SVD analysis of GPR full-wave inversion

F. Watson, WRB. Lionheart
School of Mathematics
University of Manchester
Oxford Road, Manchester M13 9PL, UK
francis.watson@manchester.ac.uk, bill.lionheart@manchester.ac.uk

Abstract—Full-wave inversion (FWI) is an imaging approach in which we find the quantitative subsurface parameters (such as the dielectric permittivity) which would best fit the recorded GPR data. This optimisation problem is nonlinear and ill-posed, and there have been numerous successes in applying FWI to GPR data. The dominant properties of the FWI inversion process can be observed in the Jacobian matrix of partial derivatives of the forward map for each acquisition system. Here, we use singular value decomposition (SVD) as a tool to analyse the Jacobian, to help us understand what a given acquisition system is capable of imaging.

We believe FWI could have great benefits in anti-personnel landmine detection, primarily because the additional quantitative information gained could help to reduce the rate of false positives. For humanitarian de-mining there is a need to produce cheaper hand-portable GPR equipment. We therefore ask whether a small array is suitable for FWI, if taking more measurements can compensate for a lack of multi-offset data, using singular value decomposition as a tool to guide our answer.

Index Terms—Full-wave inversion (FWI), landmine detection, singular value decomposition (SVD).

I. INTRODUCTION

GPR has been successfully used for landmine detection and can greatly reduce the rate of false positives (for example [1], [2]), safely speeding the detection process. Our aim is to use state of the art imaging techniques for GPR landmine detection, which provide us with quantitative information about detected targets. We hope that this additional information will contribute to a further reduction in the rate of false positives, as de-mining personnel will be able to determine that more targets are not landmines since they are not made of the right material(s) as well as the wrong shape.

To gain quantitative information, we pose the task of imaging the subsurface as Full Wave Inversion (FWI), that is to solve

$$m_{m} = \arg \min_{m} \frac{1}{2} \| F[ m ] - d \|_2^2 + \lambda R( m ), \quad (1)$$

where $m$ is a set of parameters describing the subsurface, $d$ is the recorded GPR data, and $F[ m ]$ is an operator simulating the data for the parameters $m$. The final term $R$ is a regularisation term which incorporates a priori knowledge and prevents over solving, with $\lambda > 0$. Equation 1 can be read as find the set of subsurface parameters which would give the best compromise between matching the data and incorporating a priori knowledge.

Our research into FWI is part of a wider body of research into improving landmine detection, funded by the charity Find A Better Way. Since we are working alongside the development of new GPR equipment, it is useful for us to explore what form the equipment should take for best results in FWI, or indeed what form it needs to take for acceptable results. In this paper we will use singular value decomposition (SVD) as a tool to evaluate the map from data to image, helping us understand what it is possible for a given antenna array to resolve. Oberröhrmann et al [3] have, similarly, used a chequerboard analysis to determine optimal acquisition setup for coverage and resolution of a cross-borehole experiment. Meles et al [4] also used singular values as a tool to measure the reliability of inverted GPR images from full-waveform data, for 1 to 4-sided experiments. Our results differ in that we will also use the singular vectors to analyse the imaging and null-spaces of the FWI system. Silvestrov and Tcheverda [5] carry out a similar analysis for a cross-borehole seismic experiment, though they use the method to analyse an appropriate parametrisation of the problem rather than the acquisition system.

The landmine detection problem motivates us to get the best possible information out of the most affordable equipment, and less complex acquisition systems with fewer antenna are of course more affordable. Acknowledging that FWI is generally carried out using multi-static arrays for data redundancy, we ask whether enough information can be gained from bi-static systems (or multi-static systems with few receive antenna) for a reliable FWI by taking measurements at many more source locations. To put the question another way, do these additional measurements, from many more physical locations and at many frequencies, yield enough independent information suitable for an inversion? We of course expect that we will be unable to fully resolve the sub-surface parameters using only bi-static data, but it may still be possible to distinguish objects through the additional quantitative information gained. While we concentrate on the specific case of a hand-held GPR system, our methodology can readily be used to analyse the suitability of other acquisition systems.

