The Magical Inner Product

Martin J. Mohlenkamp

Department of Mathematics

OHIO
UNIVERSITY

http://www.ohio.edu/people/mohlenka/

April 2014

Abstract

If you can define and compute an appropriate inner product, then you can approximate a function of many variables by a sum of separable functions using the alternating least squares (ALS) algorithm. In the simplest case, the inner product just defines the ordinary least-squares error. More exotic inner products allow one to regularize, incorporate symmetries, or fit to data. Non-separable structures and operators can be included as long as one can compute inner products involving them. I will explain the central role of the inner product and describe some fun ones I have worked with.

Outline

Approximation by Sums of Separable Objects

Try to approximate

$$f(x_1, x_2, \ldots, x_d) \approx$$

$$g(x_1, x_2, \ldots, x_d) = \sum_{j=1}^{r} g^j(x_1, x_2, \ldots, x_d) = \sum_{j=1}^{r} \prod_{i=1}^{d} g_i^j(x_i)$$

for $x_i \in X_i$. If the domains X_i are finite, then these are "tensors".

For fixed r, find $\{g_i^j\}$ to minimize the least-squares error function

$$E(g) = \|f - g\|^2 = \langle (f - g), (f - g) \rangle = \langle f, f \rangle - 2\langle f, g \rangle + \langle g, g \rangle .$$

The Alternating Least Squares Algorithm

Make an initial guess for g.

Loop until happy:

Loop through the directions $k = 1, \ldots, d$:

Solve a linear least-squares problem for new $\{g_k^j\}_j$
while fixing $\{g_i^j\}$ for $i \neq k$.

If $E(g)$ stabilizes but is too large, then try again,
perhaps with larger r.

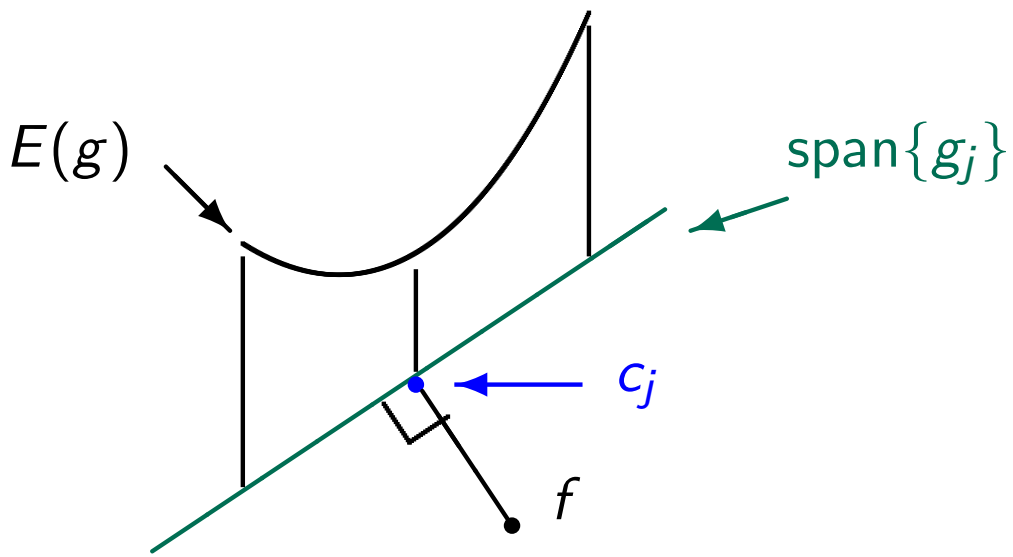
Refresher on Ordinary Linear Least-Squares

If g is a linear combination of fixed elements with unknown coefficients and we want to minimize

$$E(g) = \left\| f - \sum_j c_j g_j \right\|_2^2 = \left\langle f - \sum_j c_j g_j, f - \sum_j c_j g_j \right\rangle ,$$

the solution is the orthogonal projection of f onto $\mathrm{span}\{g_j\}$, with coefficients determined by the normal equations

