Probabilistic inverse Theory. Assignment 1 – Rayleigh Criterion

J. Alejandro Duran, ETH 14/11/2017

1 Objective

Rayleigh criterion can be defined as the minimum distance that two adjacent signals can be such that an observer is able to distinguish (resolve) between them. In seismic reflection, this criteria is commonly used. Then, when the Earth is approximated as an horizontal layered model, the Rayleigh criterion tells the minimum layer thickness than one can resolve for a given dominant wavelength. This limit will be revised in this assignment, from the point of view of probabilistic inverse theory.

2 Prior probability density

The prior $\rho(\mathbf{m})$ contains information about the model parameters $\mathbf{m}$=$m_i, ..., m_M$ that we set in order to constrains all possible results. In our case, the model parameters are the reflection coefficients $m_i \equiv r_i$ and there are some strong constrains that I describe in the following:

Constrains

1. All reflection coefficients are zero, except two of them, which are in the boundaries. In other words $r_k = 0 \; \forall \; k \neq i, j$. Where i, j are the unique points where the reflection coefficients are non zero.
2. All locations i, j are equally likely, but $j \geq i$. This means that even thought we do not know the positions i, j of the non zero reflection coefficients, we know that the location of the boundary of r_j is deeper than the one of r_i.
3. Both reflectors are assumed to be equal in magnitude, but with opposite sign. This is $r_i = -r_j = r$
4. The magnitude of the reflection coefficients follow a Gaussian distribution, with mean value zero, and standard deviation s. We can think carefully about it. If the reflection coefficient is non-zero only at two points (i, j), then the most likely value is zero. Therefore, a Gaussian distribution centered at $r = 0$ makes totally sense. This is clearer when looking at figure 1.

$$\rho(r) = \frac{1}{s\sqrt{2\pi}} \exp(\frac{-r^2}{2s^2}). \tag{1}$$

With these constrains, we will not work with the whole set of model parameters $\mathbf{m}$=$m_i, ..., m_M$, but with the smaller set $\mathbf{m}$=(r, i, j) which contains information of the constrains. It means that we built our prior as a joint prior distribution by using the constrains' information.

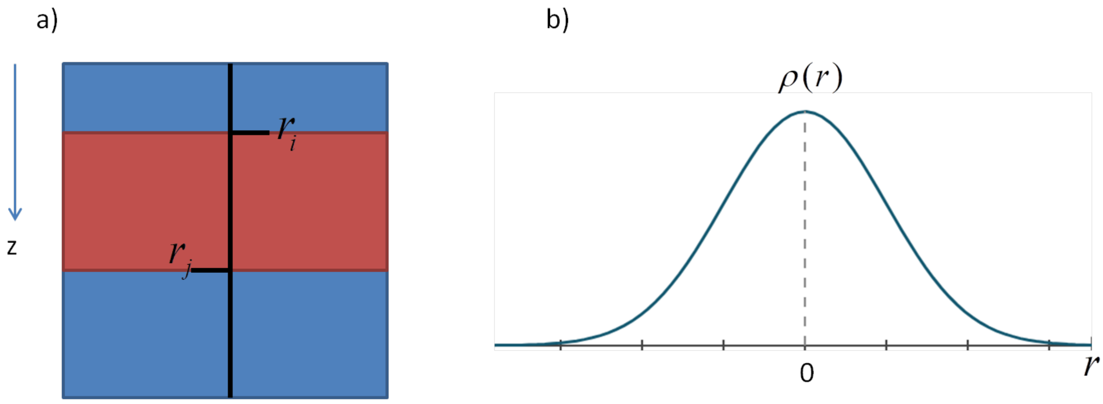

Figure 1: (a)Reflectiviy model. (b)Reflection coefficients represented by a Gaussian distribution.

