

ARPACK APPLICATIONS

This page contains a sampling of some of the research projects which have utilized the ARPACK and P_ARPACK software packages.

- ◆ [Large-Scale SVD for Image Reconstruction](#)
 - ◆ [Large Scale SVD for Molecular Dynamics Simulations](#)
 - ◆ [Stability Analysis of Crystal Growth](#)
 - ◆ [Reactive Scattering](#)
 - ◆ [Iterative Determination of Vibrational Energy Levels](#)
 - ◆ [Large-Scale Computation of Pseudospectra](#)
 - ◆ [Large-Scale Electromagnetic Eigenvalue Problems](#)
-

Large Scale SVD for 3-D Reconstruction of Images

Large Scale SVD for 3-D Reconstruction of Images from 2-D Projections Through Analysis of Projections from Electron Micrographs

We are currently working with [Dr. Wah Chiu](#) and his colleagues at the [Baylor College of Medicine](#) to increase the efficiency of their 3-dimensional image reconstruction software. Their project involves the identification and classification of 2-dimensional electron micrograph images of biological macromolecules and the subsequent generation of the corresponding high resolution 3-D images. The underlying algorithm is based upon the statistical technique of principle component analysis. An SVD of the data set is performed to extract the largest singular vectors which are then used in the classification procedure.

There are typically several thousand particle images in a working set, and in theory, each one characterizes a different viewing angle of the molecule. Once each image is aligned and processed to eliminate as much noise as possible the viewing angle must be determined. The set is then Fourier transformed, and each transformed image is added to a Fourier composite in a position appropriate to this angle. The composite is then transformed back to yield a high resolution 3-D image.

If we consider each 2-D image to be a vector in N dimensional space, where N is the number of pixels, we know that we can describe each image as a weighted sum of the singular vectors of the space defined by the set. The distribution of the images in this N dimensional coordinate system is then used to break up the set into classes of viewing angles, and the images in a class are averaged. Determining the left singular vectors of the space of images is presently the computational bottleneck of the process. As is true in many applications, the classification can be accomplished with relatively few of the singular vectors corresponding to the largest singular

values. The initial effort has been to replace the existing algorithm for computing the SVD with ARPACK. Use of ARPACK has increased the speed of the analysis by a factor of 7 on a Iris workstation and in addition has increased the accuracy of the results dramatically.

Research Project Web Site

(View animations!)

● [Herpes Simplex Virus: A-capsid](#)

● [Herpes Simplex Virus: B-capsid](#)



Large Scale SVD for Molecular Dynamics Simulations

Large Scale SVD for Data Compression and Analysis of Molecular Dynamics Simulations

Knowledge of the motion of proteins is important to understanding their function. Proteins are not rigid static structures, but exist as a complex dynamical ensemble of closely related substates. It is the dynamics of the system that permit transitions between these conformational substates and there is evidence that this dynamic behavior is also responsible for the kinetics of ligand entry and ligand binding in systems such as myoglobin.

X-ray crystallography is able to provide high resolution maps of the average structure of a protein, but this structure only represents a static conformation. Molecular dynamics, which solves the Newtonian equations of motion for each atom, is not subject to such approximations and can be used to model the range of motions available to the protein on short (nanosecond) time-scales. Time-averaged refinement extends the molecular dynamics concept, restraining the simulation by requiring the time-averaged structure to fit the observed X-ray data. With this methodology, an ensemble of structures that together are consistent with the dynamics force-fields and the observed data is generated. This ensemble represents a range of conformations that are accessible to proteins in the crystalline state.

One difficulty with time-averaged refinements is the plethora of generated conformations. No longer is one structure given as the model, rather thousands of structures collectively describe the model. Although it is possible to analyze this data directly by manually examining *movies* of each residue in the simulation until some pattern emerges, this is not a feasible approach. Nor is it feasible to watch an animation of the entire protein and discern the more subtle global conformation states that may exist. If this ensemble of conformations is thought of as a matrix however, then analytical tools from linear algebra can be used. Here the tool of choice is the SVD. The SVD can be used for both compressing the data required to represent the simulation and for locating interesting functional behavior.

Research Project Web Site

Stability Analysis of Crystal Growth

Imagine that a piece of crystal is placed in some undercooled liquid. As the crystal freezes, the interface between the solid and the liquid starts to change. The solid phase grows rapidly from the crystal by sending out branching fingers. This phenomenon is also called *dendritic solidification*, and is responsible for the complicated interface observed in snowflakes.

Both analytical and numerical solutions of the above problem are difficult to obtain because of the moving boundary. We are interested in analyzing the stability of a well known stationary solution that corresponds to a simple parabolic shaped moving front. The standard techniques of linear stability analysis lead to an eigenvalue problem. In particular, the rightmost eigenvalues of some convection diffusion operator are of interest. The stationary solution is said to be stable if there is no eigenvalue of positive real part.

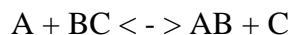
Research Project Web Site



[Crystal Growth](#)

Reactive Scattering

Reactive Scattering(**RS**) yields state-to-state information for chemical reactions providing the most detailed approach to the study of chemical reactivity. Atom-diatom (e.g. 3 atom) reactions of the type



are central in modeling of combustion systems, atmospheric chemistry, and gas-phase synthesis in interstellar clouds. Some examples are HeH_2^+ , HO_2 , LiHF , ClO_2 , FH_2 , NeH_2^+ .

RS calculations involve solving the Schroedinger equation for nuclear motion in an electronic field. In order to treat all nuclear arrangements equally (6 permutations of ABC), **A**diabatically adjusting, **P**incipal axes, **H**yperspherical (**APH**) coordinates are employed.

