Wavelets and sparse analysis

Albert Cohen

Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie
Paris

Journée Yves Meyer, Cachan, 20-06-2017

Wavelet theory from continuous to discrete : a golden decade

1981-1983 : theory of the continuous wavelet transform (Morlet-Grossmann)

1984 : stable discretization using frame theory (Daubechies)

1985 : construction of an orthonormal wavelet basis (Meyer)

1986-1987 : multiresolution analysis framework (Mallat-Meyer)

1988 : compactly supported orthonormal wavelets (Daubechies)

Sequences of nested spaces $\cdots \subset V_j \subset V_{j+1} \subset \cdots$ of the particular form

$$V_j := \mathrm{span}\{\varphi(2^j \cdot -k) \ : \ k \in \mathbb{Z}\},$$

are the natural framework for the construction of wavelet bases.

From $V_0 \subset V_1$, the scaling function φ should satisfy a two-scale equation

$$\varphi(x) = 2 \sum_{n \in \mathbb{Z}} h_n \varphi(2x - n),$$

with a sequence $(h_n)_{n \in \mathbb{Z}}$ such that $\sum h_n = 1$.

Example : $\varphi = \chi_{[0,1]} = \varphi(2x) + \varphi(2x-1)$, that is $h_0 = h_1 = \frac{1}{2}$.

The coefficients h_n play a key role, as filters, in fast wavelet transform algorithms.

Idea : any function that satisfies a two-scale equation of the above type is a natural candidate to generate a multiresolution analysis and in turn a wavelet basis. So let us design the coefficients h_n in such way that the solution φ has desirable properties.

Such functions are called "refinable function". They were independently identified in computer-aided geometric design as limits of refinement algorithms called subdivision schemes (Cavaretta-Dyn-Levin, Dahmen-Michelli, 1980's).

Sequences of nested spaces $\cdots \subset V_j \subset V_{j+1} \subset \cdots$ of the particular form

$$V_j := \mathrm{span}\{\varphi(2^j \cdot -k) \;:\; k \in \mathbb{Z}\},$$

are the natural framework for the construction of wavelet bases.

From $V_0 \subset V_1$, the scaling function φ should satisfy a two-scale equation

$$\varphi(x) = 2 \sum_{n\in\mathbb{Z}} h_n \varphi(2x-n),$$

with a sequence $(h_n)_{n\in\mathbb{Z}}$ such that $\sum h_n = 1$.

Example : $\varphi = \chi_{[0,1]} = \varphi(2x) + \varphi(2x-1)$, that is $h_0 = h_1 = \frac{1}{2}$.

The coefficients h_n play a key role, as filters, in fast wavelet transform algorithms.

Idea : any function that satisfies a two-scale equation of the above type is a natural candidate to generate a multiresolution analysis and in turn a wavelet basis. So let us design the coefficients h_n in such way that the solution φ has desirable properties.

Such functions are called "refinable function". They were independently identified in computer-aided geometric design as limits of refinement algorithms called subdivision schemes (Cavaretta-Dyn-Levin, Dahmen-Michelli, 1980's).

Sequences of nested spaces $\cdots \subset V_j \subset V_{j+1} \subset \cdots$ of the particular form

$$V_j := \mathrm{span}\{\varphi(2^j \cdot -k) \ : \ k \in \mathbb{Z}\},$$

are the natural framework for the construction of wavelet bases.

From $V_0 \subset V_1$, the scaling function φ should satisfy a two-scale equation

$$\varphi(x) = 2 \sum_{n\in\mathbb{Z}} h_n \varphi(2x - n),$$

with a sequence $(h_n)_{n\in\mathbb{Z}}$ such that $\sum h_n = 1$.

Example : $\varphi = \chi_{[0,1]} = \varphi(2x) + \varphi(2x-1)$, that is $h_0 = h_1 = \frac{1}{2}$.

The coefficients h_n play a key role, as filters, in fast wavelet transform algorithms.

Idea : any function that satisfies a two-scale equation of the above type is a natural candidate to generate a multiresolution analysis and in turn a wavelet basis. So let us design the coefficients h_n in such way that the solution φ has desirable properties.

Such functions are called "refinable function". They were independently identified in computer-aided geometric design as limits of refinement algorithms called subdivision schemes (Cavaretta-Dyn-Levin, Dahmen-Michelli, 1980's).

Sequences of nested spaces $\cdots \subset V_j \subset V_{j+1} \subset \cdots$ of the particular form

$$V_j := \mathrm{span}\{\varphi(2^j \cdot -k) \ : \ k \in \mathbb{Z}\},$$

are the natural framework for the construction of wavelet bases.

From $V_0 \subset V_1$, the scaling function φ should satisfy a two-scale equation

$$\varphi(x) = 2 \sum_{n \in \mathbb{Z}} h_n \varphi(2x - n),$$

with a sequence $(h_n)_{n\in\mathbb{Z}}$ such that $\sum h_n = 1$.

Example : $\varphi = \chi_{[0,1]} = \varphi(2x) + \varphi(2x-1)$, that is $h_0 = h_1 = \frac{1}{2}$.

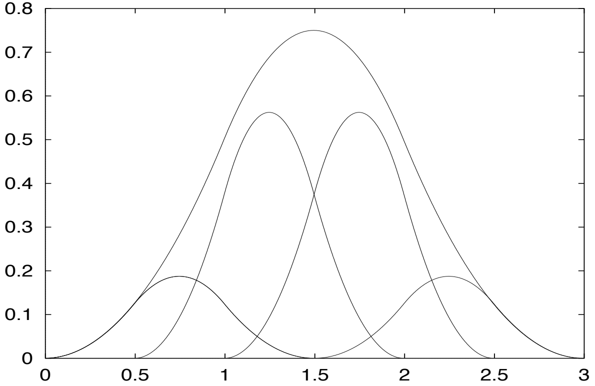
The coefficients h_n play a key role, as filters, in fast wavelet transform algorithms.

Idea : any function that satisfies a two-scale equation of the above type is a natural candidate to generate a multiresolution analysis and in turn a wavelet basis. So let us design the coefficients h_n in such way that the solution φ has desirable properties.

Such functions are called "refinable function". They were independently identified in computer-aided geometric design as limits of refinement algorithms called subdivision schemes (Cavaretta-Dyn-Levin, Dahmen-Michelli, 1980's).

Example : the B-splines

The choice $h_0 = h_3 = 1/8$ and $h_1 = h_2 = 3/8$ gives the quadratic B-spline $\varphi = B_2$.

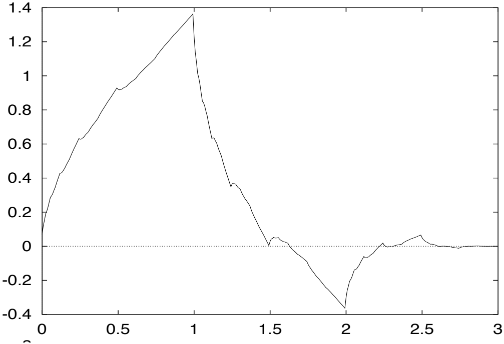


The B-spline of degree n is given by $B_n = (*)^{n+1}\chi_{[0,1]}$ (piecewise polynomials of degree n and globally C^{n-1}).

Two-scale equation obtained by convolution of the equation for $B_0 = \chi_{[0,1]}$.

Example : Daubechies orthonormal scaling functions

With a judicious choice of (h_0, h_1, h_2, h_3), Ingrid Daubechies constructs a scaling function φ that has orthonormal integer translates.



This function has no simple explict expression, but the coefficients h_n are explicitely given. It is not C^1 but has C^s Hölder smoothness for $s \sim .55$. Arbitrarily smooth such functions can be constructed up to raising their support length.

