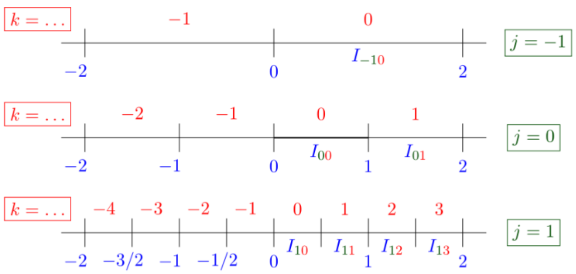
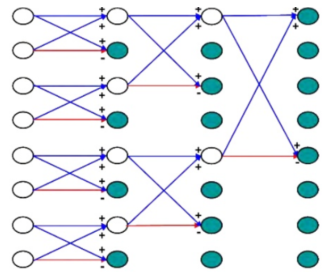


This text describes Assignment 1 (of 2) for the Master course Wavelets. Students are encouraged to work on this assignment in couples and to hand in together as a team, although this is not mandatory. Students who work together as a team get the same grade. For Assignment 2 you can choose again to work alone or with someone else, possibly a different person than for Assignment 1. Each of the two assignments counts for 15% towards your final grade.

Warning: to hand in together with somebody who did not contribute to the assignment with the intention to supply this person with a sufficient grade is considered fraudulent behavior that goes against academic integrity. Your lecturer is obliged to mention founded suspicions of fraud or plagiarism to the Examinations Board, which usually asks the students involved to appear before them. In case of sufficient evidence, it imposes sanctions.

So, please do not be tempted to "work together" with someone who either is not able or not willing to contribute to the solution of the assigment in a similar amount as yourself!

Uploading a self-contained video presentation

You are asked to compose and edit a coherent video in which you address the issues and questions that are raised in the assignment text. Perhaps the best way to do so is to consider it as a seminar presentation of your team in front of an audience having the same background as yourself, in the sense that they followed the same lectures, but that is not aware of the assignment. In fact, try not to mention you're doing an assignment, but simply share with the audience a piece of interesting material. You are invited to use LaTeX slides in combination with demonstrations of your computer codes and discussions of the coding itself, in order to show you understand the theoretical exercises, that you did the coding, and that the codes are actually functioning and producing numbers and pictures. Try to make the video self-contained: do not just answer the questions that are described in the exercises, but briefly and to the point sketch the considerations that led to the questions. As in any presentation, have a short introduction of what you're going to do and why. The expected length of the video is roughly 25 minutes. You are allowed to do editing in any way you like, gluing together smaller parts if desired. If you work as a couple, make sure that both of you have equal roles in the video, like being visible and talking about the work approximately equally much.

Disclaimer: As much as I have tried to make this text accurate and correct, it may well be that typo's, smaller errors, and larger errors have entered it. If you have doubts about anything in the text, please let me know as quickly as possible in order to prevent potential damage as much as possible. Many thanks!

And have fun! ☺ Jan Brandts

1. The Discrete Haar Transform and the Fast Haar Transform

For integer $j \geq 0$, let $n = 2^j$ and consider the points $x_{jk} = k2^{-j}$ for $k \in \{0, \ldots, n\}$. Write $I_{jk} = [x_{jk}, x_{j,k+1})$ and define the space

$$\mathcal{S}_j^0 = \Big\{ v : [0,1) \to \mathbb{R} \;:\; \forall k \in \{0, \ldots, n\} : \quad v_{|I_{jk}} \in \mathcal{P}^0(I_{jk}) \Big\}$$

where $\mathcal{P}^\ell(X)$ is the space of all polynomials on X of degree at most ℓ.

1.1. The standard basis of $\mathcal{S}_j^0$

With $\chi : \mathbb{R} \to \mathbb{R}$ the characteristic function of $[0,1)$, set

$$\chi_{jk}(x) = 2^{j/2}\chi(2^j x - k)$$

for all $j \geq 0$ and $0 \leq k \leq n-1$. Then χ_{jk} is the positive multiple of the characteristic function of I_{jk} with L_2-norm equal to one. Hence, for each $j \geq 0$ the elements of

$$\mathcal{B}_j = \{\chi_{jk} \;:\; 0 \leq k \leq 2^j - 1\}$$

form an orthonormal basis for $\mathcal{S}_j^0$. We will call it the standard basis of $\mathcal{S}_j^0$.

