

**Toward using direct methods in seismic tomography: Computation of the full
resolution matrix using high performance computing and sparse QR
factorization**

Petros Bogiatzis^{1*}, Miaki Ishii¹ and Timothy A. Davis²

¹*Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA*

²*Department of Computer Science and Engineering, Texas A&M University, TX 77843, USA*

**Corresponding author.*

Corresponding author details:

Petros Bogiatzis,
Department of Earth & Planetary Sciences, Harvard University
20 Oxford Street, Cambridge, MA-02138, USA.
Telephone: +1 617 495 2350 / Fax: +1 617 496 1907
e-mail: petrosbogiatzis@fas.harvard.edu

February 3, 2016

Keywords: Seismic Tomography, Inverse Theory, Computational Seismology.

Abbreviated title: Toward using direct methods in seismic tomography.

Summary

For more than two decades, the number of data and model parameters in seismic tomography problems has exceeded the available computational resources required for application of direct computational methods, leaving iterative solvers the only option. One disadvantage of the iterative techniques is that the inverse of the matrix that defines the system is not explicitly formed, and as a consequence, the model resolution and covariance matrices cannot be computed. Despite the significant effort in finding computationally affordable approximations of these matrices, challenges remain, and methods such as the checkerboard resolution tests continue to be used. Based upon recent developments in sparse algorithms and high performance computing resources, we show that direct methods are becoming feasible for large seismic tomography problems. We demonstrate the application of QR factorization in solving the regional P-wave structure and computing the full resolution matrix with 267,520 model parameters.

Introduction

Seismic tomography is the most powerful tool for imaging the interior of the Earth, and many geophysical problems, including seismic tomography, are typically reduced to solving a linear system of equations (e.g., Aster et al. 2005; Nolet 2008; Rawlinson et al. 2010). It is desirable to use accurate and robust direct methods or direct solvers (e.g., Demmel 1997) to obtain solutions. However, when the size of these inverse problems is large, the application of the direct solvers becomes computationally difficult. Over the last decade, advances in theory and methodology of forward and inverse problems, as well as increase in computational power and the number of available data, enabled the exploration of massive tomographic problems. For example, data availability has been accelerated with the introduction of dense networks such as the EarthScope USArray project (e.g., Meltzer et al. 1999), High Sensitivity

Seismograph Network in Japan (Okada et al. 2004), and WOMBAT Seismic Array in Australia (Rawlinson et al. 2008). These data allow finer spatial resolution of the tomographic models, yielding inverse problems with $O(10^5) - O(10^7)$ observations and $O(10^5) - O(10^6)$ unknown parameters (e.g., Nolet 2008; Rawlinson et al. 2010, 2014).

As an alternative, iterative solvers such as the LSQR method (Paige & Saunders 1982a,b), are used to avoid significant memory or computational run time. A disadvantage of the iterative techniques is that the inverse of the matrix that defines the system is not explicitly formed, hence measures of the robustness and the quality of the solution, such as the model resolution and covariance matrices, cannot be obtained. Consequently, a paradoxical situation arises. Although modern tomographic studies with large numbers of data and model parameters produce potentially better images in terms of spatial resolution, robustness and uncertainty of these models become obscured. In other words, improvements cannot be accurately assessed or confirmed.

Different solutions have been proposed to overcome this challenge. Attempts have been made to assess the spatial resolution of the model by means of synthetic reconstruction using checkerboard models or the results of the inversion as inputs (e.g., Rawlinson et al. 2014). However, such tests are misleading as they consider only the linear combinations of the columns of the resolution matrix (Lévêque et al. 1993; Rawlinson et al. 2014). Alternatively, methods based on LSQR can produce an effective resolution matrix (Zhang & McMechan 1995, 1996; Yao et al. 1999; Berryman 2000a,b; Zhang & Thurber 2007), but this matrix may differ significantly from the true resolution matrix (Deal & Nolet, 1996; Berryman 2000a,b; Zhang & Thurber 2007). Nolet et al. (1999) present another option of a one-step back-projection method, based upon the Moore-Penrose conditions (Jackson 1972), to obtain an explicit expression of the approximate inverse matrix to compute the full resolution matrix. For this approach, the accuracy of the resulting resolution matrix depends upon certain assumptions about the sensitivity matrix (Nolet et al. 1999; Yao et al. 2001). On the other hand, Soldati et al. (2006) use parallel Cholesky factorization on the normal equations in shared memory systems, explicitly forming the product of the matrix and its

transpose. One issue with such product is that it acquires a high condition number (square of the condition number of the initial matrix which is ill-conditioned), implying that the rounding errors can cause loss of information.

Recently, stochastic matrix probing techniques (Hutchinson 1990; Bekas et al. 2007) have been introduced to estimate the main diagonal elements of the resolution matrix (MacCarthy et al. 2011; An 2012; Trampert & Fichtner 2013). This method provides accurate assessment of the diagonal elements, but it does not calculate the off-diagonal elements that are useful in assessing the coupling between different model parameters, i.e., in identifying artifacts such as leakage between different parts of the model space. Furthermore, depending upon the structure of the resolution matrix, these methods require a large number of iterations to yield satisfactory approximations (Bekas et al. 2007). Another approach available for adjoint tomography applications is to utilize the second-order adjoints for extracting resolution information (Fichtner & Trampert 2011). This approach can be considered a generalization of the ray density tensor (Kissling 1988) that quantifies the space-dependent azimuthal coverage, and therefore serves only as a proxy for resolution. It is more efficient than other techniques, but is still costly, and the output is a direction-dependent resolution length, not the full resolution matrix with trade-offs between model parameters.

In this work, we revisit the application of direct methods in solving tomographic problems in view of significant developments in computer science and hardware in the last decade. For example, efficient rank-revealing algorithms based upon QR factorization that takes advantage of new architectures such as memory hierarchy and parallelism now exist (e.g., Davis 2011). Furthermore, new fill-reducing ordering algorithms effectively propagate sparsity throughout common factorizations (e.g., Karypis & Kumar 1998; Davis et al. 2004a,b; Amestoy et al. 2004). These advances are beginning to make direct methods feasible for problems with large matrices such as seismic tomography. We demonstrate that they can be used to solve a teleseismic tomography problem, and to compute the full resolution matrix. We also show that error propagation and the covariance matrix can be calculated with minimum additional cost.

