WS09 D01

Mitigating the gradient artefacts of Migration Velocity Analysis by Gauss-Newton update

R. Soubaras* (CGG), B. Gratacos (CGG)

Summary

This paper shows how the artefacts present in the gradient of the cost function when performing a Migration Velocity Analysis can be strongly attenuated by using a second-order Gauss-Newton scheme. The artefacts on the velocity gradient are strongly attenuated when the total gradient is deconvolved by the approximate total Hessian. At each iteration, a least-squares migration is produced rather than a migration. The proposed algorithm is illustrated on the Marmousi dataset.

Introduction

Full waveform inversion (FWI) is a method of velocity estimation which minimizes the misfit between recorded and modeled data, the parameters of the minimization being the velocity model. If the velocity model is smooth, then only the refracted waves are modeled and used, if the velocity model has discontinuities, the reflected waves created by the two-way modeling can be used to model the reflected waves. This highly non-linear algorithm can provide impressive results, in particular high resolution velocity from low-frequency data, but is very sensitive to its starting point.
Migration velocity analysis (MVA) uses a migration to estimate a reflectivity from the data, then uses a criterion on the reflectivity to find the best velocity model. This method is not as non linear as FWI: it is less sensitive to the starting model, but has less potential to find a very detailed velocity from low-frequency data.
In recent years, FWI has made a lot of progress to solve its inherent problems, and MVA techniques have been lagging behind. The main reason is that the gradient of the MVA cost functions exhibits artefacts that perturb the convergence.

Migration Velocity Analysis: theoretical justification and encountered problems

Let's first introduce the notations: d_0 is the recorded data, v is the velocity model, r is the reflectivity, and $G(v)$ is the modeling operator that maps the reflectivity r in the data space by producing the modeled data d with the linear modeling equation:

$$d = G(v)r \tag{1}$$

r and v can be estimated by minimizing the energy of the data misfit $e = d - d_0$ between the modeled data d and the recorded data d_0:

$$M(r,v) = \frac{1}{2}e^*e = \frac{1}{2}\left[G(v)r - d_0\right]^*\left[G(v)r - d_0\right] \tag{2}$$

The minimization in r can be performed first as it is a quadratic problem, giving $r = (G^*G)^{-1}G^*d_0$, and the minimized misfit as a function of v, $M(v) = d_0^*d_0 - d^*G(G^*G)^{-1}G^*d_0$. r is the least-squares migration of the data d_0. The migrated data m is defined by $m(v) = G(v)^*d_0$ and minimizing the data misfit is equivalent to maximizing:

$$E(v) = m(v)^*[G(v)^*G(v)]^{-1}m(v) \tag{3}$$

which is maximizing among all velocity models the energy of the migrated data with a $(G^*G)^{-1}$ weight. MVA avoids the step of computing $(G^*G)^{-1}$ by maximizing the energy of the migrated data with an ad hoc weighting. One possible weighting is to use the semblance (Soubaras and Gratacos (2007)). However, all these schemes suffers from "gradient artefacts". They appear as well when minimizing the energy of the data misfit, with the exact weight $(G^*G)^{-1}$, which is the optimal cost function, so the problem cannot be solved by changing the cost function. Finally, the term gradient artefacts is not well-suited because they do not come from the gradient being wrongly computed: they are really part of the exact gradient.

A Gauss-Newton scheme for the penalized data misfit cost function

In conclusion, the energy of the data misfit is the optimal cost function but its gradient cannot be directly utilized as a velocity model update, hence all first order schemes are ruled out. We proceed therefore to use a second-order scheme.
The advantage of the Gauss-Newton scheme is that it is second-order but doesn't need the computation of any second order derivative, under the assumption that the data misfit is small. In order to ensure this even for the first iteration, we use an extended space for the reflectivity r. Indeed, the extended space (Symes (2008)) gives extra degrees of freedom to match the data with a wrong velocity model, these extra degrees of freedom being controlled by a term $||Ar||^2$ added to the energy of the data misfit that penalizes unfocused energy. We therefore use the following cost function, r being now the extended reflectivity:

$$C(r,v) = M(r,v) + \frac{\sigma}{2}||Ar||^2 \tag{4}$$

where $||Ar||^2$ is the penalization term for the extended reflectivity and $M(r,v)$ the data misfit (equation (2)). The gradient and Hessian in r of M are:

$$\frac{\partial M}{\partial r^*} = \frac{\partial e^*}{\partial r^*} e = G^* e, \qquad \frac{\partial^2 M}{\partial r^* \partial r} = G^* G \tag{5}$$

