Operators and Fast Algorithms

Gregory Beylkin

University of Colorado at Boulder

IPAM
September 14, 2004

Plan

1. Multiresolution analysis, telescopic expansion of operators, non-standard form, selection of bases for numerical algorithms, examples.

2. Separated representations and multiresolution algorithms in high dimensions

3. Examples of separated multiresolution representations: the Poisson kernel and the projector on the divergence free functions in $\mathbb{R}^3$; a brief description of the Fast Adaptive Poisson solver in $\mathbb{R}^3$ as time permits.

Multiresolution Analysis

Chain of subspaces:

$$\ldots \subset V_0 \subset V_1 \subset V_2 \subset \ldots$$

$\bigcap_j V_j = \{0\}$ and $\overline{\bigcup_j V_j} = L^2(R^d)$.

Examples: piecewise-constant functions, Daubechies' scaling functions, polynomials up to degree $\mathfrak{m} - 1$ on a collection of intervals.

Detail spaces W_j : $V_{j+1} = V_j \oplus W_j$

Orthonormal bases in V_j and W_j are defined by the scaling function ϕ and wavelet ψ.

Examples: Haar basis, Daubechies' wavelets, multiwavelets.

Projectors: P_j onto V_j, Q_j onto W_j.

Choice of basis

The original goal of wavelets was a smooth generalization of the Haar basis.

Although they appear natural, smooth wavelets have a number of practical difficulties and limitations

These difficulties appear in signal processing but they are much more pronounced in numerical analysis.

Issues:

- boundary conditions
- interpolating property
- representation(s) of differential operators

Boundary conditions

Smoothness leads to overlapping supports.

To adapt such bases to bases on an interval necessarily involves boundary operators.

The condition number of these boundary operators grows rapidly with the order of the bases. As a result, such constructions in numerical analysis work satisfactorily only for low order schemes.

In signal processing this problem appears where it is necessary to process finite data, for example, near the edge of an image.

The problem is much more difficult if high precision is required.

Interpolating property

PDEs often involve pointwise multiplication of functions and so interpolating scaling functions are very convenient.

Let the subspace of a multiresolution analysis be defined as a linear span of functions $\{\phi(x-n)\}_{n\in Z}$, such that $\phi(n) = \delta_{n0}$. Then if $f \in \mathbf{V_0}$, we have

$$f(x) = \sum_n f(n)\phi(x-n).$$

However, the combination of smoothness, orthogonality and interpolating property leads to the non-compact support of the scaling functions.

Interpolating bases are available, (e.g., Butterworth wavelets) but do not have a compact support, so it becomes even more difficult to adapt them to life on an interval.

Representations of differential operators

The smoothness of basis functions limits the availability of scale consistent derivative operators.

The coefficients of representation of d/dx are defined by

$$r_l = \int_{-\infty}^{\infty} \phi(x-l)\, \frac{d}{dx}\, \phi(x)\, dx,$$

and, for sufficiently smooth scaling functions, these integrals are absolutely convergent.

Such a representation is a "central" difference operator and, because of the uniqueness, there are no "forward" or "backward" differences.

Example: Daubechies' wavelets with two vanishing moments,

$$\{r_l\}_{l=-2,\ldots,2} = \{-1/12, 2/3, 0, -2/3, 1/12\}.$$

Naively, this does not appear as an inconvenience, but in numerical analysis forward and backward difference operators are really useful.

Polynomial bases on intervals

Instead of a smooth generalization of the Haar basis, let us proceed in a different direction, towards multiwavelet bases.

Remark on terminology: We will consider *only* non-overlapping multiwavelets and use the term only for such bases.

Multiwavelet bases:

- are useful for representing integral operators.
- are well suited for the high-order adaptive solvers of partial differential equations (perhaps counter intuitively).
- can accommodate the boundary conditions for high-order methods
- can be orthonormal and interpolating

There is a family of scale consistent derivative operators which may be viewed as weak (non-unique) representations. Non-uniqueness is an **advantage**.

Telescopic expansion: the non-standard form

Consider $T : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$

Define $T_j = P_j T P_j, \quad A_j = Q_j T Q_j, \quad B_j = Q_j T P_j \quad C_j = P_j T Q_j,$ for $j \in Z$.

We have the telescopic expansion,

$$T_n - T_0 = \sum_{j=1}^{n} (P_j T P_j - P_{j-1} T P_{j-1}) = \sum_{j=1}^{n} (T_j - T_{j-1}) = \sum_{j=1}^{n} (A_j + B_j + C_j).$$

Rate of decay

For Calderon-Zygmund operators the entries (or blocks entries) of A_j, B_j, and C_j decay away from the diagonal according to the number of vanishing moments of the basis. Let the kernel satisfy

$$|K(x,y)| \le \frac{C_0}{|x-y|}, \text{ and } |\partial_x^M K(x,y)|, \; |\partial_y^M K(x,y)| \le \frac{C_1}{|x-y|^{1+M}},$$

for some $M \ge 1$. Then by choosing a wavelet basis with M vanishing moments the entries of A_j, B_j, and C_j satisfy the estimate

$$|\alpha_{i,l}^j|, \; |\beta_{i,l}^j|, \; |\gamma_{i,l}^j| \le \frac{C_M}{1+|i-l|^{M+1}}$$

for all $|i-l| \ge 2M$.

Multiwavelet bases

On each scale the scaling functions are orthogonal polynomials of degree up to $m-1$ on subintervals.

Choices:

1. The Legendre polynomials

2. The Lagrange interpolating polynomials with the Legendre nodes

Useful properties

1. Vanishing moments for multiwavelets

2. Interpolating property (up to rescaling)

3. Boundary conditions do not affect the order of the approximation

Selection of scaling functions

The scaling functions ϕ_i are the normalized Legendre polynomials on the interval $[0,1]$,

$$\phi_i(x) = \begin{cases} \sqrt{2i+1}P_i(2x-1), & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases},$$

where P_i are the Legendre polynomials on $[-1,1]$.

Alternatively, one can use interpolating polynomials.

Given the Gauss-Legendre nodes $x_0, \ldots, x_{\mathfrak{m}-1}$, the Lagrange interpolating polynomials are defined as

$$l_j(x) = \prod_{\substack{i=0, \\ i \neq j}}^{k-1} \left(\frac{x - x_i}{x_j - x_i} \right), \qquad j = 0, \ldots, k-1,$$

and characterized by $l_j(x_i) = \delta_{ij}$.

Interpolating scaling functions

Given the Gauss-Legendre nodes $x_0, \ldots, x_{\mathfrak{m}-1}$ and the associated Gauss-Legendre quadrature weights $w_0, \ldots, w_{\mathfrak{m}-1}$, the functions $R_j(x) = \frac{1}{\sqrt{w_j}} l_j(x)$ have the following properties:

1. They are orthonormal on $[-1, 1]$ with the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) dx.$$

2. R_j is a linear combination of Legendre polynomials,

$$R_j(x) = \sqrt{w_j} \sum_{i=0}^{k-1} \left(i + \frac{1}{2} \right) P_i(x_j) P_i(x).$$

3. Any polynomial f of degree less than $\mathfrak{m} - 1$ can be represented as

$$f(x) = \sum_{j=0}^{k-1} d_j R_j(x),$$

with $d_j = \sqrt{w_j} f(x_j)$.

Multiwavelets

Define V_j as a space of piecewise polynomial functions:

$$V_j = \begin{cases} f: & \text{a polynomial of degree less than } \mathfrak{m}-1 \text{ on the intervals} \\ & [2^{-j}l, 2^{-j}(l+1)] \text{ for } l = 0, 1, \ldots 2^j - 1 \\ 0 & \text{elsewhere} \end{cases}$$

V_j is spanned by $\mathfrak{m}$ functions obtained from $\phi_0, \ldots, \phi_{\mathfrak{m}-1}$ by dilation and translation,

$$\phi^j_{k,l}(x) = 2^{j/2}\phi_k(2^j x - l)$$

where $k = 0, \ldots, \mathfrak{m} - 1$ and $l = 0, \ldots, 2^n - 1$.

