

The Dulmage-Mendelsohn Permutation in Seismic Tomography

Petros Bogiatzis^{1*}, Miaki Ishii² and Timothy A. Davis³

1. Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, MA 02138, USA., now at Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, European Way, Southampton, SO143ZH, UK. E-mail: p.bogiatzis@soton.ac.uk.

2. Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, MA 02138, USA.,

3. Department of Computer Science and Engineering, Texas A&M University, TX 77843, USA

Abstract

Seismic tomography inverse problems are among the largest high-dimensional parameter estimation tasks in Earth science. Although iterative algorithms can be used to efficiently solve these problems, their size gives rise to several issues such as the intractability of the computation of the model resolution and the model posterior covariance matrices that provide the means of assessing the robustness of the solution. In this work, we utilize methods from combinatorics and graph theory to study the structure of typical regional seismic body-wave tomography problems, and to effectively decompose them into subsets that can be solved efficiently by means of the least squares method. In combination with recent high performance direct sparse algorithms, this reduction in dimensionality allows for an efficient computation of the model resolution and covariance matrices using limited resources. We apply this methodology to a moderate size imaging of the structure of the crust and the upper mantle beneath Japan using deep local earthquakes recorded by the High Sensitivity Seismograph Network stations. Among the prominent features that are being imaged is a strong low-velocity region beneath the subducting Pacific slab along the entire Japan trench.

1. Introduction

Seismic tomography problems can be considered as high dimensional optimization problems, where a set of model parameters are sought to satisfy observations and *a priori* constraints. Seismic tomography problems are non-linear in principle, in that the inferred model perturbations alter the sensitivity kernels, although this non-linearity is often weak and is neglected (Ellsworth, 1977; Bijwaard & Spakman, 2000; Mercerat & Nolet, 2013). Due to the significant computational cost of the forward calculations, gradient-based optimization methods are the most convenient in solving seismic tomography problems. Commonly, a cost function, that combines the data misfit penalty and prior constraints, is locally approximated about an initial Earth model, and perturbation-based methods are used to acquire successive linear updates to the starting model until a convergence or termination criterion is met. Computationally, this is equivalent to solving a linear system of equations once or multiple times (e.g., Aster et al., 2005; Nolet, 2008). Modern tomography problems involve large numbers of observations and model parameters, and thus require the inversion of large matrices. Iterative solvers, such as Krylov subspace methods (e.g., Saad, 2003) allow the derivation of the solution without explicitly forming the inverse operator, and consequently, without large requirements in memory or computational time. One of the disadvantages of using iterative methods for the inversion is that the computation of the basic tools for assessing the quality and the first order uncertainty of the solution, namely the model resolution and the posterior covariance matrices (Backus & Gilbert, 1968; Wiggins, 1972; Menke, 1989; Parker, 1994; Tarantola, 1987, 2005), is no longer straightforward. It should be noted that the resolution and covariance matrices are typically defined for linear inverse problems, however their validity can be extended on linearized inverse problems (e.g., Snieder, 1991). Other frequently used resolution proxies, such as the

checkerboard tests have proven to be misleading (e.g., Lévêque et al., 1993), although under certain conditions can be useful (e.g., Rawlinson & Spakman, 2016).

Although several approaches based upon the LSQR (Paige & Saunders, 1982a,b) algorithm have been proposed to estimate the full resolution, and in some cases the covariance, matrix (e.g., Zhang & McMechan, 1995; 1996; Yao et al., 1999; 2001; Zhang & Thurber, 2007), such estimations can significantly depart from the actual resolution matrix. This is usually because only a small subspace of the full model space is explored, due to the limited number of iterations that is required for efficiency (Deal & Nolet, 1996; Berryman, 2000a, b; Zhang & Thurber, 2007). For example, the method proposed by Nolet et al., (1999), although efficient, is accurate only under certain assumptions about the structure of the forward operator. Alternatively, stochastic approaches using matrix probing techniques with randomized vectors (Hutchinson, 1990; Bekas et al., 2007) have been introduced to estimate the diagonal elements of the resolution matrix (MacCarthy et al., 2011; An, 2012; Trampert et al., 2013). They provide relatively accurate assessment of the diagonal elements (Bogiatzis et al., 2016), but ignore the off-diagonal elements that are crucial in assessing the trade-offs between different model parameters. Furthermore, depending upon the structure of the resolution matrix that may be unknown *a priori*, a large number of iterations may be necessary to yield satisfactory approximations (Bekas et al., 2007). Fichtner & Trampert (2011) propose a new method for extracting resolution information in full waveform inversion by utilizing the second order adjoints to compute the second derivatives of the data with respect to model parameters. This approach provides a generalization of the ray density tensor (Kissling, 1988) that quantifies the space-dependent azimuthal coverage i.e., a direction-dependent resolution length, but it does not address the trade-offs between model parameters. Recently, Voronin et al., (2014) proposed a

technique that clusters near parallel ray paths in global tomography problems, and then further gains in memory can be achieved by using singular value decomposition (SVD) to project the data onto a subspace associated with the largest few singular values of each cluster. This method is powerful in the case of global tomography studies and can potentially be used to derive the model resolution and covariance matrices from the compressed system. However, its application in the case of regional or local tomography problems is not straight-forward, as the ray-paths can diverge significantly in both the sources' and the receivers' side, and also due to the fact that the hypocenters parameters of the earthquakes are also among the unknowns to be estimated.

Several researchers have suggested overcoming computational limitations of the direct methods by utilizing high performance computers that incorporate many multicore nodes under a distributed memory environment. For example, parallel dense Cholesky factorization using such system has been utilized to solve large, at the time, seismic tomography inverse problems (Boschi 2003; Soldati & Boschi 2005; Soldati et al., 2006). More recently, a dense Cholesky solver that stores and fetches the matrices in significantly slower external memory such as hard drives (also known as out-of-the-core capability; D'Azevedo & Dongarra, 2000), has been used to calculate the resolution and covariance matrix for a global tomography model with ~230K parameters (Hipp et al., 2011; Ballard et al., 2016). This particular study required about 12 hours of computation time, using 400 threads distributed over 10 computational nodes, that had access to main memory that varied from 64 GB to 768 GB per node. Bogiatzis et al. (2016) have showed that similar problems can be solved efficiently with limited computational resources by using fill-reducing ordering algorithms to propagate sparsity throughout common matrix factorizations methods (e.g., Karypis & Kumar, 1998; Davis et al., 2004a, b; Amestoy et al., 2004), and efficient direct sparse solvers that take advantage of memory hierarchy and

parallelism (Davis, 2011). In order to make large tomography problems more tractable, this manuscript examines a graph-based method that permutes the non-zero elements of the seismic tomography matrix to produce a block, upper triangular matrix. This permutation divides the initial inverse problem into significantly smaller ones, achieving substantial dimensionality reduction that allows more efficient handling of large tomography problems.

2. Reducing large linear problems with the Dulmage-Mendelsohn permutation

Linear systems of equations can be represented with “bipartite graphs” (e.g., Pothen & Fan, 1990). For example, let an $m \times n$ matrix \mathbf{G} to be the matrix of a linear system of m equations with n unknowns. For convenience, let us assume $m \geq n$ with no loss of generality, since in the case that $m \leq n$ we only need to consider the transpose of \mathbf{G} . If \mathcal{R} and \mathcal{C} are the set of vertices that represent the indexes of the rows and the columns of \mathbf{G} , respectively, then we can define a bipartite graph that describes \mathbf{G} as an ordered triple $\mathbb{G}(\mathcal{R}, \mathcal{C}, \mathcal{E})$, where \mathcal{E} is the set of edges between \mathcal{R} and \mathcal{C} . Edge (i, j) corresponds to the non-zero element $\mathbf{G}(i, j)$, and connects the row i (member of \mathcal{R}), with the column j (member of \mathcal{C}) (Fig. 1).

In a bipartite graph, we call “matching” \mathcal{M} , a subset of edges with no common vertices (Fig. 2). A vertex is considered matched (or covered) if there is an edge of the matching that is incident on it. If such an edge does not exist, then the vertex is called unmatched (Fig. 2). When a matching has the largest number of vertices, then it is called “maximum matching”.

Following the graph formulation, a system of equations is structurally under-determined (or under-constrained), i.e., there are fewer equations than unknowns, when the system’s associate

bipartite graph has a maximum matching covering all nodes of \mathcal{R} and $|\mathcal{R}| < |\mathcal{C}|$ where $|\cdot|$ denotes the number of its members. Similarly, it is structurally over-determined, i.e., there are more equations than unknowns, when its associate bipartite graph has a maximum matching covering all nodes of \mathcal{R} and $|\mathcal{R}| > |\mathcal{C}|$. Finally, it is well-determined when its associate bipartite graph has a maximum matching covering all nodes of \mathcal{R} and $|\mathcal{R}| = |\mathcal{C}|$.

Dulmage and Mendelsohn, in a series of papers (1958, 1959, 1967), first described a unique decomposition of a bipartite graph that separates the graph into a structurally under-, well-, and over-determined parts that is now known as the Dulmage-Mendelsohn coarse permutation (see Appendix). The well-determined part can be further decomposed into smaller parts with fine decomposition (Pothen, 1984; Pothen & Fan, 1990; Davis, 2006), however, we will focus on the coarse decomposition that is more applicable to tomographic problems. The algorithm takes a bipartite graph $\mathbb{G}(\mathcal{R}, \mathcal{C}, \mathcal{E})$ as its input, and outputs three sub-graphs (i.e., subsystems) that correspond to an under-determined component \mathcal{U} , a well-determined component \mathcal{W} , and an over-determined component \mathcal{O} . In matrix notation, the application of the Dulmage-Mendelsohn permutation produces the following block triangular form of \mathbf{G} ,

$$\mathbf{P} \cdot \mathbf{G} \cdot \mathbf{P}' = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \mathbf{G}_{13} & \mathbf{G}_{14} \\ 0 & 0 & \mathbf{G}_{23} & \mathbf{G}_{24} \\ 0 & 0 & 0 & \mathbf{G}_{34} \\ 0 & 0 & 0 & \mathbf{G}_{44} \end{bmatrix}, \quad (1)$$

where \mathbf{P} and \mathbf{P}' are $m \times m$ and $n \times n$ permutation matrices obtained by permuting the rows of the identity matrix. \mathbf{G}_{12} , \mathbf{G}_{23} and \mathbf{G}_{34} are square matrices with non-zero diagonals. Furthermore, the diagonals of \mathbf{G}_{12} , \mathbf{G}_{23} and \mathbf{G}_{34} define a maximum marching of \mathbb{G} . The block $[\mathbf{G}_{11} \ \mathbf{G}_{12}]$ represents the structurally under-determined subsystem of equations \mathcal{U} and, if present, it is always rectangular with more columns than rows. \mathbf{G}_{23} is the structurally well-determined

subsystem \mathcal{W} , and it is always square. Finally, $[\mathbf{G}_{34}^T \ \mathbf{G}_{44}^T]^T$ is the structurally over-determined subsystem \mathcal{O} , which, if present, is always rectangular with more rows than columns. The columns of \mathbf{G}_{11} correspond to the unmatched unknowns, and the rows of \mathbf{G}_{44} correspond to the unmatched equations. It should be noted that although the Dulmage-Mendelsohn permutation is unique in terms of the members of \mathcal{U} , \mathcal{W} , and \mathcal{O} , the exact ordering of the rows \mathcal{U}_r , \mathcal{W}_r and \mathcal{O}_r as well as the columns \mathcal{U}_c , \mathcal{W}_c and \mathcal{O}_c may change within these three sub-graphs provided that they still return a valid Dulmage-Mendelsohn permutation. Since there may be more than one maximum matchings, selecting a different maximum matching may swap columns between \mathbf{G}_{11} and \mathbf{G}_{12} and rows between \mathbf{G}_{34} and \mathbf{G}_{44} . The Dulmage-Mendelsohn method yields the structural rank of the matrix that is the maximum possible numerical rank of the matrix based on its non-zero pattern.

The Dulmage-Mendelsohn permutation offers significant advantages in solving linear systems of the form,

$$\mathbf{G} \cdot \mathbf{x} = \mathbf{d}, \quad (2)$$

where \mathbf{x} is the $n \times 1$ vector with the unknowns to be determined, and \mathbf{d} is a $m \times 1$ vector with the observations that are related to \mathbf{x} through \mathbf{G} . For example, it allows division of the initial problem into smaller problems, which, depending upon the structure of \mathbf{G} , may significantly reduce the required computational resources needed. By applying the Dulmage-Mendelsohn permutation to Eq. 2, it can be converted as,

$$\mathbf{P} \cdot \mathbf{G} \cdot \mathbf{P}' \cdot \mathbf{x}' = \mathbf{d}', \quad (3)$$

where $\mathbf{x}' = \mathbf{P}'^{-1} \cdot \mathbf{x}$, and $\mathbf{d}' = \mathbf{P} \cdot \mathbf{d}$. It should be noted that \mathbf{P}'^{-1} and \mathbf{P}^{-1} always exist since \mathbf{P}' and \mathbf{P} are non-singular as a permuted version of the identity matrix. Using Eq 1, we can rewrite Eq. 3 in matrix form,

$$\begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \mathbf{G}_{13} & \mathbf{G}_{14} \\ 0 & 0 & \mathbf{G}_{23} & \mathbf{G}_{24} \\ 0 & 0 & 0 & \mathbf{G}_{34} \\ 0 & 0 & 0 & \mathbf{G}_{44} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \mathbf{x}'_3 \\ \mathbf{x}'_4 \end{bmatrix} = \begin{bmatrix} \mathbf{d}'_1 \\ \mathbf{d}'_2 \\ \mathbf{d}'_3 \\ \mathbf{d}'_4 \end{bmatrix}, \quad (4)$$

where the indices in \mathbf{x}'_j and \mathbf{d}'_i indicate the elements of these vectors that correspond to the j th column and the i th row of the block \mathbf{G}_{ij} , respectively. The general steps for solving Eq. 4, by taking advantage the Dulmage-Mendelsohn form are the following (e.g., Davis, 2006):

- I. Solve the subsystem $\begin{bmatrix} \mathbf{G}_{34} \\ \mathbf{G}_{44} \end{bmatrix} \cdot \mathbf{x}'_4 = \begin{bmatrix} \mathbf{d}'_3 \\ \mathbf{d}'_4 \end{bmatrix}$ to obtain $\hat{\mathbf{x}}'_4$ where $\hat{}$ denotes an estimation of the model vector.
- II. Solve $\mathbf{G}_{23} \cdot \mathbf{x}'_3 = \mathbf{d}'_2 - \mathbf{G}_{24} \cdot \mathbf{x}'_4$ to obtain $\hat{\mathbf{x}}'_3$.
- III. Solve $\begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \end{bmatrix} = \mathbf{d}'_1 - \mathbf{G}_{13} \cdot \hat{\mathbf{x}}'_3 - \mathbf{G}_{14} \cdot \hat{\mathbf{x}}'_4$ to obtain $\hat{\mathbf{x}}'_1$ and $\hat{\mathbf{x}}'_2$.

