

Abstract

By representing functions of many variables as sums of separable functions, one obtains a method to bypass the curse of dimensionality. I will discuss efforts to develop, understand, and use this method, both in a general context and for applications in quantum mechanics.

Computing with Sums of Separable Functions, with Applications in Quantum Mechanics

Martin J. Mohlenkamp

Department of Mathematics



Unifying Theme: Sums of Separable Functions

The Curse of Dimensionality can be bypassed if we can approximate

$$f(\mathbf{x}) = f(x_1, \dots, x_d) \approx \sum_{l=1}^r s_l \prod_{i=1}^d f_i^l(x_i)$$

well with small separation rank r .

Why should this approximation be effective?

How do we construct and use it within an application?

“Why” has us mostly stumped, so we concentrate on “how” and hope it will eventually help with “why”.

Main Branches of Activity

- Feebly exploring “why” .
- General tools for scientific computing.
In $d \sim 3$ provides acceleration, for $d \gg 3$ enables new areas.
- Regression (machine learning, classification, control).
- Quantum Mechanics.

Exploring Why: Classes of Functions

Temlyakov shows that for functions in the class W_2^k , which is characterized using partial derivatives of order k , there is a separated representation with separation rank r that has error

$$\epsilon = \mathcal{O}(r^{-kd/(d-1)}).$$

However, a careful analysis of the proof shows that the 'constant' in the $\mathcal{O}(\cdot)$ is at least $(d!)^{2k}$ and the inductive argument can only run if $r \geq d!$.

Challenge: Give a non-trivial characterization of functions with low separation rank.

Hint: Do not use derivatives.

Exploring Why: Example: Additive Model

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^d g_i(x_i) = \frac{d}{dt} \left[\prod_{i=1}^d (1 + tg_i(x_i)) \right]_{t=0} \\ &= \lim_{h \rightarrow 0} \frac{1}{2h} \left(\prod_{i=1}^d (1 + hg_i(x_i)) - \prod_{i=1}^d (1 - hg_i(x_i)) \right), \end{aligned}$$

so we can approximate a function that naively would have $r = d$ using only $r = 2$.

This formula provides a reduction of addition to multiplication; it is connected to exponentiation, since one could use $\exp(\pm hg_i(x_i))$ instead of $1 \pm hg_i(x_i)$.

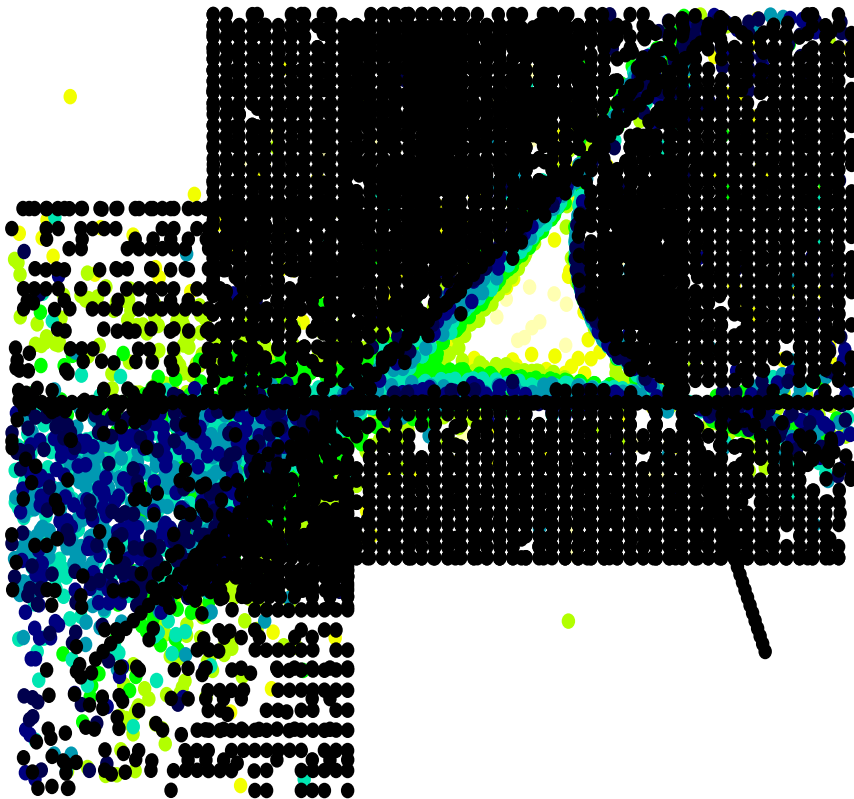
Conjecture: This mechanism is the key.

