Compresión de datos y restauración de imágenes con ruido según David Donoho

(Data compression and restoration of images with noise according to David Donoho)

Meyer, Yves
Eusko Ikaskuntza
Mª Díaz de Haro, 11-1º
48013 Bilbao

BIBLID [1137-4411 (1997), 4; 199-210]

Las ondículas ofrecen interesantes posibilidades en el tratamiento eliminación del ruido de señales e imágenes y forman una base incondicional universal apta para todos los espacios de Besov, los cuales proveen una descripción detallada y precisa de cientras clases de señales e imágenes permitiendo evaluar la calidad de la imagen reconstruida tras la compresión no lineal. Los algoritmos de compresión óptima están en el corazón de los problemas de eliminación de ruido no lineal. Se presenta una descripción de los descubrimientos de Donoho sobre imágenes de tipo geométrico. El reto consiste en encontrar una base ortonomal privilegiada del espacio de señales de energía finita.

Palabras Clave: Ondículas. Ruido de señales. Imágenes geométricas. Espacios de Besov.

Uhin txikiek ahalbide interesgarriak eskaintzen dituzte seinale eta irudien zarata ezabatzeko tratamendurako eta kondiziogabeko oinarri unibertsal egokia osatzen dute Besov-en espazio guztietarako. Espazio horiek seinale eta irudi jakin batzuen deskripzio zehatza eta egokia ematen dute eta, konpresio ez-linialaren ondoren, irudia berregitea bideratzen dute. Konpresio optimorako algoritmoak zarata ez-linealaren ezabaketaren muinean daude. Donoho-k irudi geometrikoen inguruan egindako aurkikuntzen deskripzioa ematen da. Energia finituko espazioaren oinarri ortonomal pribilejiatu bat aurkitzea da oraingo erronka.

Giltz-Hitzak: Uhineskak. Zarata. Irudi geometrikoak. Besov-en espazioa.

Les ondicules offrent d'intéressantes possibilités dans le traitement d'élimination du bruit de signaux et images et forment une base inconditionnelle universelle apte à tous les espaces de Besov, lesquels fournissent une description détaillée et précise de certaines sortes de signaux et images permettant d'évaluer la qualité de l'image reconstituée après la compression non linéaire. Les algorithmes de compression optimum sont au coeur des problèmes d'élimination du bruit non linéaire. On présente une description des découvertes de Donoho sur des images de type géométrique. Le pari consiste à trouver une base ortonomale privilégiée de l'espace de signaux d'énergie finie.

Mots Clés: Ondicules. Bruit de signaux. Images géometriques. Espaces de Besov.

1. INTRODUCCION

David Donoho, especialista del tratamiento estadlstico de la señal, fue conducido a interesarse por las posibilidades que ofrecıan las *ondículas** para *eliminar el ruido de señales e imágenes* con una estructura geométrica particular.

Donoho no conocía las ondículas y necesitaba con urgencia una herramienta muy particular. Con ocasión de una escuela de verano de Estadística en St. Flour, uno de sus colegas, Dominique Picard, le hizo saber que esa herramienta acababa precisamente de ser creada y que los detalles se encontraban en mi libro aparecido en la editorial Hermann.

Donoho pasó el resto del verano leyendo esta obra y se interesó particularmente por el hecho de que las ondículas forman una *base incondicional* universal apta para todos los espacios de Besov (sección 6). Por otra parte, Donoho sabía, que gracias a las diferentes regulaciones permitidas por los tres índices que intervienen en su definición, los espacios de Besov proveen una descripción delicada y precisa de ciertas clases de señales y de imágenes (sección 5). Los espacios de Besov permiten evaluar la calidad de la imagen reconstruida tras la *compresión no lineal* (sección 9).

Donoho demostró que la *eliminación óptima de ruido* por una algoritmo no lineal necesita una *compresión no lineal óptima* de las señales e imágenes implicadas. Dedujo de ello que el umbral dulce de los coeficientes de *ondículas* (“soft tresholding”) proporciona esta eliminación óptima de ruido para todas las señales e imágenes para las que el conocimiento a priori se modeliza por el orden de magnitud de una cierta norma de Besov (sección 10).

Los algoritmos de compresión óptima están, por tanto, en el corazón de los problemas de eliminación de ruido no lineal y están emparentados con otros algoritmos de compresión de imágenes, como el algoritmo de Oslo o “knot removal” (sección 9).

Los trabajos de Donoho, de los que muchos están escritos en colaboración con I. Johnstone, permiten finalmente dar una formulación científica precisa a conclusiones empíricas aparecidas en el seno de la “comunidad de ondiculistas”, pero que no habían recibido aún una justificación seria.

Los métodos utilizados por A. Lannes y S. Roques para eliminar el ruido de las imágenes del telescopio espacial Hubble estarán emparentados con los trabajos de Donoho, teniendo siempre en cuenta la especifidad del problema considerado.