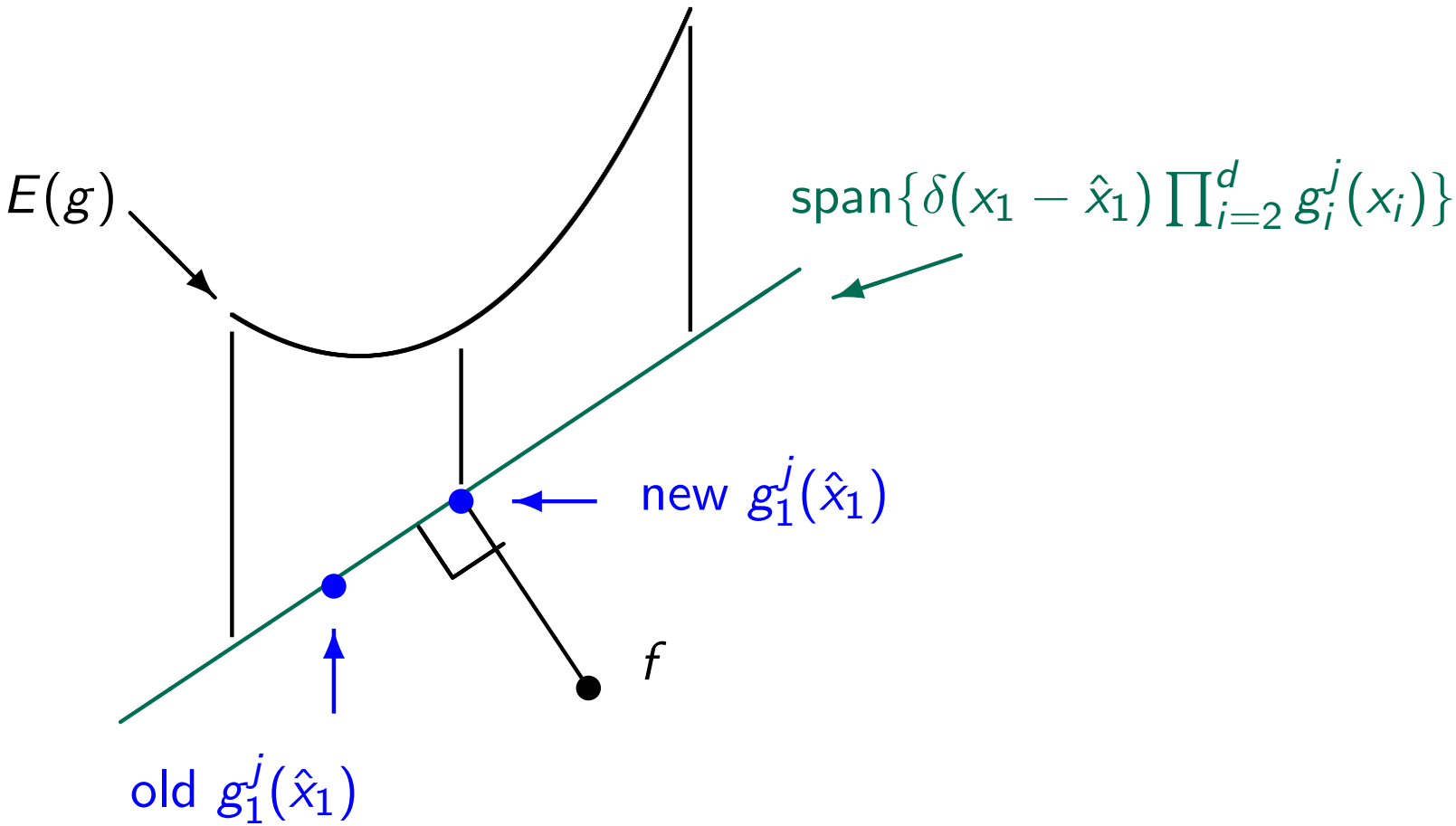
$$\langle \mathrm{span}\{g_j\}, (g - f) \rangle = 0 \quad \Leftrightarrow \quad \sum_j \langle g_k, g_j \rangle c_j = \langle g_k, f \rangle \quad \text{for all } k.$$



Each ALS Step is a Linear Least-Squares Problem

g is a linear combination of fixed elements with unknown coefficients

$$g = \sum_{j=1}^{r} \prod_{i=1}^{d} g_i^j(x_i) = \sum_{j=1}^{r} \int g_1^j(\hat{x}_1)\delta(x_1 - \hat{x}_1) \prod_{i=2}^{d} g_i^j(x_i) d\hat{x}_1 \, .$$



ALS Normal equations via inner products

$$\sum_{j=1}^{r} \int \left\langle \delta(x_1 - x_1') \prod_{i=2}^{d} g_i^k(x_i), \delta(x_1 - \hat{x}_1) \prod_{i=2}^{d} g_i^j(x_i) \right\rangle g_1^j(\hat{x}_1) d\hat{x}_1$$

$$= \sum_{j=1}^{r} \left(\prod_{i=2}^{d} \left\langle g_i^k, g_i^j \right\rangle \right) g_1^j(x_1') = \left\langle \delta(x_1 - x_1') \prod_{i=2}^{d} g_i^k(x_i), f \right\rangle \quad \text{for all } k,\ x_1'.$$

If $f = \sum_{q=1}^{Q} \prod_{i=1}^{d} f_i^q(x_i)$ and we set $\mathbf{A}(k,j) = \prod_{i=2}^{d} \left\langle g_i^k, g_i^j \right\rangle$, then the solution is

$$g_1^j = \sum_{k=1}^{r} \mathbf{A}^{-1}(j,k) \sum_{q=1}^{Q} f_1^q \prod_{i=2}^{d} \left\langle g_i^k, f_i^q \right\rangle .$$

We never needed to discretize in x_1, choose a basis, etc.

