Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 18 (2013) 581 – 590

Procedia Computer Science

International Conference on Computational Science, ICCS 2013

A scalable parallel LSQR algorithm for solving large-scale linear system for tomographic problems: a case study in seismic tomography

He Huang[a,*], John M. Dennis[b], Liqiang Wang[a], Po Chen[c]

[a]*Department of Computer Science, University of Wyoming. Laramie, WY 82071, U.S.A.*
[b]*National Center for Atmospheric Research. Boulder, CO 80301, U.S.A.*
[c]*Department of Geology and Geophysics, University of Wyoming. Laramie, WY 82071, U.S.A.*

Abstract

Least Squares with QR-factorization (LSQR) method is a widely used Krylov subspace algorithm to solve sparse rectangular linear systems for tomographic problems. Traditional parallel implementations of LSQR have the potential, depending on the non-zero structure of the matrix, to have significant communication cost. The communication cost can dramatically limit the scalability of the algorithm at large core counts. We describe a scalable parallel LSQR algorithm that utilizes the particular non-zero structure of matrices that occurs in tomographic problems. In particular, we specially treat the *kernel* component of the matrix, which is relatively dense with a random structure, and the *damping* component, which is very sparse and highly structured separately. The resulting algorithm has a scalable communication volume with a bounded number of communication neighbors regardless of core count. We present scaling studies from real seismic tomography datasets that illustrate good scalability up to $O(10,000)$ cores on a Cray XT cluster.

Keywords: tomographic problems; seismic tomography; structural seismology; parallel scientific computing; LSQR; matrix vector multiplication; scalable communication; MPI

1. Introduction

Least Squares with QR factorization (LSQR) algorithm [1] is a member of the Conjugate Gradients (CG) family of iterative Krylov algorithms and is typically reliable when a matrix is ill-conditioned. The LSQR algorithm, which uses a Lanczos iteration to construct orthonormal basis vectors in both the model and data spaces, has been shown to converge faster than other algorithms in synthetic tomographic experiments [2].

Noninvasive tomographic problems that focuses on determining characteristics of an object (its shape, internal constitution, etc.) based on observations made on the boundary of the object is an important subject in the broad mathematical field known as inverse problems. Each observation d made on the boundary can usually be expressed as a projection of the unknown image $m(x)$ onto an integration kernel $K(x)$, whose form is highly problem-dependent. In mathematical form, this type of problems can often be expressed as $\int K(x)m(x)\,\mathrm{d}V(x)$. This integration equation can be transformed to a linear algebraic equation of the unknown image by discretizing x. The resulting inverse problem can be highly under-determined, as the number of observations can be much less

*Corresponding author. Email address: hhuang1@uwyo.edu

Selection and peer review under responsibility of the organizers of the 2013 International Conference on Computational Science
doi:10.1016/j.procs.2013.05.222

than the total number of unknown parameters. Under such conditions, the inverse problem needs to be regularized and a regularization matrix, which is usually highly sparse with diagonal or block-diagonal structure, is appended below the kernel matrix.

Unfortunately, it can be computationally very challenging to apply LSQR to such tomographic matrix with a relatively dense kernel component appended by a highly sparse damping component, because it is simultaneously compute-, memory-, and communication-intensive. The coefficient matrix is typically very large and sparse. For example, a modest-sized dataset of the Los Angeles Basin (ANGF) for structural seismology has a physical domain of $496 \times 768 \times 50$ grid points. The corresponding coefficient matrix has 261 million rows, 38 million columns, and 5 billion non-zero values. The number of non-zeros in kernel is nearly 90% of the total while damping takes approximately 10%. The nearly dense rows within the coefficient matrix can generate excessive communication volume for a traditional row-based partitioning approach. Advances in structural seismology will likely increase the order of the design matrix by a factor of three. Clearly, an algorithm that scales with both problem size and core count is necessary.

In this paper, we address the computational challenges of using LSQR in seismic tomography which is made as a representative of tomographic problems. We propose a partitioning strategy and a computational algorithm that is based on the special structure of the matrix. Specially, SPLSQR contains a novel data decomposition strategy that treats different components of the matrix separately. SPLSQR algorithm results in an algorithm with scalable communication volume between a fixed and modest number of communication neighbors. SPLSQR algorithm enables scalability to $O(10,000)$ cores for the ANGF dataset in seismic tomography.

2. Related Work

LSQR is applied in a wide range of fields that involve reconstruction of images from a series of projections. It has been widely used in geophysical tomography to image subsurface geological structures using seismic waves, electromagnetic waves, or Bouguer gravity anomalies, etc. In structural seismology, the coefficient matrix is usually composed of kernel and damping component. For ray-theoretical travel-time seismic tomography, each row of the kernel component is computed from the geometry of the ray path that connects the seismic source and the seismic receiver, which usually results in a very sparse kernel component [3]. For full-wave seismic tomography, each row of the kernel which represents Frechet derivative of each misfit measurement vector is nearly dense [4, 5, 6]. The purpose of the damping is to regularize the solution of the linear system. The damping can be computed from the inverse of the model covariance. In practice, to penalize the roughness and the norm of the solution, a combination of the Laplacian operator implemented through finite-differencing and the identity operator can be used as the damping component.

The LSQR method is one of the most efficient algorithms so far for solving very large linear tomography systems, whether they are under-determined, over-determined or both [7, 2, 8]. There are several existing implementations of parallelized LSQR. Baur and Austen [9] presented a parallel implementation of LSQR by means of repeated vector-vector operations. PETSc [10] has an optimized parallel LSQR solver. PETSc is a well-optimized and widely-used scientific library, but it does suffer performance issues when using sparse matrices with random non-zero structure [11]. Liu *et al.* [12] proposed a parallel LSQR approach for seismic tomography. They partitioned the matrix into blocks by row and gave an approach to compute matrix-vector multiplication in parallel based on distributed memory. Their approach requires reduction on both vector x and y in each iteration. An MPI-CUDA implementation of LSQR (PLSQR) is described in [13]. The matrix vector multiplication uses a transpose free approach and requires only one reduction on the relative smaller vector x. Its major computation portions have been ported to GPU, and considerable speedup has been achieved.

3. Algorithm Overview

Figure 1 summarizes the idea of SPLSQR algorithm based on Message Passing Interface (MPI) programming model. The pseudo code describes the behavior of each MPI task. In particular, we describe the necessary matrix reordering, *i.e.*, line (01) to (04) in Section 4, data partitioning strategy, *i.e.*, line (05) to (08) in Section 5, and the calculation of the sparse matrix-vector multiplication, *i.e.*, line (10) to (26), in Section 6. Communication, *i.e.*, line (16) and (22), incurred during matrix-vector multiplication is detailed in Section 7.