ITINERIS

The meaning of inversion in applied geophysics

- Eusebio Maria Stucchi, UniPI

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell’Università e della Ricerca
Italiadomani

A gentle and brief Introduction to inverse problems

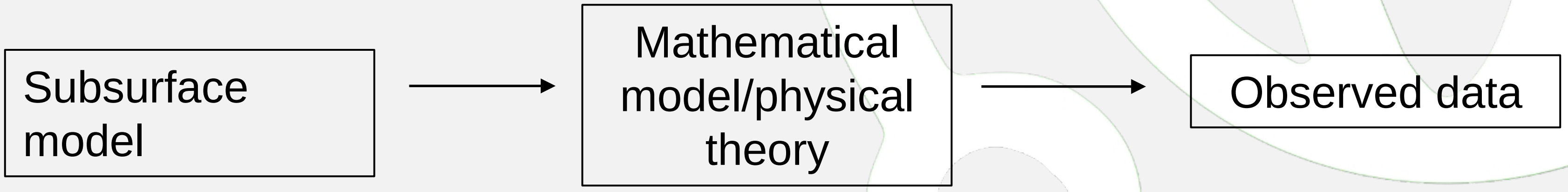
Aleardi, Stucchi

Slides from Prof. M. Aleardi

University of Pisa

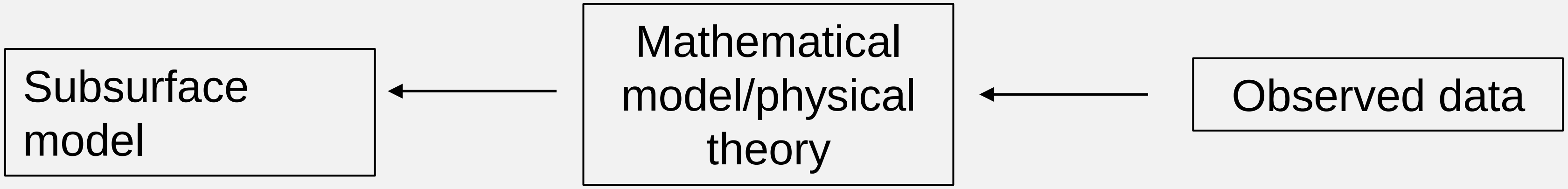
Forward and Inverse problems

Forward problem



Subsurface model → Mathematical model/physical theory → Observed data

Inverse problem



Subsurface model ← Mathematical model/physical theory ← Observed data

Problem formulation

m -> model parameters, subsurface properties (seismic velocities, resistivity, density…)

d -> observed data (seismograms, gravity anomalies, apparent resistivities ...)

G -> some mathematical operator that relates the model to the data.

Discrete (Matrix) form: data and model are expressed by vectors

$$d = G(m) \longrightarrow \begin{pmatrix} d_1 \\ \vdots \\ d_M \end{pmatrix} = G \begin{pmatrix} m_1 \\ \vdots \\ m_N \end{pmatrix}$$

Problem formulation

Three different cases

Forward problem → $d = G(m)$

- Given the model compute the data.
- Analytical or numerical solution.
- For example given the subsurface Vp, Vs and density compute the seismic data -> Numerical solution of the elastic wave equation.
- Unique solution. For a given model it exists only one data.

