k} \quad (4)$$

donde  $d_{j,k} = \langle f, \psi_{j,k} \rangle \triangleq \int_{-\infty}^{\infty} f(t) 2^{-j/2} \psi^*(2^{-j}t-k) dt$  es el producto escalar en  $L_2(\mathbb{R})$ .

### Multiresolución

La construcción de una familia de wavelets que formen una base completa de  $L_2(\mathbb{R})$  se puede realizar a partir del concepto de análisis multiresolución. Se define así a una secuencia de subespacios  $\{V_j\}_{j \in \mathbb{Z}}$  de  $L_2(\mathbb{R})$  que verifican

1. Los  $V_j$  están generados por una función de escala  $\varphi(t) \in L_2(\mathbb{R})$ , lo que quiere decir que  $\varphi(t)$  da lugar a la familia de funciones  $\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k)$ ,  $k \in \mathbb{Z}$  que para cada  $j \in \mathbb{Z}$  constituyen una base de Riesz de  $V_j$ , esto es

$$f(t) \in V_j \Rightarrow f(t) = \sum_{n=-\infty}^{\infty} a(n) \varphi(2^j t - n)$$

y

$$A \|f\|^2 \leq \sum_{n=-\infty}^{\infty} |a(n)|^2 \leq B \|f\|^2, \quad A, B > 0$$

2. Cada subespacio contiene al anterior:  $V_j \subset V_{j+1}$
3. El operador proyección  $P_j$  sobre  $V_j$  verifica  $\forall f(t) \in L^2(\mathbb{R})$

$$\lim_{j \rightarrow \infty} P_j f(t) = f(t) \quad \text{y} \quad \lim_{j \rightarrow -\infty} P_j f(t) = 0$$

De esta definición se deduce que  $f(t) \in V_j$  equivale a  $f(2t) \in V_{j+1}$  y que  $V_j$  es invariante frente a traslaciones de  $2^j$ . De la condición 2 se deduce que  $\varphi = \varphi_0$  se obtiene de la solución a la ecuación

$$\phi(t) = \sum_{n=-\infty}^{\infty} 2 \langle \phi(t), \phi(2t-n) \rangle \phi(2t-n) = \sum_{n=-\infty}^{\infty} h(n) \phi(2t-n) \quad (5)$$

La secuencia o señal  $\{h(n)\}$  la podemos asociar a un filtro discreto: aquel cuya respuesta a un impulso es  $h(n)$ . A partir de la función  $\varphi(t)$  de la definición se puede generar una función ortonormal que es una multiresolución de  $L_2(\mathbb{R})$ , por lo que de ahora en adelante suponemos que la familia  $\{\varphi_{j,k}(t)\}_{j,k \in \mathbb{Z}}$  es ortonormal.

Como  $V_j \subset V_{j+1}$  denominamos por  $W_j$  al complemento ortogonal de  $V_j$  en  $V_{j+1}$ :  $V_{j+1} = V_j \oplus W_j$ , siendo  $V_0$  el subespacio generado

por la base ortonormal  $\{\varphi(t-k)\}_{k \in \mathbb{Z}}$  y  $W_0$  su complemento ortogonal en  $V_1$ . Se tienen entonces las relaciones:

$$V_0 = V_{-1} \oplus W_{-1} = V_{-2} \oplus W_{-2} \oplus W_{-1} = \dots = \sum_{j<0} W_{-j}$$

y

$$V_1 = V_0 \oplus W_0, \quad V_2 = V_1 \oplus W_1 = V_0 \oplus W_0 \oplus W_1, \dots$$

$$V_{\infty} = L_2(\mathbb{R}) = V_0 \oplus \sum_{j=0}^{\infty} W_j = \sum_{j=-\infty}^{\infty} W_j$$

