Visualizing the misfit landscape for full waveform inversion

3 authors:

Muhammad Izzatullah
King Abdullah University of Science and Technology
10 PUBLICATIONS   **6** CITATIONS

SEE PROFILE

Tristan van Leeuwen
Utrecht University
112 PUBLICATIONS   **1,159** CITATIONS

SEE PROFILE

Daniel Peter
King Abdullah University of Science and Technology
102 PUBLICATIONS   **1,092** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   SRVM-based Uncertainty Quantification in FWI View project

Project   Artificially Intelligent seismic Full-Waveform Inversion (AI-FWI) View project

Visualizing the misfit landscape for full waveform inversion

Muhammad Izzatullah, Tristan van Leeuwen[†], and Daniel Peter**
** King Abdullah University of Science and Technology, Saudi Arabia* [†] *Utrecht University, The Netherlands*

SUMMARY

Non-linear optimization plays a big role in many tasks in geophysics, such as full waveform inversion (FWI). Visualization of the objective function is useful for the analysis and development of algorithms and formulations. However, in high-dimensional problems, we do not have the capabilities to perform such visualization. Instead, one often works with one or two predefined directions in which to slice the objective. In this work we present an approach to visualizing the misfit landscape together with the optimization trajectories based on principal component analysis. Here, the directions along which to slice the objective are chosen in accordance with the optimization trajectory. We demonstrate the approach through a numerical example using the Marmousi model.

INTRODUCTION

In low-dimensional optimization problems with up to 3 parameters the misfit landscape is readily visualized. For high-dimensional problems, we need alternatives to study objective. A prime example of such a high-dimensional optimization problem is full waveform inversion (FWI) (Tarantola and Valette, 1982; Pratt et al., 1998; Mulder and Plessix, 2008; Virieux and Operto, 2009; Virieux et al., 2017). The objective function is typically non-convex and exhibits local minima. Hence, to conduct a thorough analysis of the objective function via visualization is not trivial. The common practice to validate the results obtained from the optimization process is by plotting the convergence history. However this gives us no information about the structure or curvatures of the objective function.

In this paper, we adapt an idea from (Goodfellow et al., 2015; Keskar et al., 2016; Im et al., 2016; Li et al., 2018) to explore the misfit landscape around a given model $\mathbf{m}^*$ using component analysis (PCA). We use the principal vectors to construct a low-dimensional visualization of the misfit landscape. This enables us to study the convexity of the misfit landscape and asses performance of the optimization method.

The outline of this paper as follows. First, we discuss the theoretical foundation for visualizing the misfit landscape. Next, we demonstrate the idea on a regularized frequency-domain FWI problem. Finally, we discuss the results and conclude this research work.

THEORY

Non-linear optimization

Non-linear optimization plays a major role in many tasks especially in geophysics community and understanding the ob-

jective function is important for the development of algorithms and formulations. The commonly implemented objective function for the nonlinear optimization problems like FWI can be expressed as

$$V(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\right\|_2^2 + \frac{\alpha}{2}\|\mathbf{L}\mathbf{m}\|_2^2, \qquad (1)$$

where $\mathbf{F}(\mathbf{m})$ represents the non-linear forward map which takes the model $\mathbf{m}$ as an input and returns the data $\mathbf{d}$ as an output. The regularization operator $\mathbf{L}$ is typically taken to be a first or second order derivative operator and α is the regularization parameter.

The resulting optimization problem is typically solved using a gradient-based method which generates iterates of the form

$$\mathbf{m}_{k+1} = \mathbf{m}_k - B_k \nabla V(\mathbf{m}_k), \qquad (2)$$

where B_k includes appropriate scaling / smoothing of the gradient.

Visualizing misfit landscape along optimization trajectories

For visualizing the misfit landscape of the objective function along the optimization trajectories, one chooses a center point $\mathbf{m}$ (typically the optimal model $\mathbf{m}^*$), and chooses two direction vectors, $\mathbf{p}_1$ and $\mathbf{p}_2$. One then plots a surface based on function of the form

$$f(\alpha_1, \alpha_2) = V(\mathbf{m}^* + \alpha_1\mathbf{p}_1 + \alpha_2\mathbf{p}_2), \qquad (3)$$

where α_1 and α_2 are the scalar parameters. The key insight of (Goodfellow et al., 2015; Li et al., 2018) is to choose the direction vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ to be the principle components of the error matrix,

$$\mathbf{E} = [\mathbf{m}_0 - \mathbf{m}_n, \cdots, \mathbf{m}_{n-1} - \mathbf{m}_n] \qquad (4)$$

where $\mathbf{m}_n = \mathbf{m}^*$ is the optimal model or the converged solution at the last iteration n. The two principal components are essentially the largest eigenvectors of $\mathbf{E}\mathbf{E}^T$ and they contain the dominant variation in the iterates generated by (2). Through this approach, we are able to visualize the effective misfit landscape as experienced by the optimization method, allowing us to compare algorithms in a qualitative fashion.

