Real-Time Monitoring at CO2 Sequestration Sites: Fast Data Assimilation and Risk Evaluation

Investigators

Prof. Peter K. Kitanidis, Department of Civil and Environmental Engineering and Institute for Computational and Mathematical Engineering, Stanford University

Prof. Eric F. Darve, Department of Mechanical Engineering and Institute for Computational and Mathematical Engineering, Stanford University

Sivaram Ambikasaran, Doctoral candidate, Institute for Computational and Mathematical Engineering, Stanford University

Judith Yue Li, Doctoral candidate, Department of Civil and Environmental Engineering, Stanford University

Abstract

The availability of effective and reliable monitoring is recognized as a requirement for the acceptance of geologic sequestration of CO_2. The focus of this project is research that develops ultra fast computational methods for the real-time monitoring of CO_2 plumes and the evaluation of risks of leakage. The ultimate objective is the development of computational tools for data assimilation and uncertainty quantification based on sound fundamentals and numerical methods but adapted to specific problems.

The emphasis of this exploratory research is on demonstrating the potential of these methods through specific examples. We have developed and implemented algorithms that can process large data sets and estimate large number of unknowns, which will be of prime importance in real-time monitoring and optimal control at CO_2 sequestration sites. The methods assimilate data and also quantify uncertainty, which is important in weighing data of different types and in taking decisions that minimize the probability of failure. The algorithms speed up the time taken to solve large-scale problems by orders of magnitude compared to conventional methods. These algorithms have been applied and tested for synthetic data sets. These algorithms are becoming part of two software packages, under development, to enable solving inverse problems in real time.

The first package implements a novel algorithm for stochastic inversion. It is based on the fast multipole method and allows computing matrix vector products in a number of operations that increases linearly with m operations compared to compared with a much larger number of operations m^2 with a direct product. We coupled our novel fast method with the sparsity (zero fill-ins) of the underlying measurement operator. The final algorithm for static inversion has cost that increases with $m \times N$ where N is the number of measurements (compare with the direct method that has cost that increases with $m^2 \times N$.

The second package implements two fast direct solvers for a class of linear systems, which is relevant to linear inversion problems, with computational complexity with order of magnitude $m \times log^2 m$ and $m \times log m$ as opposed to the complexity of conventional direct solvers that is m^3.

This means that, as the size of the problem, *m*, increases, the above-mentioned methods become much faster than conventional methods.

In addition, we have been developing a Fast Kalman Filter for the sequential processing of data in time for a special case encountered in applications and have been comparing its speed and performance to traditional Kalman Filter (KF) and Ensemble Kalman Filter (EnKF) for the linear dynamic case. As can be shown from synthetic cases, the traditional KF, that is the optimal filter, is computationally very expensive to apply, particularly in updating covariance matrix at each time step. The EnKF reduces the cost of updating large covariance matrices by using sample covariance to approximate the true state error covariance; however, its performance is suboptimal. The method under development in this project has the potential to be less expensive than KF while being more accurate and versatile than EnKF.

Introduction

DOE's Research and Development Roadmap [1] affirms that "…the United States has a vast potential of geologic storage options (for CO2 sequestration)... *However, it is important to demonstrate and confirm the safe, effective, long-term geologic storage (permanence) of CO2"* (emphasis added). There is a pressing need to *assess risks* associated with decisions taken on the basis of guidance provided by data with limited information content that drive models with limited predictive ability. This is explicitly recognized [ibid.] "*Identifying and quantifying risks are also key to developing effective risk management strategies and permitting CCS projects"* (emphasis added). The IPCC report [2] stresses that "monitoring is a very important part of the overall risk management strategy for geological storage projects. Standard procedures or protocols have not been developed yet but they are expected to evolve as technology improves, depending on local risks and regulations."

Thus, it is generally recognized that monitoring at CO2 sequestration projects is indispensable in optimizing performance and minimizing risks. Performance and risks are evaluated and decisions are adjusted on the basis of mathematical models such as TOUGH2 or TOUGH2+CO2. Such models describe the state of the system on an ongoing basis through variables such as pressure and saturation and parameters of the geologic model, such as permeability. However, these models must be updated to take into account information from observations, and this updating must be performed at the points in time that the data are obtained. Essentially, monitoring is achieved by assimilating into models data in real time. The new information from the sensors must be weighed against the current estimates from these models, which reflect information from previous observations of the system. To weigh properly, and to also evaluate probability of risks, one must quantify the reliability of model projections.