Claramente los subespacios  $W_j$  son ortogonales entre sí. Dado que  $W_0 \subset V_1$ , una función  $\psi(t) \in W_0$  se puede poner en la forma:

$$\psi(t) = \sum_{n=-\infty}^{\infty} 2 \langle \psi(t), \phi(2t-n) \rangle \phi(2t-n) = \sum_{n=-\infty}^{\infty} g(n) \phi(2t-n) \quad (6)$$

Al igual que para  $h(n)$ , la secuencia  $\{g(n)\}$  la podemos asociar a la respuesta a un impulso de entrada a un filtro discreto. Si consideramos el filtro con coeficientes  $g(n) = (-1)^{1-n} h(1-n)$ , entonces la señal  $\psi(t) \in W_0$  que se obtiene da lugar a una familia  $\{\psi_{j,k}(t)\}_{j,k \in \mathbb{Z}}$  que es una base ortonormal completa de  $L_2(\mathbb{R})$ . Las funciones así obtenidas forman una familia de wavelets ortonormales. El filtro  $H(z)$  asociado a  $\{h(n)\}$  lo podemos considerar un filtro pasa baja, mientras que el filtro  $G(z)$  asociado a  $\{g(n)\}$  lo podemos considerar un filtro pasa alta.

Dada una señal  $f(t)$ , si representamos por  $A_j(t)$  su proyección en  $V_j$ , y por  $D_j(t)$  su proyección en  $W_j$  se verifica:

$$A_{j+1}(t) = A_j(t) + D_j(t) \quad (7)$$

donde

$$A_j(t) = \sum_{k=-\infty}^{\infty} 2^j \langle f, \varphi_{j,k} \rangle \varphi_{j,k} = \sum_{k=-\infty}^{\infty} a_j[k] \varphi_{j,k} \quad (8)$$

y

$$D_j(t) = \sum_{k=-\infty}^{\infty} 2^j \langle f, \psi_{j,k} \rangle \psi_{j,k} = \sum_{k=-\infty}^{\infty} d_j[k] \psi_{j,k} \quad (9)$$

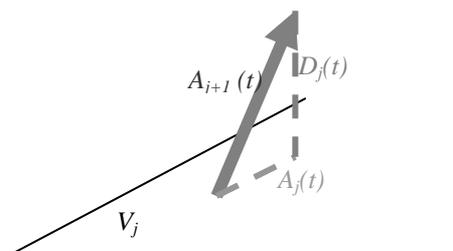


Fig. 1 Relación entre las proyecciones de una señal  $f(t)$  en los subespacios  $V_j$ ,  $W_j$  y  $V_{j+1}$

### Representación de la transformada discreta de wavelets

Las señales  $D_j(t)$  puede interpretarse como variaciones de detalle de la señal  $f(t)$  a la escala  $2^j$ . Estas capas de detalle se suman para mejorar progresivamente las aproximaciones a  $f(t)$  hasta su perfecta reconstrucción. Las señales  $A_j(t)$  se denominan aproximaciones y cada aproximación con mayor grado de resolución se forma de la aproximación anterior añadiendo un nuevo nivel de detalle  $D_j$ . A partir de las ecuaciones (5)-(9) se obtiene la siguiente relación entre coeficientes

$$a_{j-1}[n] = \sum_{k=-\infty}^{\infty} a_j[k]h[k-2n] \quad (10)$$

$$d_{j-1}[n] = \sum_{k=-\infty}^{\infty} a_j[k]g[k-2n]$$

De este modo los coeficientes de la proyección sobre  $V_{j-1}$  y  $W_{j-1}$  se obtienen a partir de los coeficientes de la proyección sobre  $V_j$  filtrando, respectivamente, con los filtros asociados a las secuencias  $h[-n]$  y  $g[-n]$  y posteriormente eliminando una de cada dos muestras ("downsampling" por 2). Repitiendo el proceso con  $a_{j-1}[n]$  se obtiene la proyección sobre  $V_{j-2}$  y  $W_{j-2}$ , y así sucesivamente.

