How much should we rely on Besov spaces as a framework for the mathematical study of images?

C. Sinan Güntürk

Princeton University, Program in Applied and Computational Mathematics

Abstract

Relations between smoothness spaces (Sobolev and in general Besov) and their wavelet based approximation properties are well known and have been proposed to explain several issues of image compression. However, the validity and accuracy of this approach as a method to analyze and classify images still need to be explored for several reasons. Among them, we investigate the following:

- Asymptotical aspects of the characterizations, like their observability and/or importance. Issues about the range of validity of asymptotical estimates. (e.g., approximation rate in the presence of low level noise or cases with no definite rate estimate. Also images with regions of poor regularity.)

- Since belonging to a certain (Besov) space is far from being an image, what is the extent to which Besov spaces are good model classes for images? In particular, because wavelets form unconditional bases for them, Besov spaces are "immune" to sign and position changes of wavelet coefficients, but images do not have this property, thus raising questions about how edges (carrying most of

the information) and other kinds of singularities interact with the sign structure of the expansions.

- The norms of these spaces measure global smoothness, could a local regularity analysis help us understand images better? Noting that image coders usually exploit local features for adaptivity, coding decisions based solely on global measures should be sub-optimal.

In the deterministic setting, the multifractal formalism is taken into consideration as a finer tool to analyse smoothness of images. An interpretation of the Hölder spectrum as a means to inspect the approximation rate of a signal is given.

In the stochastic setting, the piecewise smooth stochastic image model proposed by Cohen and d'Ales is known to demonstrate the effectiveness of wavelet bases in non-linear approximation against optimally decorrelating Karhunen-Loeve basis. We investigate almost sure properties of this model. A variant of a theorem of Kolmogorov leads to the result that for this model, the rate of expected approximation error is not superior to the rate of approximation error valid for almost any realization.

INTRODUCTION and BACKGROUND

Mathematical Models of Image Compression

Besov spaces have recently been proposed [1] as a framework for the mathematical study of image compression. These spaces could be regarded as refined versions of the classical Sobolev spaces and depending on their smoothness index values they can contain functions of varying smoothness. Moreover they turn out to be the characterizing spaces for many problems in approximation theory.

Framework:

- Error/distortion measured in L^p, mostly in L^2.
- $f \in X \subset L^2$, for some class X.

- Approximations are (truncated) basis expansions wrt an orthonormal wavelet basis.

- Wavelets are compactly supported, with sufficient number of vanishing moments and regularity.

- Approximation spaces: $\{V_j\}_{j=0}^{\infty}$

$$V_j \subset V_{j+1} \subset ...$$

  Linear approximation from V_j

$$f_j = \mathbf{Proj}_{V_j} f = \sum_{j' \leq j,\, k} < f, \psi_{j',k} > \psi_{j',k}$$

- Approximation error: $\|f - f_j\|_{L^2}$.

Error decay vs Smoothness

Characterization of function spaces by means of wavelet coefficients/approximation error. Let $c_{j,k}$ denote $< f, \psi_{j,k} >$.

- Lipα spaces (Lip$^*\alpha$ for $\alpha \in \mathbf{Z}$.)

$$f \in \text{Lip}\alpha \iff |c_{j,k}| \leq C.2^{-j(\alpha+1/2)} \iff \|f - f_j\|_\infty \leq C.2^{-j\alpha}$$

- Sobolev spaces $H^\alpha = W_2^\alpha$

$$f \in H^\alpha \iff \sum_{j,k} |c_{j,k}|^2 2^{2\alpha j} < \infty \iff \{\|f - f_j\|_{L^2} 2^{j\alpha}\}_{j\geq 0} \in l^2$$

More generally,

- Besov spaces $B_{p,q}^\alpha$

$$f \in B_{p,q}^\alpha \iff \left\{\|f - f_j\|_p 2^{j\alpha}\right\}_{j\geq 0} \in l^q$$

Besov classes include all above: Lipα, W_p^α, Lip(α, L_p).

