Full Waveform Inversion

Vincent De Comarmond Institute of Mine Seismology

Full Waveform Inversion

Vincent De Comarmond
Institute of Mine Seismology

institute of mine seismology

www.imseismology.org

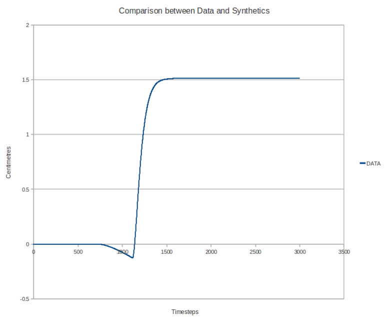
Full Waveform Inversion- Contents

Introducing Full Waveform Inversion

- **The idea of Full Waveform Inversion is quite a simple one.**

Introducing Full Waveform Inversion

- **The idea of Full Waveform Inversion is quite a simple one.**
- **Imagine that we record a controlled event .**



Using what we know

What does this model tell us about the mine?

- **We spend some time building an accurate computer model of the mine.**

Using what we know

What does this model tell us about the mine?

- **We spend some time building an accurate computer model of the mine.**
- **Once the model has been built we want to use it. However we first need to make sure that the model is reasonable.**

Using what we know

What does this model tell us about the mine?

- **We spend some time building an accurate computer model of the mine.**
- **Once the model has been built we want to use it. However we first need to make sure that the model is reasonable.**
- **The controlled event gives us the perfect opportunity to test the model of the mine.**

Using what we know

What does this model tell us about the mine?

- **We spend some time building an accurate computer model of the mine.**
- **Once the model has been built we want to use it. However we first need to make sure that the model is reasonable.**
- **The controlled event gives us the perfect opportunity to test the model of the mine.**
- **Let us then check to see how the seismograms computed by our model compare to the seismograms we recorded.**

Using what we know

What does this model tell us about the mine?

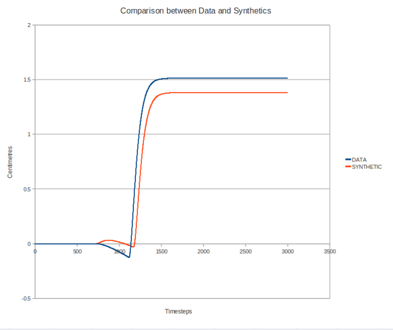
- **We spend some time building an accurate computer model of the mine.**
- **Once the model has been built we want to use it. However we first need to make sure that the model is reasonable.**
- **The controlled event gives us the perfect opportunity to test the model of the mine.**
- **Let us then check to see how the seismograms computed by our model compare to the seismograms we recorded.**
- **Henceforth the seismograms produced by the model shall be called synthetics, whilst the recorded seismograms will be called data.**

Comparing the Model to our data.

Comparing the seismograms...

- **Consider the following picture showing how the two seismograms overlap or rather don't**

Figure: The output of our model vs. what we recorded

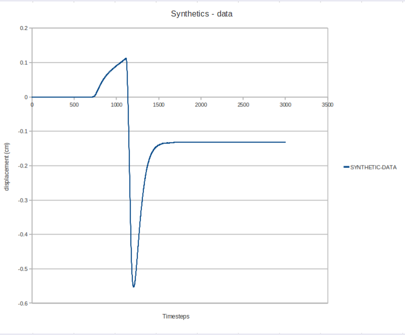


Comparing the Model to our data II

Comparing the seismograms...

- **Something more informative to look at is the difference between the data and the synthetics.**

Figure: Model waveform - Recorded waveform



Quantifying the comparison.

How do we measure the difference.

- **Perhaps we want to give our model a score. To do this all the differences need to be positive. Consider the way we measure the length between two points $\vec{p}$ and $\vec{p}_0$.**

$$(\vec{p} - \vec{p}_0)^2 = (p_x - p_{0,x})^2 + (p_y - p_{0,y})^2 + (p_z - p_{0,z})^2$$

Quantifying the comparison.

How do we measure the difference.

- **Perhaps we want to give our model a score. To do this all the differences need to be positive. Consider the way we measure the length between two points $\vec{p}$ and $\vec{p}_0$.**

$$(\vec{p} - \vec{p}_0)^2 = (p_x - p_{0,x})^2 + (p_y - p_{0,y})^2 + (p_z - p_{0,z})^2$$

- **Let us try the same thing with our seismograms.**

Quantifying the comparison.

How do we measure the difference.

- **Perhaps we want to give our model a score. To do this all the differences need to be positive. Consider the way we measure the length between two points $\vec{p}$ and $\vec{p}_0$.**

$$(\vec{p} - \vec{p}_0)^2 = (p_x - p_{0,x})^2 + (p_y - p_{0,y})^2 + (p_z - p_{0,z})^2$$

- **Let us try the same thing with our seismograms.**
- **But simply subtracting the seismograms from one another is not reasonable, as this will give a time-dependent quantity.**

Quantifying the comparison.

How do we measure the difference.

- **Perhaps we want to give our model a score. To do this all the differences need to be positive. Consider the way we measure the length between two points $\vec{p}$ and $\vec{p}_0$.**

$$(\vec{p}-\vec{p}_0)^2 = (p_x - p_{0,x})^2 + (p_y - p_{0,y})^2 + (p_z - p_{0,z})^2$$

- **Let us try the same thing with our seismograms.**
- **But simply subtracting the seismograms from one another is not reasonable, as this will give a time-dependent quantity.**
- **Let us just add the values together for all times - i.e integrate over time.**

The Cost

- **The above procedure results in the following measure for "cost" where $\vec{s}$ represents the components of the displacement seismogram for the synthetics and $\vec{d}$ represents the components of the displacement seismogram for the data.**

$$\chi = \frac{1}{2}\int_0^T \|\vec{s}(t) - \vec{d}(t)\|^2 dt$$

- **Here [0,T] is the chosen observation time.**

The Cost

- **The above procedure results in the following measure for "cost" where $\vec{s}$ represents the components of the displacement seismogram for the synthetics and $\vec{d}$ represents the components of the displacement seismogram for the data.**

$$\chi = \frac{1}{2}\int_0^T \|\vec{s}(t) - \vec{d}(t)\|^2 dt$$

- **Here [0,T] is the chosen observation time.**
- **It is quite likely that the same event is recorded at many stations, so we should allow for this possibility. Also the cost depends on the model parameters used. So the above definition can be extended.**

The Cost II

The cost of our model of the mine.