First Modifications of the Inner Product

Regularization as a redefined inner product

For $\lambda > 0$ the regularized least-squares error is

$$E_\lambda(g) = E_\lambda(g^1, \ldots, g^r) = \|f - g\|^2 + \lambda \sum_{j=1}^{r} \|g^j\|^2 .$$

Using E_λ keeps the approximation problem well-posed and controls loss-of-precision errors due to cancellations.
Defining an inner product on vectors of functions by

$$\langle [a_1(x_1), a_2(x_2) \ldots], [b_1(x_1), b_2(x_2) \ldots] \rangle = \sum_{k=1} \langle a_k, b_k \rangle ,$$

E_λ is the ordinary least-squares error of approximating

$$[f, 0, \ldots, 0] \approx \left[g, \sqrt{\lambda} g^1, \ldots, \sqrt{\lambda} g^r \right] = \sum_{j=1}^{r} g^j \left(\mathbf{e}_1 + \sqrt{\lambda} \mathbf{e}_{j+1} \right) .$$

We can use the framework we already have.

Antisymmetric Inner Products

Electrons are (observed to be) fermions, which means wavefunctions for the multiparticle Schrödinger equation must be antisymmetric: e.g.

$$f(x_1, x_2, \ldots, x_d) = -f(x_2, x_1, \ldots, x_d).$$

We can project a function onto its antisymmetric part by averaging over all permutations with the proper signs by applying the *antisymmetrizer*

$$\mathcal{A} = \frac{1}{d!} \sum_{p \in S_d} (-1)^p \mathcal{P}_p,$$

where S_d is the permutation group on d elements.
Pose the approximation problem as

$$\mathcal{A}f(x_1, x_2, \ldots, x_d) \approx \mathcal{A} \sum_{j=1}^{r} g^j(x_1, x_2, \ldots, x_d) = \mathcal{A} \sum_{j=1}^{r} \prod_{i=1}^{d} g_i^j(x_i)$$

with error measure

$$E(g) = \langle \mathcal{A}(f-g), \mathcal{A}(f-g) \rangle = \langle \mathcal{A}f, \mathcal{A}f \rangle - 2\langle \mathcal{A}f, \mathcal{A}g \rangle + \langle \mathcal{A}g, \mathcal{A}g \rangle .$$

Computing Antisymmetric Inner Products

The antisymmetric inner product between g^k and g^j can be computed via

$$\left\langle \mathcal{A}\prod_{i=1}^{d} g_i^k, \mathcal{A}\prod_{i=1}^{d} g_i^j \right\rangle = \left\langle \mathcal{A}\prod_{i=1}^{d} g_i^k, \prod_{i=1}^{d} g_i^j \right\rangle =$$

$$\int \cdots \int \frac{1}{d!} \begin{vmatrix} g_1^k(x_1) & \cdots & g_1^k(x_d) \\ \vdots & \ddots & \vdots \\ g_d^k(x_1) & \cdots & g_d^k(x_d) \end{vmatrix} g_1^j(x_1) \cdots g_d^j(x_d) dx_1 \cdots dx_d$$

$$= \frac{1}{d!} \begin{vmatrix} \langle g_1^k, g_1^j \rangle & \cdots & \langle g_1^k, g_d^j \rangle \\ \vdots & \ddots & \vdots \\ \langle g_d^k, g_1^j \rangle & \cdots & \langle g_d^k, g_d^j \rangle \end{vmatrix}.$$

With this inner product, we can again use the framework we already have.

Fitting Data (regression)

Suppose f is only a data set

$$f = \{(\mathbf{x}_n, y_n) = (x_1^n, \cdots, x_d^n; y_n)\}_{n=1}^N ,$$

so our approximation problem becomes

$$y_n \approx \sum_{j=1}^{r} \prod_{i=1}^{d} g_i^j(x_i^n) \quad \text{for all } n.$$

We can define a data-driven (pseudo) inner product

$$\langle f, g \rangle = \sum_{n=1}^{N} f(\mathbf{x}_n) g(\mathbf{x}_n)$$

and try to minimize

$$E(g) = \langle (f-g), (f-g) \rangle = \langle f, f \rangle - 2\langle f, g \rangle + \langle g, g \rangle .$$

ALS with a Data-Driven Inner Product

Our ALS framework had us split

$$g_1^j(x_1) = \int g_1^j(\hat{x}_1)\delta(x_1 - x_1')d\hat{x}_1$$

and integrate

$$\int f(\mathbf{x})\delta(x_1 - x_1')d\hat{x}_1\,,$$

which no longer makes sense, since f is only known at some points.
Instead we have to write

$$g_1^j(x_1) = \sum_{m=1}^{M} c_m^j \phi_m(x_1)$$

in some basis $\{\phi_m\}_{m=1}^M$ with unknown coefficients c_m^j.
We can then run ALS and also get a g that we can evaluate anywhere.