An explicit expression

Application of Fourier transform to $\varphi(x) = 2\sum_{n\in\mathbb{Z}} h_n \varphi(2x-n)$ gives

$$\widehat{\varphi}(\omega) = m(\omega/2)\widehat{\varphi}(\omega/2), \quad m(\omega) = \sum_{n\in\mathbb{Z}} h_n e^{-in\omega}.$$

and by iterating and using that $m(0) = 1$, we obtain

$$\widehat{\varphi}(\omega) = \prod_{j\geq 1} m(2^{-j}\omega).$$

Smooth φ can be constructed by taking $m(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^{n+1} p(\omega)$, so that

$$\widehat{\varphi}(\omega) = \widehat{B}_n(\omega)\widehat{P}(\omega), \quad \widehat{B}_n(\omega) = \left(\frac{1-e^{-i\omega}}{i\omega}\right)^{n+1}, \quad \widehat{P}(\omega) := \prod_{j\geq 1} p(2^{-j}\omega).$$

The construction of scaling function and wavelets by this strategy has been intensively studied, it led to the construction of the often used biorthogonal wavelets in 1992.

At that time, the graduate course of Yves Meyer was a “real time workshop” with new constructions and results being anounced and presented every week.

An explicit expression

Application of Fourier transform to $\varphi(x) = 2\sum_{n\in\mathbb{Z}} h_n \varphi(2x-n)$ gives

$$\widehat{\varphi}(\omega) = m(\omega/2)\widehat{\varphi}(\omega/2), \quad m(\omega) = \sum_{n\in\mathbb{Z}} h_n e^{-in\omega}.$$

and by iterating and using that $m(0) = 1$, we obtain

$$\widehat{\varphi}(\omega) = \prod_{j\geq 1} m(2^{-j}\omega).$$

Smooth φ can be constructed by taking $m(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^{n+1} p(\omega)$, so that

$$\widehat{\varphi}(\omega) = \widehat{B}_n(\omega)\widehat{P}(\omega), \quad \widehat{B}_n(\omega) = \left(\frac{1-e^{-i\omega}}{i\omega}\right)^{n+1}, \quad \widehat{P}(\omega) := \prod_{j\geq 1} p(2^{-j}\omega).$$

The construction of scaling function and wavelets by this strategy has been intensively studied, it led to the construction of the often used biorthogonal wavelets in 1992.

At that time, the graduate course of Yves Meyer was a "real time workshop" with new constructions and results being anounced and presented every week.

An explicit expression

Application of Fourier transform to $\varphi(x) = 2\sum_{n\in\mathbb{Z}} h_n \varphi(2x-n)$ gives

$$\widehat{\varphi}(\omega) = m(\omega/2)\widehat{\varphi}(\omega/2), \quad m(\omega) = \sum_{n\in\mathbb{Z}} h_n e^{-in\omega}.$$

and by iterating and using that $m(0) = 1$, we obtain

$$\widehat{\varphi}(\omega) = \prod_{j\geq 1} m(2^{-j}\omega).$$

Smooth φ can be constructed by taking $m(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^{n+1} p(\omega)$, so that

$$\widehat{\varphi}(\omega) = \widehat{B}_n(\omega)\widehat{P}(\omega), \quad \widehat{B}_n(\omega) = \left(\frac{1-e^{-i\omega}}{i\omega}\right)^{n+1}, \quad \widehat{P}(\omega) := \prod_{j\geq 1} p(2^{-j}\omega).$$

The construction of scaling function and wavelets by this strategy has been intensively studied, it led to the construction of the often used biorthogonal wavelets in 1992.

At that time, the graduate course of Yves Meyer was a "real time workshop" with new constructions and results being anounced and presented every week.

An explicit expression

Application of Fourier transform to $\varphi(x) = 2\sum_{n\in\mathbb{Z}} h_n \varphi(2x-n)$ gives

$$\widehat{\varphi}(\omega) = m(\omega/2)\widehat{\varphi}(\omega/2), \quad m(\omega) = \sum_{n\in\mathbb{Z}} h_n e^{-in\omega}.$$

and by iterating and using that $m(0) = 1$, we obtain

$$\widehat{\varphi}(\omega) = \prod_{j\geq 1} m(2^{-j}\omega).$$

Smooth φ can be constructed by taking $m(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^{n+1} p(\omega)$, so that

$$\widehat{\varphi}(\omega) = \widehat{B}_n(\omega)\widehat{P}(\omega), \quad \widehat{B}_n(\omega) = \left(\frac{1-e^{-i\omega}}{i\omega}\right)^{n+1}, \quad \widehat{P}(\omega) := \prod_{j\geq 1} p(2^{-j}\omega).$$

The construction of scaling function and wavelets by this strategy has been intensively studied, it led to the construction of the often used biorthogonal wavelets in 1992.

At that time, the graduate course of Yves Meyer was a “real time workshop” with new constructions and results being anounced and presented every week.

Construction de fonctions multifractales ayant un spectre de singularités prescrit.

Stéphane JAFFARD

1. Introduction et énoncés des résultats.

Soit $s > 0$ un exposant ; nous supposerons que s n'est pas un entier. Si x_0 est un nombre réel, nous définirons l'espace de Hölder ponctuel $C^s(x_0)$ par la condition

$$(1.1) \quad |f(x) - P(x - x_0)| \leq C|x - x_0|^s$$

où C est une constante, P un polynôme et où x appartient à un certain voisinage de x_0. On peut évidemment supposer que le degré de P ne dépasse pas la partie entière de s.

Si maintenant $s > 0$ est un nombre réel arbitraire, nous écrirons $f \in \Gamma^s(x_0)$ si

$$(1.2) \quad f \in \bigcap_{\varepsilon > 0} C^{s-\varepsilon}(x_0) \quad \text{mais} \quad f \notin \bigcup_{\varepsilon > 0} C^{s+\varepsilon}(x_0).$$

Calcul symbolique pour les opérateurs de Calderón-Zygmund généralisés

Ph. Tchamitchian

1. Introduction et énoncé du théorème

Nous désignons par $H^1 = H^1(\mathbb{R}^n)$ l'espace de Hardy généralisé de Stein et Weiss dont le dual est l'espace BMO de John et Nirenberg. Suivant Coifman et Weiss toute fonction $f \in H^1$ s'écrit $f = \sum_0^\infty \lambda_k a_k(x)$ où les λ_k sont des scalaires vérifiant $\sum_0^\infty |\lambda_k| < +\infty$ et où les $a_k(x)$ sont des atomes. Cela signifie qu'il existe des cubes Q_k tels que $\|a_k\|_\infty \leq \frac{1}{|Q_k|}$ support $a_k \subset Q_k$ et $\int a_k(x)dx = 0$.

Nous appelons $\mathcal{B} \subset \mathcal{L}(L^2, L^2)$, $L^2 = L^2(\mathbb{R}^n; dx)$, l'algèbre des opérateurs T bornés sur L^2 qui sont, de plus, bornés sur H^1 et sur BMO. Cela signifie qu'il existe une constante C telle que, pour tout atome $a(x)$, on ait $\|T(a)\|_{H^1} + \|T^*(a)\|_{H^1} \leq C$. Nous avons désigné par T^* l'adjoint de T.

La "version ondelettes" du théorème du Jacobien

Sylvia DOBYINSKY

1. Introduction.

On désigne par $\mathcal{H}^1(\mathbb{R}^2)$ l'espace de Hardy dans la version définie par E. Stein et G. Weiss : c'est à dire que f appartient à $\mathcal{H}^1(\mathbb{R}^2)$ si et seulement si f et les transformées de Riesz $R_1 f$ et $R_2 f$ appartiennent toutes trois à $L^1(\mathbb{R}^2)$.

