Abstract

We describe techniques that we are implementing to move inhomogeneous dynamical mean-field theory simulations from two- to three-dimensions. Two-dimensional simulations typically run on 2,000–10,000 lattice sites, while three-dimensional simulations typically need to run on 1,000,000 or more lattice sites. The inhomogeneous dynamical mean-field theory requires the diagonal of the inverse of many sparse matrices with the same sparsity pattern, and a dimension equal to the number of lattice-sites. For two-dimensional systems, we have employed general dense LAPACK routines since the matrices are small enough. For three-dimensional systems, we need to employ sparse matrix techniques. Here, we present one possible strategy for the sparse matrix routine, based on the well-known Lanczos technique, with a long run of the algorithm and (partial) reorthogonalization. This approach is about two-times faster than the LAPACK routines with identical accuracy, and hence will become the standard we use on the two-dimensional problems. We illustrate this approach on the problem of increasing the efficiency for pre-forming dipolar molecules in K-Rb mixtures on a lattice. We compare the local density approximation to inhomogeneous dynamical mean-field theory to illustrate how the local density approximation fails at low-temperature, and to illustrate the benefits of the new algorithms. For a three-dimensional problem, a speed-up of 1,000 or more times is needed. We end by discussing some options that are promising toward reaching this goal.

1. Introduction

Strongly correlated electron materials are some of the most interesting materials found in nature. They are composed of electrons that interact so strongly with neighboring electrons (via the Coulomb repulsion) that the motion of each electron is correlated with the motion of all other ones (analogous to how one moves squares in a Rubik’s cube or in the so-called “15” puzzle). These materials have novel properties like high-temperature superconductivity, magnetism, metal insulator transitions, etc., and the properties can be tuned by varying the temperature, pressure, external electric or magnetic field, and so on. Hence, these materials have the potential for use in a wide range of important devices ranging from sensors, to ultra-fast digital electronics, and novel field effect devices. Unfortunately, these systems require one to use quantum mechanics to describe the properties of these systems, and the quantum many-body problem is one of the most challenging problems to solve, as the Hilbert space grows exponentially with the number of particles, and rapidly becomes intractable. Typically, we say these problems are NP, meaning that they are non-polynomial on a classical computer, but if one has the answer, it can be checked in polynomial time. These problems should be QMA, which means they can be solved on a quantum computer. In the absence of having a general purpose “digital” quantum computer, one is left with a range of different approximation techniques, which sometimes work, and sometimes fail.

In 1982, Richard Feynman (1982) suggested an alternative way to solve this problem by using an analog quantum computer. He proposed that one set up a
controlled quantum mechanical system that describes the problem of interest, and then one simply lets the system evolve forward in time, and reads off the results of the simulation via different experimental probes, similar to an analog computer. At the time, there were no known systems to try to carry out this goal, but recently this has changed with new experiments on ultra-cold atomic systems in optical lattices. Defense Advanced Research Projects Agency (DARPA) is running the optical lattice emulator program to build the first generation of these quantum analog simulators, Abo-Shaeer (2010). In order to benchmark these simulators, one needs them to examine numerically-tractable problems which can be simulated on conventional computers, and verify that they produce the correct results. Our Challenge Project involves performing these conventional numerical simulations.

These optical lattice emulators are constructed by placing ultra-cold clouds of alkali atoms in the egg-carton-like potential energy surface created by two retro-reflected laser beams, which have a lattice spacing that is typically about 0.5 microns. Such, a lattice spacing is about one thousand times larger than the lattice spacing of real materials. Nevertheless, the atoms which hop between lattice sites exhibit the same kind of quantum mechanical effects as electrons do in real materials. Hence, they are ideal for use as optical lattice emulators. The one complication that these systems possess is that the atoms need to be placed in a (harmonic) trap that holds them at a fixed-location in space. The presence of the trap introduces an inhomogeneous potential in addition to the periodic lattice potential, which complicates their use as a simulator for electrons in real materials, which move on a lattice that has no trapping potential. Disentangling the effects of the inhomogeneity from those of the lattice is one of the challenges we face in carrying the optical lattice emulator program to fruition.