Observe that we're being slightly inaccurate here. As defined above, the functions χ_{jk} have domain $\mathbb{R}$, whereas we interpret them as basis functions for a space of functions with domain $[0,1)$. We will not be pedantic about this (although usually we enjoy that very much ☺).

1.2. The Haar basis of $\mathcal{S}_j^0$

With $H : \mathbb{R} \to \mathbb{R}$ the Haar function

$$H = \frac{1}{2}\sqrt{2}\,(\chi_{10} - \chi_{11}),$$

which, due to the scaling of χ_{jk} above, indeed takes values 1 and -1 on $[0,1)$, set

$$H_{jk}(x) = 2^{j/2}H(2^j x - k)$$

for all $j \geq 0$ and $0 \leq k \leq n-1$. Then, for each $j \geq 0$ the elements of

$$\mathcal{H}_j = \{\chi\} \cup \{H_{ik} \;:\; 0 \leq i < j, \quad 0 \leq k \leq 2^i - 1\}$$

form an orthonormal basis for $\mathcal{S}_j^0$ as well. This basis is called the Haar basis.

A quick count of the elements of $\mathcal{H}_j$ should convince us that we got the indices correct. Given i there are 2^i elements H_{ik} in the right-hand (of the $\cup$) set. Because

$$\sum_{i=0}^{j-1} 2^i = 1 + 2 + 4 + \ldots + 2^{j-1} = 2^j - 1$$

together with χ this gives indeed 2^j elements in $\mathcal{H}_j$. So at least we got that right ☺

1.3. Coordinate vectors and coordinate trees

Given a function v, we will write

$$c_{jk} = \langle v, \chi_{jk} \rangle \quad \text{and} \quad h_{jk} = \langle v, H_{jk} \rangle, \quad h = \langle v, \chi \rangle \quad (= c_{00}).$$

If we develop $v \in \mathcal{S}_j^0$ in the standard basis,

$$v = \sum_{k=0}^{2^j-1} c_{jk} \chi_{jk}$$

we see that only functions from one level, level j, are needed. Developing v in the Haar basis

$$v = h\chi + \sum_{i=0}^{j-1} \sum_{k=0}^{2^i-1} h_{ik} H_{ik}$$

shows that (in addition to χ) functions from j increasingly finer levels $0, 1, \ldots, j-1$ are used. The Haar basis is therefore called a multi-level basis.

The set of coordinates c_{jk} of $v \in \mathcal{S}_j^0$ can conveniently be put in a list, a vector if you wish, due to its linear structure. The coordinates h_{ik} with $0 \le i < j$ and $0 \le k \le 2^i - 1$ are most naturally represented as a binary tree, where the node $h_{i\ell}$ has children $h_{i+1,2\ell}$ and $h_{i+1,2\ell+1}$.

1.4. The Discrete Haar Transform and its computational complexity

Here is the definition of the Discrete Haar Transform.

The Discrete Haar Transform (DHT) is the transformation of the coordinates of a function $v \in \mathcal{S}_j^0$ with respect to the standard basis into its coordinates with respect to the Haar basis.

As with any basis transformation of a finite dimensional space, it can be represented by a matrix T_j, which in this current setting will have size $2^j \times 2^j$.

Exercise 1. Compute T_2 by hand. For this, you need to fix an order of the four elements of the Haar basis. Also give the sparsity pattern of T_3, i.e. indicate which entries of this matrix are non-zero. Derive an expression for the number of non-zero elements of T_j.

Of course, the number of nonzero elements of the transformation T_j determines how many multiplications of two scalars need to be computed in order to perform the transformation. We will use this number of multiplications of scalars as a measure of its complexity.

Exercise 2. What is the order of this complexity in terms of the size $n = 2^j$ of the matrix? Recall that generally a matrix-vector product requires $\mathcal{O}(n^2)$ multiplications.

A coordinate transformation is not complete without its inverse transformation. Therefore:

Exercise 3. Answer the questions in Exercise 1 and 2 also for the inverse transformations.

1.5. The Fast Haar Transform and its computational complexity

Instead of working with the basis transformation matrices T_j, one may alternatively try to implement the transformation level by level. Here we will describe what we mean by that.