Visualization based on the gradients

As an alternative to the approach sketched above, we may construct the directions $\mathbf{p_i}$ from the gradient matrix

$$\mathbf{G} = [\mathbf{g}_0, \cdots, \mathbf{g}_{n-1}, \mathbf{g}_n], \qquad (5)$$

where $\mathbf{g}_k = \nabla V(\mathbf{m}_k)$. The resulting principal components capture the variation in curvature. This approach is closely linked to the *active subspace method* for dimension reduction introduced by Russi (2010) and Constantine (2015). The difference between our approach and the active subspace method, is that we consider the gradients at each iteration instead of drawing them as samples from a distribution.

By combining these two approaches, we are able to visualize the misfit landscape along the optimization's trajectories based on the variation of the descent path and the curvature in the model space.

NUMERICAL EXAMPLES

Toy model

First, we illustrate the proposed methodology a simple example in order to provide an intuitive sense for the reader. Consider a quadratic function of 3 variables, $V(\mathbf{m}) = \mathbf{m}^T \mathbf{A} \mathbf{m}$, where $\mathbf{A}$ is symmetric and positive definite. We performed 100 conjugate gradient method (Saad, 2009; Chong and Zak, 2013) starting with the initial model $\mathbf{m}_0 = [1, 1, 1]^T$.
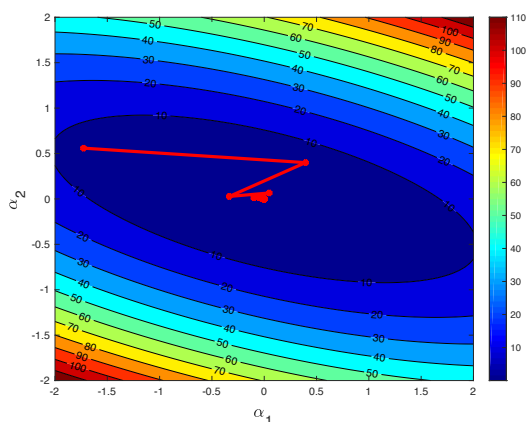
Figure 1: The contour of quadratic function misfit landscape along the optimization trajectories constructed with principal component vectors of error matrix. The trajectories plotted in red.
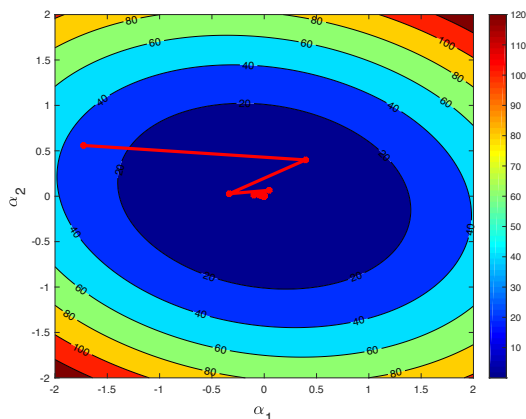

Figure 2: The contour of quadratic function misfit landscape constructed with principal component vectors of gradient matrix. The trajectories plotted in red.

Having stored all iterates and gradients, we can extract the

dominant directions from both error (4) and gradient matrices (5). Here we use the principal component analysis (PCA) method to extract the first two dominant principal component directions for misfit landscape visualization purposes.
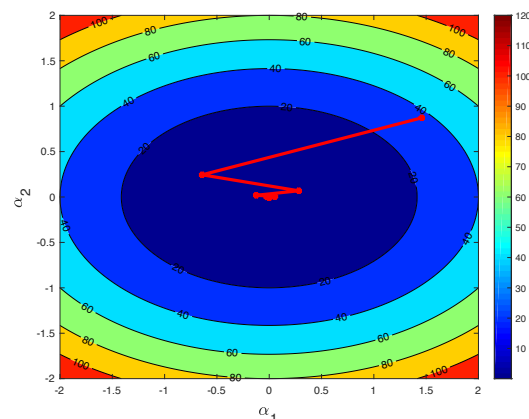

Figure 3: The contour of quadratic function misfit landscape constructed with eigenvectors of the quadratic function Hessian. The trajectories plotted in red.