- **This leads to the following definition of cost, which measures how bad the model is.**

$$\chi[\vec{m}] = \frac{1}{2}\sum_{r=1}^{N}\int_0^T \|\vec{s}(\vec{x}_r, t, \vec{m}) - \vec{d}(\vec{x}_r, t)\|^2 dt$$

Where: $\vec{m}$ **is the model vector, containing all the parameters describing our model.**

$\vec{s}(\vec{x}_r, t, \vec{m}) =$ **The Synthetic seismograms i.e.: the output of our model**

$\vec{d}(\vec{x}_r, t) =$ **The Data i.e.: the seismograms we record.**

$[0, T] =$ **The observation time**

How good/bad is this cost?

Have we made a good choice for our cost?

- **Although the above cost is very intuitive, and it tells us exactly what we want to know it has the following problems:**

How good/bad is this cost?

Have we made a good choice for our cost?

- **Although the above cost is very intuitive, and it tells us exactly what we want to know it has the following problems:**

1. **It is not a robust measure - outliers can become dominant.**
2. **The numerical value is determined largely by the large-amplitude waveforms, this means that data contained in small amplitude P wave phase changes can easily be lost.**
3. **This measure (the L_2 norm) emphasises the non-linearity that is inherent in the seismic wave equation. This is particularly problematic for the inversion procedure used.**

Outline

1. Introducing the Idea of Full Waveform Inversion.
2. Using the costs.
   - Getting useful expressions out of the cost.
   - Understanding the Adjoint wavefield.
   - Back to using the cost.
3. Home cooked examples of the Algorithm.

How is this used?

Now that we know the cost of everything how do we proceed?

- **So in the story, we've made a model, we know that it is wrong and we know how much it costs.**

How is this used?

Now that we know the cost of everything how do we proceed?

- **So in the story, we've made a model, we know that it is wrong and we know how much it costs.**
- **We could build many models of the same mine and see which model has the lowest cost. However this would not be a very efficient way to proceed.**

How is this used?

Now that we know the cost of everything how do we proceed?

- **So in the story, we've made a model, we know that it is wrong and we know how much it costs.**
- **We could build many models of the same mine and see which model has the lowest cost. However this would not be a very efficient way to proceed.**
- **A far better way to proceed is ask how the cost changes when the model changes.**

How is this used?

Now that we know the cost of everything how do we proceed?

- **So in the story, we've made a model, we know that it is wrong and we know how much it costs.**
- **We could build many models of the same mine and see which model has the lowest cost. However this would not be a very efficient way to proceed.**
- **A far better way to proceed is ask how the cost changes when the model changes.**
- **Thus what we seek is the functional derivative $\frac{\delta\chi[m]}{\delta m}$. It is easy to see that this is given by:**

$$\delta\chi = \sum_{r=1}^{N} \int_{0}^{T} [s_i(\vec{x}_r, t, \vec{m}) - d_i(\vec{x}_r, t)] \delta s_i(\vec{x}_r, t, \vec{m})$$

Turning the wheels - the Born Approximation:

Evaluating $\delta s_i(\vec{x}_r, t, \vec{m})$

- **To evaluate $\delta s_i(\vec{x}_r, t, \vec{m})$, one needs to make use of the Born Approximation.**

Turning the wheels - the Born Approximation:

Evaluating $\delta s_i(\vec{x}_r, t, \vec{m})$

- **To evaluate $\delta s_i(\vec{x}_r, t, \vec{m})$, one needs to make use of the Born Approximation.**
- **Recall the Seismic wave equation in a source-free region:**

$$\rho \frac{\partial^2}{\partial t^2} u_i^{(0)} - \frac{\partial}{\partial x_j} c_{ijkl} \epsilon_{kl}^{(0)} = 0$$

This can we written as: $\mathcal{L}^{(0)} u_i^{(0)} = 0$

Turning the wheels - the Born Approximation:

Evaluating $\delta s_i(\vec{x}_r, t, \vec{m})$

- **To evaluate $\delta s_i(\vec{x}_r, t, \vec{m})$, one needs to make use of the Born Approximation.**
- **Recall the Seismic wave equation in a source-free region:**

$$\rho \frac{\partial^2}{\partial t^2} u_i^{(0)} - \frac{\partial}{\partial x_j} c_{ijkl} \epsilon_{kl}^{(0)} = 0$$

This can we written as: $\mathcal{L}^{(0)} u_i^{(0)} = 0$

- **The Born approximation says that a small small perturbation in the medium ($\mathcal{L}^{(0)} \rightarrow \mathcal{L}^{(0)} + \delta\mathcal{L}$) will cause a small perturbation in the wavefield ($u_i^{(0)} \rightarrow u_i^{(0)} + \delta u_i$).**

Evaluating $\delta s_i(\vec{x}_r, t, \vec{m})$

- **Thus the following equation must be satisfied:**

$$
\begin{aligned}
(\mathcal{L}^{(0)} + \delta\mathcal{L})(u_i^{(0)} + \delta u_i) &= 0 \\
\Rightarrow \mathcal{L}^{(0)}\delta u_i &= -\delta\mathcal{L}u_i^{(0)} + \mathcal{O}(\delta^2)
\end{aligned}
$$

Evaluating $\delta s_i(\vec{x}_r, t, \vec{m})$

- **Thus the following equation must be satisfied:**

$$(\mathcal{L}^{(0)} + \delta\mathcal{L})(u_i^{(0)} + \delta u_i) = 0$$
$$\Rightarrow \mathcal{L}^{(0)}\delta u_i = -\delta\mathcal{L}u_i^{(0)} + \mathcal{O}(\delta^2)$$

- **And hence** $\delta u_i = -(\mathcal{L}^{(0)})^{-1}\delta\mathcal{L}u_i^{(0)} + \mathcal{O}(\delta^2)$ **.**