Our numerical approach is based on dynamical mean-field theory, which is a formal technique to solve the many-body problem which becomes exact in the limit of infinite number of spatial dimensions. In two or three dimensions, the approach is necessarily approximate, but it is believed that the approximation is quite good for many systems, and this approach is the only numerically tractable method that can describe fermionic systems at low-temperature without facing the so-called sign problem. We generalize the original formulation of dynamical mean-field theory to inhomogeneous systems (called IDMFT) to take into account the trap. The approach works with objects called Green’s functions which, describe the amplitude and phase for creating a particle at one lattice site at time \( t_1 \) and removing the particle from another lattice site at time \( t_2 \). The Green’s function can be parameterized by an object called the self-energy, which describes how the energy of the excitations is modified by the interactions between the particles, and how long-lived they are. The algorithm for solving IDMFT involves the following steps: 1) we start with a guess for the self-energy on each lattice site; 2) we calculate the local Green’s function by finding the diagonal of the inverse of a matrix (whose dimension is the number of lattice sites) which includes the self-energy and trapping potential for each lattice site on the diagonal, and the connections between neighboring lattice sites on the off-diagonal elements; 3) we extract the local effective medium by using Dyson’s equation to remove the local self-energy from the local Green’s function; 4) we solve a quantum impurity problem in the extracted dynamical mean-field for each lattice site to find the new local Green’s function; 5) we use Dyson’s equation with the new Green’s function and the old effective medium to find the new self-energy; and 6) we use the new self-energy in step (2) to continue the iterative algorithm. One repeats these steps until the functions have converged, which typically takes from 10 to 1,000 iterations. The two numerically-intensive parts of the algorithm are finding the diagonal of the inverse of the large sparse-matrix and solving the local impurity problem on each lattice-site. The former problem grows as \( N^3 \) with \( N \) the number of lattice sites, while the latter can range from being very-fast for some classes of models (like the Falicov-Kimball model studied here) to slow for more complicated quantum models (like the Hubbard model, which we plan to study in the near future), but grows linearly with the system size \( N \). Our focus in this paper is on describing how to improve the efficiency for calculating the diagonal of the inverse of a matrix, as the impurity solver part of the algorithm requires specialized knowledge and is an active research program within the many-body physics community.

2. Numerical Algorithms

Our initial algorithm described in Freericks (2009), uses conventional LAPACK and BLAS routines to solve the matrix inversion problem. It is very efficient (achieving about 2 Gflops per CPU on the Cray XT5) and can scale up to at least 3,000 CPUs. The efficiency derives primarily from the high-efficiency of the BLAS and LAPACK packages, which are highly portable and optimized for many different systems. But dense matrix routines do not take into account the sparsity of these matrices. In two-dimensional (2D) problems, one has at most 4 non-zero off-diagonal matrix elements for each row of the matrix that needs to be inverted, while for the three-dimensional (3D) system we have at most 6 non-zero off-diagonal matrix elements. This makes a 100x100x100 lattice in three-dimensions have a fraction
of 6/1,000,000 non-zero off-diagonal matrix elements, which is exceedingly sparse. In addition, the sparsity pattern of the matrix is identical to the pattern one uses for a discrete approximation of the Laplacian operator, which often is employed in differential equation solvers. Hence, much work that has been devoted to those problems can be adapted to our problem.