We define the multiwavelet subspace W_j, $j = 0, 1, 2, \ldots$ as the orthogonal complement of V_j in V_{j+1},

$$V_j \oplus W_j = V_{j+1}, \quad W_j \perp V_j \, .$$

Multiwavelets

Therefore, the multiwavelets are piecewise polynomial functions $\psi_0, \ldots, \psi_{\mathfrak{m}-1}$

$$\int \psi_i(x)\psi_k(x)dx = \delta_{ik}.$$

Since $W_j \perp V_j$, the first $\mathfrak{m}$ moments of $\psi_0, \ldots, \psi_{\mathfrak{m}-1}$ vanish,

$$\int \psi_k(x)\, x^i dx = 0, \qquad i, k = 0, 1, \ldots, \mathfrak{m} - 1\,.$$

W_j is spanned by $\mathfrak{m}$ functions obtained from $\psi_0, \ldots, \psi_{\mathfrak{m}-1}$ by dilation and translation,

$$\psi^j_{k,l}(x) = 2^{j/2}\psi_k(2^j x - l)$$

where $k = 0, \ldots, \mathfrak{m} - 1$ and $l = 0, \ldots, 2^n - 1$.

The cross-correlation functions of scaling functions

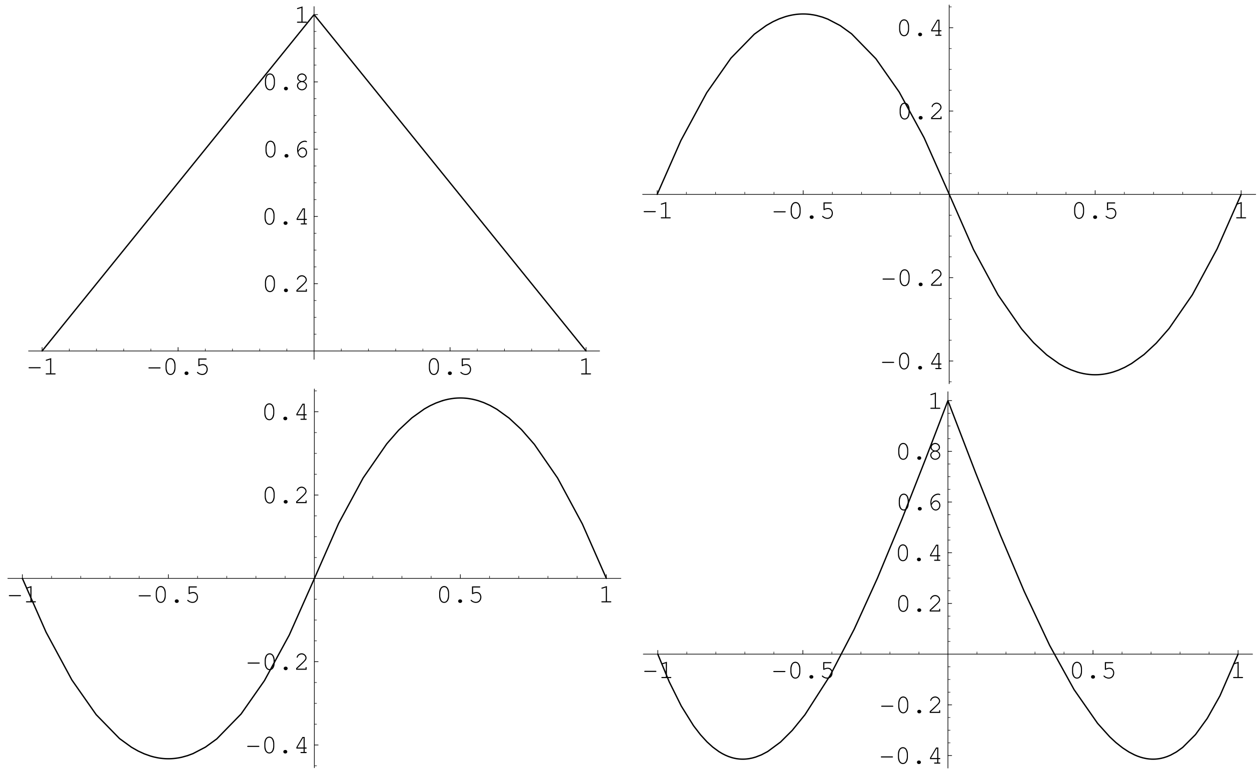
For convolution operators we only need the cross-correlation functions of the scaling functions, namely,

$$\Phi_{ii'}(x) = \begin{cases} \Phi^{+}_{ii'}(x), & 0 \leq x \leq 1, \\ \Phi^{-}_{ii'}(x), & -1 \leq x < 0, \\ 0, & 1 < |x|, \end{cases}$$

where $i, i' = 0, \dots, \mathfrak{m}-1$, $\mathfrak{m}$ is the order of the basis, and

$$\Phi^{+}_{ii'}(x) = \int_0^{1-x} \phi_i(x+y)\,\phi_{i'}(y)dy\,, \quad \Phi^{-}_{ii'}(x) = \int_{-x}^{0} \phi_i(x+y)\,\phi_{i'}(y)dy\,.$$

This implies that the functions $\Phi_{ii'}$ are piecewise polynomials of degree $i+i'+1$ with the support in $[-1,1]$.



The first four cross-correlation functions $\Phi_{00}, \Phi_{01}, \Phi_{10}$ and Φ_{11}.