Le théorème du Jacobien est l'énoncé suivant. Si $f(x,y)$ et $g(x,y)$ appartiennent à $L^1_{loc}(\mathbb{R}^2)$ et si les quatre dérivées (prises au sens des distributions) $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial g}{\partial x}$ et $\frac{\partial g}{\partial y}$ appartiennent à $L^2(\mathbb{R}^2)$, alors

$$(1.1) \qquad J(f,g) = \frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x} \in \mathcal{H}^1(\mathbb{R}^2).$$

COMPRESSION DES DONNEES ET RESTAURATION D'IMAGES BRUITEES

d'après **DAVID DONOHO**

1. Introduction.

David Donoho avait déjà acquis une très forte réputation en traitement statistique du signal lorsqu'il fut conduit à s'intéresser aux possibilités offertes par les ondelettes pour résoudre un de ses problèmes favoris. Nous allons décrire dans les pages qui suivent ce problème particulier et l'influence que sa solution a eue sur le développement contemporain de la recherche sur les ondelettes.

Ce qui m'enchante dans la démarche de Donoho est sa totale objectivité scientifique: il ne connaissait pas les ondelettes et avait un besoin urgent d'un outil très particulier. Lors d'une école d'été de statistiques à St.Flour, une de ses collègues, Dominique Picard, lui apprend que cet outil vient précisément d'être créé et que les détails se trouvent dans mon livre paru chez Hermann.

D.Donoho a passé le reste de l'été à lire cet ouvrage et a particulièrement aimé un des chapitres les plus controversés: celui où je montre que les ondelettes constituent un mode de représentation optimal pour une gamme d'espaces fonctionnels que je trouve délectables. Ces espaces sont les fameux espaces de Besov, que beaucoup de scientifiques rejettent avec mépris car trois indices différents apparaissent dans leur définition et que trois indices, c'est trop! En fait, Donoho nous apprend que ce florilège d'espaces fonctionnels permet une description délicate et précise de larges classes de signaux et d'images dont les propriétés génériques deviennent alors des propriétés fonctionnelles. Il en est résulte une suite de papiers retentissants sur le débruitage de signaux et d'images pour lesquels la connaissance a priori se modélise par l'ordre de grandeur d'une certaine norme Besov.

The space $B^s_{p,q}$ describes functions with s derivatives in L^p (fine tuning by q). Here for simplicity we take $q = p \in]0, \infty[$ and $s > 0$.

Oleg Besov (1959) : let $\Omega \subset \mathbb{R}^d$ be a domain, and $\Delta_h : f \mapsto f(\cdot + h) - f$ be the finite difference operator and Δ_h^m its power for m an integer such that $m > s$. A function $f \in L^p(\Omega)$ belongs to $B^s_{p,p}(\Omega)$ if and only if,

$$\left(\int_{t\geq 0} \left(t^{-s} \sup_{|h|\leq t} \|\Delta_h^m f\|_{L^p(\Omega_h)}\right)^p \frac{dt}{t}\right)^{1/p} < \infty$$

Here $\Omega_h := \{x \in \Omega : [x, x + mh] \subset \Omega\}$. The above-quantity added to $\|f\|_{L^p}$ defines a norm on $B^s_{p,p}(\Omega)$.

Yves Meyer (1990) : let (ψ_λ) be an orthonormal wavelet basis of $L^2(\Omega)$ with generating wavelet of class C^r for $r \geq s$. A function $f = \sum d_\lambda \psi_\lambda$ belongs to $B^s_{p,p}(\Omega)$ if and only if,

$$\left(\sum_\lambda 2^{ap|\lambda|}|d_\lambda|^p\right)^{1/p} < \infty, \quad a := s + \frac{d}{2} - \frac{d}{p}$$

Here $\psi_\lambda \sim 2^{dj/2}\psi(2^j \cdot -k)$, and we use the notation $|\lambda| = j$ for the scale level. The above-quantity defines an equivalent norm on $B^s_{p,p}(\Omega)$. This shows that wavelets are unconditional bases of these spaces.

The space $B^s_{p,q}$ describes functions with s derivatives in L^p (fine tuning by q). Here for simplicity we take $q = p \in]0, \infty[$ and $s > 0$.

Oleg Besov (1959) : let $\Omega \subset \mathbb{R}^d$ be a domain, and $\Delta_h : f \mapsto f(\cdot + h) - f$ be the finite difference operator and Δ_h^m its power for m an integer such that $m > s$. A function $f \in L^p(\Omega)$ belongs to $B^s_{p,p}(\Omega)$ if and only if,

$$\left(\int_{t\geq 0} \left(t^{-s} \sup_{|h|\leq t} \|\Delta_h^m f\|_{L^p(\Omega_h)}\right)^p \frac{dt}{t}\right)^{1/p} < \infty$$

Here $\Omega_h := \{x \in \Omega : [x, x + mh] \subset \Omega\}$. The above-quantity added to $\|f\|_{L^p}$ defines a norm on $B^s_{p,p}(\Omega)$.

Yves Meyer (1990) : let (ψ_λ) be an orthonormal wavelet basis of $L^2(\Omega)$ with generating wavelet of class C^r for $r \geq s$. A function $f = \sum d_\lambda \psi_\lambda$ belongs to $B^s_{p,p}(\Omega)$ if and only if,

$$\left(\sum_\lambda 2^{ap|\lambda|} |d_\lambda|^p\right)^{1/p} < \infty, \quad a := s + \frac{d}{2} - \frac{d}{p}$$

Here $\psi_\lambda \sim 2^{dj/2}\psi(2^j \cdot - k)$, and we use the notation $|\lambda| = j$ for the scale level. The above-quantity defines an equivalent norm on $B^s_{p,p}(\Omega)$. This shows that wavelets are unconditional bases of these spaces.

Consequences

In particular, when $s = \frac{d}{p} - \frac{d}{2}$ for some $p < 2$, then $a = 0$ so that $\|f\|_{B^s_{p,p}} \sim \|(d_\lambda)\|_{\ell^p}$.

Simple expression of classical results : with $s = \frac{d}{p} - \frac{d}{2}$ for $p < 2$, the critical Sobolev embedding $B^s_{p,p} \subset L^2$ is equivalent to the obvious observation that $\ell^p \subset \ell^2$

This embedding is not compact, yet an important approximation result holds : the best n-term approximation of f is defined $f_n = \sum_{\lambda \in \Lambda_n} d_\lambda \psi_\lambda$ where Λ_n corresponds to the n largest $|d_\lambda|$. Observation by Stechkin :

$$(d_\lambda) \in \ell^p \implies \|f - f_n\|_{L^2} \leq C n^{-r}, \quad C =: \|(d_\lambda)\|_{\ell^p}, \quad r = \frac{1}{p} - \frac{1}{2}.$$

Proof : use $(d_k)_{k \geq 1}$ the decreasing rearrangement of $(|d_\lambda|)$ and combine

$$\|f - f_n\|_{L^2}^2 = \sum_{k>n} d_k^2 \leq d_n^{2-p} \sum_{k>n} d_k^p \leq C^p d_n^{2-p} \quad \text{and} \quad n d_n^p \leq \sum_{k=1}^{n} d_k^p \leq C^p.$$

This result can be improved according to

$$(d_\lambda) \in \ell^p_w \iff \|f - f_n\|_{L^2} \leq C n^{-r}, \quad r = \frac{1}{p} - \frac{1}{2},$$

where $\ell^p_w := \{(d_\lambda) \; : \; \sup_{k \geq 1} k^{1/p} d_k < \infty\}$.

Consequences

In particular, when $s = \frac{d}{p} - \frac{d}{2}$ for some $p < 2$, then $a = 0$ so that $\|f\|_{B^s_{p,p}} \sim \|(d_\lambda)\|_{\ell^p}$.