We begin by describing how one can use the Lanczos algorithm to find the diagonal of the inverse of the matrices we work with in the IDMFT algorithm. An efficient and accurate method for calculating the diagonal of the inverse of a symmetric (or even hermitian) \( N \times N \) matrix, \( G^{-1} \), is based on the Lanczos algorithm, Lanczos (1950) and Parlett (1998). The Lanczos algorithm uses a three-term recurrence to progressively reduce a real symmetric matrix into tridiagonal form. At the \( j \)th step, where \( j \) does not exceed the dimension \( N \) of the matrix \( G^{-1} \), the algorithm produces an orthogonal basis \( q_1, q_2, \ldots, q_j \) such that the representation of the original matrix \( G^{-1} \) on the subspace represented by this basis is a tri-diagonal submatrix \( T_j \). Using this basis and the representation \( T_j \), one can then approximate the \( j \)th-element of the diagonal of the inverse of the matrix \( G_{ij} \) using a series of transformations briefly introduced below [see Sidje and Saad, (2008)]. The process ends once the Lanczos algorithm reaches \( N \) (unless spatial symmetry operations are taken into account which allow us to equate the inverse matrix elements for symmetry equivalent lattice sites when there is no spontaneous symmetry breaking in the system and thereby terminate the algorithm with fewer steps), which then determines the entire diagonal of the inverse matrix. Once the algorithm reaches the full-diimension \( N \) of the matrix, the resulting diagonal of the inverse of \( G^{-1} \) is determined to high-precision (in practice, machine epsilon is reached). Unfortunately, often it is required to use the full \( N \) steps and this means that a fairly large amount of memory might be required when the size of the matrix is large, due to additional reorthogonalization of the Lanczos vectors (see, for instance, the discussion in Parlett (1998) on re-orthogonalization for more details). A parallel out of core memory algorithm which uses efficient input/output protocols is being implemented for this purpose.

Within dynamical mean-field theory, the matrix \( G^{-1}(\omega) \) includes a contribution from the complex valued self-energy \( \Sigma(\omega) \), on the diagonal plus a contribution also on the diagonal from the frequency, \( \omega \), that is either purely imaginary [a Matsubara frequency \( \omega = i\omega_n = \pi T(2n + 1) \)] or purely real (the real-time case using \( \omega \)); the off-diagonal elements are those defined by the (real-valued) hopping matrix on a square lattice. In other words, the inverse Green’s function \( G^{-1}(\omega) \) is neither real symmetric nor hermitian, as required by the Lanczos algorithm described above. Note; however, that the complex part of the inverse Green’s function is only on its diagonal, i.e., we have \( G^{-1} = S + iC \), where \( S \) is real-symmetric and \( C \) is a real-diagonal matrix that encodes the imaginary part. Then the Green’s function can be reformulated using the above decomposition into the following form: \( G = \text{diag}^{-1}(R + i\text{diag})C^{-1/2} \) with \( R = \text{diag}^{-1/2}SC^{1/2} \). The Lanczos tridiagonal matrix \( T_n \) can thus be built using the real-symmetric matrix \( R \) instead of the original complex-symmetric matrix \( G^{-1} \). The diagonal of \( G \) is retrieved at the end of the Lanczos steps using the shift on the complex plane: \( T_n + iI \) (associated with \( R + iI \), above), combined with an \( \text{LDL}^T \) transformation of \( T_n + iI \). Explicitly, one has the transformation:

\[
(R + iI)^{-1} \approx Q_j(T_j + iI)^{-1}Q^T_j = Q_j(L_jD_jL_j^T)^{-1}Q^T_j,
\]

where \( L_j \) is a sub-diagonal matrix with \( \eta_j \) being its sub-diagonal elements and with \( 1 \)’s on its diagonal, while \( D_j \) is a diagonal matrix with \( \delta_j \) on its diagonal, both at iteration \( j \) of the Lanczos procedure. After some linear algebra based on the equation above, one finds the following algorithm for the diagonal of the inverse at iteration \( j \) of the Lanczos process:

\[
p = q - \eta_j p; \quad \delta_j = (\alpha_j + i\eta_j) - \beta_j \eta_j; \quad \text{diaginv} = \text{diaginv} + \text{diag} \left[ \frac{p^T}{\delta_j} \right]; \quad \eta_{j+1} = \beta_{j+1}/\delta_j.
\]

The \( \alpha_j \)’s and \( \beta_j \)’s are respectively the diagonal and off-diagonal elements of the tridiagonal matrix \( T_j \); the vectors \( q \) are the updated Lanczos vectors at iteration \( j \) and the vectors \( p \) are intermediary vectors solution of the system of equations \( P_j = Q_jL_j^T \) taken into account by the \( \text{LDL}^T \) decomposition.