Method

Seismic tomography is a non-linear and high-dimensional optimization problem where a set of model parameters is sought to satisfy observations and additional *a priori* constraints. In many cases, the problem is locally approximated in the vicinity of an optimal Earth model, and perturbation-based methods are used to acquire successive linear updates to the starting model until a convergence criterion is met (e.g., Rawlinson et al. 2014). The linearized problem in each iteration is constructed using the forward operator \mathbf{G} such that

$$\mathbf{G} \cdot \mathbf{x} = \mathbf{d}, \quad (1)$$

where \mathbf{x} is a model vector with m unknown parameters, \mathbf{d} is the data vector containing n observations or the residuals between observations and synthetics, and \mathbf{G} is an $n \times m$ matrix that, in most seismic tomography approaches, is sparse. A solution $\hat{\mathbf{x}}$ can be obtained by applying a linear matrix inverse operator, \mathbf{G}^{-} , i.e.,

$$\hat{\mathbf{x}} = \mathbf{G}^{-} \cdot \mathbf{d}. \quad (2)$$

The true model parameters \mathbf{x} and the estimated parameters $\hat{\mathbf{x}}$, are related through the model resolution matrix \mathcal{R} as (Jackson 1972; Menke 1989),

$$\hat{\mathbf{x}} = \mathcal{R} \cdot \mathbf{x}, \quad (3)$$

and

$$\mathcal{R} \equiv \mathbf{G}^{-} \cdot \mathbf{G}. \quad (4)$$

Eq. (3) suggests that the retrieved solution can be considered as a projection of the true model. The diagonal elements of \mathcal{R} describe the independent resolvability of each parameter, and the off-diagonal entries show the trade-offs between model parameters. In the ideal case, $\mathcal{R} = \mathbf{I}$ and $\hat{\mathbf{x}} = \mathbf{x}$, but in real tomographic applications, the resolution matrix deviates from the identity matrix with non-zero off-diagonal elements. The exact form of \mathcal{R} is related with the geometry of the tomographic problem (e.g., locations of sources and stations), the theory and assumption used for the forward operator (e.g., rays, sensitivity kernels), and the number and type of observations (e.g., arrival time measurements).

Computational aspect

Commonly, \mathbf{G}^- is chosen such that $\hat{\mathbf{x}}$ is a least squares solution to eq. (1), i.e., it is obtained by minimizing $\|\mathbf{G} \cdot \hat{\mathbf{x}} - \mathbf{d}\|^2$, where $\|\cdot\|$ is the Euclidian norm (e.g., Menke 1989; Aster et al. 2005; Nolet 2008). This yields the normal equations

$$\mathbf{G}^T \cdot \mathbf{G} \cdot \mathbf{x} = \mathbf{G}^T \cdot \mathbf{d}. \quad (5)$$

In most seismic tomography problems, the forward operator \mathbf{G} is rank deficient or nearly rank deficient even when $n > m$, since some rows are either exactly linearly dependent or nearly-linearly dependent in the presence of round-off errors and modeling uncertainties. Consequently, additional conditions are imposed such as some form of Tikhonov regularization operator \mathbf{L} of size $k \times m$ with $n + k \geq m$ (Tikhonov & Arsenin 1977). The regularized least squares problem is the same as eq. (1) except that \mathbf{G} is replaced by its regularized version \mathbf{A} and vector \mathbf{d} by \mathbf{b} where

$$\mathbf{A} = \begin{bmatrix} \mathbf{G} \\ a\mathbf{L} \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}, \quad (6)$$

and a is a scalar that controls the weight of the regularization term. The corresponding regularized normal equations are

$$(\mathbf{G}^T \cdot \mathbf{G} + a^2 \mathbf{L}^T \cdot \mathbf{L}) \cdot \mathbf{x} = \mathbf{G}^T \cdot \mathbf{d}. \quad (7)$$

A potential problem in normal equations approach is that it yields less accurate solutions than direct methods when the initial problem is ill-conditioned, because the accuracy depends on the condition number of $\mathbf{G}^T \cdot \mathbf{G}$, which is the square of the condition number of \mathbf{G} . Additionally, although \mathbf{G} is typically a sparse matrix, $\mathbf{G}^T \cdot \mathbf{G}$ is significantly denser and requires additional memory for storage. The most accurate approach in solving both eq. (1) or its regularized version is through singular value decomposition (SVD) (Lanczos 1961; Demmel 1997; Snieder & Trampert 1999; Aster et al. 2005) or generalized singular value decomposition (e.g., Hansen 1990, 1992, 1998, 2007; Aster et al. 2005). However, as data and model spaces become larger, the computational cost of SVD in floating point operations, and in particular, memory requirements, makes it intractable. This arises because the sparseness of the numerical null space of \mathbf{G} is not exploited.

Sparse QR factorization

In this study, we select QR factorization since it avoids the explicit formation of $\mathbf{G}^T \cdot \mathbf{G}$. Although modern QR algorithms can be used to solve rank-deficient problems (Davis 2011; Foster & Davis 2013), we consider the regularized least-squares problem to be consistent with conventional seismic tomography modeling such as that presented in MacCarthy et al. (2011). The forward operator \mathbf{A} is therefore a full-rank, rectangular matrix with more rows than columns, i.e., $n + k > m$. The QR factorization of \mathbf{A} decomposes the matrix into a product $\mathbf{A} = \tilde{\mathbf{Q}} \cdot \tilde{\mathbf{R}}$ consisting of an orthogonal, square matrix $\tilde{\mathbf{Q}}$ and a trapezoidal matrix $\tilde{\mathbf{R}}$. The trapezoidal matrix $\tilde{\mathbf{R}}$ has the form $\tilde{\mathbf{R}} = [\mathbf{R}^T \quad \mathbf{0}^T]^T$, where \mathbf{R} is an $m \times m$ upper triangular matrix. The square orthogonal matrix $\tilde{\mathbf{Q}} = [\mathbf{Q} \quad \mathbf{Q}_0]$, consists of the ‘‘economy’’ QR factor \mathbf{Q} of size $m \times m$ that corresponds to \mathbf{R} , and \mathbf{Q}_0 that contains the remaining $n + k - m$ columns of

$\tilde{\mathbf{Q}}$. QR factorization solves the least-squares problem through minimization of an equivalent problem (e.g., Björck 1996; Gander et al. 2014), i.e.,

$$\left\| \tilde{\mathbf{Q}}^T \cdot \mathbf{A} \cdot \mathbf{x} - \tilde{\mathbf{Q}}^T \cdot \mathbf{b} \right\| = \left\| \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \cdot \mathbf{x} - \begin{bmatrix} \mathbf{Q}^T \\ \mathbf{Q}_0^T \end{bmatrix} \cdot \mathbf{b} \right\|. \quad (8)$$

Since \mathbf{R} is full rank, the least-squares solution of eq. (8) is the same as the solution to the linear system $\mathbf{R}\mathbf{x} = \mathbf{Q}^T\mathbf{b}$, and the objective function is at its minimum when $\|\mathbf{Q}_0^T\mathbf{b}\| = \left(\|\mathbf{b}\|^2 - \|\mathbf{Q}^T\mathbf{b}\|^2\right)^{1/2}$. Similar to SVD, the 2-norm is preserved in QR factorization since $\tilde{\mathbf{Q}}$ is orthogonal. Norm preservation implies no amplification of numerical error, ensuring that inaccuracies in \mathbf{A} or \mathbf{G} are not amplified (e.g., Demmel 1997).