Two scale difference equations

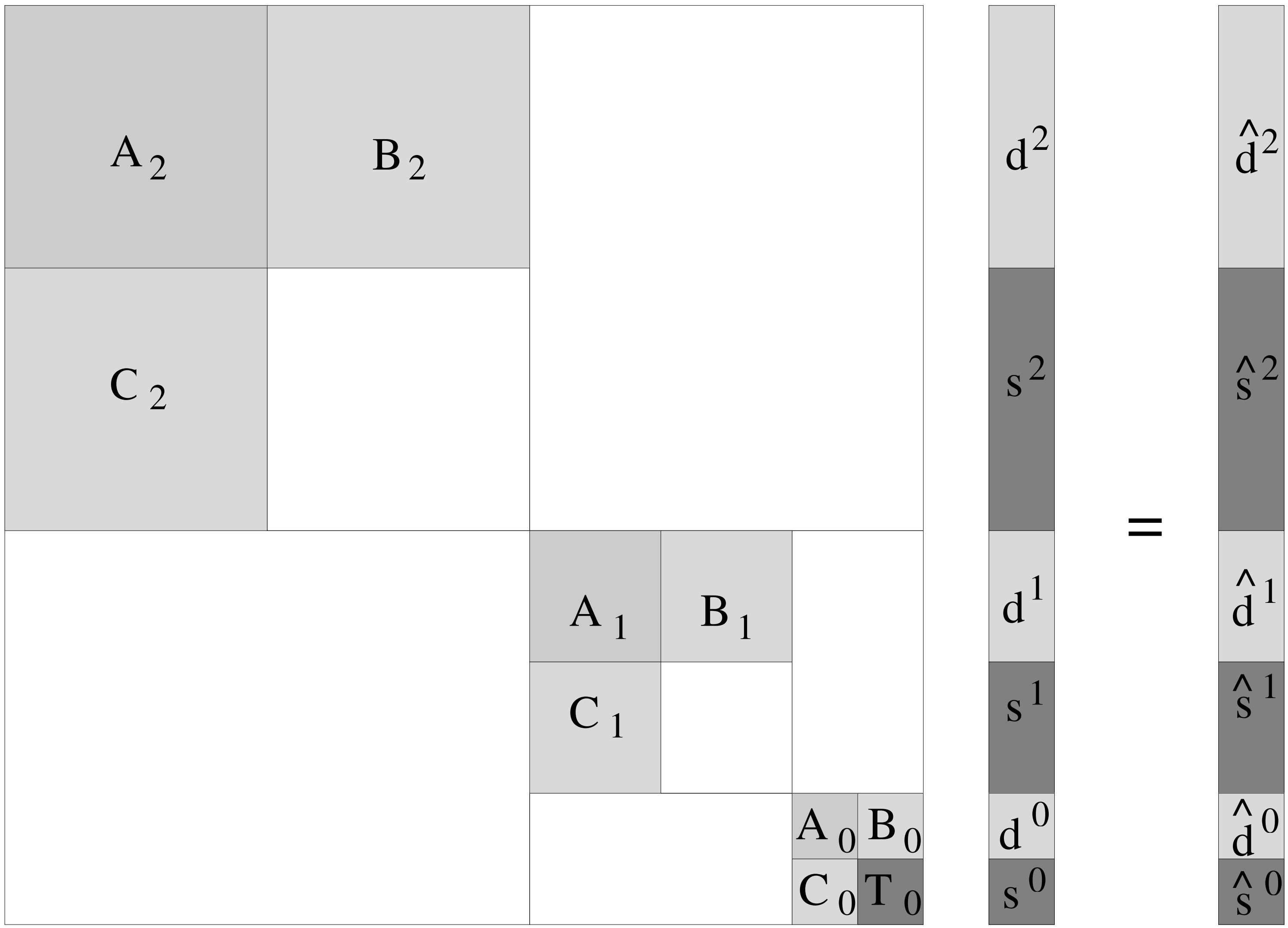
We have

$$\phi_i(x) = \sqrt{2}\sum_{k=0}^{\mathfrak{m}-1}(h_{ik}^{(0)}\phi_k(2x) + h_{ik}^{(1)}\phi_k(2x-1)),$$

$$\psi_i(x) = \sqrt{2}\sum_{k=0}^{\mathfrak{m}-1}(g_{ik}^{(0)}\phi_k(2x) + g_{ik}^{(1)}\phi_k(2x-1)),$$

where $i = 0, \ldots, \mathfrak{m}-1$. The coefficient matrices are computed from these identities.

Non-standard representation of operators in matrix form



Examples

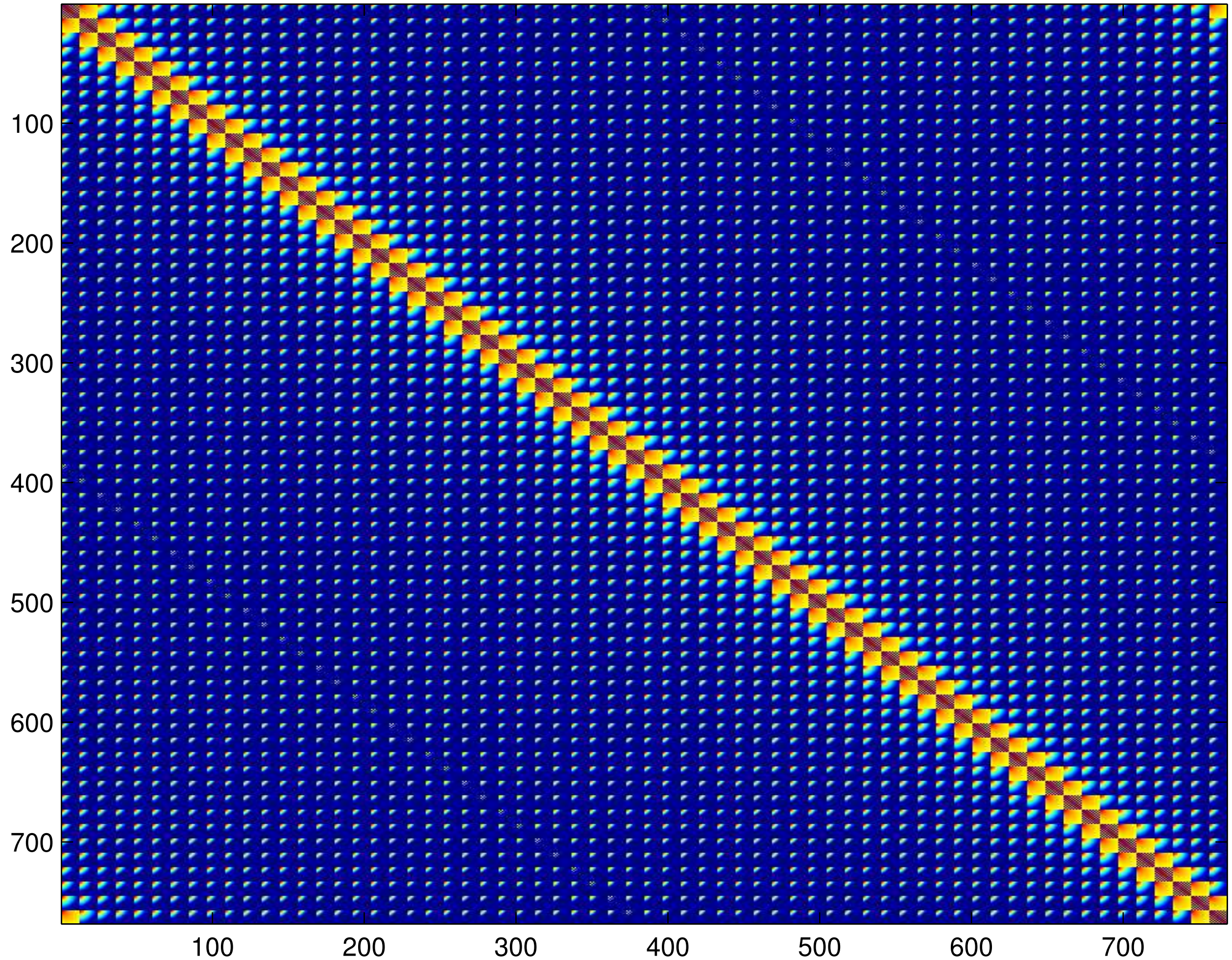
The Neumann to Dirichlet (N-to-D) map for the circle is given by

$$u(\theta) = \frac{R}{2\pi} \int_{-\pi}^{\pi} \log(\sin^2 \frac{(\theta - \theta')}{2}) \frac{\partial}{\partial r} u(r, \theta')\bigg|_{r=R} d\theta' + Const,$$