Simple expression of classical results : with $s = \frac{d}{p} - \frac{d}{2}$ for $p < 2$, the critical Sobolev embedding $B^s_{p,p} \subset L^2$ is equivalent to the obvious observation that $\ell^p \subset \ell^2$

This embedding is not compact, yet an important approximation result holds : the best n-term approximation of f is defined $f_n = \sum_{\lambda \in \Lambda_n} d_\lambda \psi_\lambda$ where Λ_n corresponds to the n largest $|d_\lambda|$. Observation by Stechkin :

$$(d_\lambda) \in \ell^p \implies \|f - f_n\|_{L^2} \le Cn^{-r}, \quad C =: \|(d_\lambda)\|_{\ell^p}, \quad r = \frac{1}{p} - \frac{1}{2}.$$

Proof : use $(d_k)_{k \ge 1}$ the decreasing rearrangement of $(|d_\lambda|)$ and combine

$$\|f - f_n\|_{L^2}^2 = \sum_{k>n} d_k^2 \le d_n^{2-p} \sum_{k>n} d_k^p \le C^p d_n^{2-p} \quad \text{and} \quad n d_n^p \le \sum_{k=1}^{n} d_k^p \le C^p.$$

This result can be improved according to

$$(d_\lambda) \in \ell^p_w \iff \|f - f_n\|_{L^2} \le Cn^{-r}, \quad r = \frac{1}{p} - \frac{1}{2},$$

where $\ell^p_w := \{(d_\lambda) \ : \ \sup_{k \ge 1} k^{1/p} d_k < \infty\}$.

In particular, when $s = \frac{d}{p} - \frac{d}{2}$ for some $p < 2$, then $a = 0$ so that $\|f\|_{B^s_{p,p}} \sim \|(d_\lambda)\|_{\ell^p}$.

Simple expression of classical results : with $s = \frac{d}{p} - \frac{d}{2}$ for $p < 2$, the critical Sobolev embedding $B^s_{p,p} \subset L^2$ is equivalent to the obvious observation that $\ell^p \subset \ell^2$

This embedding is not compact, yet an important approximation result holds : the best n-term approximation of f is defined $f_n = \sum_{\lambda \in \Lambda_n} d_\lambda \psi_\lambda$ where Λ_n corresponds to the n largest $|d_\lambda|$. Observation by Stechkin :

$$(d_\lambda) \in \ell^p \implies \|f - f_n\|_{L^2} \leq Cn^{-r}, \quad C =: \|(d_\lambda)\|_{\ell^p}, \quad r = \frac{1}{p} - \frac{1}{2}.$$

Proof : use $(d_k)_{k \geq 1}$ the decreasing rearrangement of $(|d_\lambda|)$ and combine

$$\|f - f_n\|_{L^2}^2 = \sum_{k>n} d_k^2 \leq d_n^{2-p} \sum_{k>n} d_k^p \leq C^p d_n^{2-p} \quad \text{and} \quad n d_n^p \leq \sum_{k=1}^{n} d_k^p \leq C^p.$$

This result can be improved according to

$$(d_\lambda) \in \ell^p_w \iff \|f - f_n\|_{L^2} \leq Cn^{-r}, \quad r = \frac{1}{p} - \frac{1}{2},$$

where $\ell^p_w := \{(d_\lambda) \ : \ \sup_{k \geq 1} k^{1/p} d_k < \infty\}$.

In particular, when $s = \frac{d}{p} - \frac{d}{2}$ for some $p < 2$, then $a = 0$ so that $\|f\|_{B^s_{p,p}} \sim \|(d_\lambda)\|_{\ell^p}$.

Simple expression of classical results : with $s = \frac{d}{p} - \frac{d}{2}$ for $p < 2$, the critical Sobolev embedding $B^s_{p,p} \subset L^2$ is equivalent to the obvious observation that $\ell^p \subset \ell^2$

This embedding is not compact, yet an important approximation result holds : the best n-term approximation of f is defined $f_n = \sum_{\lambda \in \Lambda_n} d_\lambda \psi_\lambda$ where Λ_n corresponds to the n largest $|d_\lambda|$. Observation by Stechkin :

$$(d_\lambda) \in \ell^p \implies \|f - f_n\|_{L^2} \leq C n^{-r}, \quad C =: \|(d_\lambda)\|_{\ell^p}, \quad r = \frac{1}{p} - \frac{1}{2}.$$

Proof : use $(d_k)_{k \geq 1}$ the decreasing rearrangement of $(|d_\lambda|)$ and combine

$$\|f - f_n\|_{L^2}^2 = \sum_{k>n} d_k^2 \leq d_n^{2-p} \sum_{k>n} d_k^p \leq C^p d_n^{2-p} \quad \text{and} \quad n d_n^p \leq \sum_{k=1}^{n} d_k^p \leq C^p.$$

This result can be improved according to

$$(d_\lambda) \in \ell^p_w \iff \|f - f_n\|_{L^2} \leq C n^{-r}, \quad r = \frac{1}{p} - \frac{1}{2},$$

where $\ell^p_w := \{(d_\lambda) \ : \ \sup_{k \geq 1} k^{1/p} d_k < \infty\}$.

In particular, when $s = \frac{d}{p} - \frac{d}{2}$ for some $p < 2$, then $a = 0$ so that $\|f\|_{B^s_{p,p}} \sim \|(d_\lambda)\|_{\ell^p}$.

Simple expression of classical results : with $s = \frac{d}{p} - \frac{d}{2}$ for $p < 2$, the critical Sobolev embedding $B^s_{p,p} \subset L^2$ is equivalent to the obvious observation that $\ell^p \subset \ell^2$

This embedding is not compact, yet an important approximation result holds : the best n-term approximation of f is defined $f_n = \sum_{\lambda \in \Lambda_n} d_\lambda \psi_\lambda$ where Λ_n corresponds to the n largest $|d_\lambda|$. Observation by Stechkin :

$$(d_\lambda) \in \ell^p \implies \|f - f_n\|_{L^2} \leq Cn^{-r}, \quad C =: \|(d_\lambda)\|_{\ell^p}, \quad r = \frac{1}{p} - \frac{1}{2}.$$

Proof : use $(d_k)_{k \geq 1}$ the decreasing rearrangement of $(|d_\lambda|)$ and combine

$$\|f - f_n\|_{L^2}^2 = \sum_{k>n} d_k^2 \leq d_n^{2-p} \sum_{k>n} d_k^p \leq C^p d_n^{2-p} \quad \text{and} \quad n d_n^p \leq \sum_{k=1}^{n} d_k^p \leq C^p.$$

This result can be improved according to

$$(d_\lambda) \in \ell^p_w \iff \|f - f_n\|_{L^2} \leq Cn^{-r}, \quad r = \frac{1}{p} - \frac{1}{2},$$

where $\ell^p_w := \{(d_\lambda) \ : \ \sup_{k \geq 1} k^{1/p} d_k < \infty\}$.

Sparsity

Small dimensional phenomenon in high dimensional context.

A few numerically significant coefficients concentrate most of the energy.

Measuring sparsity

In the previous results $r = \frac{1}{p} - \frac{1}{2} \to \infty$ as $p \to 0$. The value of p quantifies the amount of sparsity in the coefficient sequence (d_λ).

This give the rate $n^{-s/d}$ for best n-term wavelet approximation in L^2 when $f \in B^s_{p,p}$ with $d = \frac{d}{p} - \frac{d}{2}$, however in the above observations, (ψ_λ) could be any orthonormal (or Riesz) basis.

Sparsity play a key role in various applications :

signal and image compression : allocate more bits to the large coefficients.

denoising and estimation : retain the observed coefficients that exceed noise level.

adaptive numerical simulation : approximate adaptively the solution to a PDE.

David Donoho (1993) : "Unconditional bases are optimal basis for data compression and for statistical estimation".

The hunt for sparsity is open

Sparse approximation is an instance of nonlinear approximation theorized in the 1980's by Ronald Devore, Pencho Petrushev, Vasil Popov.

Measuring sparsity

In the previous results $r = \frac{1}{p} - \frac{1}{2} \to \infty$ as $p \to 0$. The value of p quantifies the amount of sparsity in the coefficient sequence (d_λ).

This give the rate $n^{-s/d}$ for best n-term wavelet approximation in L^2 when $f \in B^s_{p,p}$ with $d = \frac{d}{p} - \frac{d}{2}$, however in the above observations, (ψ_λ) could be any orthonormal (or Riesz) basis.

