Separated Representations and Fast Adaptive Algorithms in Multiple Dimensions

Gregory Beylkin
(with M. Mohlenkamp, L. Monzón, R. Cramer, V. Cheruvu and R. Harrison & G. Fann (ORNL))

University of Colorado at Boulder

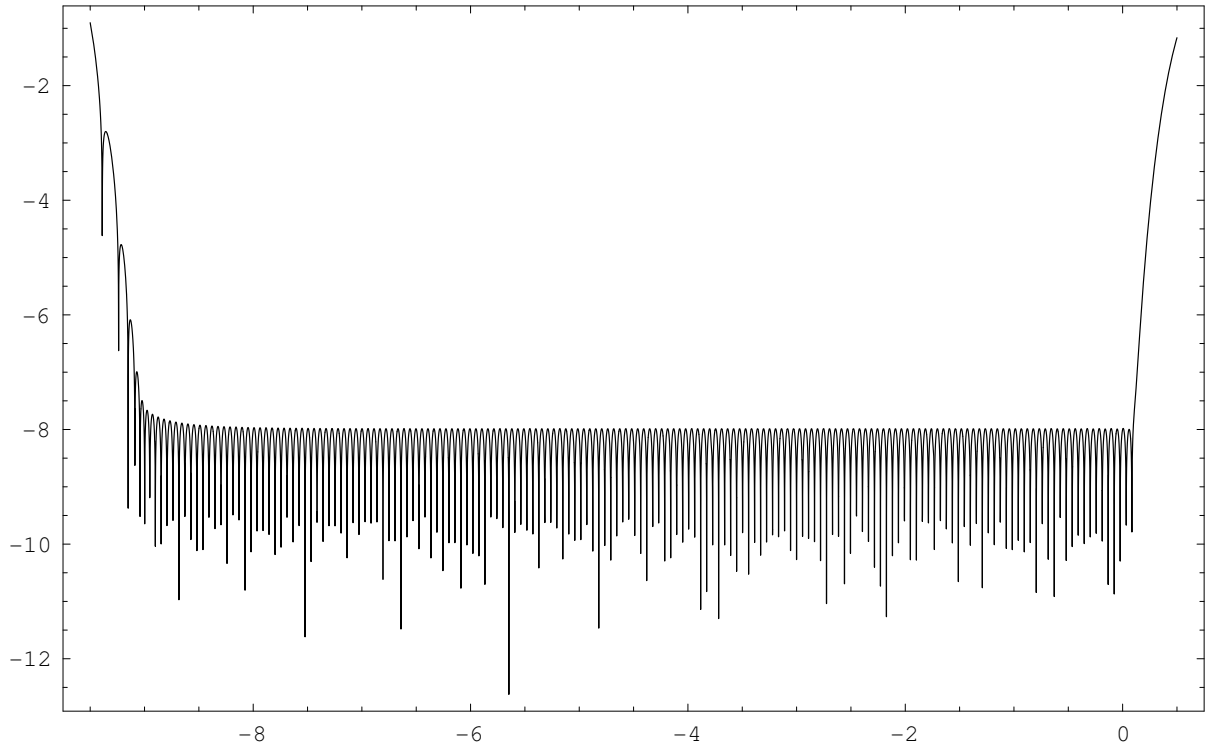
CSCAMM Program
FMM & Related Algorithms
University of Maryland
April 28, 2004

