Teoría básica de los filtros de ondas.

Un filtro es simplemente un operador definido entre dos espacios de señales (funciones, distribuciones, etc.) que tiene las propiedades de ser lineal, continuo, e invariante frente a traslaciones. Esta última propiedad significa que el efecto producido en la salida del sistema por un retardo en la señal de entrada es sencillamente el mismo retardo para la señal de salida. Si la señal exponencial compleja $e_w(t) = e^{iwt}$ se toma como entrada del filtro L, entonces es sencillo probar que $L(e_w(t)) = \lambda(w) e_w(t)$ para cierto número complejo $\lambda(w)$. Esto significa que el operador L tiene un comportamiento muy simple sobre las señales que podamos representar como superposición de exponenciales complejas y, por tanto, resulta una de las motivaciones más naturales que podamos dar para el estudio del análisis de Fourier en sus diversas variantes.

Evidentemente, dependiendo de la naturaleza de los espacios de señales implicados, los filtros tendrán una expresión analítica especial. En el caso más sencillo de todos, que es cuando las señales de entrada forman un espacio de Hilbert y las de salida son sencillamente señales acotadas, los filtros son siempre operadores de convolución. Es decir, son operadores del tipo $L(x)(t) = \int_{-\infty}^{\infty} x(s)h(t-s)ds$, en el caso analógico, o bien del tipo $L(x)[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k]$, en el caso digital. Esta representación hace muy sencillo caracterizar la causalidad o la realizabilidad física de los filtros de ondas en términos de la respuesta al impulso unidad h, en el dominio del tiempo. Para ello, basta tener en cuenta que, si queremos que el filtro sea causal, $L(x)$ sólo puede depender de los valores pasados de la entrada x, y, por tanto, h debe anularse para los valores negativos de su variable. La realizabilidad física se traduce en utilizar a lo sumo una porción acotada del futuro de la entrada, lo que está garantizado por el hecho de que h se anule para todos los valores menores a uno dado.

Si tomamos la transformada de Fourier a ambos lados de la igualdad $L(x)(t) = \int_{-\infty}^{\infty} x(s)h(t-s)ds$, entonces el filtro L queda caracterizado en el dominio de la frecuencia como un operador del tipo $L(X)(w) = X(w)H(w)$, donde $X(w), H(w)$ representan las transformadas de Fourier de $x(t), h(t)$, respectivamente (una propiedad totalmente análoga se da para el caso digital). El filtro queda, pues, determinado por la "función de transferencia" $H(w)$. Lo que es más, esta función da una descripción muy intuitiva de la acción del filtro sobre las distintas componentes frecuenciales de la señal de entrada del sistema. Así pues, lo más natural para un ingeniero eléctrico es diseñar sus filtros en el dominio de la frecuencia precisamente dando una descripción adecuada de las funciones de transferencia asociadas. Sin embargo, esto plantea una cuestión nada sencilla de responder: dada una función $H(w)$, ¿cómo sabemos si ésta se corresponde con la función de transferencia de un filtro causal o, al menos, de un filtro físicamente realizable? Es a esta cuestión precisamente a la que Paley y Wiener respondieron en 1933. Ellos probaron que la condición necesaria y suficiente para que $H(w)$ sea la función de transferencia de un filtro físicamente realizable es que $\int_{-\infty}^{\infty}\frac{\left|\log|H(w)|\right|}{1+w^2}dw<\infty$. En particular, es evidente que la función de transferencia no puede tener demasiados ceros. Concretamente, sus ceros en el eje real deben ser forzosamente aislados. Una de las consecuencias de esto es que los filtros ideales no son físicamente realizables y, por tanto, el problema de su aproximación mediante el uso de filtros realizables es uno de los más significativos e importantes de la ingeniería eléctrica.

Al principio, cualquier ingeniero se sorprenderá ante el enunciado del Teorema de Paley-Wiener porque en él no aparece para nada la fase de la función de transferencia (el teorema depende exclusivamente de su módulo). Sin embargo, el resultado es correcto porque cuando nos restringimos a los filtros causales la fase y el módulo de $H(w)$ no son independientes sino que, por el contrario, conocida una de estas funciones, podemos decir quién es la otra con total precisión. Este hecho es muy importante y también debemos a Wiener su descubrimiento.