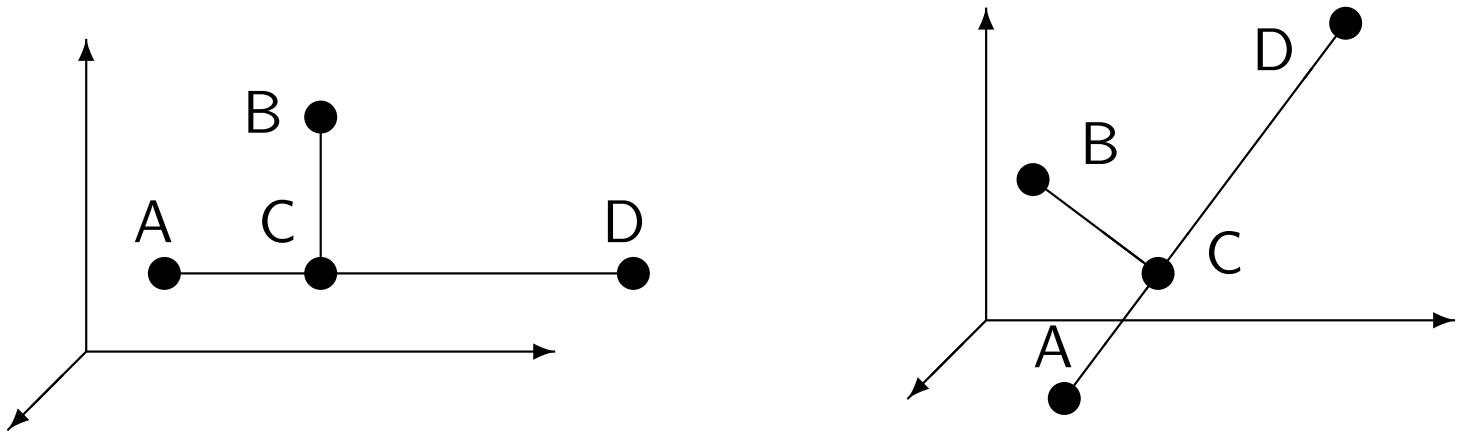
More Exotic Inner Products

Properties of Materials

The data is $f = \{(\sigma_n, \rho_n)\}_{n=1}^{N}$ where

- σ_n is a material/molecular *structure*, which is an unordered set of *atoms* $a = (t, \mathbf{r})$, where t is a species type (e.g. $t = \mathrm{Mo}$) and $\mathbf{r}$ is a location in 3-dimensional space and
- $\rho_n = \rho(\sigma_n)$ is some useful physical property of σ_n.

Two structures are *equivalent* if one can be mapped to the other by a translation and/or rotation.



We assume ρ is *consistent*, giving the same value to equivalent structures, and require our approximation g to be consistent.

Consistent Functions of Structures

For an ordered list of atoms, we can construct a function

$$g([a_1, a_2, \ldots]) := g([a_1, a_2, \ldots, a_d]) = \sum_{l=1}^{r} \prod_{i=1}^{d} g_i^l(a_i)\,.$$

To make g consistent, we map a structure to a set V_σ whose elements (w, v) are a weight w and an ordered list of atoms v called a *view*, and work with

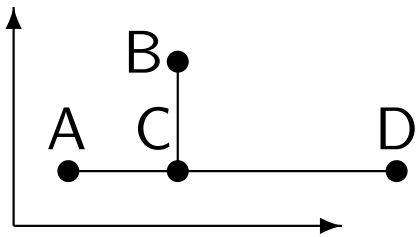
$$\mathcal{C}g(\sigma) = \sum_{(w,v)\in V_\sigma} wg(v)\,.$$

We then define a pseudo inner product by

$$\langle f, g\rangle = \frac{1}{N}\sum_{j=1}^{N} \mathcal{C}f(\sigma_j)\mathcal{C}g(\sigma_j)$$

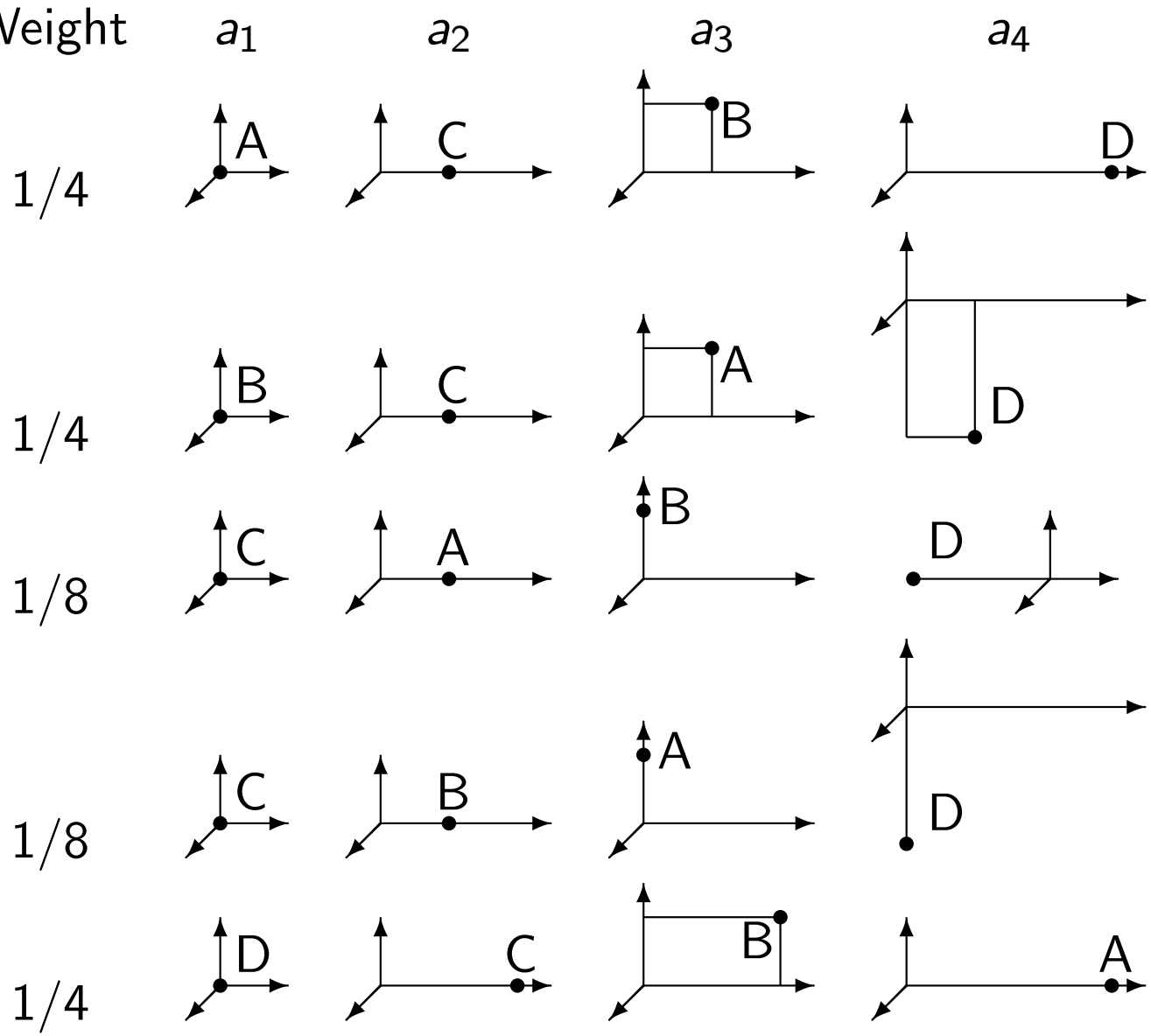
and use the framework we already have.