Equivalent Besov norm: $|f|_{B^\alpha_{p,q}} \approx \left\| 2^{j\alpha} \|f - f_j\|_p \right\|_{l^q}$

Roughly, all above says:

f has α order of smoothness in L_p

$$\Longleftrightarrow$$

$\left\|\text{approximation error from } V_j\right\|_p$ decays like $2^{-j\alpha}$ (or slightly faster)

$$\Longleftrightarrow$$

$$\|\text{first N-term approximation error}\|_p = O(N^{-\alpha/d})$$

where f is defined on $[0,1]^d$.

Nonlinear Theory:

$\mathcal{H}$ be a Hilbert space and $\{\varphi_n\}_{n\geq 0}$ be an ONB of $\mathcal{H}$.

Define Σ_N, N-term functions:

$$\Sigma_N := \left\{ f = \sum c_k \varphi_k ; \#\{k; c_k \neq 0\} = N \right\}$$

$\mathbf{A}_N f :=$ best approximation to f from Σ_N

$$= \sum_{\Lambda_N(f)} < f, \varphi_n > \varphi_n,$$

where $\Lambda_N(f)$ is the set of indices of N largest coefficients.

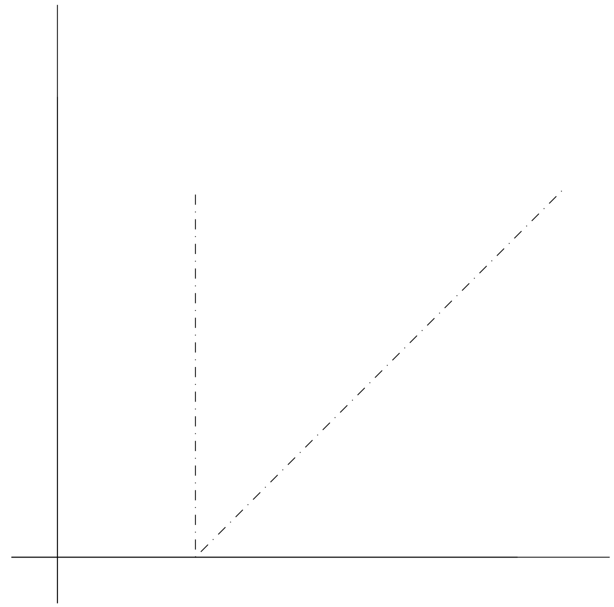
$$\left\{ 2^{j\alpha} \|f - \mathbf{A}_{2^{jd}} f\|_2 \right\}_{j\geq 0} \in l^q \iff f \in B^{\alpha}_{q,q} \quad (\frac{1}{q} = \frac{1}{2} + \frac{\alpha}{d})$$

$$\iff \left\{ \sum_{N=1}^{\infty} [N^{\alpha/d} \|f - \mathbf{A}_N f\|_2]^q \frac{1}{N} \right\}^{1/q} < \infty$$