Sparsity play a key role in various applications :

signal and image compression : allocate more bits to the large coefficients.

denoising and estimation : retain the observed coefficients that exceed noise level.

adaptive numerical simulation : approximate adaptively the solution to a PDE.

David Donoho (1993) : "Unconditional bases are optimal basis for data compression and for statistical estimation".

The hunt for sparsity is open

Sparse approximation is an instance of nonlinear approximation theorized in the 1980's by Ronald Devore, Pencho Petrushev, Vasil Popov.

Measuring sparsity

In the previous results $r = \frac{1}{p} - \frac{1}{2} \to \infty$ as $p \to 0$. The value of p quantifies the amount of sparsity in the coefficient sequence (d_λ).

This give the rate $n^{-s/d}$ for best n-term wavelet approximation in L^2 when $f \in B^s_{p,p}$ with $d = \frac{d}{p} - \frac{d}{2}$, however in the above observations, (ψ_λ) could be any orthonormal (or Riesz) basis.

Sparsity play a key role in various applications :

signal and image compression : allocate more bits to the large coefficients.

denoising and estimation : retain the observed coefficients that exceed noise level.

adaptive numerical simulation : approximate adaptively the solution to a PDE.

David Donoho (1993) : "Unconditional bases are optimal basis for data compression and for statistical estimation".

The hunt for sparsity is open

Sparse approximation is an instance of nonlinear approximation theorized in the 1980's by Ronald Devore, Pencho Petrushev, Vasil Popov.

Measuring sparsity

In the previous results $r = \frac{1}{p} - \frac{1}{2} \to \infty$ as $p \to 0$. The value of p quantifies the amount of sparsity in the coefficient sequence (d_λ).

This give the rate $n^{-s/d}$ for best n-term wavelet approximation in L^2 when $f \in B^s_{p,p}$ with $d = \frac{d}{p} - \frac{d}{2}$, however in the above observations, (ψ_λ) could be any orthonormal (or Riesz) basis.

Sparsity play a key role in various applications :

signal and image compression : allocate more bits to the large coefficients.

denoising and estimation : retain the observed coefficients that exceed noise level.

adaptive numerical simulation : approximate adaptively the solution to a PDE.

David Donoho (1993) : "Unconditional bases are optimal basis for data compression and for statistical estimation".

The hunt for sparsity is open

Sparse approximation is an instance of nonlinear approximation theorized in the 1980's by Ronald Devore, Pencho Petrushev, Vasil Popov.

Images and BV functions

$f \in BV$ if and only if $f \in L^1$ and ∇f is a finite measure.

Prototype : χ_Ω with Ω of a set of finite perimeter.

$BV([0,1]^2)$ was proposed as a model for images : "piecewise smooth" with edge singularities of finite length.

Cohen-DeVore-Petrushev-Xu (1993) : for the wavelet representation $f = \sum d_\lambda \psi_\lambda$,

$$f \in BV([0,1]^2) \Rightarrow (d_\lambda) \in w\ell^1,$$

i.e. $d_k \leq Ck^{-1}$.

Sharp estimate over BV : if $f = \chi_\Omega$ then $d_k \geq ck^{-1}$.

Optimal estimate for all bases

Best result for Fourier coefficients (Bourgain) : $\sum_{n\in\mathbb{Z}^2}(1+|n|)^{-1}|c_n| < \infty$.

Yves Meyer : "In a world where images are the BV functions and the eye measures error in L^2, wavelets are the best tool".

Images and BV functions

$f \in BV$ if and only if $f \in L^1$ and ∇f is a finite measure.

Prototype : χ_Ω with Ω of a set of finite perimeter.

$BV([0,1]^2)$ was proposed as a model for images : "piecewise smooth" with edge singularities of finite length.

Cohen-DeVore-Petrushev-Xu (1993) : for the wavelet representation $f = \sum d_\lambda \psi_\lambda$,

$$f \in BV([0,1]^2) \Rightarrow (d_\lambda) \in w\ell^1,$$

i.e. $d_k \leq Ck^{-1}$.

Sharp estimate over BV : if $f = \chi_\Omega$ then $d_k \geq ck^{-1}$.

Optimal estimate for all bases

Best result for Fourier coefficients (Bourgain) : $\sum_{n \in \mathbb{Z}^2} (1 + |n|)^{-1} |c_n| < \infty$.

Yves Meyer : "In a world where images are the BV functions and the eye measures error in L^2, wavelets are the best tool".

Cubist images

Edges of *BV* functions have finite length, but no geometric smoothness. One could hope for sparser representation for classes of images with geometrically smooth edges.

One of the simplest classes consists of piecewise constant images with straight edges.

Yves Meyer (2001) : let $f = \chi_\Omega$ where Ω is a polygon of $[0,1]^2$, then the decreasingly rearranged Fourier coefficients $c_k = c_k(f)$ decay at rate

$$c_k \leq Ck^{-1}\log(k),$$

quite similar to wavelet coefficients.

In higher dimension $d > 2$, when Ω is a polyhedron of $[0,1]^d$, this property persists with

$$c_k \leq Ck^{-1}\log(k)^{d-1},$$

and Fourier representations are then sparser than wavelet representations.

Cubist images

Edges of BV functions have finite length, but no geometric smoothness. One could hope for sparser representation for classes of images with geometrically smooth edges.

One of the simplest classes consists of piecewise constant images with straight edges.

Yves Meyer (2001) : let $f = \chi_\Omega$ where Ω is a polygon of $[0,1]^2$, then the decreasingly rearranged Fourier coefficients $c_k = c_k(f)$ decay at rate

$$c_k \leq Ck^{-1}\log(k),$$

quite similar to wavelet coefficients.

In higher dimension $d > 2$, when Ω is a polyhedron of $[0,1]^d$, this property persists with

$$c_k \leq Ck^{-1}\log(k)^{d-1},$$

and Fourier representations are then sparser than wavelet representations.

Cartoon images

The C^2-C^2 model : u is of the form $\sum_{j=1}^m u_j \chi_{\Omega_j}$, where the u_j are C^2 functions and Ω_j have piecewise C^2 boundaries.

Candes-Donoho (1998) introduce multiscale representation systems with directional selectivity : curvelets frames of the form

$$\psi_\lambda = 2^{3j/2}\psi(D^j R_l \cdot - k), \quad j \geq 0,\ k \in \mathbb{Z}, \quad l = 0, \dots, 2^j,$$

where $D = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$ and R_l is a rotation by $\pi l 2^{-j}$.

They prove that the curvelet representation $\sum d_\lambda \psi_\lambda$ of u obeying the C^2-C^2 model satisfies

$$(d_\lambda) \in \ell^p, \quad p > \frac{2}{3},$$

improving on wavelet and Fourier representation of such functions. The exponent 2/3 is optimal.

Variants : contourlets, shearlets, bandlets, anisotropic finite elements.

Cartoon images

The C^2-C^2 model : u is of the form $\sum_{j=1}^{m} u_j \chi_{\Omega_j}$, where the u_j are C^2 functions and Ω_j have piecewise C^2 boundaries.

Candes-Donoho (1998) introduce multiscale representation systems with directional selectivity : curvelets frames of the form

$$\psi_\lambda = 2^{3j/2}\psi(D^j R_l \cdot - k), \quad j \geq 0,\ k \in \mathbb{Z}, \quad l = 0, \dots, 2^j,$$

where $D = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$ and R_l is a rotation by $\pi l 2^{-j}$.

They prove that the curvelet representation $\sum d_\lambda \psi_\lambda$ of u obeying the C^2-C^2 model satisfies

$$(d_\lambda) \in \ell^p, \quad p > \frac{2}{3},$$

improving on wavelet and Fourier representation of such functions. The exponent $2/3$ is optimal.

Variants : contourlets, shearlets, bandlets, anisotropic finite elements.

Cartoon images

The C^2-C^2 model : u is of the form $\sum_{j=1}^m u_j \chi_{\Omega_j}$, where the u_j are C^2 functions and Ω_j have piecewise C^2 boundaries.

