# TRATAMIENTO DIGITAL DE SEÑALES

Ingeniería de Telecomunicación (4º, 2º c)

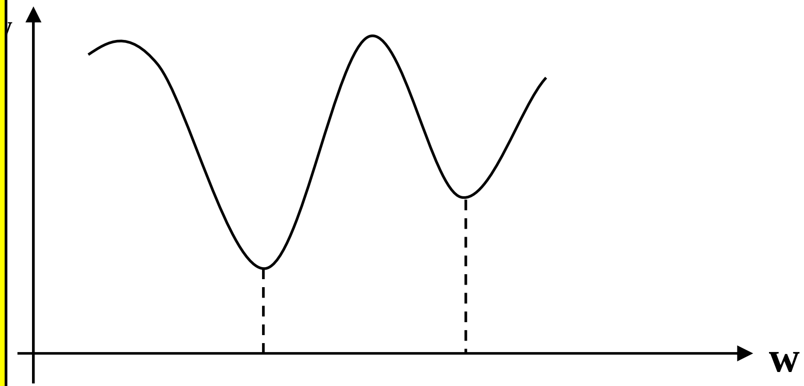
## Unidad 9ª: Búsqueda


Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

Área de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

## ...eda. Tipos de Búsqueda

...e ha visto que, en la aproximación máquina, se trata de minimizar un ...lefinido a partir de las salidas deseadas, $d^{(k)}$, y las que proporciona la ...a, $o^{(k)} = F_{\mathbf{w}}(\mathbf{x}^{(k)})$ (u $o^{(k)} = f_{\mathbf{w}}(\mathbf{x}^{(k)})$), acumulando sus valores según k; en ...va, una forma $C_{\mathbf{w}}$, cuya variación con **w** es arbitraria:




...ta de buscar el <u>mínimo absoluto</u>, o, en su defecto, uno <u>relativo</u> aceptable.

No siempre es posible obtener una solución cerrada (como en el caso de ón lineal, p. ej.): en cuyo caso hay que recurrir a aplicar un algoritmo de da de dicho mínimo.

E, incluso cuando sea accesible una solución cerrada, puede interesar una búsqueda para conseguir generalización: ya que la minimización s muestras no equivale a la minimización para todo el espacio de ación, y puede producirse sobreajuste; lo que debe evitarse:

gularizando la solución;

giendo un **coste** adecuado;

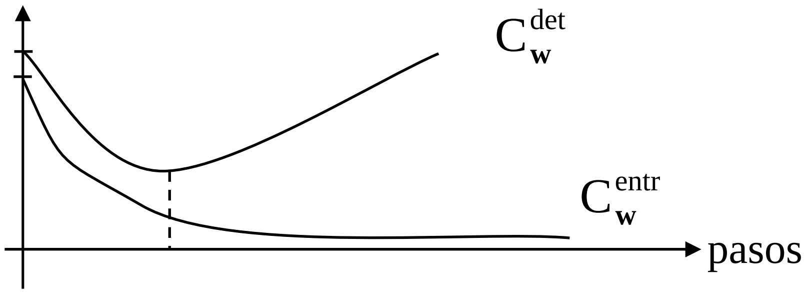
bien, entrenando hasta un cierto límite: lo que es posible, p.ej., mediante detención de un **algoritmo iterativo** de búsqueda.

Jn algoritmo iterativo va modificando (por pasos) **w** con objeto de $C_{\mathbf{w}}$: si se dividen las muestras disponibles en dos conjuntos:

ntrenamiento: con cuyos elementos se hace la búsqueda;

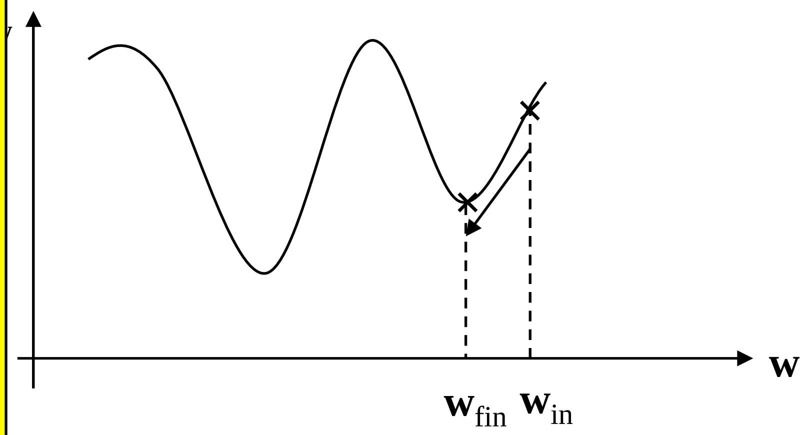
etención (o validación): que simplemente se usa para ir midiendo $C_{\mathbf{w}}$

mente, el 80% y 20% de las muestras, respectivamente), se puede ar que, antes de llegar al mínimo coste en el conjunto de entrenamiento, el de validación (al perder capacidad de generalización): ahí se puede la búsqueda