provided that

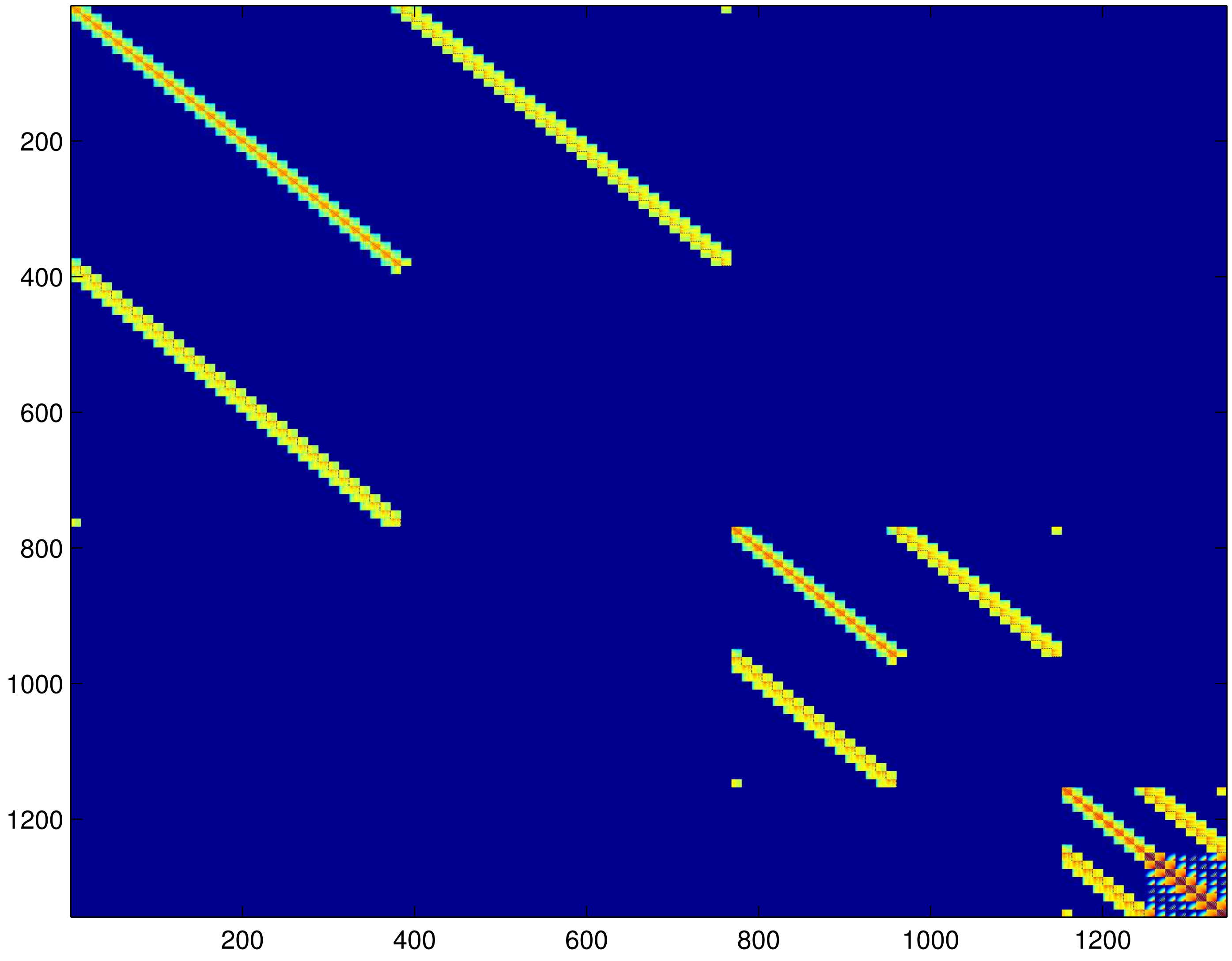
$$\int_{-\pi}^{\pi} \frac{\partial}{\partial r} u(r, \theta')\bigg|_{r=R} d\theta' = 0.$$