The structural properties of each subsystem, such as the block upper triangular form of \mathbf{G}_{23} , can be exploited by selecting the appropriate solving approach for efficiency (e.g., Davis, 2006). It should be noted that the subsystems are characterized from useful properties such as full structural rank and zero-free diagonal that are crucial in reducing the amount of work when using direct sparse solvers such as those based upon Cholesky, LU and QR factorizations. Furthermore, it ensures that the symbolic analysis, which precedes the factorization, and aims in reducing fill-ins as well as pre-allocating the data structures to hold the result of the numerical computation, is as effective as possible. Note that the Dulmage-Mendelsohn permutation is

meaningful only in the case when \mathbf{G} is sparse or can be converted into a sparse matrix so that it can be decomposed into the upper block-triangular form of Eq. 1. Also note that the concept of over-, well-, and under-determined subsystems should not be considered as a cheap, quantized version of the singular value decomposition where an ordered sequence of model-space vectors is derived based upon the singular values. In contrast, as it will be demonstrated in the next section, different blocks of the Dulmage-Mendelson permutation generally contain both large and small singular values, without any clear differentiation.

3. A simple synthetic example

To demonstrate the Dulmage-Mendelsohn analysis, and to explore its connections to other methods such as the singular value decomposition, we use a simple example of a 2-D seismic attenuation tomography problem (Fig. 3a). For simplicity, we assume that the velocity is constant throughout the model, therefore, the ray paths are straight (Fig. 3a). The forward operator of the problem, \mathbf{G} (Eq. 2; Fig. 3b) has 20 rows, equal to the number of the ray paths, and 16 columns, corresponding to each of the model blocks. Since each ray path samples only a subset of the model space, \mathbf{G} is a sparse matrix, i.e., a large number of its elements is 0, which is common for seismic tomography problems (e.g., Nolet, 2008). From the matrix \mathbf{G} , we can construct the associated bipartite graph \mathbb{G} by considering its non-zero elements (Fig. 3c), and then to perform the Dulmage-Mendelsohn permutation. The new permuted matrix is shown in Fig. 3d, and the associated bipartite graph \mathbb{G} in Fig. 3e. In this example, all blocks of the decomposition (i.e., every \mathbf{G}_{ij} of Eq. 4) are present.

The Dulmage-Mendelsohn decomposition shows that the number of edges of a maximum matching of \mathbb{G} is 14 (note that different maximum matchings will also have the same number of edges, which also states the structural rank r_s of the problem is 14. The parameters 10, 11, 12, 14, 15 and 16 make up the structurally under-determined part of the system, implying that, regardless the numerical values of the non-zero elements of \mathbf{G} , there is a null space that can produce an infinite number of solutions that fit the observations. The parameters 13 and 9 belong to the structurally well-determined subsystem, and the parameters 1 through 8 belong to the structurally over-determined subsystem. Because the structural rank poses only an upper bound to the numerical rank, it is impossible to conclude if the structurally well-determined and over-determined subsystems are of full numerical rank and well-posed problems without applying a more sophisticated analysis such as the singular value decomposition. To further explore this issue, we calculate and examine the singular values of different block combinations of the Dulmage-Mendelsohn permutation of \mathbf{G} (Fig. 4). The singular values of the full matrix \mathbf{G} show a large drop between the 13th and 14th singular value. The small singular values from the 14th onwards, are below the machine's precision, using a typical tolerance of $\max(m, n) \times \epsilon \times \sigma_{max}$ where ϵ is the machine's floating-point relative accuracy (e.g., of the order of 10^{-16} for a double precision machine), and σ_{max} is the largest singular value (e.g., Gander et al., 2014). This implies that the numerical rank r_n of the problem is 13, which is close to the structural rank $r_s = 14$ inferred from the maximum matching. It should be noted that small singular values occur even when the over-determined or the combination of the over- and well-determined subsystems are examined (Fig. 4). In both cases, the numerical rank is smaller than the structural column rank of the corresponding subsystem obtained from the Dulmage-Mendelsohn permutation. However, in each case, the condition number of the subsystems is about an order of

magnitude smaller than the condition number of the whole matrix. The condition number is calculated as the ratio of the largest over the smallest singular value, after omitting the tiny singular values based upon the tolerance described earlier. The under-determined subsystem, with structural column rank 4, happens to have 4 significant singular values too, but its condition number is similar to the condition number of the original system.

Using the singular value decomposition and the tolerance described above, the orthonormal basis for the null space of \mathbf{G} can be obtained. The three (i.e., $n - r_n = 3$) n -dimensional vectors \mathbf{n}_i are shown in Fig. 5. Any linear combination of these vectors lies in the null space of \mathbf{G} . With the exception of the minimum norm solution that does not include components from the null space of the problem, most regularized solutions commonly used in geophysical tomography (e.g., Aster et al., 2005; Nolet, 2008), are characterized by arbitrary components in the null space that alter the solution, but not the misfit. Therefore, being able to determine the null space is particularly useful for evaluating the robustness of the solution (e.g., Deal & Nolet, 1996; de Wit et al., 2012). In the case of our synthetic example, the parameters corresponding to the structurally under-determined part of the system present significant components in the null space of \mathbf{G} (i.e., large absolute values for the elements corresponding to these parameters in the null-space vectors shown in Fig. 5), indicating that they are not adequately constrained. This is also the case for some parameters that belong to the structurally well- and over-determined parts of \mathbf{G} reflecting the linear dependence of the rows. Consequently, the practicality of the permutation as a diagnostic tool is limited for floating point arithmetic problems that involve errors, noisy data and ill-conditioned matrices, such as those in geophysical tomography problems.

In order to investigate the accuracy of the Dulmage-Mendelsohn decomposition in solving a system of linear equations against typical methods, we generate a synthetic data set based upon the model shown in Fig. 6a (i.e., the true model) as $\mathbf{d}_{syn} = \mathbf{G} \cdot \mathbf{m}_{real}$, and add white Gaussian noise with a signal-to-noise ratio of 2dB. Using the synthetic data, we seek the minimum norm solution $\hat{\mathbf{m}}$ with the Moore (1920) and Penrose (1955) pseudo-inverse operator, with the tolerance as discussed previously. One approach is to utilize three submatrices obtained through the Dulmage-Mendelsohn permutation and follow the procedures described in Section 2. For comparison, the full system is also solved by applying the Moore-Penrose pseudo-inverse operator of the whole matrix \mathbf{G} . The results are identical within the machine's precision limits (Fig. 6). This is to be expected as the Dulmage-Mendelsohn approach constitutes simply a blocked version of backward substitution in solving upper triangular systems.

4. Application to a real seismic tomography problem

We apply the Dulmage-Mendelsohn technique to a seismic tomography problem where a regional model of P-wave variations is sought using travel-time observations. We use 87,616 P-wave travel times recorded at the Hi-net station (Okada et al., 2004) from 125 local earthquakes of $M_w > 4.4$, with depth greater than 150 km (Fig. 7). The picking is performed automatically using the algorithm presented in Bogiatzis & Ishii (2015). Arrivals with signal-to-noise ratio greater than 2 are considered, and the weighted standard deviation of the picks along different scales of the continuous wavelet transform is used as the measure of uncertainty of the picks.

The three-dimensional space spans from 125°E to 152°E in longitude, 20°N to 50°N in latitude, and from 4 km to -889.5 km elevation, where positive values are above sea level. It

is parameterized using a grid of 234,000 nodes with $90 \times 100 \times 26$ nodes along latitude, longitude and depth, respectively. The spacing along the latitude and longitude directions is constant, equal to 0.303° , while the spacing in depth varies smoothly using a tapered cosine function. The depth spacing is about 30 km at 11 km depth and increases gradually to 60 km at the deepest layer of 313 km depth. For depths shallower than 11km, there are three layers at - 5km, 0km and 4km to accommodate topography and near station conditions.

4.1. Forward Problem

The ray paths required for determining the P-wave travel times are calculated using the high frequency approximation of the elastic wave equation in two steps. First, Dijkstra's algorithm (e.g., Dijkstra, 1959; Moser, 1991; Bogiatzis, 2010) is applied to compute a first estimation of the ray path, in the sense of the fastest path that connects the source and the station following the Fermat's principle (e.g., Bóna & Slawinski, 2003; Červený, 2005). The ray path is constructed from linear segments that connect different nodes of the model, and the time delay along each segment is approximated by considering its length and the average of the velocities between the nodes it connects. The connectivity stencil is formed by the 27 closest nodes, except for those near the edges of the model, where the number is reduced accordingly. Next, the initial ray path is further optimized by means of the pseudo-bending method following the Fermat's principle (e.g., Moser et al., 1992; Papazachos & Nolet, 1997; Bogiatzis, 2010), and using Polak-Ribiere non-linear conjugate-gradient method (Press et al., 1997). To avoid being trapped into local minima and/or regions where the gradient of the travel time functional is relatively flat, a multi-start approach is followed. After optimizing the path generated from the Dijkstra's method, two more paths are considered as starting points. One is a straight path in the Cartesian space that connects the source and the station, and the other is the average of the straight path with the

output of the first optimization run that started from Dijkstra's algorithm path. Furthermore, the optimization algorithm for each starting path is repeated 3 times using different perturbation levels to compute the partial derivatives of the travel-time functional with respect to the location of the ray path points. The ray optimization problem is solved in local, vertical, Cartesian coordinates with the origin set to a point within the model, at $(latitude, longitude, elevation) = (35.358^{\circ}, 138.731^{\circ}, 3.776 \text{ km})$. The velocity between the nodes of the medium is interpolated by means of tri-linear interpolation utilizing the 8 surrounding nodes in the geodetic domain (i.e., latitude, longitude, elevation), rather than the curved Cartesian domain, under the assumption that within this small region between adjacent nodes, the effect of curvature is negligible. Both steps of the ray path determination are parallelized. In the case of Dijkstra's algorithm, each earthquake is assigned to a different processor, while at the second step, the total number of ray paths are divided into N subsets of approximately equal number of rays, where N is the number of processing units.

4.2. Inverse problem

We setup a nonlinear inverse problem to simultaneously determine the P-wave slowness at the nodes and the hypocentral parameters (i.e., the three spatial coordinates and the origin time). The arrival time $T(\mathbf{m})$ is a function of \mathbf{m} which is the model vector where the first 234,000 elements correspond to the slowness nodes of the model, and the last 500 elements correspond to the earthquake parameters. $T(\mathbf{m})$ is then expanded about a starting model \mathbf{m}_0 that is constructed from a one-dimensional Earth model, IASP91 (Kennett & Engdahl, 1991), and the hypocentral parameters from the Japan Meteorological Agency catalogue (JMA; <http://www.jma.go.jp>; last accessed on February 2018). The deviation of the observed travel time from those predicted using a starting model is related to model update, i.e.,

$$T(\mathbf{m}) - T(\mathbf{m}_0) = \left. \frac{\partial T}{\partial \mathbf{m}} \right|_{\mathbf{m}_0} \cdot d\mathbf{m} + h, \quad (5)$$

where h corresponds to higher order terms that are neglected under the assumption that T can be approximated as a linear function about \mathbf{m}_0 . After ignoring h , Eq. 5 has the form of Eq. 2 with $d \equiv T(\mathbf{m}) - T(\mathbf{m}_0)$, $\mathbf{G} \equiv \left. \frac{\partial T}{\partial \mathbf{m}} \right|_{\mathbf{m}_0}$ and $\mathbf{x} \equiv d\mathbf{m}$. The partial derivatives matrix \mathbf{G} has the structure

$$\mathbf{G} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{H}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{S}_2 & \mathbf{0} & \mathbf{H}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_N & \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_N \end{bmatrix}, \quad (6)$$

where \mathbf{S}_i is a $M_i \times n$ submatrix containing the partial derivatives of the slowness for the M_i ray paths associated with the i^{th} event. \mathbf{H}_i is $M_i \times 4$ submatrix containing the partial derivatives with respect of the three spatial coordinates and the origin time, of the i^{th} event.

The non-linearity of the problem is incorporated through calculation and application of successive updates to the model vector, until no further considerable improvements in misfit can be made.

4.3. Analysis of the sensitivity matrix with the Dulmage-Mendelsohn permutation

At each iteration, the Dulmage-Mendelsohn permutation of the \mathbf{G} matrix is calculated, and solutions are obtained. For example, in the first iteration, \mathbf{G} is a sparse matrix with only 0.9% of non-zero elements (Fig. 8). The Dulmage-Mendelsohn permutation reveals that all blocks \mathbf{G}_{ij} are present, and that the structural rank of \mathbf{G} is 39,015 which provides an upper limit on its numerical rank (i.e., number of recoverable model parameters in model space coordinates). The structurally under-determined subsystem has dimensions $7,019 \times 202,504$ and the dimensions of the structurally over-determined subsystem are $80,092 \times 31,491$. The submatrix \mathbf{G}_{23} that

corresponds to the structurally well-determined subsystem is substantially smaller compared to any of the \mathbf{G}_{ij} submatrices, with dimensions of 505×505 . For simplicity, we combine the well-determined and over-determined submatrices into 80597×31996 matrix, and thus merge the first two steps of the solution process (see section 2). Compared to the initial \mathbf{G} matrix, the new submatrix contains 92% of the rows (i.e., observations), but only 14% of the columns (i.e., unknown parameters). The spatial distribution of the model parameters that belong to the well/over-determined system, is consistent with the regions of good ray path coverage (Fig. 9).