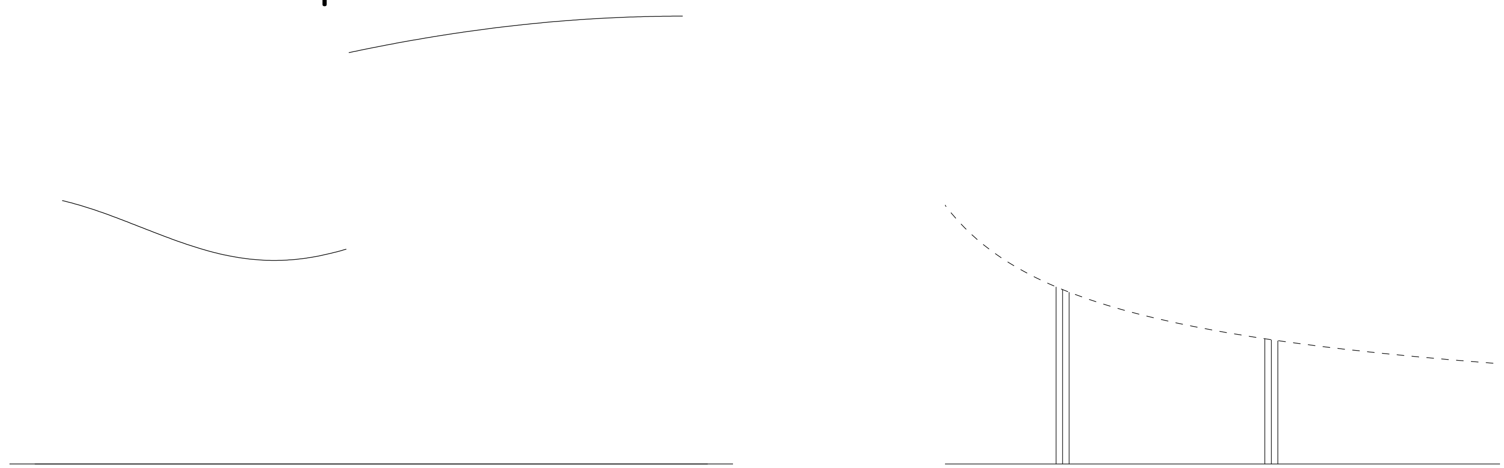
Linear vs Nonlinear

approximation	space	criteria
from a linear subspace of dim 2^{jd}	$B^{\alpha}_{2,q}$	$\left\{ 2^{j\alpha} \|f - \mathbf{P}_{2^{jd}} f\|_2 \right\} \in l^q$
from a nonlinear manifold of dim 2^{jd}	$B^{\alpha}_{q,q}$	$\left\{ 2^{j\alpha} \|f - \mathbf{A}_{2^{jd}} f\|_2 \right\} \in l^q$

in both cases: $\epsilon(N) \sim O(N^{-\alpha/d})$.



An important example:

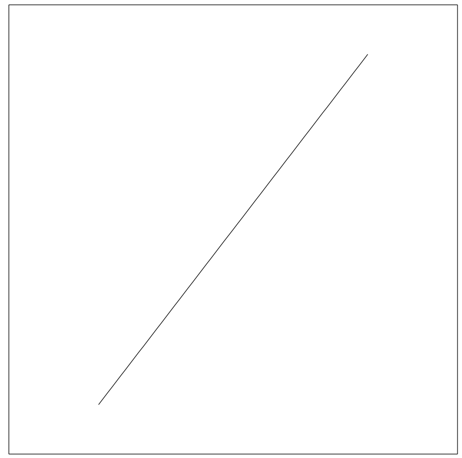


$f \in B^{1/q}_{q,\infty}$ for all q.

Linear: $f \in B^{1/2}_{2,\infty} \Rightarrow \epsilon(N) = O(N^{-1/2})$. $(2^j \mapsto 2^{-j/2})$

Nonlinear: $f \in B^{\alpha}_{q,q}, 1/q = 1/2 + \alpha$, for arbitrarily big α's so ANY rate is achievable. $(j \mapsto 2^{-j/2})$ If C^∞ is replaced by C^α, any rate $< \alpha$ is achievable.

What happens in 2-D?

Consider an edge:  $f \in B^{\min(\alpha,1/q)}_{q,\infty}$, for all q.

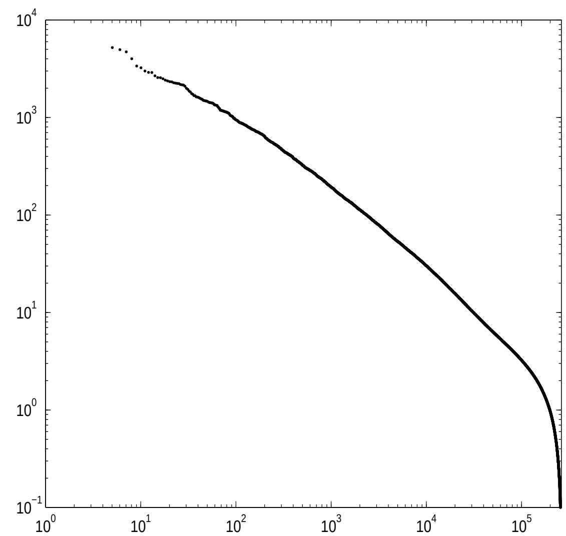
linear: $f \in B^{1/2}_{2,\infty}, \quad (\alpha \geq 1/2) \Rightarrow \epsilon(N) = O(N^{-1/4})$.

nonlinear: $c.2^j$ wavelets hit the edge, of size $2^{-j} \Rightarrow \epsilon(N) = O(N^{-1/2})$.