Evaluating $\delta s_i(\vec{x}_r, t, \vec{m})$

- **Thus the following equation must be satisfied:**

$$(\mathcal{L}^{(0)} + \delta\mathcal{L})(u_i^{(0)} + \delta u_i) = 0$$
$$\Rightarrow \mathcal{L}^{(0)}\delta u_i = -\delta\mathcal{L}u_i^{(0)} + \mathcal{O}(\delta^2)$$

- **And hence** $\delta u_i = -(\mathcal{L}^{(0)})^{-1}\delta\mathcal{L}u_i^{(0)} + \mathcal{O}(\delta^2)$ **.**
- **Further , as we can solve the seismic wave equation (at least numerically), we have :**

$$\delta s_i(\vec{x}, t) = -\int_0^t \int_V [\delta\rho(\vec{x}')G_{ij}(\vec{x}, \vec{x}'; t - t')\frac{\partial^2}{\partial t'^2}s_j(\vec{x}', t')$$
$$+\delta c_{jklm}(\vec{x}')\frac{\partial}{\partial x'_k}G_{ij}(\vec{x}, \vec{x}'; t - t')\frac{\partial}{\partial x'_l}s_m(\vec{x}', t')]\, d^3\vec{x}'\, dt'$$

Putting everything together.

What do we find when putting everything together?

- **Putting everything together one finds:**

$$\delta\chi = \sum_{r=1}^{N} \int_0^T [s_i(\vec{x}_r, t, \vec{m}) - d_i(\vec{x}_r, t)] \int_0^t \int_V [\delta\rho(\vec{x}')G_{ij}(\vec{x}_r, \vec{x}'; t - t') \frac{\partial^2}{\partial t'^2} s_j(\vec{x}', t')$$

$$+\delta c_{jklm}(\vec{x}') \frac{\partial}{\partial x'_k} G_{ij}(\vec{x}_r, \vec{x}'; t - t') \frac{\partial}{\partial x'_l} s_m(\vec{x}', t')] d^3\vec{x}' dt'$$

Putting everything together.

What do we find when putting everything together?

- **Putting everything together one finds:**

$$\delta\chi = \sum_{r=1}^{N} \int_0^T [s_i(\vec{x}_r, t, \vec{m}) - d_i(\vec{x}_r, t)] \int_0^t \int_V [\delta\rho(\vec{x}') G_{ij}(\vec{x}_r, \vec{x}'; t - t') \frac{\partial^2}{\partial t'^2} s_j(\vec{x}', t')$$

$$+\delta c_{jklm}(\vec{x}') \frac{\partial}{\partial x'_k} G_{ij}(\vec{x}_r, \vec{x}'; t - t') \frac{\partial}{\partial x'_l} s_m(\vec{x}', t')] d^3\vec{x}' dt'$$

- **After plenty of algebra one finds that this can be written as :**

$$\delta\chi = \int_V \left[K_\rho(\vec{x})\delta \ln(\rho(\vec{x})) + K_{c_{ijkl}}(\vec{x})\delta \ln(c_{ijkl}(\vec{x}))\right] d^3\vec{x}$$

Putting everything together II.

Following from above ...

- **Where the kernels $K_\rho(\vec{x})$ and $K_{c_{ijkl}}(\vec{x})$ are defined as follows:**

$$K_\rho(\vec{x}) = -\int_0^T \rho(\vec{x}) s_i^\dagger(\vec{x}, T-t) \frac{\partial^2}{\partial t^2} s_i(\vec{x}, t) dt$$

$$K_{c_{ijkl}}(\vec{x}) = -\int_0^T \epsilon_{jk}^\dagger(\vec{x}, T-t) c_{jklm}(\vec{x}) \epsilon_{lm}(\vec{x}, t) dt \qquad \textbf{[NO } \sum \textbf{]}$$

Putting everything together II.

Following from above ...

- **Where the kernels $K_\rho(\vec{x})$ and $K_{c_{ijkl}}(\vec{x})$ are defined as follows:**

$$K_\rho(\vec{x}) = -\int_0^T \rho(\vec{x}) {s_i}^\dagger(\vec{x}, T-t) \frac{\partial^2}{\partial t^2} s_i(\vec{x}, t) dt$$

$$K_{c_{ijkl}}(\vec{x}) = -\int_0^T \epsilon_{jk}^\dagger(\vec{x}, T-t) c_{jklm}(\vec{x}) \epsilon_{lm}(\vec{x}, t) dt \qquad \textbf{[NO } \sum \textbf{]}$$

- **The magic of full waveform inversion really lies in the adjoint wavefield ${s_i}^\dagger$.**

Putting everything together II.

Following from above ...

- **Where the kernels $K_\rho(\vec{x})$ and $K_{c_{ijkl}}(\vec{x})$ are defined as follows:**

$$K_\rho(\vec{x}) = -\int_0^T \rho(\vec{x}) {s_i}^\dagger(\vec{x}, T-t) \frac{\partial^2}{\partial t^2} s_i(\vec{x}, t) dt$$

$$K_{c_{ijkl}}(\vec{x}) = -\int_0^T \epsilon_{jk}^\dagger(\vec{x}, T-t) c_{jklm}(\vec{x}) \epsilon_{lm}(\vec{x}, t) dt \qquad \text{[NO } \textstyle\sum \text{]}$$

- **The magic of full waveform inversion really lies in the adjoint wavefield ${s_i}^\dagger$.**

- **${s_i}^\dagger$ arises naturally in carrying out the calculation. But what is it and how is it interpreted?**

Outline

1. Introducing the Idea of Full Waveform Inversion.
2. Using the costs.
   - Getting useful expressions out of the cost.
   - Understanding the Adjoint wavefield.
   - Back to using the cost.
3. Home cooked examples of the Algorithm.

The Adjoint Wavefield.