Start with an example

We have

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

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



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

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

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

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

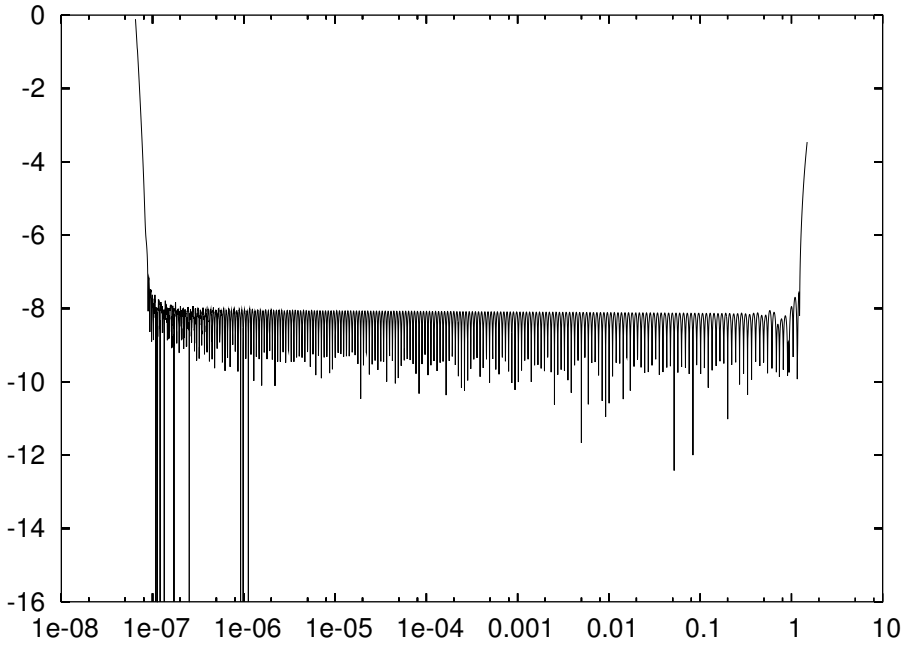
$b = \sqrt{3} + ||\mathbf{l}||$, and δ, $M = O(-\log \delta) + O(-\log \epsilon)$, p_m, w_m, $m = 1, \ldots, M$ are from the separated representation of the kernel.

The projector on the divergence-free functions

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 $10^{-7} \leq ||r|| \leq 1$.

Other Examples

1. $e^{iKr}/r \approx \sum_m w_m e^{\tau_m r^2}$ for the Helmholtz kernel, $K \sim 100$
2. $J_0(x)$ ....
3. $\sin(x)/x$
4. $D_N(x)$
5. $\pi \cot(\pi x)$
6. $\log(\sin(\pi x))$
7. $\tanh(x)$
8. ...

