

**Representation of operators in MADNESS
(Multiresolution ADaptive Numerical Environment for
Scientific Simulation)**

Gregory Beylkin

University of Colorado at Boulder

IPAM: Big Data Meets Computation workshop

February 1, 2017

MULTIRESOLUTION ADAPTIVE NUMERICAL ENVIRONMENT FOR SCIENTIFIC SIMULATION

MADNESS is a high-level software environment for the solution of multidimensional integral and differential equations using adaptive, fast multiresolution methods with guaranteed precision.

MADNESS implements a method whose key components were developed in 1988 (BCR paper, 1991) and in 2002-2004 based on an integral equation formulation of problems of quantum chemistry (and later nuclear physics). See e.g. R.J. Harrison, G.I. Fann, T. Yanai, Z. Gan and G. B., J. Chem. Phys. v. 121, n. 23, 14, 7, 2004.

Code written at ORNL (mostly by R. Harrison and G. Fann), efficiently runs on $4 \cdot 10^4$ processors and was recognized by R&D 100 Award in 2011 (R&D Magazine, 53).

Multiresolution approach

The multiresolution approach provides:

- Complete elimination of the basis error, correct scaling of the cost with system size
- Implementation for one-electron models (HF, DFT)
- Nuclear Physics modeling (G. Fann et.al.)
- Most accurate computations up to now within these models
- Much smaller computer code than “Gaussians” (← R. Harrison)

Topics for the talk

We consider several important features that allow systematic software development:

- Integral equation formulation
- Accurate separated representation of operators via Gaussians
- Multiresolution analysis using multiwavelets

We also briefly discuss

- Additional classes of operators (not yet implemented)
- New developments

Multiparticle Schrödinger operator

The Hamiltonian for the multiparticle Schrödinger operator is the sum of three terms

$$\mathcal{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N V_i + \frac{1}{2} \sum_{i=1}^{N-1} \sum_{m=i+1}^N W_{im},$$

where the 3D Laplacian corresponding to electron i is defined as $\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$, the nuclear potential V_i is operator of multiplication by $Z_\alpha / \|r_i - R_\alpha\|$ and the electron-electron potential W_{im} is multiplication by $1 / \|r_i - r_m\|$. The problem to solve:

$$\mathcal{H}\psi = E\psi.$$

The Schrödinger operator does not account for spin and, for this reason, the wave function is required to be antisymmetric, e.g., $\psi(\gamma_2, \gamma_1, \gamma_3, \dots) = -\psi(\gamma_1, \gamma_2, \gamma_3, \dots)$, where $\gamma = ((x, y, z), \sigma)$ and σ is the spin. Solutions are functions of $3N$ variables (positions of nuclei are fixed: the so-called Born–Oppenheimer approximation).

One particle theories

Solving the multiparticle Schrödinger equation directly is a grand challenge.

Currently quantum chemists mostly use the so-called one particle theories (Hartree-Fock (HF), Density Functional Theory (DFT), Local Density Approximation (LDA), etc.), which assume that the wave function is a product $\phi_1(\gamma_1)\phi_2(\gamma_2)\cdots\phi_N(\gamma_N)$.

This assumption yields a coupled system of N equations (in our case integral equations) that are then solved.

The difference between various one particle theories is in the coupling potentials.

Example

The Kohn-Sham equations result from minimization of the DFT energy functional with respect to variation of the occupied orbitals $\phi_i(r)$ which define the electron density $\rho(r) = 2 \sum_{i=1}^N |\phi_i(r)|^2$. The occupied orbitals are the lowest N eigenfunctions of the Kohn-Sham operator

$$\left(-\frac{1}{2} \nabla_i^2 + V(r) \right) \phi_i(r) = E_i \phi_i(r),$$

which implicitly depends on the orbitals through the density,

$$V(r) = - \sum_{\alpha} \frac{Z_{\alpha}}{\|r - R_{\alpha}\|} + \int \frac{\rho(r')}{\|r - r'\|} dr' + V_{xc}(r).$$

The first term accounts for the attraction of the electrons to the nuclei, the second term describes the repulsion between electrons, and the third term is the so-called exchange-correlation potential (e.g. a scalar function that depends only upon ρ).