Podemos también reconstruir los coeficientes  $a_{j+1}[n]$  a partir de los coeficientes  $a_j[n]$  y  $d_j[n]$  mediante la ecuación

$$a_{j+1}[n] = 2 \sum_{k=-\infty}^{\infty} h[n-2k]a_j[k] + 2 \sum_{k=-\infty}^{\infty} g[n-2k]d_j[k] \quad (11)$$

En la ecuación (11) las secuencias  $a_j[n]$  y  $d_j[n]$  se han redefinido con la inserción de ceros entre cada dos muestras ("upsampling" por 2). Siguiendo con este proceso se puede reconstruir los coeficientes  $a_{j+2}[n]$ , y así sucesivamente. Los métodos de descomposición y reconstrucción descritos se deben a Mallat [4].

Vemos pues que por la ecuación (4) las wavelets nos permiten realizar aproximaciones a una función mediante truncamiento de los elementos de la serie. Por otra parte, las ecuaciones (10) y (11) nos muestran que los coeficientes se pueden obtener a partir de filtros adecuados, por lo que el análisis multiresolución equivale a la elección de filtros pasa baja y pasa alta adecuados.

### 3. MÁQUINAS DE VECTORES SOPORTE

#### Caso separable linealmente

El objetivo de la clasificación de un conjunto de objetos es encontrar una regla que, en base a observaciones externas o elementos de entrenamiento, permita asignar cada objeto a alguna de varias posibles clases. Consideramos el caso más sencillo de objetos o puntos  $\mathbf{x}_i$  separables que pueden pertenecer a una de dos clases, a las que asignamos las etiquetas  $y_i \in \{-1, 1\}$ . Tenemos pues un conjunto de observaciones  $\{\mathbf{x}_i, y_i\}$ ,  $i = 1, \dots, l$ ,  $\mathbf{x}_i \in \mathbf{R}^d$ . El objetivo es establecer la ecuación de un hiperplano que divida a los elementos de ambas clases, de modo que todos los elementos de una clase estén en el mismo lado. Los puntos que pertenecen al hiperplano de separación verifican la ecuación:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

donde  $\mathbf{w}$  es un vector perpendicular al hiperplano. La distancia con signo de un punto  $\mathbf{x}_i$  al hiperplano es:

$$d_i = \frac{\mathbf{w} \cdot \mathbf{x}_i + b}{\|\mathbf{w}\|} \quad (12)$$

donde  $\|\mathbf{w}\|$  es la norma Euclídea de  $\mathbf{w}$ , y  $b/\|\mathbf{w}\|$  es la distancia con signo al origen. Representamos por  $|d_+|$  y por  $|d_-|$  a las distancias mínimas de las clases +1 y -1, respectivamente, al hiperplano separador. Se define el margen de separación del hiperplano a la distancia  $|d_+| + |d_-|$ . Para dos clases linealmente separable el objetivo que se persigue es encontrar el hiperplano de separación con mayor margen. Para encontrar una

formulación tratable matemáticamente, suponemos que todos los puntos que pertenecen a la clase +1 verifican

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq 1$$

y los que pertenecen a la clase -1 verifican

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1$$

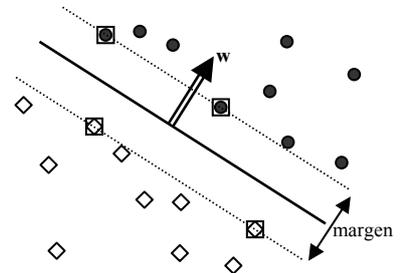


Fig. 2: Clasificación lineal en un caso separable; hiperplano de separación y margen de separación. Los vectores soporte aparecen encuadrados

Ambas ecuaciones pueden combinarse en el conjunto de desigualdades siguientes:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad \forall i \quad (13)$$