Conversion of a structure to its views (illustration)



maps to

the views:



Including Potentials in Antisymmetric Inner Products

As a step in an iterative algorithm to construct an approximate wavefunction, we need to compute antisymmetric inner products including the nuclear potential operator

$$\left\langle \mathcal{A}g^k, \mathcal{A}\mathcal{V}f^q \right\rangle = \left\langle \mathcal{A}\prod_{i=1}^{d} g_i^k(x_i), \mathcal{A}\left(\sum_{m=1}^{d} V(x_m)\right)\prod_{i=1}^{d} f_i^q(x_i) \right\rangle$$

and including the electron-electron potential operator

$$\left\langle \mathcal{A}g^k, \mathcal{A}\mathcal{W}f^q \right\rangle = \left\langle \mathcal{A}\prod_{i=1}^{d} g_i^k(x_i), \mathcal{A}\left(\frac{1}{2}\sum_{m=1}^{d}\sum_{n\neq m} \frac{1}{\|x_m - x_n\|}\right)\prod_{i=1}^{d} f_i^q(x_i) \right\rangle .$$

These operators interfere with the antisymmetric inner product, but only in a few variables at a time.

Prototype Computation Involving $\mathcal{W}$

Suppressing the k and q indexes and selecting the $m = 1$, $n = 2$ term, we need to compute

$$\left\langle \mathcal{A} \prod_{i=1}^{d} g_i(x_i), \frac{1}{\|x_1 - x_2\|} \prod_{i=1}^{d} f_i(x_i) \right\rangle =$$

$$\int \cdots \int \frac{1}{\|x_1 - x_2\|} \frac{1}{d!} \begin{vmatrix} g_1(x_1) & \cdots & g_1(x_d) \\ \vdots & \ddots & \vdots \\ g_d(x_1) & \cdots & g_d(x_d) \end{vmatrix} f_1(x_1) \cdots f_d(x_d) dx_1 \cdots dx_d$$

$$= \frac{1}{d!} \int \int \frac{f_1(x_1) f_2(x_2)}{\|x_1 - x_2\|} \begin{vmatrix} g_1(x_1) & g_1(x_2) & \langle g_1, f_3 \rangle & \cdots & \langle g_1, f_d \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_d(x_1) & g_d(x_2) & \langle g_d, f_3 \rangle & \cdots & \langle g_d, f_d \rangle \end{vmatrix} dx_1 dx_2 \, .$$

Prototype Computation Involving $\mathcal{W}$ (continued)

Defining $\mathbf{L}(i,j) = \langle g_i, f_j \rangle$ and $\begin{bmatrix} \tilde{g}_1 \\ \vdots \\ \tilde{g}_d \end{bmatrix} = \mathbf{L}^{-1} \begin{bmatrix} g_1 \\ \vdots \\ g_d \end{bmatrix}$ we can multiply by $|\mathbf{L}||\mathbf{L}^{-1}|$, merge determinants, multiply matrixes, and get

$$\frac{|\mathbf{L}|}{d!} \int\int \frac{f_1(x_1) f_2(x_2)}{\|x_1 - x_2\|} \begin{vmatrix} \tilde{g}_1(x_1) & \tilde{g}_1(x_2) & 0 & \cdots & 0 \\ \tilde{g}_2(x_1) & \tilde{g}_2(x_2) & 0 & \cdots & 0 \\ \tilde{g}_3(x_1) & \tilde{g}_3(x_2) & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{g}_d(x_1) & \tilde{g}_d(x_2) & 0 & \cdots & 1 \end{vmatrix} dx_1 dx_2$$

$$= \frac{|\mathbf{L}|}{d!} \int\int \frac{f_1(x_1) f_2(x_2)}{\|x_1 - x_2\|} \begin{vmatrix} \tilde{g}_1(x_1) & \tilde{g}_1(x_2) \\ \tilde{g}_2(x_1) & \tilde{g}_2(x_2) \end{vmatrix} dx_1 dx_2 \,.$$

