A misfit function tolerating inconsistent data

*Yunsong Huang**
King Abdullah University of Science and Technology

SUMMARY

Inversion methods that do not model all of the physical events in the observed data lead to inconsistent sets of equations. This often results in serious artifacts in the inverted model. To mitigate this problem, I propose a misfit function that partly ignores the unexplainable arrivals in the data due to modeling limitation. This is tested on least squares migration of synthetic P–P arrivals, albeit P–S reflections exist in the observed data in addition, to demonstrate that in this case the new misfit is more robust than the standard L_2 misfit.

INTRODUCTION

The choice of misfit functionals is crucial for a successful application of full waveform inversion (FWI) (Lailly, 1984; Tarantola, 1984, 2005). Studies on misfit functionals may be divided into two categories. The first category, replaces the standard misfit functional with one associated with 'skeletonized' data (Luo and Schuster, 1991) that depend quasi-linearly on the velocity models. In the second category, the misfit functional measuring the L_2 norm of the data error, which is the L_2 distance between the predicted data and the observed data, is replaced with robust alternatives. Crase et al. (1990) investigated a series of robust misfit functions, such as the L_1 criterion and the Cauchy criterion. This work was extended by Brossier et al. (2010) for elastic frequency-domain FWI. With a robust misfit function, the inversion process is less sensitive to seismic noise.

This abstract belongs to the second category, except that it addresses the problem of tolerating unexplainable arrivals in the data.. As is often the case, with limitations in the computational resources and perhaps also in the theory, the numerical simulation does not take into account all of the physics contributing to the observed data. Consequently, some parts of the data may appear extraneous, or inconsistent, with respect to the numerical simulations. For example, P–S reflections are in the observed data but only P–P arrivals are modeled in the numerical simulation. In trying to minimize the L_2 distance between the observed and the predicted data, the numerical inversion will generate an artificial model as a compromise to fit both the correctly modeled data (P–P in this example) and the extraneous data (P–S in this example). The final model will be corrupted with serious artifacts.

One way to mitigate this problem is to relax the L_2 criterion, in that if the numerical simulation has a difficulty in fitting unexplainable parts of the observed data, then the unexplainable arrivals should play a weaker role than do the explainable parts. The criterion measuring the fitting difficulty is when the predicted arrival is nearly zero despite a non-zero observed event at this arrival time. This weakening of the role can be accomplished by reducing the misfit gradient with respect to the prediction associated with the unexplainable observation. With a weaker gradient, these predicted data components are less compelled to update. This property of the misfit function is dubbed 'tolerating inconsisent data' (TID).

The TID misfit is tested with least-squares migration[1] (LSM) (Nemeth et al., 1999; Duquet et al., 2000) on a synthetic data set consisting of both P–P and P–S waves, while only P–P modeling is enabled in the LSM operators. This test suggests the robustness of the TID misfit function.

THEORY

Expressions of TID misfit functions will be given. But firstly, the blue curve in Figure 1 (upper right panel) serves to convey the intuition. The function $J(d^{pred})$ has a global minimum at $d^{pred} = d^{obs}$ and that the gradient of $J_{TID}(d^{pred})$ is weaker when d^{pred} lies between 0 and d^{pred}. Here, d^{pred} represents a component of the predicted data, and d^{obs} represents the corresponding component of the observed data. Since it is assumed to be a known parameter, d^{obs} does not explicitly appear as an argument of $J(\cdot)$.

There are three points worth noting before we begin the derivation. First, we shall restrict our attention to the misfit function as a function of $\mathbf{d}^{pred}$. Later, to derive the Fréchet derivative with respect to the trial model $\mathbf{m}$, the dependence of $\mathbf{d}^{pred}$ on $\mathbf{m}$ is employed in the chain rule. If this dependence is given by

$$\mathbf{d}^{pred} = \mathbf{L}\mathbf{m}, \tag{1}$$

an exercise in calculus leads to the Fréchet derivative

$$\nabla_{\mathbf{m}} J(\mathbf{d}^{pred}) = \mathbf{L}^T \nabla_{\mathbf{d}^{pred}} J(\mathbf{d}^{pred}). \tag{2}$$

Because $\mathbf{d}^{obs}$ must be a minimum of the misfit function $J(\mathbf{d}^{pred})$, we have the gradient equal to zero at the global minimum:

$$\nabla_{\mathbf{d}^{pred}} J(\mathbf{d}^{pred})|_{\mathbf{d}^{pred}=\mathbf{d}^{obs}} = 0. \tag{3}$$