De las ecuaciones (12) y (13) se sigue el conjunto de desigualdades:

$$y_i d_i \geq \frac{1}{\|\mathbf{w}\|}, \quad \forall i \quad (14)$$

La ecuación (14) nos muestra que la inversa de la norma del vector  $\mathbf{w}$  es un límite inferior a la distancia entre los puntos observados  $\mathbf{x}_i$  y el hiperplano de separación definido por  $(\mathbf{w}, b)$ . El margen de separación lo establecen los elementos de ambas clases con distancia mínima al hiperplano y que por tanto verifican la igualdad en la ecuación (13); luego teniendo en cuenta la definición del margen de separación y la relación (14) tenemos:

$$d_+ - d_- = \frac{2}{\|\mathbf{w}\|} \quad (15)$$

En la figura 1 aparecen en un cuadrado aquellos elementos que definen el margen. Los elementos que verifican la igualdad en la condición (14) y cuya ausencia haría que cambiase el hiperplano solución se denominan vectores soporte (VS).

Así pues, podemos encontrar el hiperplano que da el máximo margen resolviendo el siguiente problema de optimización cuadrática:

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w} \cdot \mathbf{w} \quad (16)$$

sujeto a  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, l$

La solución de este problema de optimización se puede obtener a partir de la formulación dual. Para ello se introducen un conjunto de multiplicadores de Lagrange,  $\alpha_i \geq 0, i = 1, \dots, l$ , uno para cada una de las restricciones en (16), y se forma la función Lagrangiana

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1] \quad (17)$$

La solución de (16) equivale a determinar el punto de silla de la función Lagrangiana. En el punto de silla la Lagrangiana tiene un mínimo con respecto a  $\mathbf{w}$  y a  $b$ , y un máximo con respecto a  $\alpha$ . Luego se deben cumplir las ecuaciones

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l \alpha_i y_i = 0 \quad \text{y} \quad \frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i = 0 \quad (18)$$

Sustituyendo estas relaciones en (17) se obtiene el problema dual de optimización cuadrática siguiente:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{sujeto a} \quad & \alpha_i \geq 0, \quad i=1, \dots, l \\ & \sum_{i=1}^l \alpha_i y_i = 0 \end{aligned} \quad (19)$$

A los parámetros  $\alpha_i$  solución de (19) los representamos por  $\alpha_i^*$ ,  $i=1, \dots, l$ . Con estos valores obtenemos el vector  $\mathbf{w}$  solución de la ecuación (16) utilizando la segunda condición de la ecuación (18):

$$\mathbf{w}^* = \sum_{i=1}^l \alpha_i^* y_i \mathbf{x}_i \quad (20)$$

Los puntos  $\mathbf{x}_i$  correspondientes a los parámetros  $\alpha_i^*$  que son mayores que cero se denominan vectores soporte, y solamente ellos son necesarios para determinar el vector solución dado por (20). Las condiciones de Karush-Kuhn-Tucker [13] son condiciones necesarias, y en algunos casos suficientes, para una solución óptima a un problema de optimización. Una de estas condiciones establece que si el multiplicador  $\alpha_i \geq 0$ , está asociado a la restricción  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b^*) \geq 1$  entonces se debe cumplir la relación:

$$\alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b^*) - 1] = 0 \quad (21)$$

De lo que se deduce que los vectores soporte corresponden a aquellas restricciones de la ecuación (16) para las que se verifica la igualdad, así

$$\alpha_i^* > 0 \Rightarrow y_i(\mathbf{w}^* \cdot \mathbf{x}_i + b^*) = 1 \quad (22)$$

El parámetro  $b^*$  se puede obtener a partir de la relación (22) con cualquiera de los vectores soporte, sin embargo una solución más robusta se obtiene calculando el valor medio para todos los vectores soporte.

La clasificación de un nuevo punto  $\mathbf{x}$  se realiza mediante la siguiente función de clasificación:

$$c(\mathbf{x}) = \text{signo}(\mathbf{w}^* \cdot \mathbf{x} + b^*)$$

Un valor positivo clasifica a  $\mathbf{x}$  como perteneciente a la clase +1, y un valor negativo lo clasifica en la clase -1.