The resulting **P**artial **D**ifferential **E**quation(**PDE**) contains angles and hyperradius and is approximately split into a **PDE** involving angles only and **ODE** (**O**rdinary **D**ifferential **E**quation) involving hyperradius. This approach is similar to method of lines. The PDE depends

parametrically on hyperradius which defines sectors within which hyperradius varies slowly, and is therefore approximately constant. **The PDE** solution on each sector is independent of solution on any other sector, thus providing obvious opportunity for parallel processing.

The ODE in hyperradius is a well-known Riccati equation which is solved by log-derivative propagation method for each scattering energy. This method is chosen because a solution is needed for large radius limit only.

Discretization of the **PDE** is achieved by simplifying the potential energy part of the Hamiltonian to be diagonalized. Thus, the angular momentum eigenfunction basis of spherical harmonics is transformed into **Discrete Variable Representation (DVR)**. This discretization is similar to collocation method. The original **PDE** is thereby transformed into an algebraic eigenvalue problem of real symmetric matrix nicely solvable by ARPACK. The matrix diagonalization step is frequently the most time consuming in the whole RS calculation. The order of our matrices are typically at least in tens of thousands and can scale up to millions.

Vibrational Energy Levels

Iterative Determination of Vibrational Energy Levels of Four-Atom Molecules

[Rich Lehoucq](#), [MCS Division](#) and Steve Gray, Theoretical Chemistry Division, Argonne National Laboratory, have developed a parallel computer code that utilizes the implicitly restarted Arnoldi method to determine the vibrational eigenstates of four-atom molecules.

The determination of vibrational eigenstates from first principles, i.e., quantum mechanics, is conceptually simple but is a major computational challenge for most molecules, even ones with as few as four atoms. Such eigenstates determine the observable spectroscopy of molecules, and an understanding of their nature is essential for understanding how molecules behave.

Initial efforts have been to study the vibrational eigenstates of HOCO (Hydrogen-Oxygen-Carbon-Oxygen). This four-atom system is important in combustion and atmospheric chemistry. The presence of three relatively massive atoms (O and C are 16 and 12 times as massive as H, respectively), makes the problem particularly challenging. So far a series of experiments have been successfully performed which compute the lowest 41-52 vibrational states. The associated matrix eigenvalue problems were of order 1,092,420, 1,638,630 and 2,278,044. These problems took approximately 6, 8.4 and 11 hours, respectively, using 54 nodes of the IBM SPsystem at Argonne. The current goal is to generate the first 100 vibrational eigenstates of this system. Once completed, work will begin on other challenging four-atom systems, e.g., acetylene and hydrogen peroxide.

The [MPI](#) based parallel implementation of the implicitly restarted Arnoldi algorithm in the software package [ARPACK](#) was used to compute the vibrational levels.

Research Project Paper

- [Vibrational Eigenstates of Four-Atom Molecules: A Parallel Strategy Employing the Implicitly Restarted Lanczos Method](#)
-

Large Scale Computation of Pseudospectra

Large Scale Computation of Pseudospectra Using ARPACK and eigs

[Nick Trefethen](#) and [Tom Wright](#) of the [Oxford University Computing Laboratory Numerical Analysis Group](#) have shown that good estimates of pseudospectra in addition to eigenvalues can be obtained as a by-product of Arnoldi iteration with implicit restarts.

For some matrices, non-normality (non-orthogonality of the eigenvectors) may be physically important. In extreme cases, non-normality combined with the practical limits of machine precision can lead to difficulties in accurately finding the eigenvalues. Perhaps the more common and more important situation is when the non-normality is pronounced enough to limit the physical significance of eigenvalues for applications, without rendering them uncomputable. The pseudospectrum provides information on the physical significance of the eigenvalues of a matrix.

The pseudospectrum of a matrix A can be defined as

$\Lambda_\epsilon(A) = \{z \in \mathbb{C} : \sigma_{\min}(zI - A) \leq \epsilon\}$. This definition is the basis of many algorithms for computing pseudospectra. However, the computational work for determining the minimum

singular value of a general matrix of dimension N is $\mathcal{O}(N^3)$. Two important cost reducing techniques for computing the pseudospectrum of a matrix are projection of the matrix onto a lower dimensional invariant subspace and matrix reduction (to Hessenberg or triangular form). In combining these techniques Wright and Trefethen have developed a method that is much more

efficient, reducing the computational work to $\mathcal{O}(N^2)$. Their method uses the upper Hessenberg matrix constructed after successive iterations of the implicitly restarted Arnoldi algorithm given by ARPACK.

Research Project Website

- [Pseudospectra Gateway](#)
 - [Large-scale computation of pseudospectra using ARPACK and eigs \(SIAM J. Sci. Comp., to appear\)](#)
-

Large-Scale Electromagnetic Eigenvalue Problems

Large-Scale Electromagnetic Eigenvalue Problems

[Daniel White](#) and [Joseph Koning](#) of the Center for Applied Scientific Computing, [CASC](#), at [Lawrence Livermore National Lab](#) work on calculating the electromagnetic fields within closed perfectly conducting cavities that may contain dielectric or magnetic materials. The appropriate partial differential equation for this problem is the vector Helmholtz equation. Using the Galerkin method with H(curl) basis functions, EMSolve converts the Helmholtz equations to a generalized eigenvalue problem which is then solved by ARPACK.

This problem has direct applications in the analysis of linear accelerator components. To test their method they computed the lowest eigenmodes of a linear accelerator induction cell. Of particular interest was the magnitude of the electric field in the accelerating gap, as this determines whether or not the particular mode will couple with the electron beam. A 33024 cell hexahedral mesh was used to model the induction cell, creating an eigenvalue problem of dimension $N = 90237$. A total of 5 IRAM iterations were required for convergence of all 20 eigenmodes.

Research Project Website

 [EMSolve](#)