Using that $I_{jk} = I_{j+1,2k} \cup I_{j+1,2k+1}$ we can derive that

$$v = c_{j+1,2k} \cdot \chi_{j+1,2k} + c_{j+1,2k+1} \cdot \chi_{j+1,2k+1}$$

if and only if

$$v = \frac{1}{2}\sqrt{2}(c_{j+1,2k} + c_{j+1,2k+1}) \cdot \chi_{jk} + \frac{1}{2}\sqrt{2}(c_{j+1,2k} - c_{j+1,2k+1}) \cdot H_{jk}$$

where the factors $\frac{1}{2}\sqrt{2}$ are a consequence of the chosen normalization of the basis functions. Therefore,

$$c_{jk} = \frac{1}{2}\sqrt{2}(c_{j+1,2k} + c_{j+1,2k+1}) \quad \text{and} \quad h_{jk} = \frac{1}{2}\sqrt{2}(c_{j+1,2k} - c_{j+1,2k+1}).$$

This shows that the coordinates c_{jk} with $0 \le k < 2^j - 1$ of $v \in \mathcal{S}_j^0$ can be transformed into coordinates $c_{j-1,k}$ with $0 \le k < 2^{j-1} - 1$ of an element in $\mathcal{S}_{j-1}^0$ and coordinates $h_{j-1,k}$ with $0 \le k < 2^{j-1} - 1$ of an element in $\mathcal{W}_{j-1}^0$, the orthogonal complement of S_{j-1}^0 in $\mathcal{S}_j^0$.

The repetitive application of the above computation scheme yields the coordinates of $v \in \mathcal{S}_j^0$ with respect to the Haar basis, given its coordinates with respect to the standard basis. This implementation of the transformation is called the Fast Haar Transform (FHT). Its inverse, that computes $c_{j+1,2k}$ and $c_{j+1,2k+1}$ from $c_{j,k}$ and $h_{j,k}$ can be easily derived from the above.

Exercise 4. Count the number of multiplications of scalars that is needed to descend from level j to $j-1$ and use this to derive an expression for the total number of multiplications of scalars that is needed to perform the complete basis transformation for $v \in \mathcal{S}_j^0$. Express it in terms of j, but also give the complexity as an order of $n = 2^j$, as in Exercise 2.

Exercise 5. Write four computer codes that implement the DHT and the FHT and their inverses. Both codes should only perform the multiplications with scalars that you have counted in previous exercises. In particular, do not implement them as matrix-vector multiplication, as this includes multiplications by zeros that increase the complexity unnecessarily!

To test your codes, you need input data. For this, write a simple code that given a continuous function f on $[0,1]$ samples its 2^j values at the midpoints of the intervals I_{jk} for some given j. These function values can be associated with an approximation $f_j \in \mathcal{S}_j^0$ of f. Note that the function values of f_j differ by a normalization factor from the coordinates c_{jk} of f_j.

Exercise 6. Although not strictly necessary, it is very helpful to be able to visualize functions in $\mathcal{S}_j^0$, such as f_j, together with f itself, for the simple reason that it's often easier to see if things work out fine by a picture than by staring at hundreds of coefficients. So, support your codes by suitable visualization tools that help to understand what is going on.

2. Lossy data compression by the FHT

The standard basis $\mathcal{B}_j$ for $\mathcal{S}_j^0$ is very simple, and to compute the coordinates c_{jk} of a given $v \in \mathcal{S}_j^0$ with respect to $\mathcal{B}_j$ it suffices to evaluate v on each interval and apply a proper scaling. This shows that these coordinates contain information about the local magnitude of v. The coordinates w_{jk} with respect to the Haar basis $\mathcal{H}_j$, alternatively, give information about the local variation of v, which has to do with smoothness. Here we will investigate that property.

2.1. Decay of coefficients in relation to smoothness

A major advantage of expressing a given function f in the Haar basis is that the magnitudes of the coordinates h_{jk} of f tend to zero as j increases with a speed that depends on the smoothness of f. In other words, if f is smooth, the majority of the information[1] it carries is contained in the lower levels (smaller values of j).