Integral equation formulation

Consider

$$\psi_i = 2G_\mu V\psi_i,$$

where G_μ is the Green's function

$$(-\nabla^2 + \mu^2) G_\mu(r) = \delta(r - r').$$

If $\mu = \sqrt{-2E_i}$ then $\psi_i = \phi_i$. We solve this system with a fixed $\mu^{(n)}$ and, using the result, compute a new $\mu^{(n+1)}$ and solve again, etc. It turns out that these steps constitute a quadratically convergent iteration that allows us to compute E_i and the orbitals ϕ_i .

We avoid a number of problems of the original differential formulation, e.g. we do not need preconditioners, achieve full accuracy control, etc.

This iteration was first suggested by Kalos in 1962 for a Monte-Carlo approach.

The Green's function

The Green's function G_μ for one particle in free space is readily available,

$$G_\mu(r) = \frac{1}{4\pi} \frac{e^{-\mu r}}{r},$$

as well as the Coulomb potential

$$\frac{Z_\alpha}{\|r - R_\alpha\|}.$$

The question is how to apply these operators efficiently since we need to account for a very large dynamic range of these functions.

For example, we need a multiresolution representation of the kernel via wavelets but its straightforward representation in a wavelet basis is too expensive to be practical.

Separated representation of functions and operators

Standard separation of variables: $f(x_1, x_2, \dots, x_d) = \phi_1(x_1) \cdot \phi_2(x_2) \cdot \dots \cdot \phi_d(x_d)$

Definition: For a given ϵ , we represent a function $f = f(x_1, x_2, \dots, x_d)$ in dimension d as

$$\sum_{l=1}^r s_l \phi_1^l(x_1) \phi_2^l(x_2) \cdots \phi_d^l(x_d),$$

where s_l is a scalar, $s_1 \geq \dots \geq s_r > 0$, and ϕ_i^l are functions of norm one. We require the error to be less than ϵ ,

$$\left\| f - \sum_{l=1}^r s_l \phi_1^l \cdot \phi_2^l \cdot \dots \cdot \phi_d^l \right\| \leq \epsilon \|f\|.$$

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

For operator kernels, $\mathbb{A} = \mathbb{A}(x_1, x'_1, x_2, x'_2, \dots, x_d, x'_d)$, separated representation splits them as $\sum_{l=1}^r s_l A_1^l(x_1, x'_1) A_2^l(x_2, x'_2) \cdots A_d^l(x_d, x'_d)$.

Separated representations of kernels via Gaussians

Discretizing integrals

$$r^{-\alpha} = \frac{1}{\Gamma(\alpha/2)} \int_{-\infty}^{\infty} e^{-r^2 e^t + \frac{\alpha}{2} t} dt,$$

and

$$\frac{e^{-\mu r}}{r} = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-r^2 e^t/4 - \mu^2 e^{-t} + \frac{1}{2} t} dt$$