The gradient and Hessian in v of M are:

$$\frac{\partial M}{\partial v^T} = \frac{\partial e^*}{\partial v^T} e = \frac{\partial (Gr)^*}{\partial v^T} e, \qquad \frac{\partial^2 M}{\partial v^T \partial v} = \frac{\partial e^*}{\partial v^T}\frac{\partial e}{\partial v} + \frac{\partial^2 e^*}{\partial v^T \partial v} e = \frac{\partial (Gr)^*}{\partial v^T}\frac{\partial (Gr)}{\partial v} + \frac{\partial^2 e^*}{\partial v^T \partial v} e \tag{6}$$

The Gauss-Newton method consists in using an approximate Hessian by neglecting the second term $\frac{\partial^2 e^*}{\partial v^T \partial v} e$. The justification is that this term is small as long as the data misfit e is small. Also, the resulting approximate Hessian can be computed by first derivatives only and is guaranteed to be a positive matrix. The cross term of the Hessian, assuming again e is small, is:

$$\frac{\partial^2 M}{\partial v^T \partial r} = \frac{\partial e^*}{\partial v^T}\frac{\partial e}{\partial r} = \frac{\partial (Gr)^*}{\partial v^T} G \tag{7}$$

The total gradient in r and v is, adding the gradient of the penalty term:

$$g = \begin{bmatrix} \frac{\partial C}{\partial r^*} \\ \frac{\partial C}{\partial v^T} \end{bmatrix} = \begin{bmatrix} G^* \\ \frac{\partial (Gr)^*}{\partial v^T} \end{bmatrix} e + \sigma \begin{bmatrix} A^* A r \\ 0 \end{bmatrix} \tag{8}$$

and the total Hessian in r and v with the penalty term is:

$$H = \begin{bmatrix} \frac{\partial^2 C}{\partial r^* \partial r} & \frac{\partial^2 C}{\partial r^* \partial v} \\ \frac{\partial^2 C}{\partial v^T \partial r} & \frac{\partial^2 C}{\partial v^T \partial v} \end{bmatrix} = \begin{bmatrix} G^* \\ \frac{\partial (Gr)^*}{\partial v^T} \end{bmatrix} \begin{bmatrix} G & \frac{\partial (Gr)}{\partial v} \end{bmatrix} + \sigma \begin{bmatrix} A^* A & 0 \\ 0 & 0 \end{bmatrix} \tag{9}$$

which can be computed by the first derivatives of $G(v)$ only and has no dependence on the data d_0. The Gauss-Newton scheme consists in updating simultaneously the velocity and the reflectivity by:

$$\begin{bmatrix} r_{n+1} \\ v_{n+1} \end{bmatrix} = \begin{bmatrix} r_n \\ v_n \end{bmatrix} - \lambda_n H_n^{-1}(r_n, v_n) g_n(r_n, v_n) \tag{10}$$

where g_n is the total gradient which can be computed by the adjoint-state method, and $h_n = H_n^{-1} g_n$ the total deconvolved gradient which can be obtained by solving $H_n h_n = g_n$ with a linear conjuguate gradient algorithm, applying the Hessian H to a vector h being done by cascading a "direct-state" method and an adjoint state method. λ_n can be determined by a line-search after h_n is computed.
Because r_n minimizes the energy of the data misfit, it is the least-squares migration of the data d_0 for the velocity v_n. Using a second order Gauss-Newton scheme therefore yields an Inversion Velocity Analysis (IVA) scheme (Liu et al. (2014), Chauris et al. (2015)) where each iteration provides a velocity model and the least square migration associated with that model, instead of the migration provided by MVA.