### Caso no separable linealmente

En el caso de que las dos clases no sean separables de forma lineal, no habrá solución factible al problema de optimización. Una forma de abordar el problema es introduciendo variables de holgura para relajar las restricciones en (16)

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \beta_i, \quad \beta_i \geq 0, \quad i=1, \dots, l$$

De este modo, si un  $\mathbf{x}_i$  verifica la condición (13) entonces  $\beta_i$  es cero y la ecuación (22) equivale a (13); pero si no se verifica la condición (13) se considera el término  $-\beta_i$  de modo que se permite la violación de (13). No obstante, para que se minimice el número de violaciones se añade a la función de coste un término adicional en la forma  $C \sum_{i=1}^l \beta_i$ , de modo que el problema a resolver es ahora:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^l \beta_i \\ \text{sujeto a} \quad & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \beta_i, \quad i=1, \dots, l \\ & \beta_i \geq 0, \quad i=1, \dots, l \end{aligned} \quad (23)$$

El parámetro  $C > 0$  determina un compromiso entre el máximo margen de separación y el número de puntos clasificados erróneamente. Un valor elevado de  $C$  hace que se considere una gran penalización en el número de errores y un valor pequeño da una pequeña penalización. El problema dual se plantea ahora de la forma:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{sujeto a} \quad & 0 \leq \alpha_i \leq C, \quad i=1, \dots, l \\ & \sum_{i=1}^l \alpha_i y_i = 0 \end{aligned} \quad (24)$$

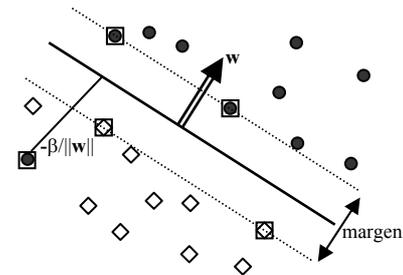


Fig. 3: Clasificación lineal en un caso no separable.

Nuevamente representamos por  $\alpha_i^*$ ,  $i=1, \dots, l$  a los parámetros solución de (24), y el vector del hiperplano solución está dado por

$$\mathbf{w}^* = \sum_{i=1}^l \alpha_i^* y_i \mathbf{x}_i \quad (25)$$

Como se ve en la ecuación (24), la diferencia con el planteamiento del hiperplano óptimo para clases separadas es la limitación en los valores máximos de los multiplicadores de Lagrange,  $\alpha_i \leq C$ . Las condiciones de KKT aplicadas a este problema llevan a las siguientes relaciones

$$\begin{aligned} \alpha_i = 0 & \Rightarrow y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \text{y} \quad \beta_i = 0 \\ 0 < \alpha_i < C & \Rightarrow y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1 \quad \text{y} \quad \beta_i = 0 \\ \alpha_i = C & \Rightarrow y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \leq 1 \quad \text{y} \quad \beta_i \geq 0 \end{aligned} \quad (26)$$

Al igual que para el caso separable, se denominan vectores soporte a aquellos  $\mathbf{x}_i$  para los que  $\alpha_i > 0$ , pero ahora cabe distinguir dos casos, aquellos vectores soporte para los que  $\alpha_i < C$  y que corresponden a los vectores soporte que delimitan el margen, por lo que se les suele denominar vectores margen, y aquellos para los que  $\alpha_i = C$  se suelen denominar errores. Dentro de estos se pueden diferenciar aquellos casos situados dentro del margen  $\beta_i \leq 1$  y que están correctamente clasificados,

de aquellos para los que  $\beta_i > 1$  que corresponden a una clasificación errónea.

Al igual que en el caso separable, la clasificación de un nuevo punto  $\mathbf{x}$  se realiza examinando el signo de la relación  $\mathbf{w}^* \cdot \mathbf{x} + b^*$ .

