Computation with Wavelets in Higher Dimensions

Jan-Olov Strömberg[1]

Abstract.

In dimension d, a lattice grid of size N has N^d points. The representation of a function by, for instance, splines or the so-called non-standard wavelets with error ϵ would require $O(\epsilon^{-ad})$ lattice point values (resp. wavelet coefficients), for some positive a depending on the spline order (resp. the properties of the wavelet). Unless d is very small, we easily will get a data set that is larger than a computer in practice can handle, even for very moderate choices of N or ϵ.

I will discuss how to organize the wavelets so that functions can be represented with $O((\log(1/\epsilon))^{a(d-1)}\epsilon^{-a})$ coefficients. Using wavelet packets, the number of coefficients may be further reduced.

1991 Mathematics Subject Classification: Primary Secondary
Keywords and Phrases: Wavelets, high dimension .

1 Introduction

Although we live in a three-dimensional space it is often useful to consider spaces with much higher dimension. For example, describing the positions in a system with P particles we may use a $3P$ dimensional space. Although we in theory can work very high dimesnsion, it is very limited what we can do in practical numerical computations. The are properties of the geometry in very high dimension that may be surprising.. For example, consider a geometric object as simple as a cube in $\mathbf{R}^d$. A cube with side length as small as a finger nail may still contain a sets as large as the earth on a three-dimensional subspaces, provided d is large enough.

The fundamental issue of analysis in high dimensions involves the approximation to prescribed accuracy of transformations of high dimensional data. Approximating functions with a grid of size N in dimension d means that we have the N^d grid points with data. With the limited amount of data we can handle in

[1]Professor at University of Tromsø, Norway until 1998. Supported by Norwegian Research Counsil

practice, this imposes strong restrictions both on N and on d. The current state of approximation theory is essentially useless in dimensions larger than 10.

One may think that this problem would be solved with faster and faster computers. But there are limitations how fast computers can be.. Let us illustrate this by the following example of "ultimate massive parallel super computer": (I have taken physical constants from a standard physics handbook.)

Let the number or parallel processors be as many as the estimate of total number of protons in universe, let clock cycle speed on each processor defined by the time to travel the distance of a nuclear radius at the speed of light, and finally let running time be as long as estimated live time of universe. Totally this will be about 10^{120} cycles. This correspond to a the number of grid points N^d with $N = 256$ and with $d \approx 50$, For a systems or P particles ($d = 3P$), this means that P can not be larger than 17. Let f function be in the unit cube in $\mathbf{R}^{3P}$ with first order derivatives bounded by 1. Then f may approximated in this grid with accuracy $1/10$ (when $P = 17$). In reality, it seems to be beyond our reach to handle $P > 3$ or maybe even $P > 2$ particles.

The theory for numerical computation in high dimensions is in a premature state, but the approach of Jones; Davis and Semmmes([5]), is the first indication that a powerful theory for high dimensions exists.

In the rest of this paper, I will discuss some ideas which in practice are useful only in rather low dimensions (≤ 10).

In recent years wavelet methods have appeared as useful tools for reducing complexity in numerical computations. By expanding functions in wavelet coefficients one has been able to compress the data to be handled. For example, consider Singular Integral Operator on functions on $\mathbf{R}$ bounded on L^2 and with kernel $K(x, y)$. Assume the kernel satisfies the standard decay properties away from the diagonal:

$$|\partial_x^\alpha \partial_y^\beta K(x,y)| \leq C|x-y|^{-d-\alpha-\beta} \tag{1}$$

for $\alpha, \beta| \leq m$. Representing the operator with a $N \times N$ matrix, we need to use essentially all N^2 matrix elements, even if the elements far away from the diagonal are very small - the total contribution of all the matrix elements on distance $\approx 2^j$ from diagonal will not decay by j. In the famous paper by Beylkin, Coifman, and Rokhlin([2], the authors has shown how the Singular Integral Operators can be expressed in a wavelet basis with a matrix, where the elements decays much faster away from the diagonal. In fact, if the accuracy level is ϵ one may use a diagonal band limited matrix with bandwidth proportional to $\epsilon^{\frac{1}{m}}$. Here m is the order of the wavelets. Thus one need only to use $NO(\epsilon^{-\frac{1}{m}})$ non-zero matrix elements.