Exercise 7. Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is called Hölder continuous with coefficient $\alpha \in (0, 1]$ if there exists a $C \in \mathbb{R}$ such that for all $x, y \in \mathbb{R}$,

$$|f(x) - f(y)| \leq C|x - y|^\alpha.$$

Prove that if f is Hölder continuous with coefficient $\alpha \in (0, 1]$ then there exists a $C \in \mathbb{R}$ such that

$$|h_{jk}| = |\langle f, H_{jk}\rangle| \leq C 2^{-j(\alpha+1/2)} \tag{1}$$

for all elements H_{jk} of the Haar basis.

The above result can be computationally verified using the available codes and a little more administration. Given some f, a correct way to compute all h_{ik} with $0 \leq i \leq j - 1$ is:

- fix a value of j;
- compute the orthogonal projection $P_j f$ of f on S_j^0;
- express $P_j f$ in the standard basis;
- apply the FHT to get the coefficients h_{ik} for each level $0 \leq i \leq j - 1$.

Indeed, because $H_{ik} \in S_j^0$ for all $0 \leq i < j$ we have by definition of orthogonal projection that

$$\langle f - P_j f, H_{ik}\rangle = 0$$

and thus the coefficients h_{jk} of f and $P_j f$ coincide for all levels $0 \leq i < j$.

The above recipe is straightforward apart from computing $P_j f$, which requires to integrate f over each interval I_{jk}. It is not always possible to do so exactly, so we choose to approximate $P_j f$ by the function f_j from Exercise 5, which interpolates f at the midpoints of I_{jk}. As

$$\|f_j - P_j\|_2 = \mathcal{O}(2^{-2j})$$

this leads to a perturbation that is of higher order than the right-hand side in (1) and will become negligible when j becomes large enough. The proof of the above estimate we omit.

[1] Basically, its L_2-norm, which is the square root of the sum of the squared coefficients

Exercise 8. Fix $j = 10$. Use the above recipe with $P_j f$ replaced by its simple approximation f_j to compute[a] all coefficients h_{ik} with $0 \le i \le j-1$. For each such i, set

$$m_i = \max_{0 \le k \le 2^i - 1} \{|h_{ik}|\}$$

the maximum absolute coefficient on level i. Because the numbers m_i of course also satisfy (1), their consecutive quotients m_i/m_{i+1} should tend to $2^{(\alpha+1/2)}$.

Verify this for $f(x) = \cos(2\pi x)$, which has Hölder smoothness $\alpha = 1$ and also for $g(x) = \sqrt{|\cos(2\pi x)|}$ which has Hölder smoothness $\alpha = \frac{1}{2}$. We depict g in Figure 1.

[a] or better phrased: to approximate, up to higher order perturbation, because $P_j f$ is replaced by f_j

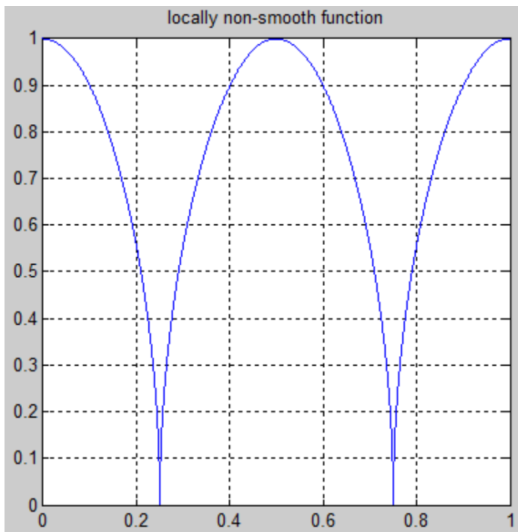


Figure 1. The graph of $g(x) = \sqrt{|\cos(2\pi x)|}$, which is Hölder $\frac{1}{2}$-continuous.

Since g is only locally non-smooth, this non-smoothness will manifest itself only in the coefficients h_{jk} of H_{jk} whose support contain that non-smoothness. For g this will concern the h_{jk} on intervals around the cusps of its graph at $x = \frac{1}{4}$ and $x = \frac{3}{4}$. This is shown in Figure 2.