### Maquinas de vectores soporte no lineales

El método expuesto se puede extender a casos en los que se necesitan funciones no lineales para delimitar las clases de objetos. La generalización se basa en el hecho de que los puntos de entrenamiento aparecen en el problema de optimización y en la función de clasificación en forma de producto escalar ( $\mathbf{x}_i \cdot \mathbf{x}_j$  o  $\mathbf{w}^* \cdot \mathbf{x}$ ). Supongamos que hacemos una transformación  $\Phi$  de los puntos de entrenamiento a un espacio distinto  $F$ , denominado espacio de características, de mayor dimensión, incluso, quizás, de dimensión infinita,

$$\mathbf{x} \in \mathbb{R}^d \xrightarrow{\Phi} \Phi(\mathbf{x}) \in F$$

Consideramos ahora el método anterior al problema de la clasificación en el espacio  $F$ , con los datos de observación  $\{\Phi(\mathbf{x}_i), y_i\}$ ,  $i=1, \dots, l$ . Lo interesante de este método es que el aprendizaje para clasificar en este espacio puede llevar a reglas de decisión más sencillas.

Al igual que en el caso lineal, en el algoritmo se van a necesitar el cálculo de productos escalares  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ , lo que puede resultar un problema para espacios de gran dimensión. No obstante, se puede encontrar una solución sencilla mediante el uso de las denominadas funciones núcleo, que verifican  $k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ . De modo que el problema de la clasificación en el espacio  $F$  se plantea ahora, en su forma más general, como la resolución del problema de optimización

$$\begin{aligned} \max_a \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{sujeto a} \quad & 0 \leq \alpha_i \leq C, \quad i=1, \dots, l \\ & \sum_{i=1}^l \alpha_i y_i = 0 \end{aligned} \quad (27)$$

y la clasificación de un nuevo punto  $\mathbf{x}$  se realiza mediante la función

$$c(\mathbf{x}) = \text{signo} \left( \sum_{\substack{\text{vectores} \\ \text{soporte}}} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b^* \right)$$

Tabla 1. Funciones núcleo normalmente utilizadas en máquinas de vectores soporte no lineales. Los parámetros se deben elegir de la forma:  $d \in \{1, 2, \dots\}$ ,  $\sigma \in \mathbb{R}$ ,  $v \in \mathbb{R}$ ,  $c \in \mathbb{R}$ .

Polinomial	$k(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^d$
Función de base radial	$k(\mathbf{x}_i, \mathbf{x}_j) = \exp \left( -\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2} \right)$
Red neuronal sigmoideal	$k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(v(\mathbf{x}_i \cdot \mathbf{x}_j) + c)$
Inversión multicuadrática	$k(\mathbf{x}_i, \mathbf{x}_j) = \left( \sqrt{\ \mathbf{x}_i - \mathbf{x}_j\ ^2 + c} \right)^{-1}$

El problema principal a la hora de utilizar este método está en la elección de la función núcleo y en el valor del parámetro o parámetros asociados.

## 4. EXPERIMENTOS

Vamos a mostrar algunos resultados preliminares que se han obtenido al aplicar las técnicas descritas en el reconocimiento de formas de la base de datos de señales del TJII ubicado en el CIEMAT (Madrid, España). El TJ-II es un dispositivo del tipo stellarator [14] (tipo heliac,  $B(0) \leq 1.2$  T,  $R(0) = 1.5$  m,  $\langle a \rangle \leq 0.22$  m) y puede explorar un amplio rango de transformada rotacional ( $0.9 \leq \text{iota}/2p \leq 2.2$ ). En el TJII los plasmas se producen y se calientan con ECRH (2 girotrones de 300 kW cada uno, 53.2 GHz, 2nd armónico, modo de polarización X) y con NBI (300 kW). Dispone de 940 canales digitales para medidas experimentales.

En los experimentos realizados se han utilizado señales pertenecientes a algunas de las clases que se muestran en la tabla 2.

Tabla 2. Clases de señales de la base de datos del TJII

BOL5	Señal de bolometría
ECE7	Emisión electrón-ciclotrón
RX306	Rayos-X blandos
ACTON275	Señal espectroscópica
HALFAC3	Emisión de la línea ALFA de hidrógeno
Densidad2	Densidad electrónica de línea

Se calcula la TW de las señales para reducir su dimensión. La figura 4 muestra una señal original y su transformada wavelet para aproximaciones con tres niveles distintos.