System identification problem

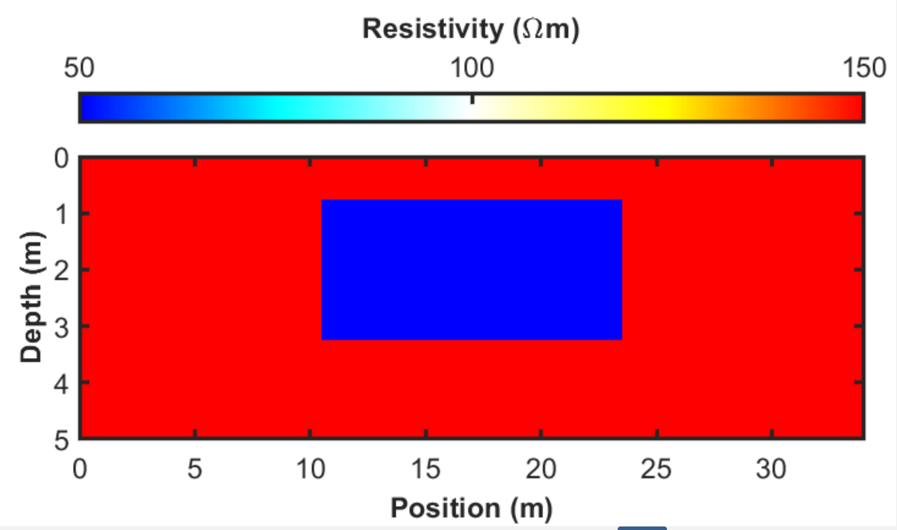
- Given **d** and **m** find **G:**

Inverse problem → $m = G^{-1}(d)$

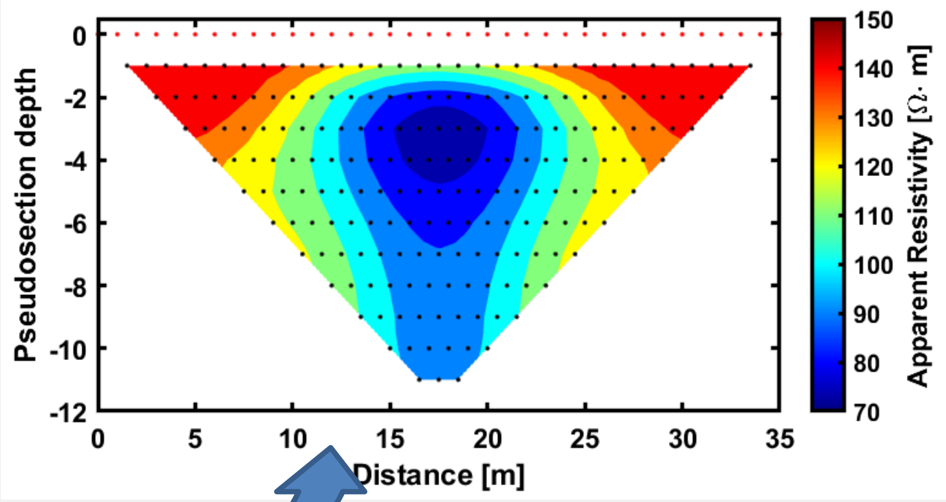
- Find **m** from **d** and **G**

Example: Forward problem

Actual resistivity model

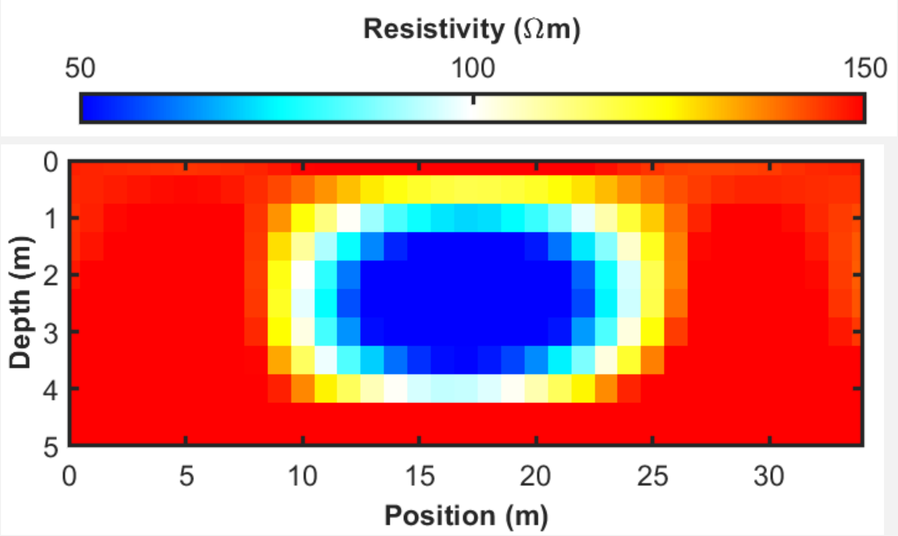


Apparent resistivity section

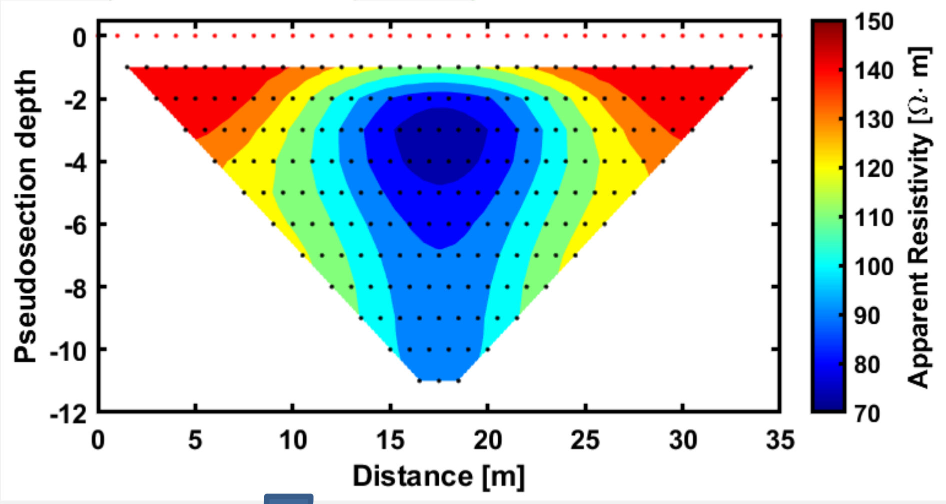


Example: Inverse problem

Predicted resistivity model



Apparent resistivity section



Types of inverse problems

Linear -> The inverse problem is linear if and only if:

$$G(m_1 + m_2) = G(m_1) + G(m_2) \longrightarrow \text{Superposition law}$$

$$G(am) = aG(m) \longrightarrow \text{Scaling law}$$

For a discrete inverse problem **m** e **d** are vectors and **G** is a matrix. They define a system of linear equations:

$$\begin{pmatrix} d_1 \\ \vdots \\ d_M \end{pmatrix} = \begin{pmatrix} G_{11} & \dots & G_{1N} \\ \vdots & \ddots & \vdots \\ G_{M1} & \dots & G_{MN} \end{pmatrix} \begin{pmatrix} m_1 \\ \vdots \\ m_N \end{pmatrix}$$