For visualizing the misfit landscape along the optimization trajectories, we need to evaluate the objective function once again in the form of equation (3). Here, the $\mathbf{p}_1$ and $\mathbf{p}_2$ are the first and the second principal component vectors of the error and gradient matrix respectively. The resulting visualizations are shown in Figures 1 and 2. We plot the misfit landscape constructed using the first and the second eigenvectors of the Hessian in Figure 3 for comparison.

Based on this toy model example, the misfit landscapes constructed from the principal component vectors of the error and gradient matrices as well as the one constructed using the eigenvectors of the Hessian, give us the same information. We observe that the visualization captures the dominant directions of the Hessian as well. This gives us an intuition that this approach able to inform us information about the misfit in high dimensional problems.

Marmousi Model

In this numerical example, we use the 2D Marmousi model, shown in Figure 4, to demonstrate the proposed methodology. All the computations were performed in MATLAB using a standard finite-difference discretization of the Helmholtz equation (van Leeuwen, 2019). The grid size is 61×220, with 50m gridspacing. A total of 50 sources and 100 receivers are placed near the top of the model. The data are generated for frequencies 0.5 to 4 Hz with a spacing of 0.5 Hz.

As regularization we use the Tikhonov regularization method with regularization operator $\mathbf{L}$ is taken to be a first order derivative operator and regularization parameter α equals to 0.1. We invert all frequencies simultaneously using a nonlinear conjugate gradient method (Saad, 2009; Chong and Zak, 2013; Yang, 2014) starting from the initial model shown in Figure 4. The resulting model after 100 iterations is shown in Figure 4.
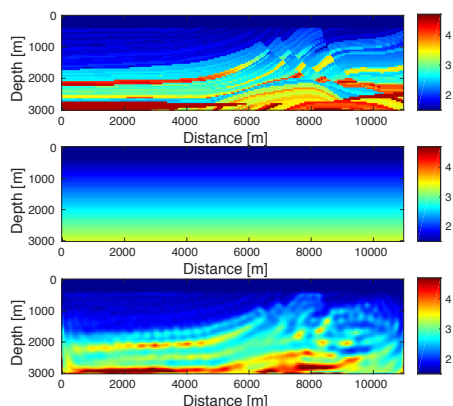
Figure 4: *Top*: True velocity model. *Middle*: Initial velocity model. *Bottom*: Reconstructed velocity model ($\mathbf{m}^*$) with 100 nonlinear CG iterations.

The convergence history in terms of the objective and norm of the gradient is shown in Figure 5. The scripts and data to produce the figures for this numerical example can be found at https://github.com/izzatum/Visualizing-The-Misfit-Landscape-of-Nonlinear-Optimization-Objective-Function.
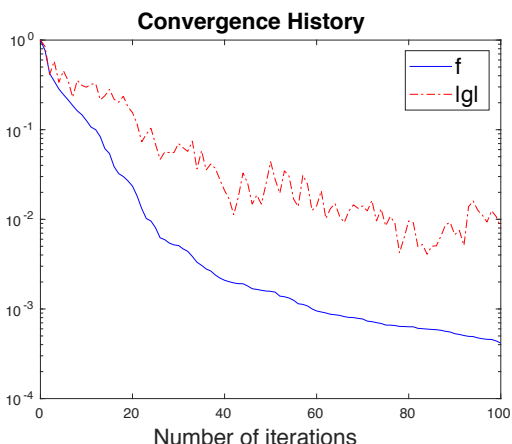

Figure 5: Convergence history plot of the misfit (blue) and the norm of gradient (red) with 100 nonlinear CG iterations.

The singular values of the error and gradient matrices are shown in figure 6. The decay gives an indication of how well the first two principal components capture the variation. The first two principal components of the error and gradient matrices are shown in 7 and 8. These two principal components of the error and gradient matrices are representing the amount of variation captured in the descent path and in the curvature, respectively. The resulting visualizations are shown in Figures 9 and 10.

