Add, Multiple, Divide...and Conquer: Factorization Algorithms for Large-Scale Many-Body Problems

> Collaborators: W. Erich Ormand, Lawrence Livermore Plamen G. Krastev, SDSU Hai Ah Nam, SDSU/ Oak Ridge Collaborators-in-training: Joshua Staker & Micah Schuster, SDSU

THE KEY IDEAS

Basic problem: find extremal eigenvalues of very large, very sparse Hermitian matrix

Lanczos algorithm

fundamental operation is *matrix-vector multiply*

Despite sparsity, nonzero matrix elements can require TB of storage

Only a fraction of matrix elements are unique; **most are reused.** Reuse of matrix elements understood through *spectator* particles.

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

Find extremal eigenvalues of very large, very sparse Hermitian matrix Lanczos algorithm fundamental operation is matrix-vector multiply

The basic *science question* is to model detailed quantum structure of many-body systems, such the electronic structure of an atom, or structure of an atomic nucleus.

Find extremal eigenvalues of very large, very sparse Hermitian matrix Lanczos algorithm fundamental operation is matrix-vector multiply

To answer this, we attempt to solve *Schrödinger's equation*:

$$\left(\sum_{i} -\frac{\hbar^2}{2m} \nabla^2 + U(r_i) + \sum_{i < j} V(\vec{r}_i - \vec{r}_j)\right) \Psi(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) = E \Psi$$

or

$$\hat{\mathbf{H}} |\Psi\rangle = E |\Psi\rangle$$

Find extremal eigenvalues of very large, very sparse Hermitian matrix Lanczos algorithm fundamental operation is matrix-vector multiply

This differential equation is too difficult to solve directly

$$\begin{split} \left(\sum_{i} -\frac{\hbar^{2}}{2m} \nabla^{2} + U(r_{i}) + \sum_{i < j} V(\vec{r}_{i} - \vec{r}_{j}) \right) \Psi(\vec{r}_{1}, \vec{r}_{2}, \vec{r}_{3} \dots) &= E \Psi \\ \text{so we use the matrix formalism} \\ \hat{\mathbf{H}} \left| \Psi \right\rangle &= E \left| \Psi \right\rangle \end{split}$$

Find extremal eigenvalues of very large, very sparse Hermitian matrix Lanczos algorithm fundamental operation is matrix-vector multiply

$$\begin{split} |\Psi\rangle &= \sum_{\alpha} c_{\alpha} |\alpha\rangle \qquad H_{\alpha\beta} = \langle \alpha | \hat{\mathbf{H}} |\beta\rangle \\ \sum_{\beta} H_{\alpha\beta} c_{\beta} &= Ec_{\alpha} \quad \text{if} \quad \langle \alpha |\beta\rangle = \delta_{\alpha\beta} \end{split}$$

so we use the matrix formalism

$$\hat{\mathbf{H}}|\Psi\rangle = E|\Psi\rangle$$

Find extremal eigenvalues of very large, very sparse Hermitian matrix Lanczos algorithm

fundamental operation is *matrix-vector multiply*

$$H_{\alpha\beta} = \left\langle \alpha | \hat{\mathbf{H}} | \beta \right\rangle$$

* **H** is generally a very large matrix – dimensions up to 10^{10} have been tackled.

- * **H** is generally very sparse.
- * We usually only want a few low-lying states



Lanczos algorithm!

Find extremal eigenvalues of very large, very sparse Hermitian matrix Lanczos algorithm fundamental operation is matrix-vector multiply

Standard algorithm to obtain *all* eigenvalues of a real, symmetric matrix **A**: <u>Householder</u>

Find orthogonal matrix U such that $U^T A U = B$, a tridiagonal matrix

The Lanczos algorithm is similar, in that it also uses an orthogonal matrix to take **A** to a tridiagonal matrix **B**.....

Lanczos algorithm!