Hay dos tipos de métodos de búsqueda:

**ales**: con los que se puede acceder a un mínimo relativo, a partir del punto en que se inicialice el algoritmo, que minimiza paso a paso el coste; con lo que sólo es posible llegar al mínimo que se encuentre en la <u>cuenca de atracción</u> del punto inicial




**bales**: que persiguen alcanzar el mínimo absoluto.

Los métodos de búsqueda globales son, como es fácil comprender,
jos, delicados y lentos: computacionalmente muy costosos; por lo que
aplican en casos en que merezca la pena perseguir el mínimo absoluto
, si basta un buen mínimo relativo, suele ser suficiente repetir búsquedas
con distintas inicializaciones y seleccionar la mejor solución). Tienen
entos muy distintos: pero todos incluyen un mecanismo de **exploración**
acio de soluciones (**w**) y uno de **explotación** de los resultados favorables
a exploración.

<u>…da local</u>

<u>…oritmos de gradiente</u>

… trabajase mediante procedimientos analíticos, es decir, conociendo $\overline{C}(\mathbf{w})$ …e medio, que incluye la promediación sobre todo el espacio muestral), y …$_{\mathbf{w}}\overline{C}(\mathbf{w})$ es su gradiente, está claro que proceder de la forma

$$\mathbf{w}^{nuevo} = \mathbf{w}^{ant} - \eta \mathbf{g}\left(\mathbf{w}^{ant}\right) \qquad (\eta > 0)$$

…irá a un mínimo local, si $\eta$ es suficientemente pequeño.

…procedimiento se conoce como **algoritmo del descenso más pendiente** …est descent”): requiere conocer $\overline{C}(\mathbf{w})$, lo que no es habitual.

En situaciones muestrales, se puede trabajar con el coste para las
as; aunque puede procederse para la totalidad (o subconjuntos del total),
uente y recomendable hacerlo secuencialmente (muestra a muestra:
o éstas tantas veces como se precise, y extendiendo de este modo {k}):

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \eta \mathbf{g}^{(k)}\left(\mathbf{w}^{(k)}\right)$$

el **algoritmo (secuencial) de gradiente**. En esta regla, $\mathbf{g}^{(k)}$ es el
te del coste evaluado para la k-ésima muestra.

Nótese que es adaptativo si k es el tiempo).

Discutiremos su comportamiento <u>en media</u>, sustituyendo $\mathbf{g}^{(k)}$ por $\mathbf{g}$: pero
e aclarar que al coste mínimo teórico se añadirá un **error de desajuste**
al efecto del muestreo.

En las cercanías del mínimo buscado, $\mathbf{w}_0$, se podrá escribir:

$$C(\mathbf{w}) = C(\mathbf{w}_0) + (\mathbf{w} - \mathbf{w}_0)^T \mathbf{g}(\mathbf{w}_0) + \frac{1}{2}(\mathbf{w} - \mathbf{w}_0)^T H_{\mathbf{w}}(\mathbf{w}_0)(\mathbf{w} - \mathbf{w}_0)$$

$H_{\mathbf{w}}$ la matriz hessiana de C($\mathbf{w}$)

$$\left.H_{\mathbf{w}}\right]_{i_1 i_2} = \frac{\partial^2 C(\mathbf{w})}{\partial w_{i_1} \partial w_{i_2}}$$

ser $\mathbf{w}_0$ un mínimo (local), el segundo término de la derecha es nulo;
poniendo $H_{\mathbf{w}}(\mathbf{w}_0)$ en sus vectores propios

$$H_{\mathbf{w}}(\mathbf{w}_0)\, \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

biendo

$$\mathbf{w} - \mathbf{w}_0 = \sum_i \alpha_i \mathbf{v}_i$$

e

$$C(\mathbf{w}) = C(\mathbf{w}_0) + \frac{1}{2}\left(\sum_i \alpha_i \mathbf{v}_i\right)^T H_{\mathbf{w}}(\mathbf{w}_0)\left(\sum_i \alpha_i \mathbf{v}_i\right) = C(\mathbf{w}_0) + \frac{1}{2}\sum_i \lambda_i \alpha_i^2$$