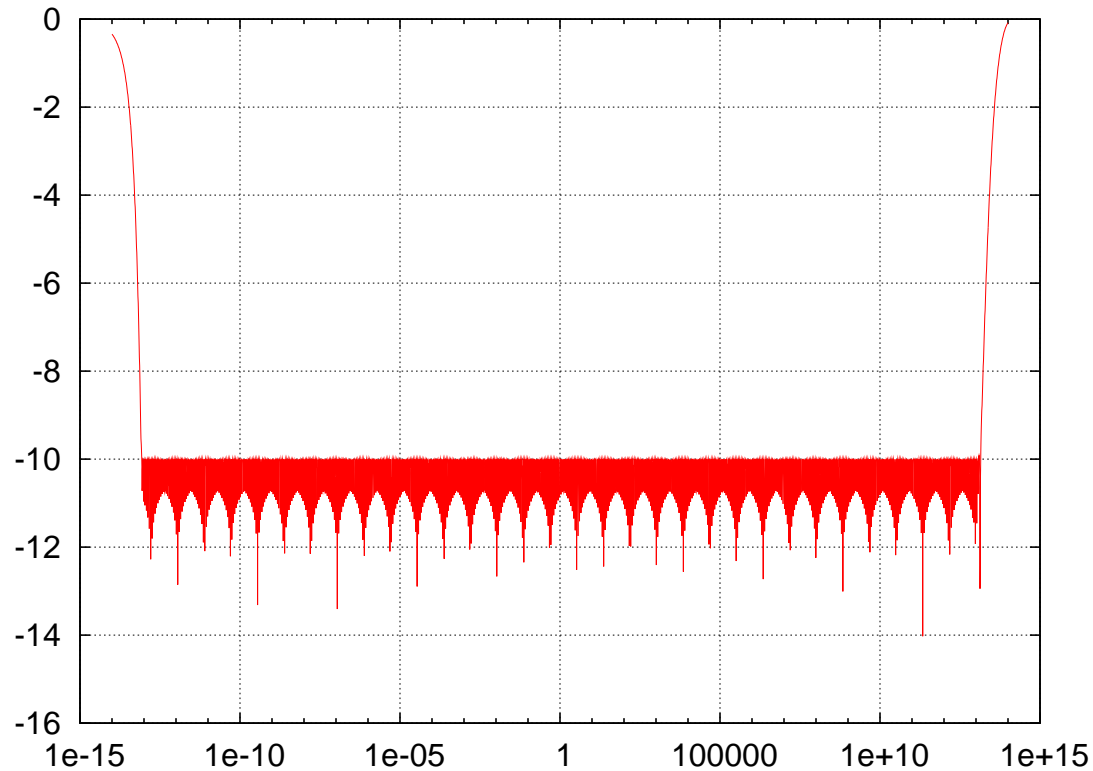
via the trapezoidal rule yields efficient separated representations. Estimates of accuracy and the number of terms are obtained using the Poisson summation, see G.B. and L. Monzón, *Approximation by exponential sums revisited*, ACHA, 28, (2010).

Example: the Poisson kernel

We have

$$\left| \frac{1}{\|\mathbf{r}\|} - \sum_{m=1}^M w_m e^{-\tau_m \|\mathbf{x}\|^2} \right| \leq \frac{\epsilon}{\|\mathbf{r}\|}, \quad \tau_m, w_m > 0, \quad \delta \leq \|\mathbf{r}\| \leq R,$$

where $M = \mathcal{O}(-\log \delta)$. Example: $\epsilon = 10^{-10}$, $\delta = 10^{-13}$, $R = 10^{13}$, $M = 303$.



Estimates using Poisson summation

Computing $\int_{\mathbb{R}} f(t)dt$:

- For any $h > 0$ and real shift s , by Poisson summation, we have

$$h \sum_{n \in \mathbb{Z}} f(s + nh) = \sum_{n \in \mathbb{Z}} \hat{f}\left(\frac{n}{h}\right) e^{2\pi i s \frac{n}{h}}, \quad \text{and}$$

$$\left| \int_{\mathbb{R}} f(t)dt - h \sum_{n \in \mathbb{Z}} f(s + nh) \right| \leq \sum_{n \neq 0} \left| \hat{f}\left(\frac{n}{h}\right) \right|, \quad \text{since } \hat{f}(0) = \int_{\mathbb{R}} f(t)dt.$$

- Fast decay of \hat{f} imply that we can choose h to achieve a small error.
- Fast decay of f yields a finite sum approximation.

Applying the idea to $r^{-\alpha}$

We have $r^{-\alpha} = \int_{-\infty}^{\infty} f(t) dt$ with

$$f(t) = \frac{e^{\alpha t}}{\Gamma(\alpha)} e^{-e^t r}, \quad \hat{f}(\xi) = \frac{\Gamma(\alpha - 2\pi i \xi)}{\Gamma(\alpha)} r^{2\pi i \xi - \alpha}.$$

Both f and \hat{f} have exponential or super exponential decay at $\pm\infty$. A relative error estimate (independent of r) follows choosing h such that

$$\sum_{n \neq 0} \frac{|\Gamma(\alpha - 2\pi i \frac{n}{h})|}{\Gamma(\alpha)} < \epsilon.$$

The choice of h depends only on ϵ and α ,

$$h \leq \frac{2\pi}{\log 3 + \alpha \log(\cos 1)^{-1} + \log \epsilon^{-1}}.$$

Multiwavelet bases

We use multiwavelet bases

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

Choice of bases:

1. The Legendre polynomials
2. The Lagrange interpolating polynomials with the Legendre nodes

Many useful properties of multiwavelet bases make this choice better than alternatives.