Los operadores de convolución del tipo $L(x)(t)=\int_{-\infty}^{\infty}x(s)h(t-s)ds$ siempre satisfacen la propiedad de invarianza respecto del tiempo y, por tanto, podrían ser de interés tambien cuando cambiamos los espacios de señales de entrada y de salida. Por ejemplo, para garantizar la estabilidad de este operador (y, por tanto, su carácter de filtro de ondas) cuando tanto las entradas como las salidas son señales acotadas, bastará exigir que la respuesta al impulso unidad sea una función sumable. Este tipo de estabilidad se llama habitualmente estabilidad BIBO (del inglés: “Bounded Input, Bounded Output”) y, en el caso de que la función de transferencia $H(w)$ sea una función racional, el sistema asociado es estable BIBO si y solo si los polos de $H(w)$ tienen todos parte real negativa.

Los filtros de ondas son sin duda muy útiles en ingeniería. Vamos a explicar aquí una de sus aplicaciones más sencillas, para que el lector comprenda su importancia en la vida diaria. Supongamos que queremos emitir a través de una antena de radio varios canales de forma simultánea, de modo que un receptor sea capaz posteriormente de recibir cualquiera de estos canales de forma independiente y libre. ¿Cómo se hace esto? La idea es muy sencilla. Si la señal emitida por el canal k-ésimo es la función $x_k(t)$, la cual supondremos que es una señal con ancho de banda finito menor o igual a 20.000 Herzios –pues ese es el límite impuesto por nuestra sensibilidad auditiva- entonces hacemos que nuestra antena emita la señal $x(t)=\sum_{k=1}^{n}e^{iw_kt}x_k(t)$, de modo que la señal emitida por cada canal se modula, en el dominio del tiempo, pasando a ser una señal del tipo $e^{iw_kt}x_k(t)$, que es una señal limitada en banda con ancho de banda el de la señal original $x_k(t)$ pero con intervalo de frecuencias centrado en el punto w_k. Esto tiene el efecto de que si tomamos los valores w_k de modo que los intervalos

$$[w_k-20.000,w_k+20.000];\ k=1,2,\cdots,n$$

son disjuntos dos a dos, entonces podemos hacer que nuestro aparato receptor filtre la señal recibida $x(t)$ con intervalo paso banda $[w_k-20.000,w_k+20.000]$ y luego multiplique la señal recibida por e^{-iw_kt}, obteniendo de esta forma precisamente la señal original $x_k(t)$. A este proceso se le denomina modulación en amplitud y es el método empleado para emitir en frecuencia AM. Evidentemente, aunque el filtrado ideal impuesto aquí no es físicamente realizable, resulta que éste se puede aproximar muy bien con filtros que sí lo son y, además, teniendo en cuenta que las señales emitidas ya están limitadas en banda desde el principio, el proceso se obtiene con precisión absoluta.

Retroalimentación y oscilación.

Lo primero que vamos a hacer es explicar cuándo se produce la estabilidad de un filtro de ondas con función de transferencia una función racional, y en qué se traduce en términos del comportamiento del filtro. Luego veremos cómo aplicar estos resultados a ciertos filtros retroalimentados.

Se sabe que si $L(x)(t)=(x*h)(t)=\int_{-\infty}^{+\infty}x(s)h(t-s)ds$ es un operador de convolución, su estabilidad BIBO (es decir, la propiedad de que éste transforma, con continuidad, señales acotadas en señales acotadas) equivale a que $\int_{-\infty}^{\infty}|h|<\infty$. En particular, existe la transformada de Fourier de la respuesta al impulso unidad del sistema, y ésta es una función continua que se anula en el infinito. Por tanto, sabemos que el filtro se puede describir en términos de su función de transferencia $H(w)$, y que ésta se puede representar como $H(w)=R(iw)$, donde $R(s)=\int_{-\infty}^{\infty}h(t)e^{-st}dt$ es la transformada de Laplace de h. Por otra parte, la causalidad del filtro se traduce en que la región de convergencia de la transformada de Laplace es precisamente el semiplano de números complejos cuya parte real sea mayor que la parte real del polo más a la derecha de la función de transferencia. Si ésta es una función racional, el criterio se convierte en asegurar que todos sus polos tienen parte real negativa.