Suppose the gradient can be expressed as

$$\nabla_{\mathbf{d}^{pred}} J(\mathbf{d}^{pred}) = \mathbf{Q}(\mathbf{d}^{pred} - \mathbf{d}^{obs}), \tag{4}$$

where $\mathbf{Q}$ is a diagonal matrix (to be defined later in equation 11) and $\mathbf{Q}$ depends on $\mathbf{d}^{pred}$. Inserting this equation into equation 2 yields

$$\nabla_{\mathbf{m}} J(\mathbf{d}^{pred}) = \mathbf{L}^T \mathbf{Q}(\mathbf{d}^{pred} - \mathbf{d}^{obs}). \tag{5}$$

This form allows the implementation of the optimization method known as Iteratively Reweighted Least Squares (IRLS) (Scales et al., 1988). Equation 5 can be implemented with three steps:

[1] As the misfit function is TID rather than L_2, 'least-squares' in this context is an oxymoron. Nevertheless this terminology is adopted here for its popularity.

DOI http://dx.doi.org/10.1190/segam2014-1470.1

1) compute the data error $\mathbf{Lm}-\mathbf{d}^{obs}$, 2) weight the data error by $\mathbf{Q}$, and finally 3) migration. In the TID approach, $\mathbf{Q}$ serves to reduce the gradient when desired, by weighting down components of the $\mathbf{Lm}-\mathbf{d}^{obs}$. This contrasts with the L_2 misfit, where $\mathbf{Q}$ remains as the identity matrix.

Second, the misfit function of a vector argument is defined through the function of each component of the vector. Recall the conventional squared misfit function of a vector argument, given as

$$\begin{aligned}\chi(\mathbf{d}^{pred}) &= ||\mathbf{d}^{pred}-\mathbf{d}^{obs}||^2 \\ &= \sum_i \chi(d_i^{pred}) = \sum_i (d_i^{pred}-d_i^{obs})^2. \end{aligned} \tag{6}$$

Likewise, the TID misfit function can be expressed as

$$J(\mathbf{d}^{pred}) = \sum_i J(d_i^{pred}). \tag{7}$$

For notational economy, we reuse the name of the function J.

Third, the function $J(\mathbf{Lm})$ is convex in $\mathbf{m}$ if $J(d)$ is convex in d. This can be seen by computing the Hessian matrix of $J(\mathbf{Lm})$ with respect to $\mathbf{m}$, given as

$$\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}J(\mathbf{Lm}) = \mathbf{L}^T\nabla_{\mathbf{d}}\nabla_{\mathbf{d}}J(\mathbf{d})\mathbf{L}, \tag{8}$$

$$= \mathbf{L}^T\mathbf{H}\mathbf{L}, \tag{9}$$

where $\mathbf{H}$ is a diagonal matrix with $H_{ii} = \partial^2 J(d_i)/\partial d_i^2$. If $J(d_i)$ is convex, then $\mathbf{H}$ is positive semi-definite, and it follows from equation 9 that $\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}J(\mathbf{Lm})$ is positive semi-definite. Therefore $J(\mathbf{Lm})$ is convex in $\mathbf{m}$.

Next, we proceed to construct $J(d_i^{pred})$ by a convex piece-wise polynomial continuous in its derivative. Suppose $d_i^{obs} > 0$, we construct $J(d_i^{pred})$ as

$$J(x) = \begin{cases} (x-h)^2/2 & \text{if } x > (1-\alpha)h, \\ -\alpha h x + (\alpha - \frac{\alpha^2}{2})h^2 & \text{if } 0 < x \le (1-\alpha)h, \\ (x-\alpha h)^2/2 + (\alpha-\alpha^2)h^2 & \text{if } x \le 0. \end{cases} \tag{10}$$

Here, $x = d_i^{pred}$, $h = d_i^{obs}$, and $0 \le \alpha \le 1$ is a parameter. The $J(x)$ for $x \in [0,(1-\alpha)h]$ is replaced by an affine linear function. When $\alpha \to 1$, this replacement is minimal, whereas if $\alpha \to 0$, this replacement effect is the maximal. For $d_i^{obs} \le 0$, the $J(d_i^{pred})$ can be likewise constructed.