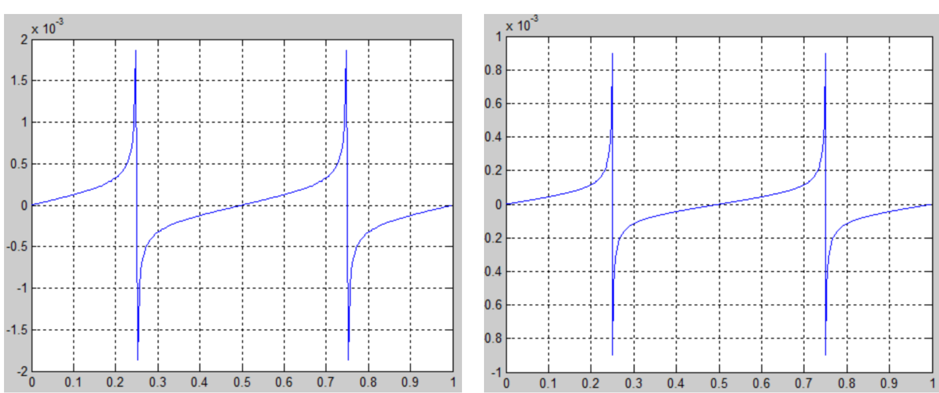


Figure 2. Left: the 256 coefficients h_{jk} for $j = 8$. Right: the 512 coefficients h_{jk} for $j = 9$. Note that the scaling on the vertical axis differs by a factor of two in both pictures.

On the left in Figure 2, we see for $j = 8$ the 256 coefficients h_{jk} against the position of the supports I_{jk} of H_{jk} within the interval $[0, 1]$. On the right the same for $j = 9$. The largest coefficients in the right picture are indeed approximately a factor two smaller than in the left, as predicted by (1). The sharper localization of the peaks in the right picture also suggest that the coefficients in the smoother region became smaller by a factor more than two. Indeed, by (1) they should become a factor $2\sqrt{2}$ smaller, as g is differentiable on these intervals.

Exercise 9. Reproduce the pictures in Figure 2, and add the picture belonging to $j = 7$.

2.1. Relation with vanishing moments

Recall that the number of vanishing moments of any wavelet W is 1 plus the largest integer ℓ such that

$$\int_{\mathbb{R}} x^q W(x)dx = 0 \quad \text{for all integer } q \text{ with } \quad 0 \le q \le \ell.$$

In other words, W is orthogonal to the space $\mathcal{P}^\ell(\mathbb{R})$ of all polynomials up to degree ℓ.

Exercise 10. Show that this property is invariant under translation and dilation, and hence that all elements

$$W_{jk}(x) = 2^{j/2}W(2^jx - k)$$

of the wavelet basis of $L_2(\mathbb{R})$ that W induces are also orthogonal to $\mathcal{P}^\ell(\mathbb{R})$.

Assume now additionally that W has a compact support K with width κ. Then W_{jk} has support K_{jk} with width $2^{-j}\kappa$, and for all $p \in \mathcal{P}^\ell(I)$ we find

$$|w_{jk}| = |\langle f, W_{jk}\rangle| = |\langle f - p, W_{jk}\rangle| \le \|f - p\|_{2,K_{jk}}\|W_{jk}\|_{2,K_{jk}} = \|f - p\|_{2,K_{jk}}.$$

If f is ℓ times continuously differentiable, choosing for p the best L_2-approximation of f on K_{jk} gives that

$$\|f - p\|_{2,K_{jk}} \le C|K_{jk}|^\ell\|f^{(\ell)}\|_{2,K_{jk}} \le C|K_{jk}|^{(\ell+\frac{1}{2})}\|f^{(\ell)}\|_{\infty,K_{jk}} \le C2^{-j(\ell+\frac{1}{2})}\|f^{(\ell)}\|_{\infty,K_{jk}}$$

where the first bound is standard approximation theory and the second bound comes from the assumed continuity of $f^{(\ell)}$ and switching to the uniform norm. Finally, combining the above,

$$|w_{jk}| \le C2^{-j(\ell+\frac{1}{2})}\|f^{(\ell)}\|_\infty$$

generalizing the result in Exercise 7 whenever $\ell > 1$. Thus, decay of the coefficients w_{jk} increases with the smoothness of f and the number of vanishing moments of the wavelet W.