Los argumentos anteriores son un tanto hetéreos, pues no nos dan una clave sencilla de lo que significa la estabilidad del sistema. Es por ello que, a continuación, vamos a dar una argumentación diferente que, en nuestra opinión, sí arroja luz sobre esta cuestión, pues nos permite visualizar perfectamente cómo es la respuesta al impulso unidad de un sistema (que supondremos causal) cuya función de transferencia es una función racional.

Supongamos, pues, que nuestro filtro tiene función de transferencia una función racional que, por conveniencia, escribiremos como

$$R(iw)=\frac{P(iw)}{Q(iw)}=\frac{b_m(iw)^m+b_{m-1}(iw)^{m-1}+\cdots+b_0}{a_n(iw)^n+a_{n-1}(iw)^{n-1}+\cdots+a_0}.$$

Entonces es un ejercicio muy sencillo comprobar que esta función se corresponde precisamente con la función de transferencia del sistema descrito por la ecuación diferencial lineal de coeficientes constantes dada por

$$a_0y(t)+a_1y'(t)+\cdots+a_ny^{(n)}(t)=b_0x(t)+b_1x'(t)+\cdots+b_mx^{(m)}(t).$$

(para verlo, basta tomar la transformada de Laplace a ambos lados de la ecuación, despejar, y evaluarla en $s=iw$). Además, para que $R(iw)$ pueda ser la función de transferencia de un filtro estable, es necesario que se anule en el infinito, por lo que sabemos que necesariamente $m<n$. Ahora bien, descomponiendo $R(s)$ como suma

de fracciones simples (algo que podemos hacer gracias al teorema fundamental del álgebra), llegamos a que

$$R(s)=\sum_{k=1}^{N}\sum_{j=1}^{m_k}\frac{C_{kj}}{(s-\alpha_k)^j}$$

para ciertos números complejos C_{kj}, donde $\{\alpha_k\}_{k=1}^{N}$ son los polos de $R(s)$ y $\{m_k\}_{k=1}^{N}$ son sus correspondientes multiplicidades. Lo interesante de esta descomposición es que nos permite hallar fácilmente la transformada de Laplace inversa, que necesariamente coincide con h, y está dada por

$$h(t)=\sum_{k=1}^{N}\sum_{j=1}^{m_k}\frac{C_{kj}}{j!}t^j e^{\alpha_k t}u(t),$$

donde $u(t)$ es la función escalón unidad de Heaviside. Ahora bien, si $\alpha=a+ib$ es un número complejo, entonces

$$t^j e^{\alpha t}u(t)=t^j(\cos(bt)+i\sin(bt))e^{at}u(t),$$

por lo que cada uno de los sumandos de la expresión anterior es una función oscilante y el tamaño de sus oscilaciones o bien crece exponencialmente (si $\mathrm{Re}(\alpha_k)>0$) o bien decrece exponencialmente (si $\mathrm{Re}(\alpha_k)<0$) o permanece acotado (si $\mathrm{Re}(\alpha_k)=0$) para $t\to+\infty$.

Así pues, la respuesta del sistema al impulso unidad es una función absolutamente integrable si y solo si $\mathrm{Re}(\alpha_k)<0$ para $k=1,2,\cdots,N$; que es lo que queríamos demostrar.