Find extremal eigenvalues of very large, very sparse Hermitian matrix Lanczos algorithm fundamental operation is matrix-vector multiply

$$\mathbf{A}\vec{v}_{1} = \alpha_{1}\vec{v}_{1} + \beta_{1}\vec{v}_{2}$$

$$\mathbf{A}\vec{v}_{2} = \beta_{1}\vec{v}_{1} + \alpha_{2}\vec{v}_{2} + \beta_{2}\vec{v}_{3}$$

$$\mathbf{A}\vec{v}_{3} = \beta_{2}\vec{v}_{2} + \alpha_{3}\vec{v}_{3} + \beta_{3}\vec{v}_{4}$$

$$\mathbf{A}\vec{v}_{4} = \beta_{3}\vec{v}_{3} + \alpha_{4}\vec{v}_{4} + \beta_{4}\vec{v}_{5}$$

Lanczos algorithm!

Find extremal eigenvalues of very large, very sparse Hermitian matrix Lanczos algorithm fundamental operation is matrix-vector multiply

$$\mathbf{A}\vec{v}_{1} = \alpha_{1}\vec{v}_{1} + \beta_{1}\vec{v}_{2}$$

$$\mathbf{A}\vec{v}_{2} = \beta_{1}\vec{v}_{1} + \alpha_{2}\vec{v}_{2} + \beta_{2}\vec{v}_{3}$$

$$\mathbf{A}\vec{v}_{3} = \beta_{2}\vec{v}_{2} + \alpha_{3}\vec{v}_{3} + \beta_{3}\vec{v}_{4}$$

$$\mathbf{A}\vec{v}_{4} = \beta_{3}\vec{v}_{3} + \alpha_{4}\vec{v}_{4} + \beta_{4}\vec{v}_{5}$$

matrix-vector multiply

Lanczos algorithm!

Despite sparsity, nonzero matrix elements can require TB of storage

I need to quickly cover:

- How the basis states are represented
- How the Hamiltonian operator is represented
- •Why most matrix elements are zero
- Typical dimensions and sparsity

Despite sparsity, nonzero matrix elements can require TB of storage

• How the basis states are represented

This differential equation is too difficult to solve directly

$$\left(\sum_{i} -\frac{\hbar^2}{2m} \nabla^2 + U(r_i) + \sum_{i < j} V(\vec{r}_i - \vec{r}_j)\right) \Psi(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) = E \Psi$$

Can only really solve 1D differential equation

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + U(r)\right)\phi_i(r) = \varepsilon_i\phi_i(r)$$

Despite sparsity, nonzero matrix elements can require TB of storage

• How the basis states are represented

Can only really solve 1D differential equation

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dr^2}+U(r)\right)\phi_i(r)=\varepsilon_i\phi_i(r) \quad \Longrightarrow \quad \left\{\phi_i(\vec{r})\right\}$$

Single-particle wave functions labeled by, *e.g.*, *n*, *j*, *l*, *m*

Atomic case: 1s, 2s, 2p, 3s, 3p, 3d etc

Nuclear: $0s_{1/2}$, $0p_{3/2}$, $0p_{1/2}$, $0d_{5/2}$, $1s_{1/2}$, $0d_{3/2}$, *etc*

Despite sparsity, nonzero matrix elements can require TB of storage

• How the basis states are represented

Can only really solve 1D differential equation

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dr^2}+U(r)\right)\phi_i(r)=\varepsilon_i\phi_i(r) \quad \Longrightarrow \quad \left\{\phi_i(\vec{r})\right\}$$

Product wavefunction ("Slater Determinant") $\Psi(\vec{r}_1, \vec{r}_2, \vec{r}_3...) = \phi_{n_1}(\vec{r}_1)\phi_{n_2}(\vec{r}_2)\phi_{n_3}(\vec{r}_3)...\phi_{n_N}(\vec{r}_N)$

Despite sparsity, nonzero matrix elements can require TB of storage

• How the basis states are represented

Product wavefunction ("Slater Determinant")

$$\Psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, ...) = \phi_{(n)}(\vec{r}_1)\phi_{(n)}(\vec{r}_2)\phi_{(n)}(\vec{r}_3)...\phi_{(n)}(\vec{r}_N)$$

Each many-body state can be *uniquely* determined by a list of "occupied" single-particle states = "occupation representation"

$$|\alpha\rangle = \hat{a}_{n_1}^+ \hat{a}_{n_2}^+ \hat{a}_{n_3}^+ \dots \hat{a}_{n_N}^+ |0\rangle$$