To compensate for the lack of multi-offset data we must finely sample the frequency spectrum of data in our experiment, and so take 100 frequencies between 1 and 3 GHz to ensure a correct coverage of imaged wavenumbers [6].
II. STATEMENT OF THE PROBLEM

A. Full-waveform inversion

One of the difficulties of solving equation (1) is that many different parameter sets \( \mathbf{m} \) can give the same measurements, up to the noise level, so we need sufficiently strong a priori information to obtain a unique and stable solution. Much emphasis has been placed on developing stable and efficient optimisation procedures for solving the FWI problem, particularly in the seismic community [6], [7]. However, unlike in the seismic imaging case, where data is recorded on large arrays of receivers but over a small frequency band, in the hand-held GPR case we can have only a small array of antennas but a wider frequency band. We therefore wish to analyse what such an acquisition system is able to resolve, and if a more complex antenna array (i.e. large multi-static) is required for good results.

Equation (1) is solved iteratively with a suitable nonlinear optimisation scheme, such as the (l-)BFGS method [8, p. 177]. Any such scheme requires calculation of the gradients of the function \( \mathcal{J} = \| \mathcal{F}[\mathbf{m}] - \mathbf{d} \|^2_2 \) with respect to the parameters \( \mathbf{m} \), which is given by [7]

\[
\nabla_{\mathbf{m}} \mathcal{J} = \Re \{ \mathbf{J}^\dagger \delta \mathbf{d}^* \},
\]

(2)

where \( \delta \mathbf{d} = \mathcal{F}[\mathbf{m}] - \mathbf{d} \) are the data residuals, and \( \mathbf{J}(\mathbf{m}) \) is the Jacobian matrix of partial derivatives \( J_{ij} = \partial \mathcal{F}_i / \partial m_j \), while \( \dagger \) denotes a (non-conjugate) transpose and \( * \) denotes complex conjugation. Note that computation of the Jacobian matrix is not mandatory in order to calculate the gradient, which can be efficiently computed via the adjoint state method [9]. The Jacobian matrix is also calculated via the adjoint state method for this work.

If we assume the use of a (quasi-) Newton method then the descent direction \( \mathbf{p} \) is given by

\[
H \mathbf{p} = -\nabla_{\mathbf{m}} \mathcal{J}(\mathbf{m}),
\]

(3)

where \( H \) is (an approximation to) the Hessian matrix.

The solution to the linearised inverse problem, neglecting regularisation, is [10]

\[
\mathbf{m}_{\text{lin}} = \mathbf{m}_0 + \Re\{\delta \mathbf{m}\},
\]

\[
\mathbf{J} \delta \mathbf{m} = \delta \mathbf{d},
\]

(4)

where \( \mathbf{m}_0 \) is the initial estimate, and \( \delta \mathbf{d} \) is the data residual \( \mathcal{F}[\mathbf{m}_0] - \mathbf{d} \). The least-squares solution is given by

\[
\delta \mathbf{m} = (\mathbf{J}^\dagger \mathbf{J})^{-1} \mathbf{J}^\dagger \delta \mathbf{d}.
\]

(5)

The similarities between a linear inversion, gradient calculation and iterative updates to the solution of the nonlinear inverse problem are made clear on comparison of equations (2), (3) and (5), with the gradient and solution to the linearised inverse problem differing by a scale factor [9], [11]. Note also that \( \mathbf{J}^\dagger \mathbf{J} \) is the Gauss-Newton approximation to the Hessian matrix \( H \). In our example of the (l-)BFGS method, the first iteration is a gradient descent in which the linear inversion is realised. Since we expect the dominant part of the data to be from first order reflections, we assert that the Jacobian matrix exhibits the dominant features of the map from data to image, and in particular the Jacobian of the first iteration of the inversion. Since our optimisation method will always update the image in a direction close to \( \mathbf{J}^\dagger \delta \mathbf{d} \), we cannot expect to be able to resolve anything in the nullspace of \( \mathbf{J}^\dagger \).