N-to-D operator

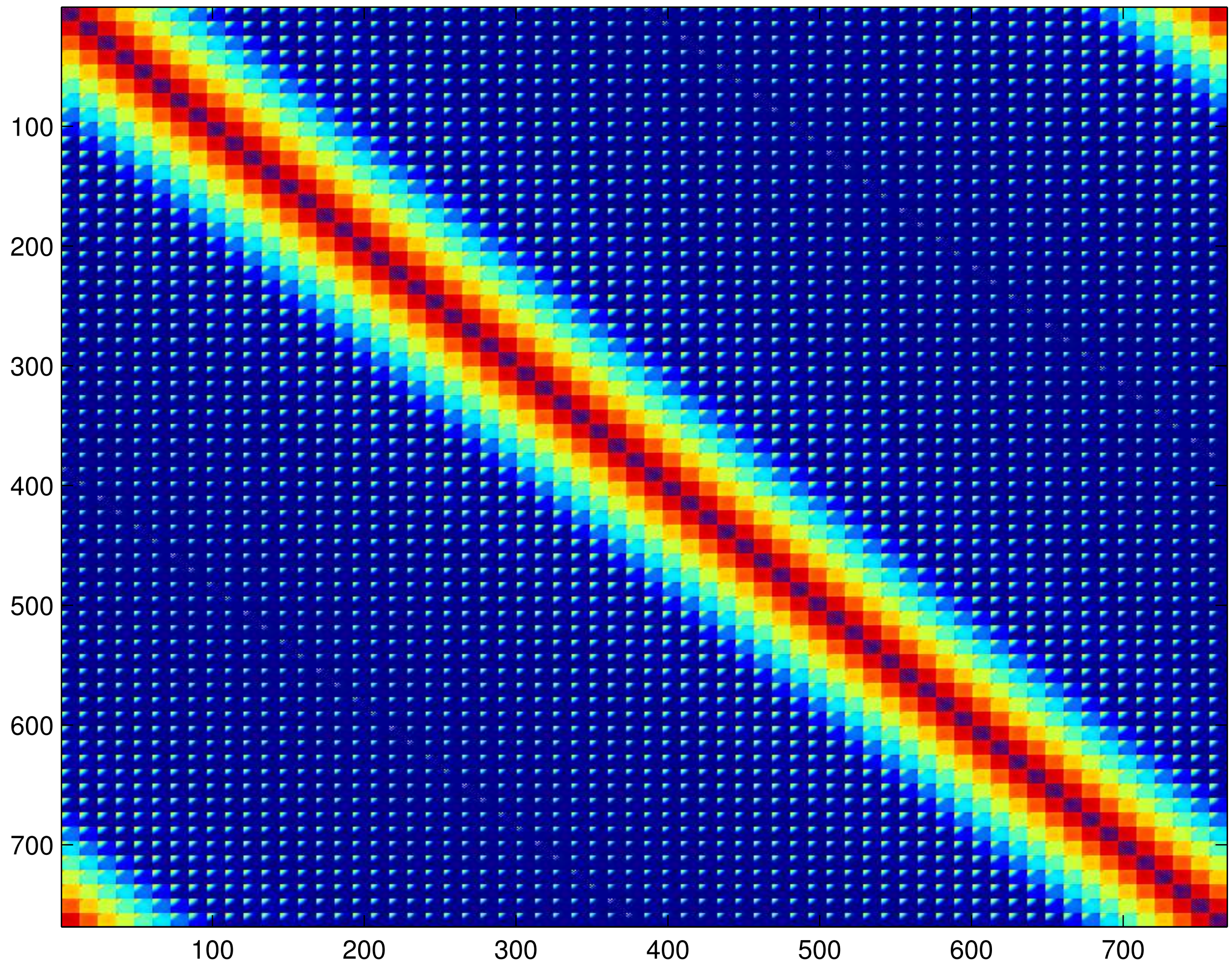


N-to-D operator and its non-standard form

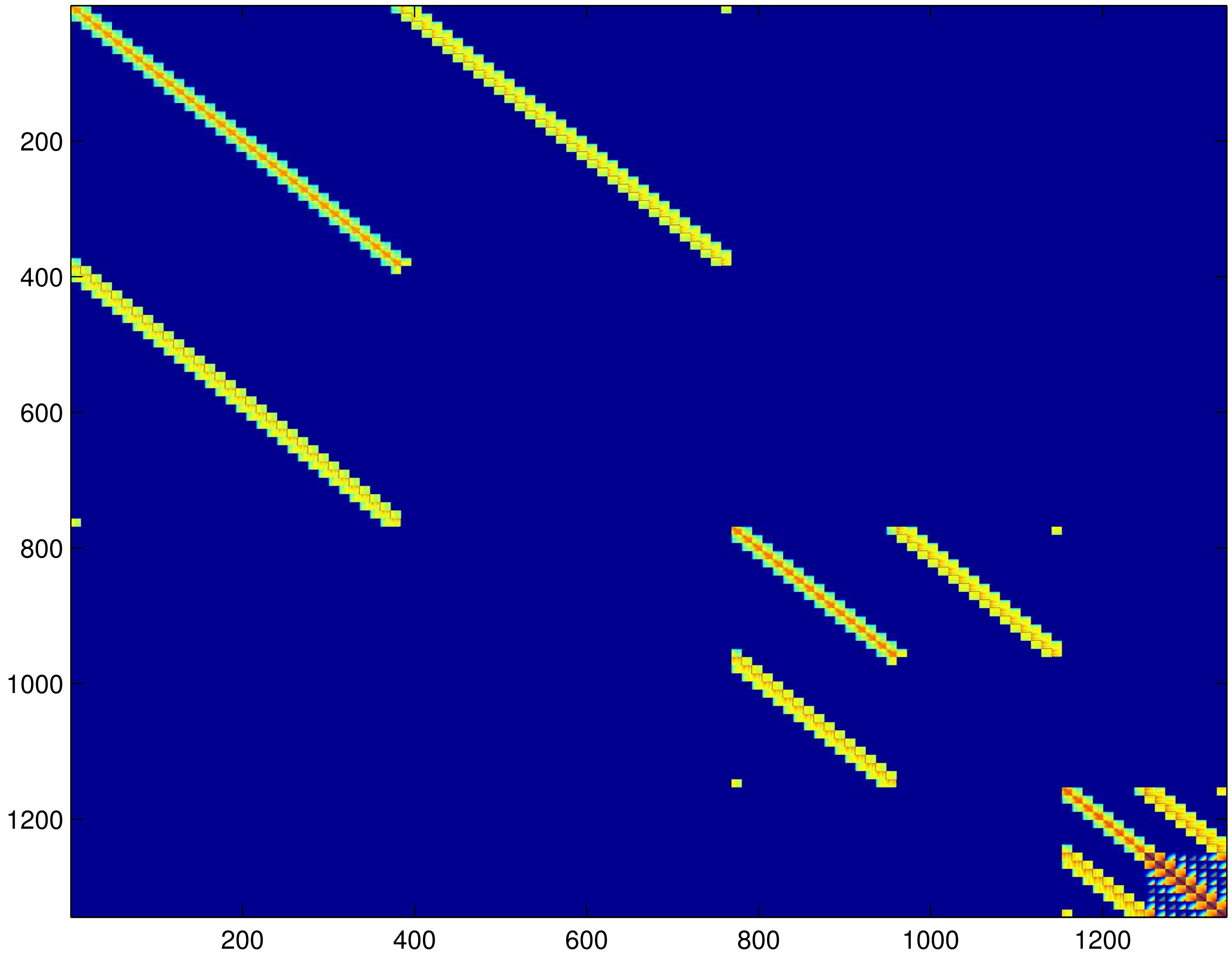
The non-standard form of N-to-D operator



Inverse of N-to-D operator (D-to-N)



The non-standard form of D-to-N operator



Problem of extending to multiple dimentions

- **Multiresolution representation of operators**
  - Classes of operators represented by banded matrices acting at different scales
- **Curse of dimensionality**
  - Number of entries in a banded matrix: $\mathcal{O}(bM)$
  - Cost of multiplication of two banded matrices: $\mathcal{O}(b^2M)$
  - Number of entries in a banded operator in dimension d: $\mathcal{O}(b^dM^d)$
  - Cost of multiplication of two banded operators in dimension d: $\mathcal{O}(b^{2d}M^d)$

The Separated Representation

The standard separation of variables: $f(x_1, x_2, \ldots, x_d) = \phi_1(x_1) \cdot \phi_2(x_2) \cdot \ldots \cdot \phi_d(x_d)$

Definition: For a given ϵ, we represent a matrix $\mathbb{A} = A(j_1, j_1'; j_2, j_2'; \ldots; j_d, j_d')$ in dimension d as

$$\sum_{l=1}^{r} s_l \, A_1^l(j_1, j_1') \, A_2^l(j_2, j_2') \cdots A_d^l(j_d, j_d'),$$

where s_l is a scalar, $s_1 \geq \cdots \geq s_r > 0$, and $\mathbb{A}_i^l$ are matrices with entries $A_i^l(j_i, j_i')$ and norm one. We require the error to be less than ϵ:

$$||\mathbb{A} - \sum_{l=1}^{r} s_l \, \mathbb{A}_1^l \otimes \mathbb{A}_2^l \otimes \cdots \otimes \mathbb{A}_d^l|| \leq \epsilon.$$

We call the scalars s_l separation values and the rank r the separation rank.

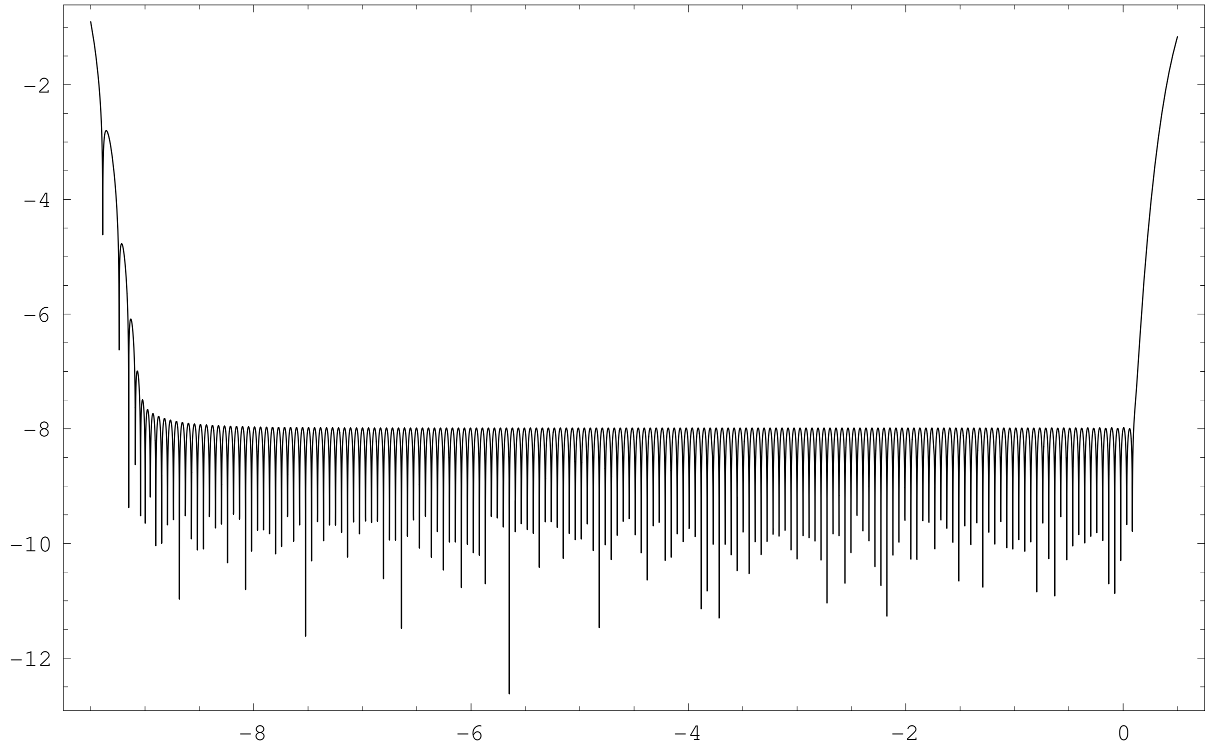
The smallest r that yields such a representation for a given ϵ is the optimal separation rank.