Despite sparsity, nonzero matrix elements can require TB of storage

• How the Hamiltonian is represented "occupation representation" $|\alpha\rangle =$

$$|\alpha\rangle = \hat{a}_{n_1}^{\dagger} \hat{a}_{n_2}^{\dagger} \hat{a}_{n_3}^{\dagger} \dots \hat{a}_{n_N}^{\dagger} |0\rangle$$

"creation operator"

$$\hat{H} = \sum_{ij} T_{ij} \hat{a}_{i}^{\dagger} \hat{a}_{j} + \frac{1}{4} \sum_{ijkl} V_{ijkl} \hat{a}_{i}^{\dagger} \hat{a}_{j}^{\dagger} \hat{a}_{l} \hat{a}_{k}$$

motion of a single particle ("one-body operator") interaction of two particles ("two-body operator")

$$V_{ijkl} = \iint \phi_i(\vec{r})\phi_j(\vec{r}')V(\vec{r},\vec{r}')\phi_k(\vec{r})\phi_l(\vec{r}')d^3rd^3r'$$

Despite sparsity, nonzero matrix elements can require TB of storage

• How the Hamiltonian is represented

"occupation representation"

$$|\alpha\rangle = \hat{a}_{n_1}^+ \hat{a}_{n_2}^+ \hat{a}_{n_3}^+ \dots \hat{a}_{n_N}^+ |0\rangle$$

n _i	1	2	3	4	5	6	7
α=1	1	0	0	1	1	0	1
α=2	1	0	1	0	0	1	1
α=3	0	1	1	1	0	1	0

$$\hat{H} = \sum_{ij} T_{ij} \hat{a}_{i}^{\dagger} \hat{a}_{j} + \frac{1}{4} \sum_{ijkl} V_{ijkl} \hat{a}_{i}^{\dagger} \hat{a}_{j}^{\dagger} \hat{a}_{l} \hat{a}_{k}$$

Despite sparsity, nonzero matrix elements can require TB of storage

• How the Hamiltonian is represented



$$\hat{a}_3^+ \hat{a}_6^+ \hat{a}_4 \hat{a}_5 |\alpha = 1\rangle = |\alpha = 2\rangle$$

Despite sparsity, nonzero matrix elements can require TB of storage

• How the Hamiltonian is represented

"occupation representation" $|\alpha|$

lpha angle	$=\hat{a}_{n_1}^+\hat{a}_{n_2}^+\hat{a}_{n_2}$	$\hat{a}_{n_3}^+ \dots \hat{a}_{n_N}^+ 0\rangle$

	n _i	1	2	3	4	5	6	7
	α=1	1	0	0	1	1	0	1
	α=2	1	0	1	0	0	1	-1
•	α=3	0	1	1	1	0	1	0

$$\hat{a}_{2}^{+}\hat{a}_{4}^{+}\hat{a}_{1}\hat{a}_{7}|\alpha=2\rangle = |\alpha=3\rangle$$

Despite sparsity, nonzero matrix elements can require TB of storage

• Why most matrix elements are zero $|\alpha\rangle = \hat{a}_{n_1}^+ \hat{a}_{n_2}^+ \hat{a}_{n_3}^+ \dots \hat{a}_{n_N}^+ |0\rangle$ "occupation representation" n_i **α=1 α=2** $\alpha = 3$

 $\hat{a}_{2}^{+}\hat{a}_{4}^{+}\hat{a}_{6}^{+}\hat{a}_{1}\hat{a}_{5}\hat{a}_{7}|\alpha=1\rangle = |\alpha=3\rangle$ need 3 particles to interact simultaneously!

Despite sparsity, nonzero matrix elements can require TB of storage

•Typical dimensions and sparsity

Nuclide	valence space	valence Z	valence N	basis dim	sparsity (%)
20 Ne	"sd"	2	2	640	10
^{25}Mg	"sd"	4	5	44,133	0.5
⁴⁹ Cr	"pf"	4	5	6M	0.01
⁵⁶ Fe	"pf"	6	10	500M	2x10-4

This corresponds to 2 Tb of data!

RECYCLED MATRIX ELEMENTS

Only a fraction of matrix elements are unique; **most are reused.** Reuse of matrix elements understood through *spectator* particles.