By defining a q as $\partial J(x)/\partial x = q(x-h)$, we find

$$q = \begin{cases} 1 & \text{if } x > (1-\alpha)h, \\ \frac{\alpha h}{h-x} & \text{if } 0 < x \le (1-\alpha)h, \\ \frac{x-\alpha h}{x-h} & \text{if } x \le 0. \end{cases} \tag{11}$$

This q is the ith diagonal of the $\mathbf{Q}$ introduced earlier. We see that if $0 < x \le (1-\alpha)h$, then $q = \frac{\alpha h}{h-x} < \frac{\alpha h}{h-(1-\alpha)h} = 1$, suggesting a reduction of the gradient magnitude. This regime is where the TID condition is met, because while the predicted data component x is closer to 0, compared to the observed data component h.

A TOY EXAMPLE OF DATA INCONSISTENCY

How the TID misfit copes with inconsistent data is now illustrated with a toy example, and compared to the behaviors of alternative misfit functions.

Take a model $\mathbf{m}$ and a forward modeling operator $\mathbf{L}$ as

$$\mathbf{m} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} 0.9 & 0.5 \\ -0.9 & 0.5 \\ 0.5 & 0.9 \\ 0.7 & -1.5 \end{bmatrix}. \tag{12}$$

Then the consistent data would be

$$\mathbf{d} = \mathbf{L}\,\mathbf{m} = \begin{bmatrix} 2.3 \\ -1.3 \\ 1.9 \\ -0.1 \end{bmatrix}. \tag{13}$$

Make the fourth data component inconsistent by setting it to 1, and we have the observed data:

$$\mathbf{d}^{obs} = \begin{bmatrix} 2.3 \\ -1.3 \\ 1.9 \\ 1 \end{bmatrix}. \tag{14}$$

Three types of misfit functions L_2, L_1 and TID are examined, by plotting the misfit values in a 2D space of trial model vectors, as shown in Figures 1. (The 1D functions are plotted in the upper right panel.) Here, the minimization of the TID misfit function leads to the proximity of the actual model. The reconstructed models in the cases of both L_2 and L_1 misfit functions, however, are heavily affected by the equation in the last row of $\mathbf{L}\,\mathbf{m} = \mathbf{d}^{obs}$, because this row of $\mathbf{L}$ has the largest norm. In contrast, in the TID approach, the last data component, which is incongruous with the other three, is partly ignored.

A SYNTHETIC EXAMPLE WITH LSM

To demonstrate the effectiveness of the proposed TID misfit, I test it on simplistic synthetic data, where the observed data contain both P and S waves while the forward modeling and migration operators of LSM only model P–P reflections. The model size is 1.6 km in depth $\times$ 3.2 km in width, with 16 sources and 160 receivers evenly spaced, respectively, on the surface. The background homogeneous velocity model has $v_p = 2.28$ km/s and $v_s = 1.31$ km/s, and the source time history is a Ricker wavelet with a peak frequency at $f_0 = 13$ Hz. The P–P reflectivity model $\mathbf{m}_P$ is shown in Figure 2(a). In addition, P-to-S conversion occurs at the reflectors. For simplicity, the P–S reflection coefficients are assumed to be half of the P–P reflectivity coefficients. Consequently, the observed data $\mathbf{d}^{obs}$, generated by Kirchhoff modeling, contains both P and S waves, written as $\mathbf{d}_P^{obs}$ and $\mathbf{d}_S^{obs}$, respectively.

The LSM images, obtained after 25 iterations, computed from the conventional least-squares misfit and the proposed TID misfit are shown in Figures 2(b) and (c), respectively. The latter image exhibits improvements over the former.