Non linear -> d and m are related by a non linear equation. The problem can not be written in matrix form

Types of inverse problems: Example

$T = t + x/V$ Analytical equation (e.g., refracted events)

$m_a = [t_a; V_a]$ $m_b = [t_b; V_b]$ Two different model vectors

$G(m_a) = t_a + x/V_a$ $G(m_b) = t_b + x/V_b$ Forward equations

Let’s verify the superposition law

$G(m_a + m_b) = (t_a + t_b) + x/(V_a + V_b)$ is different from …

$$G(m_a) + G(m_b) = (t_a + t_b) + x\left(\frac{1}{V_a} + \frac{1}{V_b}\right)$$

NON LINEAR PROBLEM

… from velocities to slownesses

$T = t\ + x/V \longrightarrow T = t\ + Sx$ Analytical equation (e.g., refracted events)

$m_a = [t_a; S_a] \quad m_b = [t_b; S_b]$ Two different model vectors

$$\Downarrow$$

$G(m_a) = t_a + xS_a \quad G(m_b) = t_b + xS_b$ Forward equations

Let’s verify the superposition law

$G(m_a + m_b) = (t_a{+}t_b) + x(S_a + S_b)$ is equal to …

$$G(m_a) + G(m_b) = (t_a{+}t_b) + x(S_a + S_b)$$

… continue

… from velocities to slownesses

$T = t + x/V \longrightarrow T = t + Sx$ Analytical equation (e.g., refracted events)