RECYCLED MATRIX ELEMENTS

Only a fraction of matrix elements are unique; **most are reused.** Reuse of matrix elements understood through *spectator* particles.

of nonzero matrix elements vs. # unique matrix elements

Nuclide	valence space	valence Z	valence N	# nonzero	# unique
$^{28}\mathrm{Si}$	"sd"	6	6	$26 \ge 10^6$	3600
⁵² Fe	"pf"	6	6	90 x 10 ⁹	21,500

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

A quantum number is the eigenvalue of an operator

generally a operator that exactly commutes with the Hamiltonian

e.g. angular momentum \mathbf{J}^2 and z-component \mathbf{J}_z

$$\hat{J}^2 |\Psi\rangle = J(J+1) |\Psi\rangle \quad \hat{J}_z |\Psi\rangle = M |\Psi\rangle$$

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

A quantum number is the eigenvalue of an operator

For composite systems, one can apply the operator to each component separately:

$$\hat{O}|\Psi\rangle = \left(\hat{O}_1 + \hat{O}_2 + \hat{O}_3 + \ldots\right) \left(|\Psi_1\rangle \otimes |\Psi_2\rangle \otimes |\Psi_3\rangle \otimes \ldots\right)$$

Sometimes the total quantum number is a simple sum/product as is the case for \mathbf{J}_z or parity....

$$\hat{J}_{z}|\Psi\rangle = M|\Psi\rangle = (m_{1} + m_{2} + m_{3} + \ldots)|\Psi\rangle$$

...but in other cases the addition is complicated (e.g. for J^2)

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

I consider composite many-fermion systems, in particular those with 2 major components protons and neutrons

or

spin-up and *spin-down* electrons

$$\left|\Psi\right\rangle = \left|\Psi_{1}\right\rangle \otimes \left|\Psi_{2}\right\rangle$$

Each component itself is a Slater determinant which is composed of many particles

$$\hat{J}_{z} |\Psi\rangle = M |\Psi\rangle \qquad M = M_{1} + M_{2}$$
$$M_{1} = m_{1}^{(1)} + m_{1}^{(2)} + m_{1}^{(2)} + \dots$$

Reuse can be exploited using exact factorization enforced through additive/multiplicative quantum numbers

Because the M values are discrete integers or half-integers $(-3, -2, -1, 0, 1, 2, \dots \text{ or } -3/2, -1/2, +1/2, +3/2...)$ we can organize the basis states in discrete sectors

Example: 2 protons, 4 neutrons, total M = 0

$$M_{z}(\pi) = -4$$

$$M_{z}(\upsilon) = +4$$

$$M_{z}(\pi) = -3$$

$$M_{z}(\upsilon) = +3$$

$$M_{z}(\tau) = -2$$

$$M_{z}(\upsilon) = +2$$

CSRC SEMINAR -- FACTORIZATION Algorithms -- Sept 24, 2010

IVI-(II)

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

In fact, we can see an example of factorization here because all proton Slater determinants in one M-sector *must* combine with all the conjugate neutron Slater determinants

Example: 2 protons, 4 neutrons, total M = 0

$$M_z(\pi) = -4: 2 \text{ SDs}$$
 $M_z(\upsilon) = +4: 24 \text{ SDs}$ 48 combined $M_z(\pi) = -3: 4 \text{ SDs}$ $M_z(\upsilon) = +3: 39 \text{ SDs}$ 156 combined $M_z(\pi) = -2: 9 \text{ SDs}$ $M_z(\upsilon) = +2: 60 \text{ SDs}$ 540 combined

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

In fact, we can see an example of factorization here because all proton Slater determinants in one M-sector *must* combine with all the conjugate neutron Slater determinants



29

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*



Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

Factorization allows us to keep track of all basis states without writing out every one explicitly -- we only need to write down the proton/neutron components

The same trick can be applied to matrix-vector multiply



Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*



There are potentially 48×48 matrix elements But for H_{pp} at most 4×24 are nonzero and we only have to look up 4 matrix elements

Advantage: we can store 98 matrix elements as 4 matrix elements and avoid 2000+ zero matrix elements.

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

 $M_{z}(\pi) = -4: 2 \text{ SDs} \qquad M_{z}(\upsilon) = +4: 24 \text{ SDs} \qquad 48 \text{ combined}$ $\begin{vmatrix} v_{1} \rangle \\ |v_{2} \rangle \\ |\pi_{2} \rangle \qquad H_{pp} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \qquad \begin{vmatrix} v_{1} \rangle \\ |v_{2} \rangle \\ |v_{3} \rangle \\ |v_{4} \rangle \\ \vdots \\ |v_{24} \rangle$

Advantage: we can store 98 matrix elements as 4 matrix elements and avoid 2000+ zero matrix elements.