More generally, if $\Omega = [0,1]^2 = \bigcup_i \Omega_i$, where each Ω_i has a boundary of dim ≤ 1, and f is C^{γ_i} on Ω_i, then the best nonlinear approximation rate is $O(N^{-\tilde{\alpha}/2})$ where $\tilde{\alpha} = \min\{1, \min_i \gamma_i\}$

Measuring smoothness of images: an example

Log-log plot of ordered wavelet coefficients of the Lena image

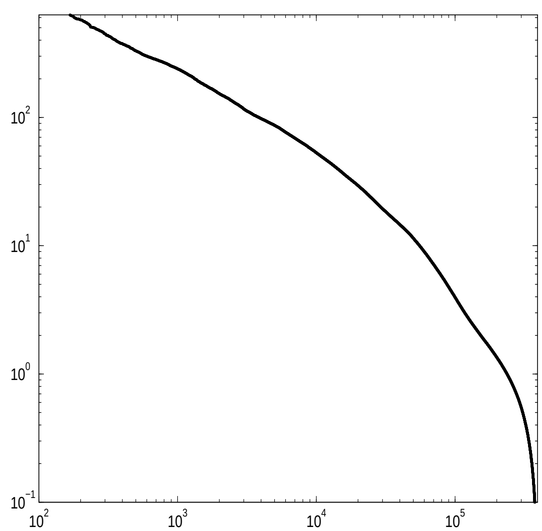


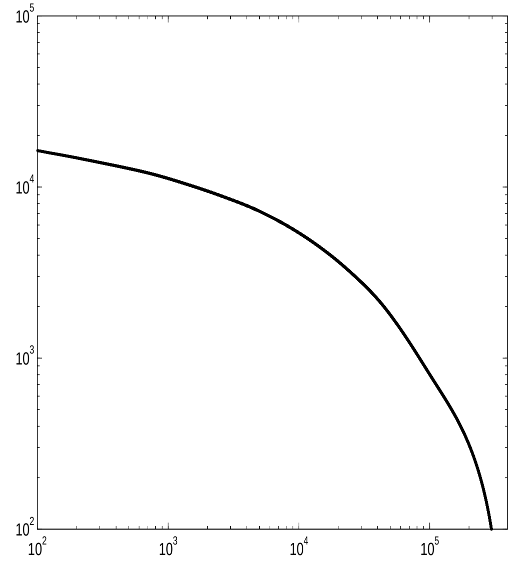
slope $\approx 0.8 \Rightarrow \epsilon(N) = O(N^{-0.3})$. So the image has 0.6 order of smoothness in $L^{5/4}$.

GLOBAL SMOOTHNESS SPACES MAY NOT BE THE COMPLETE EXPLANATION

Asymptotical aspects:

- There is not always a definite slope.





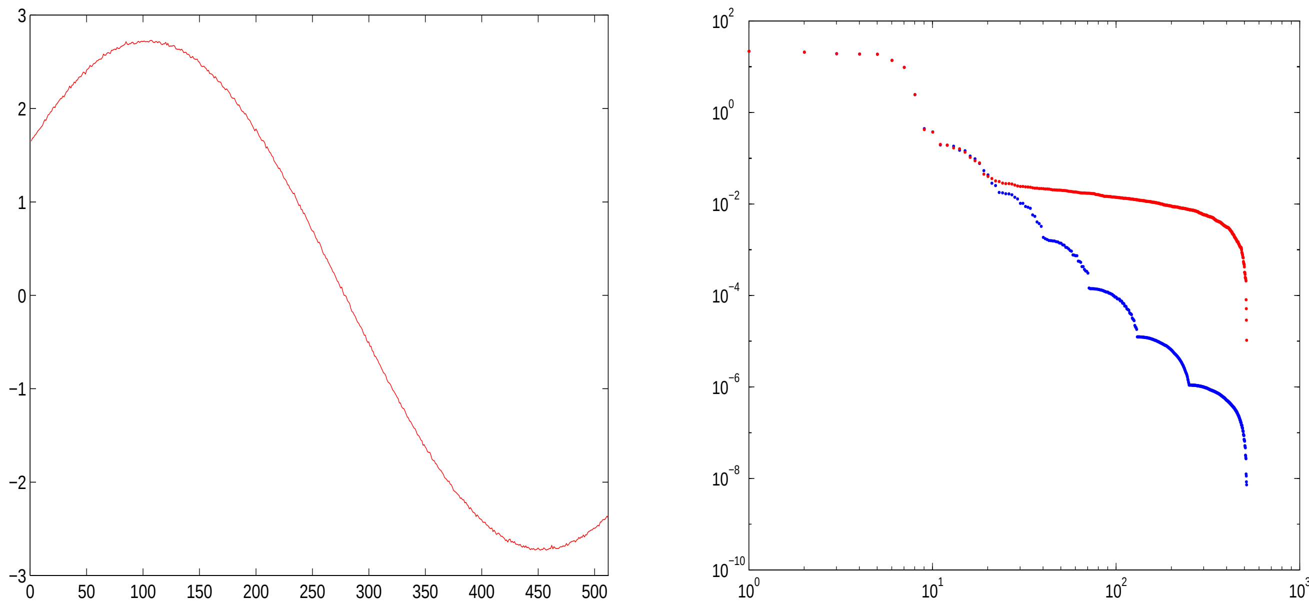
Log-log plot of ordered wavelet coefficients and nonlinear approximation error for the Window image