$\mathbf{w}_0$ es un mínimo si $H_{\mathbf{w}}(\mathbf{w}_0)$ es definida positiva: $\lambda_i > 0, \ \forall$ i ando ahora gradientes sobre

$$C(\mathbf{w}) = C(\mathbf{w}_0) + \frac{1}{2}(\mathbf{w} - \mathbf{w}_0)^T H_{\mathbf{w}}(\mathbf{w}_0)(\mathbf{w} - \mathbf{w}_0)$$

$$\mathbf{g}(\mathbf{w}) = H_{\mathbf{w}}(\mathbf{w}_0)(\mathbf{w} - \mathbf{w}_0) = \sum_i \lambda_i \alpha_i \mathbf{v}_i$$

que el algoritmo $\left(\mathbf{w}^{nuevo} = \mathbf{w}^{ant} - \eta \mathbf{g}\left(\mathbf{w}^{ant}\right)\right)$ resulta

$$\sum_i \alpha_i^{nueva} \mathbf{v}_i = \sum_i \alpha_i^{ant} \mathbf{v}_i - \eta \sum_i \lambda_i \alpha_i^{ant} \mathbf{v}_i$$

do los vectores propios linealmente independientes,

$$\alpha_i^{nueva} = (1 - \eta \lambda_i) \alpha_i^{ant}$$

para la correspondiente $\mathbf{v}_i$, la búsqueda empieza en un cierto punto $\alpha_i^{(0)}$ una progresión geométrica de razón $1\text{-}\eta\lambda_i$; que converge si

$$|1 - \eta \lambda_i| < 1 : \qquad \eta < 2/\lambda_i \quad (< 2/\lambda_{max})$$

Nótese que la convergencia de la sucesión de $\mathbf{v}_i$ es lo más rápida posible $\eta=1/\lambda_i$. Sin embargo, ésta puede ser una mala elección para otras nentes.

*Ejercicio:*

*Definiendo la velocidad de convergencia como*

$$v_\eta = \min_i \left| \frac{1}{1-\eta\lambda_i} \right|$$

*stre que el valor de $\eta$ que maximiza $v_\eta$ es*

$$\eta_{opt} = \arg\left\{ \max_\eta \left\{ v_\eta \right\} \right\} = \frac{2}{\lambda_{min} + \lambda_{max}}$$

La velocidad no es el único factor a tener en cuenta en la elección del n general, cuanto mayor sea η, mayor será el **error de desajuste** debido esado muestra a muestra.

caso de coste cuadrático y estimador lineal:

$$C(\mathbf{w}) = \frac{1}{2} E\left\{ \left( s - \mathbf{w}^T \mathbf{x} \right)^2 \right\}$$

$$\frac{\partial C}{\partial w_i} = -E\left\{ \left( s - \mathbf{w}^T \mathbf{x} \right) x_i \right\}$$

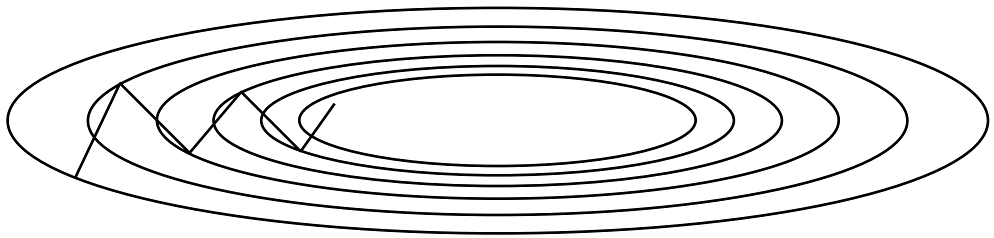
$$\frac{\partial^2 C}{\partial w_i \partial w_j} = E\left\{ x_i x_j \right\}$$

iano es la matriz de autocorrelación $R_{xx}$; y el algoritmo secuencial de te

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \eta \left( s^{(k)} - \mathbf{w}^{(k)T} \mathbf{x}^{(k)} \right) \mathbf{x}^{(k)}$$

conoce como **LMS** ("Least Mean Squares") o de **Widrow-Hoff**: muy temente aplicado para tiempo discreto k, y del que hay que destacar su z y su notable capacidad de seguimiento en casos no estacionarios.

La principal debilidad de estos algoritmos radica en su ineficiencia H tiene autovalores muy dispersos (recuérdese la discusión de gencia): lo que ocurre entonces es que la superficie del coste en el del mínimo tiene aspecto fuertemente hiperelíptico, y los sucesivos tes implican una ineficaz búsqueda en “zig-zag”.



*mas de evitar este inconveniente de los algoritmos de gradiente.*

Hay muchas variantes de estos algoritmos para evitar (otras) iones: p. ej., para acelerar la búsqueda cuando se atraviesa una zona e C(**w**)

ede aplicarse el **algoritmo del momento**

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \eta \mathbf{g}^{(k)}\left(\mathbf{w}^{(k)}\right) + \beta\left[\mathbf{w}^{(k)} - \mathbf{w}^{(k-1)}\right]$$

que el último término proporciona "inercia" a la actualización de los rámetros

ede realizarse gestión del escalón: aumentándolo suavemente cuando el or va decreciendo, y reduciéndolo rápidamente si el error crece

*tudiar y aplicar algoritmos de gestión del escalón.*

También es popular el **algoritmo NLMS** ("Normalized LMS"), en el que
a la dependencia del tamaño de las muestras utilizando como escalón

$\mathbf{x}^{(k)}\|_2^2\Big)$ (ε evita problemas numéricos)