In dimension $d > 1$ one has often used the so called non-standard tensor extension , using tensor product combinations of the one-dimensional wavelets and and its scale functions, with all factors of the same scale. Let $M = N^d$, the number of grid points in which the functions arthe represented. Instead of using all M^2 elements to represent the Singular Integral Operator, one need only use a matrix with non-zero element limited to a band around the diagonal ($x = y$), For accuracy level ϵ one need to use $MO(\epsilon^{-\frac{d}{m}})$ non-zero matrix elements. When d is large and m small the number of terms $O(\epsilon^{-\frac{d}{m}})$ increases very fast as ϵ gets smaller.

Even in as low dimension as $d = 3$,4 or 5 we feel the restriction on how small ϵ may be. We shall see than under certain circumstances, with good control of the mixed variation of f, the exponential dependence of d in $O((1\epsilon)^{\frac{d}{m}})$ may be replaced with the somewhat better, but still exponential expression; $O(\log\frac{1}{\{}\epsilon^{\frac{1}{m}})^{d-1}\epsilon^{-\frac{1}{m}})$. The error in the approximation, in sup norm, is of magnitude $O(\log\frac{1}{\{}\epsilon^{\frac{1}{m}})^{d-1}\epsilon)$. More exactly, if the smoothness condition on f is stated with the expression $\alpha_1 + \cdot +$ $\alpha_d <= m$ replaced by $\max \alpha_i \leq m$. Here $\alpha = (\alpha_1, .., \alpha_d)$ is the multi-index for the derivative $D^\alpha f$. This is a smoothness condition that is especially suitable to use for functions which are tensor products as $f_1(x_1)\cdots f_d(x_d)$, or for functions that behaves almost like such tensor products.

The ideas ,which are presented here ,comes from some very trivial observations I did trying to work with wavelets in dimension $d > 2$: First, as said above, we very easily get a terrible amount of data.

Second, the full tensor extension of the wavelets to higher dimensions seems to give better compression of the data than the, now classical, non-standard tensor wavelet extension.

This is certainly not a new observation. There have, for example, been arguments for using the full tensor wavelet expansion on R^2 in image compression. I have also seen a mixed tensor wavelet representation for Operators on functions on $\mathbf{R}^d$: The non-standard wavelet tensor basis on $\mathbf{R}^d$ was extended to a basis on $\mathbf{R}^d \times \mathbf{R}^d$ by a full tensor extension of the d dimensional wavelet basis.

I have, so far, only made rather trivial estimates using ideas with full wavelet tensor extension. My Ph.D student Øyvind Bjørkås has done some more detailed studies in his Cand. Scient. Thesis. ([3]).The implementations would be longer future projects Some of the ideas presented in here were communicated to R.R. Coifman, who in a joint paper with D.L. Donoho ([4] has used them in the setting of stochastic variables and their distribution functions. A tensor wavelets expansion in 3-dimension has been used by Averbuch, Israeli and Vozovoi ([1]) to implement a fast PDE solver.

We will, in this presentation, only consider the case $m = 1$. In this case the wavelet functions are the classical Haar functions. However, it is not difficult to generalize the corresponding statements to general $m > 0$. One may also, without much difficulty, generalize to the situation with fractional smoothness conditions (as multi-Lipschitz conditions).

2 PRELIMINARIES

Let χ be characteristic function of interval $[0, 1]$, and the Haar function

$$H(x) = \begin{matrix} -1 & \text{when} & 0 \leq x \leq \frac{1}{2} \\ 1 & \text{when} & \frac{1}{2} < x \leq 1 \end{matrix}$$