The total prior is (from Equation 2.30 of the course notes):

$$\rho(r, i, j) = \rho(r|i, j)\rho(i, j). \tag{2}$$

The term $\rho(r|i, j)$ is a conditional probability density. This expression can be read as: *"Given the reflector positions i, j , what is the probability density of r (r is the magnitude of the non-zero reflection coefficients)?"* Nevertheless, we observe that the reflector positions are independent of the magnitude of r. It is useful to observe again the constrains and realize that they are not related, therefore $\rho(r|i, j) = \rho(r)$.

The second term in equation 2 can be obtained from constraint number 2. It says that all locations i, j are equally likely but $j \geq i$. Then, if there is a maximum of M positions, we should consider all possible combinations of i, j which satisfy these constraints. This is equivalent to consider two balls (i, j) which are allowed to drop inside M buckets. If we take the order into account i, j is the same as j, i and we don't count it twice; in this way we include the condition $j \geq i$. The number of possible locations is given by the combinations expression:

$$\binom{M}{2} = \frac{M!}{(M-2)!2!} = \frac{M(M-1)}{2}. \tag{3}$$

Thus, $\rho(i, j)$ is equal to one event (the one in which I am situated over the non-zero reflectors) over all possible events.

$$\rho(i, j) = \frac{1}{\binom{M}{2}} = \frac{2}{M(M-1)}. \tag{4}$$

From equation 2 we can now write explicitly the total prior:

$$\rho(r, i, j) = \rho(r)\rho(i, j) = \frac{2}{M(M-1)} \frac{1}{s\sqrt{2\pi}} \exp(\frac{-r^2}{2s^2}). \tag{5}$$

3 Posterior probability density

Using Tarantola-Valette expression, we define the posterior probability density as:

$$\sigma(\mathbf{m}) = c \cdot \rho(\mathbf{m})L(\mathbf{m}) \tag{6}$$

The prior was defined in equation 5 using the constraints. Now our task is to obtain the Likelihood. The likelihood is an indication of how well the data fits the model.

Then, by considering the following assumptions:

i)the inverse problem is linear

ii)the noise probability density is Gaussian with zero mean,

we can write the likelihood as:

$$L(\mathbf{m}) = c \cdot \exp(-\frac{1}{2}(\mathbf{d} - \mathbf{Gm})^T \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{Gm})) \tag{7}$$

Where $\mathbf{d}$ is the observed (measured) data, $\mathbf{d}_{pred} = \mathbf{Gm}$ is the predicted data (based on the Physics of the problem) and $\mathbf{C}_d$ is the covariance matrix of the noise. Noise is considered as $\mathbf{n} = \mathbf{d} - \mathbf{Gm}$. The constant c is an unknown normalization factor.

By using the constraints, we said that our model parameters were $\mathbf{m} = (r, i, j)$. If the predicted data is the convolution between the source wavelet and the reflection coeficients, we can parametrize the predicted data as: $\mathbf{d}_{pred} = r(\mathbf{w}_i - \mathbf{w}_j)$. Here, $\mathbf{w}_i$ and $\mathbf{w}_j$ are time delayed copies of the source wavelet.

We are able now to write the posterior as the multiplication of the prior (Eq. 5) and the likelihood (Eq. 7). We will skip the normalization factor because is unknown.

$$\sigma(r, i, j) = \frac{2}{M(M-1)} \frac{1}{s\sqrt{2\pi}} \exp(\frac{-r^2}{2s^2}) \exp(-\frac{1}{2}(\mathbf{d} - r(\mathbf{w}_i - \mathbf{w}_j))^T \mathbf{C}_d^{-1}(\mathbf{d} - r(\mathbf{w}_i - \mathbf{w}_j))) \tag{8}$$

Equation 8 is a probability density function. It is made by multiplying density functions, the likelihood and the prior, each of them having a particular maximum. Additionally the posterior can be evaluated for each observed data d_k, given a different result at each data point, in the general case. Because of these reasons, the posterior in 8 is a function which may have more than one maximum.

4 Maximum Posterior

Let's consider the linear inverse problem: $\mathbf{d} = \mathbf{Gm}$. In our case, for the fixed i, j non-zero reflection positions, $\mathbf{G} = \mathbf{w}_i - \mathbf{w}_j$ is a N x 1 matrix. On the other hand, $\mathbf{m}$ is the magnitude of the reflection coefficient r (a positive scalar value). The covariance of our prior will be simply $\mathbf{C}_m = s^2$.