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \frac{1}{\varepsilon + \left\|\mathbf{x}^{(k)}\right\|_2^2}\left(s^{(k)} - \mathbf{w}^{(k)T}\mathbf{x}^{(k)}\right)\ \mathbf{x}^{(k)}$$

que se satisface "naturalmente" la condición de convergencia, ya que
$> \lambda_{max}$, y $\left\|\mathbf{x}^{(k)}\right\|_2^2$ es la traza muestral).

es de otras búsquedas locales

queda en línea

e una dirección de búsqueda $\mathbf{d}^{(k)}$, y se aplica

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \eta^{(k)}\mathbf{d}^{(k)}$$

do $\eta^{(k)}$ para la máxima reducción del coste

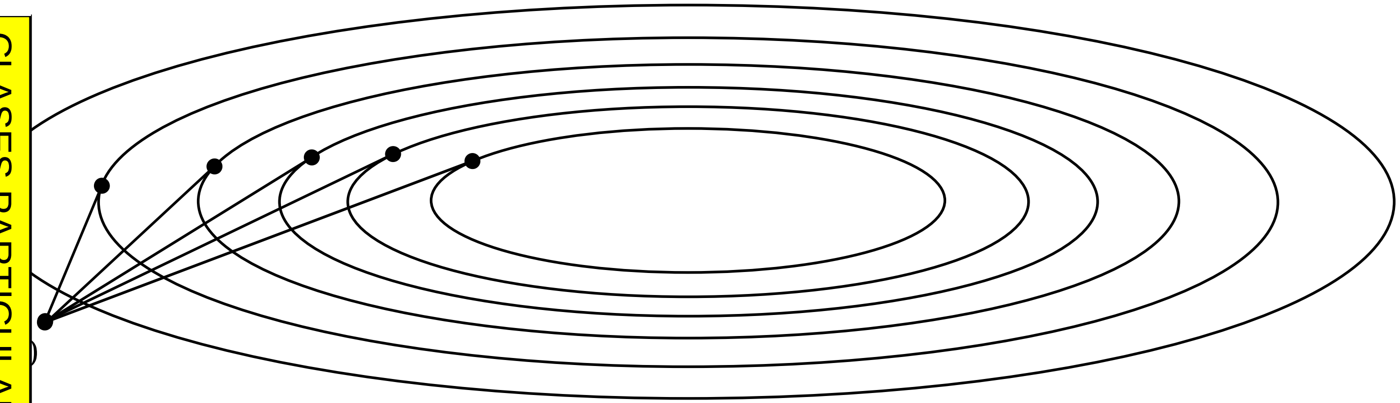
$$\frac{\partial C\left(\mathbf{w}^{(k+1)}\right)}{\partial \eta^{(k)}} = \left[\frac{\partial C\left(\mathbf{w}^{(k+1)}\right)}{\partial \mathbf{w}^{(k+1)}}\right]^{T} \frac{\partial \mathbf{w}^{(k+1)}}{\partial \eta^{(k)}} = \mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\frac{\partial\left(\mathbf{w}^{(k)} + \eta^{(k)}\mathbf{d}^{(k)}\right)}{\partial \eta^{(k)}} =$$

$$= \mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\mathbf{d}^{(k)} = 0$$

ndose así los **algoritmos de descenso más rápido**.

En la práctica, $\eta^{(k)}$ se determina minimizando $C(\mathbf{w}^{(k)} + \eta^{(k)}\mathbf{d}^{(k)})$ por da directa unidimensional.

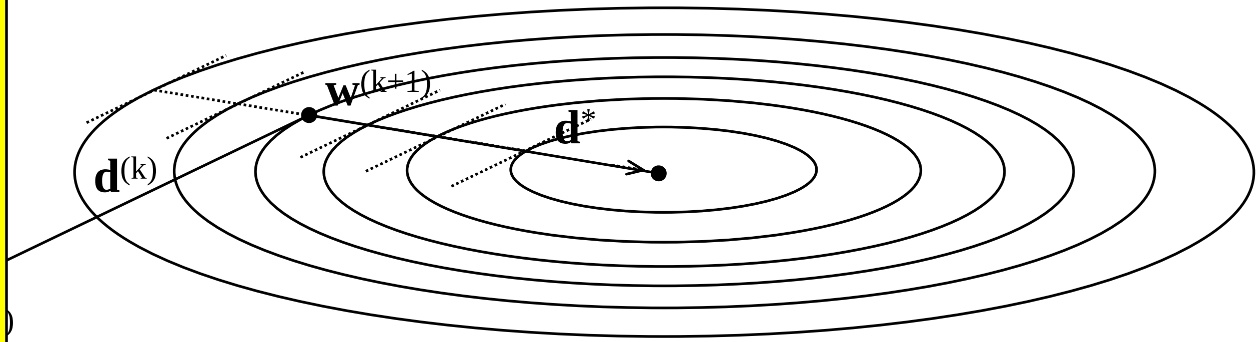
a condición $\mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\mathbf{d}^{(k)} = 0$ implica que, en el punto de destino, la ón de búsqueda es siempre tangente a la curva de nivel:



Cómo determinar la dirección de búsqueda?

bsérvese cómo al elegir $\mathbf{d}^{(k)} = -\,\mathbf{g}(\mathbf{w}^{(k)})$, los gradientes consecutivos son nales, lo que suele llevar a búsqueda en zig-zag.