2. BORDES Y TEXTURAS

Presentamos, en el marco del tratamiento de la imagen, una primera descripción de los descubrimientos de Donoho. Este trabajo corresponde a *imágenes de tipo geométrico.* Veamos de lo que se trata. Una imagen real, como la de la sala en la que enseño actualmente mi curso sobre las ondículas, se compone aproximadamente de formas geométricas, limitadas por contornos bastante precisos. Estas formas geométricas están “llenas” sea por fluctuaciones muy regulares de la intensidad luminosa, sea por variaciones rápidas de esta intensidad, llamadas texturas. Por ejemplo, algunos estudiantes llevan jerseys y el examen atento de es-

* Utilizamos esta palabra como equivalente del francés *ondelette* y del inglés wavelet.

tos jerseys revela motivos periódicos o casi periódicos precisos, de alta frecuencia espacial, pero de intensidad débil. Es decir, las fluctuaciones correspondientes de la iluminación son muy rápidas pero permanecen débiles si se comparan con las variaciones mucho más brutales de esta misma iluminación de la que son responsables los bordes de las siluetas de nuestros estudiantes. Si se pidiese a un dibujante hábil hacer un croquis de esta aula, las líneas de los contornos serían muy netas, mientras que las texturas serían reproducidas con mucha menos fidelidad. Estas texturas así como las creadas por los cabellos serán más evocadas que dibujadas con cuidado. Es de este modo como procede un pintor y A. Durero era célebre por simular, de una sola pincelada, una cabellera de la que se podría creer que había sido dibujada cabello a cabello.

Esto conducirá a la idea de simular automáticamente texturas naturales por técnicas algorítmicas, imitando el pincel de Durero. En el momento actual, se usan versiones bilineales del browniano fraccionario para simular texturas naturales. Estas simulaciones se efectúan con ayuda de la represetación del browniano fraccionario en serie de ondículas apropiadas. Los detalles aparecerán en la tesis de Fabrice Sellan (Matra, Les Miroirs).

Conviene volver a los contornos e imitar a Ingres, en su visión pictórica. Es ahí donde interviene el algoritmo de Donoho, que se puede presentar como un algoritmo de contornos. Se puede decir que la reflexión de Donoho prolonga y completa la que fue iniciada por D. Marr. Está claro que el paradigma según el cual contornos y texturas serían las únicas componentes de una imagen es una simplificación. Pero es una simplificación muy útil, que se adapta a numerosas situaciones reales.

Consideremos ahora el caso de una *imagen geométrica con ruido.* La hipótesis de trabajo fundamental es que, en este caso, ruido y texturas serán indistinguibles y que el algoritmo debe extraer el dibujo, despreciando las texturas y el ruido. El algoritmo de Donoho podría ser comparado al trabajo paciente y cuidadoso del arqueólogo que a partir de fragmentos de cerámica, estropeados y rotos, reconstruye reflexivamente los trozos que faltan y deduce los hábitos alimenticios de una civilización. Estas informaciones están ausentes para el profano y no están presentes más que para el especialista que dispone de un *conocimiento a priori* sobre lo que busca. Los restos de cerámica permiten entonces elegir una pista entre varias, en un universo de posibilidades bastante restringido, en todo caso.

3. MODELIZACION POR ESPACIOS FUNCIONALES

Pero, ¿cuál es el conocimiento a priori de que disponemos sobre la imagen? Yo pensaba, siguiendo a David Marr, que este conocimiento sería una especie de transcripción simbólica (en forma de ideogramas), en el lenguaje de las líneas, de las leyes que rigen la organización geométrica de los objetos en el espacio tridimensional. Es el punto de vista adoptado por David Mumford quien, tras brillantes trabajos en Geometría algebraica, se ha convertido en uno de los especialistas del tratamiento de la imagen. Mumford trata de codificar eficazmente las reglas de incidencia de las líneas (o de los bordes) que testimonian la disposición relativa de unos objetos con respecto a otros. El punto de vista adoptado por Donoho está muy alejado del favorecido por Mumford.

Para Donoho, que es estadístico y espacialista en el tratamiento de la señal, el reto consiste en encontrar una *base ortonormal privilegiada del espacio de señales de energía finita*.

Esta base ortonormal es privilegiada si permite presentar, de manera más contrastada, lo que se busca con respecto a lo que se está destinado a despreciar. Vamos a detenernos un momento sobre esta idea y recordar al lector en qué sentido la base de Karhunen-Loève estaba considerada como la base privilegiada en el marco del estudio estadístico de las señales.

4. LA BASE DE KARHUNEN-LOEVE

Vamos a recordar primero lo que es el algoritmo de Karhunen-Loeve (KL) que, erróneamente, se considera como el algoritmo óptimo para la compresión de datos. Supongamos que cada dato X contiene n informaciones y se puede, por tanto, codificar como un vector de R^n. Supongamos que disponemos de N de tales vectores, denotados X_1, X_2, ..., X_n. Estos N vectores constituyen una nube de puntos en R^n. Nuestro propósito es encontrar la representación más cómoda de estos N puntos. La hipótesis que sostiene esta cuestión es la de un orden escondido en el desorden aparente de la información bruta que nos proporcionan estos N puntos. La búsqueda de este orden escondido es, evidentemente, el fin fundamental de la actividad científica.

Se comienza por examinar si estos N puntos representan el equivalente a un pueblo muy concentrado (alrededor de su iglesia o de su ayuntamiento) o a un hábitat disperso. Para ello se determina el centro (de gravedad) E_o de la nube, que minimiza la suma de los cuadrados de las distancias a los puntos X_1, X_2, ..., X_n. Hecho esto, se determina a continuación el eje principal (o eje de inercia) de la nube; es la recta D tal que la suma de los cuadrados de las distancias de los puntos de la nube a D sea mínima. Esta recta afín, denotada E_1, contiene al centro de gravedad de la nube. Se considera el plano afín E_2 tal que la suma de los cuadrados de las distancias de los puntos de nuestra nube a E_2 sea mínima. Este plano E_2 contiene al eje de inercia precedente y se continúa hasta construir E_n.