4.4. Calculation of the singular values

The singular value decomposition (e.g., Lanczos, 1961; Golub & Reinsch, 1970; Demmel, 1997) of the sensitivity kernel is a crucial diagnostic tool about the nature of the inverse problem. It also provides the most effective tool for solving ill-posed problems (e.g., Hansen, 1990, 1992, 1998, 2007; Demmel, 1997; Snieder & Trampert, 1999; Aster et al., 2005). However, as data and model spaces become larger, the computational cost of singular value decomposition in floating point operations, and in particular, memory requirements, makes it intractable. Over the years, several algorithms have been developed to extend the applicability of singular value analysis, or some of its features to larger problems. For example, iterative methods such as SVDPACK (Berry, 1992) and PROPACK (Larsen, 1998) are very fast when only a small, *a-priori* fixed number of the singular values of large matrices is needed. More recently, algorithms based upon randomized sampling have been used to derive partial and approximate singular value decompositions (e.g., Drineas et al., 2006; Halko et al., 2010; Voronin et al., 2014). In all these methods, only a small subset of the largest singular values is returned, and the user must *a-priori* define this number.

We have chosen the PIRO_SKY algorithm (PIpelined plane ROtations of sparse SKYline matrices; Rajamanickam, 2009) to calculate the singular values of the seismic tomography problem. This algorithm uses SPQR (Davis, 2011; also see next paragraph) to compute the QR factorization and reduces the sparse upper triangular \mathbf{R} matrix to bidiagonal form. The bidiagonal matrix is further reduced to a diagonal form through LAPACK (Anderson et al., 1999), which contains the singular values of the original matrix. Among the advantages of this method is the efficient calculation of all singular values of the matrix, and it can be applied to large sparse matrices such as one considered for this study. All the singular values for both the well/over-determined and under-determined subsystems are computed through this approach. The total computation time using a node on Harvard's Odyssey supercomputer with 32 CPUs and ~200 GB of memory for the well/over-determined system was about 4.5 days, with the most time-consuming part (4.4 days) being the bi-diagonalization of the \mathbf{R} matrix returned from the QR decomposition. The under-determined system took a few seconds as there are only 7019 rows. The results from both subsystems show gradual decay of the singular values (Fig. 10). In the case of the well/over-determined part, the singular values decay gradually and smoothly towards zero. The under-determined part of the system exhibits similar behavior, but there are a few abrupt slope changes. The gradual decay of singular values is a known property of discrete ill-posed problems, and in such cases, not only the determination, but the definition of the rank of the problem is ambiguous (e.g., Hansen, 2007). The singular values curve gives a rough estimate of the low limit for the regularization parameter as the point where the singular values start decreasing rapidly which correspond to singular values between ~50 and ~100 for the well/over-determined part and ~100 for the under-determined subsystem. (Fig. 10). It should be noted that the computation of the singular values is a completely independent task, and it is not required for

the calculation of the model vector, the model resolution or the posterior model covariance matrices.

4.5. Calculation of the Solution

Inversion of both under-determined and well/over-determined subsystems are ill-posed problems, although the condition number of the well/over-determined submatrix is often smaller compared to the under-determined subsystem. Consequently, regularization is required to yield meaningful solutions. By assuming Tikhonov type regularization operators (Tikhonov & Arsenin, 1977), this corresponds to solving

$$\begin{bmatrix} \mathbf{G}_{23} & \mathbf{G}_{24} \\ \mathbf{0} & \mathbf{G}_{34} \\ \mathbf{0} & \mathbf{G}_{44} \\ & \mathbf{D}_O \\ & \mathbf{L}_{OO} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}'_3 \\ \mathbf{x}'_4 \end{bmatrix} = \begin{bmatrix} \mathbf{d}'_3 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad (7a)$$

followed by

$$\begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{D}_U \\ \mathbf{L}_{UU} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \end{bmatrix} = \left\{ \begin{array}{l} \mathbf{d}'_1 - \mathbf{G}_{13} \cdot \hat{\mathbf{x}}'_3 - \mathbf{G}_{13} \cdot \hat{\mathbf{x}}'_4 \\ \mathbf{0} \\ -\mathbf{L}_{UO} \cdot \begin{bmatrix} \mathbf{x}'_3 \\ \mathbf{x}'_4 \end{bmatrix} \end{array} \right\}, \quad (7b)$$

where \mathbf{D}_O and \mathbf{D}_U are the norm damping operators for the well/over-determined and under-determined systems, respectively. \mathbf{L}_{OO} , \mathbf{L}_{UU} and \mathbf{L}_{UO} are finite-difference, smoothing operators that implement the second order Tikhonov regularization (i.e., Laplacian; Tikhonov & Arsenin, 1977). Initially, the smoothing operator of the whole model is calculated as

$$\mathbf{L} = \mathbf{A} - \mathbf{D}, \quad (8)$$

where \mathbf{A} is the adjacency matrix of the grid describing the connectivity from the node i to the node j . More specifically, \mathbf{A}_{ij} is the distance between nodes i and j , if these nodes are connected,

or null otherwise. \mathbf{A} has been formed based upon a connectivity stencil of 27 neighboring grid points, in an equivalent manner to the matrix used for calculating the shortest paths in the forward problem. \mathbf{D} is the degree matrix, i.e., a diagonal matrix with $\mathbf{D}_{ii} = \sum_{j=1}^n \mathbf{A}_{ij}$. The matrices \mathbf{L}_{OO} and \mathbf{L}_{UU} are created by considering the rows and columns that correspond to well/over-determined and under-determined parameters, respectively, and \mathbf{L}_{UO} is obtained by considering the rows of the under-determined matrix and the columns of the well/over-determined matrix. Note that \mathbf{L}_{UO} in Eq. (7b) incorporates the contribution of the model estimate obtained from the previous step of solving the well/over-determined problem. If only the norm damping is chosen to regularize the inversion, then the bottom row in Eq. (7a) and Eq. (7b) is omitted.

The Dulmage-Mendelsohn decomposition into smaller submatrices makes the application of direct methods in solving Eq. (7a) and Eq. (7b) a tractable problem. We utilize the SPQR algorithm (Davis, 2011), which is a high-performance, parallel, multifrontal sparse solver based upon the QR factorization method (Duff & Reid, 1983). It breaks the factorization problem of large sparse matrices into many small dense submatrices using a dendritic organization that is suitable for parallelism. Furthermore, to reduce fill-ins, it includes a symbolic analysis phase that considers only the non-zero pattern of the matrix and tests the effectiveness of several fill-in ordering algorithms (e.g., Karypis & Kumar, 1998; Davis et al., 2004a,b; Amestoy et al., 2004) to select the one with the best performance. The applicability of this algorithm to seismic tomography problems has been demonstrated by Bogiatzis et al. (2016). The average computation time is slightly more than an hour for one inversion requiring less than 24 GB of memory. The algorithm is used here to produce several solutions with different types and strengths of regularization in order to define the optimum regularization by means of the L-curve

approach (e.g., Hansen, 2007; Aster et al., 2005). Alternatively, LSQR or similar iterative techniques can be used in solving Eq. (7), and they can accelerate the L-curve/surface construction. We determine that the optimum values correspond to damping weight close to 300 and smoothing weight close to 20. Furthermore, one of the advantages of using a direct method, especially after the optimum regularization parameters have been determined, is the ability to calculate the model resolution and the posterior model covariance matrices (e.g., Bogiatzis et al., 2016). The breakdown of the problem into two significantly smaller matrices with Dulmage-Mendelsohn permutation further improves the computation of the solution vector and the model resolution matrix by limiting both the symbolic analysis and factorization only once as opposed to a more involved approach described by Bogiatzis et al. (2016).

4.6. Results

The final solution of regional seismic tomography model of Japan is obtained after 6 iterations, keeping the regularization values constant to those determined in the previous section. Both misfit in the sense of the root mean square error as well as the shape of the error distribution improve substantially with each iteration (Fig. 11). The corrections of the earthquake hypocentral locations and origin times are on the order of tens of km for the location and up to a few seconds for the origin time (Fig. 12). The most prominent feature in the model is a linear high-velocity anomaly observed along the trench that dips to the west, corresponding to the cold subducting Pacific plate (Figs. 13 and 14). Our results are consistent with several previous regional and global studies of the region (e.g., Stern, 2002; Gorbatov & Kennett, 2003; Kennett & Furumura, 2010; Koulakov, 2011; Wei et al., 2012, 2015; Zhao et al., 1992; 1994; 2012; Liu & Zhao, 2016). The thickness of the Pacific plate is about 100 km. The upper boundary is imaged clearly almost everywhere, but the lower boundary of the slab becomes more

complicated at latitudes higher than $\sim 37^\circ$ and at depths greater than 300 km (Figs 13 and 14). Another high-velocity anomaly with narrower width and weaker amplitude exists from Kyushu through Shikoku to Kansai, and this can be attributed to the Philippine plate subduction. It presents a complex structure and remains visible down to at least 300 km, which agrees with other recent results (e.g., Liu & Zhao, 2016). An interesting feature is the upwards bending of the slab at the Japan Trench and depth of about 100-120 km (e.g., Fig 13d and e).

In contrast, the mantle wedge beneath the back-arc areas above the subducting Pacific plate include smaller-scale low-velocity anomalies. The surficial extent of these low velocity bodies reaches the active volcanoes (Figs. 13 and 14), suggesting that the low velocity anomalies may be the sources of arc magmatism and volcanism. These features are consistent with mineralogical and geodynamical results (e.g., Stern, 2002; Hacker et al., 2003) and previous tomographic studies (e.g., Zhao et al., 1992; 2012; Hasegawa et al., 2013; Liu & Zhao 2016) showing that the combination of corner flow in the mantle wedge (bringing hot asthenosphere from depths) and release of volatiles from the slab (lowering the melting temperature) lead to the presence of partial melts and lower seismic speed.

Another prominent low velocity region occurs beneath the subducting Pacific slab, close to the bending axis of the Pacific lithosphere (Figs. 13 and 14). This feature becomes visible below the depth of 170 km, initially at latitudes greater than 35° and spreads along the entire Japan trench. A similar anomaly has been reported in Cascadia subduction zone and has been associated with the accretion of a thin asthenospheric layer that resists subduction due to its buoyancy (Hawley et al., 2016). It has also been argued that similar structures can be artifacts due to the isotropic velocity assumption in the presence of significant anisotropy (Bezada et al., 2016).

To assess the robustness of the model, the model resolution matrix (Fig. 15) and the posterior model covariance matrix (Fig. 16) are calculated as part of the last iteration. Instead of calculating the resolution and covariance for all model parameters, only the well- and over-determined parameters are examined based upon the Dulmage-Mendelsohn decomposition. This substantially reduced the required computational resources and the overall computational time. For example, it would be impossible to calculate the full model resolution matrix, and in particular, the model covariance matrix, which is typically a dense matrix requiring more than 200 GB of memory assuming only the upper triangular part is considered. With only the well- and over-determined components, it required less than 24 GB of memory and about one day to calculate both matrices. Regions with good ray path coverage are associated with better resolution and lower variances (Figs. 15 and 16).

We choose not to interpret features that are in regions associated with the under-determined parameters even though similarities with previous studies may be noted. Such features are generally outside the resolving capability of our tomographic data and they cannot be robustly determined.

5. Discussion and conclusions

We show that through the Dulmage-Mendelsohn permutation, a matrix with different rank deficiencies can be quickly decomposed, allowing the separation of the structurally under-determined and over-determined parts of the system. It should be noted that the Dulmage-Mendelsohn technique is applicable only in the case where the matrix to be inverted is sparse. This translates to tomographic problems where the model is parameterized with local basis

functions such as grid points or cells, as opposed to global parameterizations such as spherical harmonics. Similarly, the observations should have localized sensitivity kernels as opposed to data such as normal-mode eigenfrequencies, where the sensitivity is distributed across the whole model. However, even in such cases, the initial matrix can be transformed to a sparse matrix using various sparse representation methods such as the wavelet basis functions (e.g., Simons et al., 2011; Voronin et al., 2015).

Among the advantages of the Dulmage-Mendelsohn approach is a fast dimensionality reduction of the initial problem, which, in many cases, makes the application of direct solvers feasible. Nevertheless, in cases of very large problems with millions of data and/or unknowns the application of such solvers, depending the structure of the initial matrix, may still require unaffordable large amounts of memory. In such problems, techniques such as those presented by Voronin et al., (2014; 2015) can be used to compress and break the initial matrix to smaller blocks, before the application of the method presented in this manuscript.

With direct methods, the model uncertainty can be evaluated with the model resolution matrix and the model posterior covariance matrix. It should be noted that the ability to calculate the model posterior covariance matrix, is a first step towards putting error bars in moderate size tomographic models, and thus allowing more accurate quantitative interpretation of the tomographic models. This further requires realistically determining the prior data errors and model covariance (e.g., Tarantola, 1987; Nolet et al., 1999; Rawlinson et al., 2014), and recently, significant effort has been focused on improving the determination of these priors (e.g., Bodin et al., 2012; Duputel et al., 2012, 2014; Rodi & Myers, 2013; Voronin et al., 2014; Ballard et al., 2016). Furthermore, we show that the matrix decomposition, in combination with a recently developed singular value decomposition algorithm, allow the computation of the entire range of

singular values of both the well/over-determined and the under-determined subsystems giving insight into the problem. For example, it reveals the exact nature of the gradual decay of the singular values, which is of considerable interest for the efficient and accurate dimensionality reduction of the problem by means of low rank approximations (e.g., Voronin et al., 2014, 2015 Gu 2015).