Sin embargo, nótese también que, admitiendo forma del coste de 2º en cualquier punto situado en la dirección óptima, **d***, el gradiente es la dirección de búsqueda anterior.




Lo anterior puede imponerse como condición: elíjase $\mathbf{d}^{(k+1)}$ de tal modo

$$\mathbf{g}^{T}\left(\mathbf{w}^{(k+2)}\right)\mathbf{d}^{(k)}=0$$

el fundamento de los algoritmos de gradiente conjugado.

## ɔritmos de gradiente conjugado

ı búsquedas en línea (óptimas) en que se eligen las $\mathbf{d}^{(k)}$ de modo que no urbe lo progresado en pasos anteriores: concretamente, obligando en aso a que el gradiente en el destino sea ortogonal a la dirección de da previa

$$^{T}\left(\mathbf{w}^{(k+2)}\right)\mathbf{d}^{(k)} = \mathbf{g}^{T}\left(\mathbf{w}^{(k+1)} + \eta^{(k+1)}\mathbf{d}^{(k+1)}\right)\mathbf{d}^{(k)} = 0$$

de, admitiendo forma de 2º orden:

$$\left(\mathbf{w}^{(k+2)}\right) = \mathrm{H}_{\mathbf{w}}\left(\mathbf{w}_0\right)\left(\mathbf{w}^{(k+2)} - \mathbf{w}_0\right) =$$

$$= \mathrm{H}_{\mathbf{w}}\left(\mathbf{w}_0\right)\left(\mathbf{w}^{(k+1)} + \eta^{(k+1)}\mathbf{d}^{(k+1)} - \mathbf{w}_0\right) =$$

$$= \mathrm{H}_{\mathbf{w}}\left(\mathbf{w}_0\right)\left(\mathbf{w}^{(k+1)} - \mathbf{w}_0\right) + \eta^{(k+1)}\mathrm{H}_{\mathbf{w}}\left(\mathbf{w}_0\right)\mathbf{d}^{(k+1)} =$$

$$= \mathbf{g}\left(\mathbf{w}^{(k+1)}\right) + \eta^{(k+1)}\mathrm{H}_{\mathbf{w}}\left(\mathbf{w}_0\right)\mathbf{d}^{(k+1)}$$

que la condición queda

$$\mathbf{d}^{(k)^{T}}\mathbf{g}\left(\mathbf{w}^{(k+2)}\right) = \mathbf{d}^{(k)^{T}}\mathbf{g}\left(\mathbf{w}^{(k+1)}\right) + \eta^{(k+1)}\mathbf{d}^{(k)^{T}}\mathrm{H}_{\mathbf{w}}\left(\mathbf{w}_0\right)\mathbf{d}^{(k+1)} = 0$$

o 0 el primer sumando del segundo miembro,

$$\mathbf{d}^{(k)^T} H_{\mathbf{w}}(\mathbf{w}_0)\,\mathbf{d}^{(k+1)} = 0$$

la condición de conjugación. Si se aplica consecutivamente en un de dimensión P (nº de parámetros) bajo la hipótesis de superficie de 2º da lugar a P direcciones conjugadas que permiten resolver el problema sos (no es así si la superficie no es de 2º orden).

Las formas prácticas de estos algoritmos se obtienen de aplicar la ón de conjugación a

$$\mathbf{d}^{(k+1)} = -\mathbf{g}\left(\mathbf{w}^{(k+1)}\right) + \beta^{(k)}\mathbf{d}^{(k)}$$

yendo en $\mathbf{d}^{(k+1)^T} H_{\mathbf{w}}(\mathbf{w}_0)\,\mathbf{d}^{(k)} = 0$ , se puede despejar

$$\beta^{(k)} = \frac{\mathbf{g}\left(\mathbf{w}^{(k+1)}\right) H_{\mathbf{w}}(\mathbf{w}_0)\,\mathbf{d}^{(k)}}{\mathbf{d}^{(k)^T} H_{\mathbf{w}}(\mathbf{w}_0)\,\mathbf{d}^{(k)}}$$