The process of assimilating large data sets from periodic seismic surveys and almost continuous monitoring of pressure into large and nonlinear models is computationally very expensive. However, fast-linear algebra methods allow the exploitation of structure in the mathematical objects involved (such as sparsity or hierarchical data-sparsity in matrices) with potentially dramatic improvements in computational cost. This exploratory research demonstrates the potential of these methods through specific examples.

Background

More and more large-scale storage projects are initiated or become operational under the DOE Regional Carbon Sequestration Partnerships (RCSPs) and the Industrial Carbon Capture and Storage (ICCS) Programs in the US. In the rest of the world, similar programs encourage the development of geologic carbon sequestration projects. Meanwhile, developments and refinements in data collection techniques, such as 4D seismic surveys, are being reported throughout the literature. Recently, extensive monitoring of an overlying aquifer at CO2CRC Otway Project site was able to show that the risks are low, well-understood and manageable, whereas groundwater, soil, gas, and atmospheric monitoring provide full assurance at larger distances from the reservoir [3].

Results

We present here a numerical benchmark for a synthetic data set (static case) and a real data set (dynamic case, Kalman filter). The objective is not to capture the most realistic case or to show the complete potential of the methods we use but rather to test against standard methods.

Static

We first consider the static synthetic case, where the true slowness is shown in Figure 1. The results discussed below are the results obtained using the FLIPACK package we developed as part of this project.

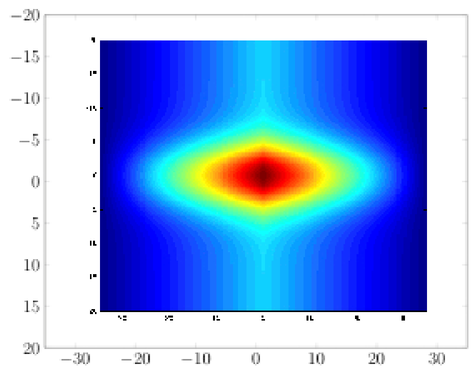

Figure 1: The image of actual slowness used in the synthetic case.

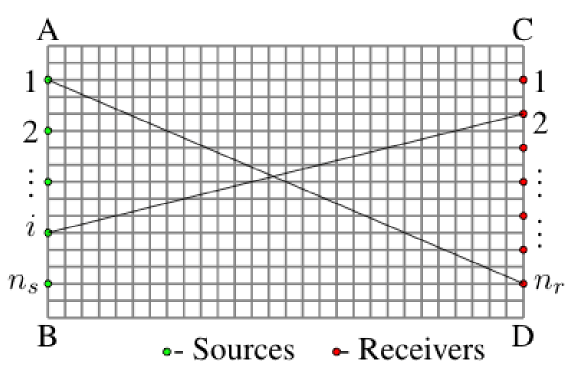

Figure 2: Cross section showing setup for seismic survey.

The domain is a 70m by 40m rectangular domain, as shown in Figure 2. A cross-well tomography survey is set up with sources along the vertical left line and receivers along the vertical right line. We image the slowness of the medium to determine the concentration of CO_2, where slowness is the inverse of the seismic velocity. If the CO_2 saturation and/or plume thickness increase along a given ray-path, the travel-time decreases, thereby allowing detection of the CO_2 plumes.

We have 3 uniformly spaced sources along the vertical left line and 12 uniformly spaced receivers along the vertical right line, giving us a total of 3*12 = 36 measurements. The rectangular domain is discretized into a total of *m* cells.

Typically, *m* is much larger than the number of measurements. This under-determined inverse problem is solved by our stochastic Bayesian approach using the novel fast algorithms. Advantages of the stochastic approach include that it allows more flexibility in using other information and also it allows the evaluation of the accuracy of the estimates. The novelty of the fast direct solver stems from an inventive application of the fast multipole method [4]. In addition, an improved version also exploits the underlying sparsity (i.e., the zero fill-ins) of the system. The sparsity is related to the fact that each ray crosses only elements placed on a straight line. The results are compared with a conventional direct solver.

Table 1: Comparison of performance.