Synthetic example

A one-way wave-equation propagation was used for the operator $G(v)$. Starting from the Marmousi velocity model, with a 1 km water layer on top of it, the shots were recomputed on a sparse shot acquisition, shown in Figure 1: 24 shots separated by 240 m, 96 receivers, 192 m minimum offset, 2472 m maximum offset. The minimum and maximum frequencies are 2.5 Hz and 30 Hz. The sparse interval between shots makes the gradient artefacts more visible. The initial velocity model is a simple 1D gradient. The velocity computed with a maximum frequency $f_{max} = 10$ Hz is shown in Figure 2. It is quite good until 2.5km. The velocity part of the gradient computed at that point with $f_{max} = 20$ Hz is shown in Figure 3. The gradients artefacts are clearly visible. The gradient shows the deep high velocity layers on the left and right of the section, which should be part of the update because they cannot be seen on the velocity model. The gradient artefacts are the wavepaths that propagate this update up to the surface. The problem of using this gradient as update is that it damages the shallow velocity which is correct. The velocity part of the deconvolved gradient is shown in Figure 4. The gradient artefacts can be seen to be strongly attenuated and only the desired features remain, so that Figure 4 is a much better update for Figure 2 than Figure 3. The final velocity model, obtained with $f_{max} = 30$ Hz, with the associated reflectivity is shown in Figure 5.

Conclusion

We have shown that performing MVA with an extended reflectivity and the data misfit plus penalization cost function allows the use of a second-order Gauss-Newton algorithm. The artefacts of the gradient are still present because they are real. The Hessian deconvolution performed during the Gauss-Newton update removes these artefacts, so that they are not included in the velocity update and do not perturb the convergence.

References

Chauris, H., Lameloise, C. and Cocher, E. [2015]. Inversion Velocity Analysis: the importance of regularization. *77th EAGE Conference & Exhibition* Expanded Abstracts, WS05-A02.

Liu, Y., Symes, W.W. and Li, Z. [2014]. Inversion velocity analysis via Differential Semblance Optimization. *76th EAGE Conference & Exhibition*, Expanded Abstracts, Th P02-07.

Soubaras, R., Gratacos, B. [2007]. Velocity model building by semblance maximization of modulated-shot gathers. *Geophysics*, **72**, U67-U73.

Symes, W. W. [2008]. Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, **56**, n 6, 765-790.