Obviamente, $\beta^{(k)}$ no se puede obtener de esta expresión teórica, ya que no oce $H_{\mathbf{w}}(\mathbf{w}_0)$:

se sustituye $H_{\mathbf{w}}(\mathbf{w}_0)\mathbf{d}^{(k)}$ por $[\mathbf{g}(\mathbf{w}^{(k+1)}) - \mathbf{g}(\mathbf{w}^{(k)})]/\eta^{(k)}$ ualdad que se obtiene de la expresión de $\mathbf{g}(\mathbf{w}^{(k+2)})$ en la p. 9.18, mbiando k+2 por k+1), se llega al **algoritmo de Hestenes-Stiefel**

$$\beta_{HS}^{(k)} = \frac{\mathbf{g}^{\mathbf{T}}\left(\mathbf{w}^{(k+1)}\right)\left[\mathbf{g}\left(\mathbf{w}^{(k+1)}\right) - \mathbf{g}\left(\mathbf{w}^{(k)}\right)\right]}{\mathbf{d}^{(k)^{T}}\left[\mathbf{g}\left(\mathbf{w}^{(k+1)}\right) - \mathbf{g}\left(\mathbf{w}^{(k)}\right)\right]}$$

ahora se considera la expresión de $\mathbf{d}^{(k+1)}$ en la p. 9.19, se traspone y se ltiplican por $\mathbf{g}(\mathbf{w}^{(k+1)})$ ambos términos, se tiene:

$$\mathbf{d}^{(k+1)^{T}}\mathbf{g}\left(\mathbf{w}^{(k+1)}\right) = -\mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\mathbf{g}\left(\mathbf{w}^{(k+1)}\right) + \beta^{(k)}\mathbf{d}^{(k)^{T}}\mathbf{g}\left(\mathbf{w}^{(k+1)}\right)$$

como el segundo sumando del término de la derecha es nulo (por scenso más rápido), se deduce cambiando k+1 por k:

$$\mathbf{d}^{(k)^{T}}\mathbf{g}\left(\mathbf{w}^{(k)}\right) = -\mathbf{g}^{T}\left(\mathbf{w}^{(k)}\right)\mathbf{g}\left(\mathbf{w}^{(k)}\right)$$

e, llevada al denominador de $\beta_{HS}^{(k)}$, considerando además que )$^{T}$**g(w**$^{(k+1)}$)=0, proporciona el **algoritmo de Polak-Ribiere**

$$\beta_{PL}^{(k)} = \frac{\mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\left[\mathbf{g}\left(\mathbf{w}^{(k+1)}\right)-\mathbf{g}\left(\mathbf{w}^{(k)}\right)\right]}{\mathbf{g}^{T}\left(\mathbf{w}^{(k)}\right)\mathbf{g}\left(\mathbf{w}^{(k)}\right)}$$

r último, si de la condición de descenso más rápido de la p. 9.15

$$\mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\mathbf{d}^{(k)} = 0$$

inserta la expresión de **d**$^{(k)}$ de la p. 9.19, se tiene

$$\mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\left[-\mathbf{g}\left(\mathbf{w}^{(k)}\right)+\beta^{(k-1)}\mathbf{d}^{(k-1)}\right] = 0$$

donde

$$-\mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\mathbf{g}\left(\mathbf{w}^{(k)}\right)+\beta^{(k-1)}\mathbf{g}^{T}\left(\mathbf{w}^{(k+1)}\right)\mathbf{d}^{(k-1)} = 0$$

n lo que, como es nulo el segundo sumando del primer término, lo es el mero: eliminándolo en el numerador de $\beta_{PL}^{(k)}$, queda establecido el **goritmo de Fletcher-Reeves:**

$$\beta_{FR}^{(k)} = \frac{\mathbf{g}^T\left(\mathbf{w}^{(k+1)}\right)\mathbf{g}\left(\mathbf{w}^{(k+1)}\right)}{\mathbf{g}^T\left(\mathbf{w}^{(k)}\right)\mathbf{g}\left(\mathbf{w}^{(k)}\right)}$$

tos algoritmos

n eficaces si la aproximación cuadrática es razonable: si no, pueden ser ores que uno de gradiente;

ben "reinicializarse" (mediante gradiente) cada cierto número de pasos, ra evitar los efectos de acumulación de errores.

## El Método de Newton

upuesto que nos encontramos en **w*** y que es aceptable

$$C(\mathbf{w}) = C(\mathbf{w}^*) + (\mathbf{w} - \mathbf{w}^*)^T \mathbf{g}(\mathbf{w}^*) + \frac{1}{2}(\mathbf{w} - \mathbf{w}^*)^T H_{\mathbf{w}}(\mathbf{w}^*)(\mathbf{w} - \mathbf{w}^*)$$

lo gradientes

$$\mathbf{g}(\mathbf{w}) = \mathbf{g}(\mathbf{w}^*) + H_{\mathbf{w}}(\mathbf{w}^*)(\mathbf{w} - \mathbf{w}^*)$$

á nulo si **w** es el mínimo buscado; de donde

$$\mathbf{w} = \mathbf{w}^* - H_{\mathbf{w}}^{-1}(\mathbf{w}^*)\mathbf{g}(\mathbf{w}^*)$$

licación de forma iterada

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - H_{\mathbf{w}}^{-1}(\mathbf{w}^{(k)})\mathbf{g}(\mathbf{w}^{(k)})$$

goritmo correspondiente a este método.