The family of Haar functions in dimension $d = 1$ is defined by

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

We also define the corresponding set of so called scale functions by

$$\chi_k j(x) = 2^j \chi(2^j x - k)$$

With this notation the set of Haar functions H_{kj}, $0 \leq k < 2^j, j \geq 0$. together with function χ_{00} is an orthonormal basis on the unit interval $[0,1]$ in $\mathbf{R}$ In this paper I prefer to work the the L^∞ normalized Haar functions $h_{j,k} = 2^{-j/2} H_{jk}$ and the L^1 normalized dual functions $\tilde{h}_{jk} = 2^{j/2} H_{jk}$. We also write $\psi = 2^{-j/2} \chi_{jk}$ and $\tilde{\psi}_{jk} = 2^{j/2} \chi_{jk}$ The expansion of a function f on $[0,1]$ with the Haar wavelets then may be written

$$f = < f, \tilde{\psi} > \psi + \sum_{jk} < f, \tilde{h}_{jk} > h_{jk}.$$

Clearly, $| < f, \tilde{\psi} > | \leq |||f||_\infty$. Also, if there is a constant $A > 0$ such that

$$|f(x) - f(y)| \leq A|x - y| \tag{2}$$

then

$$| < f, \tilde{h}_{jk} > | \leq \frac{1}{4} A 2^{-j}. \tag{3}$$

Note that the factor 2^{-j} is equal to the length of the supporting interval of $\tilde{h}_{jk}$

3 THE NON-STANDARD TENSOR WAVELET EXTENSION

I non-standard tensor extension, all the 2_d combinations of the wavelets h_{jk} and the scale functionswhere for each scale j are used exept for the tensor product where all factors are scale functions. The latter tensor product is unly used on coarses scale. Estimating the wavelets coefficients we have the the worst cases tensor products with only on wavelet factor and thus only one directions where we have the estimate decreasing with j as above . The number of functions needed for accuracy level ϵ is $(1\epsilon)^d$. We will later compare this with the full wavelet tensor expansion (See figure 1)

4 MIXED VARIATION

With good control of the mixed variation ,we will get better estimates for the wavelet coefficients in the full tensor wavelet expansion. To define what we mean with mixed variation, we need some definition. Let R be a rectangle in $\mathbf{R}^d$ of dimension s, $0 \leq s \leq d$, which is parallel to the axes. Let $Corner(R)$ be the set of corners of R. For $p = (x_1, \ldots, x_d) \in Corner(R)$ we associate the number δ_p equal to $+1$ or -1. We do this by setting $p = 1$ at the point p for which $x_1 + \cdot + x_d$ is maximal and by changing the sign of the value of p as me move along each of the edges (of dimension 1) of R. The difference operator $\triangle_R$ is defined bye

$$\triangle_R f = \sum_{p \in Corner(R)} \delta_p f(p)$$

(In the case $s = 0$ R is a point and $p = R$ and $\triangle_R f = f(p)$.)

Definition 1 *A function f on the unit cube I^d in $\mathbf{R}^d$ is of bonded mixed variation of order $m = 1$ and with constant A if*

$$|\triangle_R f| \leq A(|R|_s) \tag{4}$$

for each rectangles R parallel to the axes and of dimension $s, 0 \leq s \leq d,$. We use notation $|R|_s$ for the s dimensional volume of R. (In the case $s = 0$ R is a point and $|R|_s = 1$.)

Let $M(\underline{x}_1, \dots, \underline{x}_m) = \underline{x}_1 + \dots + \underline{x}_m$ be a mapping $\mathbf{R}^{md}$ to $\mathbf{R}^d$

Definition 2 *The function f onf I^d has bounded mixed variation higher order m with constant A if for the function $F(\underline{x}) = f(M(\underline{x})$ we have*

$$|\triangle_R F| \leq A|R|_s \tag{5}$$

for each sub-rectangle R in $\mathbf{R}^{md}$ of of dimension $s, 0 \leq s \leq md$ contained in I^d

As a direct consequence of the classical mean value theorem, the condition 4 holds for any function f satisfying the mixed derivative condition $\partial^\alpha f / \partial x^\alpha$ satisfying

$$|\partial^\alpha f / \partial x^\alpha| \leq A \tag{6}$$

for all multi index $\alpha = (\alpha_1, \dots, \alpha_d)$ with

$$\max \alpha_i \leq m \tag{7}$$

5 The full tensor wavelet expansion

In the full tensor wavelet expansion we use tensor products

$$\eta_J = \eta_{j_1,k_1} \otimes \dots \otimes \eta_{j_d,k_d},$$

where $\eta_R = \eta \in \{h, \chi\}$ with $0 \leq < k_i < 2^{j_i}$. Here the scale index in the i-direction $j_i \geq 0$ when $\eta = h$ and $j_i = 0$ when $\eta = \chi$. R indicates the supporting rectangle of the this wavelet function. For the full tensor wavelet extension of the one dimensional wavelet we get the estimate of the of the coefficients related to the volume of the supporting rectangle. In case of the Haar wavelet extension we have, more precisely,

$$| < f, \tilde{\eta}_R > | \leq (1/4)^s A |R|_d \tag{8}$$

where s is the number of h factors in the tensor product η_R. This is much better estimate than the estimate for the non-standard extension of the wavlets where a coefficients in the worst case are related to the side length of the supporting cube.