Exploring Why: Topology and Geometry

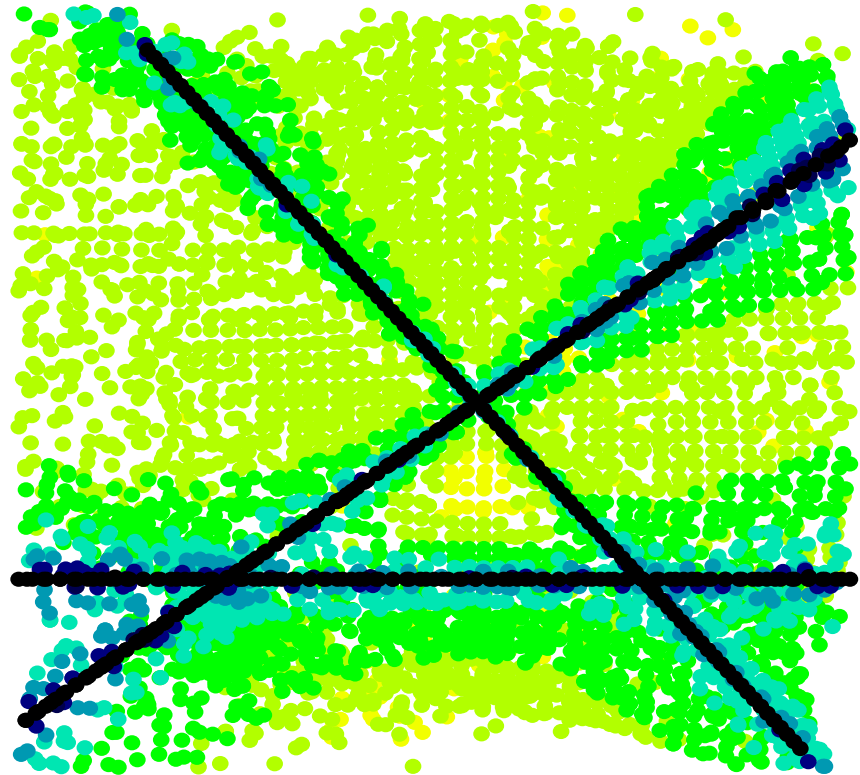
The set of $r = 2$ functions/tensors is neither open nor closed, therefore it is interesting.

Challenge: Describe the geometry and/or topology of this set.

$r = 2$ slice through $2 \times 2 \times 2$.



$r = 2$ slice through $3 \times 3 \times 3$.



Exploring Why: Lessons Learned

- The “obvious” (analytic) separated representation may be woefully inefficient.
- Representations are non-unique, and that is good.
- It is **essential** that $f_i^l(x_i)$ not be constrained. Orthogonality is bad.

General Tools in High Dimensions

“Strong” computational paradigm: Apply a sum of separable operators to a sum of separable functions, get a sum of separable functions with more terms, then reduce the number of terms with a least-squares fitting.

“Weak” computational paradigm: Fit a sum of separable functions to what you would get if you applied some (ugly) operator to a sum of separable functions.

Insight: You do not need a thing explicitly in order to fit to it; you only need to be able to compute inner products with it.

The weak paradigm allows a wider class of operators, but usually does not allow measurement of the fitting error.

General Tools: Fitting Algorithm

All our fitting is based on Alternating Least Squares (ALS), which is robust but

- slow and
- prone to local minima.

There has been work on other algorithms, but they are not convincingly better.

Challenge: Produce a convincingly better algorithm or concrete improvement within ALS.

General Tools: an Opinionated Opinion

To understand high dimensions we should study functions and operators, and **not** vectors, matrices, and tensors.

- That is what we really have. Ditch Galerkin, go adaptive!
- Gets to the intrinsic issues and gives cleaner proofs.
- Avoids the “false friend” of flattening.

Regression

Given scattered data

$$\{((x_1^j, \dots, x_d^j), y_j)\}_{j=1}^N = \{(\mathbf{x}^j, y_j)\}_{j=1}^N$$

construct a function f so that $f(\mathbf{x}^j) \approx y_j$ and $f(\mathbf{x})$ is reasonable for other \mathbf{x} . Using sums of separable functions enables an $\mathcal{O}(r^2 d N)$ algorithm.

Classification: Let y_j be class labels.

Learning Physics: Let \mathbf{x}^j be a representation of a molecular or material structure and y_j be a physical property.

Control: Let \mathbf{x}^j be a situation and y_j a control parameter that we experienced as having a good result.

Quantum Mechanics: Overview

My main project, supported by the NSF (**Thanks!**).

- Why does (might) it work? (connects to general why)
- Antisymmetry constraint and interelectron interaction operator require weak formulation. (done)
- Size-consistency requires hierarchy of sums of products. (in progress, is painful with antisymmetry.)
- Interelectron cusp requires geminals. (painful, on hold)

The **multiparticle Schrödinger equation** is the basic governing equation in Quantum Mechanics.

The **wavefunction** has one 3D spatial variable $\mathbf{r} = (x, y, z)$ **per electron**, and so looks like $\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$.

The **kinetic energy operator** is $\mathcal{T} = -\frac{1}{2} \sum_{i=1}^N \Delta_i$.

The **nuclear potential operator** is $\mathcal{V} = \sum_{i=1}^N V(\mathbf{r}_i)$.

The **electron-electron interaction operator** is

$$\mathcal{W} = \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|}$$

Find the Low(est) Eigenvalues to get Energies

$$\mathcal{H}\psi = (\mathcal{T} + \mathcal{V} + \mathcal{W})\psi = \lambda\psi$$

subject to an **antisymmetry constraint**, e.g.