How does one understand the Adjoint Wavefield

- **The adjoint wavefield satisfies the following equation:**

$$\rho \frac{\partial^2}{\partial t^2} s_i^{\dagger}(\vec{x}, t) - \frac{\partial}{\partial x_j} c_{ijkl} \frac{\partial}{\partial x_k} s_l^{\dagger}(\vec{x}, t) =$$

$$= \sum_{r=1}^{N} \left[s_i(\vec{x}_r, T-t) - d_i(\vec{x}_r, T-t) \right] \delta(\vec{x} - \vec{x}_r) = f_i^{\dagger}(\vec{x}, t)$$

The Adjoint Wavefield.

How does one understand the Adjoint Wavefield

- **The adjoint wavefield satisfies the following equation:**

$$\rho \frac{\partial^2}{\partial t^2} s_i^{\dagger}(\vec{x}, t) - \frac{\partial}{\partial x_j} c_{ijkl} \frac{\partial}{\partial x_k} s_l^{\dagger}(\vec{x}, t) =$$

$$= \sum_{r=1}^{N} \left[s_i(\vec{x}_r, T - t) - d_i(\vec{x}_r, T - t) \right] \delta(\vec{x} - \vec{x}_r) = f_i^{\dagger}(\vec{x}, t)$$

- **But that is just the seismic wave equation with a source term :** $f_i^{\dagger}(\vec{x}, t)$

The Adjoint Wavefield.

How does one understand the Adjoint Wavefield

- **The adjoint wavefield satisfies the following equation:**

$$\rho \frac{\partial^2}{\partial t^2} s_i{}^{\dagger}(\vec{x}, t) - \frac{\partial}{\partial x_j} c_{ijkl} \frac{\partial}{\partial x_k} s_l{}^{\dagger}(\vec{x}, t) =$$

$$= \sum_{r=1}^{N} \left[s_i(\vec{x}_r, T-t) - d_i(\vec{x}_r, T-t) \right] \delta(\vec{x} - \vec{x}_r) = f_i{}^{\dagger}(\vec{x}, t)$$

- **But that is just the seismic wave equation with a source term :** $f_i{}^{\dagger}(\vec{x}, t)$

- **Where the waveform adjoint source** $f_i{}^{\dagger}(\vec{x}, t)$

$$f_i{}^{\dagger}(\vec{x}, t) = \sum_{r=1}^{N} \left[s_i(\vec{x}_r, T-t) - d_i(\vec{x}_r, T-t) \right] \delta(\vec{x} - \vec{x}_r)$$

Interpreting the Adjoint wavefield

How can we understand this ?

- **As the adjoint field satisfies the seismic wave equation, we can just think of it as seismic waves set up by the waveform adjoint source.**

Interpreting the Adjoint wavefield

How can we understand this ?

- **As the adjoint field satisfies the seismic wave equation, we can just think of it as seismic waves set up by the waveform adjoint source.**
- **But that is just the seismic wave equation with a source term :** $f_i^{\dagger}(\vec{x}, t)$

Interpreting the Adjoint wavefield

How can we understand this ?

- **As the adjoint field satisfies the seismic wave equation, we can just think of it as seismic waves set up by the waveform adjoint source.**
- **But that is just the seismic wave equation with a source term :** $f_i^{\dagger}(\vec{x}, t)$
- **Note that the waveform adjoint source** $f_i^{\dagger}(\vec{x}, t)$ **is just the time-reversed differences between the synthetics and the data, injected at the locations at the sensor locations.**

$$f_i^{\dagger}(\vec{x}, t) = \sum_{r=1}^{N} \left[s_i(\vec{x}_r, T - t) - d_i(\vec{x}_r, T - t) \right] \delta(\vec{x} - \vec{x}_r)$$

Interpreting the Adjoint wavefield II.

Following from above ...

- **Thus the waveform adjoint field can be seen as the seismic wave generated by injecting the mistakes (differences in data and synthetics) we've made back into the model from the sensor where those mistakes were made with reversed time.**

Interpreting the Adjoint wavefield II.

Following from above ...

- **Thus the waveform adjoint field can be seen as the seismic wave generated by injecting the mistakes (differences in data and synthetics) we've made back into the model from the sensor where those mistakes were made with reversed time.**
- **You may be asking yourself "so what ". This is of course a very good question. The answer to the question lies in something I made mention to earlier.**

Interpreting the Adjoint wavefield II.

Following from above ...

- **Thus the waveform adjoint field can be seen as the seismic wave generated by injecting the mistakes (differences in data and synthetics) we've made back into the model from the sensor where those mistakes were made with reversed time.**
- **You may be asking yourself "so what ". This is of course a very good question. The answer to the question lies in something I made mention to earlier.**
- **So let us look back at how we use the cost.**

Outline

How did the cost change with model perturbations again?

Recall the expressions :

- **Recall that given a model perturbation, the change in cost was :**

$$\delta\chi = \int_V \left[K_\rho(\vec{x})\delta \ln(\rho(\vec{x})) + K_{c_{ijkl}}(\vec{x})\delta \ln(c_{ijkl}(\vec{x}))\right] d^3\vec{x}$$

How did the cost change with model perturbations again?

Recall the expressions :

- **Recall that given a model perturbation, the change in cost was :**

$$\delta\chi = \int_V \left[K_\rho(\vec{x})\delta\ln(\rho(\vec{x})) + K_{c_{ijkl}}(\vec{x})\delta\ln(c_{ijkl}(\vec{x}))\right] d^3\vec{x}$$

- **Where the kernels $K_\rho(\vec{x})$ and $K_{c_{ijkl}}(\vec{x})$ were defined as follows:**

$$K_\rho(\vec{x}) = -\int_0^T \rho(\vec{x}) s_i^{\dagger}(\vec{x}, T-t)\frac{\partial^2}{\partial t^2} s_i(\vec{x}, t)dt$$

$$K_{c_{ijkl}}(\vec{x}) = -\int_0^T \epsilon_{jk}^{\dagger}(\vec{x}, T-t) c_{jklm}(\vec{x})\epsilon_{lm}(\vec{x}, t)dt \qquad \textbf{[NO } \textstyle\sum \textbf{]}$$

So what are these so called Kernels then?

What are the kernels?

- **The kernels as defined above are just the functional derivatives of the cost with respect to the various model parameters.**

So what are these so called Kernels then?