An example of separated representation

We have

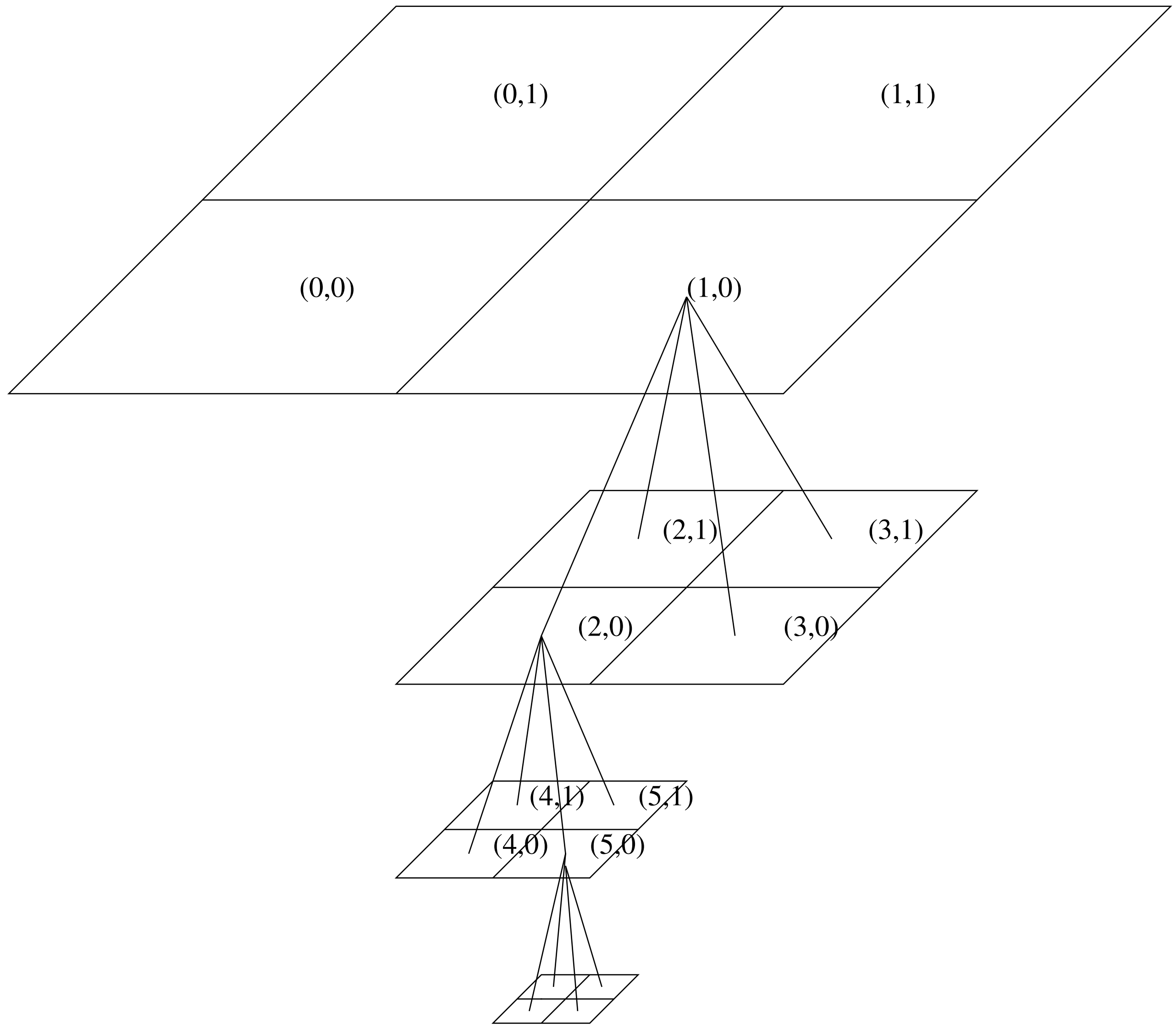
$$\left| \frac{1}{||r||} - \sum_{m=1}^{M} w_m e^{-p_m ||r||^2} \right| \le \frac{\epsilon}{||r||},$$

for $\delta \le ||r|| \le 1$, where $p_m, w_m > 0$ and $M = \mathcal{O}(-\log \delta)$.