Various direct solvers exist that take advantage of the sparsity of the \mathbf{A} matrix (e.g., Davis 2006), but they are considerably more complicated and difficult to implement than their generic counterparts. Among the complications is the need for efficient handling of the fill-in (i.e., non-zero entries) in the factors during the factorization, and to ensure that the sparseness propagates throughout the procedure. For example, appropriate permutation of rows and columns of the input matrix results in the \mathbf{R} matrix with vast difference in its sparseness. An example shown in Fig. (1) considers a matrix that is formed from an upper rectangular block and a lower diagonal matrix, a typical structure of a tomographic problem with 0th order Tikhonov regularization. Only 0.9% of the entries of this matrix are non-zero, but QR factorization of this matrix yields an upper triangular \mathbf{R} matrix that consists of 46% non-zero entries. A different ordering of the columns of the initial matrix, on the other hand, produces significantly sparser \mathbf{R} factor with only 9.4% of the factors being non-zero (Fig. 1).

In this study, we use the SPQR algorithm from the SuiteSparseQR package (Davis 2011), a high-performance, parallel algorithm based on the multifrontal method (Duff & Reid 1983). The factorization of large sparse matrices is broken into multiple factorizations of smaller dense submatrices, and the structure of the algorithm has a dendritic organization, suitable for parallel computing. Each node in the

tree is assigned to the factorization of a dense submatrix, a frontal matrix. The results of lower order nodes are gathered to assemble frontal matrices of higher order nodes through irregular, one-way data communication that occurs only from the lower to higher order nodes (Davis 2011).

The SPQR algorithm includes the ability to represent and store the orthogonal matrix \mathbf{Q} in an efficient sparse format using the Householder transformations (Householder 1958; Davis 2011). Nevertheless, a more efficient alternative is to directly compute the product $\mathbf{Q}^T \cdot \mathbf{b}$ during the calculation of \mathbf{Q} to avoid storing \mathbf{Q} , which is an option in the SuiteSparseQR package. We can obtain both the solution vector and the resolution matrix \mathcal{R} of the regularized problem by using

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{B}, \text{ where } \mathbf{B} = \begin{bmatrix} \mathbf{d}_{m \times 1} & \mathbf{G}_{m \times n} \\ \mathbf{0}_{k \times 1} & \mathbf{0}_{k \times n} \end{bmatrix}, \quad (9)$$

as the right-hand side of the regularized version of eq. (1). This formulation has the advantage that both the analysis and the factorization are performed only once. The QR factors are used for retrieving the model estimation that corresponds to the first column of \mathbf{B} , and the resolution matrix is obtained, in a column-wise manner, from the remaining columns of \mathbf{B} .

Following this approach, the error propagation can also be performed with minor additional cost. In real applications, the data vector \mathbf{d} is contaminated with errors due to noise and other uncertainties, and eq. (1) is written as

$$\mathbf{G} \cdot \mathbf{x} = \mathbf{d} + \mathbf{e}, \quad (10)$$

where \mathbf{e} is the vector containing the errors. The error propagates to the model estimation as

$$\hat{\mathbf{x}} = \mathbf{G}^{-} \cdot \mathbf{d} + \mathbf{G}^{-} \cdot \mathbf{e}. \quad (11)$$

The overall departure of the true model \mathbf{x} from the estimated model $\hat{\mathbf{x}}$ is the result of both limited resolution and error propagation (e.g., Snieder & Trampert 1999), i.e.,

$$e(\hat{\mathbf{x}}) = \mathbf{x} - \hat{\mathbf{x}} = (\mathbf{I} - \mathbf{R}) \cdot \mathbf{x} - \mathbf{G}^{-} \cdot \mathbf{e}. \quad (12)$$

The error propagation can be calculated with minimum cost by incorporating the vector \mathbf{e} as an additional column of \mathbf{B} (eqn. 9). In addition to the error propagation, the model covariance matrix, defined as $\mathbf{C} = \mathbf{A}^{-} \cdot (\mathbf{A}^{-})^T$ provides a measure of model uncertainty (e.g., Nolet et al. 1999). In QR factorization, due to orthogonality, \mathbf{C} can be rewritten as, $\mathbf{C} = \mathbf{R}^{-} \cdot (\mathbf{R}^{-})^T = (\mathbf{R}^T \cdot \mathbf{R})^{-}$ (e.g., Björck 1996; Gander et al. 2014), and using its symmetry, only the upper triangular part needs to be computed. There are several algorithms that efficiently compute \mathbf{C} or certain elements of it such as the diagonal elements (e.g., Björck 1996; Gander et al. 2014).

Application

To demonstrate the features and applicability of the sparse QR approach, we use the teleseismic data set from CREST (Colorado Rockies Experiment and Seismic Transects) (Aster et al. 2009; MacCarthy 2010) and obtain P-wave structure beneath Colorado. This data set has been used to estimate the diagonal elements of the resolution matrix using the matrix probing techniques (MacCarthy et al. 2011) and provides a good comparison basis. The model space is parameterized in the same manner as in MacCarthy et al. (2011) using 267,520 constant velocity blocks of $0.25^\circ \times 0.25^\circ \times 25$ km in size. We further follow MacCarthy et al. (2011), and regularize the least-squares problem, utilizing the Tikhonov regularization (Tikhonov & Arsenin 1977). The data vector \mathbf{d} consists of 19,608 mean-removed teleseismic P-wave travel-time residuals from 167 stations (Fig. 2), and the forward operator \mathbf{G} has the dimensions of $19,608 \times 267,520$. The regularization kernel $\mathbf{L}_{k \times m}$, with $k = 570,232$, is a combination of damping and Laplacian-smoothing in equal amounts, both scaled by the Lagrangian multiplier a that is set to 0.5 (MacCarthy 2010; MacCarthy et al. 2011). The sizes of the matrices are large, and using

conventional seismic tomography techniques and algorithms, this problem is considered unfeasible to solve directly.

Our setup is based on SuiteSparse 4.4.0 (Davis, 2009, 2011; available online at <http://faculty.cse.tamu.edu/davis/suitesparse.html>) with MATLAB interface compiled with METIS 4.0.1 (Karypis & Kumar 1998) and Intel's Threading Building Blocks 4.3 (Reinders 2007). Various algorithms in the SuiteSparse package are tested for the best permutation of the input matrix for sparseness preservation, and we find that METIS algorithm (Karypis & Kumar 1998) yields the best result for the QR factors. All computations are performed in one node at Harvard's Odyssey Cluster, using 32 CPU cores, and total shared memory of about 250 GB. The parameters for SPQR algorithm are selected such that the "nthreads" parameter is set to the number of cores and the "grain" parameter is set to 2 times the number of cores. Both these parameters control the multitasking performance of Intel's Threading Building Blocks, and their values comply with the recommended configuration of the algorithm (Davis 2009). The tolerance for treating a column as zero is set to have a 2-norm less or equal with 0.1. The matrix \mathbf{A} is then successfully factorized, and the factors \mathbf{R} and \mathbf{Q} are stored, the latter in the economic form of Householder reflections. The resulting \mathbf{R} factor is sparse with only 2.2% of elements being non-zero (Fig. 3a).