Candes-Donoho (1998) introduce multiscale representation systems with directional selectivity : curvelets frames of the form

$$\psi_\lambda = 2^{3j/2}\psi(D^j R_l \cdot - k), \quad j \geq 0,\ k \in \mathbb{Z}, \quad l = 0, \dots, 2^j,$$

where $D = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$ and R_l is a rotation by $\pi l 2^{-j}$.

They prove that the curvelet representation $\sum d_\lambda \psi_\lambda$ of u obeying the C^2-C^2 model satisfies

$$(d_\lambda) \in \ell^p, \quad p > \frac{2}{3},$$

improving on wavelet and Fourier representation of such functions. The exponent $2/3$ is optimal.

Variants : contourlets, shearlets, bandlets, anisotropic finite elements.

Sparsity in high dimensions

Numerical approximation of functions of many variables suffers from the curse of dimensionality. Example : best n-term wavelet approximation rate for $B^s_{p,p}$ is $n^{-s/d}$.

A source of infinite dimensional problems : parametrized PDEs. Elementary example :

$$-\mathrm{div}(a\nabla u) = f \ \text{ in } \ \Omega \subset \mathbb{R}^m \ \text{ and } \ u = 0 \ \text{ on } \ \partial\Omega,$$

where $f \in L^2$ if fixed and the diffusion function is parametrized according to

$$a = a(y) = \overline{a} + \sum_{j>0} y_j \psi_j \quad y = (y_j)_{j>0} \in U := [-1,1]^{\mathbb{N}},$$

where $\overline{a}$ and $(\psi_j)_{j\geq 1}$ are in L^∞, such that $0 < r \leq a(y) \leq R < \infty$ for all $y \in U$.

Polynomial expansions of the solution map from U to $V = H^1_0(\Omega)$:

$$y \mapsto u(y) = \sum_\nu u_\nu y^\nu, \quad y^\nu = \prod_{j\geq 1} y_j^{\nu_j}, \quad \nu = (\nu_1, \nu_2, \dots).$$

Cohen-DeVore-Schwab (2011) : for any $p < 1$,

$$\sum_j \|\psi_j\|^p_{L^\infty} < \infty \implies \sum_\nu \|u_\nu\|^p_V < \infty.$$

Other approaches for high dimensional problems : sparse tensor formats, sparse grids.

Numerical approximation of functions of many variables suffers from the curse of dimensionality. Example : best n-term wavelet approximation rate for $B^s_{p,p}$ is $n^{-s/d}$.

A source of infinite dimensional problems : parametrized PDEs. Elementary example :

$$-\mathrm{div}(a\nabla u) = f \ \text{ in } \ \Omega \subset \mathbb{R}^m \ \text{ and } \ u = 0 \ \text{ on } \ \partial\Omega,$$

where $f \in L^2$ if fixed and the diffusion function is parametrized according to

$$a = a(y) = \overline{a} + \sum_{j>0} y_j \psi_j \quad y = (y_j)_{j>0} \in U := [-1,1]^{\mathbb{N}},$$

where $\overline{a}$ and $(\psi_j)_{j\geq 1}$ are in L^∞, such that $0 < r \leq a(y) \leq R < \infty$ for all $y \in U$.

Polynomial expansions of the solution map from U to $V = H^1_0(\Omega)$:

$$y \mapsto u(y) = \sum_\nu u_\nu y^\nu, \quad y^\nu = \prod_{j\geq 1} y_j^{\nu_j}, \quad \nu = (\nu_1, \nu_2, \dots).$$

Cohen-DeVore-Schwab (2011) : for any $p < 1$,

$$\sum_j \|\psi_j\|^p_{L^\infty} < \infty \implies \sum_\nu \|u_\nu\|^p_V < \infty.$$

Other approaches for high dimensional problems : sparse tensor formats, sparse grids.

Numerical approximation of functions of many variables suffers from the curse of dimensionality. Example : best n-term wavelet approximation rate for $B^s_{p,p}$ is $n^{-s/d}$.

A source of infinite dimensional problems : parametrized PDEs. Elementary example :

$$-\mathrm{div}(a\nabla u) = f \text{ in } \Omega \subset \mathbb{R}^m \text{ and } u = 0 \text{ on } \partial\Omega,$$

where $f \in L^2$ if fixed and the diffusion function is parametrized according to

$$a = a(y) = \bar{a} + \sum_{j>0} y_j \psi_j \quad y = (y_j)_{j>0} \in U := [-1,1]^{\mathbb{N}},$$

where $\bar{a}$ and $(\psi_j)_{j\geq 1}$ are in L^∞, such that $0 < r \leq a(y) \leq R < \infty$ for all $y \in U$.

Polynomial expansions of the solution map from U to $V = H^1_0(\Omega)$:

$$y \mapsto u(y) = \sum_\nu u_\nu y^\nu, \quad y^\nu = \prod_{j\geq 1} y_j^{\nu_j}, \quad \nu = (\nu_1, \nu_2, \dots).$$

Cohen-DeVore-Schwab (2011) : for any $p < 1$,

$$\sum_j \|\psi_j\|^p_{L^\infty} < \infty \implies \sum_\nu \|u_\nu\|^p_V < \infty.$$

Other approaches for high dimensional problems : sparse tensor formats, sparse grids.

Numerical approximation of functions of many variables suffers from the curse of dimensionality. Example : best n-term wavelet approximation rate for $B^s_{p,p}$ is $n^{-s/d}$.

A source of infinite dimensional problems : parametrized PDEs. Elementary example :

$$-\mathrm{div}(a\nabla u) = f \text{ in } \Omega \subset \mathbb{R}^m \text{ and } u = 0 \text{ on } \partial\Omega,$$

where $f \in L^2$ if fixed and the diffusion function is parametrized according to

$$a = a(y) = \overline{a} + \sum_{j>0} y_j \psi_j \quad y = (y_j)_{j>0} \in U := [-1,1]^{\mathbb{N}},$$

where $\overline{a}$ and $(\psi_j)_{j\geq 1}$ are in L^∞, such that $0 < r \leq a(y) \leq R < \infty$ for all $y \in U$.

Polynomial expansions of the solution map from U to $V = H^1_0(\Omega)$:

$$y \mapsto u(y) = \sum_{\nu} u_\nu y^\nu, \quad y^\nu = \prod_{j\geq 1} y_j^{\nu_j}, \quad \nu = (\nu_1, \nu_2, \dots).$$

Cohen-DeVore-Schwab (2011) : for any $p < 1$,

$$\sum_j \|\psi_j\|^p_{L^\infty} < \infty \implies \sum_\nu \|u_\nu\|^p_V < \infty.$$

Other approaches for high dimensional problems : sparse tensor formats, sparse grids.

Numerical approximation of functions of many variables suffers from the curse of dimensionality. Example : best n-term wavelet approximation rate for $B^s_{p,p}$ is $n^{-s/d}$.

A source of infinite dimensional problems : parametrized PDEs. Elementary example :

$$-\mathrm{div}(a\nabla u) = f \ \text{ in } \ \Omega \subset \mathbb{R}^m \ \text{ and } \ u = 0 \ \text{ on } \ \partial\Omega,$$

where $f \in L^2$ if fixed and the diffusion function is parametrized according to

$$a = a(y) = \bar{a} + \sum_{j>0} y_j \psi_j \quad y = (y_j)_{j>0} \in U := [-1,1]^{\mathbb{N}},$$

where $\bar{a}$ and $(\psi_j)_{j\geq 1}$ are in L^∞, such that $0 < r \leq a(y) \leq R < \infty$ for all $y \in U$.

Polynomial expansions of the solution map from U to $V = H^1_0(\Omega)$:

$$y \mapsto u(y) = \sum_{\nu} u_\nu y^\nu, \quad y^\nu = \prod_{j\geq 1} y_j^{\nu_j}, \quad \nu = (\nu_1, \nu_2, \dots).$$

Cohen-DeVore-Schwab (2011) : for any $p < 1$,

$$\sum_j \|\psi_j\|^p_{L^\infty} < \infty \implies \sum_\nu \|u_\nu\|^p_V < \infty.$$

Other approaches for high dimensional problems : sparse tensor formats, sparse grids.