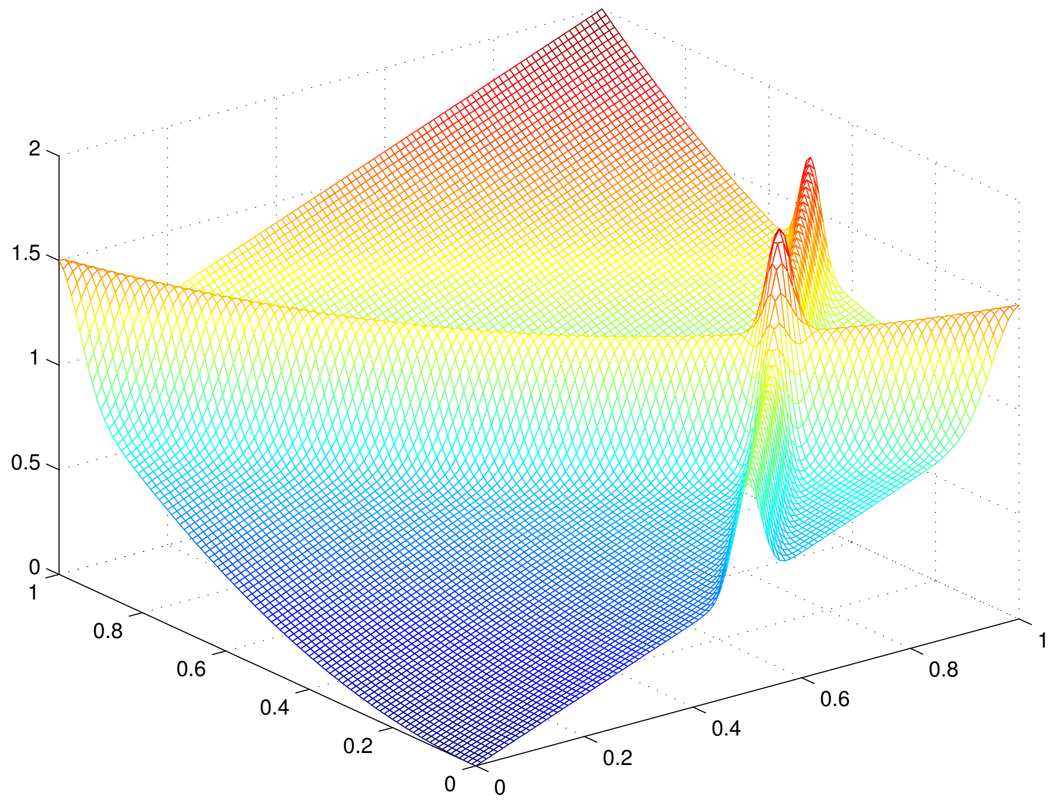


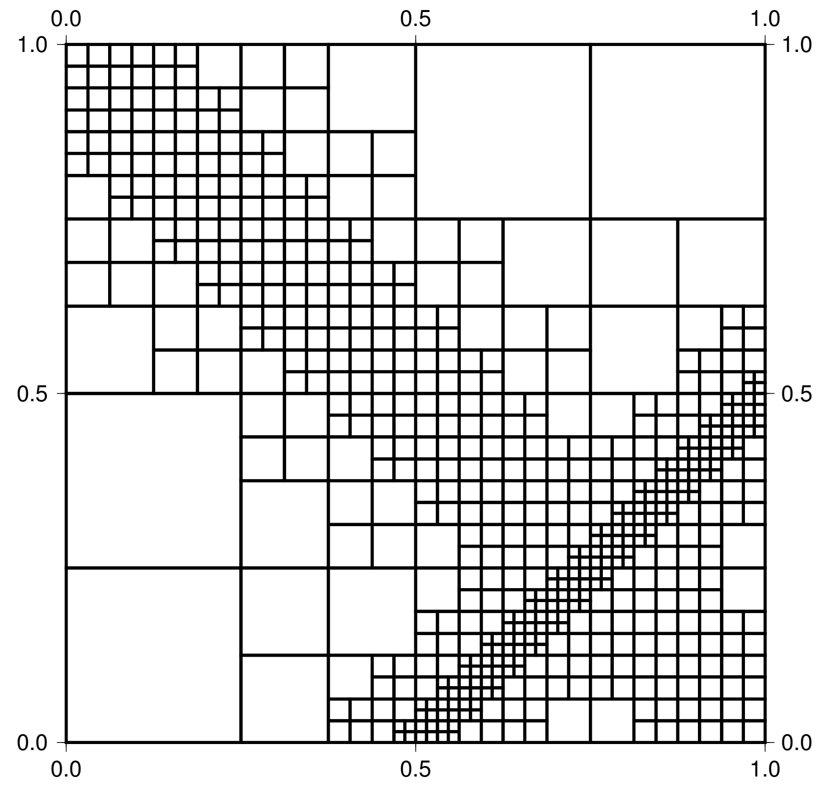
Error ($\log_{10}$) of approximating the Poisson kernel for $10^{-9} \le ||r|| \le 1$, $M = 89$.

Adaptive subdivision



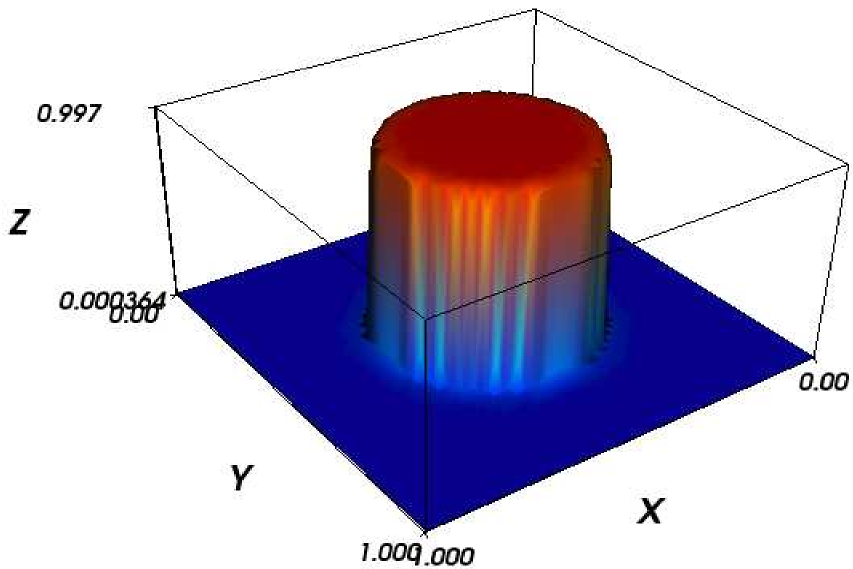
Adaptive representation of a function



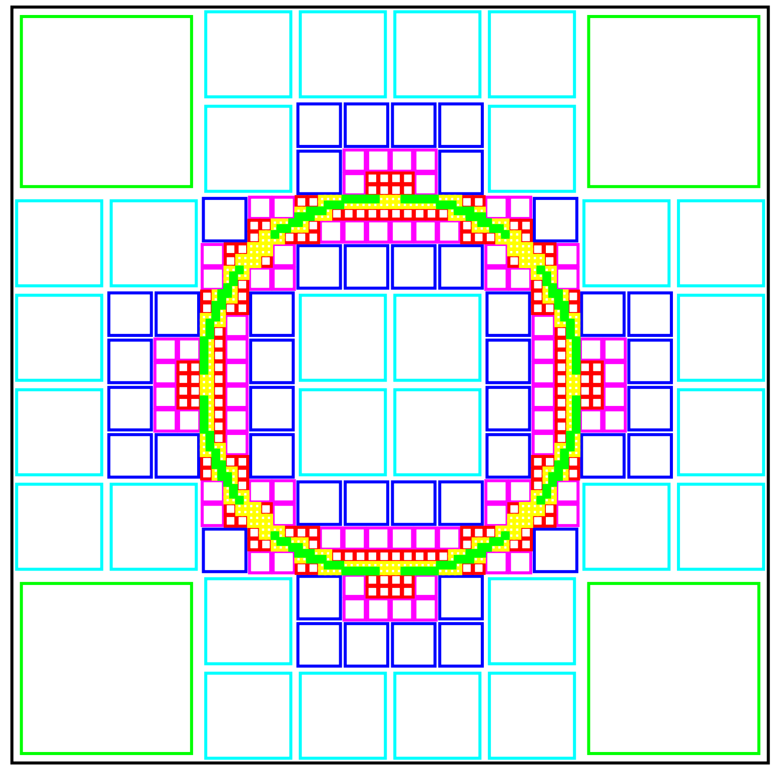


Another example

Consider the characteristic function of a disk



$N_{nod} = 8, \epsilon = 1.0e - 02, N_{blocks} = 1276$



The Poisson kernel

Due to the homogenuity of the Poisson kernel, we have

$$t^{n;\mathbf{l}}_{ii',jj',kk'} = 2^{-2n}\, t^{\mathbf{l}}_{ii',jj',kk'} \ ,$$

where

$$t^{\mathbf{l}}_{ii',jj',kk'} = t^{l_1,l_2,l_3}_{ii',jj',kk'} = \frac{1}{4\pi}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} \frac{1}{||\mathbf{x}+\mathbf{l}||}\, \Phi_{ii'}(x_1)\, \Phi_{jj'}(x_2)\, \Phi_{kk'}(x_3)\, d\mathbf{x},$$

and

$$\Phi_{ii'}(x) \ = \int_0^1 \ \phi_i(x+y)\, \phi_{i'}(y)dy\,, \quad i,i' = 0,\ldots,k-1\,,$$

are the cross-correlation functions of the scaling functions of the multiwavelet basis.

Separated representation of the Poisson kernel

Theorem: For any $\epsilon > 0$ the coefficients $t^{\mathbf{l}}_{ii',jj',kk'}$ have an approximation with a low separation rank,

$$r^{\mathbf{l}}_{ii',jj',kk'} = \sum_{m=1}^{M} \frac{w_m}{b} F^{m,l_1}_{ii'} F^{m,l_2}_{jj'} F^{m,l_3}_{kk'} ,$$

such that

$$\begin{array}{lll} |t^{\mathbf{l}}_{ii',jj',kk'} - r^{\mathbf{l}}_{ii',jj',kk'}| & \leq & \frac{2\epsilon}{\pi} \qquad\qquad \max_i |l_i| \geq 2 \\ |t^{\mathbf{l}}_{ii',jj',kk'} - r^{\mathbf{l}}_{ii',jj',kk'}| & \leq & C\delta^2 + \frac{2\epsilon}{\pi} \qquad \max_i |l_i| \leq 1 \end{array}$$

$$F^{m,l}_{ii'} = \int_{-1}^{1} e^{-p_m/b^2(x+l)^2} \Phi_{ii'}(x)\, dx ,$$

$b = \sqrt{3} + ||\mathbf{l}||$, and δ, $M = O(-\log\delta) + O(-\log\epsilon)$, p_m, w_m, $m = 1, \ldots, M$ are from the separated representation of the kernel.