The model vector \mathbf{x} and the full resolution matrix \mathcal{R} should be attainable from these factors, or more efficiently, by using the SPQR_SOLVE algorithm (Davis 2011) where \mathbf{Q} is not stored at all (eq. 9). If the resolution matrix is sparse enough, this procedure is carried out efficiently with one call of the SPQR_SOLVE algorithm. However, the resolution matrix in this case is not sufficiently sparse and exceeds the available memory. To circumvent this problem, the \mathbf{B} matrix is broken into 10 equally divided columnar strips. If a strip fails to produce results due to a lack of memory, the strip is further divided in half. The model parameters and the full resolution matrix (Fig. 3b) are obtained using this approach. The resolution matrix with $267,520 \times 267,520$ elements contain ~40% of non-zero entries and requires over 200GB of storage space. Reduction in storage can be achieved by eliminating entries with

absolute values smaller than 10^{-3} . These elements are practically zero, and the application of the threshold leads to significant gains in hard disk space, decreasing the “non-zero” entries to $\sim 0.5\%$ and required storage to about 3GB. This thresholded version of \mathcal{R} is considered for the remainder of this manuscript.

Both the estimated model and diagonal elements of the resolution matrix are compatible with the results from MacCarthy et al. (2011). The relative difference ε_x between the estimated models following MacCarthy et al. (2011) approach that utilizes the LSQR algorithm, \mathbf{x}_{LSQR} , and the QR approach described here, \mathbf{x}_{QR} , calculated as $\varepsilon_x = \|\mathbf{x}_{LSQR} - \mathbf{x}_{QR}\| / \|\mathbf{x}_{LSQR}\|$, is 0.03% (Fig. 4; Supplementary Figure 1). Furthermore, the relative difference ε_{diag} , of the diagonal elements of the resolution matrix between the stochastic approximation of MacCarthy et al. (2011), \mathcal{R}_{sto}^{diag} , and the QR approach, \mathcal{R}_{QR}^{diag} , calculated as, $\varepsilon_{diag} = \|\mathcal{R}_{sto}^{diag} - \mathcal{R}_{QR}^{diag}\| / \|\mathcal{R}_{sto}^{diag}\|$, is 5.40%. This result does not change, up to the second decimal place, if we omit values smaller than 10^{-3} from \mathcal{R}_{sto}^{diag} . The relative difference between the initial version of \mathcal{R}_{sto}^{diag} and that after the application of the 10^{-3} threshold is 0.27%, confirming that these small entries, even for the diagonal elements, do not contribute significantly to the structure of the resolution matrix.

The availability of the full resolution matrix allows the evaluation of the trade-off between model parameters. This can be visualized by plotting the specific rows of the resolution matrix (Fig. 5). The example in Fig. 5(a) and 5(b) illustrates that trade-offs extent over large distances, well exceeding the region where the diagonal element of the resolution matrix is relevant. Even if the absolute values of individual off-diagonal elements are relatively small, the cumulative effect of these values associated with large number of model parameters can be significant. This implies that estimations of uncertainty are inaccurate, especially when the diagonal elements of the resolution matrix have small values.

Conclusions

Advances in computational science and hardware infrastructure are at a stage where large tomographic problems can be efficiently solved using direct methods. In addition, most seismic tomography matrices are sparse, and this feature can be used to further accommodate the analysis. Depending upon the algorithm and the implementation, the users can adjust some parameters of the algorithms such as the column 2-norm tolerance to determine when a column is ignored, but many are hard-wired. For example, “zero threshold” is typically defined as the tolerance threshold below which an element is considered zero and dropped during numerical operations. Software (including Sparse QR) and hardware that complies with IEEE-754 format for the floating point numbers use the smallest denormalized number as the zero threshold, which is $4.9407e-324$ for double precision calculations. This means that any element of a matrix above this value would be considered nonzero, and therefore stored in memory, even if it becomes as small as $1e-323$. This threshold is unrealistically small for many numerical applications, including seismic tomography, and results in pseudo-dense matrices, for example, during the computation of the resolution matrix. Introducing modifications to existing solvers based upon the particular needs and properties of seismic tomography problems, such as an option to change the “zero threshold”, can significantly increase the computational efficiency. For example, as mentioned previously, the requirement to break \mathbf{B} into smaller strips to fit the memory, and therefore to repeat the factorization procedure, is necessary, because the product of \mathbf{Q}^T with \mathbf{B} is dense, or pseudo-dense. Changing the “zero threshold” parameter would allow small values to be dropped as they are computed individually, instead of storing the dense result and removing entries using the 10^{-3} threshold. This would accelerate the procedure dramatically, since it would allow the computation of both the inversion solution and the full resolution matrix with only one factorization, requiring total computation time slightly more than one factorization time (~ 12 hours). Furthermore, it would significantly reduce the total required memory, allowing the computations to be carried out on a desktop computer.

One of the main advantages of direct sparse methods is the assessment of the model resolution through accurate calculation of the full resolution matrix, including both diagonal and off-diagonal elements. The

large number of parameters (i.e., columns), however, poses challenges in efficient examination and exploration of this matrix. One possible future direction is using algorithms similar to those applied to large networks and graphs (e.g., Batagelj & Mrvar 2003; Bullmore & Sporns 2009; Cohen & Havlin 2010; Davis & Hu 2011; Kenett & Havlin 2015). The resolution matrix can be considered as a graph where each element \mathcal{R}_{ij} describes the trade-off relationship between the parameters i and j .

Acknowledgements

The authors would like to thank the editor, Ingo Grevemeyer, and two reviewers Adrià Meléndez and Jonathan MacCarthy for their constructive comments and suggestions that helped to improve the manuscript.

References

- Amante, C. & Eakins, B.W., 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. *NOAA Technical Memorandum NESDIS NGDC-24, National Geophysical Data Center, NOAA*, doi:10.7289/V5C8276M, last accessed July, 2015.
- Amestoy, P.R., Davis, T.A. & Duff, I.S., 2004. Algorithm 837: AMD, an approximate minimum degree ordering algorithm, *ACM Trans. Math. Softw.*, 30(3), 381--388.
- An, M., 2012. A simple method for determining the spatial resolution of a general inverse problem, *Geophys. J. Int.*, 191, 849--864.
- Aster, R., Borchers, B. & Thurber, C., 2005. *Parameter Estimation and Inverse Problems*, Elsevier Academic, Burlington, MA.