$m_a = [t_a; S_a]$ $m_b = [t_b; S_b]$ Two different model vectors

$\Downarrow$

$G(m_a) = t_a + xS_a$ $G(m_b) = t_b + xS_b$ Forward equations

Let's verify the scaling law

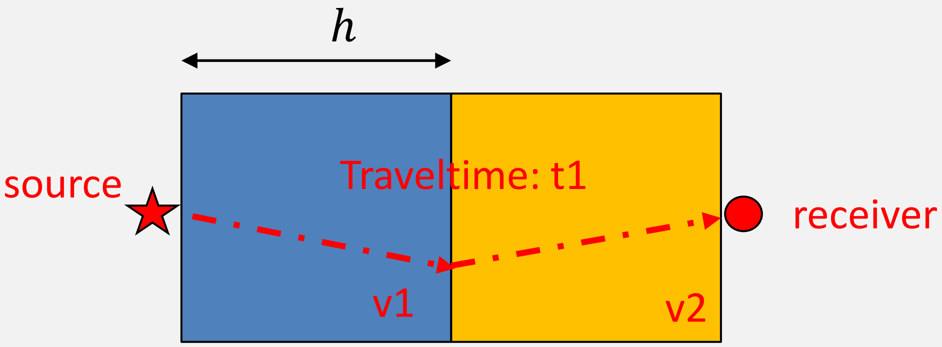
$G(\alpha m_a) = \alpha t_a + x\alpha S_a$ is equal to …

$$\alpha G(m_a) = \alpha(t_a + xS_a)$$

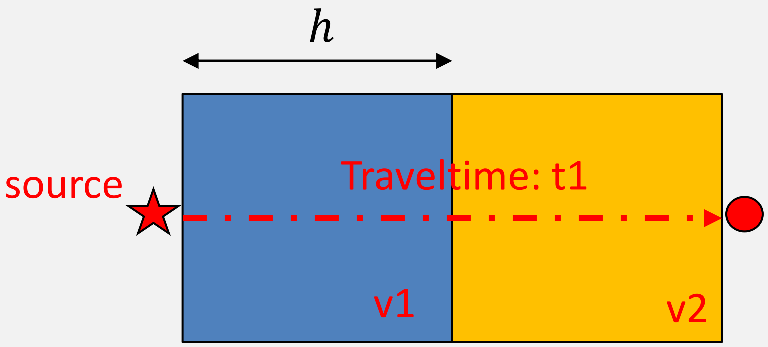
LINEAR PROBLEM!!

An example of a linearized problem

Taking into accont the Snell’ law: Non linear problem. The ray trajectory as expressed by the forward modelling depends on the velocity



Neglecting the Snell’ law: Linear problem. The ray trajectory is independed from the velocity (straight trajectory)



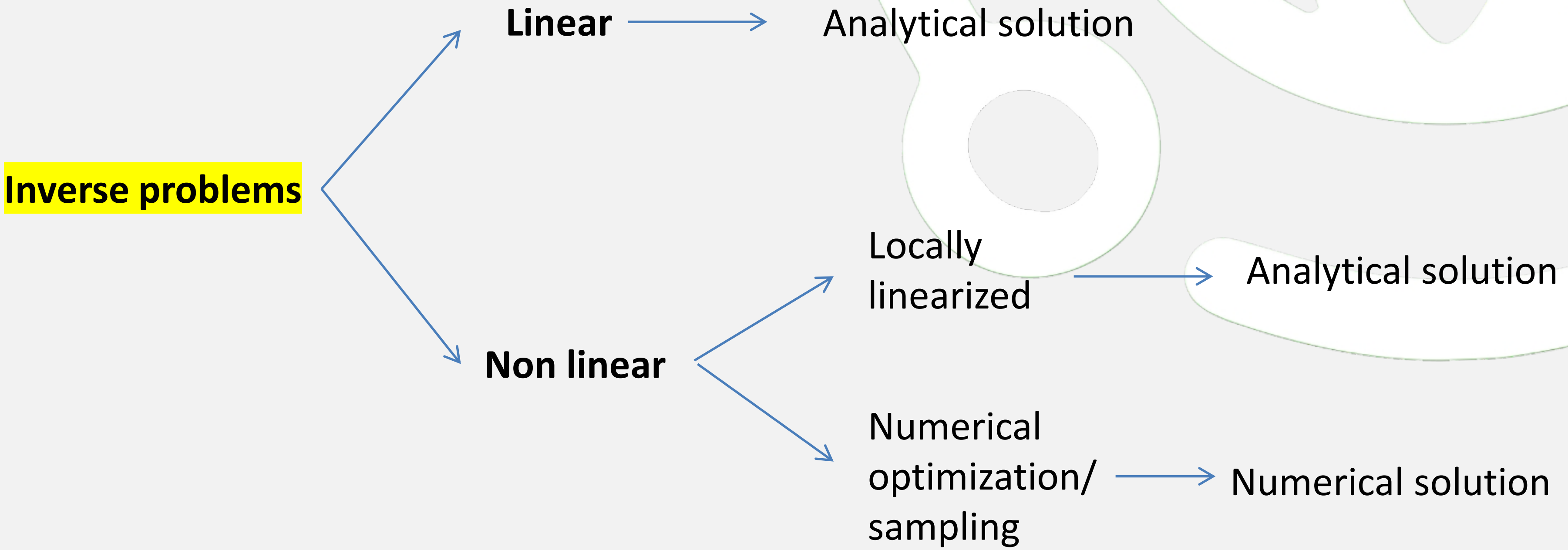
$$t_1 = \frac{h}{v_1} + \frac{h}{v_2} = hs_1 + hs_2$$