Quantum Chemistry

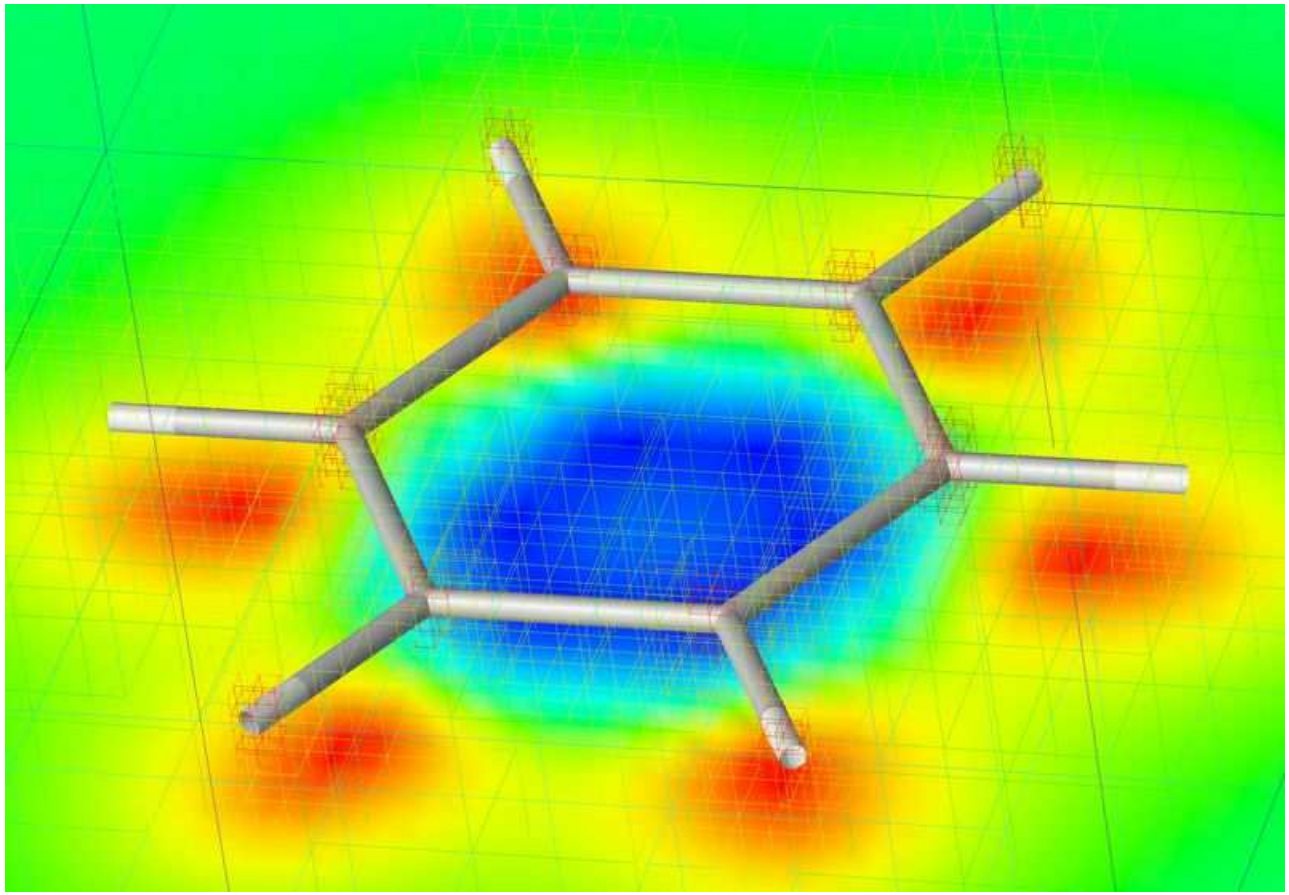
With R. Harrison, G. Fann and T. Yanai

- Complete elimination of the basis error
- Implementation for one-electron models (HF, DFT)
- Most accurate computations up to now (within these models)
- Correct scaling of cost with system size
- Much smaller computer code than "Gaussians" (<— R. Harrison)

With R. Cramer, V. Cheruvu and F. Pérez

- Adaptive PDE solvers
- Operator calculus in 3D

Examples: elements, small molecules...



Adaptive subdivision of space for the benzene molecule C_6H_6
(from R. Harrison, G. Fann and G. Beylkin)

Current 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.

All 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
t/block (s)	0.07	0.1	0.16	0.16
MFLOPS	171	317	430	505

Some comments:

- MADNESS (R. Harrison et. al, ORNL) already implements improvements which make it at least $50X$ faster than our code (our improvements will be different but should produce a similar speedup)

- Already very competitive with multigrid codes for small ϵ, where multigrid slows down dramatically.

- Need to test against FMM-adaptive results in 2d (Greengard & Etheridge 2001).

- Python overhead measured at $\sim 1 - 2\%$.

Problem

- **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.

Beyond the definition

- How to construct such representations?
- How to use them?
- What is the class of operators that yields a small separation rank for a given ϵ?

Trivial examples:

- Identity: $\mathbb{I} = \mathbb{I}_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \mathbb{I}_d$
- Laplacian: $\mathcal{T} = \Delta_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \mathbb{I}_d + \mathbb{I}_1 \otimes \Delta_2 \otimes \cdots \otimes \mathbb{I}_d + \ldots + \mathbb{I}_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \Delta_d$

Non-trivial examples to follow...

First example: a new trigonometric identity

Consider the function $\sin(x_1 + x_2 + \ldots + x_d)$ and represent it in the separated form. The usual trigonometric formulas produce 2^{d-1} terms.

Lemma:

$$\sin(\sum_{j=1}^{d} x_j) = \sum_{j=1}^{d} \sin(x_j) \prod_{k=1, k\neq j}^{d} \frac{\sin(x_k + \alpha_k - \alpha_j)}{\sin(\alpha_k - \alpha_j)},$$

for all choices of $\alpha_1, \alpha_2, \ldots, \alpha_d$, such that $\sin(\alpha_k - \alpha_j) \neq 0$ for all $j \neq k$.

Observations:

- Only d terms
- Separated representation is not unique
- It can be ill-conditioned!

Condition number of separated representation

Definition:

We call the ratio

$$\kappa = \frac{\sum_{l=1}^{r} s_l}{||\mathbb{A}||},$$

the condition number of separated representation.