The algorithm above has been tested using both imaginary and real-time DMFT codes, for matrices up to \( N=71 \times 71 \) and for temperatures as low as 0.025 \( t \), with \( t \) being the hopping integral. In every simulation, the Lanczos algorithm gives the same accuracy (of the order of \( 10^{-11} \)) compared to direct dense solvers (i.e., LU-decomposition plus inversion using a system of equations resolving \( U^T L^T \)) based on LAPACK routines. The computing time of the Lanczos algorithm using partial-reorthogonalization [Lanczos (1950) and Parlett (1998)] is about 2–3 times faster than that of the direct method of LAPACK with the same accuracy. Note that no loss of accuracy is observed between full- and partial-reorthogonalization implementations of the Lanczos code.

3. Results

The problem that we have been examining recently is the problem of how to form dipolar molecules at ultra-cold temperatures on 2D square lattices. Dipolar molecules are of interest because they possess long-range forces, which could lead to interesting quantum states of matter (and perhaps even be used to make general-purpose quantum computers). There has been significant
progress in experiments which begin by cooling mixtures of fermionic \(^{40}\)K and bosonic \(^{87}\)Rb atoms in a trap, Ni et al. (2008). They can cool down to just above the Bose-Einstein condensation temperature of the Rb atoms, because below that temperature, the Boson cloud shrinks, and becomes spatially separated from the Fermion cloud. Next, the magnetic field is swept through a Feshbach resonance, which binds the K and Rb atoms together into a weakly bound halo molecule. Finally, a process called stimulated Raman adiabatic passage (STIRAP) is employed to coherently move the molecules from the excited state to the ground state without releasing any energy as heat. The overall efficiency of this process is about 20%, due primarily to the poor overlap between the K and Rb atoms during the Feshbach sweep. In a previous Challenge Project, Freericks et al. (2010), we showed how placing the atoms on an optical lattice (plus a trap) prior to doing the Feshbach sweep could increase the efficiency of dipolar molecule formation up to nearly 100%. The basic idea is that if we place the atoms on the lattice and tune the attractive interaction between the species, then we can achieve a situation with exactly one K and one Rb atom on each lattice site, which is perfectly primed for molecule formation via the Feshbach sweep.

Originally we ran our code in a Capability Applications Project (CAP) on the Cray XT% at Arctic Region Supercomputing Center (ARSC) (Pingos) and then completed it during our Challenge Project. We calculated the efficiency versus the entropy per particle and found that reducing the entropy per particle down to about 0.75 to 1 \(k_B\) is sufficient to have a significant increase in the efficiency for forming dipolar molecules. All of these runs took place with the LAPACK codes. We have subsequently examined these codes with the new Lanczos algorithm, and on the imaginary axis, we have achieved nearly a factor of 2 speed-up, with identical accuracy. This approach will become our standard in the future for these two-dimensional problems.