In our current work, our analysis is applied to the 2D Helmholtz equation,

\[
(\nabla^2 + k^2) u(\mathbf{x}, \omega) = s(\mathbf{x}, \omega),
\]

(6)

where \( k = \omega/c(\mathbf{x}) \) is the wavenumber and \( s \) a source term. That is, the operator \( \mathcal{F} \) solves the boundary value for the 2D Helmholtz equation for each frequency of interest and evaluates the solution at receiver locations, taking \( \mathbf{m} \) to be some discretisation of \( 1/c(\mathbf{x}) \). We assume that the Helmholtz equation captures the main phenomena of the electromagnetic wave scattering, ignoring polarization. We solve the Helmholtz equation using a 2nd order finite difference scheme.

B. Inverse problems and singular value decomposition

A singular value decomposition (SVD) of \( \mathbf{J} \in \mathbb{R}^{m \times n} \) is given by

\[
\mathbf{J} = \mathbf{U} \Sigma \mathbf{V}^T = \sum_{i=1}^{q} \mathbf{u}_i \sigma_i \mathbf{v}_i^T,
\]

(7)

where \( \mathbf{U} \in \mathbb{R}^{m \times m} \) and \( \mathbf{V} \in \mathbb{R}^{n \times n} \) are orthonormal matrices with columns \( \mathbf{u}_i \) and \( \mathbf{v}_i \), respectively, and \( \Sigma \in \mathbb{R}^{m \times n} \), \( q = \min(m, n) \), is a diagonal matrix of non-negative values \( \Sigma_{ii} = \sigma_i \), ordered such that

\[
\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_q \geq 0.
\]

(8)

The \( \sigma_i \) are called the singular values, the \( \mathbf{u}_i \) and \( \mathbf{v}_i \) are called the left- and right- singular vectors of \( \mathbf{J} \), respectively. The left singular vectors span the data space, and the right span the image space.

Linear ill-posed problems are characterised by the singular values decaying rapidly to zero, so that [12], [13]

\[
\sigma_1 \geq \ldots \geq \sigma_a > \mu \geq \sigma_{a+1} \geq \ldots \geq \sigma_b > \epsilon \geq \ldots \geq \sigma_q \geq \sigma_{q+1} = \ldots = \sigma_q = 0,
\]

(9)

where \( \mu \) is the accuracy threshold set by the measurement instrumentation, and \( \epsilon \) is the threshold for which observations are negligible. The singular values \( \sigma_1, \ldots, \sigma_a \) thus relate to reliable measurements which will be correctly mapped by the Jacobian to a descent direction introducing model parameters responsible for the recorded data. The singular values \( \sigma_{a+1}, \ldots, \sigma_b \) relate to unreliable measurements, and the rest are null (to machine precision). The former will result in noise being introduced to the solution, and the latter will have no effect.

Note that while \( \mathbf{J} \) may represent an over-determined system (4) (i.e. we have more measurements than parameters we are trying to find, \( m > n \) ), \( a \) will typically be far less than the number of parameters \( n \). In this sense, we are left with an effective Jacobian,

\[
\mathbf{J}_{\text{eff}} = \sum_{i=1}^{a} \mathbf{u}_i \sigma_i \mathbf{v}_i^T,
\]

(10)
which maps measurements to the image space spanned by the first \( a \) singular vectors \( v_i \) only. Therefore, we can in fact only expect to be able to resolve targets represented by the vectors in the row space of \( J_{\text{eff}} \), rather than the full matrix \( J \).

Given measurements of relative accuracy \( r \), we can estimate the accuracy threshold \( \mu \) by using the ‘rule of thumb’ that the condition number of \( J_{\text{eff}} \) should not be greater than \( 1/r \), and noting that \( \text{cond}(J_{\text{eff}}) = \sigma_1/\sigma_a \). This rule of thumb ensures that \( J_{\text{eff}} \) satisfies the discrete Picard condition \([12, \text{p. 82}]\).

More precisely though, the meaning of \( \mu \) is \([12, \text{p. 22}]\)

\[
\text{rank}(J_{\text{eff}}) \equiv \min_{\|E\|_2 \leq \mu} \text{rank}(J + E),
\]

in which we may consider the arbitrary \( m \times n \) matrices \( E \) to represent modeling errors.