La *base de Karhunen-Loève* se construye utilizando los subespacios encajados precedentes: e_1 es un vector unitario de E_1, (e_1, e_2) es una base ortonormal de E_2, etc.

El significado de esta base es muy claro cuando se dispone, por ejemplo, de un gran número de testimonios sobre un criminal. Se hace el “retrato robot” tratando de dar la razón a todo el mundo y se considera seguidamente un primer retoque posible (el más probable) que se podría añadir a este retrato medio, etc.

Veamos un ejemplo en el que el punto de vista de Karhunen-Loève está completamente inadaptado. Supongamos que $N = n$ y consideremos primero el caso en que los vectores X_1, X_2, ..., X_n constituyen una base ortonormal del espacio R^n. Entonces la elección de la base de Karhunen-Loève es completamente arbitraria. Pero, evidentemente, la base ortonormal más adecuada para representar de manera eficaz nuestro n vectores es precisamente la base compuesta por estos vectores. Si nuestros n vectores están muy próximos numéricamente a una base ortonormal $(e_1, e_2, ..., e_n)$, se pueden elegir para que la base de Karhunen-Loève esté muy alejada de $(e_1, e_2, ..., e_n)$, mientras que X_j se puede escribir en ciertos ejemplos con ayuda solamente de e_j y de e_{j+1}. Un ejemplo muy simple viene dado por $X_j = e_j + ee_{j+1}$ ($\varepsilon > 0$ es arbitrariamente pequeño) con la convención de que $e_{n+1} = e_1$. La base ortonormal provista por el algoritmo KL conduce a $f_1 = n^{-1/2}(e_1 + ... + e_n)$ que está lejos de cada uno de los vectores considerados y que “agradando a todos, se ha alejado de cada un o”.

Esta patología viene de que el algoritmo de Karhunen-Loève es inestable cuando los vectores dados están próximos a una base ortonormal. De modo que el punto de vista de Karhunen-Loève consiste en preferir una base ortonormal "óptima" en la que todos los datos están mal representados a una base ortonormal casi óptima en la que cada vector posee una representación muy simple, que depende del vector considerado. En el ejemplo que hemos dado, está claro que la base de partida (e_1, e_2, ..., e_n) es mucho mejor que la base KL.

Veamos una ilustración del mismo fenómeno, escrita en el lenguaje del tratamiento de la señal. Se considera un proceso en el que el azar viene dado por la elección aleatoria de un punto denotado ω cuya ley de distribución está equidistribuida en (0,1). Se pone entonces $X(\omega, t) = t$ si $0 \leq t \leq \omega$ mientras que $X(\omega, t) = t - 1$ si $\omega \leq t < 1$. Es fácil calcular la base de Karhunen-Loève en este caso y se obtiene la sucesión sin (πkt), k = 1,2, ... De hecho, se vuelve a encontrar aquí la paradoja de que tratando de agradar a todo el mundo, no se agrada a nadie. Ninguna de las realizaciones $X(\omega, t)$ está bien representada en la base KL. Si se quiere aproximar $X(\omega, t)$ en norma L^2 por una suma parcial $c_1 \sin(\pi t) + \ldots + c_N \sin(\pi Nt)$ con un error que no sobrepase ε, es necesario tomar en cuenta $N = \varepsilon^{-2}$ términos de la serie de Karhunen-Loève y todas las realizaciones del proceso están tratadas igualmente mal. Si, por el contrario, se toman las ondículas de Ingrid Daubechies de soporte de longitud igual a tres, con dos momentos nulos, entonces cada realización no utiliza más que unos pocos vectores de la base, para obtener una representación de la misma calidad que la precedente. De hecho $N = C \log(1/\varepsilon)$ vectores bastan, pero este ramillete de vectores depende ahora de ω. Se vuelve a encontrar exactamente, en este caso continuo, la situación presentada en el caso discreto: el algoritmo de Karhunen-Loève nos lleva, bajo forma de compromiso, a hacer una elección mala para todos, buscando minimizar la suma de los cuadrados de los descontentos. Vale más aumentar un poco el descontento general y que cada realización del proceso está bien representada en la base ortonormal escogida. *Esto es tanto más cierto cuanto que en ciertos casos (astronomía extragaláctica), no se dispone más que de una realización del proceso.*

En las situaciones de tratamiento de la imagen que vamos a considerar, aparece el mismo fenómeno: la elección de una base de KL es mala, comparada a una base subóptima que, para cada realización, ofrece una compresión muy buena.

Estas consideraciones tienen una importancia crucial cuando se busca una eliminación de ruido óptima. Si para representar una señal se necesita dispersar su energía en un número muy grande de vectores ortogonales, tal representación resulta muy sensible al ruido. En efecto, si el nivel de ruido es δ, entonces todos los coeficientes de la señal de módulo inferior a δ estarán ahogados en el ruido y ni se podrán distinguir ni podrán ser utilizados para la reconstrucción de la señal sin ruido. Pero si la convergencia es lenta, gran parte de la energía se perderá si se desprecian todos estos pequeños coeficientes.