The other project we have been working on is an interim project for the 3D systems which sidesteps the issue of needing the diagonal of the inverse of a large sparse matrix during the computational algorithm. This approach is called the local density approximation, and just like how it is used in density functional theory, it assumes that we start with a system that is homogeneous, and we solve this system for a range of different chemical potentials at a fixed temperature. Then we assume that the Green’s function on the \(i\)th lattice site is determined by the homogeneous solution with a chemical potential given by the local chemical potential, equal to the chemical potential minus the trapping potential at site \(i\). This approach never needs the diagonal of the inverse of a large matrix, because it does not use the Dyson equation on the inhomogeneous system, but uses the Dyson equation of the homogeneous system, which becomes diagonal in momentum space. One of the problems with the local density approximation (LDA) is that if the homogeneous system has order, then it assumes the inhomogeneous system will also have order, at least for those sites where the chemical potential lies in the regime that is ordered in the bulk. In general, one would expect the inhomogeneous system to order at a lower temperature as the coherence of that order is “felt” over the inhomogeneous system. Hence, we expect that the LDA will be an accurate approach at high temperature only. In addition, the LDA assumes that the phases of the inhomogeneous system vary along contour lines given by equi-potentials of the trap potential, since that is how the local chemical potential varies. Hence, this approach will not typically show any kind of faceting, which often can occur in inhomogeneous systems at low-temperature due to the underlying lattice structure.

The model that we examined was the spinless fermionic Falicov-Kimball model which has two types of particles—light particles which can hop to nearest neighbors and heavy particles, which cannot, Falicov and Kimball (1969). When a light particle and a heavy particle sit on the same lattice site, they interact with an interaction \(U\), which can be attractive or repulsive (we consider the repulsive interaction here). We define the characteristic length of the trap to be equal to the length at which the trap potential energy is equal to the hopping integral. Previously, we studied these systems on a 51\(\times\)51 lattice with characteristic lengths varying from 12.9 to 30 lattice sites, Maska (2008). We found, in those cases, that the density of the particles did not smoothly go to zero before we reached the edge of our system (where the trap potential can be viewed to jump to infinity). So we now studied these systems on 101\(\times\)101 size lattices, where the system, with 625 light and 625 heavy particles does nicely in the lattice for temperatures up to \(T=0.5t\) and the interaction \(U=5t\). We compared the solutions for the LDA to those of the IDMFT to see when the simplified approximation breaks down. The LDA calculations are very fast for this model, and can be run on a laptop in a few minutes to an hour of CPU time. The calculations with the IDMFT, typically take tens-of-thousands of CPU hours to complete a low temperature run, and are run on high performance computing (HPC) resources. In Figure 1, we show a comparison of these two approaches at a temperature \(T=0.15t\), where the system is close to, but has not yet ordered. One can see that a radial density plot for the light (red) and heavy (black) particles is virtually indistinguishable for the LDA versus the IDMFT. The pattern of the heavy particles, with the symbol size being proportional to the density on a given site, and the plots showing just the inner 50\(\times\)50 section of the 101\(\times\)101 lattice, is plotted in Figure 2. They also agree very well between the two techniques (i.e., there is no faceting occurring yet).
Figure 1. Radial density slices for the LDA (left) and IDMFT (right) at T=0.15t for a characteristic length of 12.9 for the light fermions and 30.0 for the heavy fermions. The black curve is the density profile for the heavies, and the red curve is the density profile for the lights. Note how the two curves appear to be virtually identical.

Figure 2. Density plots on the plane for the heavy particles. The symbol size is proportional to the density. The LDA is on the left and the IDMFT is on the right. The parameters are the same as in Figure 1. Note the close agreement of the two images.

Figure 3. Radial density profiles for T=0.05t and the same traps as before. The LDA is on the left and the IDMFT is on the right. Note the larger disagreement between these results now.

Figure 4. Density plots at T=0.05. The LDA is on the left and the IDMFT is on the right. Note how the IDMFT profiles are rounder, thinner, and farther out. The LDA solution appears to show some faceting effects, but this arises from the fact that the ordered phase on the edges is a period 2 density wave phase which has stripes along the diagonals, which give the effect of faceting.

Hence, one can see that the LDA works very well at temperatures above the ordering temperature, but it fails below. We will be using this approach next when we examine the Hubbard model in three-dimensions. Here, we will examine the LDA at high temperatures, with the confidence that it should agree with the IDMFT result, which needs to await newer algorithm development for fast sparse matrix algorithms to find the diagonal of the inverse of a matrix. We also are working on a generalized gradient expansion correction to the LDA, which will allow for faceting, and might improve upon the calculation results without adding a significant load. Because the impurity solver for the Hubbard model requires significantly more time than the solver for the Falicov-Kimball model, even the LDA will require HPC resources to carry out the calculations in three-dimensions.