Analytical form

$$[t_1] = [h\ h]\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \longrightarrow \mathbf{d} = \mathbf{Gm}$$

Matrix form

Types of inverse problems



Inverse problems

Linear → Analytical solution

Non linear

Locally linearized → Analytical solution

Numerical optimization/ sampling → Numerical solution

Why inverse problems are "hard"

Existence of the solution-> There is no model that exactly fits the data. This happens because the forward model G is usually approximated and/or the data are contaminated by noise.

The solution is not unique-> Often there are infinite models that equally fit the observed data. This happens because some model parameters do not influence the data. Or in other terms, these parameters are not illuminated by the data.

Instability -> Small variations in the data produces significant variations in the recovered model. For discrete problems, this issue is often call ill-conditioning. We need some methods to stabilize the solution.

Data are usually noise contaminated -> Ambiental noise, instrumental noise affect the data. The noise statistic is often difficult to define. The noise contamination translates into multiple solutions that equally fit the data or it can introduce bias/artifacts in the solution

A very brief review of linear algebra

ITINERIS

Matrices and vectors

Row vector: $v = [1\ 2\ 3]$

Column vector: $v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

Matrix: $M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

Transpose of a matrix: $M^T = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$

Symmetric matrix: $M = M^T$

Identity matrix: $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Matrix summation: $M = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$ $N = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}$ $M + N = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 3 & 0 \\ 3 & 0 & 4 \end{bmatrix}$

Matrix multiplication:

$M = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 2 & 2 & 3 & 3 \end{bmatrix}$ $N = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}$ $Q = MN \rightarrow Q_{ij} = \sum_{k=1}^{K} M_{ik} N_{ki}$ $Q = \begin{bmatrix} 9 & 9 \\ 15 & 15 \end{bmatrix}$

$\downarrow$ 2x4

$\downarrow$ 4x2

$\downarrow$ 2x2

Matrices and vectors

Associative property: $(AB)C = A(BC)$

Commutative property: $AB \neq BA$

Transposition of a product: $(AB)^T = B^T A^T$

Inverse: $A^{-1}A = I$

Derivative of a matrix operator: $v = My \longrightarrow \frac{dv}{dy} = M$

Some other useful properties:

$$(A + B)^T = A^T + B^T$$

$$\frac{\partial(v^T B v)}{\partial v} = (B^T + B)\, v$$

$$A^T B = (B^T A)^T$$

Linear systems

$$Gm = d \longrightarrow \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \longrightarrow \begin{cases} G_{11}m_1 + G_{12}m_2 + G_{13}m_3 = d_1 \\ G_{21}m_1 + G_{22}m_2 + G_{23}m_3 = d_2 \\ G_{31}m_1 + G_{32}m_2 + G_{33}m_3 = d_3 \end{cases}$$

Inconsistent systems: there are no solutions

$$\begin{cases} x + y = 1 \\ \\ x + y = 0 \end{cases}$$

Consistent systems

- Unique solution $\begin{cases} x + y = 1 \\ x + 2y = 2 \end{cases}$
- Infinite solutions $\begin{cases} x - y = 0 \\ 2x - 2y = 0 \end{cases}$

Inverse of a matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$AB = I$ → B is the matrix inverse of A

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$AB = I \longrightarrow \begin{cases} a_{11}b_{11} + a_{12}b_{21} = 1 \\ a_{11}b_{12} + a_{12}b_{22} = 0 \\ a_{21}b_{11} + a_{22}b_{21} = 0 \\ a_{21}b_{12} + a_{22}b_{22} = 1 \end{cases}$$

To find the matrix inverse a system of equations must be solved.