D. Donoho demuestra en un trabajo notable *la equivalencia entre eliminación de ruido óptima* (1) *y compresión óptima* (2) en el marco hilbertiano de una sucesión (χ_n) de cuadrado sumable:

(1) la *restauración óptima* en norma l^2 de esta sucesión, cuando ha sido corrompida por un ruido no estocástico δz_n, se puede hacer con un error que no supera $C\delta\alpha$;

(2) el *reordenamiento decreciente* (χ_n^*) de la sucesión $(|\chi_n|)$ verifica $\chi_n^* \leq Cn^{-\beta}$ donde β está relacionado con α de un manera muy sencilla.

El ruido no estocástico se define simplemente por una sucesión arbitraria de números complejos z_n, cuyos módulos no son mayores que 1. Este descubrimiento de Donoho tiene una implicación fundamental: conservar, de una vez para siempre, el orden natural de la base ortonormal e_o, e_1, ..., e_n, ... conduce a colocar los coeficientes en ese mismo orden, lo que no permite capturar correctamente los problemas de eliminación óptima de ruido. Hay que estar dispuesto a efectuar cualquier permutación en la sucesión de vectores e_o, e_1, ..., e_n, ... lo que es de nuevo absolutamente contrario al espíritu del algoritmo de Karhunen-Loève. El problema de la eliminación de ruido estocástico nos conducirá a una formulación un poco más precisa de nuestras exigencias, que se dará en la seción 10.

Antes de pasar a este estudio del caso ruidoso, conviene hacer una nota aclaratoria esencial. En el tratamiento de la imagen las estadísticas pueden intervenir a dos niveles totalmente distintos. En primer lugar, algunos investigadores consideran a las imágenes como realizaciones de un proceso estocástico bidimensional; por ejemplo, estarán abocados a utilizar un browniano fraccionario bidimensional para modelizar texturas naturales. Por otro lado, la adquisición de imágenes naturales (imágenes spot, por ejemplo) suministra forzosamente imágenes ensuciadas por un ruido de diversos orígenes. Esta misma distinción vale para la señal de la palabra, una de cuyas simulaciones consiste en excitar un banco de filtros por una entrada que es un ruido blanco. Se trata pues de la modelización tradicional de los procesos estacionarios. Pero una vez producida, la señal de la palabra se puede encontrar ensuciada por un ruido externo, independiente del ruido blanco de excitación. En la discusión que sigue, el trabajo estadístico no trata de la modelización de imágenes naturales. El punto de vista adoptado por el estadístico D. Donoho es que las imágenes naturales no son realizaciones del proceso, sino que obedecen a reglas de coherencia de tipo geométrico que se relacionan con ciertas normas en espacios de Besov. El ruido será “externo” o sobreimpuesto y el problema de eliminación de ruido consiste en recuperar la imagen de partida con ayuda de un estimador óptimo.

5. BORDES, TEXTURAS Y ESPACIOS DE BESOV

Consideremos primero un caso escolar, a saber, el estudio de la regularidad de la función característica de un intervalo *(a,b)* que denotaremos $\theta(\chi)$. Bajo el punto de vista habitual, tal función no es regular, pero no es ni más ni menos regular que la función $\theta\omega(\chi)$ definida por el producto $\theta(\chi)\cos(\omega\chi)$. Pero bajo el punto de vista de la teoría de los espacios de Sobolev, esta función es más regular comparada con el producto $\theta(\chi)\cos(\omega\chi)$ cuanto más grande es ω. Si, por ejemplo, el índice s del espacio de Sobolev es menor que 1/2, la norma de Sobolev de $\theta(\chi)$ permanece finita mientras que la de $\theta(\chi)\cos(\omega\chi)$ tiende a infinito como ω^s. En cierto sentido, la norma de Sobolev sabe distinguir los bordes de las texturas. Pero este contraste entre bordes y texturas se acentúa aún más si se utilizan los espacios de Besov $B^{s,q}(L^p)$. En efecto, si $p = s = 1$ y si q es infinito, entonces $\theta(\chi)$ pertenece al espacio de Besov correspondiente, mientras que $\theta(\chi)\cos(\omega\chi)$ tiene una norma del orden de ω (con los espacios de Sobolev no se puede conseguir más que $\omega^{1/2-\varepsilon}$).

Esto nos conduce a un programa consistente en discriminar los bordes de las texturas (y del ruido) utilizando el criterio ofrecido por la norma de Besov y exigiendo que el resultado final tenga una norma de Besov "pequeña". Este es el punto de vista adoptado por Donoho y contrasta evidentemente con la idea de considerar una imagen como la realización de un proceso estocástico, pero también con ideas geométricas como las expresadas por David Mumford.

6. BASES INCONDICIONALES Y ELIMINACION OPTIMA DE RUIDO

Vayamos a la definición de una base incondicional en un espacio de Banach B. Consiste en una sucesión de vectores de B, $e_0, e_1, ..., e_n, ...$, tal que se satisfacen las tres propiedades siguientes:

(1) todo vector χ de B se escribe en forma de serie convergente $\chi = \alpha_0 e_0 + \alpha_1 e_1 + \ldots + \alpha_n e_n + \ldots$ (la convergencia ocurre en el sentido de la norma de B);

(2) los coeficientes $\alpha_0 \alpha_1, \ldots, \alpha_n, \ldots$ son únicos;

(3) si la serie $\alpha_0 e_0 + \alpha_1 e_1 + \ldots$ converge en B, lo mismo ocurre para todas las series $\beta_0 e_0 + \beta_1 e_1 + \ldots$ cuando $|\beta_0| \leq |\alpha_0|$, $|\beta_1| \leq |\alpha_1|, ...$