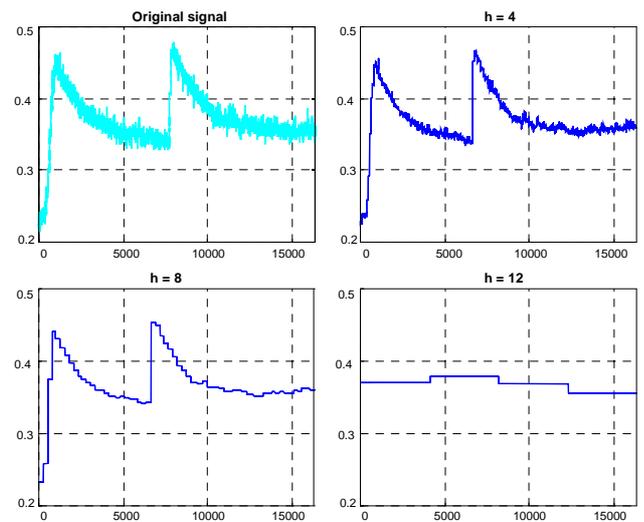


FIGURA 4: Transformación wavelet de una señal con diferentes niveles de descomposición.

Las señales se clasifican usando SVM. Se hace una clasificación en múltiples clases a partir de clasificaciones binarias de cada tipo de clase frente al resto mediante un grupo de señales de entrenamiento. Se determinan los vectores soporte y estos se usan para clasificar señales que no hayan sido utilizado en el aprendizaje.

Se muestran los resultados de dos tipos de experimentos. En el primero se consideran cuatro clases de señales: ECE7, BOL5, RX306 y Densidad2. El conjunto de entrenamiento consta de 40 señales, y el conjunto de test posterior de 32 señales. La figura 5 muestra los vectores soporte de la clase individual (+1) cuando se usa un kernel lineal. La figura 6 muestra los resultados de las

clasificaciones correctas, incorrectas y que no se pueden clasificar.

En el segundo experimento se muestran los resultados obtenidos con distintos kernels. En este caso el número de señales de entrenamiento es de 60 y el número de señales de test es de 48.

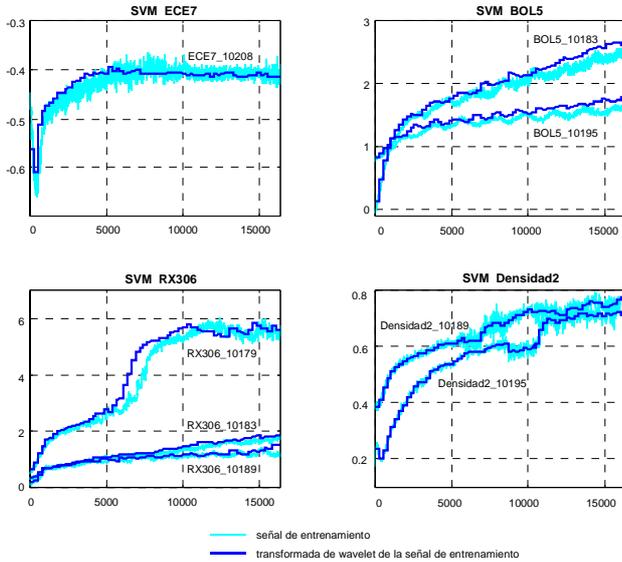


FIGURA5: Representación de los vectores soportes positivas para cada una de las cuatro clases.