Supongamos ahora que $R(iw)$ es la función de transferencia del sistema. ¿Cómo calculamos los polos de la función racional $R(z)$ que hay a la derecha o a la izquierda del eje de ordenadas? Pues hacemos uso de un conocido resultado de variable compleja que los matemáticos denominamos principio del argumento y los ingenieros llaman propiedad de circunvalación. Según este principio, si C es una curva cerrada en el plano y $f(z)$ es una función de variable compleja que es derivable en un entorno de la curva y en todos los puntos interiores a la misma, excepto quizás unos pocos puntos en los que puede tener polos, y si asumimos además que para todo punto $z\in C$, $f(z)\neq 0$, entonces:

$$\frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)}dz=N_0(f)-N_P(f),$$

donde $N_0(f), N_P(f)$ denotan, respectivamente, el número de ceros y el número de polos de f interiores a la curva C (contados con multiplicidad), y se supone que la integral se toma recorriendo la curva en el sentido contrario a las agujas del reloj.

Este resultado toma el nombre de principio del argumento porque la integral $\frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)}dz$ representa el número de vueltas que da la curva $f(C)$ en torno al origen de coordenadas (dicho número es la diferencia: número de vueltas en sentido contrario a las agujas del reloj menos número de vueltas en el sentido de las agujas del reloj).

Aplicando el principio del argumento a la función $R(s)$ sobre la curva C_M que se obtiene de unir el intervalo $[-Mi,Mi]$ con la semicircunferencia que une los extremos de este intervalo y está contenida en el semiplano derecho, $\Pi^{+} = \{ z \in \mathbb{C} : \mathrm{Re}(z) > 0 \}$, y haciendo $M \to \infty$, vemos que, a partir de cierto valor de $M > 0$, la curva $R(C_M)$ da el mismo número de vueltas alrededor del origen de coordenadas (pues dicha curva ya encierra todos los ceros y polos de $R(z)$ en el semiplano derecho Π^{+}) Además, como $\lim_{|z| \to \infty} R(z) = 0$, para M suficientemente grande, la parte de la curva C_M correspondiente a la semicircunferencia no aporta nada al número de vueltas de C_M, por lo que podríamos hallar directamente el número de vueltas dado por la curva imagen del eje de ordenadas (añadiendo el punto del infinito, que es el origen) alrededor del origen de coordenadas. Ese valor es necesariamente igual a $N_0(R) - N_P(R)$, donde $N_0(R), N_P(R)$ denotan, respectivamente, el número de ceros y el número de polos de R en el semiplano derecho Π^{+}. Esto implica que si conocemos el número de ceros de R en el semiplano derecho Π^{+}, entonces también podemos hallar el número de polos en esa zona.

Los argumentos anteriores son especialmente útiles cuando se aplican a un sistema retroalimentado con función de transferencia

$$R(s) = \frac{H(s)}{1 + \lambda G(s) H(s)}.$$

Los polos de $R(s)$ son precisamente los ceros de $\Lambda(s) = \frac{1}{\lambda} + G(s)H(s)$. Además, si aplicamos nuestro argumento anterior a la curva $\Gamma = \{ G(iw)H(iw) : w \in \mathbb{R} \}$, vemos que el número de vueltas que da Γ alrededor del punto $-\frac{1}{\lambda}$, que denotamos por $n(\Gamma, -\frac{1}{\lambda})$, verifica la fórmula:

$$n(\Gamma, -\frac{1}{\lambda}) = N_0(\frac{1}{\lambda} + GH) - N_P(\frac{1}{\lambda} + GH) = N_P(\frac{H}{1 + \lambda GH}) - N_P(GH)$$

Por tanto, el sistema retroalimentado es estable (es decir, $N_P(\frac{H}{1 + \lambda GH}) = 0$) si y sólo si

$$n(\Gamma, -\frac{1}{\lambda}) = -N_P(GH).$$

Éste es precisamente el criterio de H. Nyquist (1889-1976) para la estabilidad de sistemas retroalimentados, y es el que utiliza Wiener en sus cálculos en el capítulo de retroalimentación y oscilación de *Cibernética*. Con él a mano, Wiener calcula varios ejemplos especialmente significativos y, entre otras cosas, demuestra la necesidad de varias retroalimentaciones simultáneas para llevar a la estabilidad determinado tipo de sistemas.