The non-standard form

Let T be an operator acting on a Hilbert space, $T : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$. Given multiresolution analysis (a decomposition of the Hilbert space $L^2(\mathbb{R})$ into a chain of closed subspaces),

$$\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots,$$

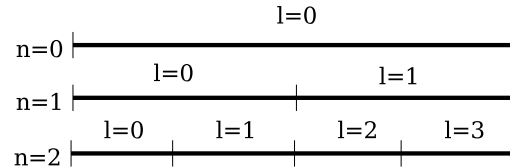
such that $V_{j+1} = V_j \oplus W_j$, we define orthogonal projection operators on subspaces V_j , $P_j : L^2(\mathbb{R}) \rightarrow V_j$ and W_j , $Q_j : L^2(\mathbb{R}) \rightarrow W_j$. Expanding operator T in a telescopic series, we obtain

$$T = \sum_{j \in \mathbb{Z}} (P_{j+1} T P_{j+1} - P_j T P_j) = \sum_{j \in \mathbb{Z}} (Q_j T Q_j + Q_j T P_j + P_j T Q_j),$$

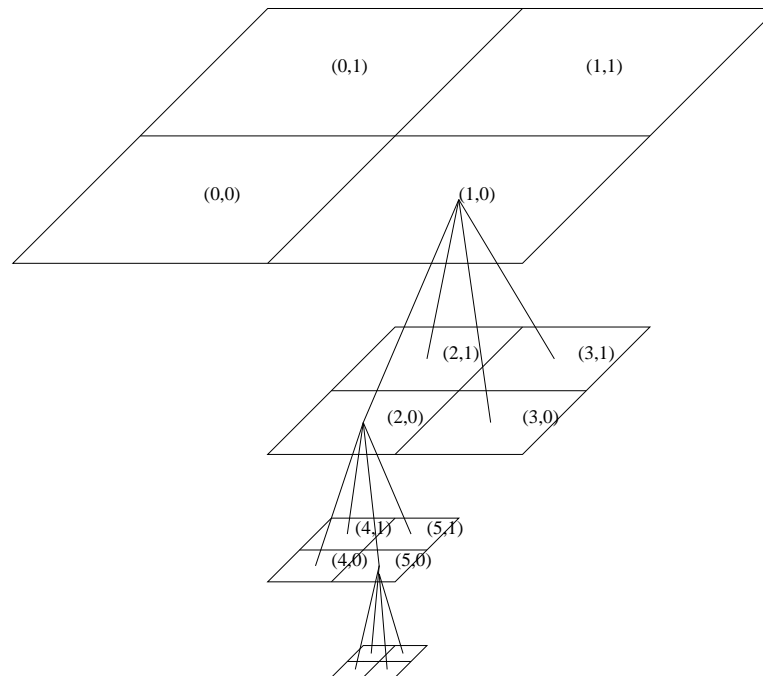
its non-standard form. In MADNESS operators are applied via their non-standard form (introduced in BCR paper, 1991).

Subdivision of space

At each scale n , divide the unit interval $[0, 1]$ into 2^n binary subintervals:

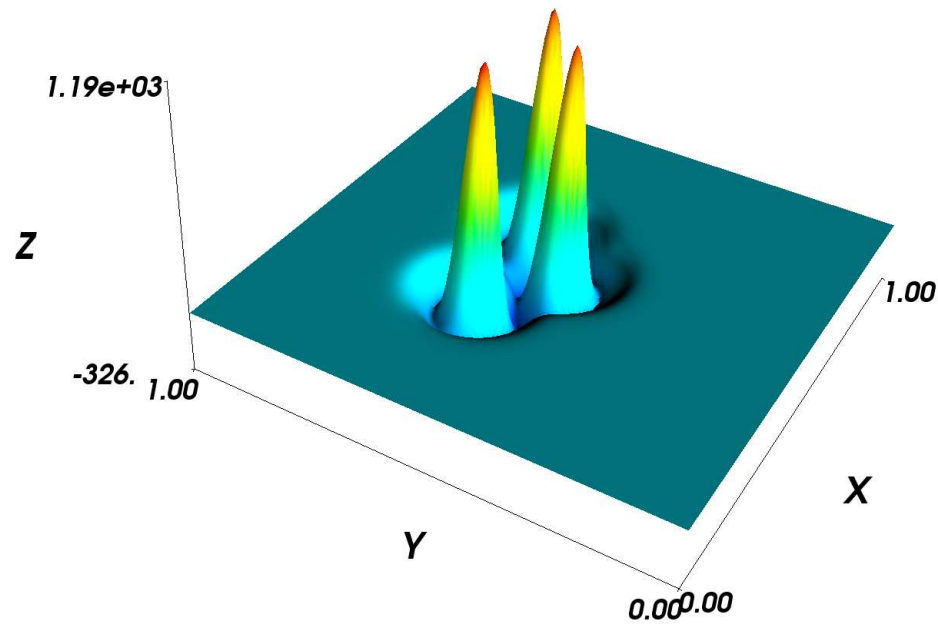


On $[0, 1] \times [0, 1]$ we have

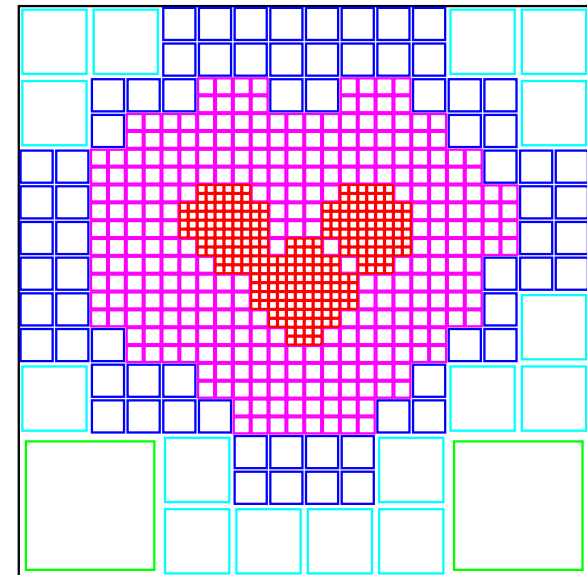


, etc.

Example of an adaptive representation



$N_{nod} = 10, \epsilon = 5.0e - 11, N_{blocks} = 634$



The cross-correlation functions of scaling functions

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

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

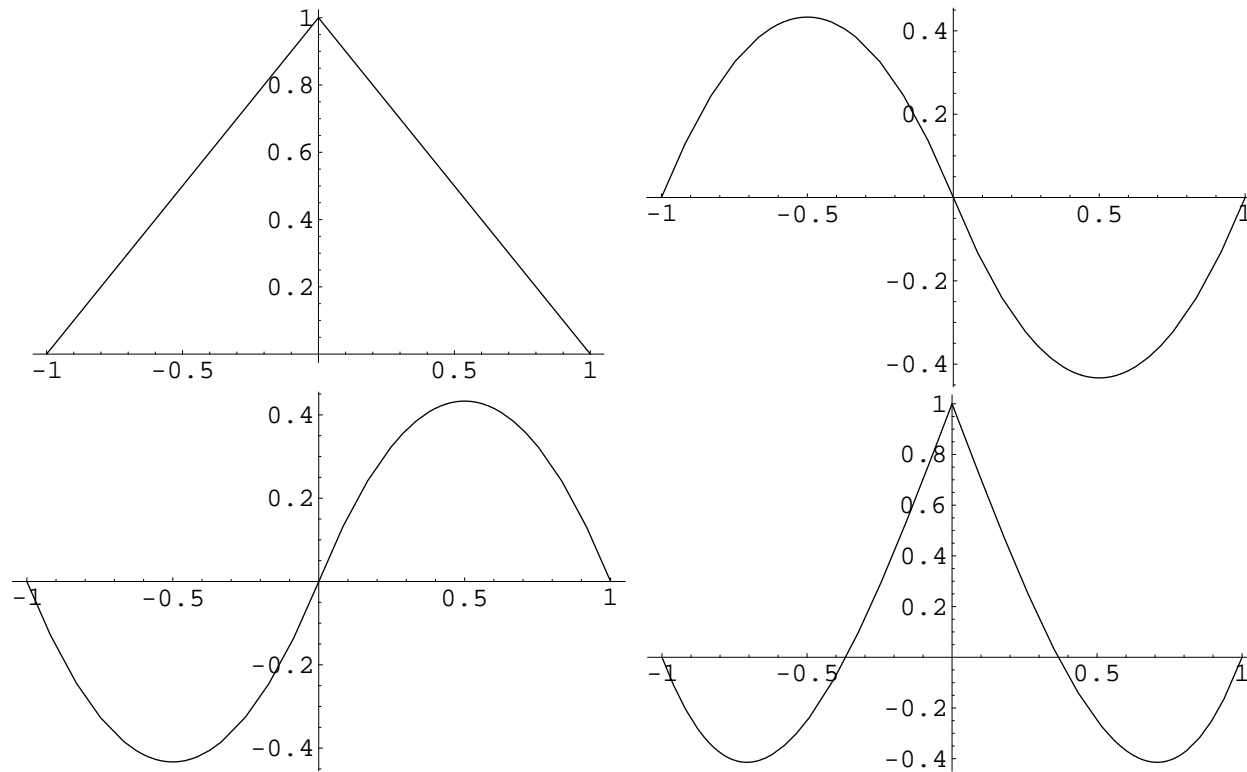
where $i, i' = 0, \dots, m-1$, m is the order of the basis, and