What are the kernels?

- **The kernels as defined above are just the functional derivatives of the cost with respect to the various model parameters.**
- **In other words, at the expense of one extra simulation (for the adjoint wave), the gradients of the model become available to use.**

So what are these so called Kernels then?

What are the kernels?

- **The kernels as defined above are just the functional derivatives of the cost with respect to the various model parameters.**
- **In other words, at the expense of one extra simulation (for the adjoint wave), the gradients of the model become available to use.**
- **So for example, If I wanted to know how the model changes with respect to changes in density it would be as simple (or difficult) as computing:**

$$\frac{\delta\chi}{\delta\rho(\vec{x})} = K_\rho(\vec{x}) = -\int_0^T \rho(\vec{x}) s_i{}^\dagger(\vec{x}, T-t)\frac{\partial^2}{\partial t^2} s_i(\vec{x}, t) dt$$

How the gradients are useful.

Why is it good to have gradients?

- **The gradients are very useful, because they allow the use of efficient techniques to minimise the cost of the model. Consider the definition for the cost once again.**

$$\chi[\vec{m}] = \frac{1}{2}\sum_{r=1}^{N}\int_0^T \|\vec{s}(\vec{x}_r, t, \vec{m}) - \vec{d}(\vec{x}_r, t)\|^2 dt$$

How the gradients are useful.

Why is it good to have gradients?

- **The gradients are very useful, because they allow the use of efficient techniques to minimise the cost of the model. Consider the definition for the cost once again.**

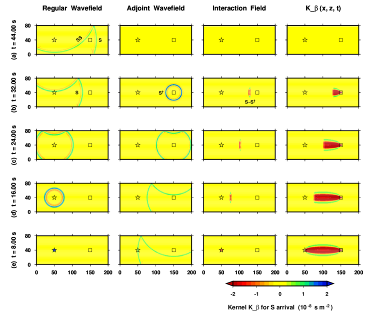
$$\chi[\vec{m}] = \frac{1}{2} \sum_{r=1}^{N} \int_0^T \|\vec{s}(\vec{x}_r, t, \vec{m}) - \vec{d}(\vec{x}_r, t)\|^2 dt$$

- **When the synthetics and the data agree, the cost is zero, a global minimum. If the model is close enough to reality, then the model parameters lie in a well surrounding this global minimum. Further we know that the derivatives should be zero at the minimum, small nearby, and larger further away. Thus the gradients indicate the direction in which the function is increasing.**

The Kernels paint a picture for us.

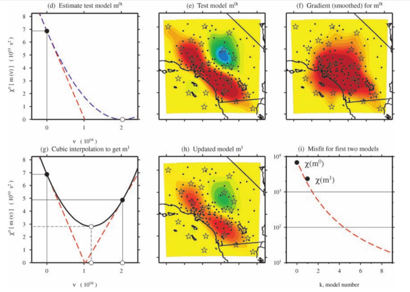
The Kernels are small where the model is accurate and large where the model is inaccurate.

- **Following the above reasoning, and noting that the kernels for the medium are functions of location, one can understand how the kernels can paint a picture of the accuracy of our model.**



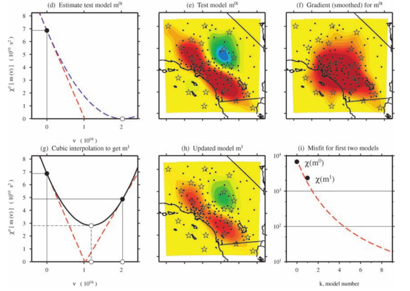
An example of what it means to be inside a potential well.

- **If we change the model parameters in the opposite manner the gradients change, then we'll be heading towards the minimum.**



An example of what it means to be inside a potential well.

- **If we change the model parameters in the opposite manner the gradients change, then we'll be heading towards the minimum.**



- **The minimisation procedure is formalised in the conjugate-gradient algorithm.**

The Conjugate-Gradient Algorithm.

Let us formalise the above

- **The above idea can be formalised in the conjugate gradient algorithm.**
- **The procedure using the Kernals to iteratively improve our model follows:**
  1. For $k = 0$, $\vec{p}^{(0)} = -\frac{\partial \chi}{\partial \vec{m}} = -g^{(0)}$. If at any stage $\|\vec{p}\| < \varepsilon$, then our model is accurate.
  2. Find a scalar $\lambda^{(k)}$ that minimises the function $\tilde{\chi}(\lambda) = \chi(m^{(0)} + \lambda \vec{p}^{(k)})$.
  3. Update the model as follows: $\vec{m}^{(k+1)} = \vec{m}^{(0)} + \lambda^{(k)} \vec{p}^{(k)}$, calculate $g^{(k+1)} = \frac{\partial}{\partial \vec{m}^{(k+1)}} \chi(\vec{m}^{(k+1)})$.
  4. Update $\vec{p}$ as follows: $\vec{p}^{(k+1)} = -g^{(k+1)} + \beta_{k+1} \vec{p}^{(k)}$. Where $\beta_{k+1} = \frac{\vec{g}^{(k+1)} \cdot \vec{g}^{(k+1)}}{\vec{g}^{(k)} \cdot \vec{g}^{(k)}}$.
  5. If $\|\vec{p}^{(k+1)}\| < \varepsilon$ then the model $\vec{m}^{(k+1)}$ is the one we want. Else we restart from 2.

How is the performance of this procedure.

Some examples...

- **Firstly consider a vector force acting at a certain location in the medium. Further imagine that we know the magnitude and location of the force, but not its orientation. In the same way that the above kernels were derived, one may derive the kernel associated with changes in the force which generates seismic waves.**

How is the performance of this procedure.

Some examples...

- **Firstly consider a vector force acting at a certain location in the medium. Further imagine that we know the magnitude and location of the force, but not its orientation. In the same way that the above kernels were derived, one may derive the kernel associated with changes in the force which generates seismic waves.**
- **The variation of cost with respect to the generating force is given by:**

$$\delta\chi = \int_0^T \int_V \delta f_i(\vec{x}, t) s_i^\dagger(\vec{x}, T-t)\, dVdt$$

Writing things in terms of angles.