Exploiting sparsity in a different way

Assume that f is a sparse signal or image (in some basis).

Classical way to encode f : retain its k largest coordinates in the basis and encode them. This requires to compute all coordinates before discarding the small one.

Compressed sensing (Donoho, Candes-Tao, 2000's) : use m linear measurements of f prescribed in advance, and exploit that f is sparse in order to reconstruct it accurately from these measurements.

In other word, we observe $y = \Phi f \in \mathbb{R}^m$ with Φ a fixed measurement matrix and we want to build $g = \Delta(y)$ close to f.

Key ingredient : Δ should be nonlinear.

Exploiting sparsity in a different way

Assume that f is a sparse signal or image (in some basis).

Classical way to encode f : retain its k largest coordinates in the basis and encode them. This requires to compute all coordinates before discarding the small one.

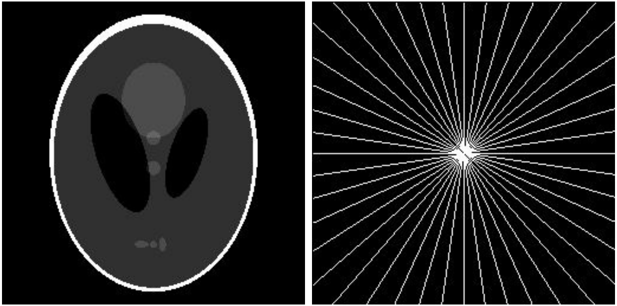
Compressed sensing (Donoho, Candes-Tao, 2000's) : use m linear measurements of f prescribed in advance, and exploit that f is sparse in order to reconstruct it accurately from these measurements.

In other word, we observe $y = \Phi f \in \mathbb{R}^m$ with Φ a fixed measurement matrix and we want to build $g = \Delta(y)$ close to f.

Key ingredient : Δ should be nonlinear.

An instructive example : 2D tomography (Candes-Romberg-Tao)

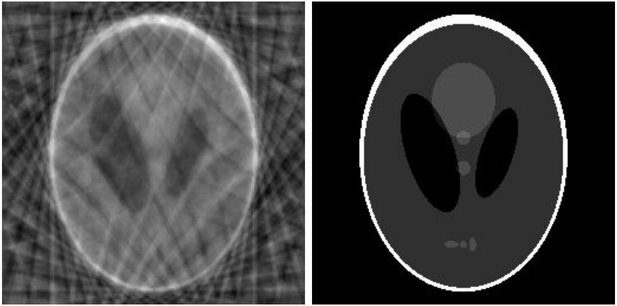
The Radon transform captures partial Fourier information.



Left : the Logan-Shep phantom test image

Right : position of the observed Fourier coefficients (white)

Two different reconstructions

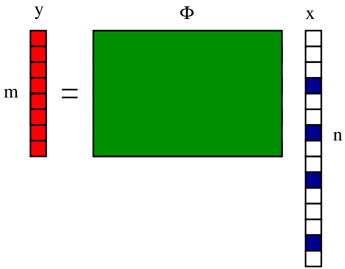


Left : put the unknown coefficient to zero (minimum ℓ^2 norm) and reconstruct the partial Fourier serie $\Rightarrow$ oscillation artifacts.

Right : adjust the unknown coefficients so to minimize the total variation of the image $|f|_{TV} = \int |\nabla f|$ $\Rightarrow$ nearly exact reconstruction !

Questions

Let Σ_k be the set of k-sparse vectors ($x \in \mathbb{R}^N$ with at most k non-zero entries).



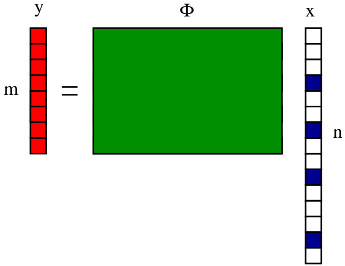
Minimal number m of measures which is sufficient to characterize any $x \in \Sigma_k$.

With which matrices Φ ? Which decodes Δ ?

Robustness ? In practice, $y = \Phi x + e$ with $\|e\|_{\ell^2} \leq \varepsilon$ and $x \in \mathbb{R}^n$ close to Σ_k.

Questions

Let Σ_k be the set of k-sparse vectors ($x \in \mathbb{R}^N$ with at most k non-zero entries).



Minimal number m of measures which is sufficient to characterize any $x \in \Sigma_k$.

With which matrices Φ ? Which decodes Δ ?

Robustness ? In practice, $y = \Phi x + e$ with $\|e\|_{\ell^2} \leq \varepsilon$ and $x \in \mathbb{R}^n$ close to Σ_k.

Reconstruction results

With $m = 2k$ measures and generic choice of Φ, one can reconstruct exactly any $x \in \Sigma_k$, but...

(i) Complex decoder : $\Delta(y) := \mathrm{Argmin}\{\|y - \Phi z\| \ : \ z \in \Sigma_k\}$, and therefore $\mathcal{O}(N^k)$ least square systems to solve. Alternative : $\Delta(y) := \mathrm{Argmin}\{\|z\|_0 \ : \ \Phi z = y\}$, with $\|z\|_0 = \#\{i \ : \ z_i \neq 0\}$, same complexity.

(ii) No robustness to noise and deviation from Σ_k.

Candes-Tao (2005) : using $m \sim ck \log(N/k)$ measures and specific Φ which satisfy restricted isometry properties (RIP), one can reconstruct exactly any $x \in \Sigma_k$, with

(i) Simple decoder : $\Delta(y) := \mathrm{Argmin}\{\|z\|_1 \ : \ \Phi z = y\}$ with $\|z\|_1 := |z_1| + \cdots + |z_n|$, convex optimization, linear programming.

(ii) Robustness : $\|x - \Delta(\Phi x)\|$ controlled by noise and deviation of x from Σ_k.

but... Φ obtained by probabilistic techniques. Example : $\Phi = (\Phi_{i,j})$ with $\Phi_{i,j}$ independant random draws of Bernoulli ± 1 or gaussians $\mathcal{N}(0, 1)$.

Deterministic constructions known for $m \sim ck^2 \log(N/k)$ (DeVore, Calderbank).

Reconstruction results

With $m = 2k$ measures and generic choice of Φ, one can reconstruct exactly any $x \in \Sigma_k$, but...

(i) Complex decoder : $\Delta(y) := \mathrm{Argmin}\{\|y - \Phi z\| \ : \ z \in \Sigma_k\}$, and therefore $\mathcal{O}(N^k)$ least square systems to solve. Alternative : $\Delta(y) := \mathrm{Argmin}\{\|z\|_0 \ : \ \Phi z = y\}$, with $\|z\|_0 = \#\{i \ : \ z_i \neq 0\}$, same complexity.

(ii) No robustness to noise and deviation from Σ_k.

Candes-Tao (2005) : using $m \sim ck \log(N/k)$ measures and specific Φ which satisfy restricted isometry properties (RIP), one can reconstruct exactly any $x \in \Sigma_k$, with

(i) Simple decoder : $\Delta(y) := \mathrm{Argmin}\{\|z\|_1 \ : \ \Phi z = y\}$ with $\|z\|_1 := |z_1| + \cdots + |z_n|$, convex optimization, linear programming.

(ii) Robustness : $\|x - \Delta(\Phi x)\|$ controlled by noise and deviation of x from Σ_k.

but... Φ obtained by probabilistic techniques. Example : $\Phi = (\Phi_{i,j})$ with $\Phi_{i,j}$ independant random draws of Bernoulli ± 1 or gaussians $\mathcal{N}(0,1)$.

Deterministic constructions known for $m \sim ck^2 \log(N/k)$ (DeVore, Calderbank).