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

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 $[-1, 1]$. 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 Φ_{00} , Φ_{01} , Φ_{10} and Φ_{11} .

The Poisson kernel in a multiwavelet basis

Due to the homogeneity of the Poisson kernel, we have

$$t_{ii',jj',kk'}^{n;1} = 2^{-2n} t_{ii',jj',kk'}^1,$$

where

$$t_{ii',jj',kk'}^1 = t_{ii',jj',kk'}^{l_1,l_2,l_3} = \frac{1}{4\pi} \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \frac{1}{\|\mathbf{x} + \mathbf{1}\|} \Phi_{ii'}(x_1) \Phi_{jj'}(x_2) \Phi_{kk'}(x_3) d\mathbf{x},$$

and

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

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 $t_{ii',jj',kk'}^l$ have an approximation with a low separation rank,

$$r_{ii',jj',kk'}^l = \sum_{m=1}^M \frac{w_m}{b} F_{ii'}^{m,l_1} F_{jj'}^{m,l_2} F_{kk'}^{m,l_3},$$

such that

$$\left| t_{ii',jj',kk'}^l - r_{ii',jj',kk'}^l \right| \leq \frac{2\epsilon}{\pi} \quad \max_i |l_i| \geq 2$$

$$\left| t_{ii',jj',kk'}^l - r_{ii',jj',kk'}^l \right| \leq C\delta^2 + \frac{2\epsilon}{\pi} \quad \max_i |l_i| \leq 1$$

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

$b = \sqrt{3} + \|\mathbf{I}\|$, and $\delta, M = \mathcal{O}(-\log \delta) + \mathcal{O}(-\log \epsilon)$, $\tau_m, w_m, m = 1, \dots, M$ come from the separated representation of the kernel.

Projector on divergence free functions

The projector on divergence free vector functions (the so-called Leray projector, a singular operator) is given by the matrix of convolution kernels,

$$\mathbf{P}_{\iota\iota'}(\mathbf{x}) = \delta_{\iota\iota'}\delta(\mathbf{x}) - \frac{1}{4\pi} \left(\frac{\delta_{\iota\iota'}}{\|\mathbf{x}\|_2^3} - \frac{3x_\iota x_{\iota'}}{\|\mathbf{x}\|_2^5} \right),$$

where $\iota, \iota' = 1, 2, 3$ and $\delta_{\iota\iota'}$ denotes the Kronecker delta function. Multiwavelet representation requires computing (only) three type of integrals

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

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

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

Example: additional convolution operators

We have for $\operatorname{Re}(\alpha) > 0$,

$$\mathcal{G}_0^\alpha = (-\nabla^2 + \mu^2 \mathcal{I})^{-\alpha} = \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{\infty} e^{e^t \nabla^2} e^{-\mu^2 e^t} e^{\alpha t} dt,$$

so that we can approximate the kernel of this operator via Gaussians by discretizing the integral and use it in the same manner as the Green's function, $\alpha = 1$.

For example, set $\alpha = 1/2$ so that

$$\mathcal{G}_0^{-1/2} = (-\nabla^2 + \mu^2 \mathcal{I})^{-1/2}.$$

If μ is small, applying this operator to functions is tricky if the function has e.g. a large bandwidth (this is a pseudo-differential operator with the symbol $1/\sqrt{\|\xi\|^2 + \mu^2}$).