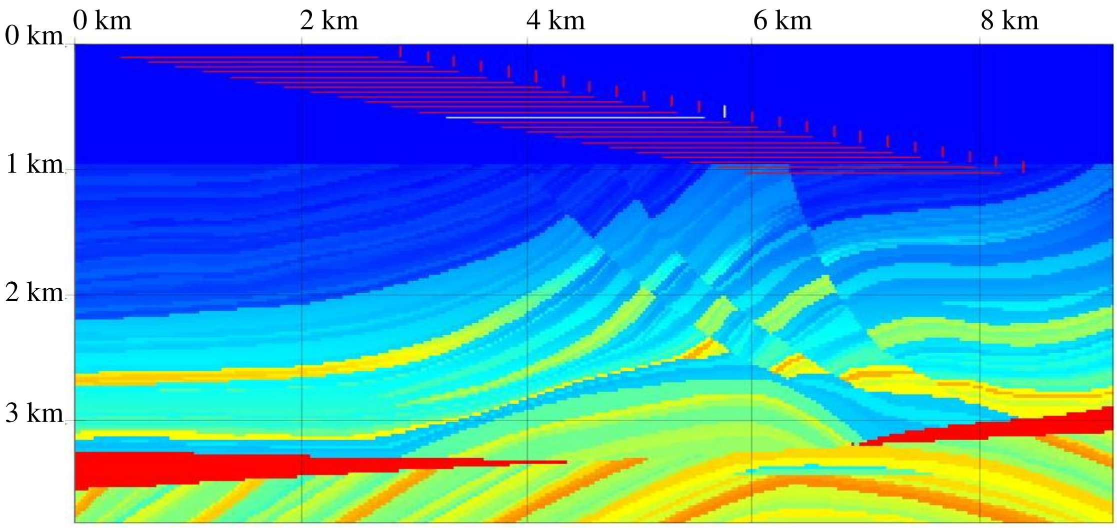

Figure 1 *Acquisition geometry and true velocity model*

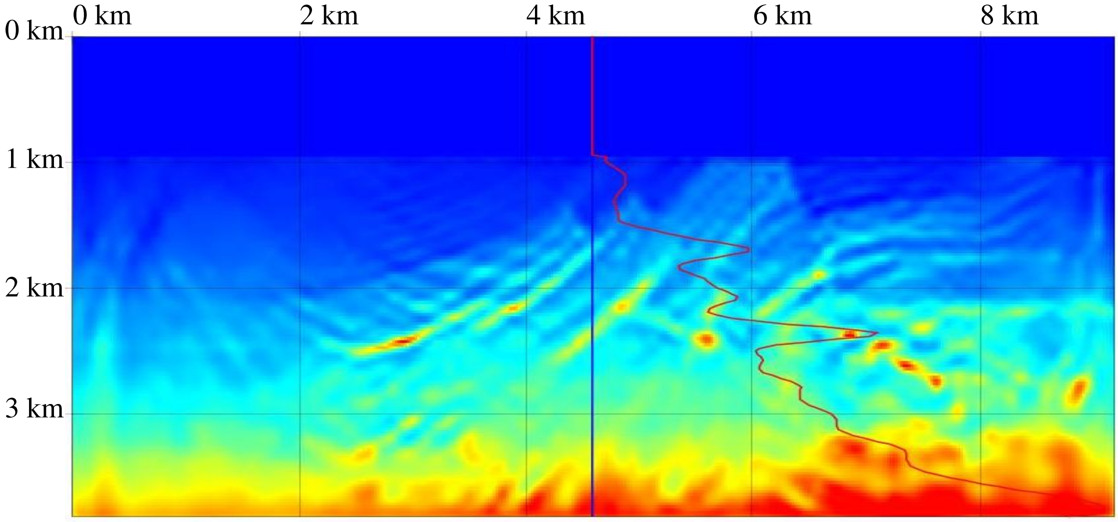

Figure 2 *Velocity model with the graph of the velocity at the location of the vertical blue line*