It is a natural definition since $||s_l \, \mathbb{A}_1^l \otimes \mathbb{A}_2^l \otimes \cdots \otimes \mathbb{A}_d^l|| = s_l$.

We need to maintain $\kappa\mu||\mathbb{A}|| \leq \epsilon$, where μ is the machine roundoff.

Similarities and differences with SVD

If $d = 2$ then the separated representation can be obtained via SVD.

(Actually, we use a much simpler algorithm since we do not insist on orthogonality between vectors in a given direction).

If $d \geq 3$ then the analogy with SVD breaks down: by changing ϵ we change all terms in the representation rather than add/delete terms

Many attempts to treat separated representation as a generalization of SVD but the construction depends on ϵ !

The Poisson kernel in higher dimensions

Theorem: For any $\epsilon > 0$, the Poisson kernel

$$\frac{1}{(d-2)\,\Omega_d}\,\frac{1}{|x|^{d-2}},$$

when applied to functions with the Fourier transform supported in the annulus $(\sum_{i=1}^{d} \xi_i^2)^{1/2} \in [\sqrt{\delta}, \sqrt{D}]$, can be approximated to within ϵ by

$$\sum_{l=1}^{r} \alpha_l \bigotimes_{i=1}^{d} \frac{1}{\sqrt{4\pi\tau_l}} \exp(-x_i^2/4\tau_l),$$

where the separation rank grows as $r = \mathcal{O}\left(\log(D/(\delta\epsilon))\right)$.

For functions with the Fourier transform in the "cubic annulus", the separation rank grows as $r = \mathcal{O}\left(\log(d^2 D/(\delta\epsilon))\right)$.

Multiparticle Schrödinger operator

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

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

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 $1/r_i$ and the electron-electron potential W_{im} is multiplication by $1/|r_i - r_m|$.

$\mathcal{T} = -\Delta_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \mathbb{I}_N - \mathbb{I}_1 \otimes \Delta_2 \otimes \cdots \otimes \mathbb{I}_N - \ldots - \mathbb{I}_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \Delta_N$

$\mathbb{V} = -V_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \mathbb{I}_N - \mathbb{I}_1 \otimes V_2 \otimes \cdots \otimes \mathbb{I}_N - \ldots - \mathbb{I}_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes V_N$

$W = \ldots$ has $\mathcal{O}(N^2)$ terms.

Thus, the nominal separation rank grows as $\mathcal{O}(N^2)$.

It turns out that ...

Separation rank of the Schrödinger operator

Theorem:

The representation of $\mathcal{T} + \mathbb{V}$ to within ϵ in the operator norm has separation rank

$$\mathrm{r}{=}\mathcal{O}\left(\frac{\log(N\,||-\Delta_1 - V_1||/\epsilon)}{\log(1/\mu) - \log(N\,||-\Delta_1 - V_1||/\epsilon)}\right).$$

Let us symmetrically separate $W_{im} = \sum_{k=1}^{r_w} W_i^k\, W_m^k + \mathcal{O}(\epsilon)$. For each value of k the operator has the form $\mathcal{A} = \sum_{i=1}^{N-1} \sum_{m=i+1}^{N} \mathcal{A}_i\, \mathcal{A}_m$.

Theorem:

The representation of $\mathcal{A}$ to within ϵ in the operator norm has separation rank

$$\mathrm{r}{=}\mathcal{O}\left(\frac{\log(N^2\,||\mathcal{A}_1||^2/\epsilon)}{\log(1/\mu) - \log(N^2\,||\mathcal{A}_1||^2/\epsilon)}\right).$$

Thus, the separation rank of the Schrödinger operator grows only as $\log(N)$

Constructive proof

Consider

$$\mathbb{G}(t) = ||-\Delta_1 - V_1|| \bigotimes_{i=1}^{N} (\mathbb{I}_i + t\frac{-\Delta_i - V_i}{||-\Delta_1 - V_1||}),$$

and note that $\mathbb{G}'(0) = \mathcal{T} + \mathbb{V}$. Using finite difference formula of order r, we approximate

$$\mathbb{G}'(0) \approx \sum_{j=1}^{r} \alpha_j \mathbb{G}'(t_j).$$

Similarly, we using $\mathcal{A}_i$ instead of $-\Delta_i - V_i$ in the definition of $\mathbb{G}$, we note that $\mathbb{G}''(0) = 2\mathcal{A}/||\mathcal{A}_1||$ and use the same approach.