D. Donoho estudia una clase de problemas de eliminación de ruido que se pueden modelizar del modo siguiente: sea H un espacio de Hilbert y B un espacio de Banach contenido en H que admite una base incondicional $e_0, e_1, ..., e_n, ..$ que igualmente es una base hilbertiana en H. Sea Θ un conjunto acotado en B y supongamos que las señales o imágenes sin ruido que estudiamos pueden ser modelizadas por puntos de Θ. Sea $Y = X + \sigma Z$ una señal observada (Z es un ruido blanco normalizado). Propongámonos el problema de construir un estimador lineal o no lineal $\theta(Y)$ de X que, cuando σ tiende hacia 0, proporcione, en norma hilbertiana, un error que no sobrepase $C\sigma^{\alpha}$, donde el exponente α (que mide la calidad del estimador) es, en el peor de los casos, el mayor posible (hablaremos entonces de un estimador óptimo).

Entonces el *teorema de Donoho* es que la base $e_0, e_1, ...$ es óptima y que la estrategia óptima es diagonal. Consiste en reemplazar los coeficientes α_n por coeficien tes $\beta_n = \theta(\alpha_n, \sigma)$ donde la función θ se escoge hábilmente por un procedimiento que será revelado en la sección 10. Se conjuga este resultado general con el hecho de que las ondículas ortogonales satisfacen las condiciones del teorema de Donoho, cuando el espacio de Banach B es un espacio de Besov.

Pero antes de avanzar en la descripción de los resultados obtenidos por Donoho, detengámonos un momento en los algoritmos deterministas correspondientes, conocidos con el nombre de aproximación no lineal.

7. APROXIMACION RACIONAL

Históricamente la aproximación no lineal se ha desarrollado a partir de trabajos de matemáticos de los paises de "Europa del Este" (Bulgaria, Rusia, etc.) sobre la aproximación racional. Comencemos recordando de qué se trata. Sea $f(\chi)$ una función definida sobre un intervalo compacto I (cerrado y de longitud finita). Si, para fijar ideas, $f(\chi)$ pertenece a $L^{\infty}(I)$, buscamos entonces, para cada entero N, la fracción racional $g_N(\chi) = P(\chi)/Q(\chi)$ cuyo grado (definido como el máximo de los grados de los polinomios P y Q) no sobrepase N, que realice la mejor aproximación de $f(\chi)$, en el sentido de la norma $L^\infty(I)$. Se trata entonces, para cada entero N, de minimizar $\|f - g_N\|_\infty$, bajo las restricciones $g_N = P/Q$, grado (P) $\leq N$, grado (Q) $\leq N$. No se hace ninguna hipótesis sobre la posición de los polos de g_N. Resulta de aquí que el conjunto G_N de funciones racionales $g_N = P/Q$ que van a ser examinadas en la búsqueda del minimo no es un subespacio vectorial de L^∞. Esta es la razón por la que el algoritmo que define la mejor aproximación no es lineal.

El objetivo perseguido es representar funciones bastante complicadas con pocos datos: los $2N + 1$ coeficientes de los polinomios P y Q. Para ello, hay que saber controlar el error cometido en la aproximacion. Se pretende, por tanto, estimar $\gamma_N(f) = \|f — g_N\|_\infty$ cuando N tiende a infinito, siendo $g_N = g_N(f)$ la mejor aproximación racional de f. La aproximación racional no representará un progreso con respecto a la aproximación polinómica más que si, en abundantes ejemplos, $\gamma_N(f)$ tiende mucho más rápidamente a 0 cuando N tiende a infinito en el caso de la aproximación racional que en el caso de la aproximacion polinómica. Podremos entonces, para un cierto umbral aceptable de error, representar nuestra función f con ayuda de muy pocos coeficientes. Este problema se estudiará también cuando la norma L^∞ se sustituye por otras normas funcionales como la norma L^2 o la norma L^p.

Contrariamente a lo que ocurre en el marco de la aproximación polinómica, la sucesión $\gamma_N(f) = \|f — g_N\|_\infty$ *puede tener decrecimiento rápido, cuando N tiende a infinito, sin que* $f(\chi)$ *sea regular en el sentido habitual, sobre el intervalo I.*

Un ejemplo de este fenómeno viene dado por las sumas finitas $f(\chi) = c_1|\chi - \chi_1|^{\alpha 1} + \ldots + c_m|\chi - \chi_m|^{\alpha m}$ en las que los exponentes $\alpha_1, \ldots, \alpha_m$ son números positivos que se pueden escoger arbitrariamente pequeños. Se tiene, sin embargo, $\gamma_N = \|f - g_N\|_\infty = O(N^{-q})$ para todo entero q. La regularidad de $f(\chi)$ se puede hacer arbitrariamente débil y esta mala regularidad no afecta a la aproximación racional. Por ejemplo, si $f(\chi) = |\chi|$ y si 0 es el centro de I, el error cometido en la mejor aproximación racional, en norma L^∞, viene dado por $\gamma_N = O(\exp(-cN^{1/2}))$ donde c es una constante positiva. Si nos contentásemos con una aproximación polinómica, con polinomios de grado inferior a N, el error cometido sería en este caso superior a cN^{-1} donde c es una constante positiva. Aquí se ve cuánto acelera la convergencia la aproximación racional. Un ejemplo en el que la aproximación racional no aporta un progreso decisivo es $f(\chi) = \chi \sin(1/\chi)$.