- **Thus we can find the derivatives with respect to the polar and azimuthal angles:**

$$\frac{\delta\chi}{\delta\theta_j} = \int_0^T \int_V h(t) s_i^\dagger(\vec{x}, T-t) \frac{\partial x_i}{\partial\theta_j} \delta\theta_j dVdt$$

Writing things in terms of angles.

- **Thus we can find the derivatives with respect to the polar and azimuthal angles:**

$$\frac{\delta\chi}{\delta\theta_j} = \int_0^T \int_V h(t) s_i^\dagger(\vec{x}, T-t) \frac{\partial x_i}{\partial\theta_j} \delta\theta_j dV dt$$

- **The procedure was performed where the initial source orientation was at an angle of 178.2° to the true source orientation. The results follow:**

Performance for the inversion of an orientation.

Performance in terms of re-orienting the source

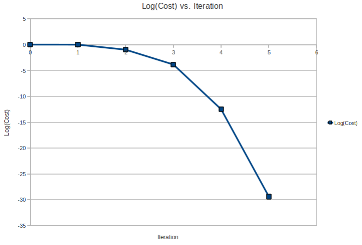


Figure: Cost: $1 \rightarrow 10^{-30}$ in 5 iterations

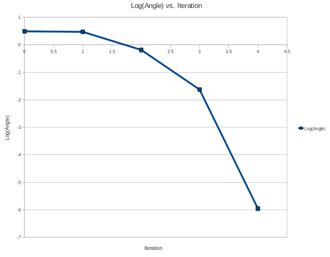


Figure: Angle $178.2^{\circ} \rightarrow 0^{\circ}$ in 5 iterations

What about having the source mis Located?

Performance in re-Locating a generating force.

- **Next consider a generating force, where we know both the orientation of the force, together with its time dependence. The only thing we do not know is the location of the source.**

What about having the source mis Located?

Performance in re-Locating a generating force.

- **Next consider a generating force, where we know both the orientation of the force, together with its time dependence. The only thing we do not know is the location of the source.**
- **In this case the variation of the Cost with respect to the model parameters are given by:**

$$\frac{\delta \chi}{\delta x_j} = \delta \vec{x} \cdot \vec{\nabla} \int_0^T h(t) s_j^{\dagger}(\vec{x}, T - t) dt$$

What about having the source mis Located?

Performance in re-Locating a generating force.

- **Next consider a generating force, where we know both the orientation of the force, together with its time dependence. The only thing we do not know is the location of the source.**

- **In this case the variation of the Cost with respect to the model parameters are given by:**

$$\frac{\delta\chi}{\delta x_j} = \delta\vec{x} \cdot \vec{\nabla} \int_0^T h(t) s_j^\dagger(\vec{x}, T-t)\,dt$$

- **The iterative procedure was performed where the true location was (0,0,0) and the initial synthetic location is given by (200,100,-50) . Thus the initial distance between the source and location is 229m.**

Performance for the inversion of an Location

Performance in terms of re-locating the source

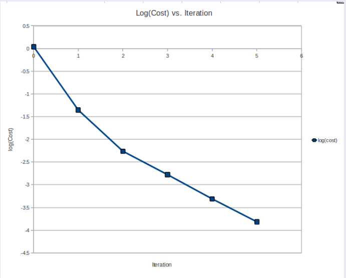


Figure: Cost: $1 \rightarrow 1.5 \times 10^{-4}$ in 5 iterations

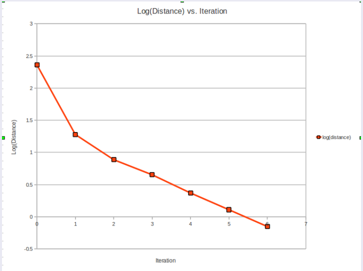


Figure: Mis-location: $229m \rightarrow 0.7m$ in 5 iterations

And sources which are both mis-located and mis-oriented?

- **As before, we assume that we know the details of the force.**

And sources which are both mis-located and mis-oriented?

- **As before, we assume that we know the details of the force.**
- **The variation of the cost with the model parameters are the same as before.**

And sources which are both mis-located and mis-oriented?

- **As before, we assume that we know the details of the force.**
- **The variation of the cost with the model parameters are the same as before.**
- **This time there is however an additional subtlety which makes the problem more complicated. The subtlety lies in the fact that the gradients of the different groups of model parameters do not necessarily reflect the contributions of the various model parameters to the total cost. In the following example the initial angle between the synthetics and data is 15°, and the initial distance between the location for the data and synthetics was 15m**

Exposition of what can go wrong.

Still looks fine here

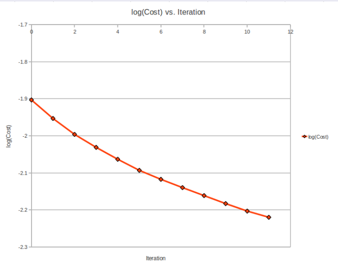


Figure: Cost : 0.01 → 0.005 in 15 iterations

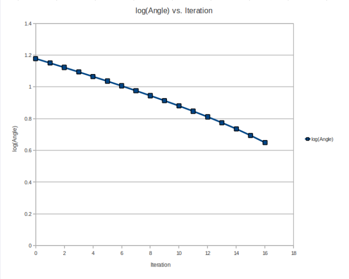


Figure: Alignement Angle: $15^{\circ} \rightarrow 4.5^{\circ}$ in 15 iterations

Exposition of what can go wrong II.

Look at the very slow convergence

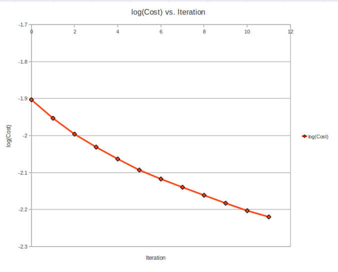


Figure: Cost : 0.01 → 0.005 in 15 iterations

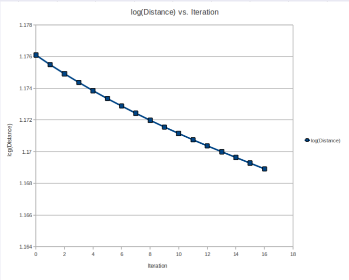


Figure: Mis-location: $15m \rightarrow 14.75m$ in 15 iterations

Another approach?