6 The Multi Scale Space Grid

In dimension one we have a sequence of nested spaces V_j which we may think of as points along a line. In the Haar case V_j is the space of functions partially constant on intervals of length 2^{-j}. In dimension $d \geq 1$ we will consider the spaces

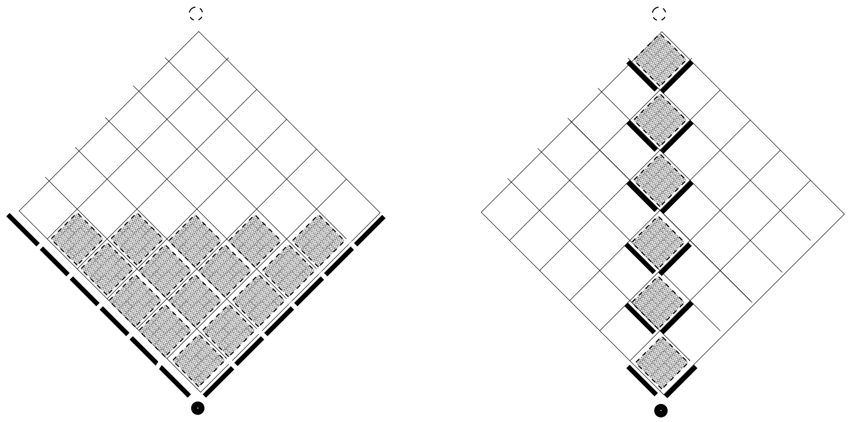

Figure 1: (*left:*) Approximation with full tensor expansion, (*right:*) Approximation with non-standard tensor expansion

$V_j = V_{j_1,\ldots j_d} = V_{j_1} \otimes \cdots \otimes V_{j_d}$ as points $j = (j_1, \ldots, j_d)$ in a d dimensional integer grid. Let P_j be the projection to V_j. We evaluate a function f at the space point j as the function $P_j f$. As above may define the mixed different $\triangle_R$ for any axes parallel rectangle R on this grid (with dimension $s, 0 \leq s \leq d$). One can show that $\triangle_R$ is a projection. In this space grid we identify rectangles R as the space corresponding to projection $\triangle_R$ Now, we may make simple algebraic rules of how any rectangle R and its lower dimensional boundaries are related. We may also add together collections of rectangles R_i and express the sum in terms of their union and its boundaries. We introduce the variable $J = j_1 + \cdots + j_d$ and turn out multi-scale grid with the J-axis (not drawn) pointing vertically upwards:

Let Q be the whole Multi Scale Space Grid as a cube in $\mathbf{R}^d$ The space corresponding to the top point of Q is $V_{n,\ldots,n}$, while the bottom point is the space $V_{0,\ldots,0}$ In the full tensor expansion of the wavelets the top point is decomposed as a direct sum of all spaces corresponding to all the small s-dimensional cubes, $0 \leq s \leq d$, lying on those s - dimensional boundary cubes of Q, which contains the bottom point. These spaces are indicated on the left part of figure 1 as all filled squares,all bold line segment and finally the bottom point.