(This approximation was first discovered numerically).

Linear algebra for separated representations

- Addition and multiplication of operators in separated form leads to operators of the same form
- $\mathcal{O}(d \cdot r \cdot M^2)$ or $\mathcal{O}(d \cdot r \cdot M)$ to store
- $\mathcal{O}(d \cdot \tilde{r}\hat{r} \cdot M^3)$ (or better) to multiply operators
- Separation rank grows, e.g. $\tilde{r}\hat{r}$ for multiplication of operators

We need to reduce the separation rank.

Reduction of separation rank via alternating least squares

Core algorithm:

We start with the initial approximation of separation rank $r' < \tilde{r}\hat{r}$,

$$\sum_{l=1}^{r'} s_l \, \mathbb{A}_1^l \otimes \mathbb{A}_2^l \otimes \cdots \otimes \mathbb{A}_d^l,$$

and then iteratively refine it. We refine one direction, k, at a time, and then alternate the directions, $k = 1, 2, \ldots, d$. To refine in direction k, we fix the vectors in the other directions and then solve for new $\mathbb{A}_k^l$ and s_l to minimize the residual. It is easy to show that the residual decreases at each step. Each refinement requires solving a linear least squares problem to minimize the residual.

We detect if the residual does not change anymore and, if it is not small enough, increase the separation rank r' , etc.

Some details

- Convergence is not guaranteed but the algorithm performs well in practice
- The condition number of separated representation is included into least squares
- The algorithm loses $1/2$ of the digits, so that the best accuracy obtainable is $\epsilon = \sqrt{\mu}$, where μ is the machine roundoff
- One iteration cost roughly $\mathcal{O}(d \cdot \tilde{r}^3 \cdot M)$.
- Several preliminary steps to reduce computational cost
- We have discovered a new trigonometric identity and an approximation technique using this algorithm

Example

Simplified single-atom N-electron model

$$\mathbb{H} = -\Delta + 2N \sum_{i=1}^{N} \cos(x_i) + \sum_{i=1}^{N} \sum_{k>i}^{N} \cos(x_i - x_k)$$

- The number of "particles" $N = 30$
- We use power method for $\mathbb{A} = C_N \mathbb{I} - \mathbb{H}$, where $C_N \approx ||\mathbb{H}||/2$
- Nominal computational cost: 10^{80}

Results of computation

$\mathbf{F}_k$ is the scalar multiple of the eigenvector corresponding to the smallest eigenvalue.

$\mathbf{F}_0$ is chosen with the separation rank one.

Accuracy is increased gradually.

ϵ	iterations	$r(\mathbf{F}_k)$	time (sec)
1e-2	521	1	2.7e+1
1e-3	2557	2	1.4e+2
1e-4	4130	5	1.5e+3
1e-5	5230	6	6.4e+3
1e-6	6160	11	2.0e+4
1e-7	6368	12	2.2e+4

Solving linear systems in separated representation

Linear system $\mathbb{A}\mathbf{F} = \mathbf{G}$, where $\mathbb{A}$ is the Laplacian in dimension 20 , size in each direction $M = 30$

$r(\mathbf{F})$	$\|\|\mathbb{A}\mathbf{F} - \mathbf{G}\|\|/\|\|\mathbf{G}\|\|$	time
1	$2.5 \cdot 10^{-2}$	13
3	$3.8 \cdot 10^{-3}$	84
5	$6.8 \cdot 10^{-4}$	213
9	$8.0 \cdot 10^{-5}$	789
13	$8.4 \cdot 10^{-6}$	2048
19	$9.5 \cdot 10^{-7}$	6121

$\mathbf{G}$ is random, with separation rank one.

Separation rank of the solution cannot exceed that of the inverse Laplacian

Antisymmetry

Since electrons are fermions, the wave function must be antisymmetric, e.g., $\psi(\gamma_2, \gamma_1, \gamma_3, \ldots) = -\psi(\gamma_1, \gamma_2, \gamma_3, \ldots)$, where $\gamma = ((x, y, z), \sigma)$ and σ is the spin.