2.3. Approximation by the best N-term approximation

Consider again the function $g(x) = \sqrt{|\cos(2\pi x)|}$. For given j we can compute g_j, the approximation of P_jg that does not integrate g over each interval I_{jk} but simply uses its values at the midpoints of the I_{jk}. We denote the coordinates of g_j with respect to the standard basis $\mathcal{B}_j$ by c_{jk}, and use the FHT to transform them to the coordinates h_{jk} of g_j with respect to the Haar basis $\mathcal{H}_j$. Then

$$\|g_j\|_2^2 = \sum_{k=0}^{2^j-1} c_{jk}^2 = c_{00}^2 + \sum_{i=0}^{j-1}\sum_{k=0}^{2^i-1} h_{ik}^2.$$

Contrary to the coordinates c_{jk}, the coordinates h_{jk} decay to zero as the level j increases. Hence, in the Haar basis one can distinguish between so-called modes $h_{jk}H_{jk}$ that matter more than others, in terms of size. In the left picture of Figure 3 and for $j = 13$ we display the coordinates c_{00} and h_{ik} with $0 \leq i \leq j-1$ and $0 \leq k \leq 2^i - 1$ sorted according to (absolute) magnitude. As visible at the far left, there is a small number of relatively large coordinates.

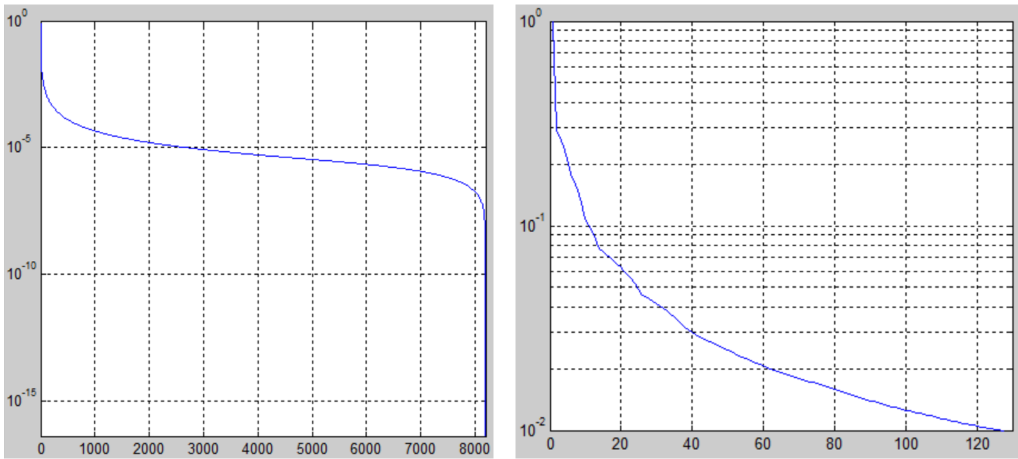


Figure 3. Left: sorted Haar basis coordinates. Right: error of best N-term approximation.

It is not hard to see that the sum of the N modes of the form $h_{jk}H_{jk}$ corresponding to the largest coefficients h_{jk} is the best approximation with N such modes that can be select from the 2^j modes in total. Indeed, its error equals the ℓ_2-norm of the vector of the remaining (or unused) coefficients, which by definition are the $2^j - N$ smallest. The approximation formed by the N modes with the largest N coefficients is called the best N-term approximation.

In the right picture of Figure 3 we display the relative error of the best N-term approximation. We see that 99% of g_j is captured by its largest 125 out of 8192 modes: just 1.5% of them! The Haar basis tells us which ones they are, and the FHT computes them in optimal complexity.