III. NUMERICAL RESULTS

We now calculate the Jacobian matrices using an adjoint-state formulation \([7]\), before taking an SVD and projecting objects of interest onto the row space of the respective effective Jacobians to illuminate whether the system would be able to resolve such a target.

Our simulated systems all have a 20 cm maximum source-receiver offset, a limit chosen to represent our need for hand-held devices. Measurements are taken at 100 frequencies linearly sampled between 1 and 3 GHz, at source positions evenly spaced between 10 cm and 70 cm along the surface of a 1 m wide by 30 cm deep domain (so that the final receiver position is 10 cm from the other end of the domain). These limitations were chosen represent possible hand-held GPR measurements.

The test domain is 1 m wide by 0.3 m deep with absorbing boundary conditions on the sides and bottom, and has a background wave speed of \( 1.5 \times 10^3 \text{m/s} \). It is discretised to 1 cm for imaging purposes, however it is refined further to 0.33 cm for both forward and adjoint solutions of the Helmholtz equation to ensure accuracy in the higher frequencies.

The test variable we are interested in is the number of receive antennas, which are evenly spaced in the 20 cm interval. We analyse both fixing the total number of measurements for our comparison, reducing the number of source locations with an increased number of receive antennas, and fixing the source locations. In the former case we keep \( n_s \times n_r = 48 \), where \( n_s \) is the number of source locations and \( n_r \) the number of receive antennas, and in the latter we keep \( n_s = 16 \). For our accuracy threshold, we accept singular vectors \( v_i \) associated with singular values \( \sigma_i \) satisfying \( \sigma_i > 10^{-2}\sigma_1 \), which represents a possible accuracy of noisy measurements with 2 significant figures accuracy. Our results are not sensitive to a (reasonable) change in this accuracy threshold, as the number of singular vectors above the threshold for each test case remain in proportion.

A. SVD analysis

Fig. 1 shows us the first 1000 singular values for Jacobian matrices calculated for the homogeneous domain, representing the first linear step of the inversion. The near straight lines in these log-log scale plots show that the singular values decay like \( i^{-\gamma} \), with \( \gamma > 0 \) a positive real constant which increases at around \( i = 500 \) (depending on the set of singular values in question). This fits the rapid decay typical of an inverse problem, as mentioned in section II-B. The steepening at \( i = 500 \) is an effective drop in rank, implying less independent information can be gained from subsequent measurements.

Fig. 1a shows a slower decay in singular values for systems with more receivers, which is as we would expect. However, when we fix the total number of measurements, Fig. 1b shows a similar rate of decay for all four systems. Particularly, the \( \sigma_i \) for the bi-static system decay almost as slowly as those for the multi-static systems, and those for the 3 and 4 receiver case are almost indistinguishable. This suggests we may be
able to at least partially compensate for a lack of multi-offset data by increasing the number of source locations, as each results in a Jacobian matrix with a similar condition number.

However, it is not just the rank and condition of $J_{\text{eff}}$ that are important, or equivalently amount of data above the noise level, but the suitability of these measurements to image our target. One can only expect to image something which makes a significant enough impact to observed data and better measurements do not necessarily just mean more observations above the noise level. To understand the suitability of our measurements, one must also look at the singular vectors $v_i$, and how they map the imaging space.

For a realistic noise level $J_{\text{eff}}$ will always have a large nullspace, and so taking more measurements to result in a better conditioned Jacobian is not going to help if what we want to be able to image still lies in this nullspace. To help us understand the image and nullspaces of $J_{\text{eff}}$, we can project test targets onto singular vectors $v_i$ corresponding to singular values above the accuracy threshold. As previously mentioned, we take the accuracy threshold to be given by the $\sigma_i$ satisfying $\sigma_i > 10^{-2}\sigma_1$.