Further study of the singular values can also provide insight into the required regularization and define a small region within which the L-curve analysis can be performed. In this manuscript we choose not to compute the orthogonal matrices that describe the row space and column space as they are typically dense matrices, requiring significant additional memory. However, one useful future direction is towards the modification of the PIRO_SKY algorithm (Rajamanickam, 2009) to directly calculate the Fourier coefficients $|u_i^T * \mathbf{d}|$ where u_i is the left singular vector and \mathbf{d} is the right-hand side of Eq. (2). This will allow the evaluation of the discrete Picard condition, and provide a more definitive way to regularize the solution.

The successful factorization of \mathbf{G} , or at least of its well/over-determined portion, as well as the efficient computation of the resolution matrix and the singular values of the problem can be useful in the application of other uncertainty quantification approaches including the null space shuttle (Deal & Nolet, 1996; de Wit et al., 2012), the extremal bounds analysis (e.g., Meju, 2009), the Lie group techniques (Vasco, 2007) and the sensitivity tests (e.g., Rawlinson & Spakman, 2016). For example, provided that the resolution matrix is available, sensitivity tests can be performed very efficiently as they are just a multiplication between a sparse matrix (resolution matrix) and the model vector. Therefore, a large number of tests such as checkerboard tests, and arbitrary realizations of the model space can be performed, and then

statistics of the posterior variability of the model parameters can be used to distinguish robust from non-robust features of the solution.

Finally, another future application of the Dulmage-Mendelsohn method is to accelerate the recently proposed formulation of the Backus-Gilbert inversion by Zoroli (2016), which takes advantage of the Subtractive Optimally Localized Averages (SOLA) technique. In this approach a system of linear equations is solved for each parameter of the model, using the LSQR solver. The Dulmage-Mendelsohn permutation can be used to break the problem into smaller ones, and make the matrix factorization affordable, using for instance the SPQR solver. Since the matrix remains the same, the analysis and the factorization is required only once, thus, the solutions for each of the unknown parameters can be retrieved rapidly, using backward substitution. The system considered by Zoroli (2016) contains 79,765 data and 38,125 unknown parameters, which is within the capabilities of the method presented in this manuscript.

Acknowledgements

The authors would like to gratefully acknowledge the editor, Lapo Boschi, Guust Nolet and one anonymous reviewer for their critical reviews and constructive comments that helped to improve the manuscript.

References

- Amante, C. and 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 on January 2017.
- Amestoy, P. R., Davis, T. A. and 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.
- Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and Sorensen. D., 1999. LAPACK Users' Guide. 3rd ed. Philadelphia: Soc. for Industrial and Applied Mathematics.
- Aster, R., Borchers, B. and Thurber, C., 2005. Parameter Estimation and Inverse Problems, Elsevier Academic.
- Backus, G. and Gilbert, F., 1968. The Resolving Power of Gross Earth Data, *Geophys. J. R. Astron. Soc.*, 16: 169-205, doi:10.1111/j.1365-246X.1968.tb00216.x.
- Ballard, S., Hipp, J. R., Begnaud, M. L., Young, C. J., Encarnacao, A. V., Chael, E. P. and Phillips, W. S., 2016. SALSA3D: A Tomographic Model of Compressional Wave Slowness in the Earth's Mantle for Improved Travel-Time Prediction and Travel-Time Prediction Uncertainty, *Bull. Seismol. Soc. Am.* 106(6), 2900 - 2916, doi: 10.1785/0120150271.
- Bekas, C., E. Kokiopoulou, and Y. Saad, 2007. An Estimator for the Diagonal of a Matrix, *Appl. Numer. Math.* 57(11-12), 1214-1229.

- Berry, M., 1992. Large Scale Singular Value Computations, *Int. J. High Perform. Comput. Appl.* 6(1), 13-49.
- 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.
- Bezada, M. J., Faccenda, M. and Toomey D. R., 2016. Representing anisotropic subduction zones with isotropic velocity models: A characterization of the problem and some steps on a possible path forward, *Geochem. Geophys. Geosyst.* 17, 3164–3189, doi: 10.1002/2016GC006507.
- Bijwaard, H., and Spakman, W., 2000. Non-linear global P-wave tomography by iterated linearized inversion, *Geophys. J. Int.* 141, 71–82, doi: 10.1046/j.1365-246X.2000.00053.x.
- Bodin, T., M. Sambridge, H. Tkalčić, P. Arroucau, K. Gallagher, and N. Rawlinson, 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. Geophys. Res.* 117, B02301, doi:10.1029/2011JB008560.
- Bogiatzis, P., 2010. Contribution to Joint Tomography of Different Types of Seismic Data, Ph.D. Thesis, Aristotle University of Thessaloniki, Thessaloniki, Greece, pp. 228, doi:10.12681/eadd/22059.
- Bogiatzis, P. and Ishii, M., 2015. Continuous wavelet decomposition algorithms for automatic detection of compressional- and shear-wave arrival times. *Bull. Seismol. Soc. Am.* 105(3), doi: 10.1785/0120140267.

- Bogiatzis, P., Ishii, M. and Davis, T.A., 2016. Toward using direct methods in seismic tomography: Computation of the full resolution matrix using high performance computing and sparse QR factorization, *Geophys. J. Int.* 205(2), 830 - 836, doi: 10.1093/gji/ggw052.
- Bóna, A. and Slawinski, M. A., 2003. Fermat's principle for seismic rays in elastic media, *J. Appl. Geophys.* 54, 445-451, doi: 10.1016/j.jappgeo.2003.08.019.
- Boschi, L., 2003. Measures of resolution in global body wave tomography, *Geophys. Res. Lett.* 30, 1978, doi:10.1029/2003GL018222.
- Červený, V., 2005. *Seismic Ray Theory*, Cambridge Univ. Press, New York, NY, pp 724.
- D' Azevedo, E. and Dongarra, L., 2000. The design and implementation of the parallel out-of-core ScaLAPACK LU, QR, and Cholesky factorization routines, *Concurrency Comput. Pract.* 12(15), 1481-1493.
- Davis, T. A., 2006. *Direct Methods for Sparse Linear Systems*, SIAM, Philadelphia.
- 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., and Hu, Y., 2011. The University of Florida sparse matrix collection, Timothy A. Davis, Yifan Hu, *ACM Trans. Math. Softw.* 38, 1:1 - 1:25.
- Davis, T. A., J. R. Gilbert, S. I. Larimore, and E. G. Ng, 2004a. Algorithm 836: COLAMD, A Column Approximate Minimum Degree Ordering Algorithm, *ACM Trans. Math. Softw.* 30 (3), 377-380.

- Davis, T. A., J. R. Gilbert, S. I. Larimore, and E. G. Ng, 2004b. A Column Approximate Minimum Degree Ordering Algorithm, *ACM Trans. Math. Softw.* 30 (3), 353-376.
- Deal, M. M. and Nolet, G., 1996. Nullspace shuttles, *Geophys. J. Int.* 124, 372–380.
- de Wit, R.W.L., Trampert, J. and van der Hilst, R.D., 2012. Toward quantifying uncertainty in travel time tomography using the null-space shuttle, *J. Geophys. Res.* 117, B03301, doi:10.1029/2011JB008754.
- Demmel, J. W., 1997. Applied Numerical Linear Algebra Soc. for Industrial and Applied Mathematics, Philadelphia, 101-138, doi: 10.1137/1.9781611971446.ch3.
- Dijkstra, E. W., 1959. A note on two problems in connection with graphs, *Numerische Mathematik* 1, 269–271. doi:10.1007/BF01386390.
- Drineas, P., Kannan, R. and Mahoney, M.W., 2006. Fast Monte Carlo Algorithms for Matrices II: Computing a Low-Rank Approximation to a Matrix. *SIAM J of Computing* 36(1):158–183.
- Duff, I. S. and J. K. Reid, 1983. The Multifrontal Solution of Indefinite Sparse Symmetric Linear Systems, *ACM Trans. Math. Softw.* 9, 302 -325, doi:10.1145/356044.356047.
- Dulmage, A. L. and Mendelsohn, N. S., 1958. Coverings of bipartite graphs, *Canad. J. Math.* 10, 517–534, doi:10.4153/cjm-1958-052-0.
- Dulmage, A. L. and Mendelsohn, N. S., 1959. A structure theory of bipartite graphs of finite exterior dimension, *Trans. Roy. Soc. Canada* 53, 1–13.
- Dulmage, A. L. and Mendelsohn, N. S., 1967. Graph Theory and Theoretical Physics: Graphs and matrices, 167–277, Academic Press, London.

- Duputel, Z., Rivera, R., Fukahata, Y. and Kanamori, H., 2012. Uncertainty estimations for seismic source inversions, *Geophys. J. Int.* 190, 1243-1256.
- Duputel, Z., Agram, P. S., Simons, M., Minson, S. E. and Beck, J. L., 2014. Accounting for prediction uncertainty when inferring subsurface fault slip, *Geophys. J. Int.* 197, 464–482.
- Dziewoński, A. M., Hager, B. H. and O’Connell, R. J., 1977. Large-scale heterogeneities in the lower mantle, *J. Geophys. Res.*, 82, 239–255.
- Ellsworth W. L., 1977. Three-Dimensional Structure of the Crust and Mantle Beneath the Island of Hawaii. PhD Thesis, Massachusetts Institute of Technology, Cambridge.
- Fichtner, A. and J. Trampert, 2011. Resolution analysis in full waveform inversion, *Geophys. J. Int.* 187: 1604–1624. doi: 10.1111/j.1365-246X.2011.05218.x.
- Fichtner, A. and T. Leeuwen, 2015. Resolution analysis by random probing, *J. Geophys. Res. Solid Earth* 120, 5549–5573, doi:10.1002/2015JB012106.
- Gander, W., Gander, M.J. and Kwok, F., 2014. Scientific Computing - An Introduction using Maple and MATLAB, Springer, New York, NY.
- Global Volcanism Program, 2013. Volcanoes of the World, v. 4.6.0. Venzke, E (ed.). Smithsonian Institution. Downloaded 22 Jun 2017, doi 10.5479/si.GVP.VOTW4-2013.
- Gu, M., 2015. Subspace iteration randomization and singular value problems. *SIAM J. Scientific Computing*, 37(3)..
- Golub, G. H. and Reinsch, C., 1970. Singular value decomposition and least squares solutions, *Numerische Mathematik*, 14 (5): 403–420. doi:10.1007/BF02163027.

- Gorbatov A. and Kennett, B., 2003. Joint bulk-sound and shear tomography for Western Pacific subduction zones, *Earth Planet. Sci. Lett.* 210, 527-543
- Hacker, B.R., Abers, G.A. and Peacock, S.M., 2003. Subduction Factory 1. Theoretical mineralogy, densities, seismic wave speeds, and H₂O contents, *J. Geophys. Res.* 108, 1-25, doi: 10.1029/2001JB001127.
- Halko, N., Martinsson, P. G. and Tropp, J. A., 2011. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Rev.* 53, 217-288.
- Hansen, P. C., 1990. Relations Between SVD and GSVD of Discrete Regularization Problems in Standard and General Form, *Linear Algebra Appl.* 141, 165-176.
- Hansen, P. C., 1992. Analysis of Discrete Ill-posed Problems by Means of the L-curve, *SIAM Rev.* 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, *Numer. Algorithms* 46, 189-194.
- Hasegawa, A., Nakajima, J., Yanada, T., Uchida, N., Okada, T., Zhao, D., Matsuzawa, T. and Umino, N., 2013. Complex slab structure and arc magmatism beneath the Japanese Islands, *J. Asian Earth Sci.* 78, 277-290, doi: 10.1016/j.jseas.2012.12.031.
- Hawley W. B., Allen, R. M. and Richards, M. A., 2016. Tomography reveals buoyant asthenosphere accumulating beneath the Juan de Fuca plate, *Science* 353(6306), 1406 – 1408.

- Hipp, J. R., Encarnacao, A. V., Young, C. J., Ballard, S., Chang, M. C., Phillips, W. S. and Begnaud, M. L., 2011. Calculating a path-dependent travel time prediction variance and covariance for a global tomographic P - velocity model, *Proc. of the 2011 Monitoring Research Review: Ground -Based Nuclear Explosion Monitoring Technologies, Orlando, Florida*, 283–291
- Hopcroft, J. E. and Karp, R. M., 1973. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs, *SIAM Journal on Computing* 2 (4), 225–231, doi:10.1137/0202019.
- Hutchinson, M., 1990. A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines, *Commun. Stat. – Sumul. Comput.* 19, 433–450.
- Karypis, G. and V. Kumar, 1998. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs, *SIAM J. Sci. Comput.* 20, 359–392.
- Kennett, B. L. N. and Engdahl, E. R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.*, 105, S. 429–465, doi:10.1111/j.1365-246X.1991.tb06724.x.
- Kennett, B. L. N. and Furumura, T., 2010. Tears or thinning? Subduction structures in the Pacific plate beneath the Japanese Islands, *Phys. Earth Planet. Inter.* 180, 52–58, doi: 10.1016/j.pepi.2010.03.001.
- Kissling, E., 1988. Geotomography with local earthquake data, *Rev. Geophys.* 26, 659–698.
- Koulakov I., 2011. High-frequency P and S velocity anomalies in the upper mantle beneath Asia from inversion of worldwide traveltime data, *J. Geophys. Res.* 116, B04301, doi: 10.1029/2010JB007938.
- Lanczos, C., 1961. *Linear Differential Operators*, Van Nostrand, New York.