V. Peller ha explicado estos resultados situándolos en un marco más general. Ha demostrado que el error $\gamma_N(f)$ cometido en la aproximación racional en norma uniforme verifica $\gamma_N = O(N^{-q})$, para todo entero q, *si y* sólo si $f(\chi)$ pertenece a todos los espacios de Besov $B^{1/p,p}(L^p)$ donde $p > 0$. Se ve que el exponente $s = 1/q$ que sirve para medir la regularidad es arbitrariamente grande, mientras que la norma utilizada para medir esta regularidad no cesa de deteriorarse.

De Vore ha relacionado el trabajo de Peller con el reordenamiento óptimo de las series de ondículas. Observó que $f(\chi)$ pertenece a todos los espacios de Besov $B^{1/p,p}(L^p)$, $p > 0$ si y sólo si los módulos $|\alpha_{j,k}|$ de los coeficientes de ondicula de $f(\chi) = \Sigma\Sigma\alpha_{j,k}\Psi_{j,k}(\chi)$, $\Psi_{j,k}(\chi) = \Psi(2^j\chi - k)$, una vez reordenados decrecientemente, constituyen una sucesión $c_1 > c_2 > \ldots > c_N > \ldots$ de decrecimiento rápido cuando N tiende a infinito. Por ello, la serie de ondículas, correctamente reordenada, tiene una convergencia rápida. Estos resultados fueron extendidos inmediatamente al caso multidimensional por Ronald De Vore, Bjorn Jawerth y V. A. Popov en el contexto de la aproximación no lineal por ondículas (sección 8).

Gracias a los trabajos de J. Peetre, V. Peller, A. Pekarskii y P. Petrushev, la aproximación no lineal se relaciona con la búsqueda de una segmentación óptima en el marco de la aproximación no lineal por splines con nodos libres.

En dimensión 1, se reemplaza L^{∞} por L^p *y* el conjunto G_N de fracciones racionales cuyos polos tienen posiciones arbitrarias, pero cuyo grado no es mayor que *N,* por el conjunto S_N de splines cuyos *N* nudos $t_1,t_2,\ldots t_N$ ocupan posiciones arbitrarias en el intervalo *I*. Se debe suponer, además, que el orden (de regularidad) γ de los splines utilizados es suficientemente grande, teniendo en cuenta la velocidad a la que se quiere que γ_N tienda hacia 0. Las caracterizaciones de las propiedades de regularidad (en el sentido dado por las normas de Besov) de $f(\chi)$ con ayuda de la velocidad de convergencia de la aproximación por "free knots splines", cuando el número de nodos tiende a infinito, serán las mismas que en el caso de la aproximación racional. Esta búsqueda de las posiciones óptimas de los *N* nodos $t_1, \ldots, t_N$ está emparentada con la de una segmentación óptima de la señal dada (o de la función *f* dada) sobre el intervalo *I*. Si la función que hay que segmentar es muy oscilante, como $e^{i\omega\chi}$ para un valor grande de ω, está claro que la segmentación óptima es decepcionante, puesto que consiste en descomponer la sinusoide en la sucesión de restricciones a los intervalos de longitud $2\pi/\omega$ y a destruir la información dada por la periodicidad. La misma observación vale para una función de la forma $|\omega|^{\alpha}\sin(1/\chi)$ que está mal aproximada por la aproximación racional o por la aproximación spline de nodos libres. Se vuelve a encontrar aquí el hecho de que las texturas estarán mal comprimidas en los algoritmos de compresión basados en el umbral de los coeficientes de ondículas.

8. APROXIMACION NO LINEAL EN DIMENSION CUALQUIERA

Demos algunas precisiones más sobre los resultados obtenidos por De Vore describiéndolos en dos casos importantes: la aproximación no lineal de las funciones que no son acotadas a priori y las de las funciones que ya poseen cierta regularidad a priori. En el primer caso, se parte de un exponente $p > 1$ y de una función "irregular" $f(\chi)$ perteneciente a $L^p(R^n)$. Pero el problema que vamos a resolver es, de hecho, local . Se quiere estimar, en dimensión cualquiera *n,* $\rho_N(f) = \inf\{\|f - f_N\|_p\}$ en los dos casos siguientes: (1) $f_N(\chi)$ pertenece a S_N que es el conjunto de "free knot splines" con a lo sumo *N* nudos o (2) $f_N = \Sigma\alpha\lambda\psi\lambda(\chi)$ y el índice *N* de f_N significa ahora que la suma comprende como mucho *N* ondículas. El conjunto S_N está definido mediante una particion del dominio *D* sobre el que se está trabajando en *N* cubos diádicos $Q_1, \ldots, Q_N$ y por el hecho de que sólo se consideran combinaciones lineales de los "basic splines" ajustados a dichos cubos diádicos *Q*. Se mide la calidad de la aproximación en n orma L^p utilizando un exponente α positivo que se supone primero inferior a *n/p.* Se define entonces *q* mediante $1/q - 1/p = \alpha/n = \beta$. Esta es la caracterización de las propiedades de regularidad de $f(\chi)$ con la ayuda de la calidad de la aproximación mediante los "free knots" o mediante ondículas.

Nuestra función *f* pertenece al espacio los coeficientes de ondículas donde se conservan únicamente los coeficientes $\alpha\lambda$ tales que $\|\alpha\lambda\psi\lambda\|_p \leq N^{-1/q}$.

Se trata efectivamente de una técnica de aproximación en la que el orden natural de la serie de ondículas es alterado en beneficio de una reorganización en la cual los términos se ordenan en orden decreciente de las normas L^p.