Another approach I tried:

- **Because of the problems of non-convergence another approach had to be tried. In this approach, in an attempt to escape the dominating effect of the orientation over the location, it was assumed that both groups of model parameters were solely responsible for generating the model cost.**

Another approach?

Another approach I tried:

- **Because of the problems of non-convergence another approach had to be tried. In this approach, in an attempt to escape the dominating effect of the orientation over the location, it was assumed that both groups of model parameters were solely responsible for generating the model cost.**
- **This approach does not do too badly, but has serious problems. In many examples, where the synthetics's initial location and orientation did not lie sufficiently close to the data's location and orientation the model parameters "ran away from " the data parameters.**

Still not quite right.

The other approach works much better:

- **Also, as we're always overcompensating for the errors in the model, there is a definite limit on how close the model can come to reality- at which point overcompensation makes the model less accurate rather than more accurate.**

Still not quite right.

The other approach works much better:

- **Also, as we're always overcompensating for the errors in the model, there is a definite limit on how close the model can come to reality- at which point overcompensation makes the model less accurate rather than more accurate.**
- **It is also obvious that the cost is not strictly decreasing, which is a problem for an algorithm which searches for minima. In the following best case scenario, used as an example, the initial distance between the source location and the synthetic location is 60m, and the initial alignment angle between the synthetics and the data is 18°.**

This approach works much better:

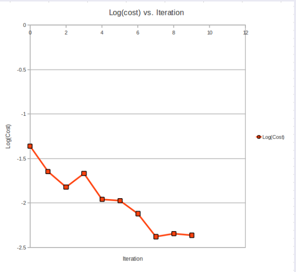


Figure: Cost : $4 \times 10^{-2} \rightarrow 4 \times 10^{-3}$ in 10 iterations

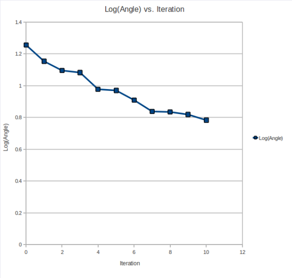


Figure: Alignement angle: $18^\circ \rightarrow 6^\circ$ in 10 iterations

Hints of problems...

This works better - but is still not quite right:

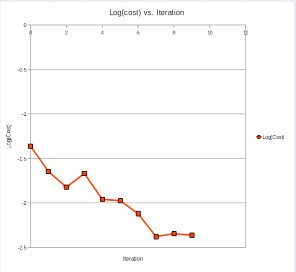


Figure: $4 \times 10^{-2} \rightarrow 4 \times 10^{-3}$ in 10 iterations

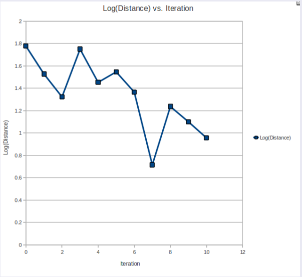


Figure: Mis-loction: $60m \rightarrow 9m$ in 10 iterations

Best results

The best solution I've come up with yet.

- **In an attempt to solve the problems present in the above approach, a modification was made, where only the changes that lead to a lower cost are made.**

Best results

The best solution I've come up with yet.

- **In an attempt to solve the problems present in the above approach, a modification was made, where only the changes that lead to a lower cost are made.**
- **This approach works well, but it does suffer from a drawback in that the computational time is significantly longer. In the following example the initial distance between the source and synthetic locations was 92m and the initial angle between the direction of the synthetics, and the direction of the data was 35°.**

Best results so far

An acceptable result:

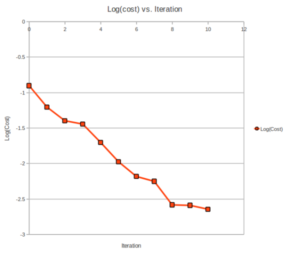


Figure: Cost: $0.12 \rightarrow 0.0022$ in 10 iterations

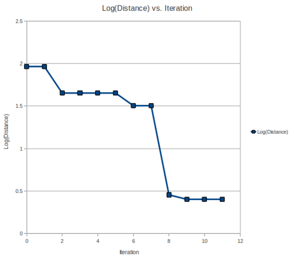


Figure: Mis-location: $92.5m \rightarrow 2.5m$ in 10 iterations

Best results so far

This works:

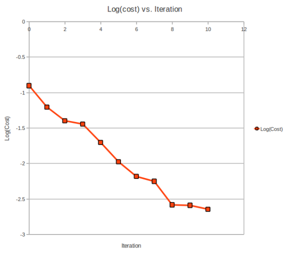


Figure: Cost: 0.12 $\rightarrow$ 0.0022 in 10 iterations

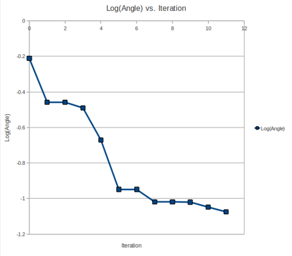


Figure: Alignment Angle: 35° $\rightarrow$ 5° in 10 iterations

What about Moment Sources.

What about extending this analysis to Moment sources

- **Next we consider a moment source, the location of which we know, but the components are unknown to us.**

What about Moment Sources.

What about extending this analysis to Moment sources

- **Next we consider a moment source, the location of which we know, but the components are unknown to us.**
- **The variation of the Cost with respect to the model parameters is given by:**

$$\delta\chi = \int_0^T \int_\Sigma \epsilon^\dagger_{ij}(\vec{x}, T - t)\delta m_{ij}(\vec{x}, t)dAdt$$

Where $\epsilon^\dagger_{ij} = \frac{1}{2}(\frac{\partial}{\partial x_i}s^\dagger_j + \frac{\partial}{\partial x_j}s^\dagger_i)$

And Σ **is the area of the fault plane.**

A Moment source example

An example with moment sources.