Reconstruction results

With $m = 2k$ measures and generic choice of Φ, one can reconstruct exactly any $x \in \Sigma_k$, but...

(i) Complex decoder : $\Delta(y) := \mathrm{Argmin}\{\|y - \Phi z\| \ : \ z \in \Sigma_k\}$, and therefore $\mathcal{O}(N^k)$ least square systems to solve. Alternative : $\Delta(y) := \mathrm{Argmin}\{\|z\|_0 \ : \ \Phi z = y\}$, with $\|z\|_0 = \#\{i \ : \ z_i \neq 0\}$, same complexity.

(ii) No robustness to noise and deviation from Σ_k.

Candes-Tao (2005) : using $m \sim ck \log(N/k)$ measures and specific Φ which satisfy restricted isometry properties (RIP), one can reconstruct exactly any $x \in \Sigma_k$, with

(i) Simple decoder : $\Delta(y) := \mathrm{Argmin}\{\|z\|_1 \ : \ \Phi z = y\}$ with $\|z\|_1 := |z_1| + \cdots + |z_n|$, convex optimization, linear programming.

(ii) Robustness : $\|x - \Delta(\Phi x)\|$ controlled by noise and deviation of x from Σ_k.

but... Φ obtained by probabilistic techniques. Example : $\Phi = (\Phi_{i,j})$ with $\Phi_{i,j}$ independant random draws of Bernoulli ± 1 or gaussians $\mathcal{N}(0,1)$.

Deterministic constructions known for $m \sim ck^2 \log(N/k)$ (DeVore, Calderbank).

Reconstruction results

With $m = 2k$ measures and generic choice of Φ, one can reconstruct exactly any $x \in \Sigma_k$, but...

(i) Complex decoder : $\Delta(y) := \mathrm{Argmin}\{\|y - \Phi z\| \; : \; z \in \Sigma_k\}$, and therefore $\mathcal{O}(N^k)$ least square systems to solve. Alternative : $\Delta(y) := \mathrm{Argmin}\{\|z\|_0 \; : \; \Phi z = y\}$, with $\|z\|_0 = \#\{i \; : \; z_i \neq 0\}$, same complexity.

(ii) No robustness to noise and deviation from Σ_k.

Candes-Tao (2005) : using $m \sim ck \log(N/k)$ measures and specific Φ which satisfy restricted isometry properties (RIP), one can reconstruct exactly any $x \in \Sigma_k$, with

(i) Simple decoder : $\Delta(y) := \mathrm{Argmin}\{\|z\|_1 \; : \; \Phi z = y\}$ with $\|z\|_1 := |z_1| + \cdots + |z_n|$, convex optimization, linear programming.

(ii) Robustness : $\|x - \Delta(\Phi x)\|$ controlled by noise and deviation of x from Σ_k.

but... Φ obtained by probabilistic techniques. Example : $\Phi = (\Phi_{i,j})$ with $\Phi_{i,j}$ independant random draws of Bernoulli ± 1 or gaussians $\mathcal{N}(0,1)$.

Deterministic constructions known for $m \sim ck^2 \log(N/k)$ (DeVore, Calderbank).

Sampling of band-limited discrete signals

Consider periodic functions on the 2d torus $[0,1]^2 \sim \mathbb{R}^2/\mathbb{Z}^2$.

We say that f is β-sparse if its support inside $[0,1]^2$ is a set S of measure $|S| \leq \beta$.

Matei-Meyer (2009) construct for any $0 < \alpha < \frac{1}{2}$, deterministic sets $\Lambda_\alpha \in \mathbb{Z}^2$ of density 2α such that any $f \in L^1([0,1]^2)$ that is β-sparse for $\beta < \alpha$ and positive, is characterized by the Fourier coefficients $c_n(f)$ for $n \in \Lambda_\alpha$.

Other constructions have been proposed by Olevskii and Ulanovskii.

The sets proposed Matei-Meyer are simple quasi-crystals. One example of such a set is

$$\Lambda_\alpha := \{n = (n_1, n_2) \in \mathbb{Z}^2 \ : \ \mathrm{dist}(n_1\sqrt{2} + n_2\sqrt{3}, \mathbb{Z}) \leq \alpha\}$$

This set has density 2α : for all $\varepsilon > 0$, there exists $R = R(\varepsilon)$ such that

$$(2\alpha - \varepsilon)\pi R^2 \leq \#\{n \in \mathbb{Z}^2 \ : \ \Lambda_\alpha \cap |n - x| \leq R\} \leq (2\alpha + \varepsilon)\pi R^2, \quad x \in \mathbb{R}^2.$$

Periodic lattices would not work, due to the phenomenon of aliasing.

Sampling of band-limited discrete signals

Consider periodic functions on the 2d torus $[0,1]^2 \sim \mathbb{R}^2/\mathbb{Z}^2$.

We say that f is β-sparse if its support inside $[0,1]^2$ is a set S of measure $|S| \leq \beta$.

Matei-Meyer (2009) construct for any $0 < \alpha < \frac{1}{2}$, deterministic sets $\Lambda_\alpha \in \mathbb{Z}^2$ of density 2α such that any $f \in L^1([0,1]^2)$ that is β-sparse for $\beta < \alpha$ and positive, is characterized by the Fourier coefficients $c_n(f)$ for $n \in \Lambda_\alpha$.

Other constructions have been proposed by Olevskii and Ulanovskii.

The sets proposed Matei-Meyer are simple quasi-crystals. One example of such a set is

$$\Lambda_\alpha := \{n = (n_1, n_2) \in \mathbb{Z}^2 \ : \ \mathrm{dist}(n_1\sqrt{2} + n_2\sqrt{3}, \mathbb{Z}) \leq \alpha\}$$

This set has density 2α : for all $\varepsilon > 0$, there exists $R = R(\varepsilon)$ such that

$$(2\alpha - \varepsilon)\pi R^2 \leq \#\{n \in \mathbb{Z}^2 \ : \ \Lambda_\alpha \cap |n - x| \leq R\} \leq (2\alpha + \varepsilon)\pi R^2, \quad x \in \mathbb{R}^2.$$

Periodic lattices would not work, due to the phenomenon of aliasing.

Sampling of band-limited discrete signals

Consider periodic functions on the 2d torus $[0,1]^2 \sim \mathbb{R}^2/\mathbb{Z}^2$.

We say that f is β-sparse if its support inside $[0,1]^2$ is a set S of measure $|S| \leq \beta$.

Matei-Meyer (2009) construct for any $0 < \alpha < \frac{1}{2}$, deterministic sets $\Lambda_\alpha \in \mathbb{Z}^2$ of density 2α such that any $f \in L^1([0,1]^2)$ that is β-sparse for $\beta < \alpha$ and positive, is characterized by the Fourier coefficients $c_n(f)$ for $n \in \Lambda_\alpha$.

Other constructions have been proposed by Olevskii and Ulanovskii.

The sets proposed Matei-Meyer are simple quasi-crystals. One example of such a set is

$$\Lambda_\alpha := \{n = (n_1, n_2) \in \mathbb{Z}^2 \ : \ \mathrm{dist}(n_1\sqrt{2} + n_2\sqrt{3}, \mathbb{Z}) \leq \alpha\}$$

This set has density 2α : for all $\varepsilon > 0$, there exists $R = R(\varepsilon)$ such that

$$(2\alpha - \varepsilon)\pi R^2 \leq \#\{n \in \mathbb{Z}^2 \ : \ \Lambda_\alpha \cap |n - x| \leq R\} \leq (2\alpha + \varepsilon)\pi R^2, \quad x \in \mathbb{R}^2.$$

Periodic lattices would not work, due to the phenomenon of aliasing.

Conclusions

Wavelet have been instrumental in putting the concept of sparsity into the forefront.

General objective : describe the properties of classes of object of interest by means of an appropriate representation.

Many applications : data compression, estimation, inverse problems, numerical simulation, compressed sensing.

Many open problems, in particular in high dimensions : what are the relevant classes ? what are the natural sparse representations for these classes ?