Number of cells (m)	Time taken by MATLAB (in seconds)	Time taken by our fast algorithm (in seconds)
10,000	$3.23\ 10^{+0}$	$3.57\ 10^{-3}$
40,000	$2.13\ 10^{+1}$	$1.56\ 10^{-2}$
250,000	-	$1.23\ 10^{-1}$
1,000,000		$6.68\ 10^{-1}$
4,000,000		$3.24\ 10^{+0}$

The time taken by the fast algorithm is compared with a conventional direct solver. The results are shown in the table above. The table shows that our fast algorithm scales much better than conventional techniques. For instance, the algorithm can solve a 4 million grid in as much the same time as a conventional algorithm would take for a meager 10,000 grids. The algorithm not only wins on the running time taken, but also on the storage requirements. This scaling of the fast new algorithm will help us to handle very large inversion problems. The accuracy of the new algorithm is more than satisfactory for practical purposes. The error in the fast algorithm is quite small, in fact negligible compared to estimation (uncertainty) errors.

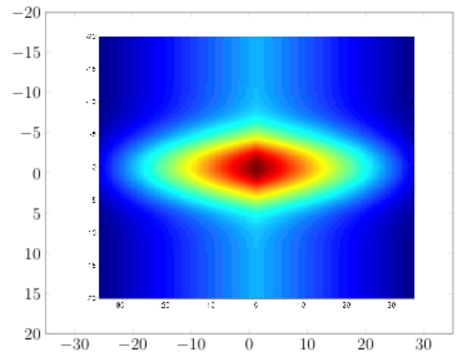

Figure 3: The image of slowness obtained based on a stochastic inversing method from cross-well seismic. Obtained for the 10^6-node grid.

Dynamic

FLIPACK handles static linear inverse problems arising from cross well tomography by applying fast multipole method exploiting the zero fill-ins. We are currently extending FLIPACK by implementing fast novel algorithms for Kalman filters to handle the dynamic case, to monitor the evolution of the CO2 plumes. The method essentially applies Kalman filter equations except that a version of the fast multipole method is used for critical matrix vector multiplications. The method is only partially developed and tested but the results so far have been encouraging.

We present some illustrative results for fast KF computed under the assumption of the "simplest possible" random walk forecast model, which essentially assumes that there is gradual change in the state between sequential, in time, images. In future applications, we intend to employ more reliable process based models, like TOUGH2. Nevertheless, a random walk forecast model seems appropriate and is useful when data is acquired rapidly at a rate faster than the discernible change of the system, in the absence of a valid physical model for state evolution [5].

For $m=3245$, $n=288$, running for 10 time steps on standard PC using Matlab, we used three different algorithms that give practically the same estimate at time step 10. The fast Kalman filter was almost two orders of magnitude faster.

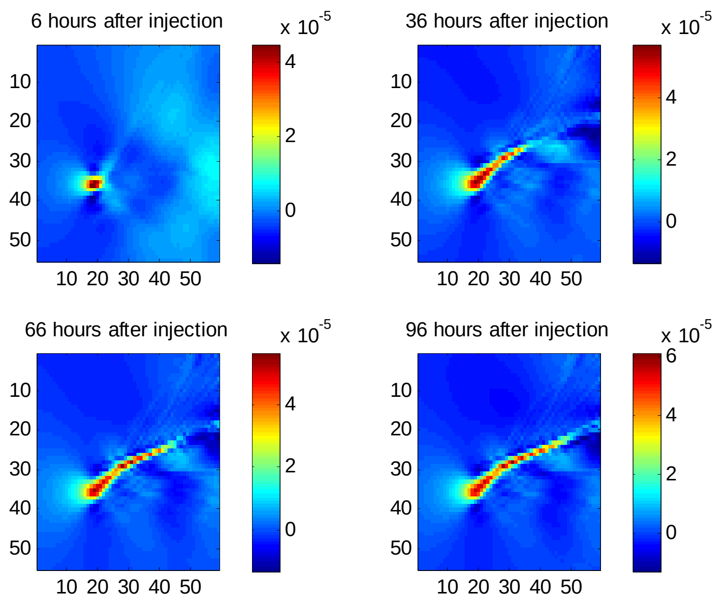

Figure 4: Change in slowness predicted by fast KF and KF

Kalman Smoother for seismic difference data

For CO_2 monitoring problem, we assume that the change in CO_2 saturation is directly related to the change in seismic velocity (or change in slowness, the reciprocal of velocity), which can be estimated by measuring the travel time difference. Then the state variable of interest is the slowness change compared to the pre-injection background slowness.