Given a function of N variables, its "antisymmetrizer" is defined by

$$\mathcal{A} = \frac{1}{N!} \sum_{p \in S_N} (-1)^p \mathcal{P},$$

where S_N is the permutation group on N elements. If $\mathcal{A}$ is applied to a separable function, then the result can be expressed as a Slater determinant,

$$\mathcal{A} \prod_{j=1}^{N} \phi_j(\gamma_j) = \frac{1}{N!} \begin{vmatrix} \phi_1(\gamma_1) & \phi_1(\gamma_2) & \cdots & \phi_1(\gamma_N) \\ \phi_2(\gamma_1) & \phi_2(\gamma_2) & \cdots & \phi_2(\gamma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\gamma_1) & \phi_N(\gamma_2) & \cdots & \phi_N(\gamma_N) \end{vmatrix}.$$

Do we have a problem?

The number of terms in $\mathcal{A}\prod_{j=1}^{N}\phi_j(\gamma_j)$ grows exponentially fast and, although this number can algebraically be reduced somewhat, are we in trouble?

If we care only about computing inner products with $\mathcal{A}\prod_{j=1}^{N}\phi_j(\gamma_j)$, then the so-called Löwdin rules provide a solution,

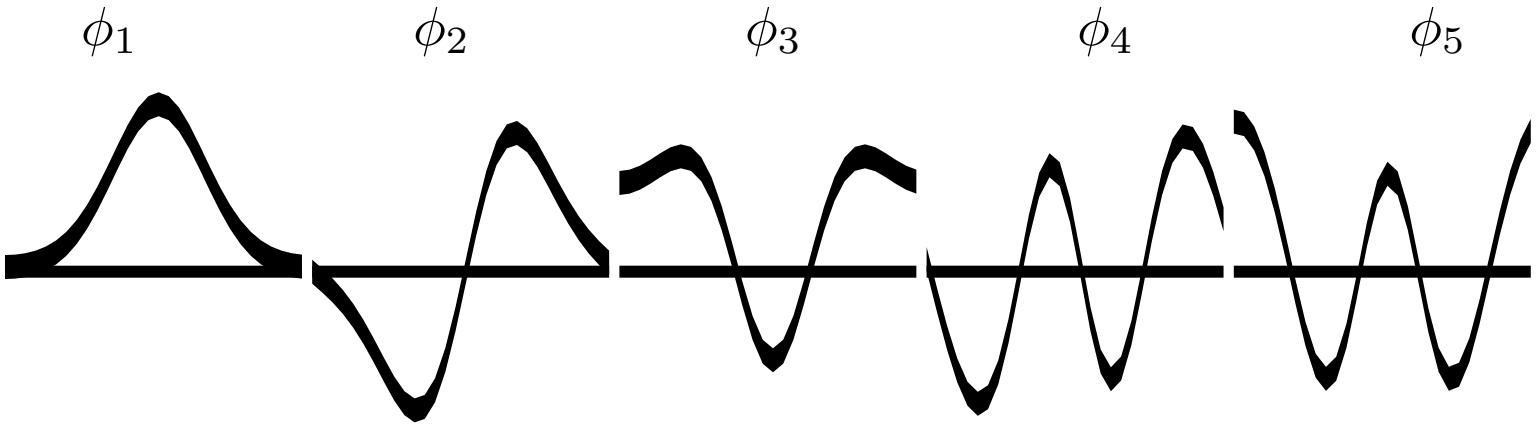
$$\langle \mathcal{A}\prod_{j=1}^{N}\phi_j(\gamma_j), \mathcal{A}\prod_{j=1}^{N}\tilde{\phi}_j(\gamma_j)\rangle = \frac{1}{N!}\begin{vmatrix} \langle\phi_1,\tilde{\phi}_1\rangle & \langle\phi_1,\tilde{\phi}_2\rangle & \cdots & \langle\phi_1,\tilde{\phi}_N\rangle \\ \langle\phi_2,\tilde{\phi}_1\rangle & \langle\phi_2,\tilde{\phi}_2\rangle & \cdots & \langle\phi_2,\tilde{\phi}_N\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\phi_N,\tilde{\phi}_1\rangle & \langle\phi_N,\tilde{\phi}_2\rangle & \cdots & \langle\phi_N,\tilde{\phi}_N\rangle \end{vmatrix}.$$

Computing determinant costs at most $\mathcal{O}(N^3)$, but for large N the matrix is banded and the cost is $\mathcal{O}(N)$, so that we are O.K.