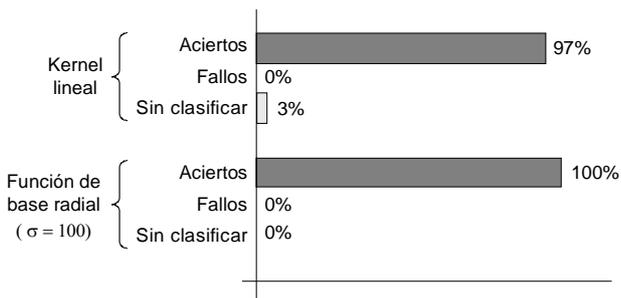


FIGURA6: Resultados para el experimento de cuatro clases de señales.

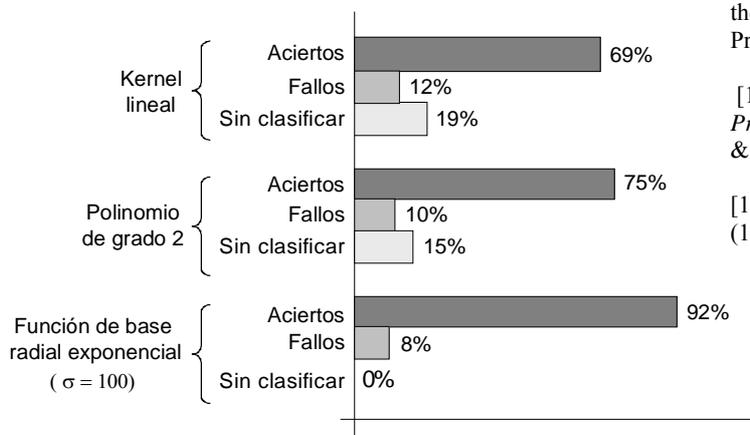


FIGURA7: Resultados para el experimento de seis clases de señales.

## 5. CONCLUSIONES

Aunque preliminares, los resultados obtenidos resultan prometedores en el uso de la TW, para reducir la información necesaria para describir las señales, y de los SVM para la clasificación y recuperación de señales de fusión.

## 6. REFERENCIAS

- [1] Vega, J. Estado estacionario, tiempo real y participación remota en el horizonte de los dispositivos de fusión. CISCI 2004.
- [2] Radiei, D., Mendelzon, A. "Efficient Retrieval of Similar Time Sequences Using DFT", Proc. 5<sup>th</sup> Inter. Conf. on Foundations of Data Organization (FODO'98), pp. 249-257, 1998.
- [3] Nakanishi, H., Hochin, T., Kojima, M. and LABCOM group, 'Search and Retrieval Methods of Similar Plasma Waveforms'. Fusion Eng. (to appear).
- [4] Mallat, S. *A Wavelet Tour of Signal Processing*". 2<sup>o</sup> Edition, Academic Press, 2001.
- [5] Vetterli, M. "Wavelets, Approximation and Compression", IEEE Signal Processing Magazine, pp. 59-73, September 2000.
- [6] Daubechies, I., *Ten Lectures on Wavelets*, SIAM, 1992.
- [7] Vapnik, V., "*Statistical Learning Theory*", John Wiley & Sons, INC, 1998.
- [8] Vapnik, V., "*The Nature of Statistical Learning Theory*", 2<sup>o</sup> Edition, Springer,2000.
- [9] Schölkopf, B., Smola, A.J., *Learning with Kernels*, MIT Press, 2002
- [10] Duda, R.O., Hart, P.E., Stork, D.G., "*Pattern Classification*", 2<sup>o</sup> Edition, John Wiley, 2001.
- [11] Burges, C.J.C., "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2, 121-167, 1998.
- [12] Sebal, J.D. Bucklew, J.A., "Support Vector Machines and the Multiple Hypothesis Test Problems". IEEE Trans. on Signal Processing, vol. 49, n° 11, pp. 2865-2872, 2001.
- [13] Bazaraa, M.S., Sherali, H.D., Shetty, C.M. *Nonlinear Programming. Theory and Algorithms*, 2<sup>o</sup> Edition, John Wiley & Sons, 1993.
- [14] Alejaldre, C. *et al.* Plasma Phys. Controlled Fusion 41, 1 (1999), pp. A539.