DOI http://dx.doi.org/10.1190/segam2014-1470.1

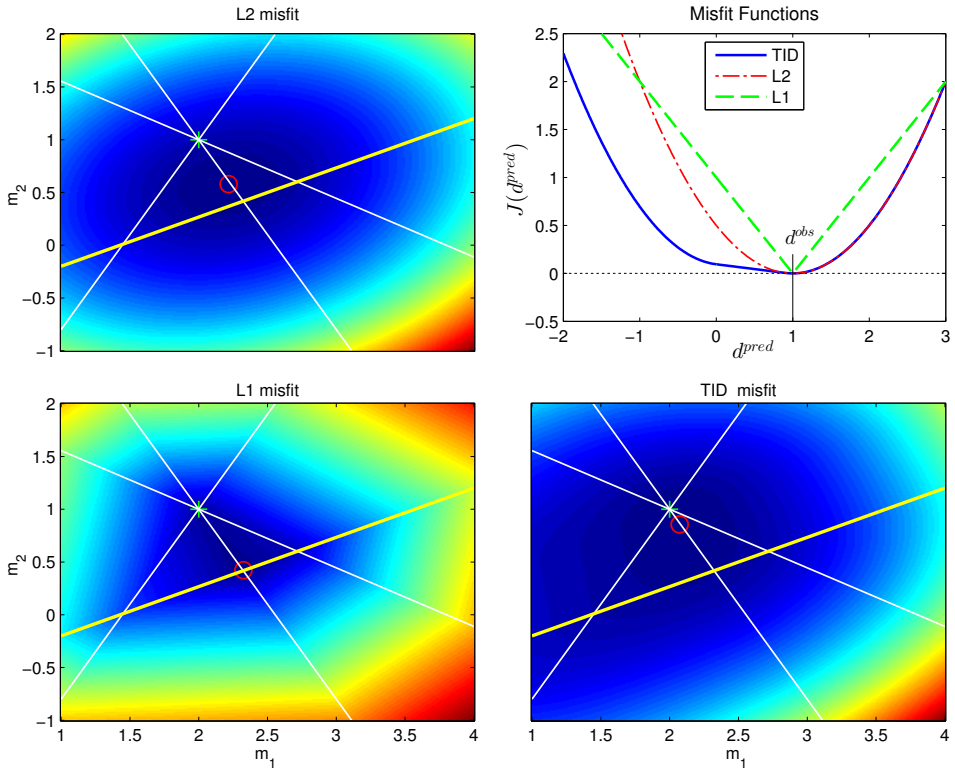

Figure 1: Shown over trial models of 2D $\mathbf{m} = [m_1, m_2]^T$ are the misfit functions: (upper left) L_2, (lower left) L_1, and (upper right) TID. Each line in white or yellow represents an equation $[L_{i1}, L_{i2}]\mathbf{m} = d_i$, for $i = 1, \ldots, 4$. The yellow line corresponds to the last equation, i.e., $i = 4$. Green '+' denotes the actual model responsible for generating the data (except for the inconsistent 4^{th} data component). Individual red '$\circ$' denotes the best inverted model to minimize each case of misfit function. (Upper right) Plot the TID, L_2, and L_1 misfit functions over a 1D predicted data component d^{pred}, when the observed counterpart is at $d^{obs} = 1$.

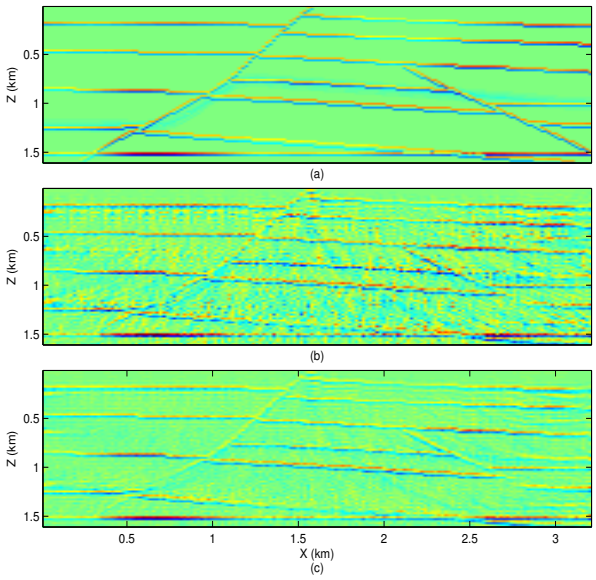

Figure 2: P–P reflectivity models, of (a) the actual model, and (b & c) inverted by LSM, using the (b) L_2 and (c) TID misfit functions. Here, the input data consists of both PP and PS reflections, but the modeling can only model PP arrivals.

The standard LSM method with L_2 misfit strives to fit $\mathbf{d}_S^{obs}$, as well as to $\mathbf{d}_P^{obs}$, by forward modeling with a P–P reflectivity model. In the standard approach, the P–P modeler has to fit the P–S arrivals, but in the proposed approach this fitting will be discouraged should the TID conditions be met. Figure 3(d) shows the predicted P–S arrivals using the standard LSM method. That is, the best fit model with a P–P modeler tends to introduce noisy artifacts (see Figure 2(b)) into the inverted reflectivity model in order to explain, as best as this method can, the P–S arrivals. In contrast, Figure 3(f) shows that the TID prediction is less concerned with fitting the P–S arrivals. Consequently, the inverted reflectivity model, shown in Figure 2(c), is less susceptible to artifacts.

