

# Introduction to Inverse Problems

*Albert Tarantola*

Institut de Physique du Globe de Paris

# Probability

Consider a manifold  $\mathfrak{M}$ , with a notion of *volume*. For any  $\mathcal{A} \subset \mathfrak{M}$ ,

$$V(\mathcal{A}) = \int_{\mathcal{A}} dv \quad .$$

A **volumetric probability** is a function  $f$  that to any  $\mathcal{A} \subset \mathfrak{M}$  associates its probability

$$P(\mathcal{A}) = \int_{\mathcal{A}} dv f \quad .$$

Example: If  $\mathfrak{M}$  is a metric manifold endowed with some coordinates  $\{x^1, \dots, x^n\}$ , then  $dv = \sqrt{\det \mathbf{g}} dx^1 \wedge \dots \wedge dx^n$ , and

$$P(\mathcal{A}) = \int_{\mathcal{A}} dx^1 \wedge \dots \wedge dx^n \underbrace{\sqrt{\det \mathbf{g}} f}_{\bar{f}} = \int_{\mathcal{A}} dx^1 \wedge \dots \wedge dx^n \bar{f} \quad .$$

Warning, the volumetric probability  $f$  is an invariant, the probability density  $\bar{f}$  is not (it is a density).

A basic operation with volumetric probabilities is their product,

$$(f \cdot g)(\mathcal{P}) = \frac{1}{\nu} f(\mathcal{P}) g(\mathcal{P}) \quad ,$$

where  $\nu = \int_{\mathfrak{M}} dv f(\mathcal{P}) g(\mathcal{P})$ .

Example: Two planes make two estimations of the geographical coordinates of a shipwrecked man, represented by the two volumetric probabilities  $f(\varphi, \lambda)$  and  $g(\varphi, \lambda)$ . The volumetric probability that combines these two pieces of information is

$$(f \cdot g)(\varphi, \lambda) = \frac{f(\varphi, \lambda) g(\varphi, \lambda)}{\int dS(\varphi, \lambda) f(\varphi, \lambda) g(\varphi, \lambda)} \quad .$$

This operation of product of volumetric probabilities extends to the following case:

- there is a volumetric probability  $f(\mathcal{P})$  defined on a first manifold  $\mathfrak{M}$ ,
- there is another volumetric probability  $\varphi(\mathcal{Q})$  defined on a second manifold  $\mathfrak{N}$ ,
- there is an application  $\mathcal{P} \mapsto \mathcal{Q} = \mathcal{Q}(\mathcal{P})$  from  $\mathfrak{M}$  into  $\mathfrak{N}$ .

Then, the basic operation is

$$g(\mathcal{P}) = \frac{1}{\nu} f(\mathcal{P}) \varphi(\mathcal{Q}(\mathcal{P})) .$$

where  $\nu = \int_{\mathfrak{M}} dv(\mathcal{P}) f(\mathcal{P}) \varphi(\mathcal{Q}(\mathcal{P}))$ .

# Inverse Problems

In a typical inverse problem, there is:

- a set of **model parameters**  $\{m^1, m^2, \dots, m^n\}$ ,
- a set of **observable parameters**  $\{o^1, o^2, \dots, o^n\}$ ,
- a relation  $o^i = o^i(m^1, m^2, \dots, m^n)$  predicting the outcome of the possible observations.

The model parameters are *coordinates* on the *model parameter manifold*  $\mathfrak{M}$ , while the observable parameters are *coordinates* over the *observable parameter manifold*  $\mathfrak{D}$ . Points on  $\mathfrak{M}$  are denoted  $\mathcal{M}, \mathcal{M}', \dots$  while points on  $\mathfrak{D}$  are denoted  $\mathcal{O}, \mathcal{O}', \dots$

Then the relation above is written  $\mathcal{M} \mapsto \mathcal{O} = \mathcal{O}(\mathcal{M})$

The three basic elements of a typical inverse problem are:

- some **a priori information** on the model parameters, represented by a volumetric probability  $\rho_{\text{prior}}(\mathcal{M})$  defined over  $\mathfrak{M}$ ,
- some **experimental information** obtained on the observable parameters, represented by a volumetric probability  $\sigma_{\text{obs}}(\mathcal{O})$  defined over  $\mathfrak{D}$ ,
- the ‘**forward modeling**’ relation  $\mathcal{M} \mapsto \mathcal{O} = \mathcal{O}(\mathcal{M})$  that we have just seen.

This leads to

$$\rho_{\text{post}}(\mathcal{M}) = \frac{1}{\nu} \rho_{\text{prior}}(\mathcal{M}) \sigma_{\text{obs}}(\mathcal{O}(\mathcal{M})) \quad ,$$

where  $\nu$  is a normalization constant.