4. Other Algorithms

In addition to improving the performance and implementing parallel versions of the Lanczos algorithm applied to IDMFT, we are also investigating other approaches for an even higher speed-up. For instance, we noticed that at sufficiently high temperatures (at least 2t and above) the inverse Green’s function in the imaginary time case is diagonally dominant, which implies that the Green’s function is sparse with its elements decaying rapidly away from the diagonal. A probing method [Tang and Saad (2010)] using multi-coloring graph techniques was developed for this case. The probing method applied to IDMFT is designed to evaluate the diagonal of the inverse of a matrix [as opposed to preconditioning methods where probing is used for estimating the inverse of the matrix, see Saad (2003)]. For diagonally dominant matrices a 10- to 100-fold improvement of computer speed can be obtained compared to the Lanczos approach. However, most of the matrices in IDMFT are indefinite.
(at low temperature) and the efficiency of the probing method remains slightly better than that of Lanczos for those indefinite cases.

The iterative solvers in the package “ZITSOL”, see, e.g., Osei-Kuffuor and Saad (2007), has been implemented and applied to the indefinite matrices at low temperature. The algorithm leads to some improvements of the speed-up compared to Lanczos, for small matrices \(N<7^2\). The algorithm is currently being tested on larger matrices. Other algorithms are also being investigated such as a divide-and-conquer technique based on the standard Sherman-Morrison formula, see e.g., Sherman and Morrison (1949) and Golub and Van Loan (1996). Domain decomposition algorithms are also under development, as well as mixed methods which combine divide-and-conquer with the Lanczos-recurrence for the remaining term of the Sherman-Morrison formula. All of these algorithms are being developed for both the two- and three-dimensional cases. Finally we will also examine techniques based on the super LU approach super LU, Demmel et al. (1999).

5. Significance to DoD

DARPA’s interest in this problem is to ultimately build a materials science emulator out of ultra-cold atomic atoms. One can then hypothesize a particular material, program the emulator to simulate its properties, see if these properties are an improvement over currently known materials, and then devise a way to make the new materials with these targeted properties. We are still far away from this goal, but have made much progress with being able to make simpler quantum many-body problem emulators in a variety of different platforms and for a variety of different models.

6. Conclusions

In this work, we have shown how one can improve the speed of the IDMFT algorithm by at least a factor-of-two, and how such speed-up is needed to examine realistic problems with fermions in three-dimensions. We also showed that a conventional approximation, called the local density approximation, works remarkably well at high-temperatures, but fails as \(T\) is lowered. Finally, we discussed a range of different techniques that we will try to implement to achieve even higher speed-ups so we can successfully simulate realistic three-dimensional systems that show Hubbard-model physics. This latter project will be carried out over the next few years.

Acknowledgements

J.K. Freericks acknowledges support from the Army Research Office Grant Number W911NF0710576 with funds from the DARPA OLE program (for the development of the LDA approaches and for the IDMFT in two-dimensions) and the Air Force Office of Scientific Research under grant number FA9559-09-1-0617 (for the dipolar molecule work). H.R. Krishnamurthy acknowledges support from the DST (India). JKF, PC, and YS acknowledge support from the National Science Foundation under grant number OCI-0904597 (for development of the new Lanczos algorithm and future sparse matrix algorithms for the three-dimensional IDMFT). DoD HPC computer time was provided by the Arctic Region Supercomputer Center (ARSC), the Air Force Engineering and Research and Development Center (AFRL), and the Army Engineering Research and Development Center (ERDC). This project was supported primarily by Challenge Project DARPA-C4J in fiscal year 2009–2010.

References