- Aster, R., MacCarthy, J., Heizler, M., Kelley, S., Karlstrom, K., Crossey, L., Dueker, K. & the CREST Team, 2009. CREST experiment probes the roots and geologic history of the Colorado Rockies, *Outcrop.*, 58(1), 6--21.
- Batagelj, V. & Mrvar, A., 2003. Pajek - Analysis and visualization of large networks in M. Jünger & P. Mutzel eds., pp. 77--103, Graph Drawing Software, Springer, Berlin.
- Bekas, C., Kokiopoulou, E. & Saad, Y., 2007. An estimator for the diagonal of a matrix, *Appl. Numer. Math.*, 57(11--12), 1214--1229.
- Berryman, J., 2000a. Analysis of approximate inverses in tomography I. Resolution analysis of Common inverses, *Optim. Eng.*, 1, 87--115.
- Berryman, J., 2000b. Analysis of approximate inverses in tomography II. Iterative inverses, *Optim. Eng.*, 1, 437--473.
- Björck, Å., 1996. *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia.
- Bullmore, E. & Sporns, O., 2009. Complex brain networks: Graph theoretical analysis of structural and functional systems, *Nature Reviews Neuroscience*, 10, 186--198.
- Cohen, R. & Havlin, S., 2010. *Complex Networks: Structure, Robustness and Function*, Cambridge University Press.
- Davis, T.A., 2006. *Direct Methods for Sparse Linear Systems*, SIAM, Philadelphia.
- Davis, T.A., 2009. Users guide for SuiteSparseQR, a multifrontal multithreaded sparse QR factorization package, available online from <http://faculty.cse.tamu.edu/davis/suitesparse.html>, last accessed July, 2015.
- Davis, T.A., 2011. Algorithm 915, SuiteSparseQR: Multifrontal multithreaded rank-revealing sparse QR factorization, *ACM Trans. Math. Softw.* 38 (1), 8:1--8:22.

- Davis, T.A. & Hu, Y.F., 2011. The University of Florida sparse matrix collection, *ACM Trans. Math. Softw.*, 38(1), doi: 10.1145/2049662.2049663
- Davis, T.A., Gilbert, J.R., Larimore, S.I. & Ng, E.G., 2004a. Algorithm 836: COLAMD, A column approximate minimum degree ordering algorithm, *ACM Trans. Math. Softw.*, 30 (3), 377--380.
- Davis, T.A., Gilbert, J.R., Larimore, S.I. & Ng, E.G., 2004b. A column approximate minimum degree ordering algorithm, *ACM Trans. Math. Softw.*, 30 (3), 353--376.
- Deal, M.M. & Nolet, G., 1996. Nullspace shuttles, *Geophys. J. Int.*, 124, 372--380.
- Demmel, J.W., 1997. *Applied Numerical Linear Algebra*, SIAM, pp. 101--138, Philadelphia, doi: 10.1137/1.9781611971446.ch3
- Duff, I.S. & Reid, J.K., 1983. The multifrontal solution of indefinite sparse symmetric linear systems, *ACM Trans. Math. Softw.*, 9, 302--325, doi:10.1145/356044.356047.
- Fichtner, A. & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophys. J. Int.*, 187, 1604—1624, doi: 10.1111/j.1365-246X.2011.05218.x
- Foster, L.V. & Davis, T.A., 2013. Algorithm 933: Reliable calculation of numerical rank, null space bases, pseudoinverse solutions, and basic solutions using SuiteSparseQR, *ACM Trans. Math. Softw.*, 40(1), 7:1--7:23.
- Gander, W., Gander, M.J. & Kwok, F., 2014. Chapter 6: Least squares problems in *Scientific Computing - An Introduction using Maple and MATLAB*, pp. 261--385, Springer, New York, NY.
- Hansen, P.C., 1990. Relations between SVD and GSVD of discrete regularization problems in standard and general form, *Linear Algebra and Its Applications*, 141, 165--176.
- Hansen, P.C., 1992. Analysis of discrete ill-posed problems by means of the L-curve, *SIAM Review*, 34(4), 561--580.

- Hansen, P.C., 1998. *Rank-Deficient and Discrete Ill-Posed Problems, Numerical Aspects of Linear Inversion*, SIAM, Philadelphia.
- Hansen, P.C., 2007. Regularization tools version 4.0 for Matlab 7.3, *Numerical Algorithms*, 46, 189-194.
- Householder, A.S., 1958. Unitary triangularization of a nonsymmetric matrix, *J. ACM*, 5(4), 339--342, doi: 10.1145/320941.320947.
- Hutchinson, M., 1990. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines, *Commun. Stat. – Simul. Comput.*, 19, 433-450.
- Jackson, D.D., 1972. Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. Astr. Soc.*, 28(2), 97–109, doi:10.1111/j.1365-246X.1972.tb06115.x.
- Karypis, G. & Kumar, V., 1998. A fast and high quality multilevel scheme for partitioning irregular graphs, *SIAM J. Sci. Comput.*, 20, 359--392.
- Kenett, D.Y. & Havlin, S., 2015. Network science: A useful tool in economics and finance, *Mind & Society*, 15, 167.
- Kissling, E., 1988. Geotomography with local earthquake data, *Rev. Geophys.*, 26, 659--698.
- Lanczos, C., 1961. *Linear Differential Operators*, Van Nostrand, New York.
- Lévêque, J.J., Rivera, L. & Wittlinger, G. 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, 115, 313–318.
- MacCarthy, J., 2010., *The structure of the lithosphere beneath the Colorado Rocky mountains and support for high elevations*, Ph.D. thesis, N.M. Inst. of Min. and Technol., Socorro.
- MacCarthy, J.K., Brochers, B. & Aster, R.C., 2011. Efficient stochastic estimation of the model resolution matrix diagonal and generalised cross-validation for large geophysical inverse problems, *J. Geophys. Res.*, 116, doi: 10.1029/2011JB008234.