He aquí un segundo enunciado correspondiente a funciones $f(\chi)$ "regulares" y a valores de α superiores a n/q. Este enunciado cubre las técnicas de "knot removal" utilizadas en tratamiento de imagen. Se parte de una función $f(\chi)$ del espacio de Besov $B^{\alpha,q}(L^q)$ donde $\alpha > n/q$. Este espacio de Besov es caracterizado muy simplemente mediante el desarrollo en serie de ondículas . Una función f pertenece a $B^{\alpha,q}(L^q)$ si y sólo si sus coeficientes de ondículas $a(j, k,l)$ una vez multiplicados por $2^{j\beta}$ pertenecen a $l^q(Z \times R^n)$. Hemos definido $\beta = \alpha - n/q$ y los coeficientes de ondícul as se han calculado aquí adoptando para las ondículas la normalización en norma L^1. Para obtener una aproximación uniforme S_N de $f(\chi)$, con un error menor que δ, se define un umbral j_0 mediante $j_0 = \alpha^{-1} \log_2(1/\delta)$ y, para empezar, se conservan en la descomposición en ondículas ortogonales de $f(\chi)$, todos los términos correspondientes a las escalas $j \geq 0$, $j < j_0$. Supongamos que el punto de partida es la aproximación a escala 1, dada por una función V_0 y que buscamos una aproximación sobre un conjunto acotado. El número de términos conservados se eleva entonces a $C2^{nj_0}$. Si $j \geq j_0$, se aplica en cada escala el umbral explícito a los coeficientes de ondículas. Este umbral consiste en sustituir por 0 todos los coeficientes tales que $|\alpha(j, k)| < \varepsilon_j = \delta 2^{-\beta(j-j_0)}$. El número N_j de coeficientes conservados en la escala 2^{-j} es estimado observando que la condición de pertenecer a l^q implica la desigualdad débil correspondiente. Se obtiene $N_j \varepsilon_j^g \leq \Sigma\ |a(j,\ k)|^q$ y se tiene entonces $N < C2^{nj_0} + \Sigma N_j < C'2^{nj_0}$. El error cometido no sobrepasa $C\delta$. Esta aproximación, en la que son suficientes N términos para obtener un error inferior a $N^{-\alpha/n}$, es sorprendente, puesto que la regularidad global de f viene dada por el exponente de Holder $\beta = \alpha - n/q$. Utilizando un algoritmo lineal, el error sería de $N^{-\beta/n}$.

9. APROXIMACION NO LINEAL: APLICACIONES

Este tipo de consideraciones es utilizado en "computer aided design" (CAD) bajo el nombre de "knot removal". Se parte de una aproximación de la imagen por splines asociados a una malla extremadamente fina. Está claro que la transmisión de la imagen es relativamente costosa ya que no se ha explotado la redundancia de la imagen. La estrategia consiste entonces en utilizar las inclusiones naturales de los espacios de splines asociados a refinamientos de malla para suprimir poco a poco los nudos parásitos, que no aportan ninguna información pertinente, progresando sistemáticamente hacia las escalas más groseras (algoritmo de tipo "fine to coarse"). Se ha compactificado entonces la información de manera óptima y el algoritmo correspondiente sigue siendo del tipo de la aproximación no lineal de De Vore et al.

Un último y muy célebre ejemplo de aproximación no lineal es proporcionado por la descomposición atómica de las funciones del espacio de Hardy $H^1(R^n)$ en su versión "variable real" debida a Stein y Weiss. La descomposición atómica se obtiene a partir de la descomposición en ondículas, haciendo un reordenamiento de los términos, análogo a los que acabamos de hacer, pero un poco más sutil, en el sentido de que se hace al mismo tiempo un reordenamiento en paquetes, proporcionando cada paquete uno de los átomos de la descomposición atómica de Coifman y Weiss. El lector que desee saber más acerca de esto podrá consultar el tomo I de "Ondelettes et Opérateurs", Hermann (1990).

Repitamos una vez más que la aproximación no lineal es mucho más eficaz que la aproximación lineal. En otras palabras, se puede con muy pocos términos representar funciones bastante irregulares utilizando la aproximación no lineal, mientras que si quisiéramos la misma calidad de aproximación utilizando un esquema lineal, deberíamos utilizar una cantidad mucho más importante de términos de la serie (o imponer mucha más regularidad a las funciones que se quiere representar). Volviendo al tratamiento de imagen, la aproximación no lineal permite obtener bordes nítidos, optimizando a la vez la asignación en número de bits.

El procedimiento de Karhunen-Loève es en realidad un procedimiento lineal en el que el orden de los términos es el mismo para todas las imágenes de una cierta clase. El orden elegido y las funciones utilizadas convienen simultáneamente a todas las realizaciones. Ahora bien, a menudo ocurre que solamente se disponga de una realización particular. Cuando estudiamos el contraejemplo del proceso con una discontinuidad de primera especie aleatoria, los términos no nulos de la serie de ondículas correspondiente proporcionan una aproximación no lineal del tipo de las consideradas por De Vore et al.