In practice seismic difference data is more valuable than the absolute measurement. Notice that the difference data is dependent on both the current state and a pre-injection state s_0, at which the baseline seismic survey was conducted. Both estimates are subject to uncertainty, which can be reduced through data assimilation techniques like Kalman Smoother (KS). KS refers to the use of the current data to infer about the current state as well as the past states, as KF only update the current state. We have made progress on developing such a smoother and results will be reported in forthcoming publications.

Fast Direct Solvers

We briefly give an example of the fast direct solver that uses the novel algorithms developed by the PIs. We calculate the factorization of a dense matrix of size N. The algorithm relies on using low-rank approximations for certain sub-blocks of the matrix. This is a hierarchical approach where the matrix is subdivided recursively in a nested fashion. The figure below indicates the rank structure of a certain dense matrix of size 128*128.

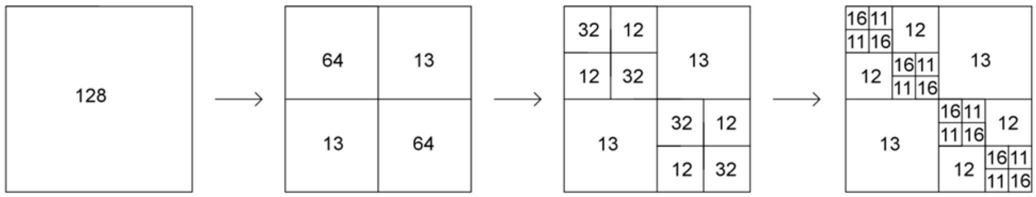

Figure 5: Rank structure of a covariance matrix.

The hierarchical low-rank property of sub-matrices is quite universal and is found in many matrices arising out of engineering applications. The result shown below is for a matrix, A, whose entries are given by $A(i,j) = 1/sqrt((x_i\text{-}x_j)^2 + a^2)$ where x_i are points in an interval in 1D. The results are obtained using FDSPACK, a package developed by us. This matrix is associated with the inverse multi-quadric bi-harmonic covariance function. This is to illustrate the accuracy and performance of the method and, in particular, to demonstrate how computational effort scales with the size of the problem. Similar results have been obtained for other covariance functions. Off-diagonal blocks are approximated using low-rank matrices. The rank is chosen large enough (rank = 15) that machine accuracy is essentially achieved. The condition number is $6{\times}10^3$. The result is still computed with enough accuracy for practical purposes, with a relative error around 10^{-12} to 10^{-9}.

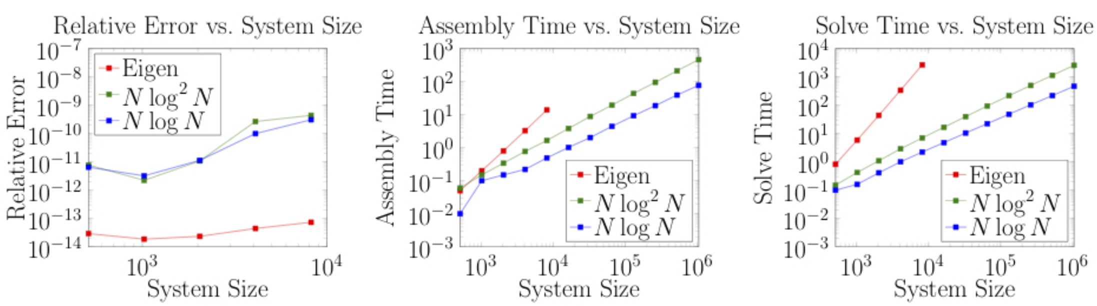

Figure 6: Performance and accuracy of fast direct solver.