7 Approximation with full tensor wavelets expansions

The main observation is, that the a priori estimate of the wavelet coefficient for a the subspace in this decomposition is essentially proportional to

$$2^{-J}$$

or the J - coordinate of the position of this subspace. On the other hand the number of bases element in the subspace is proportional to

$$2^J$$

This means that it strategic to approximate a functions in the top space by using the projection with all subspaces with J coordinate under some level, such as $J \leq n$. We get

THEOREM 1 *Let $\epsilon > 0$. Then there is a set S with $O((\log \frac{1}{\epsilon})^{d-1}\epsilon^{-1})$ Haar full tensor wavelets functions such that any function f be a function in the unit cube in $\mathbf{R}^d$ satisfying condition 7 may be approximated by*

$$\sum_{h\in S} < f, \tilde{h} > h$$

with accuracy in sup norm $O((\log \frac{1}{\epsilon})^{d-1}\epsilon)$ (in L^2 norm $O((\log \frac{1}{\epsilon})^{fracd-12}\epsilon)$). The value of f at a single point may be computed in $O((\log \frac{1}{\epsilon})^d$ steps.

8 A SPARSE SET OF RECTANGLES

The approximation in the theorem above is done with subspaces with space grid coordinate $J \leq N$. All those subspaces are in the span of the subspaces $V_{\underline{j}}$ with the coordinate $J = n$. This means in the Haar case we that the mean f_{R_0} may be computed from the mean values f_R, R dyadic rectangle (with some accuracy).

THEOREM 2 *Let f be a function on the unit cube in $\mathbf{R}^d$ satisfying condition 7. Let R_0 be a dyadic sub rectangle with volume $|R_0|_d \leq 2^{-n}$. and let $\mathcal{R}_k$ be the set of dyadic rectangle R in the unit cube with volume $|R|_d = 2^{-k}$ and let f_R the mean value of f over R Then, with errorr $O(n^{d-1}2^{-n})$ we have*

$$f_{R_0} \approx \sum_{k=0}^{d-1}(-1)^k \binom{d-1}{k} \sum_{\substack{R \in \mathcal{R}_{n-k} \\ R \supset R_0}} f_R$$

9 A SPARSE SUBSET OF GRID POINTS

We would not have much practical use of the full tensor approximation with with the sparse set of $O((\log N)^{d-1}N)$ wavelet coefficients if, in order to compute them, we need all N^d samples points of the function. However, it is not very difficult to show that that these wavelet coefficients may be calculated from a sparse set of sample values of the function.

Let G_N be the set of grid points in the unit cube in R^d where the grid size is $1/N$ Let S_N be the subset of G_N consisting of all corners of dyadic rectangles with volumes $\geq 1/N$. The number of points in S_N is $O((\log N)^{d-1}N)$.

THEOREM 3 *Let f be a function satisfying 5. Given the value of f at S_N one may also compute the the sparse set of ($O((\log N)^{d-1}N)$ coefficients $< f, \tilde{h}_{\underline{jk}>}$ with an error bounded by $O((\log N)^{d-1}\frac{1}{7}N)$ (with $\tilde{h}_{\underline{jk}}$ L^1- normalized) This can be done in $O((\log N)^{d-1}N)$ steps.*

There is also a fast algorithm to recover the values of functions at any grid point in G_N

THEOREM 4 *Let f be a function satisfying 5. Given the value of f at all points in S_N one may compute the values of f at any point $p \in G_N$ with error bounded by $O((\log N)^{d-1}\frac{1}{N})$. The complexity of such a computation is $O((\log N)^{d-1})$.*

We do not have room here to include any proofs of this.

10 TENSOR WAVELETS ON SINGULAR INTEGRAL OPERATORS

Wavelets was used with great success, by Beylkin, Coifman and Rokhlin with great success to reduce complexity for the computation of Singular Integral Operators . The operator T is assumed to be bounded on $L^2(\mathbf{R}^d)$ and its kernel $K(x,y)$ satisfies the usual smoothness conditions 1 away away from the diagonal. Let us consider the problem as to compute the inner product

$$< f, Tg >=< f \otimes g, K > .$$

We may think of K represented by a matrix ($2d$ dimensional tensor) with N^{2d} elements, $N = 2^n$. Beylkin, Coifman and Rokhlin represent this inner product by use the non-standard tensor bases extension of the wavelets (of order m) on $\mathbf{R}^{2d}$. In this bases the Kernel is represented by a matrix, which may by approximated with accuracy level ϵ to a matrix with finite diagonal bandwidth containing $N^d O(\epsilon^{\frac{d}{m}})$ non-zero elements. We see that estimate for the number of non-zero elements in this matrix is not very good when d is large. Using a hybrid of non-standard and full tensor basis extension it is possible to improve their result:

THEOREM 5 *There is a hybrid tensor wavelet extension basis on the unit cube I^{2d} in $\mathbf{R}^{2d}$, in which the kernel K is represented with accuracy level ϵ by a matrix with $N^d O(1/\epsilon)$ non-zero elements. The coefficients of K in this basis may be computed in $O(N^{2d})$ steps. The wavelet coefficients of $f \otimes g$ in this basis are simple products of coefficients taken from on set of coefficients for f and one set set of coefficients for g, Each of these two sets is order $O(N^d)$ and may be computed in $O(N^d)$ steps.*

We will not give the proof here.

11 THE HAAR PACKETS

A C^∞ function may be approximated much better with wavelet packets than by wavelets. We will for simplicity only consider wavelet packets of the Haar functions for approximating functions f satisfying

$$\frac{d^k|f}{dx^k}(x)| \leq 1 \text{ for all } k \geq 0 \tag{9}$$

Then we have

THEOREM 6 *Let $\epsilon > 0$ then there is as set of*

$$M_\epsilon = O\left(\exp\left(c\sqrt{\log(\frac{1}{\epsilon})}\right)\right)$$

Haar wavelet packet functions $\{W_k\}_{k=1}^{M_\epsilon}$ on the interval $[0,1]$, such that any function f as above is approximated with error less than ϵ by

$$f \approx \sum_{k=1}^{M_\epsilon} < f, W_k > W_k.$$

The smoothness condition 9 on the function f is very strong. However, one may get some rather similar estimates with much weaker condition on f. The wavelet packe tree starting from the top consists of nodes, which are connected by branches of low-pass and high-pass filters. At the bottom of the tree, each space is of dimensions one. Let $j_1, j_2, \ldots, j_s$ be a sequence of positve integers which indicates the levels of the branshes, where we have passed through the high-pass filter to reach a node. Then we get, by iteration of the mean value theorem, the a priori estimate

$$2^{-(j_1+\cdots+j_s)}2^{-2s}.$$

(We assume have normalized the filters so that the Low-pass filter does not increase the norm.) Let $2^{-n} = \epsilon$. To get the Theorem above we mainly hove to solve the following problem in combinatorics: Estimate the number of finite sequences of integers $(j_1, j2, \cdots j_s)$ with

$$0 < j_1 < j_2 < \cdots < j_s \leq n$$

such that

$$j_1 + j_2 + \cdots + j_s \leq n.$$

Approximating with the Haar packets on the unit cube i $\mathbf{R}^d$ leads to a similar combinatorics problem with sets where up to d indices j_k may assume the same integer values. One will get the estimate

$$M = O\left(\exp\left(c\sqrt{d\log(\frac{1}{\epsilon})}\right)\right)$$

The problem with the approximating Singular Integral Operator kernels with Haar packets will also lead to a combinatoric problem. As in the compression of the kernel K by Haar wavelets above one my use Whitney decomposition also in this case. We have not analized this in detailed yet. However, one would probably get something like, that the kernel K, may be approximated at accuracy level ϵ using a set of

$$N^d O\left(\exp\left(c\sqrt{d\log(\frac{1}{\epsilon})}\right)\right),$$

Haar Packet functions.

REFERENCES

[1] Averbush, A., Israeli, M. an Vozovoi, L. . A fast Poisson solver of arbitrary order accuracy in rectangular regions.*SIAM J. Sci. Comput. 19 (1998), no. 3, 933–952.*

[2] G. Beylkin, G., Coifman, R. and Rokhlin, V, Fast wavelet transforms and numerical algorithms. *Comm. Pure. Appl. Math., 1991.*

[3] Bjørkås, Ø. Numerical Analysis of Singular Integral Opteratos with Wavelets. Cand. Scient Thesis at University of Tromsø, 1997.

[4] Coifman, R. and Donoho, D. To appear

[5] David, G, and Semmes,S *Fractured fractals an broken dreams.*Self-similar geometry through metric and measure. The Clarendon Press, Oxford University Press,NewYork, 1997, pp. 212

Jan-Olov Strömberg
Department of Mathematics
Royal Institute of Technology
Stockholm
Sweden
janolov@math.kth.se