In Figure 9, we can see the misfit landscape with the optimization trajectories plotted in red. We observe in the first several iterations there are significant amount of variation in optimization trajectories. This is because the optimization method is
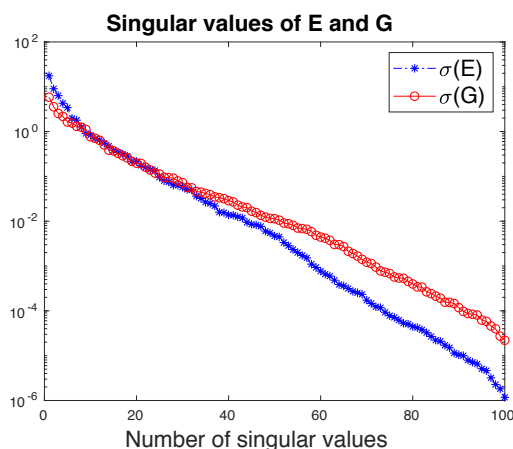

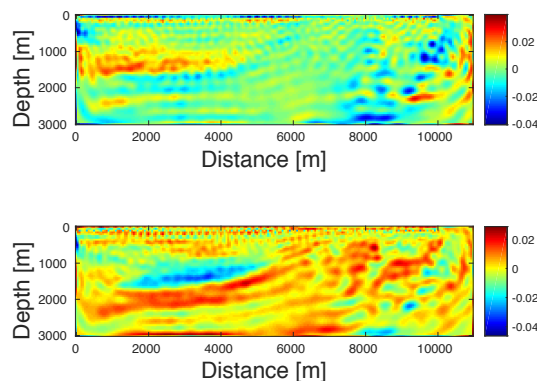Figure 6: The singular values of error matrix, $\mathbf{E}$ and gradient matrix, $\mathbf{G}$.

Figure 7: *Top*: The first principal vector of error matrix $\mathbf{E}$ acts as $\mathbf{p}_1$. *Bottom*: The second principal vector of error matrix $\mathbf{E}$ acts as $\mathbf{p}_2$

searching for the dominant descent direction in order to converge into a nearest minima. Once the dominant descent direction discovered, the error in the data misfit decreases and we can see the trajectory falls in into the local minima.
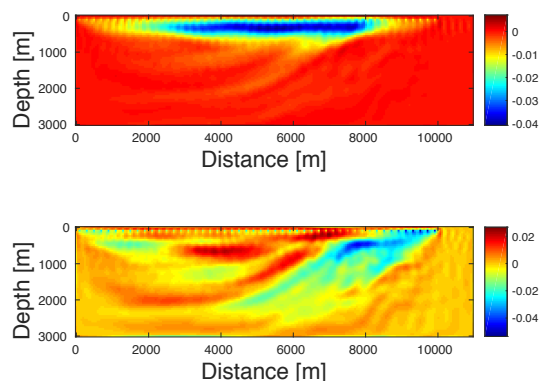

Figure 8: *Top*: The first principal vector of gradient matrix **G** acts as $\mathbf{p}_1$. *Bottom*: The second principal vector of gradient matrix **G** acts as $\mathbf{p}_2$

Meanwhile, in Figure 10 we illustrate the misfit landscape constructed from the gradient matrix in which it captures the variation of curvature in model space at each optimization iterations. We observe the same situation as in Figure 9. The trajectories illustrated here shows that the optimization method is moving towards the dominant direction of variation in the curvature. Based on Figures 9 and 10, we see that the variation in the descent path and the model space curvature are related to each other for the optimization method to converge.
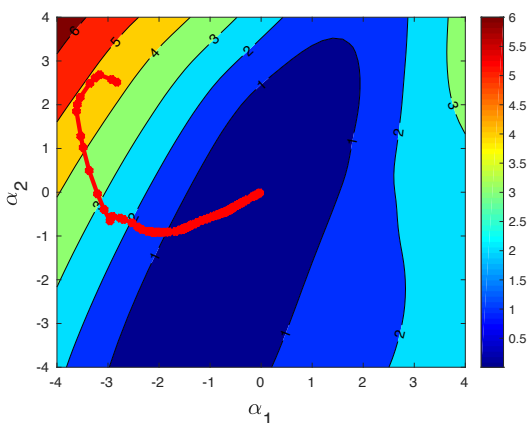

Figure 9: The contour of Marmousi model example misfit landscape along the optimization trajectories constructed with principal component vectors of error matrix.