In equation 2.36 of the notes (Hansen and Mosegaard), there is a description of a posterior built from a Gaussian likelihood and prior. This is also the case in our current exercise. Then, a multiplication of two Gaussian densities will be also Gaussian, and the equation 2.36 of the notes shows that the mean of such Gaussian posterior is:

$$\mathbf{m}_{post} = \mathbf{m}_0 + \mathbf{C}_{post}\mathbf{G}^T \mathbf{C}_n^{-1}(\mathbf{d} - \mathbf{Gm}_0) \tag{9}$$

With the posterior covariance $\mathbf{C}_{post} = (\mathbf{G}^T \mathbf{C}_n^{-1}\mathbf{G} + \mathbf{C}_m^{-1})^{-1}$

We can use these expression for our current case. Here, our posterior Gaussian density is centered at zero, i.e. it has zero mean $\mathbf{m}_0 = 0$. As we mentioned before $\mathbf{G} = \mathbf{w}_i - \mathbf{w}_j$ and $\mathbf{C}_m = s^2$. There is also a slight change in notation, for our current exercise $\mathbf{C}_n^{-1} \equiv \mathbf{C}_d^{-1}$. With this in mind we can replace these values in equation 9 and obtain the maximum of the conditional posterior reflectivity for the given (i, j) positions.

$$r_{MAP}(i,j) = \hat{\sigma}(i,j)(\mathbf{w}_i - \mathbf{w}_j)^T \mathbf{C}_d^{-1}(\mathbf{w}_i - \mathbf{w}_j) \tag{10}$$

Where $\sigma(i, j)$ is the the conditional a posteriori covariance given by:

$$\hat{\sigma}(i,j) = ((\mathbf{w}_i - \mathbf{w}_j)^T \mathbf{C}_d^{-1}(\mathbf{w}_i - \mathbf{w}_j) + \frac{1}{s^2})^{-1} \tag{11}$$

4.1 Rayleigh criterion under absence of noise

Let's assume the noise is zero. This means $\mathbf{n} = \mathbf{d} - \mathbf{d}_{pred} = 0$, therefore the predicted and observed data are both the same. If I want to obtain the covariance between these two types of data (observed and predicted) I will get:

$$COV(\mathbf{d}, \mathbf{d}_{pred}) = \sigma^2(\mathbf{d}) = \sigma^2(\mathbf{d}_{pred}) \tag{12}$$

This is the covariance of the noise $\mathbf{C}_d$. The covariance becomes a variance (σ^2) because both data are the same in the absence of noise. Then, I can use the physical relation of my predicted data $\mathbf{d}_{pred} = \mathbf{Gm}$ for the current exercise to obtain:

$$\sigma^2(\mathbf{Gm}) = s^2(\mathbf{w}_i - \mathbf{w}_j)(\mathbf{w}_i - \mathbf{w}_j)^T \tag{13}$$

By introducing the last expression in equation 11 we get:

$$\hat{\sigma}(i,j) = ((\mathbf{w}_i - \mathbf{w}_j)^T[-s^2(\mathbf{w}_i - \mathbf{w}_j)(\mathbf{w}_i - \mathbf{w}_j)^T]^{-1}(\mathbf{w}_i - \mathbf{w}_j) + \frac{1}{s^2})^{-1} \tag{14}$$

Using the properties when operating with matrices we end up to

$$\hat{\sigma}(i,j) = (\frac{-1}{s^2} + \frac{1}{s^2})^{-1} \to \infty \tag{15}$$

Then, if we put last result in equation10 we observe that the maximum of the posterior goes to infinity (I remark here that we did not normalize the posterior). This means that no matter the source wavelet we can compute the model with infinite precision. In fact, the posterior (which is the probability density of my model parameters for the given data) tends to infinity in the absence of noise.