The targets we use are 5 cm squares placed 3 cm below the surface and 2 cm apart, and have wave speeds of $1.3 \times 10^8$ and $1.4 \times 10^8$ m/s, respectively, as shown in Fig. 3. These are chosen as we are interested in whether a system is able to distinguish nearby objects based on quantitative information, particularly with a low contrast to one another. The low contrast, both between the objects and to the background, also means that the linear approximation should be a good one.

The results of the projections are shown in Fig. 2, and the

![Fig. 2](image)

- **Fixed source locations**
- **Fixed number of measurements**

![Fig. 3](image)

- Depth $z$, m
- Horizontal position $x$, m

## Table I: Relative Error of Projections of Low Contrast Targets onto the Right Singular Vectors of Acquisition Systems, for Fixed Source Locations (Column A) and Fixed Total Number of Measurements (Column B)

<table>
<thead>
<tr>
<th>Acquisition system</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-static</td>
<td>37.5%</td>
<td>32.7%</td>
</tr>
<tr>
<td>2 receivers</td>
<td>29.1%</td>
<td>26.0%</td>
</tr>
<tr>
<td>3 receivers</td>
<td>26.5%</td>
<td>26.5%</td>
</tr>
<tr>
<td>4 receivers</td>
<td>24.1%</td>
<td>26.7%</td>
</tr>
</tbody>
</table>

Note that projections are onto $\delta m = m_{\text{true}} - m_0$ for a homogeneous background $m_0$ of wave speed $1.5 \times 10^8$ m/s.
relative errors of the projection compared to the true $\delta m$ are given in Table I. As we ought to expect, for fixed source locations we get better results for the systems with more receivers, with a progressively better contrast between the two objects. The improvement going from bi-static to 2 receivers is significant, but increasing further to 3 or 4 receivers less so. Particularly, the bi-static projection gives a lower contrast and a ‘noisier’ background. This is clear from Table I, which shows the projection of the 4 receiver system is 15.9% closer to the true solution than the bi-static case, with the 3 and 4 receiver projections on par.

For the projections with a fixed total number of measurements, only the bi-static result remains poor, with all multi-static projections giving results to within 1% relative error of each other. The implication is that, provided we have some degree of multi-static data in 20 cm offset limit, we ought to be able to at least partially compensate for a lack of large array by taking more measurements.

B. Inversion results

We now invert data simulated for each of the four acquisition systems, for both 16 source locations and 48 total measurements (locations $\times$ receivers), at 100 frequencies between 1 and 3GHz, with 5% Gaussian noise added. We use a small weight of total variation regularisation [14, p. 129] for stability, and allow 20 iterations of the l-BFGS optimisation procedure.

The inversion results are shown in Fig. 4, and Fig. 6 shows the absolute model error with each iteration. It is clear that the inversion of bi-static data reaches a local minima much farther from the true solution than the multi-static cases, and we are also almost unable to distinguish the two objects from one-another in the end result. Conversely, it is easy to distinguish the two objects from the inversions of multi-static data, as highlighted by the horizontal slices of the inversion shown in Fig. 5.

When we fix the total number of measurements (an increase in measurements for 2 receivers, reduction for 4, and the same data set for 3 receivers), the results from the multi-static inversion do become slightly more similar, though there is little improvement in the bi-static case. This is most apparent from the plots of model error, Fig. 6. This backs up the result of our projections onto singular vectors, in which we saw similar results when keeping the total number of measurements fixed. We suspect the difference between the multi-static inversion results with a fixed number of measurements is due to the non-linear part of the inversion (i.e. inversion of the part of the data caused by diffraction and multiple reflections between the two objects), since this part is ignored by the SVD analysis.

IV. DISCUSSION AND FURTHER WORK

Taking an SVD of the Jacobian matrices shows us how data can be mapped to the gradient in our inversion procedure, allowing us to see what can (or can not) be resolved in a linear step of the inversion for a given acquisition system. Since the inversion procedure consists of many more non-linear updates, a linear step being unable to resolve a certain target does not mean we cannot resolve it at all, but that we will be relying on non-linear steps to do so – e.g. relying on data caused by multiply scattered or refracted waves. Further, we expect the
singly scattered waves to give the biggest contribution to the data [5], and so the first linear step ought to give the largest contribution. Because of this, our SVD analysis gives a good indication of the row- and null-spaces of the full non-linear inversion, without having to calculate the Jacobian matrix for subsequent iterations (the Jacobian $J$ changes with each iteration).