- Asymptotics can start “too late” or may not at all!

- The decay can be very slow in the presence of noise, even if the noise level is very small.

experiment:

Look at $f + \sigma\mathcal{N}(0,1)$ where f is smooth and $\sigma \ll \|f\|_2$.



Wavelets should not form an unconditional basis for the class of images:

- A classification using only the sizes of the coefficients is incomplete.

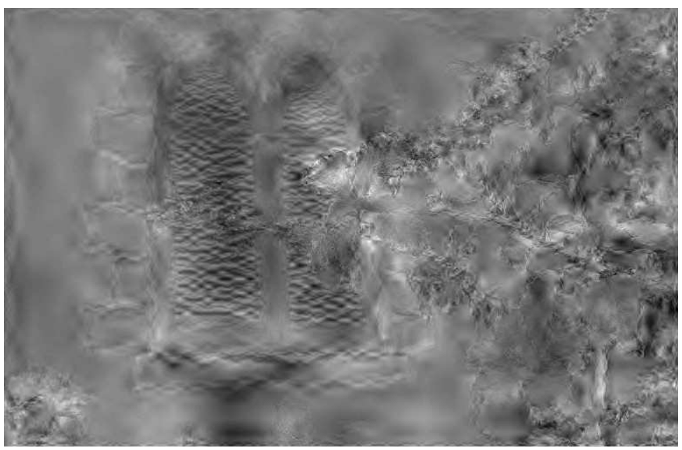
experiment:

- Given an image $f = \sum c_{j,k}\psi_{j,k}$, look at $\sum \epsilon_{j,k} c_{j,k}\psi_{j,k}$ where $\epsilon_{j,k} = \mp 1$, random.

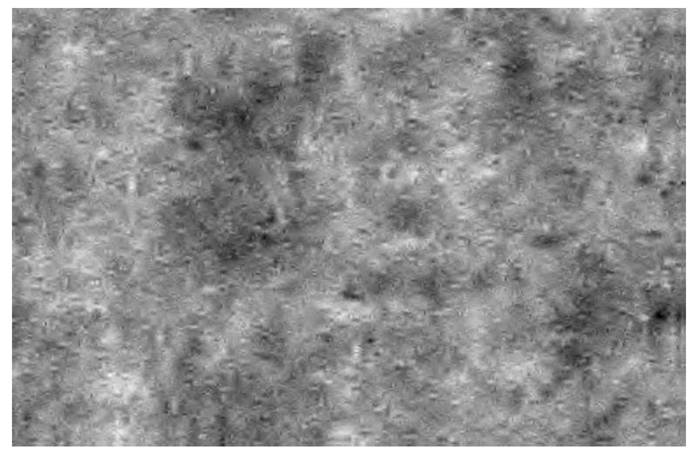
- $+$ shuffle the positions within the same scale.