- Larsen, R. M., 1998. Lanczos bidiagonalization with partial reorthogonalization, <http://soi.stanford.edu/~rmunk/PROPACK/>, last accessed on April, 2018.
- Liu, X. and Zhao, D., 2016. P and S wave tomography of Japan subduction zone from joint inversions of local and teleseismic travel times and surface-wave data, *Phys. Earth Planet.* 252, 1-22, doi: 10.1016/j.pepi.2016.01.002.
- MacCarthy, J. K., Brochers, B. and 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.
- Meju, M.A., 2009. Regularized extremal bounds analysis (REBA): an approach to quantifying uncertainty in nonlinear geophysical inverse problems, *Geophys. Res. Lett.* 36, L03304, doi:10.1029/2008GL036407.
- Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press.
- Mercerat, D. E., and Nolet, G., 2013. On the linearity of cross-correlation delay times in finite-frequency tomography, *Geophys. J. Int.* 192, 681–687, doi: 10.1093/gji/ggs017.
- Moore, E. H., 1920. On the reciprocal of the general algebraic matrix, *Bull. Amer. Math. Soc.* 26, 394–395. doi:10.1090/S0002-9904-1920-03322-7
- Moser T. J., 1991. Shortest path calculation of seismic rays, *Geophys.* 56, 1, 59-67.
- Moser, T. J., Nolet G. and Snieder R., 1992. Ray bending revisited, *Bull. Seismol. Soc. Am.* 82, 259 – 288.
- Nolet, G., R. Montelli, and J. Virieux, 1999. Explicit, Approximate Expressions for the Resolution and a Posteriori Covariance of Massive Tomographic Systems, *Geophys. J. Int.* 138(1), 36–44.

- Nolet, G., 2008. *A Breviary of Seismic Tomography*, Cambridge Univ. Press. Cambridge.
- Okada, Y., K. Kasahara, S. Hori, K. Obara, S. Sekiguchi, H. Fujiwara, and A. Yamamoto, 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. and M. A., Saunders, 1982a. LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares, *ACM Trans. Math. Softw.* 8 (1), 43-71.
- Paige, C. C. and M. A. Saunders, 1982b. Algorithm 583: LSQR: Sparse Linear Equations and Least Squares Problems, *ACM Trans. Math. Softw.* 8 (2), 195-209.
- Papazachos, C. B. and Nolet. G., 1997. P and S deep velocity structure of the Hellenic area obtained by robust non-linear inversion of travel times, *J. Geophys. Res.* 102, 8349-8367.
- Parker, R.L., 1994. *Geophysical Inverse Theory*, Princeton Univ. Press.
- Penrose, R. 1955. A generalized inverse for matrices, *Proc. Cambridge Philos. Soc.* 51, 406–413, doi:10.1017/S0305004100030401.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., 1997. *Numerical Recipes in C: The Art of Scientific Computing*, Second Ed., Cambridge Univ. Press, Cambridge, England, 394–455.
- Pothen, A., 1984. *Sparse Null Bases and Marriage Theorems*, Ph.D. thesis, Cornell Univ.
- Pothen, A. and C.-J. Fan, 1990. Computing the block triangular form of a sparse matrix, *ACM Trans. Math. Softw.* 16, 303-324.
- Rajamanickam, S., 2009. *Efficient algorithms for sparse singular value decomposition*, PhD Thesis, Computer Science Dept., Univ. of Florida, Gainesville, FL.

- Rawlinson, N., A. Fichtner, M. Sambridge, and M. K. Young 2014. Seismic Tomography and the Assessment of Uncertainty, In: R. Dmowska, Ed., Adv. in *Geophys.* Elsevier, 55, 1-76, doi: 10.1016/bs.agph.2014.08.001.
- Rawlinson, N. and Spakman, W., 2016. On the use of sensitivity tests in seismic tomography, *Geophys. J. Int.* 205, 1221–1243.
- Rodi, W. L. and S. C. Myers, 2013. Computation of traveltimes covariances based on stochastic models of velocity heterogeneity, *Geophys. J. Int.* 194, no. 3, 1582–1595.
- Saad, Y., 2003. Iterative Methods for Sparse Linear Systems, SIAM, Philadelphia.
- Simons, F. J., Loris, I., Nolet, G., Daubechies, I. C., Voronin, S., Judd, J. S., Vetter, P. A., Chárlety, J., and Vonesch, C., 2011. Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity. *Geophys. J. Int.* 187, 969–988.
- Snieder, R., 1991. An extension of Backus-Gilbert theory to non-linear inverse problems, *Inverse Problems* 7, 409-433.
- Snieder, R. and J. Trampert, 1999. Inverse Problems in Geophysics, in *Wavefield Inversion*, ed. A. Wirgin, pp. 119-190, Springer Verlag, New York.
- Soldati, G. and Boschi, L., 2005. The resolution of whole Earth Seismic Tomography models, *Geophys. J. Int.* 161, 143-153, doi:10.1111/j.1365-246X.2005.02551.x.
- Soldati, G., L. Boschi, and A. Piersanti, 2006. Global Seismic Tomography and Modern Parallel Computers, *Annals of Geophys.* 49, 4-5.
- Stern, R. J., 2002. Subduction zones, *Rev. Geophys.* 40 (4), 1012, doi:10.1029/2001RG000108.

- Tarantola, A., 1987. Inverse problem theory. Elsevier.
- Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM. ISBN 978-0-89871-572-9.
- Tikhonov, A. N. and Arsenin, V. Y., 1977. Solution of Ill-posed Problems. Washington: Winston & Sons. ISBN 0-470-99124-0, 272.
- Trampert, J., Fichtner, A. and Ritsema, J., 2013. Resolution tests revisited: The power of random numbers, *Geophys. J. Int.* 192(2), 676-680. DOI: 10.1093/gji/ggs057.
- Vasco, D.W., 2007. Invariance, groups, and non-uniqueness: the discrete case, *Geophys. J. Int.* 168, 473–490.
- Voronin, S., Mikesell, D., and Guust. N., 2015. Compression Approaches for the Regularized Solutions of Linear Systems from Large-Scale Inverse Problems, *Int. J. Geomath.* 6, 251-294, doi: 10.1007/s13137-015-0073-9.
- Voronin, S., Mikesell, D., Slezak, I. and Guust., N., 2014. Solving large tomographic linear systems: size reduction and error estimation, *Geophys. J. Int.* 199, 276-285.
- Wei, W., Xu, J., Zhao, D. and Shi, Y., 2012. East Asia mantle tomography: New insight into plate subduction and intraplate volcanism, *J. Asian Earth Sci.* 60, 88-103.
- Wei, W., Zhao, D., Xu, J., Wei, F., Liu, G., 2015. P and S wave tomography and anisotropy in Northwest Pacific and East Asia: constraints on stagnant slab and intraplate volcanism. *J. Geophys. Res.* 120, 1642–1666.
- Wiggins, R. A., 1972. The general linear inverse problem: implication of surface waves and free oscillations for earth structure, *J. Geophys. Res.* 10, 251–285.

- Yao, Z. S., Roberts, R. G. and 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. and Tryggvason, A., 2001. Comment on ‘Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems’ by G. Nolet, R. Montelli and J. Virieux, *Geophys. J. Int.* 145 (1), 307-314.
- Zaroli, C., 2016. Global seismic tomography using Backus–Gilbert inversion, *Geophys. J. Int.* 207, 876–888 doi: 10.1093/gji/ggw315.
- Zhang, H. and 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. and McMechan, G. A., 1995. Estimation of resolution and covariance for large matrix inversions, *Geophys. J. Int.* 121 (2), 409-426.
- Zhang, J. and McMechan, G. A., 1996. Reply to Comment by M. M. Deal and G. Nolet on “Estimation of resolution and covariance for large matrix inversions”, *Geophys. J. Int.* 127, 251-252.
- Zhao, D., Hasegawa, A. and Horiuchi S., 1992. Tomographic imaging of P and S wave velocity structure beneath northeastern Japan, *J. Geophys. Res.* 97(B13), 19909–19928, doi:10.1029/92JB00603.
- Zhao D. Hasegawa A. Kanamori H., 1994. Deep structure of Japan subduction zone as derived from local, regional, and teleseismic events, *J. Geophys. Res.* 99, 22 313–22 329.
- Zhao, D., Yanada, T., Hasegawa, A., Umino, N., Wei Imaging W., 2012. The subducting slabs and mantle upwelling under the Japan Islands, *Geophys. J. Int.* 190, 816-828, doi: 10.1111/j.1365-246X.2012.05550.x.

Appendix

The Dulmage-Mendelsohn algorithm

In the methodology discussed above, the coarse Dulmage-Mendelsohn permutation was introduced, using primarily the matrix notation, and focusing on the output of the permutation. Here, we present the basic algorithmic steps that are followed to perform the permutation, together with some necessary concepts of the graph-theory that are used in this method.

In a graph, two vertices are adjacent if there is an edge that connects them. A “path” is a sequence of edges that connects a sequence of vertices which are all distinct from each other i.e., neither vertices nor edges can be repeated (Fig. 17a). An “alternating path” is a path that contains alternating matched and unmatched edges (Fig. 17b). In a directed graph the edges have a direction associated with them, therefore a path has the additional restriction that the edges should all be directed in the same direction (Fig. 17c). In a directed graph, a vertex i is called “ancestor” of another vertex j , if there is a path from i to j , and then j is called “descendant” of i . Finally, a “source”, in a directed graph, is a vertex with no incoming edges (i.e., edges pointing at it), while a “sink” is a vertex with no outgoing edges (i.e., edges pointing to adjacent vertices). Based on the above, the Dulmage–Mendelsohn decomposition can be described as a partition of the vertices of a bipartite graph $\mathbb{G}(\mathcal{R}, \mathcal{C}, \mathcal{E})$, into three subsets that correspond to an under-determined component \mathcal{U} , a well-determined component \mathcal{W} , and an over-determined component \mathcal{O} , with the following properties (e.g., Pothen & Fan, 1990):

- The rows of underdetermined subsystem, \mathcal{U}_r , are reachable via alternating path from some unmatched column.

- The columns of the underdetermined subsystem, \mathcal{U}_c , are reachable via alternating path from some unmatched column.
- The rows of the overdetermined subsystem, \mathcal{O}_r , are reachable via alternating path from some unmatched row.
- The columns of the overdetermined subsystem, \mathcal{O}_r , are reachable via alternating path from some unmatched row.
- The rows and columns of \mathcal{W} are the remaining rows and columns of \mathbb{G} .

The procedures to calculate the three sub-graphs are given below.

1. Find a maximum matching \mathcal{M} of \mathbb{G} .
2. Create the directed graph $\vec{\mathbb{G}}$ from \mathbb{G} , by replacing each edge in \mathcal{M} by two directed edges of opposite direction and orienting all other edges of \mathbb{G} from the equations to the unknowns.
3. Let \mathcal{O} be the set of all descendants of sources of $\vec{\mathbb{G}}$.
4. Let \mathcal{U} be the set of all ancestors of sink of the directed graph $\vec{\mathbb{G}}$.
5. Calculate \mathcal{W} by subtracting the nodes that belong to \mathcal{O} and \mathcal{U} from \mathbb{G}

The first step controls the complexity of the entire algorithm, because it requires $O(v\sqrt{\varepsilon})$ steps using, for example, the bipartite matching algorithm of Hopcroft & Karp (1973), where v is the total number of vertices (i.e., $v = m + n$) and ε is the number of edges (i.e., the number of non-zero elements of \mathbf{G}). The procedures 2 to 5 require $O(v + \varepsilon)$ steps. The algorithm is particularly efficient, because all steps involve only binary operations between binary arrays. In this paper we use the program `cs_dmperm` for computing the Dulmage-Mendelsohn permutation,

which can be found in the SuiteSparse software (Davis and Hu, 2011; also available from <http://www.suitesparse.com>, accessed on November, 2018). It receives as input a sparse matrix (e.g., \mathbf{G}) and returns the row and column indexes of \mathcal{U} , \mathcal{O} and \mathcal{W} , as output.

Figures

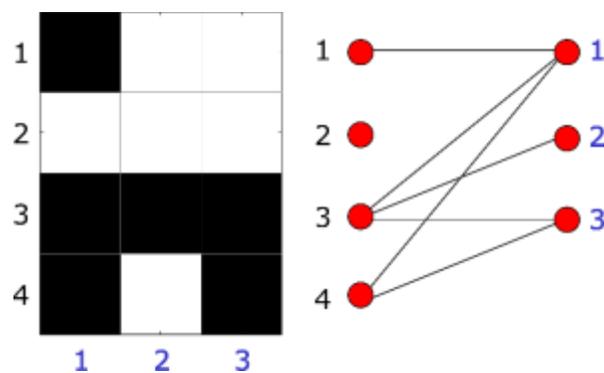


Figure 1: An example of a sparse matrix (left) and its associated bipartite graph (right). The filled cells of the matrix denote elements with non-zero values. The graph consists from two columns of vertices, shown as circles. The left column represents the row-vertices, \mathcal{C} (black labels), and the right column represents the column-vertices \mathcal{R} (blue labels). Each straight line represents an edge, and corresponds to a non-zero element of the matrix, connecting a row-vertex with a column-vertex.

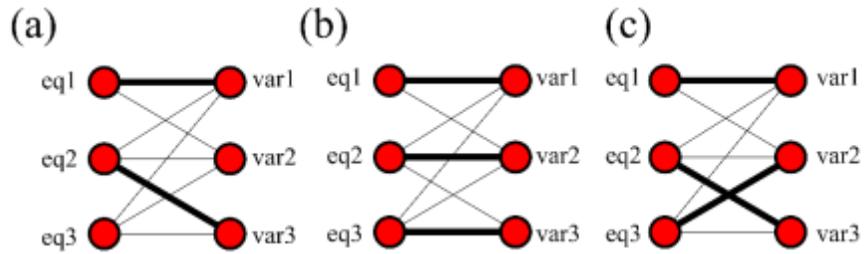


Figure 2: An example bipartite graph of a system with 3 equations and 3 unknowns. Possible matchings are shown with thick edges. (a) In this case, the vertices eq1, eq2, var1 and var2 are matched, while the vertices eq3 and var2 are unmatched. (b) and (c) These two cases show two maximum matchings with thick edges..