CONCLUSION

As illustrated in the numerical examples, we observe that in low-dimensional spaces, the misfit landscape is convex. However, we should be careful in interpreting the results because even the misfit landscape is convex in the low-dimensional subspace, it is not necessarily convex in the original high-dimensional space. In order to properly analyze the convexity of the problem, we need to perform Hessian analysis (Goodfellow et al., 2015; Keskar et al., 2016; Im et al., 2016; Li et al., 2018) at each iteration. However, this analysis is not trivial due to computational limitations and it could be a future research work.
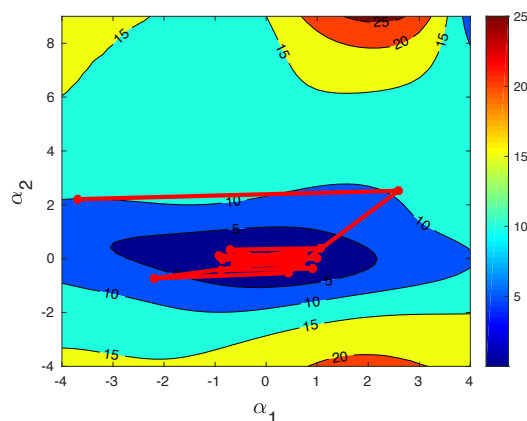

Figure 10: The contour of Marmousi model example misfit landscape along the optimization trajectories constructed with principal component vectors of gradient matrix.

To conclude, we demonstrated the visualization of the misfit landscape in a low-dimensional subspace. This visualization may aid us in further understanding our objective function and compare optimization methods.

ACKNOWLEDGMENTS

REFERENCES

Chong, E. K. P., and S. H. Zak, 2013, An introduction to optimization: Wiley.

Constantine, P. G., 2015, Active subspaces: Emerging ideas for dimension reduction in parameter studies: SIAM.

Goodfellow, I., O. Vinyals, and A. Saxe, 2015, Qualitatively characterizing neural network optimization problems: Presented at the International Conference on Learning Representations, arxiv:1412.6544.

Im, D. J., M. Tao, and K. Branson, 2016, An empirical analysis of deep network loss surfaces: Presented at the International Conference on Learning Representations, arxiv:1612.04010.

Keskar, N. S., D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, 2016, On large-batch training for deep learning: Generalization gap and sharp minima: Presented at the International Conference on Learning Representations, arXiv:1609.04836.

Li, H., Z. Xu, G. Taylor, C. Studer, and T. Goldstein, 2018, Visualizing the loss landscape of neural nets: Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS, https://doi.org/http://papers.nips.cc/paper/7875-visualizing-the-https://doi.org/ loss-landscape-of-neural-nets.

Mulder, W. A., and R.-E. Plessix, 2008, Exploring some issues in acoustic full waveform inversion: Geophysical Prospecting, **56**, 827–841, doi: https://doi.org/10.1111/j.1365-2478.2008.00708.x.

Gerhard Pratt, R., C. Shin, and G. J. Hick, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362, doi: https://doi.org/10.1046/j.1365-246X.1998.00498.x.

Russi, T. M., 2010, Uncertainty quantification with experimental data and complex system models: Ph.D. thesis, University of California.

Saad, Y., 2009, Iterative methods for sparse linear systems: SIAM.

Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics, **20**, 219–232, doi: https://doi.org/10.1029/RG020i002p00219.

van Leeuwen, T., 2019, Simple FWI, https://doi.org/https://github.com/tristanvanleeuwen/simplefwi.

Virieux, J., R. Brossier, W. Zhou, L. Metivier, A. Ribodetti, and A. Asnaashari, 2017, 6. An introduction to full waveform inversion, *in* G. Vladimir and K. Wapenaar, eds., Encyclopedia of Exploration Geophysics: SEG, R1-1–R1-40, doi: https://doi.org/10.1190/1.9781560803027.entry6.

Virieux, J., and S. Operto, 2009, An overview of FWI in Exploration Geophysics: Geophysics, **74**, no. 6, WCC127–WCC152, doi: https://doi.org/10.1190/1.3238367.

Yang, P., 2014, A numerical tour of wave propagation, http://www.ahay.org/rsf/book/xjtu/primer/paper.pdf.