Example of computing the antisymmetric ground state

Separated rank, achieved approximation, and eigenvalue estimates for the separable ($\mathbf{F}_0$) and main approximations to the wavefunction.

	r	ϵ	$\|\|\mathcal{A}\mathbb{H}\mathbf{F}$	$\langle \mathcal{A}\mathbb{H}\mathbf{F}, \mathcal{A}\mathbf{F} \rangle$
$\mathbf{F}_0$	1	$3.4 \cdot 10^{-3}$	322.6727395	322.3859013
$\mathbf{F}$	2	10^{-4}	321.8852595	321.8844158

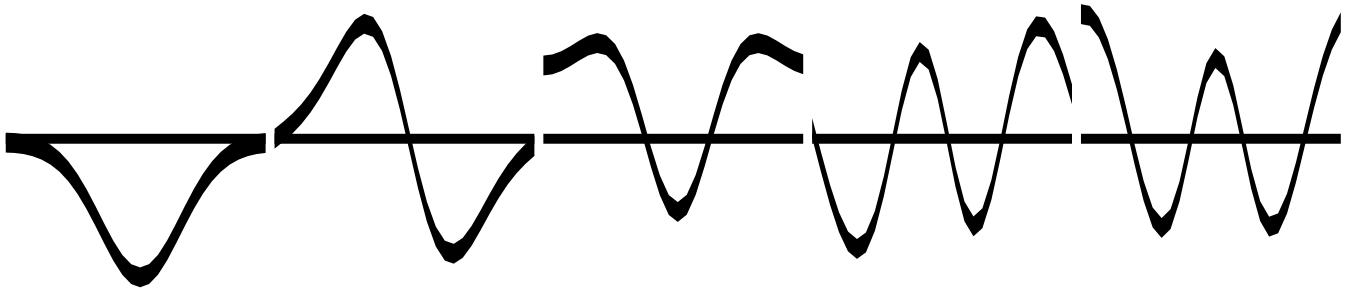


The computed separable approximation $\mathbf{F}_0$ to the wave function.

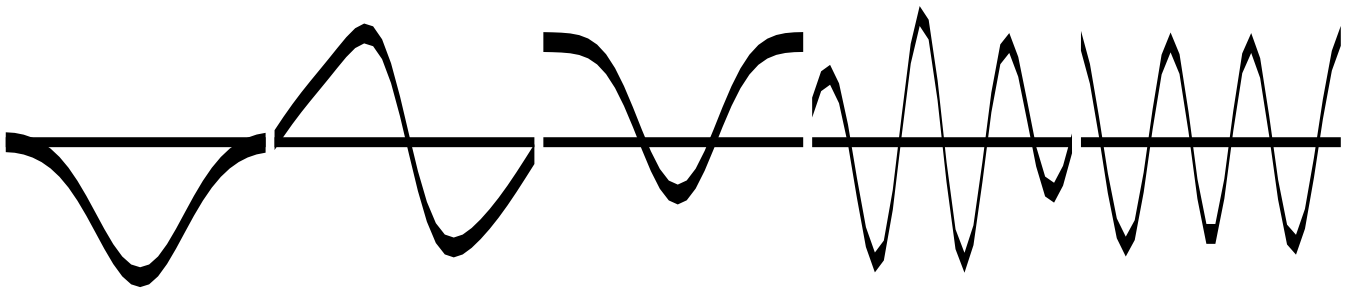
Structure of the antisymmetric ground state

$i = 1$ $i = 2$ $i = 3$ $i = 4$ $i = 5$

$l = 1;\ 0.999350$



$l = 2;\ 0.033093$



Conclusions and future work

- Powerful method for multidimensional problems
- Lattice sums
- Operators in 3D (e.g., oscillatory Green's functions)
- Operators in 6D for multiresolution quantum chemistry (two-electron models)
- We are attempting to solve the multiparticle Schrödinger equation
- Complete MADNESS (Multiresolution ADaptive NumErical Scientific Simulation)