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = -\psi(\mathbf{r}_2, \mathbf{r}_1, \dots, \mathbf{r}_N).$$

The *antisymmetrizer* \mathcal{A} converts a product to a Slater determinant, so we consider

$$\psi_{(r)} = \mathcal{A} \sum_{l=1}^r s_l \prod_{i=1}^N \phi_i^l(\mathbf{r}_i) = \frac{1}{N!} \sum_{l=1}^r s_l \begin{vmatrix} \phi_1^l(\mathbf{r}_1) & \phi_1^l(\mathbf{r}_2) & \cdots & \phi_1^l(\mathbf{r}_N) \\ \phi_2^l(\mathbf{r}_1) & \phi_2^l(\mathbf{r}_2) & \cdots & \phi_2^l(\mathbf{r}_N) \\ \vdots & \vdots & & \vdots \\ \phi_N^l(\mathbf{r}_1) & \phi_N^l(\mathbf{r}_2) & \cdots & \phi_N^l(\mathbf{r}_N) \end{vmatrix}.$$

Quantum Mechanics: Sketch of the Basic Method

1. Convert the eigenproblem to a Green's function iteration.
2. Modify the iteration to a \mathcal{A} -least-squares fitting problem.
3. Collapse that to a set of one-electron least-squares fitting problems using ALS.
4. Update the one-electron functions using:
 - an expansion of the Green's function into Gaussian convolutions,
 - formulas involving the nuclear potential and the Poisson kernel, and
 - an adaptive numerical method for operating on one-electron functions.

The basic operating unit is a function, as opposed to a number, vector, or matrix.

Quantum Mechanics: Hierarchy and Center-of-Mass

To scale well with the number of subsystems, use

$$\psi \approx \mathcal{A} \sum_{l=1}^r \prod_{\text{subsystems}} \left(\sum_{\text{electrons in the subsystem}} \prod \phi \right).$$

We must compute $\langle \cdot, \cdot \rangle_{\mathcal{A}}$ (with \mathcal{V} , \mathcal{W} , and the Green's function) without multiplying out $\prod_{\text{subsystems}}$. A center-of-mass principle

can be applied, but requires expansions of determinants of sums

$$|\mathbb{A} + \mathbb{B}| = \sum_{k=0}^{|\alpha_0|} \sum_{\substack{\alpha \subset \alpha_0, \beta \subset \beta_0 \\ |\alpha| = |\beta| = k}} (-1)^{\sigma(\alpha \subset \alpha_0) + \sigma(\beta \subset \beta_0)} |\mathbb{A}[\alpha_0 \setminus \alpha; \beta_0 \setminus \beta]| \cdot |\mathbb{B}[\alpha; \beta]|.$$

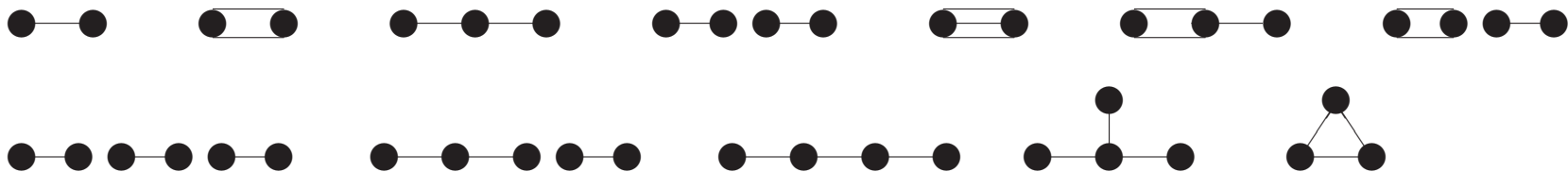
Ouch! Help!

Quantum Mechanics: Geminals

To account for the interelectron cusp, use

$$\psi \approx \mathcal{A} \sum_{p=0}^P \left[\left(\sum_{i \neq j} w_p(\|\mathbf{r}_i - \mathbf{r}_j\|) \right) \left(\sum_{l=1}^{r_p} \prod_{j=1}^N \phi_j^{lp}(\mathbf{r}_j) \right) \right]$$

When used in $\langle \mathcal{W}\psi, \psi \rangle_{\mathcal{A}}$ we get geminals connecting up to 3 pairs of variables, in the patterns



We get up to 6 entangled indices and 6 entangled variables. Disentangling them is a challenge, and has parallels to the tensor contraction problem. **Ouch! Help!**

Summary

Challenge: Give a non-trivial characterization of functions with low separation rank.

Conjecture: The additive model mechanism is the key.

Challenge: Describe the geometry and/or topology of the set of low-rank sums of separable functions.

Challenge: Produce a convincingly better algorithm or concrete improvement within ALS.

Opinion: Study functions, not tensors; flattening is a false friend.

Ouch! Help! Provide more effective methods for determinants of sums.

Ouch! Help! Provide automatic logic for contracting multiple variables and indices.

Our understanding of “why” is limited, but “how” proceeds.