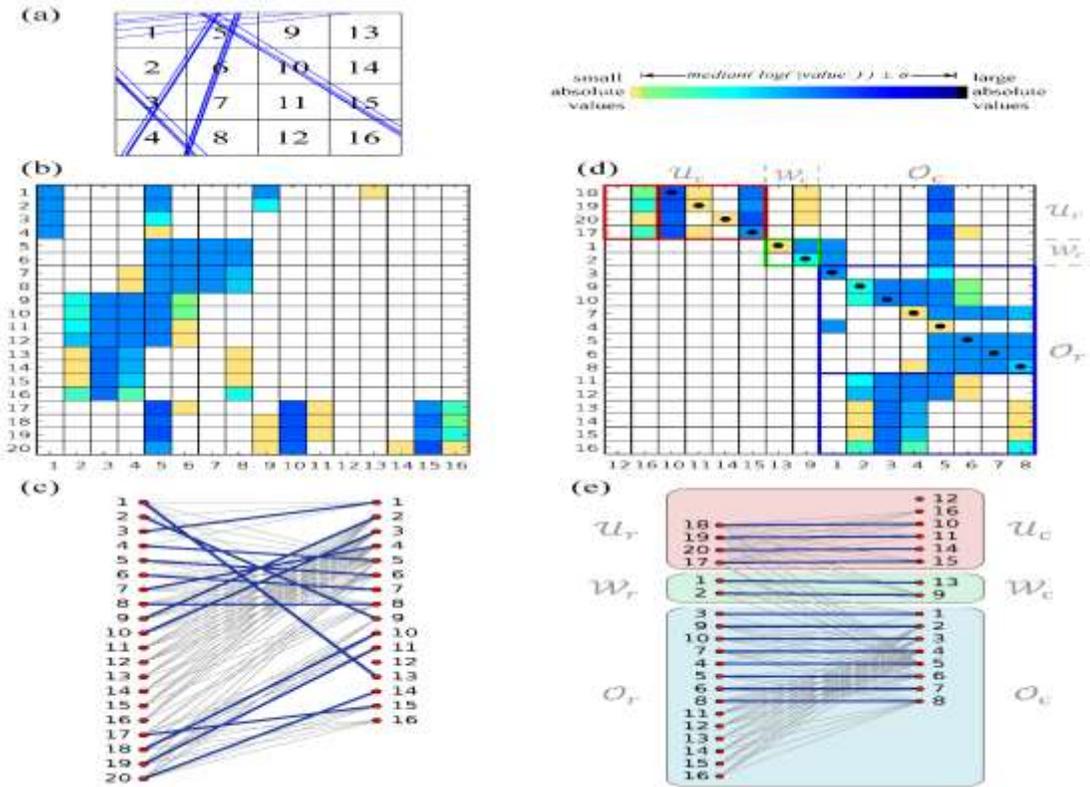


Figure 3: Simple synthetic example and its Dulmage-Mendelsohn decomposition. (a) The 2D model space is discretized into 16 blocks of constant value that is sampled by 20 ray paths. (b) The associated forward operator \mathbf{G} . The element $G(i, j)$ is the length that the i th ray path is sampling the j th parameter that is represented by color. (c) The bipartite graph associated with the sparse matrix \mathbf{G} drawn in its initial order of rows and columns, corresponding to (b). A maximum matching is shown with thick edges. d) The Dulmage-Mendelsohn decomposition of \mathbf{G} . $\mathcal{U}_r, \mathcal{W}_r$ and \mathcal{O}_r indicate the rows, and $\mathcal{U}_c, \mathcal{W}_c, \mathcal{O}_c$ the columns that belong to the structurally under-, well- and over-determined systems, respectively. (e) Bipartite graph showing the Dulmage-Mendelsohn permutation of \mathbf{G} , corresponding to (d). Thick edges correspond to the same maximum matching shown in (c), and the frames indicate the subgraphs derived from the Dulmage-Mendelsohn permutation. Note that this graph is a reordered version of the graph in (c).

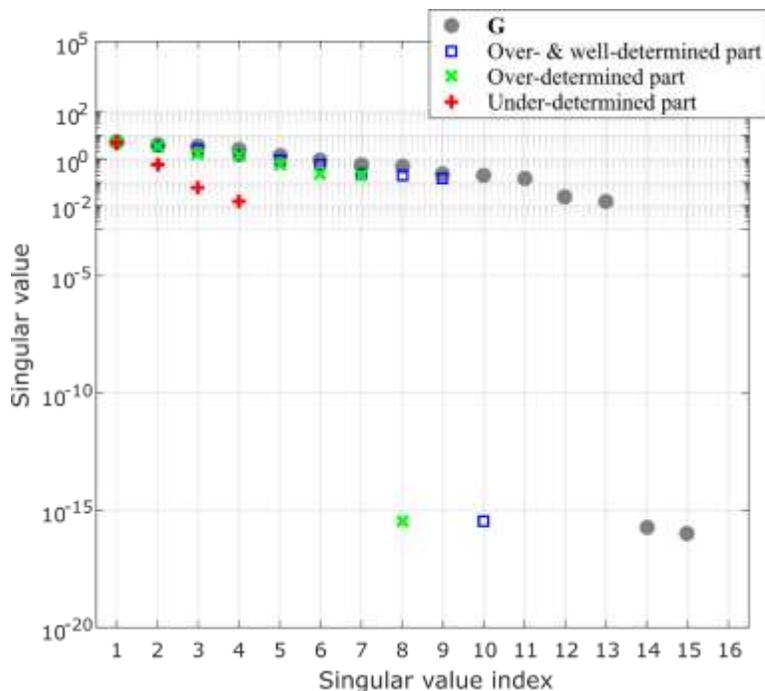


Figure 4: Singular values of \mathbf{G} and different parts of the Dulmage-Mendelsohn decomposition of \mathbf{G} obtained from the singular value decomposition (Lanczos, 1961).

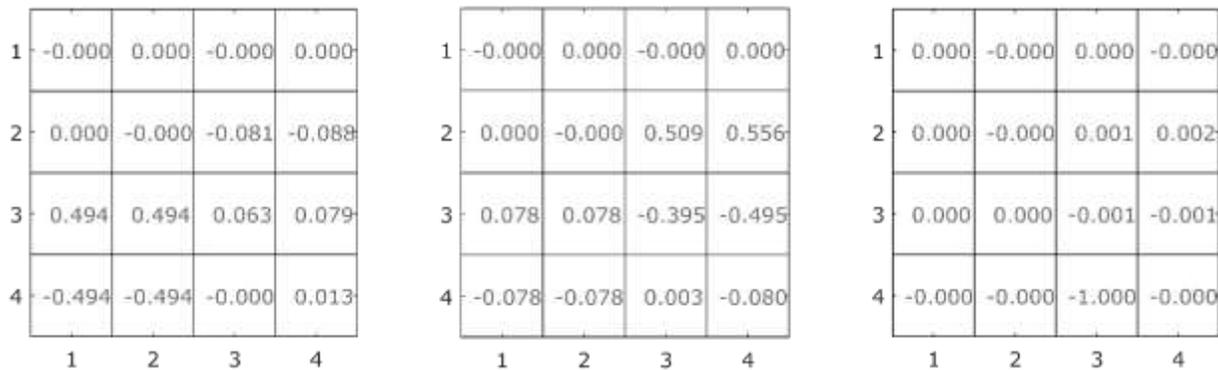


Figure 5: The null space basis vectors of the model shown in in Fig. 3(a). The vectors are arranged column-wise, the same way as the 2-D layout of the model.

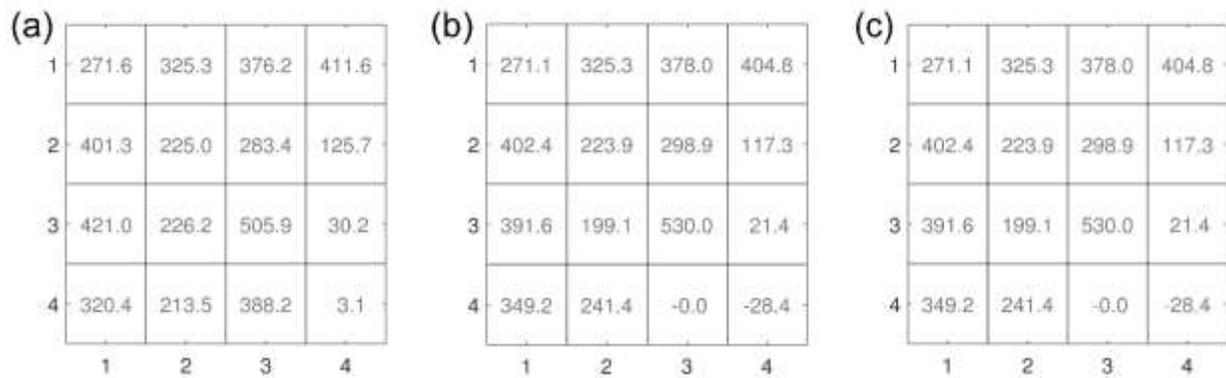


Figure 6: (a) True model. (b) Estimated solution by solving the whole system. (c) Estimated solution using the Dulmage-Mendelsohn approach and solving the sub-problems separately. In all cases, the Moore-Penrose pseudo-inverse operator (Moore, 1920; Penrose, 1955) is applied to derive the solution. The two different estimated models are identical within the machine's precision limits.

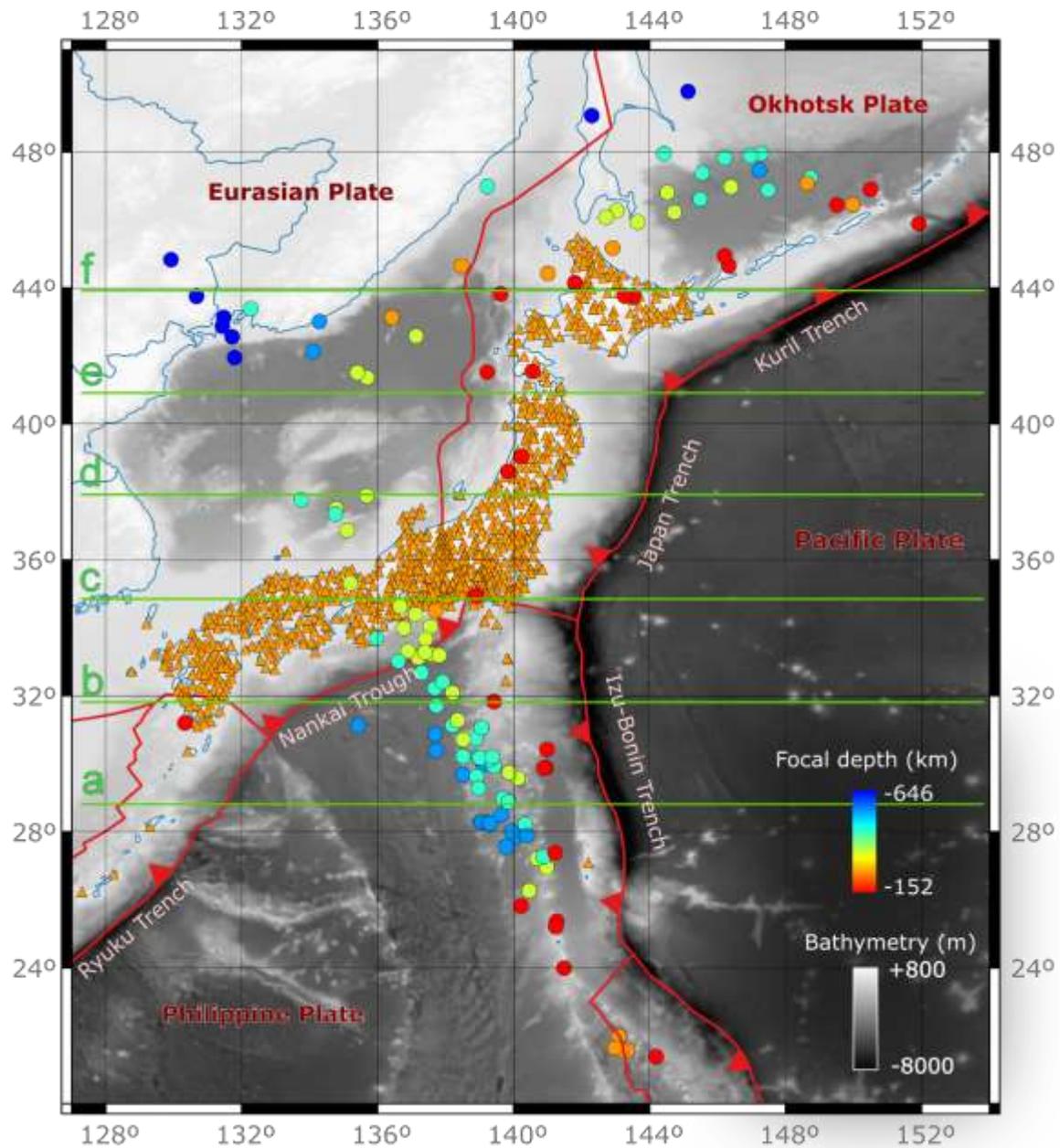


Figure 7: Topography, bathymetry (Amante & Eakins, 2008), and the major plate boundaries (Bird, 2002) of area around Japan. Convergent boundaries are shown with toothed lines. Small orange triangles show the location of Hi-net stations, and circles show the epicenters of the earthquakes with colors indicating the focal depth. Thin blue lines are the coastlines, while green horizontal lines mark the cross sections shown in Figs. 12, 14 and 15.