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

Mz	(π) = -4: 2 SDs	$M_z(v) =$	+4: 24 SDs	48 combined	
$egin{array}{c} \pi_1 angle \ \pi_2 angle \end{array}$	$H_{pp} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$	$egin{aligned} egin{aligned} egi$	$egin{aligned} &H_{pp} \pi_1 angle u_1 angle = \ &H_{pp} \pi_2 angle u_1 angle = \ &H_{pp} \pi_1 angle u_2 angle = \ &H_{pp} \pi_2 angl$	$H_{11} \pi_{1}\rangle \nu_{1}\rangle + H_{12}$ $= H_{12} \pi_{1}\rangle \nu_{1}\rangle + H_{21}$ $= H_{11} \pi_{1}\rangle \nu_{2}\rangle + H_{11}$ $= H_{12} \pi_{1}\rangle \nu_{2}\rangle + H_{12}$ $= H_{12} \pi_{1}\rangle \nu_{24}\rangle + H_{22}$	$egin{aligned} & \pi_2 angle & u_1 angle \ &_2 & \pi_2 angle & u_1 angle \ &_2 & u_2 angle \ &_{22} & u_2 angle \ &_{22} & u_2 angle \ &_{22} & u_2 angle \ &_{24} angle \ &_{22} & u_{22} & u_{24} angle \ &_{22} & u_{24} angle \ &_{24} & u$

Advantage: we can store 98 matrix elements as 4 matrix elements and avoid 2000+ zero matrix elements.

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*

Comparison of nonzero matrix storage with factorization

Nuclide	Space	Basis dim	matrix store	factorization
⁵⁶ Fe	pf	501 M	290 Gb	0.72 Gb
⁷ Li	N _{max} =12	252 M	3600 Gb	96 Gb
⁷ Li	N _{max} =14	1200 M	23 Tb	624 Gb
¹² C	N _{max} =6	32M	196 Gb	3.3 Gb
¹² C	N _{max} =8	590M	5000 Gb	65 Gb
¹² C	N _{max} =10	7800M	111 Tb	1.4 Tb
¹⁶ O	N _{max} =6	26 M	142 Gb	3.0 Gb
¹⁶ O	N _{max} =8	990 M	9700 Gb	130 Gb

PARALLEL IMPLEMENTATION

Factorization makes it easier to compute workload and distribute across multiple nodes



PARALLEL IMPLEMENTATION

Factorization makes it easier to compute workload and distribute across multiple nodes



ALGORITHMS -- SEPT 24, 2010

THE BIGSTICK CODE

Many-fermion code: 2nd generation after REDSTICK code (started in *Baton Rouge, La*.)

Arbitrary single-particle radial waveforms Allows local or nonlocal two-body interaction Applies to both nuclear and atomic cases

Runs on both desktop and parallel machines --can run at least dimension 100M+ on desktop (20 Lanczos iterations in 300 CPU minutes)

20-30k lines of codes Fortran 90 + MPI + OpenMP Partially funded by SciDAC Plans to run on 50,000-100,000 compute nodes Plans to publish code late 2011

SCIENCE APPLICATIONS

Comparison of "exact" results versus approximations -- *e.g.* mean-field, density-functional, time-dependent mean-field, projection of mean-field onto exact symmetries, *etc.*

How "unique" in the nucleon-nucleon interaction? --*i.e.*, the input V_{ijkl} look like random numbers; what happens if we actually use random numbers?

-- can we use ambiguity in the interaction to our advantage?

Looking for and using broken symmetries --isospin breaking and the unitarity of the CKM matrix

CONCLUSIONS

Basic problem: find extremal eigenvalues of very large, very sparse Hermitian matrix

Lanczos algorithm

fundamental operation is *matrix-vector multiply*

Despite sparsity, nonzero matrix elements can require TB of storage

Only a fraction of matrix elements are unique; **most are reused.** Reuse of matrix elements understood through *spectator* particles.

Reuse can be **exploited using exact factorization** enforced through *additive/multiplicative quantum numbers*