Antisymmetric Inner Product Involving $\mathcal{V}$ or $\mathcal{W}$

Suppressing k and q, we have

$$\langle \mathcal{A}g, \mathcal{A}\mathcal{V}f \rangle = \frac{|\mathbf{L}|}{d!} \sum_{m=1}^{d} \int V(x) f_m(x) \tilde{g}_m(x) dx \quad \text{and}$$

$$\langle \mathcal{A}g, \mathcal{A}\mathcal{W}f \rangle = \frac{1}{2} \frac{|\mathbf{L}|}{d!} \sum_{m=1}^{d} \sum_{n \neq m} \int \int \frac{f_m(x) f_n(y)}{\|x - y\|} \begin{vmatrix} \tilde{g}_m(x) & \tilde{g}_m(y) \\ \tilde{g}_n(x) & \tilde{g}_n(y) \end{vmatrix} dxdy \, .$$

With these inner product formulas (and a little more work), we can again use the fitting and ALS framework we already have.

It gets worse: Including Geminals

The wavefunction has an 'inter-electron cusp' whenever two variables coincide. We will need to include such structure in our approximation, with something like

$$g(\mathbf{x}) = \mathcal{A} \sum_{p=0}^{P} \left(\frac{1}{2} \sum_{m \neq n} w_p(\|x_m - x_n\|) \right) \sum_{j=1}^{r} \prod_{i=1}^{d} g_i^{p,j}(x_i) \, .$$

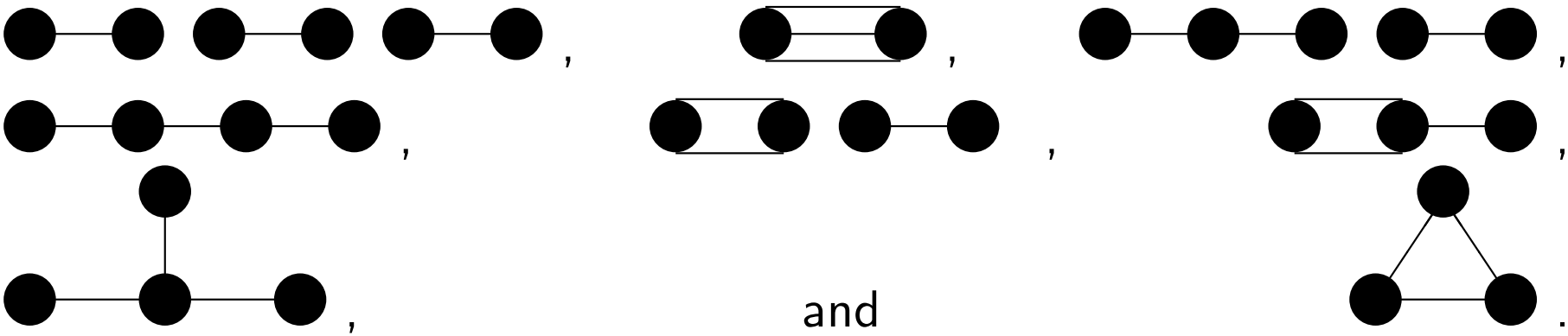
With such structure in g and f and $\mathcal{W}$, we have to compute antisymmetric inner products like

$$\sum_{m \neq n} \sum_{u \neq v} \sum_{a \neq b} \left\langle \mathcal{A} w_2(\|x_m - x_n\|) \prod_{i=1}^{d} g_i(x_i), \frac{1}{\|x_u - x_v\|} w_1(\|x_a - x_b\|) \prod_{i=1}^{d} f_i(x_i) \right\rangle$$

The method we used for $\mathcal{W}$ lets us integrate out all variables except x_m, x_n, x_u, x_v, x_a, and x_b.

It gets even worse: Many Geminal Cases

The indexes are restricted by $a \neq b$, $m \neq n$, and $u \neq v$, but we can still have overlaps such as $a = m$. Between 2 and 6 variables remain.
There are 8 distinct cases, which can be represented graphically as



Each case corresponds to a formula involving a determinant of function of size up to 6×6. To compute we expand all determinants and end up with several hundred terms to compute. (We have to automate.)

We can then use our existing framework for ALS.