Estimates of decay

Let us consider $T_j - T_{j-1}$. We need to estimate

$$G = \sum_m w_m F^{m,l_1} F^{m,l_2} F^{m,l_3} - \sum_m w_m \tilde{F}^{m,l_1} \tilde{F}^{m,l_2} \tilde{F}^{m,l_3}.$$

We have

$$\begin{aligned} ||G|| &\leq \sum_m w_m ||F^{m,l_1} - \tilde{F}^{m,l_1}|| \, ||F^{m,l_2}|| \, ||F^{m,l_3}|| \\ &+ \sum_m w_m ||\tilde{F}^{m,l_1}|| \, ||F^{m,l_2} - \tilde{F}^{m,l_2}|| \, ||F^{m,l_3}|| \\ &+ \sum_m w_m ||\tilde{F}^{m,l_1}|| \, ||\tilde{F}^{m,l_2}|| \, ||F^{m,l_3} - \tilde{F}^{m,l_3}|| \end{aligned}$$

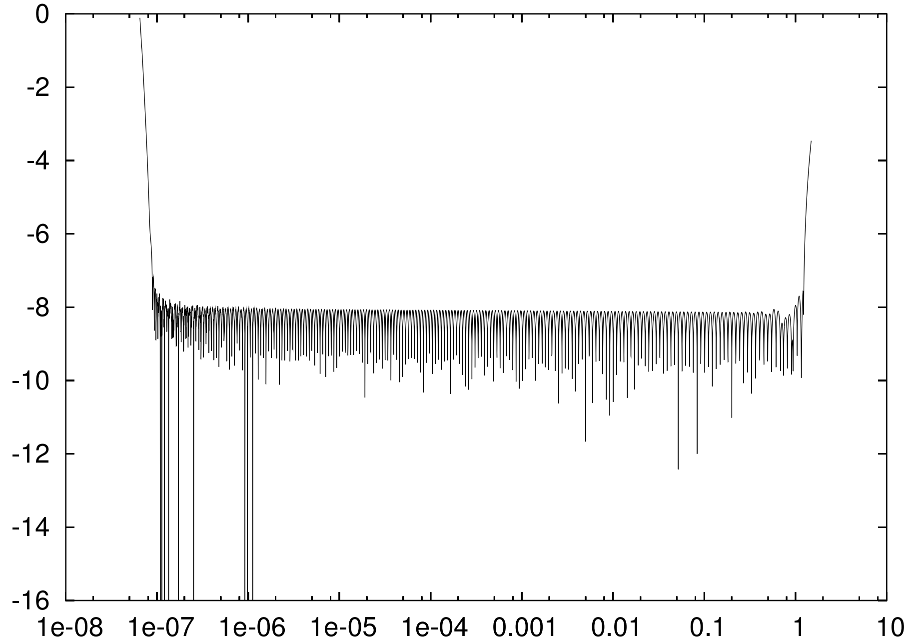
Thus, we need $F^{m,l}_{ii'}$ only for $|l| \leq l_{\max}$ (e.g. $l_{\max} = 5$).

Separated representation for the projector on the divergence-free functions

We have

$$\left| \frac{1}{||r||^3} - \sum_{m=1}^{M} w_m e^{-p_m ||r||^2} \right| \leq \frac{\epsilon}{||r||^2},$$

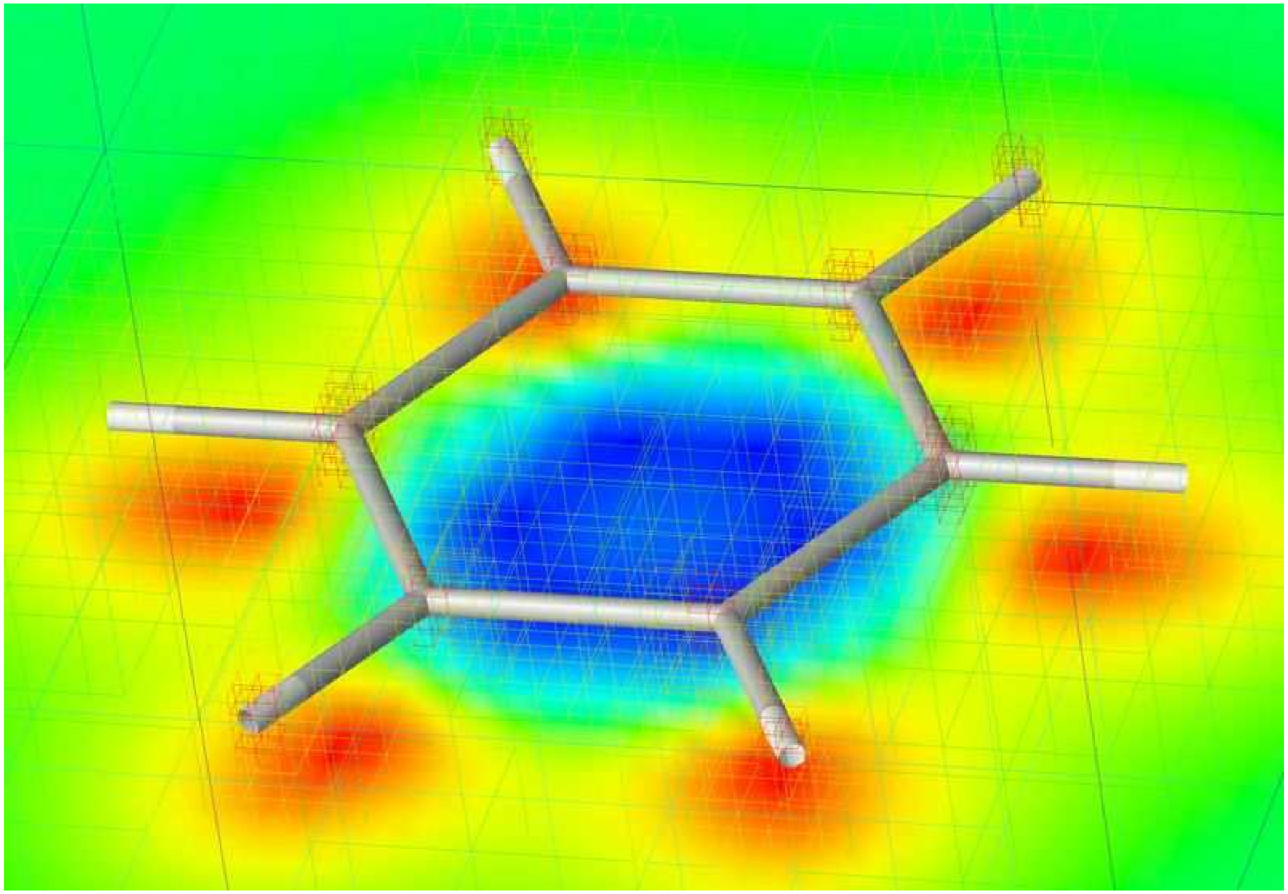
for $\delta \leq ||r|| \leq 1$, where $p_m, w_m > 0$ and $M = \mathcal{O}(-\log \delta)$.



Error of the approximation with 110 terms over the domain $10^{-7} \leq ||r|| \leq 1$.

Electron structure computations: elements, small molecules

A series of papers with R. Harrison, G. Fann, T. Yanai and Z. Gan (ORNL) in Journal of Chemical Physics



Adaptive subdivision of space for the benzene molecule C_6H_6

Timing results for 3D adaptive Poisson solver

Platform: Pentium 4-2.8 GHz with 1 GB of RAM for which `flops.c` gives ~ 950 MFLOPS for add-multiply code.

Initial timings made for $\epsilon = 5 \times 10^{-3}$

N_{nod}	6	8	10	12
N_{blocks}	512	120	120	64
$t\ (s)$	36	12.1	19.3	10.4
MFLOPS	171	317	430	505

Current timings

N_{nod}	12
N_{blocks}	64
$t\ (s)$	1.3
MFLOPS	911