- **In the following example, I let the components of the moment tensor change, without letting the magnitude change. It was also assumed that the location and source time function were known. On the following slide, the average difference between the components of the moment tensor is plotted along with the cost function.**

A Moment source example

An example with moment sources.

- **In the following example, I let the components of the moment tensor change, without letting the magnitude change. It was also assumed that the location and source time function were known. On the following slide, the average difference between the components of the moment tensor is plotted along with the cost function.**
- **For the following example the moment tensors considered were:**

Data moment tensor

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Synthetic moment tensor

$$\begin{pmatrix} 5 & -6 & 10 \\ -6 & 3 & -7 \\ 10 & -7 & 6 \end{pmatrix}$$

Some results

Not too bad:

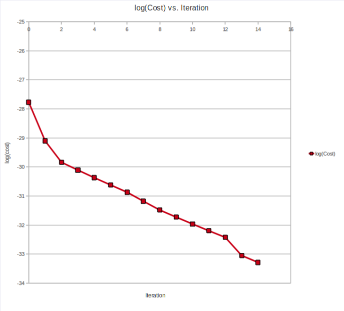


Figure: Cost: $1.7 \times 10^{-28} \rightarrow 5.21 \times 10^{-34}$ in 15 iterations

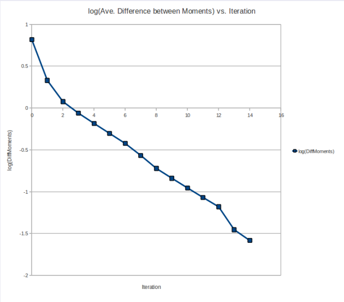


Figure: Average Difference in Moments :$6.56 \rightarrow 0.056$ in 15 iterations

More results

Not too bad:

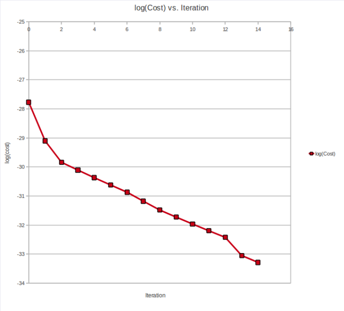


Figure: Cost: $1.7 \times 10^{-28} \rightarrow 5.21 \times 10^{-34}$ in 15 iterations

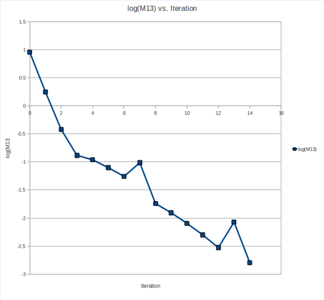


Figure: Differences in M_{13} :9 $\rightarrow$ 0.0016 in 15 iterations

Diagonal terms

Not too bad:

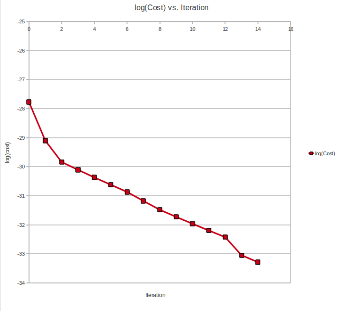


Figure: Cost: $1.7 \times 10^{-28} \rightarrow 5.21 \times 10^{-34}$ in 15 iterations

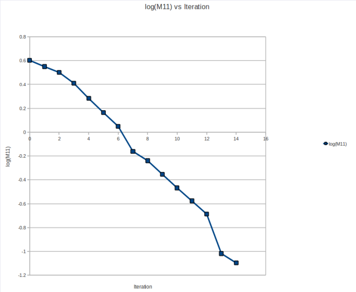


Figure: Differences in M_{11} :4 $\rightarrow$ 0.08 in 15 iterations

Some Notes.

Some final remarks :

- **The diagonal and off diagonal terms of the moment tensor appear largely independent. The behaviour of the diagonal and off-diagonal terms are somewhat reminiscent of the relationship between the orientation and location of the force vector for vector force inversion.**

Some Notes.

Some final remarks :

- **The diagonal and off diagonal terms of the moment tensor appear largely independent. The behaviour of the diagonal and off-diagonal terms are somewhat reminiscent of the relationship between the orientation and location of the force vector for vector force inversion.**
- **The iterative process is not computationally cheap as full waveform modeling in a non homogeneous medium is needed.**

Some Notes.

Some final remarks :

- The diagonal and off diagonal terms of the moment tensor appear largely independent. The behaviour of the diagonal and off-diagonal terms are somewhat reminiscent of the relationship between the orientation and location of the force vector for vector force inversion.
- The iterative process is not computationally cheap as full waveform modeling in a non homogeneous medium is needed.
- This method of iteratively improving the model is not limited. Nearly every model parameter one can think of can be inverted for using the above procedure. The more model parameters are inverted for, the more difficult and expensive the iterative procedure.

Where to from here

Some final remarks :

- **The next step to take are to include more model parameters. This will eventually lead to performing inversions for hetrogeneous media.**

Where to from here

Some final remarks :

- **The next step to take are to include more model parameters. This will eventually lead to performing inversions for hetrogeneous media.**

- **At some stage it is likely that a norm other than the L_2 norm will be needed. Other, more robust norms, which are of greater practical utility have recently become available - however they are less easy to use.**

Where to from here

Some final remarks :

- **The next step to take are to include more model parameters. This will eventually lead to performing inversions for hetrogeneous media.**
- **At some stage it is likely that a norm other than the L_2 norm will be needed. Other, more robust norms, which are of greater practical utility have recently become available - however they are less easy to use.**
- **Thank you for your time and goodbye.**

References.

- J. Tromp , C. Tape and Q. Liu,
  Seismic tomography, adjoint methods, time reversal and banana-doughnut kernals
  Geophysical Journal International **v160**(2005)
- J. Tromp , C. Tape and Q. Liu,
  Finite-frequency tomography using adjoint methods-Methodology and examples using membrane surface waves.
  Geophysical Journal International **v168**(2007)
- J. Tromp , C. Tape,
  Finite-Frequency Kernels Based on Adjoint Methods
  Bulletin of the Seismological Society of America **v96.6**(December 2007)