- Meltzer, A., Rudnick, R., Zeitler, P., Levander, A., Humphreys, G., Karlstrom, K., Ekström, G., Carlson, R., Dixon, T., Gurnis, M., Shearer, P. & van der Hilst, R. 1999. USArray initiative, *GSA Today*, 9(11), 8-10.
- Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic, San Diego, CA.
- Nolet, G., 2008. *A Breviary of Seismic Tomography*. Cambridge University Press. Cambridge.
- Nolet, G., Montelli, R. & Virieux, J. 1999. Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.*, 138(1), 36--44.
- Okada, Y., Kasahara, K., Hori, S., Obara, K., Sekiguchi, S., Fujiwara, H. & Yamamoto A., 2004. Recent progress of seismic observation networks in Japan - Hi-net, F-net, K-NET and KiK-net, *Earth Planets Space*, 56(8), 15--28.
- Paige, C.C. & Saunders, M.A., 1982a. LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.*, 8(1), 43--71.
- Paige, C.C. & Saunders, M.A., 1982b. Algorithm 583: LSQR: Sparse linear equations and least squares problems, *ACM Trans. Math. Softw.*, 8(2), 195--209.
- Rawlinson, N., Tkalčić, H. & Kennett, B.L.N., 2008. New results from WOMBAT: An ongoing program of passive seismic array deployments in Australia, *EOS, Trans. Am. Geophys. Un.*, 89(53), Fall Meet. Suppl., Abstract S22A-03.
- Rawlinson, N., Pozgay, S. & Fishwick, S., 2010. Seismic tomography: a window into deep Earth, *Phys. Earth Planet. Inter.* 178, 101--135.
- Rawlinson, N., Fichtner, A., Sambridge, M. & Young, M.K., 2014. Seismic tomography and the assessment of uncertainty, in Dmowska R., ed., *Adv. in Geophys.*, Elsevier, 55, 1-76, doi: 10.1016/bs.agph.2014.08.001.

- Reinders, J., 2007. *Intel Threading Building Blocks. Outfitting C++ for Multi-core Processor Parallelism*, O'Reilly Media, Sebastopol, CA.
- Snieder, R. & Trampert, J., 1999. Inverse problems in geophysics, in wavefield inversion, A. Wirgin ed., pp. 119--190, Springer Verlag, New York.
- Soldati, G., Boschi, L. & Piersanti, A., 2006. Global seismic tomography and modern parallel computers. *Annals of Geophys.* 49, 4--5.
- Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solution of Ill-posed Problems*, Winston & Sons. ISBN 0-470-99124-0, Washington.
- Trampert, J. & Fichtner, A., 2013. Resolution tests revisited: The power of random numbers. *Geophys. J. Int.* 192, 676--680.
- Yao, Z.S., Roberts, R.G. & Tryggvason, A., 1999. Calculating resolution and covariance matrices for seismic tomography with the LSQR method, *Geophys. J. Int.*, 138(3), 886--894.
- Yao, Z.S., Roberts, R.G. & Tryggvason, A., 2001. Comment on 'Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems' by Nolet, G., Montelli, R. & Virieux, J., *Geophys. J. Int.*, 145(1), 307--314.
- Yeralan, S.N., Davis, T.A., Ranka, S. & Sid-Lakhdar, W., 2015. Sparse QR factorization on the GPU submitted to *ACM Trans. Math. Softw.*, 2015. GPU-accelerated software available at <http://faculty.cse.tamu.edu/davis/suitesparse.html>.
- Zhang, H. & Thurber, C.H., 2007. Estimating the model resolution matrix for large seismic tomography problems based on Lanczos bidiagonalization with partial reorthogonalization. *Geophys. J. Int.*, 170(1), 337--345.
- Zhang, J. & McMechan, G.A., 1995. Estimation of resolution and covariance for large matrix inversions. *Geophys. J. Int.*, 121(2), 409--426.

Zhang, J. & McMechan, G.A., 1996. Reply to comment by Deal, M.M. & Nolet, G., on “Estimation of resolution and covariance for large matrix inversions”, *Geophys. J. Int.*, 127, 251--252.

Figures

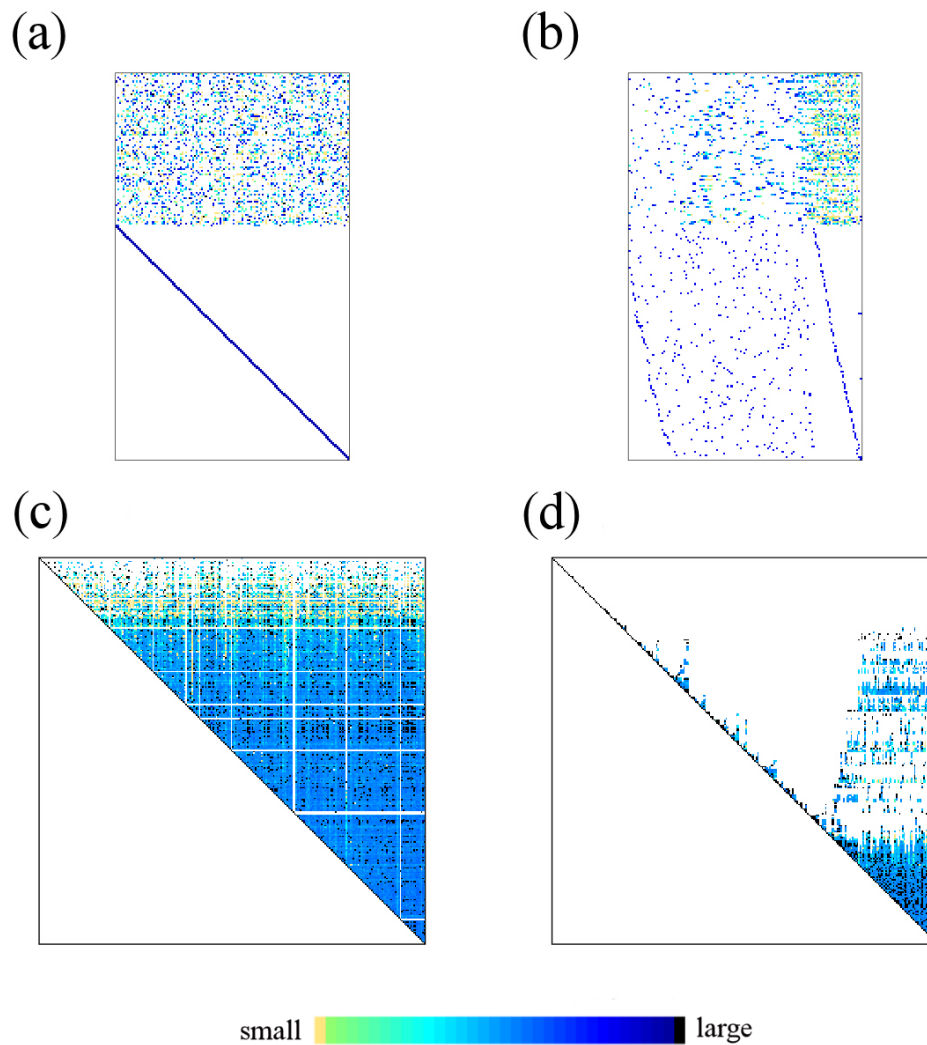


Figure 1: Effect of ordering on the sparseness. (a) An example matrix that contains 0.9% of non-zero elements. The colors show the size, i.e., $\text{median}(\log(|\text{value}|))$, of the entries with white being zero and dark blue colors indicating large absolute values. (b) The same matrix as in (a) with different ordering of its columns. (c) The upper triangular matrix \mathbf{R} obtained from QR factorization of the matrix shown in (a). (d) Same as in (c) except for the input matrix shown in (b). Note that the matrices are much larger than the pixel resolution (e.g., matrices in c and d are $267,520 \times 267,520$), hence the figures show only an impression of the structure of the matrix.