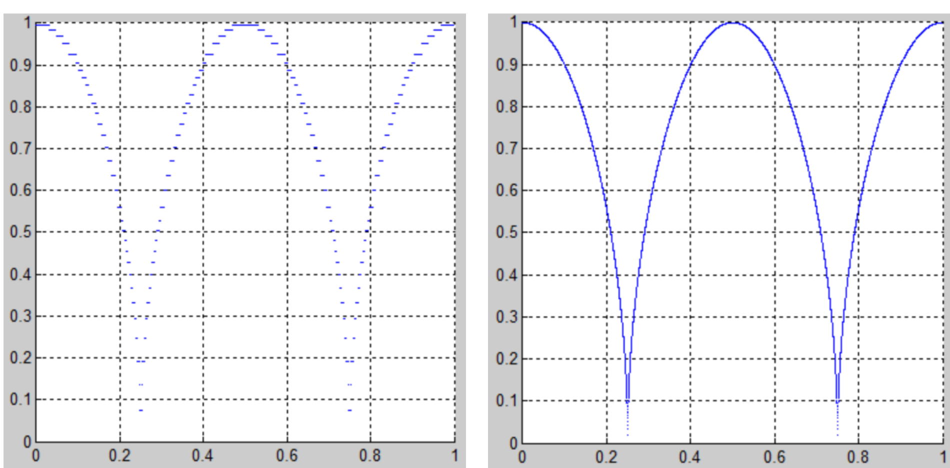


Figure 4. Left: best 125-term approximation of the original 8192-term function g_{13} (right).

In the right picture of Figure 4 we plot g_{13} and in the left its piecewise constant approximation that corresponds to its best 125-term approximation. To explicitly compute it, simply put all coefficients in the Haar basis other than the 125 largest to zero, and perform the inverse FHT. This yields its coordinates with respect to the standard basis. It can be interpreted as a piecewise constant function relative to a partition of dyadic intervals, not all of same size.

Exercise 11. Consider the function $f(x) = \arctan(50(x - \frac{1}{3}))$ on $[0, 1]$ depicted below.

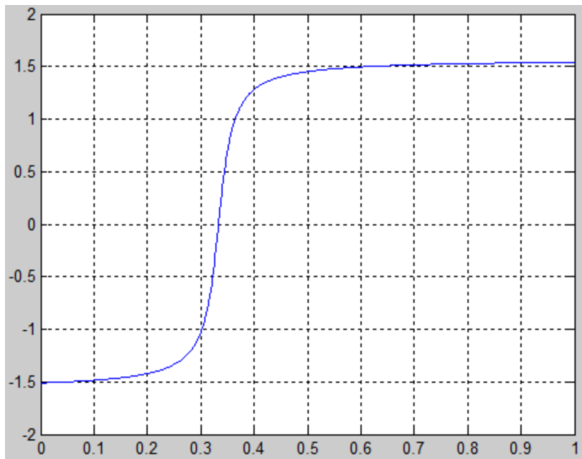


Figure 5. Graph of $f(x) = \arctan(50(x - \frac{1}{3}))$.

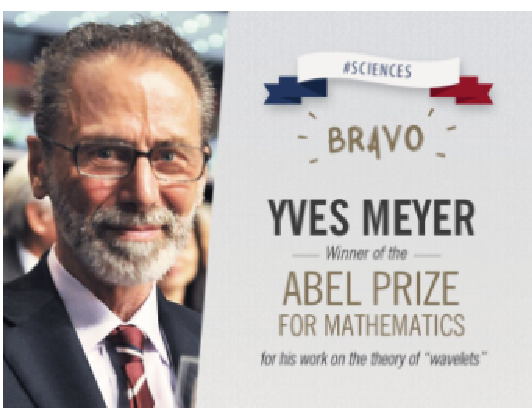
Using this function to reproduce some experiments from the previous sections. Concretely:

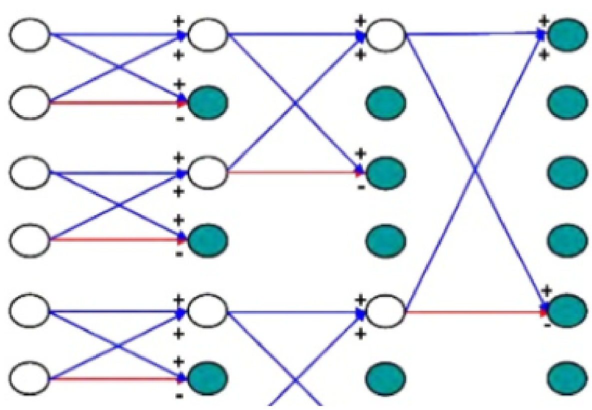
- for $j = 13$, approximate $P_j f$ with f_j;
- compute the coefficients c_{jk} of f_j with respect to $\mathcal{B}_j$;
- use the FHT to transform them into h_{ik} for $0 \leq i \leq j - 1$;
- verify how their magnitudes descend by computing the numbers m_i from Exercise 8;
- sort their absolute values in descending order and construct the equivalent of Figure 13;
- construct the best N-term approximation of f_j that captures 99% of $\|f_j\|_2$;
- what is the corresponding value of N, what percentage of coefficients does it use?
- set the remaining coordinates with respect to $\mathcal{H}_j$ to zero and transform back to $\mathcal{B}_j$;
- as in Figure 4, depict f_j and its best N-term approximation determined above.