# Example I: Sampling

- Sample the a priori volumetric probability  $\rho_{\text{prior}}(\mathcal{M})$ , to obtain (many) random models  $\mathcal{M}_1, \mathcal{M}_2, \dots$
- For each model  $\mathcal{M}_i$ , solve the forward modeling problem,  $\mathcal{O}_i = \mathcal{O}_i(\mathcal{M}_i)$ .
- Give to each model  $\mathcal{M}_i$  a probability of ‘survival’ proportional to  $\sigma_{\text{obs}}(\mathcal{O}_i(\mathcal{M}_i))$ .
- The surviving models  $\mathcal{M}_{1'}, \mathcal{M}_{2'}, \dots$  are samples of the a posteriori volumetric probability

$$\rho_{\text{post}}(\mathcal{M}) = \frac{1}{\mathcal{V}} \rho_{\text{prior}}(\mathcal{M}) \sigma_{\text{obs}}(\mathcal{O}(\mathcal{M})) \quad .$$

## Example II: Least-squares

- The model parameter manifold may be a linear space, with vectors denoted  $\mathbf{m}, \mathbf{m}', \dots$ , and the a priori information may have the Gaussian form

$$\rho_{\text{prior}}(\mathbf{m}) = k \exp\left(-\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{prior}})^t \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{prior}})\right) .$$

- The observable parameter manifold may be a linear space, with vectors denoted  $\mathbf{o}, \mathbf{o}', \dots$ , and the information brought by measurements may have the Gaussian form

$$\sigma_{\text{obs}}(\mathbf{o}) = k \exp\left(-\frac{1}{2} (\mathbf{o} - \mathbf{o}_{\text{obs}})^t \mathbf{C}_o^{-1} (\mathbf{o} - \mathbf{o}_{\text{obs}})\right) .$$

- The forward modeling relation becomes, with these notations,

$$\mathbf{o} = \mathbf{o}(\mathbf{m}) .$$



Then, the posterior volumetric probability for the model parameters is

$$\rho_{\text{post}}(\mathbf{m}) = k \exp(-S(\mathbf{m})) \quad ,$$

where the **misfit function**  $S(\mathbf{m})$  is the sum of squares

$$\begin{aligned} 2 S(\mathbf{m}) = & (\mathbf{m} - \mathbf{m}_{\text{prior}})^t \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{prior}}) \\ & + (\mathbf{o}(\mathbf{m}) - \mathbf{o}_{\text{obs}})^t \mathbf{C}_o^{-1} (\mathbf{o}(\mathbf{m}) - \mathbf{o}_{\text{obs}}) \quad . \end{aligned}$$

The **maximum likelihood** model is the model  $\mathbf{m}$  maximizing  $\rho_{\text{post}}(\mathbf{m})$ . It is also the model minimizing  $S(\mathbf{m})$ . It can be obtained using a quasi-Newton algorithm,

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \mathbf{H}_n^{-1} \boldsymbol{\gamma}_n \quad ,$$

where the *Hessian* of  $S$  is

$$\mathbf{H}_n = \mathbf{I} + \mathbf{C}_m \mathbf{O}_n^t \mathbf{C}_o^{-1} \mathbf{O}_n \quad ,$$

and the *gradient* of  $S$  is

$$\boldsymbol{\gamma}_n = \mathbf{C}_m \mathbf{O}_n^t \mathbf{C}_o^{-1} (\mathbf{o}(\mathbf{m}_n) - \mathbf{o}_{\text{obs}}) + (\mathbf{m}_n - \mathbf{m}_{\text{prior}}) \quad .$$

Here, the tangent linear operator  $\mathbf{O}_n$  is defined via

$$\mathbf{o}(\mathbf{m}_n + \delta\mathbf{m}) = \mathbf{o}(\mathbf{m}_n) + \mathbf{O}_n \delta\mathbf{m} + \dots$$

As we have seen, the model  $\mathbf{m}_\infty$  at which the algorithm converges maximizes the posterior volumetric probability  $\rho_{\text{post}}(\mathbf{m})$ .

To estimate the posterior uncertainties: the covariance operator of the Gaussian volumetric probability that is tangent to  $\rho_{\text{post}}(\mathbf{m})$  at  $\mathbf{m}_\infty$  is  $\mathbf{H}_\infty^{-1} \mathbf{C}_m$ .

## References

Tarantola, A., 2004, Inverse Problem Theory and Model Parameter Estimation, SIAM.

Mosegaard, K., and Tarantola, A., 2002, Probabilistic Approach to Inverse Problems, International Handbook of Earthquake & Engineering Seismology, Part A., pp. 237–265, Academic Press.

Material on Inverse Problems is available at the web site <http://www.ccr.jussieu.fr/tarantola>