lo se aplica directamente por la carga que supone

cular $H_w$ en cada paso

vertir la matriz

u lugar se utilizan

métodos Seudo-Newton: reduciendo $H_w$ a sus términos diagonales ciendo positivos los negativos; añadiendo una cte. ε para evitar stabilidades numéricas)

métodos Quasi-Newton: basados en aproximar iterativamente $H_w^{-1}$ poniendo la "condición Quasi-Newton"

$$\mathbf{w}^{(k+1)} - \mathbf{w}^{(k)} = H_w^{-1}\left(\mathbf{w}^{(k)}\right)\left[\mathbf{g}\left(\mathbf{w}^{(k+1)}\right) - \mathbf{g}\left(\mathbf{w}^{(k)}\right)\right]$$

*scutir los métodos Q-N*

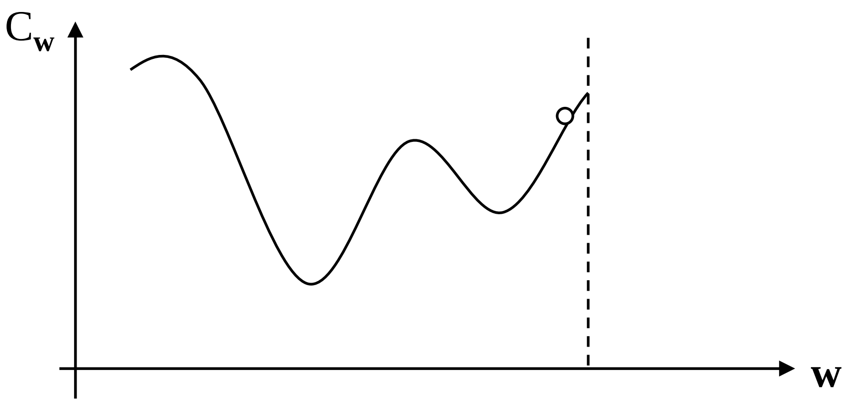
*de Davidson-Fletcher-Powell (DFP)*

*de Broyden-Fletcher-Goldfarb-Shanno (BFGS)* eda Global

**Temple Simulado** ("Simulated Annealing")

Como se procede con el temple del acero: enfriando lentamente para lcanzar un estado de mínima energía interna)

uye con el símil del iento de una bola por perficie de coste: con gitación fuerte recorre superficie, y si se va uyendo poco a poco á a quedarse en el o absoluto.



**\prendizaje de Boltzmann**

cializar **w**

cializar la temperatura T a $T_0$ (valor alto)

licar un observable al azar y obtener C

egir un parámetro al azar y modificarlo (según $G(0, T^2/2\pi)$, p. ej.) teniendo $C+\Delta C$

i $\Delta C < 0$ : aceptar el cambio

i $\Delta C > 0$ y $\exp(-\Delta C/T) > U[0,1]$ : aceptar el cambio

i $\Delta C > 0$ y $\exp(-\Delta C/T) < U[0,1]$ : no aceptarlo

lver a 4 hasta agotar los parámetros

lver a 3 hasta agotar los observables

licar un criterio de parada:

i se cumple, terminar

i no, reducir la temperatura según $T(l)=T_0/\ln(1+l)$, y volver a 3

*étodos de aceleración del Aprendizaje de Boltzmann*

**oritmos Genéticos**

Propuestos por Holland a mediados de los 70, son búsquedas Neo-ıianas, con exploración por **cruce** y **mutación** sobre poblaciones de ıtos a la solución, y explotación mediante selección para **supervivencia**.

**tmo Genético Básico** (AGB)

ımponentes de la solución (**w**) se expresan en <u>binario</u>)

<u>zación</u>:

nstruir una población inicial aleatoria de tiras binarias

<u>aluación</u>: aplicar una medida de calidad: <u>ajuste</u> (“fitness”) a un objetivo

tención: aplicar un criterio de parada; si no se cumple

<u>producción</u>: generar una nueva población, sorteando el paso de los lividuos de acuerdo con su calidad