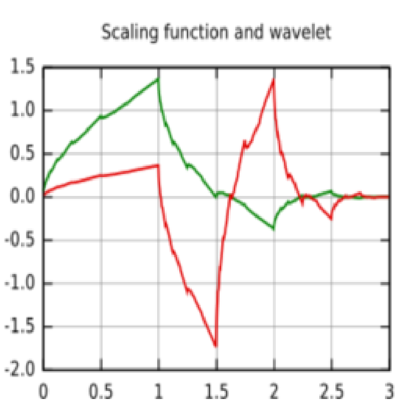
2.4. Concluding remarks

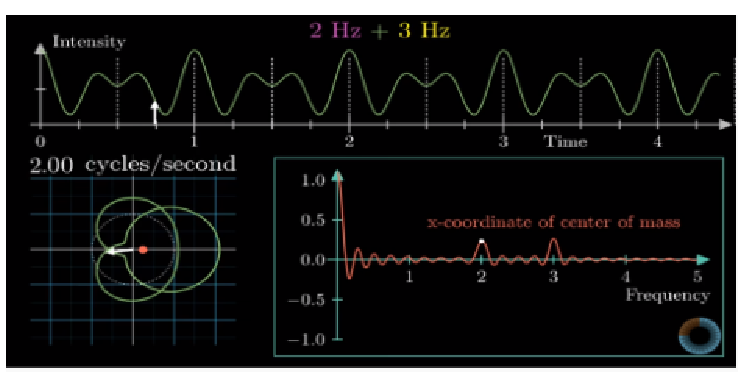
We have seen that expressing a function f in the Haar basis leads to the possibility to distinguish between modes that contribute more to f than others, as a result of the decay properties of its coefficients. As a consequence, it is possible to construct the best N-term approximations of $f_j \in \mathcal{S}_j^0$. Observe that this best N-term approximation of f_j can only be computed after f_j itself has been computed. A valid question is therefore why one would be interested in doing so. The answer is that it reduces storage and facilitates transmission. After having invested the energy to compute a suitable best N-term approximation, the original data f_j can be removed. Also, once sender and receiver have agreed on using a certain wavelet basis, the potential transmission of large amounts of data reduces to transmission of quite a small percentage of that data. The transformation to the wavelet basis followed by discarding small coefficients should be seen as a fairly automated way to compress data. This compression is of the lossy type in the sense that the original signal can never be fully recovered. If it has sufficient quality, this need not be a problem. Due to the local supports of the Haar wavelet basis functions, they are able to pick out local non-smoothnesses, something that for instance a Fourier transform is not well able to do. On the other hand, due to its discontinuity and its limited number of vanishing moments (being the bare minimum of one), the Haar wavelet is a very primitive example of what a wavelet can do in this context.

A good observer will have noticed that we have neglected to mention the computational costs of $\mathcal{O}(n\log(n))$, where $n = 2^j$, to sort the Haar wavelet coefficients in descending order, and to use the largest N of them to construct the best N-term approximation. Of course, the logarithmic factor $j = \log(n)$ for the values of j we have worked with above, will not really hurt in case all computations go within a split second, but is nevertheless present. There are two work-arounds in practice. The first one is not to sort all the coefficients element-by-element, but distribute them over pre-selected bins whose widths depend on $\|f_j\|$, a number which can be computed in $\mathcal{O}(n)$, and the expected decay of the coefficients. Because the interest is in a small number of large elements, it does not pay off the sort those elements who will be discarded anyhow. There exist reasonable strategies that reduce the complexity to $\mathcal{O}(n)$, sometimes at the cost of finding only an almost best N-term approximation. A second way to avoid sorting, somewhat more blunt, is to use thresholding and to discard all coefficients that are smaller than some threshold. Of course one should then not spend more than $\mathcal{O}(n)$ in selecting the threshold.









But what is the Fourier Transform? A visual introduction.