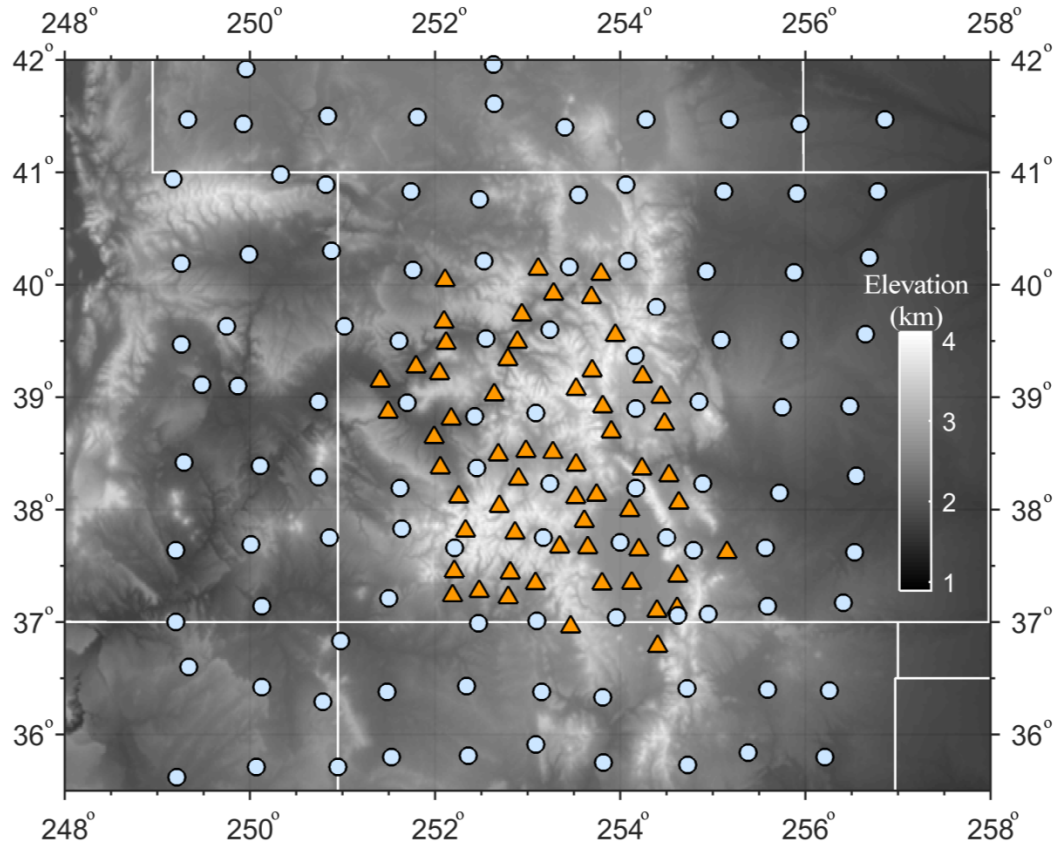


Figure 2: Map of stations from the teleseismic tomography experiment of MacCarthy et al. (2011). CREST stations are shown as orange triangles, and USArray stations are shown as blue circles. The background elevation is given in grey scale (Amante & Eakins 2009). The grey lines depict state boundaries.

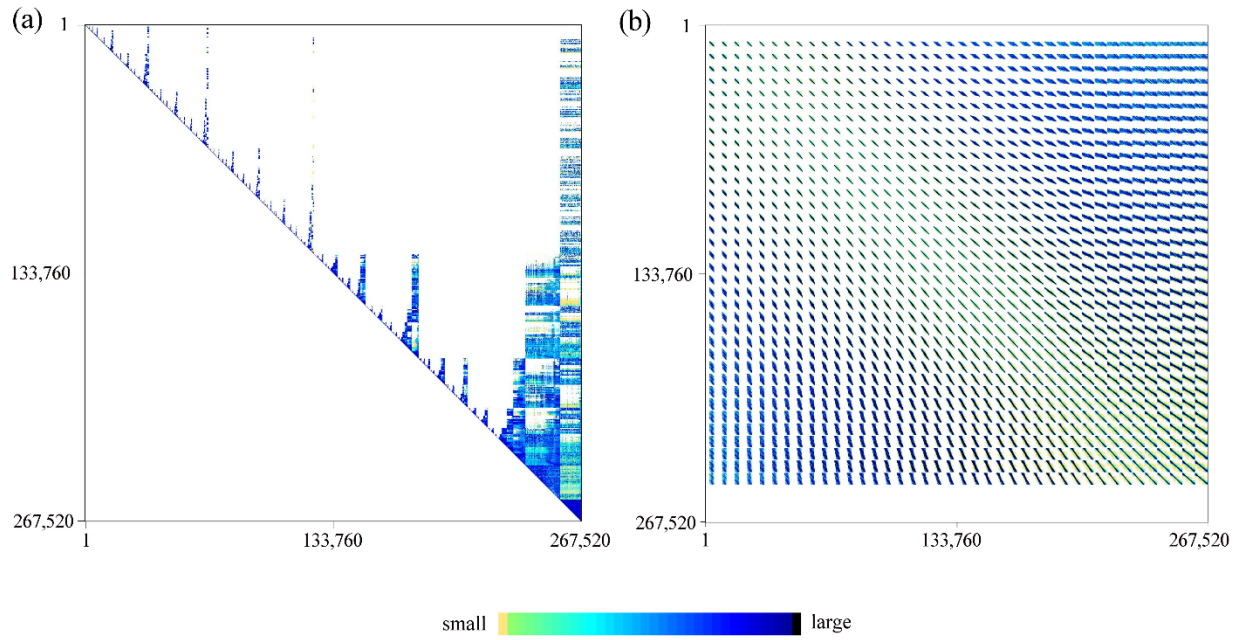


Figure 3: (a) Sparseness structure of the \mathbf{R} factor from QR decomposition. Zero elements are colored white while darker colors represent larger absolute values. (b) The full resolution matrix. The color scheme is same as in (a) except that white color includes all entries with absolute value smaller than 0.001.