Como ocurre a menudo, esta técnica de aproximación no lineai puede parecer, a posteriori, muy natural en análisis puesto que consiste en clasificar las cosas por orden de importancia, en vez de limitarse a un orden convencional, como el orden de los términos de una serie fijada de antemano. En el campo de la estadística, la aproximación no lineal es más cuestionada, porque el algoritmo de Karhunen-Loève disfruta todavía de una gran estima y parece el más adaptado a los datos experimentales. Pero pensamos que, muy a menudo, la hipótesis de trabajo que subyace bajo el algoritmo de Karhunen-Loève corresponde a una concepción científica errónea. En realidad, en el tratamiento de la señal de la palabra, los especialistas se han orientado hacia “átomos tiempo-frecuencia” adaptados. Estos “átomos tiempo-frecuencia” no vienen de ninguna manera proporcionados por la aplicación del algoritmo de Karhunen-Loève a la señal de la palabra. Lo mismo ocurre con los elementos estructurales que el algoritmo de “matching pursuit» de S. Mallat permite descubrir en una señal dada. En estos dos casos, el algoritmo de Karhunen-Loève no tendría ningún sentido puesto que no se dispone más que de una realización de la señal.

10. VUELTA AL PROBLEMA DE ELIMINACION DE RUIDO

Volvamos al problema fundamental de la eliminación de ruido de señales o imágenes “sucias”. Haremos dos hipótesis simplificadoras: por una parte, la señal o imagen es de tipo geométrico. Ello significa que la restricción a priori de tener una norma de Besov pequeña es una hipótesis de trabajo compatible con el objetivo perseguido: suprimir la texturas y el ruido, pero conservando bordes nítidos. Por otra parte, el ruido es un ruido blanco gaussiano. Ello implica que en cualquier base ortonormal, el ruido seguirá siendo un ruido blanco gaussiano. Se impone por lo tanto la utilización de una base ortonormal de ondículas pues ésta permite expresar de la manera más sencilla y eficaz la restricción a priori sobre la señal. Por otra parte, el ruido blanco está completamente incorrelado en toda base ortonormal.

El recién nacido de los algoritmos de Donoho tiene el siguiente carácter notable: su puesta en obra no depende más que de los exponentes α, p y q del espacio de Besov $B^{\alpha,q}(L^p)$ utilizado para modelizar los datos. Situémonos en el caso de la imagen e intentemos reconstruir $f(t_i)$ con la ayuda de los datos con ruido $d_i = f(t_i) + \sigma z_i$ donde z_i es un ruido blanco y donde los puntos $t_i = (k_{i,1}/N, k_{i,2}/N)$ pertenecen al mallado fino que define la imagen. He aquí cómo funciona el algoritmo: partimos de los datos con ruido, formamos los coeficientes de ondícula empíricos correspondientes (volveremos más adelante sobre la definición de estos coeficientes). Finalmente se aplica a estos coeficientes empíricos el “wavelet shrinkage” (algoritmo de estrechamiento) siguiente: todos los coeficientes inferiores, en módulo, a $\tau = 2\sigma N^{-1}(\log N)^{1/2}$ son sustituidos por 0. Aquellos que son, en módulo, superiores a τ son desplazados hacia 0 en una cantidad igual a τ; en otros términos, se sustituye cada coeficiente de ondícula, llamado χ, por $y = \theta(\chi) = \chi - \tau \mathrm{sign}(\chi)$ si $|\chi| > \tau$ y por $y = 0$, si no.

Donoho demuestra que este estimador posee las siguientes propiedades notables:

(1) el algoritmo es suboptimal, cada vez que se disponga de un conocimiento a priori de tipo Besov sobre la imagen o la señal estudiadas.

(2) el algoritmo preserva la regularidad (conocimiento a priori sobre la señal).

(3) el algoritmo hace desaparecer el ruido completamente; de hecho el umbral utilizado en el algoritmo depende del nivel de ruido.

El carácter suboptimal viene también definido por la velocidad a la que $\|f - g\|_2$ tiende a cero cuando al nivel de ruido σ tiende a cero (hemos designado por g el estimador de f dado por el algoritmo de Donoho). En el algoritmo de Donoho, se deben calcular los coeficientes de ondículas en una situación bastante diferente del caso usual de una función definida sobre toda la recta real. Disponemos en realidad de datos discretos, definidos sobre un intervalo. Las ondículas sobre el intervalo han sido construidas por I. Daubechies (en colaboración con A. Cohen y P. Vial). Grosso modo, se definen los espacios de aproximación V_j utilizando todas las funciones $\phi(2^j\chi - k)$ cuyo soporte está contenido en el intervalo I sobre el que se trabaja, pero añadiendo también una cantidad mínima de funciones de escala relativas a los extremos de I, de manera que se engendren todos los polinomios de grado menor o igual que N (en el caso de ondículas con N momentos nulos). La construcción de las ondículas sigue luego el procedimiento habitual. Simultáneamente, Ingrid Daubechies y sus colaboradores han construido los filtros necesarios para pasar de una escala a la siguiente. Son esos mismos filtros los que permitirán tratar los datos discretos que intervienen en el algoritmo de Donoho.

11. BIBLIOGRAFIA

1. DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. AMD PICARD D., Wavelet shrinkage. Tech. Rept. Statistics, Stanford, 1993.

2. DONOHO, D. L., Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, Tech. Rept. Statistics, Stanford, 1992.

3. DONOHO, D. L., Wavelet shrinkage and W.V.D.: a ten minute tour, in “Progress in wavelet analysis and applications” Y. Meyer and S. Roques ed. Editions Frontières, 1993.

4. DONOHO, D. L., Unconditional bases are optimal for data compression and statistical estimation. Apll. Comp. Harmonic Anal. 1, no. 1, 100-115.