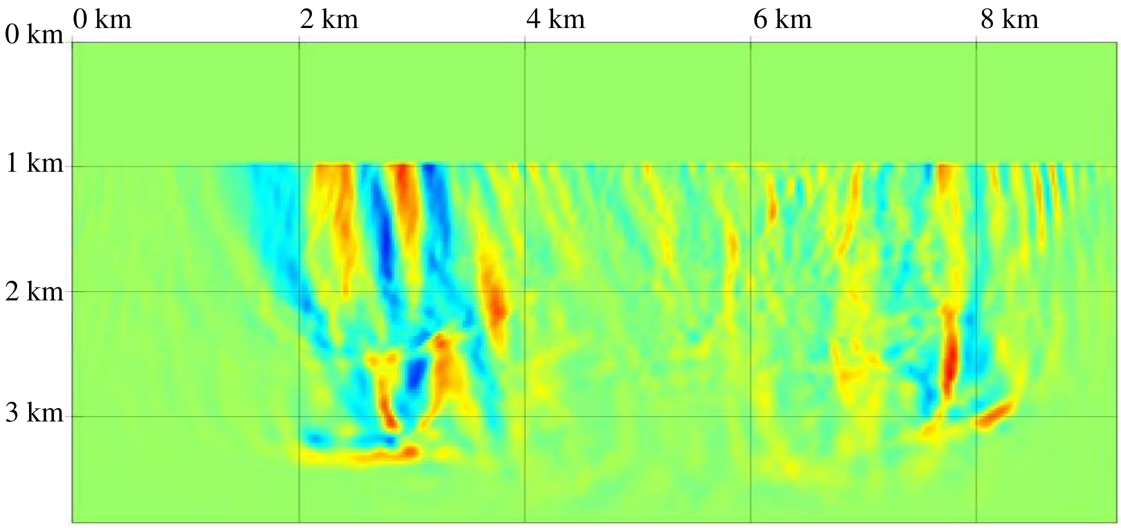

Figure 3 *Velocity part of the gradient*

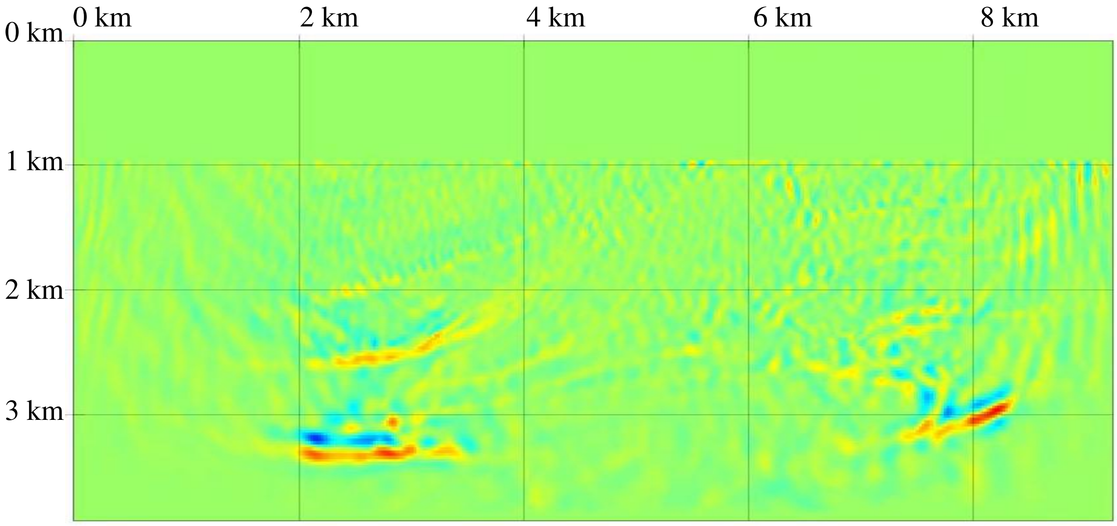

Figure 4 *Velocity part of the deconvolved gradient*

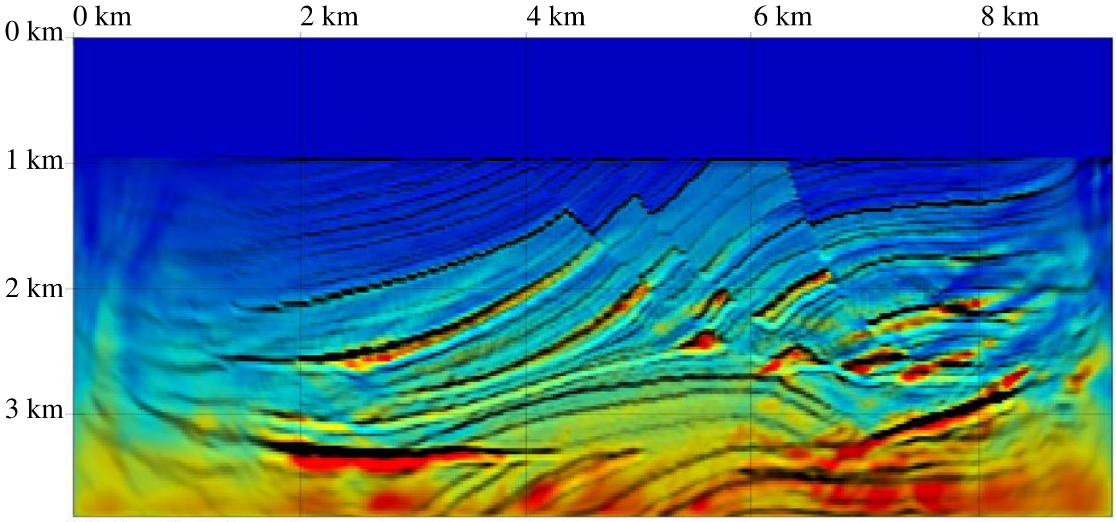

Figure 5 *Final velocity and reflectivity model*