It gets worse: Including Recursive Approximations

To handle extended quantum-mechanical systems without exponential scaling, we will need another form, such as the recursive form

$$g(\mathbf{x}) = \mathcal{A} \sum_{j=1}^{r} \prod_{k=1}^{K} \left(\sum_{j_k=1}^{r_k} \prod_{i_k=1}^{d_k} g_{k,i_k}^{j,j_k}(x_{k,i_k}) \right) .$$

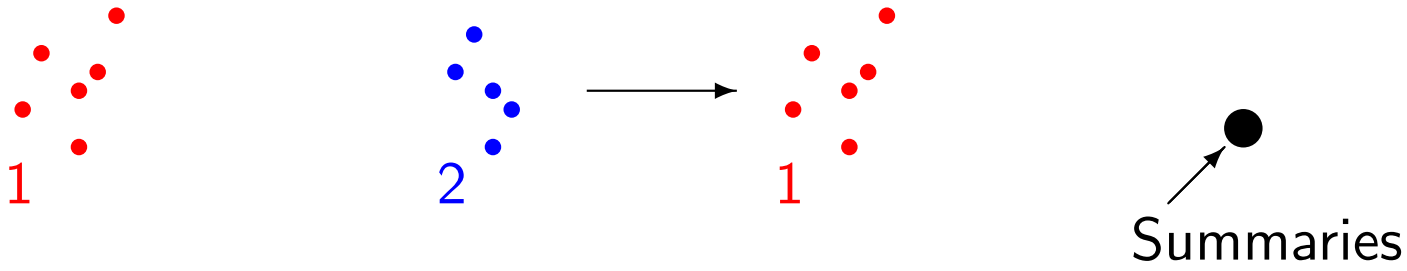
We need to compute antisymmetric inner products like

$$\left\langle \mathcal{A} \prod_{k=1}^{K} \left(\sum_{j_k=1}^{r_k} \prod_{i_k=1}^{d_k} g_{k,i_k}^{j_k}(x_{k,i_k}) \right) , \prod_{k=1}^{K} \left(\sum_{q_k=1}^{r_k} \prod_{i_k=1}^{d_k} f_{k,i_k}^{q_k}(x_{k,i_k}) \right) \right\rangle$$

without expanding out.

If successful, we can then use our existing framework for ALS.

The Classical Center-of-Mass Principle



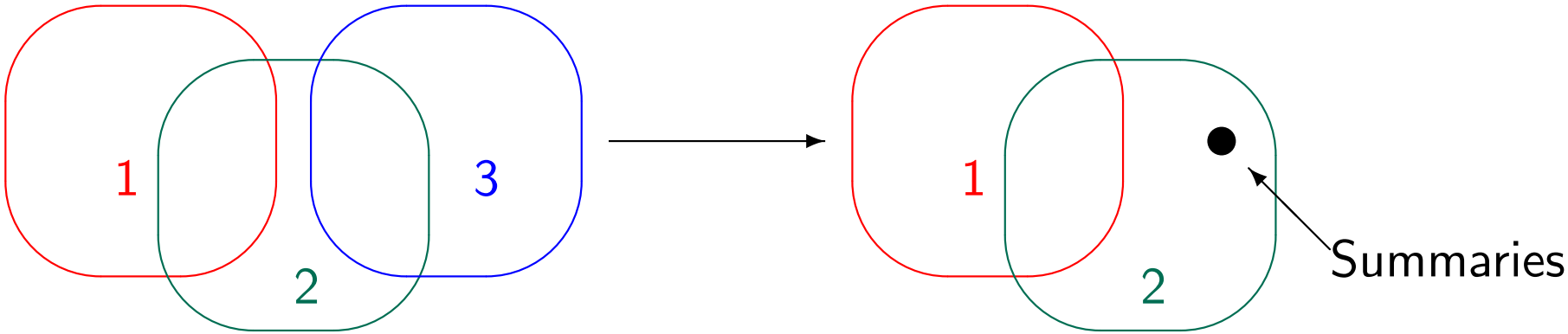
Approximate the gravitational potential energy between the two groups by

$$\sum_{i_1=1}^{M_1}\sum_{i_2=1}^{M_2}\frac{m_1^{i_1}\,m_2^{i_2}}{\|\mathbf{r}_1^{i_1}-\mathbf{r}_2^{i_2}\|}\approx\sum_{i_1=1}^{M_1}\sum_{i_2=1}^{M_2}\frac{m_1^{i_1}\,m_2^{i_2}}{\|\mathbf{r}_1^{i_1}-\mathbf{r}_2\|}$$

$$=\sum_{i_1=1}^{M_1}\frac{m_1^{i_1}}{\|\mathbf{r}_1^{i_1}-\mathbf{r}_2\|}\left(\sum_{i_2=1}^{M_2}m_2^{i_2}\right)=\sum_{i_1=1}^{M_1}\frac{m_1^{i_1}}{\|\mathbf{r}_1^{i_1}-\mathbf{r}_2\|}S_2\,,$$

where $\mathbf{r}_2$ is the center of mass of group two.
This reduces the cost from $\mathcal{O}(M_1M_2)$ to $\mathcal{O}(M_1+M_2)$.

A Quantum Center-of-Mass Principle, schematic



Even when groups 1 and 3 do not directly interact, the presence of an intermediate group couples them.

It will take a partial expansion of the determinant to decouple them.

Group 3 can then be summarized and the summary embedded in group 2.