Besov spaces are immune to these operations.

original window image



after random sign change



+ random shuffling within bands

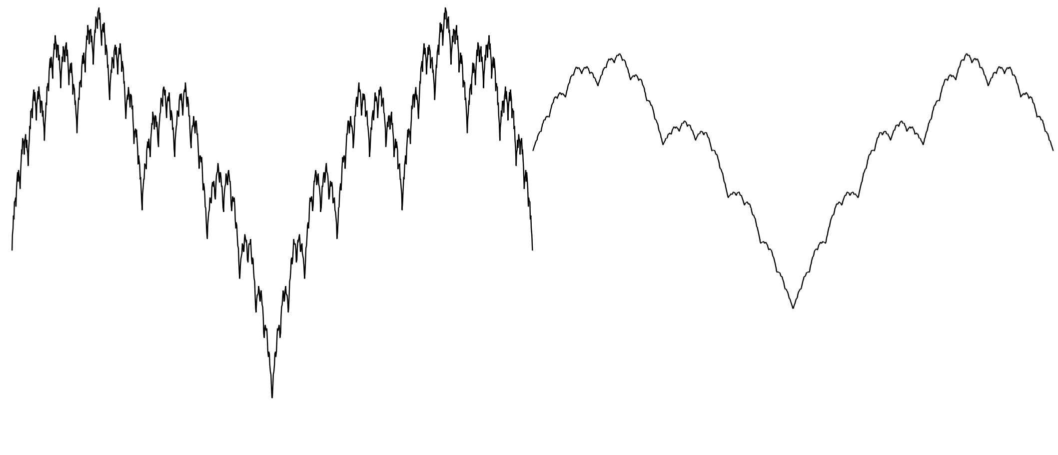
Mixtures:

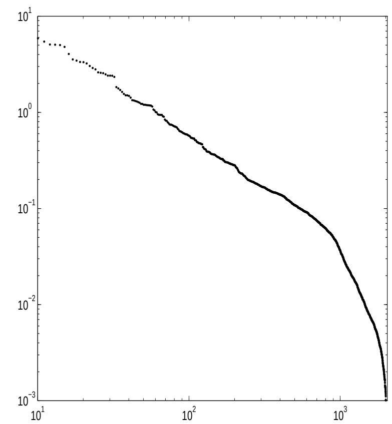
- A small region of very poor regularity (e.g. strong texture) can hide the "nice" regions, where one can exploit regularity. e.g. Lena image, hair + feather.

- In the discrete (finite) case, even very simple mixtures may result in not so simple decay profiles.

experiment:

- Consider a 1-D signal composed of two regions with **exact** C^{α_1} and C^{α_2} regularities: (say, by using lacunary Fourier series.)





If we had all the scales, the coefficient decay would forever be dominated by the smaller smoothness index, but in the finite setting, we run out of them and start seeing coefficients of the more regular region. However, since the curve is shifted, the log-log slope of the tail does not give the actual smoothness index of that region.

compare to: Sorting the mixture two finite sequences $\{n^{-\alpha_1}\}$ and $\{n^{-\alpha_2}\}$, $\alpha_1 < \alpha_2$, of length N would produce a tail that could be approximated by $\{(n-N)^{-\alpha_2}\}$. However, the log-log slope of this shifted curve is not constant any longer and changes from $-\infty$ to $-2\alpha_2$ on the corresponding interval $[N, 2N]$.

Finer tools?

Multifractal Formalism

Define pointwise Hölder spaces

$$C^{\alpha}(x_0) = \{f : \mathbf{R} \to \mathbf{R}, \exists \text{ poly. } P_{x_0} \text{ of deg } < \alpha \text{ s.t.}$$

$$|f(x) - P_{x_0}(x)| \le C|x - x_0|^{\alpha} \text{ about } x_0\}.$$

P_{x_0}: usually the Taylor poly. of f at x_0. define

$$\alpha(x_0) = \sup\{\alpha : f \in C^{\alpha}(x_0)\}$$

singularity spectrum (Hölder spectrum):

$d(\alpha) = \dim_{\text{Haus}} S_{\alpha}$, where $S_{\alpha} = \{x : \alpha(x) = \alpha\}$.