In a similar fashion to our SVD analysis, with an aim to determine the information content of data and the extent to which an image was fully resolved, Meles et al [4] formed a cumulative sensitivity plot from Jacobian matrices to help gain understanding about what a system may be capable of imaging. If there is no sensitivity in an area, then one cannot detect an object there. However, an object may lie in an area of sensitivity, but still lie (partially) in the nullspace of $J$, since the singular vectors $v_i$ above the noise level may not be able to adequately represent the shape of the object or its contrast against the surrounding ground. This is where our analysis differs, as we have also explored these singular vectors, as did Silvestrov and Tcheverda [5] when considering appropriate model parametrisation for a seismic cross-borehole experiment.

Our analysis fixed the maximum source-receiver offset at 20 cm, and it would be interesting to see how the minimal offset affects our results. We also considered only a simple test problem with two square objects. This was useful for us since one of our aims is to use FWI to distinguish objects through additional quantitative information, but other (more realistic and complex) domains should be considered. Further, we have neglected to include antenna characteristics, as well as simplifying our forward map to the 2D Helmholtz equation, and so must be careful in how we apply any conclusions to the design of a real GPR system. To fully utilise SVD as a tool to aid designing an acquisition system, we ought to work with the 3D Maxwell equations, including antenna characteristics as well as different layouts. That said, if we find poor results in this simplified case then we cannot reasonably expect better results in the more complex 3D Maxwell case, since we will be adding more degrees of freedom to the inversion. There is therefore little point in studying a bi-static system for FWI, but it would be interesting to continue the study with small-scale
V. Conclusions

We presented an SVD analysis of Jacobian matrices for acquisition systems with 1 to 4 receive antennas. This was used as to analyse the ill-posedness of the FWI problem for the resulting data set, as well as being a tool for analysing the ability of such a system to image objects of interest (or, equivalently, to explore their nullspaces and inability to image). The method can readily be applied to the analysis of other acquisition systems, provided calculation of the Jacobian matrix is not prohibitive.

In our analysis, when we fixed the source locations for the different acquisition systems, the decay of singular values showed Jacobians for systems with more receivers have a greater rank (up to a given noise level), or equivalently the system is better conditioned. However, when we fix the total number of measurements (within our hand-held limitations of up to 4 receivers in a 20 cm maximum offset), the difference between the decays of singular values for all systems is almost negligible. Since these Jacobian matrices then have approximately the same rank and condition, this implies that we ought to be able to compensate for a lack of multi-offset data by taking measurements at more locations.

When we include noise though, regardless of how many measurements we are able to take, we still have a large nullspace of the resulting effective Jacobian. It is therefore important not just to look at the singular values to analyse the condition of a system, but also to analyse the span of singular vectors above the noise level. The results of projections of our objects of interest onto these singular vectors imply that, when we have at least some degree of multi-static data (within our imposed hand-held limits), we can compensate for a having fewer offsets by taking more measurements. This is not the case for the bi-static system which, despite having approximately the same rank and condition, is unable to distinguish the objects. These results are backed up anecdotally by our inversion results, and we see that a small scale multi-static array ought to be sufficient to gain quantitative information able to distinguish objects, though not necessarily fully resolve a more realistic domain.

This highlights the fact that it is not just having more singular vectors available to span our image space that are important (i.e. obtaining a better conditioned system), and taking more measurements to average over noise. Rather, better measurements are also those that result in singular vectors better able to resolve what we are interested in, as we need our data to be sufficiently sensitive to their presence. It is the latter that has led to all the multi-static systems behaving comparably to each other, and the bi-static system falling short.

Acknowledgment

We would like to thank Find A Better Way for their support of this research. Find A Better Way is a charity established to improve the lives of individuals and communities affected by Landmines, and aims to develop technology to accelerate the detection and safe removal of landmines globally.

References