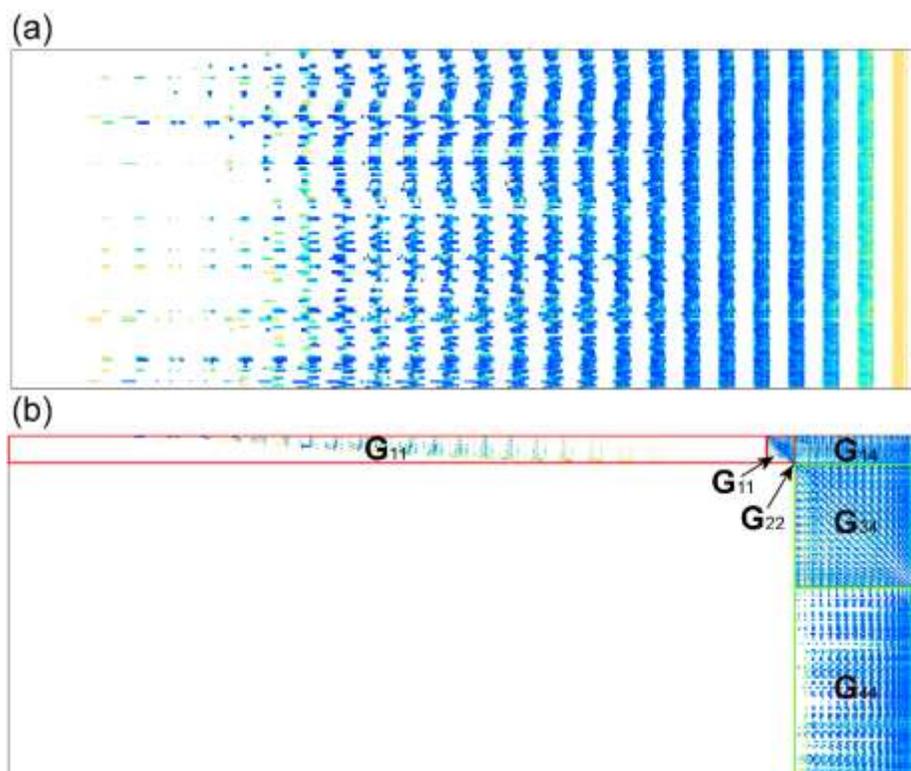


Figure 8: (a) The initial matrix G . (b) The Dulmage-Mendelsohn permutation of G (Eq. 1). The colors show the size of the entries with white being zero and dark blue indicating large absolute values.

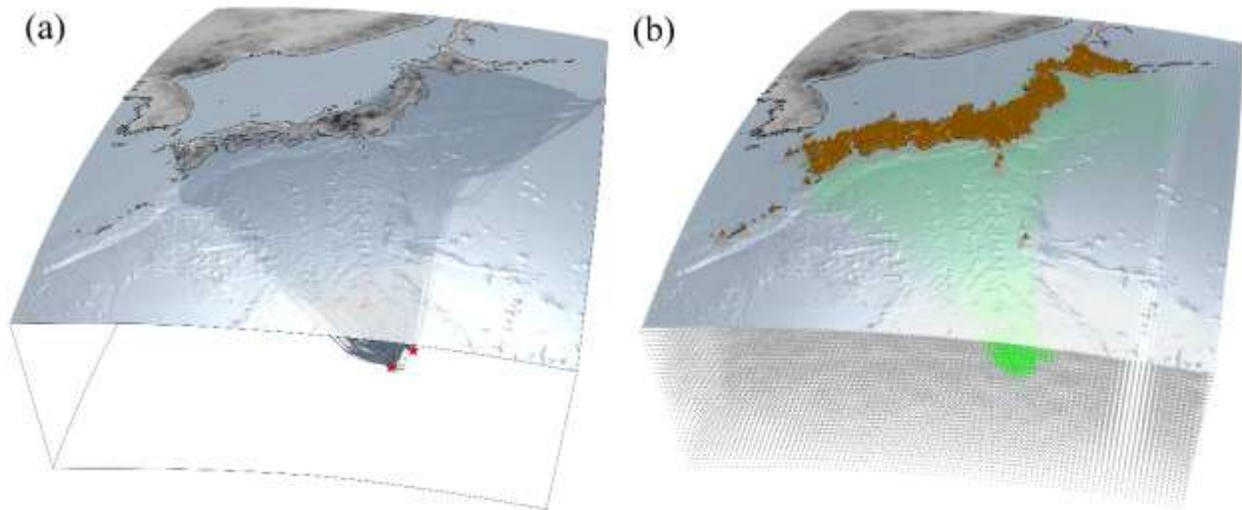


Figure 9: (Left) The model space around Japan where a total of 87,616 P-wave ray paths (gray curves) from 125 earthquakes (stars; Fig. 7) are shown. (Right) Parameterization of the three-dimensional space with 234,500 grids (gray and green spheres). Parameters that correspond to the columns of matrices \mathbf{G}_{11} and \mathbf{G}_{12} (poorly resolved) are colored in gray, while those of \mathbf{G}_{3} and \mathbf{G}_{4} (well resolved) are colored in green. Well-resolved parameters correspond to regions where there is adequate ray coverage. The elevation model shown at the top of the model space is from Amante and Eakins (2009) and Hi-net stations are shown with triangles (Okada et al., 2004).

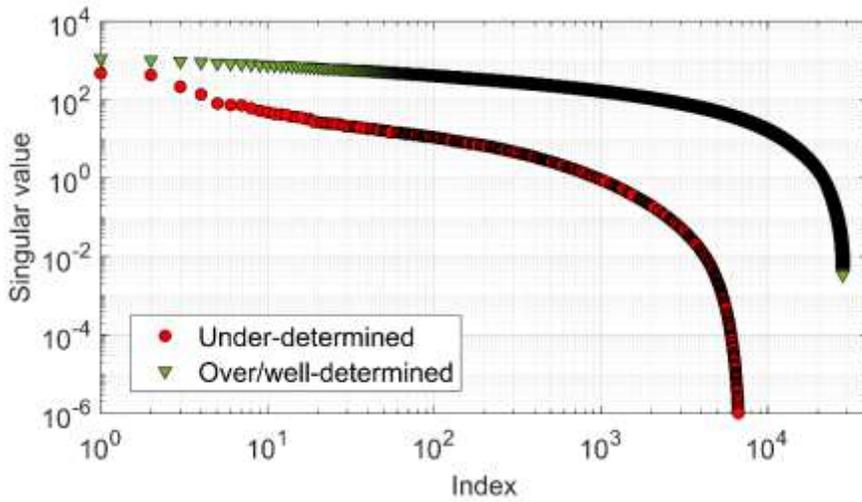


Figure 10: Spectra of the singular values for the under-determined (circles) and the well/ over-determined (triangles) subsystems. Singular values smaller than 10^{-6} are not shown in the graph for the sake of visualization.

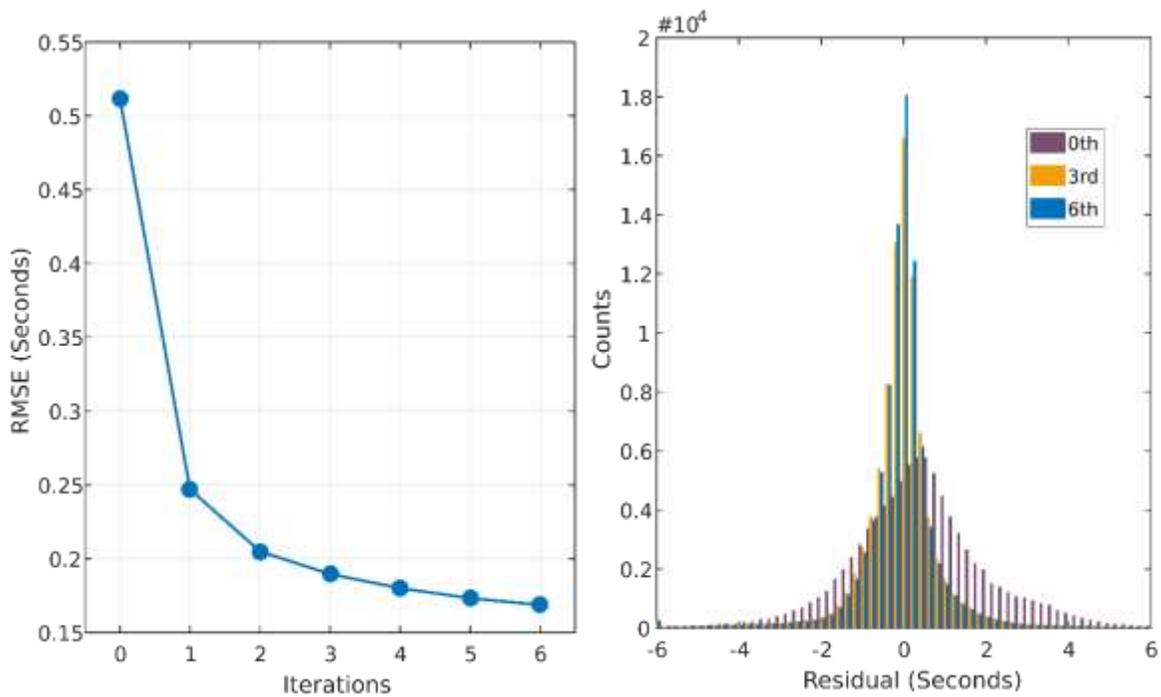


Figure 11: (Left) Evolution of the root mean square error against the iteration. (Right) Histogram of the residuals at different iterations.

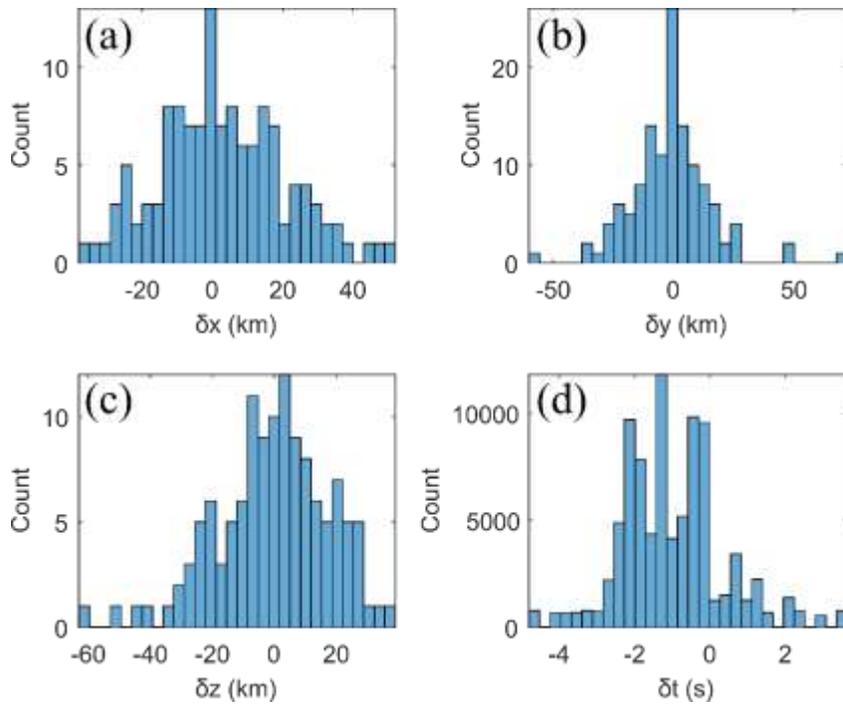


Figure 12: Perturbations of the earthquake hypocentral location (a, b and c) and origin time (d) after six iterations. The perturbations $\delta x, \delta y$ and δz refer to the local, vertical, Cartesian coordinates (see section 4.1).

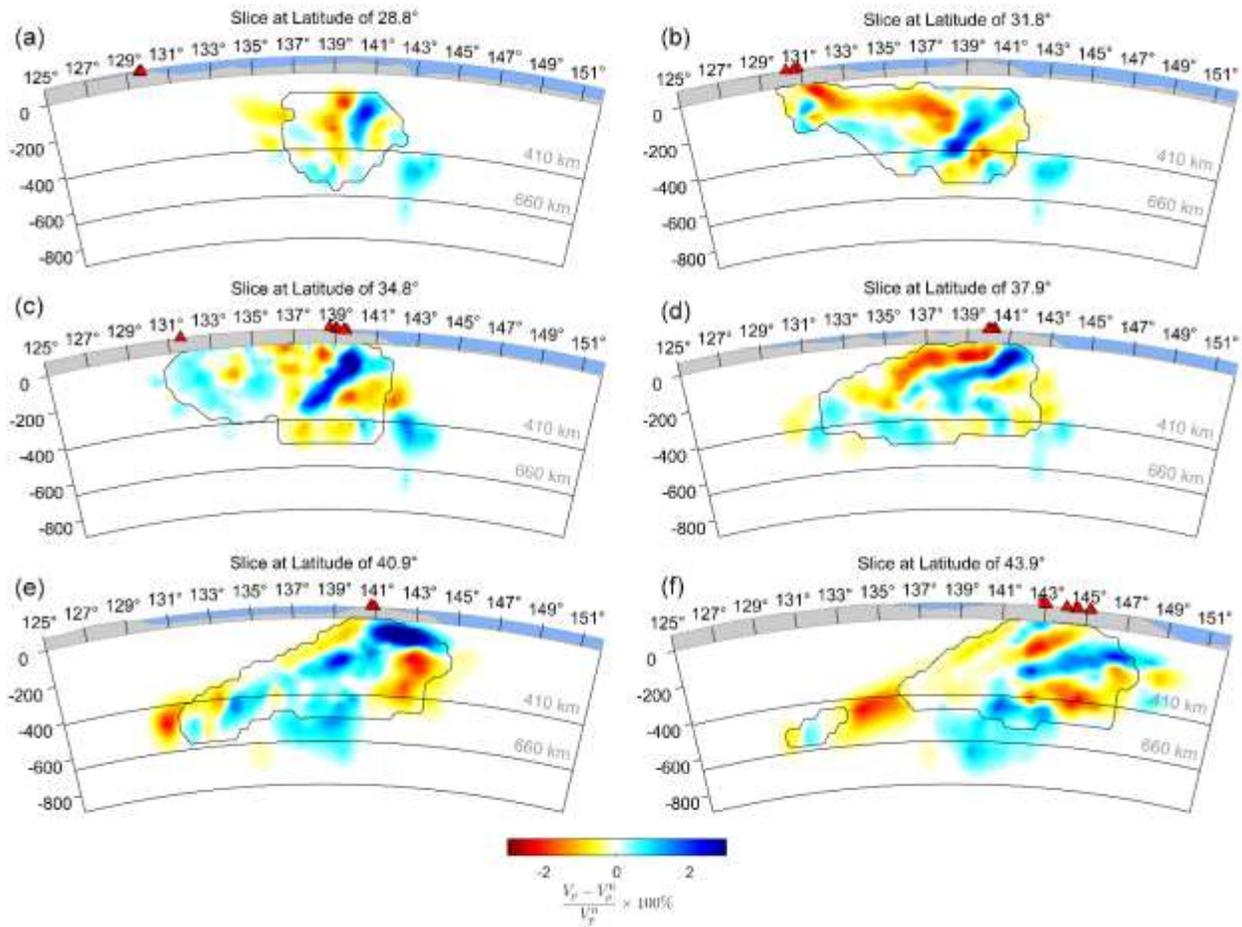


Figure 13: Longitude-depth cross sections of the P-wave velocity anomaly at various latitudes (see Fig. 7). The black contour marks the region that encloses the well/over-determined parameters. The (a) to (f) labels of this figure correspond to the labels of the green lines in Fig. 7.