- How to compute the spectrum? - heuristic derivation:

structure function: $S_q(h) = \int |f(x+h) - f(x)|^q dx$

$$\alpha(x_0) = \alpha \quad \leftrightarrow \quad |f(x_0+h) - f(x_0)| \sim |h|^\alpha$$

so, $S_q(h) \sim \sum_\alpha |h|^{q\alpha} |h|^{-d(\alpha)} |h|,$ $\qquad S_q(h) \sim |h|^{\inf_\alpha(\alpha q + 1 - d(\alpha))}.$

If the computed $S_q(h) \sim |h|^{\zeta(q)}$, then $\zeta(q) = \inf(\alpha q + 1 - d(\alpha))$.

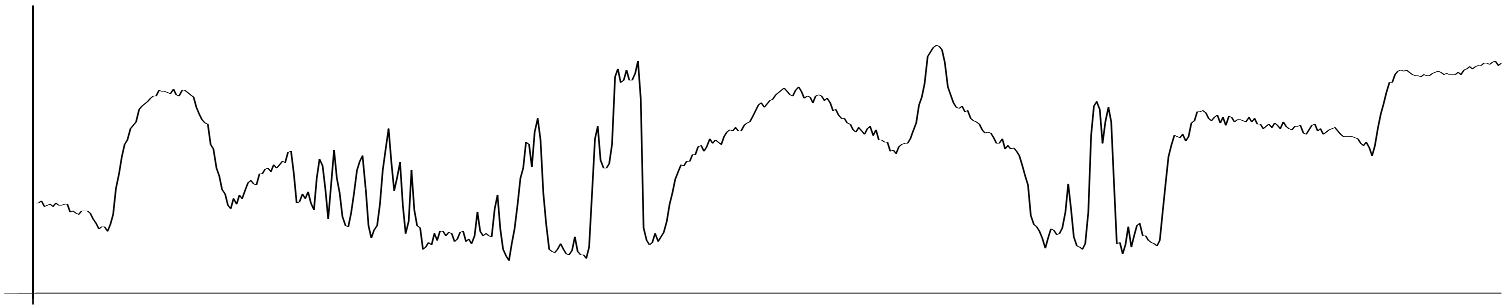
Inversion possible if $d(\alpha)$ is concave: $d(\alpha) = \inf(\alpha q + 1 - \zeta(q))$.

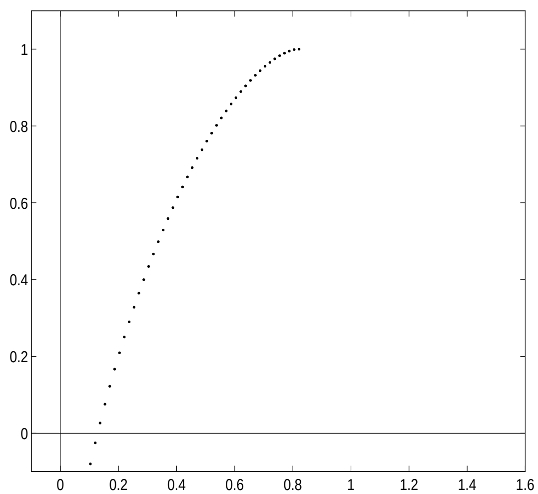
So the multifractal formalism consists of two parts none of which has to be true for a general function.

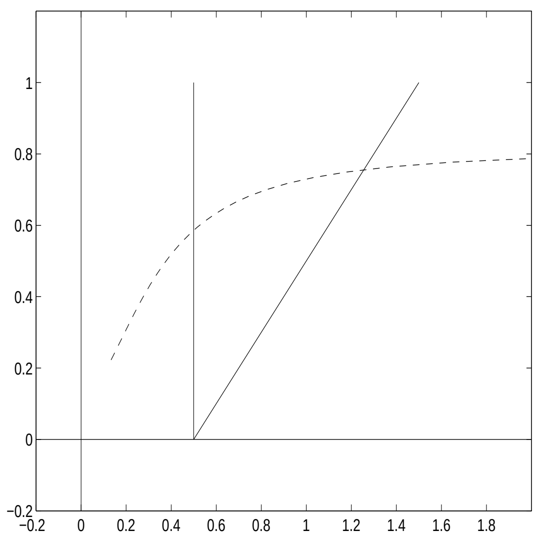
An Interpretation:

$$S_q(h) = \int |f(x+h) - f(x)|^q dx \sim |h|^{\zeta(q)}$$

$$\leftrightarrow \quad w_1(f,h)_{L^q} \le C|h|^{\zeta(q)/q} \quad \leftrightarrow \quad f \in B^{\zeta(q)/q}_{q,\infty}, \quad \forall q.$$




1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6


1
0.8
0.6
0.4
0.2
0
−0.2
−0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8

THE STOCHASTIC MODEL OF COHEN AND D'ALES: **What is the smoothness of realizations?**

Setting, theory and the model [2]:

- $f(t)$ be a stochastic process/random function on $[0,1]$ with autocorrelation $R(t,u)$.