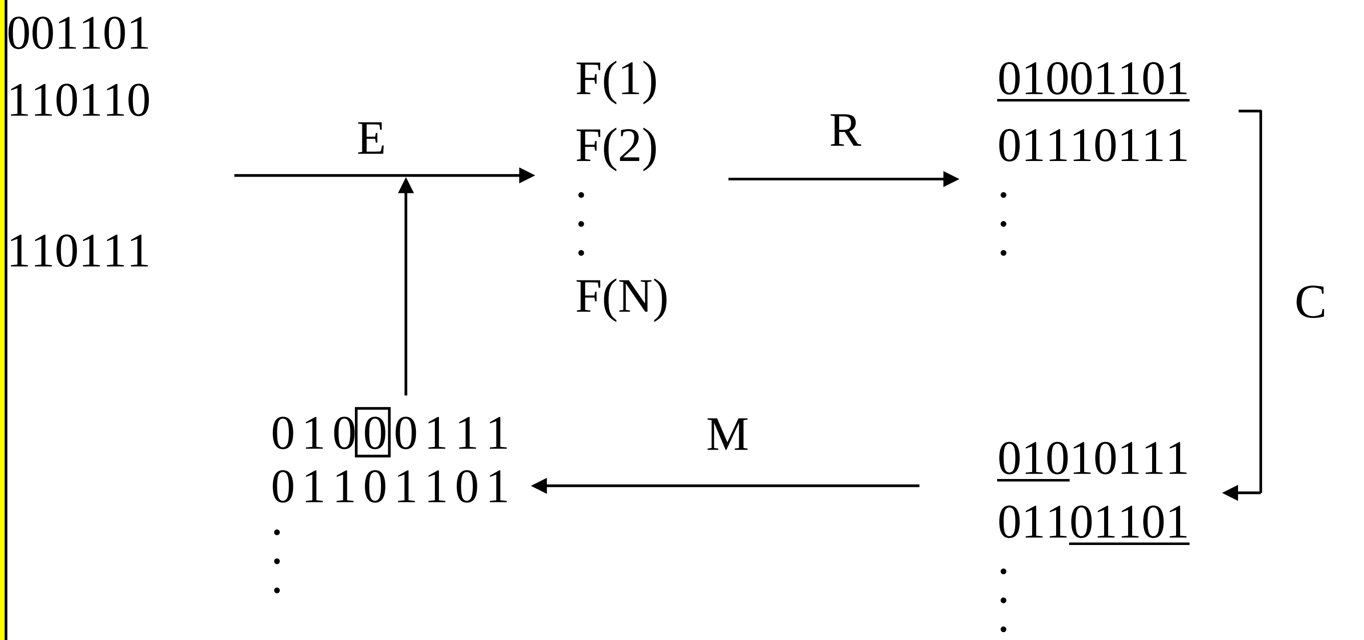
<u>uce</u>: tomar aleatoriamente dos individuos, cortarlos por un (mismo) nto aleatorio, y cruzar los trozos

<u>tación</u>: por sorteo (de baja probabilidad) para cada bit de la población

ılver a 2

oducción, cruce, mutación: <u>operadores</u>)

tración


001101
110110
110111
E
F(1)
F(2)
F(N)
R
01001101
01110111
C
M
0 1 0 0 0 1 1 1
0 1 1 0 1 1 0 1
01010111
01101101

Formatos típicos:

maño de la población: alrededor de 50 individuos

sa de cruce: 0.6 - 0.8

sa de mutación: $10^{-2}$ - $10^{-3}$

ngitud de los individuos: hasta varios cientos

Con estas características, el algoritmo busca una solución global, sando (con oscilaciones) hacia mejores valores de calidad tanto del mejor uo como en media para la población.

Tienen gran potencia de búsqueda.

erísticas de los AG

ciones

gos: no es fácil incorporar información previa

cil introducir restricciones sobre las soluciones: se suele hacer aplicando ores que las mantengan

gran carga computacional, debido a la evaluación

icación directa no ofrece generalización, al calcular directamente la

oclives al estancamiento

r pérdida de variedad “genética” (en los bits homólogos de todos los lividuos), que imposibilita la exploración;

e puede corregir: por aumento de la mutación o por renovación de parte de la población r aparición de un <u>superindividuo</u> (muy bueno para su generación, pero como solución final): que tiende a ocupar toda la población por su ntaja en la reproducción;

combate: equilibrando la reproducción; p. ej., procediendo según rango la evaluación, y no de modo directamente proporcional a la calidad.

n este tipo de precaución, se pueden aplicar estrategias <u>elitistas</u>, pasando mejor individuo de cada generación directamente a la siguiente, para tar pérdidas de difícil reparación)

...en capacidad de seguimiento.

...ueden combinarse con búsquedas locales (que, además, serán vía para ...r generalización), según principios

...marckistas: se heredan los resultados;

...ldwinianas: no se hereda la búsqueda local, pero se emplea en la evaluación.

...B es de aplicación limitada por la codificación binaria: la codificación ...requiere otros operadores para cruce (incluyendo fragmentación de cada ...n) y mutación.

as

dicha potencia de búsqueda)

quieren formulación analítica: la evaluación puede llevarse a cabo te la aplicación directa de las soluciones al problema (p. ej., haciendo s estrategias que se diseñen para un juego)

introducirse muchos otros operadores de búsqueda "ad hoc" para el na; así como codificaciones diferentes de las tiras de bits

*...os métodos de búsqueda global son:*

- *...s estrategias evolutivas*
- *...s búsqueda "Tabú"*
- *...s algoritmos "branch and bound"*

*...escríbase su aplicación y discútanse sus principales características.*