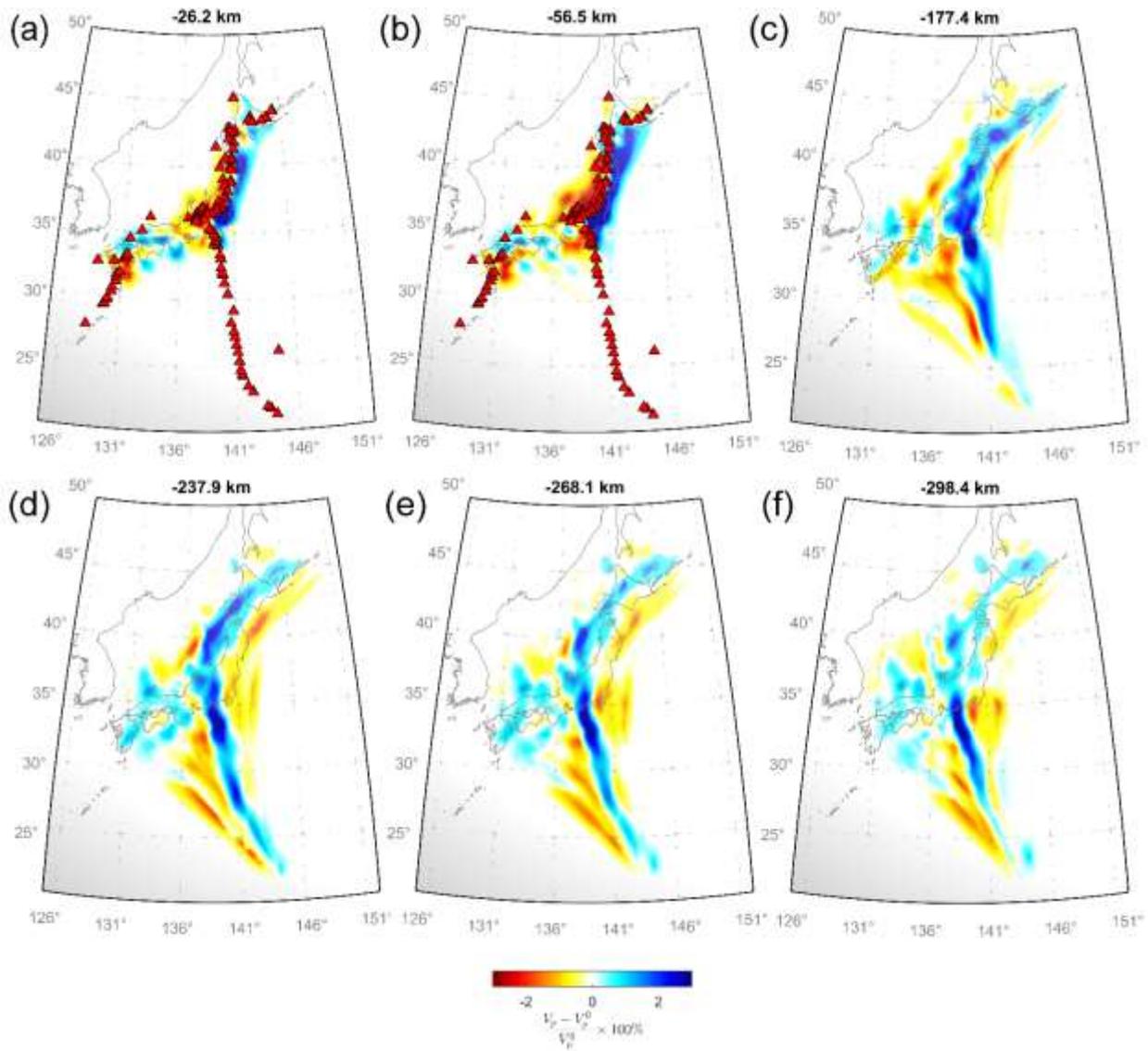


Figure 14: Depth slices showing the P-wave velocity anomaly. In (a) and (b), the locations of active volcanoes are also shown (triangles; Global Volcanism Program, 2013). The depth of the grid layer is shown on top of each slice.

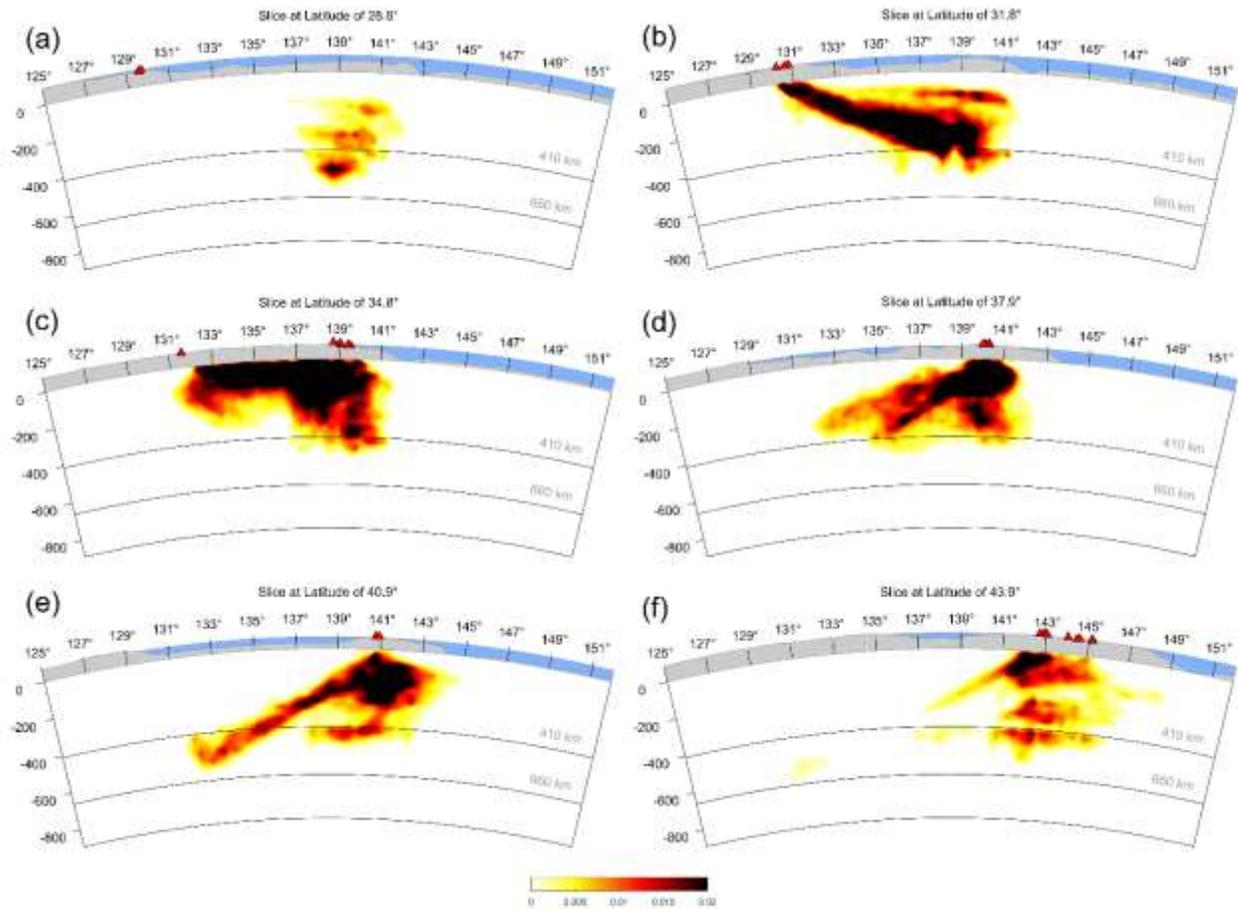


Figure 15: Diagonal model resolution after the final iteration for only the well/over-determined parameters (black contour in Fig. 13). Darker areas correspond to better resolved regions. The (a) to (f) labels of this figure correspond to the labels of the green lines in Fig. 7.

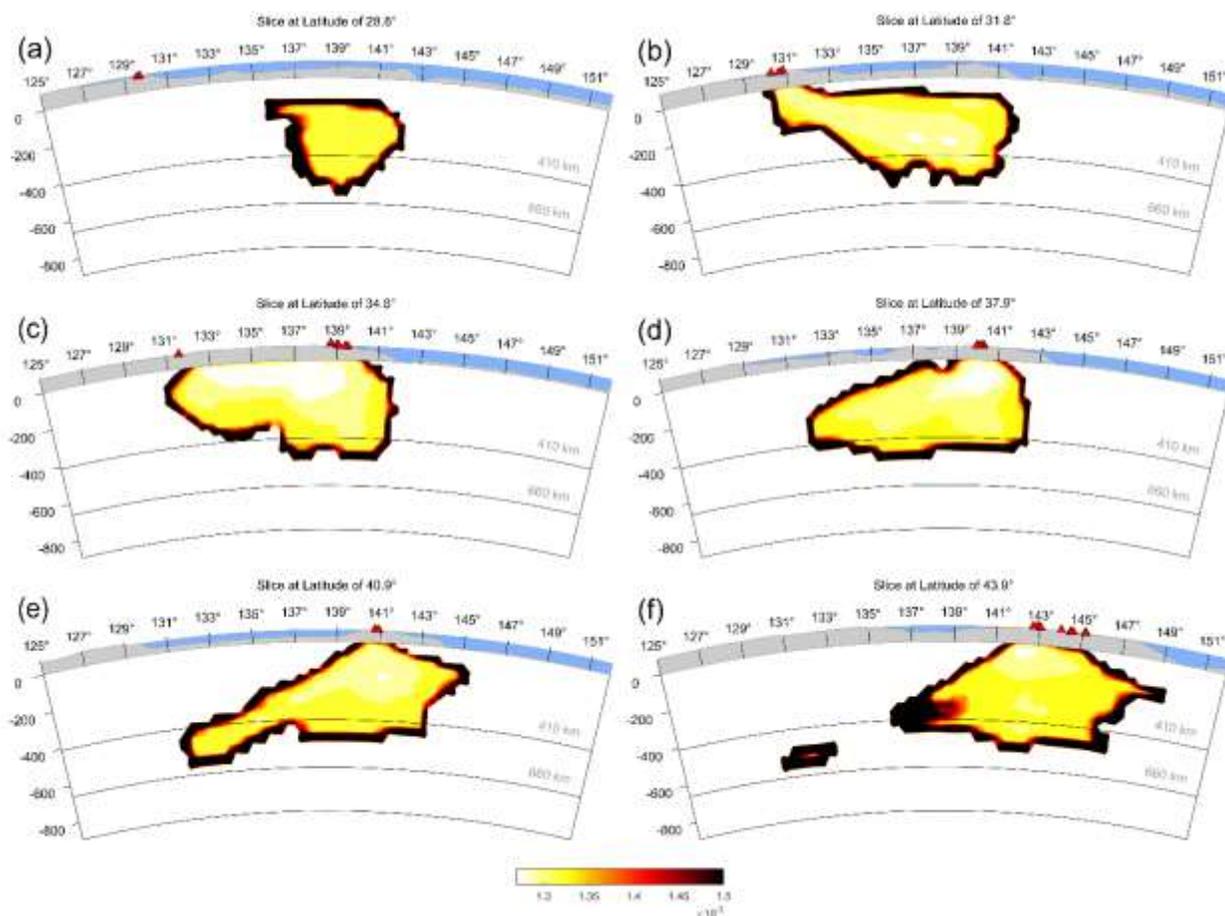


Figure 16: Posterior variance, i.e., the diagonal elements of the posterior model covariance matrix that correspond to the well/over-determined parameters (black contour in Fig. 13), calculated after the final iteration. The (a) to (f) labels of this figure correspond to the labels of the green lines in Fig. 7.

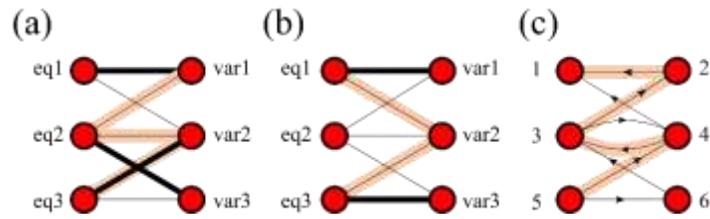


Figure 17: (a) The graph from (Fig. 2c) with an example of a path is highlighted. More precisely two paths correspond to the highlighted edges, i.e., the path (var1, eq1, var2, eq3) or the path (eq3, var2, eq2, var1). (b) An example of two alternating paths (eq1, var2, eq3), or (eq3, var2, eq1) highlighted for the matching shown with the thick lines. (c) An example of a directed graph with the path (5, 4, 3, 2, 1) highlighted. In this case vertex 1 is a sink and vertex 5 is a source.