- f_N be the first N-term approximation of f using the ONB $\{e_n\}_{n\geq 0}$ of $L^2([0,1])$. $\epsilon(N)^2 = \mathbf{E}\|f - f_N\|^2$ be the error measure.

- It is well known that the Karhunen-Loeve basis minimizes $\epsilon(N)$ among all ON bases.

- **Thm [2]: (Linear approx. with wavelets)** $R(t,u)$ be C^α on $\{(t,t), t \in [0,1]\}$ for some $\alpha > 0$. Let ψ have at least α vanishing moments. Then $\epsilon(N) = O(N^{-\alpha/2})$.

The piecewise smooth model:

jump points	$\rightarrow$	Poisson, density μ.
pieces	$\rightarrow$	independent, stationary with a common $r \in C^\alpha(0)$.

- Global process is also stationary.

- **Thm [2]: (Linear approx. of this process)** Using KL basis, $\epsilon(N) \geq C.N^{-1/2}$. (assuming $\alpha > 3/2$).

- **Thm [2]: (Nonlinear approx. of this process)** If r is C^α at the origin, and ψ has α vanishing moments, then $\epsilon(N) = O(N^{-\alpha/2})$ with the wavelet basis and nonlinear approximation. $(f_N = \mathbf{A}_N f)$

Deterministic (almost sure) properties of the model:

- **motivation:** (Kolmogorov's theorem) If a stochastic process f on $[0,1]$ satisfies, for some $\alpha, \beta > 0$,

$$\mathbf{E}|f(t) - f(s)|^{\alpha} \leq C.|t-s|^{1+\beta} \qquad \forall t, s$$

  then (there is a modification of f s.t.) f is continuous for almost all realizations. (in fact, $f \in C^{\gamma} \quad \forall \gamma < \frac{\beta}{\alpha}$)

- **more is true:** If $\mathbf{E}|\Delta_h^k f(t)|^p \leq C.|h|^{1+\gamma} \quad \forall t$, and for a suitable k, then almost surely, sample realizations are in $B_{p,p}^{\frac{1+\gamma}{p}-\epsilon} \quad \forall \epsilon > 0$.

- In particular, by the Sobolev embedding for Besov spaces, $B_{p,p}^{\frac{1+\gamma}{p}-\epsilon} \hookrightarrow B_{\infty,\infty}^{\frac{\gamma}{p}-\epsilon} = C^{\frac{\gamma}{p}-\epsilon}$ and one gets Kolmogorov's theorem.

- How does this relate to the model of Cohen and d'Ales?

Each segment of the process has autocorrelation r that is C^{α} at 0. From this one can deduce that $\mathbf{E}|\Delta_h^k f(t)|^2 \leq C.|h|^{\alpha}$ for a suitable k and all t. Hence from above, each segment is in $B_{2,2}^{\frac{\alpha}{2}-\epsilon}$ almost surely and the approximation of almost every realization of it satisfies $\epsilon(N) = O(N^{-\alpha/2+\epsilon})$. Since wavelets are used in the approximation, the whole realization has the same property. This means that the rate of expected approximation error is actually not superior to the rate valid for almost any realization.

remark: $\mathbf{E}|f|^2_{B_{2,2}^{\frac{\alpha}{2}-\epsilon}} = O(\frac{1}{\epsilon})$.

References:

[1] R.A. DeVore, B. Jawerth, and B.J. Lucier, *Image compression through wavelet transform coding*, IEEE Trans. Information Theory., Vol 38, pp. 719-746, March 1992.

[2] A. Cohen, J-P. d'Ales, *Nonlinear Approximation of Random Functions* SIAM J. Appl. Math., Vol 57, pp. 518-540, April 1997.