Features of MADNESS

In MADNESS the user can program at the level of equations, dial in the desired accuracy and obtain the appropriate solutions via fast algorithms without worrying about either discretization or methodology. These features of MADNESS are possible due to the following two key ingredients:

1. The kernels of physically significant operators (e.g. Green's functions of operators of mathematical physics) and potentials depend only on the distance r between interacting particles and both can be efficiently approximated (within any user-selected accuracy) by a linear combination of Gaussians. Importantly, the number of terms in such approximations depends logarithmically on all relevant parameters, e.g. ϵ^{-1} , δ^{-1} and R , where ϵ the desired accuracy and $\delta \leq r \leq R$ is the range of validity of the approximation.

The key advantage of using representations via Gaussians is that they yield a separated representation. Without such separated representations multiresolution operators would be too expensive to apply in high dimensions.

2. In order to apply Green's functions, i.e. compute integrals, separated representations via Gaussians are converted into separated multiwavelets representations. It turns out that multiwavelets (which are piece-wise polynomials and provide bases on intervals) offer more convenient multiresolution representations for numerical purposes than other possible choices of bases (e.g. Daubechies wavelets).

We emphasize that multiwavelets combined with separated representations permit an efficient computation of multidimensional integrals.

A simple example: Hartree-Fock for Helium Hydride Ion HeH^+

We solve the Hartree-Fock equation,

$$\left(-\frac{1}{2}\nabla^2 + V - 4\pi\Delta^{-1}(|\phi|^2) \right) \phi = E\phi,$$

with the potential

$$V(\mathbf{r}) = \frac{Z_1}{\|\mathbf{r} - \mathbf{R}_1\|} + \frac{Z_2}{\|\mathbf{r} - \mathbf{R}_2\|}.$$

We recasts Hartree-Fock as an integral equation and solve it via iteration.

Integral Equation

We have

$$(-\Delta + \mu^2) \phi = -2V_\phi \phi,$$

where $\mu^2 = -2E$ and $V_\phi = V - 4\pi\Delta^{-1}(|\phi|^2)$. We solve this via the following iteration:

$$\tilde{\phi} \leftarrow -2(-\Delta + \mu^2)^{-1}(V_\phi \phi),$$

$$E \leftarrow E + \frac{\langle \phi - \tilde{\phi}, V_\phi \phi \rangle}{\|\tilde{\phi}\|^2},$$

$$\phi \leftarrow \frac{\tilde{\phi}}{\|\tilde{\phi}\|},$$

$$\mu \leftarrow \sqrt{-2E}.$$

Operators with boundary conditions on simple domains

In G. B., G. Fann, R.J. Harrison, C. Kurcz and L. Monzon, *Multiresolution representation of operators with boundary conditions on simple domains*, ACHA, 33, (2012) we show that, once a free space convolution operator is approximated by a linear combination of Gaussians, we can construct a multiresolution approximation of corresponding Green's functions with periodic, Dirichlet or Neumann boundary conditions on simple domains (cubes). In fact, on fine wavelet scales these operators essentially coincide with their free space versions. Note that e.g. operator with zero boundary conditions is no longer a convolution but can be treated within the same multiresolution framework.

This construction has not been added yet to MADNESS.

Differential Operators

One “weakness” of using multiwavelets as a basis is that only first order derivative is available so that higher order derivatives have to be computed as powers of the first order derivative. This causes problems since small mismatches at the boundaries are amplified. In QC one needs to compute (at least) second order derivatives to obtain useful functionals of solutions.

This summer at a workshop at Stony Brook this issue was raised and it appears that an interesting solution to this problem has been found. It turns out that we can compute the derivative (or any other operator) in a different basis while computing only in multiwavelet basis. This has been tested as a stand-alone procedure and will be put into MADNESS.

Multiresolution Basis with a Gaussian as a scaling function

Recently in G. B., Lucas Monzón and Ignas Satkauskas, *On computing distributions of products of random variables via Gaussian multiresolution analysis*, arXiv:1611.08580, we constructed multiresolution basis where the scaling function is well approximated by a Gaussian.

We are working to see if multiwavelets can be replaced by such basis: the difficulty is that this is not an orthogonal basis so that additional algorithms are needed for its effective use.