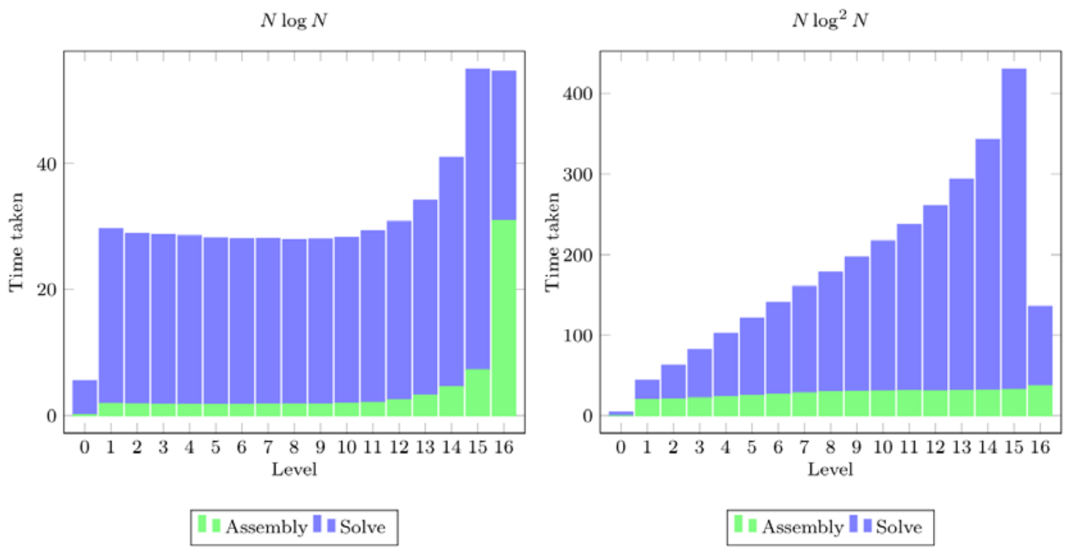

Figure 7: Time taken at each level in the hierarchy by the two fast algorithms.

The factorization of a dense matrix is computed using two fast algorithms developed by us. The first algorithm scales as N $\log^2$ N and the second one scales as N log N. The accuracy and running times are compared against a direct algorithm to solve the linear system. Eigen is a C++ template library for linear algebra. The package FDSPACK is at:

http://www.stanford.edu/~sivaambi/Fast_Direct_Solver_PACKage.html

Conclusions

We have developed and implemented a new fast direct solver (FDSPACK) which works for one-dimensional manifolds. This solver reduces the cost from order of magnitude N^3 to $N \log N$ where N is the number of unknowns. We also have developed and implemented an algorithm for fast linear inversion (FLIPACK) which reduces the cost of inversion from order of magnitude M^2 to order of magnitude M, where M is the number of grid points. We performed various numerical benchmarks to test FDSPACK and determine how to optimally choose the rank to have minimal numerical error.

We created a flexible C++ package that implements these algorithms. The packages can be found here

http://www.stanford.edu/~sivaambi/Fast_Direct_Solver_PACKage.html

http://www.stanford.edu/~sivaambi/Fast_Linear_Inversion_PACKage.html

These packages will be useful tools for the implementation of real-time monitoring at CO2 sequestration sites, where the efficient handling and processing of large data sets in real time are crucial. As the availability of effective and reliable monitoring is recognized as a requirement for the acceptance of geologic sequestration of CO2, these algorithms and packages have the potential for application at a significant scale thus providing another option in the reduction of CO2 emissions.

Publications and Patents

1. Ambikasaran, S., and Darve, E.F., (2013) "An *O(NlogN)* fast direct solver for partially hierarchical semi-separable matrices", *Journal of Scientific Computing*. Preprint/Journal version
2. Ambikasaran, S., Li, J.Y., Kitanidis, P.K., and Darve, E.F., (2013) "Large-scale stochastic linear inversion using hierarchical matrices", *Computational Geosciences*, under review.
3. Li, J. Y., Ambikasaran, S., Darve, E.F., and Kitanidis, P.K. (2013) "An accurate and computationally efficient large-scale Kalman filtering algorithm for fast data assimilation illustrated with an application in seismic imaging", in preparation.

References

1. NETL, DOE, DOE/NETL Carbon Dioxide Capture and Storage RD&D Roadmap, http://www.netl.doe.gov/technologies/carbon_seq/refshelf/CCSRoadmap.pdf. 2010, National Energy Technology Laboratory.

2. Rubin, E., et al., Technical Summary, in IPCC Special Report on Carbon Dioxide Capture and Storage, Intergovernmental Panel on Climate Change. 2005, Cambridge University Press: Cambridge, U.K.

3. Jenkins, C.R., et al., Safe storage and effective monitoring of CO2 in depleted gas fields. Proceedings of the National Academy of Sciences, 2011.

4. Fong, W. and E. Darve, The black-box fast multipole method. Journal of Computational Physics, 2009. **228**(23): p. 8712 – 8725.

5. Mitchell, V. and R. Knight, Inversion of time-lapse electrical resistivity imaging data for monitoring infiltration. The Leading Edge, 2011(February 2011): p. 140-144.

Contacts

Peter K. Kitanidis: peterk@stanford.edu

Eric F. Darve: darve@stanford.edu

Sivaram Ambikasaran: sivaambi@stanford.edu

Judith Yue Li: yuel@stanford.edu