DISCUSSIONS AND CONCLUSIONS

This abstract introduces a misfit function that can selectively diminish some of the unexplainable parts of the data. Parts of the data are considered unexplainable or inconsistent if the prediction for them is nearly 0 even though the corresponding observed parts are non-zero. This happens when inversion methods do not model all of the physical events in the observed data. The selective diminishing acts on the gradient with respect to the prediction, because reducing such a gradi-

DOI http://dx.doi.org/10.1190/segam2014-1470.1

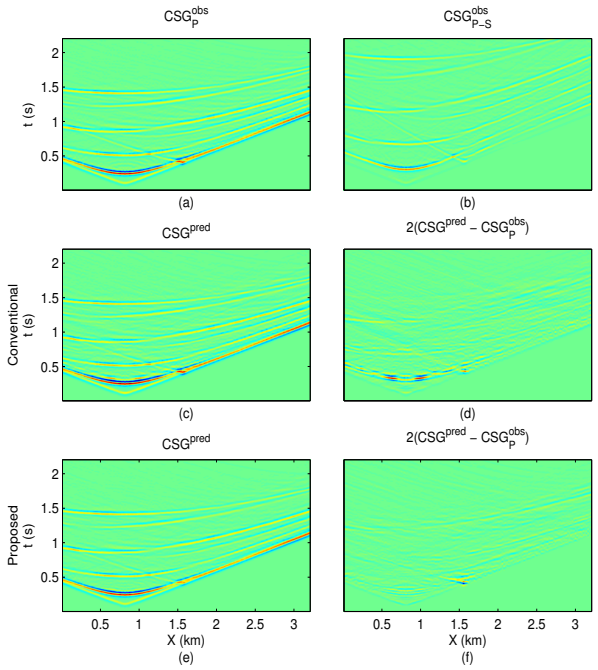

Figure 3: An example of observed CSG consisting of (a) P and (b) P–S reflections. The latter is not modeled in (c–f), implying data inconsistency. (c & e) Predicted CSG of P reflections from inverted models, (d & f) prediction residual of P reflections, using (c & d) L_2 and (e & f) TID misfits. All panels use the same dynamic range of plotting.

ent is equivalent to partly excluding the corresponding terms of the prediction error from the misfit function. This prescribed action on the gradient is accomplished by the TID misfit function. An example of data inconsistency is tested to verify that the proposed misfit function is advantageous compared to the conventional L_2 misfit.

A potential limitation of the proposed approach is that the misfit function has no expert knowledge to distinguish extraneous arrivals from modeled arrivals in the observed data. Future work will aim towards incorporating expert knowledge in the inversion through the TID misfit function. Expert knowledge can be embedded as a soft muting. The more confident the expert decision, the stronger is the muting. This soft muting exerts a preference that the TID misfit function can take advantage of, and thus helps to clarify whether the TID condition is met for certain arrivals.

ACKNOWLEDGMENTS

I wish to thank the sponsors of Center for Subsurface Imaging and Fluid Modeling (CSIM) at KAUST for their financial support.

DOI http://dx.doi.org/10.1190/segam2014-1470.1

http://dx.doi.org/10.1190/segam2014-1470.1

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2014 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Brossier, R., S. Operto, and J. Virieux, 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: Geophysics, **75**, no. 3, R37–R46, http://dx.doi.org/10.1190/1.3379323.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538, http://dx.doi.org/10.1190/1.1442864.

Duquet, B., K. J. Marfurt, and J. Dellinger, 2000, Kirchhoff modeling, inversion for reflectivity, and subsurface illumination: Geophysics, **65**, 1195–1209, http://dx.doi.org/10.1190/1.1444812.

Lailly, P., 1984, Migration methods: Partial but efficient solutions to the seismic inverse problem, *in* F. Santosa, ed., Inverse Problems of Acoustic and Elastic Waves: Society for Industrial and Applied Mathematics, 182–212.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653, http://dx.doi.org/10.1190/1.1443081.

Nemeth, T., C. Wu, and G. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221, http://dx.doi.org/10.1190/1.1444517.

Scales, J., A. Gersztenkorn, and S. Treitel, 1988, Fast l_p solution of large, sparse, linear systems: Application to seismic travel time tomography: Journal of Computational Physics, **75**, 314–333, http://dx.doi.org/10.1016/0021-9991(88)90115-5.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, http://dx.doi.org/10.1190/1.1441754.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society for Industrial and Applied Mathematics, http://dx.doi.org/10.1137/1.9780898717921.