A Quantum Center-of-Mass Principle, formula sketch

$$\sum_{j_1=1}^{r_1}\sum_{\tilde{j}_1=1}^{r_1}\sum_{j_2=1}^{r_2}\sum_{\tilde{j}_2=1}^{r_2}\sum_{j_3=1}^{r_3}\sum_{\tilde{j}_3=1}^{r_3} a(j_1,\tilde{j}_1;j_2,\tilde{j}_2;j_3,\tilde{j}_3)$$

$$\approx \sum_{j_1=1}^{r_1}\sum_{\tilde{j}_1=1}^{r_1}\sum_{j_2=1}^{r_2}\sum_{\tilde{j}_2=1}^{r_2}\sum_{j_3=1}^{r_3}\sum_{\tilde{j}_3=1}^{r_3}\left(\sum_{\alpha} b(j_1,\tilde{j}_1;j_2,\tilde{j}_2;\alpha)c(\alpha,j_2,\tilde{j}_2;j_3,\tilde{j}_3)\right)$$

$$= \sum_{j_1=1}^{r_1}\sum_{\tilde{j}_1=1}^{r_1}\sum_{j_2=1}^{r_2}\sum_{\tilde{j}_2=1}^{r_2}\sum_{\alpha} b(j_1,\tilde{j}_1;j_2,\tilde{j}_2;\alpha)\left(\sum_{j_3=1}^{r_3}\sum_{\tilde{j}_3=1}^{r_3} c(\alpha,j_2,\tilde{j}_2;j_3,\tilde{j}_3)\right)$$

$$= \sum_{j_1=1}^{r_1}\sum_{\tilde{j}_1=1}^{r_1}\sum_{j_2=1}^{r_2}\sum_{\tilde{j}_2=1}^{r_2}\sum_{\alpha} b(j_1,\tilde{j}_1;j_2,\tilde{j}_2;\alpha)S(\alpha,j_2,\tilde{j}_2)\,.$$

This reduces the cost from $\mathcal{O}(r_1^2 r_2^2 r_3^2)$ to $\mathcal{O}(r_1^2 r_2^2 + r_2^2 r_3^2)$.
For a chain of K groups, $\mathcal{O}(r^{2K})$ reduces to $\mathcal{O}(r^4 K)$.

A Quantum Center-of-Mass Principle, core determinant

Using the non-overlap of groups 1 and 3, the determinant is

$$a(j_1, \tilde{j}_1; j_2, \tilde{j}_2; j_3, \tilde{j}_3) =$$

$$\begin{vmatrix} \mathbb{L}_{11} & \mathbb{L}_{12} & 0 \\ \mathbb{L}_{21} & \mathbb{L}_{22} & \mathbb{L}_{23} \\ 0 & \mathbb{L}_{32} & \mathbb{L}_{33} \end{vmatrix} = \left| \begin{bmatrix} \mathbb{L}_{11} & \mathbb{L}_{12} & 0 \\ \mathbb{L}_{21} & \mathbb{L}_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathbb{L}_{23} \\ 0 & \mathbb{L}_{32} & \mathbb{L}_{33} \end{bmatrix} \right|$$

$$= |\mathbb{A}_{12} + \mathbb{B}_{23}| \ ,$$

where the subscripts indicate which summation indices the blocks depend upon.

To decouple groups 1 and 3 we need to expand out the determinant of a sum of two matrices.

Determinant of the Sum of Two Matrices

For $N \times N$ matrices $\mathbb{A}$ and $\mathbb{B}$,

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

where $\alpha_0 = \{1, 2, \ldots, N\}$, α and β are ordered subsets, $\sigma(\alpha)$ is the sum of the entries in α, $\alpha_0 \setminus \alpha$ is the complement of α in α_0, and $\mathbb{B}[\alpha;\beta]$ denotes the matrix formed using rows α and columns β from $\mathbb{B}$.

The expansions are unpleasant. The zeros in our matrices help.

The off-diagonal blocks $\mathbb{L}_{12}$, $\mathbb{L}_{21}$, $\mathbb{L}_{23}$, and $\mathbb{L}_{32}$ are expected to be low rank, which shortens the sums needed.

Conclusions and Comments

- You can do a lot if you can find the right inner product.
- Exotic inner products are harder (impossible?) for other function/tensor approximations.
- Since $\langle (f-g),(f-g)\rangle = \langle f,f\rangle - 2\langle f,g\rangle + \langle g,g\rangle$, we can minimize over g even if we cannot compute $\langle f,f\rangle$. This situation arises e.g. when f includes an operator like $\mathcal{W}$.
- ALS convergence is an issue.

References

See http://www.ohio.edu/people/mohlenka/ for

Numerical Operator Calculus in Higher Dimensions.

Algorithms for Numerical Analysis in High Dimensions.

Musings on Multilinear Fitting.

Multivariate Regression and Machine Learning with Sums of Separable Functions.

Learning to Predict Physical Properties using Sums of Separable Functions.

Approximating a Wavefunction as an Unconstrained Sum of Slater Determinants.

A Center-of-Mass Principle for the Multiparticle Schrodinger Equation.

Capturing the Inter-electron Cusp using a Geminal Layer on an Unconstrained Sum of Slater Determinants.

Portions were supported by DARPA and NSF.