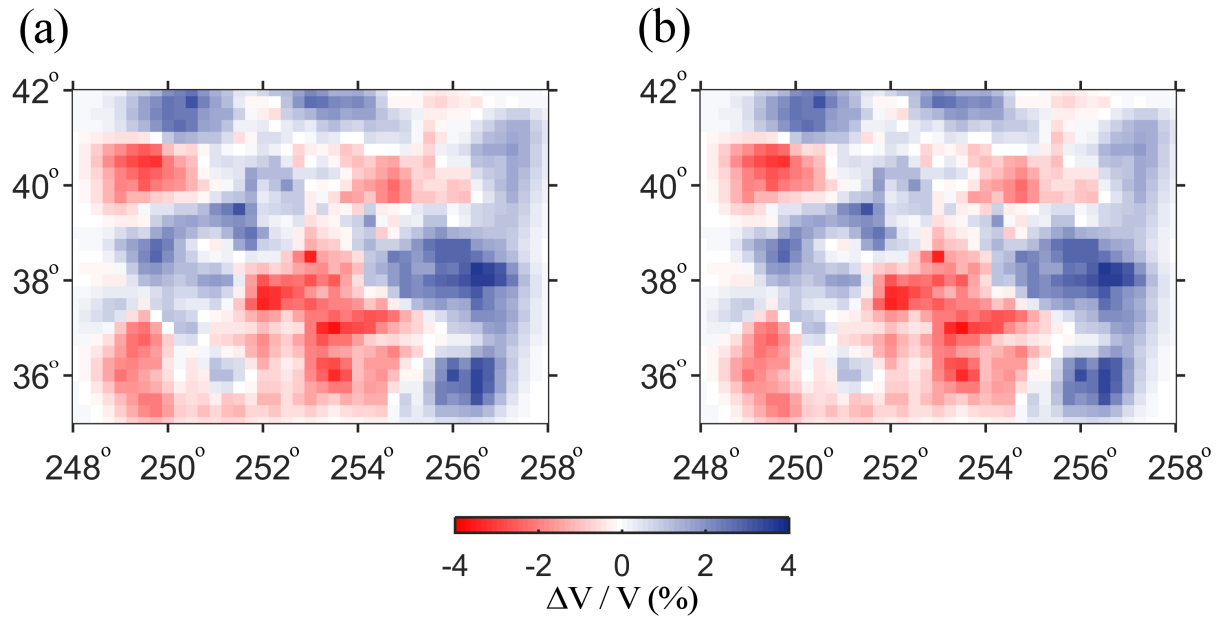


Figure 4: A comparison of P-wave perturbation models at 90 km depth obtained by using (a) the approach presented here, and (b) LSQR algorithm of the MacCarthy et al. (2011). The map area shown here corresponds to the area displayed in Figure 2. Additional comparisons at 20 and 400 km depths are presented in Supplementary Figure 1.

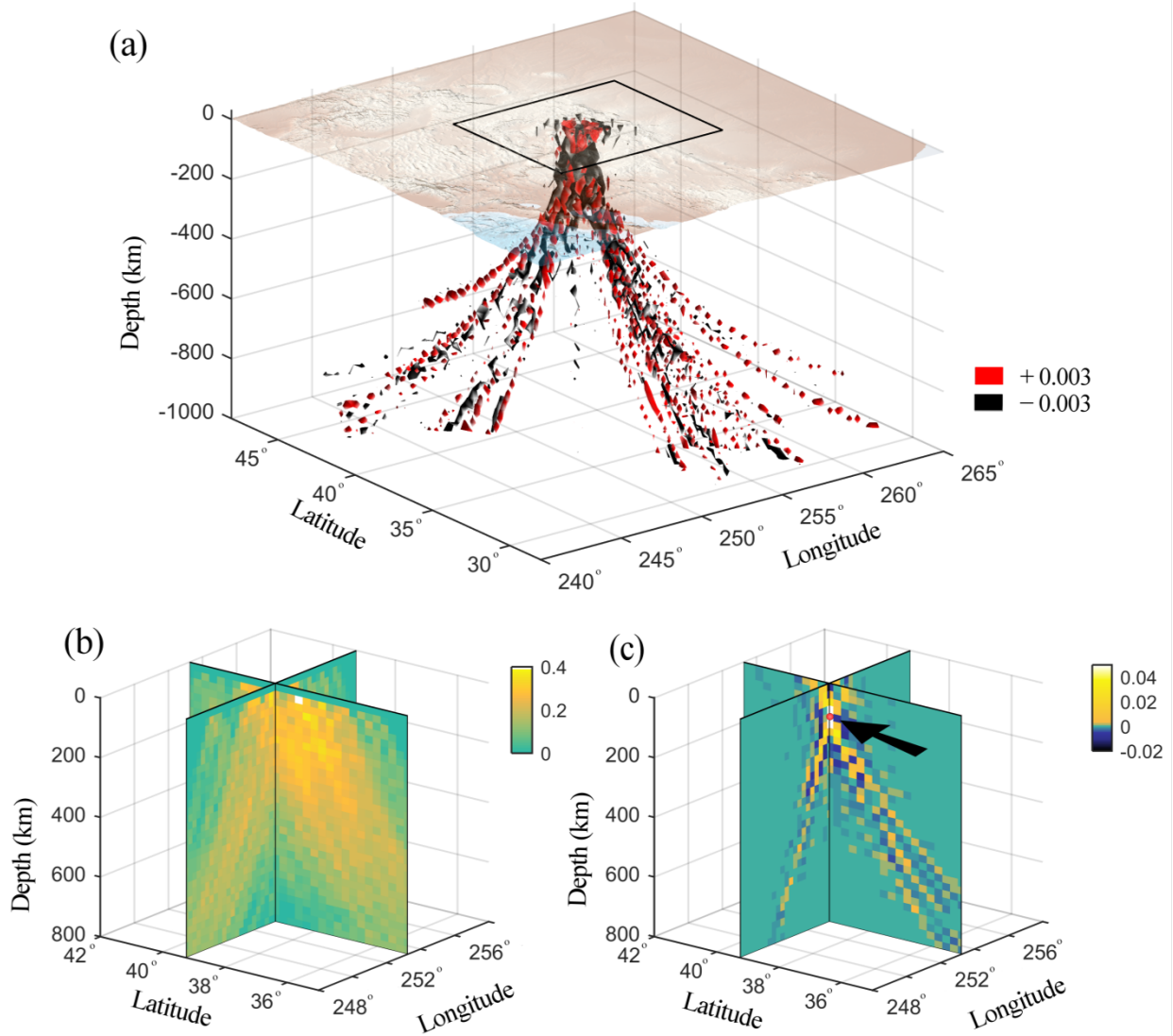


Figure 5: (a) Three-dimensional visualization of the 30,136th row of the resolution matrix, showing the isosurfaces at 0.003 (red) and -0.003 (black). In the topographic map, the rectangle with the thick grey border shows the region of the model displayed in (b) and (c) corresponding to the region interpreted by MacCarthy (2010) and MacCarthy et al. (2011). (b) Cross sections of the diagonal elements of the resolution matrix for the grey-boxed region in (a) and to a depth of 800 km. (c) Similar to (b) but for the 30,136th row of the resolution matrix. The location of the corresponding model parameter is shown by the red circle at the depth of ~100 km emphasized by the black arrow.