To be invertible a matrix A must be square and with a not-null determinant

A simple way to compute the inverse of a 2x2 matrix

$$B = A^{-1} = \begin{bmatrix} a_{22}/Det(A) & -a_{12}/Det(A) \\ -a_{21}/Det(A) & a_{11}/Det(A) \end{bmatrix}$$

Linear dependent vectors and rank of a matrix

$\{x_1, x_2, \dots, x_m\} \longrightarrow$ *m* vectors *x* (e.g. m column vectors forming the **G** matrix)

If it exists at least one non-zero α_i so that:

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m = 0$$

The *x* vectors are **linearly dependent**

Differently, if $\sum_{j=1}^{m} \alpha_j x_j = 0$ only if all the $\alpha_j=0$, then the *x* vectors are **linearly independent**

The number of linearly independent columns of a matrix is called the **rank** (**r**)

r(A)<=min(N,M) A is a NxM matrix

if N=M and r(A)=N=M A is full column rank -> then, the inverse of A exists

Vector norm

The n-*th* norm of vector a: $\|a\|_n = \left(\sum_{j=1}^{m} |a_j|^n \right)^{1/n}$

Some properties:

- The norm is always positive and the minimum norm is zero
- The norm of a is zero if all the elements of vector a are equal to zero
- $\|\alpha a\| = |\alpha| \|a\|$
- $\|a + b\| \leq \|a\| \ \|b\|$

Linear inverse problems starting from linear regressions

Linear regression

Goal -> find the parameters that define a best-fitting curve with respect to the observed experimental data. If the regression model is linear we can use the matrix formalism to express the relation between model parameters and observed data.

$$y = xp + q + e \longrightarrow \textit{noise}$$

Dependent variable (observed data)

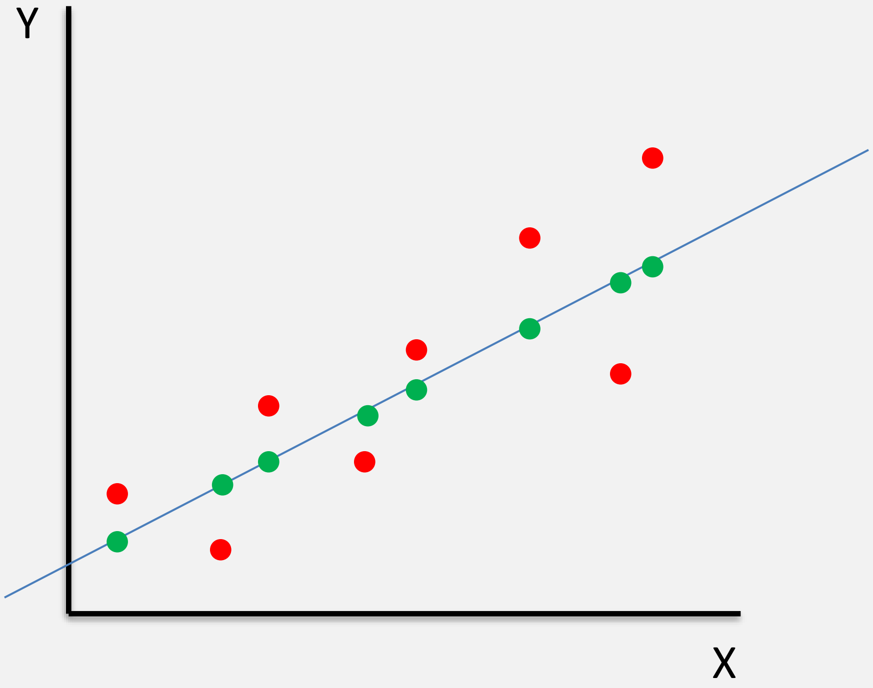
Independent variable

Regression coefficients (model parameters)

$$\begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_M \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix}$$

$$d = Gm + e$$

Linear regression as an inverse problem



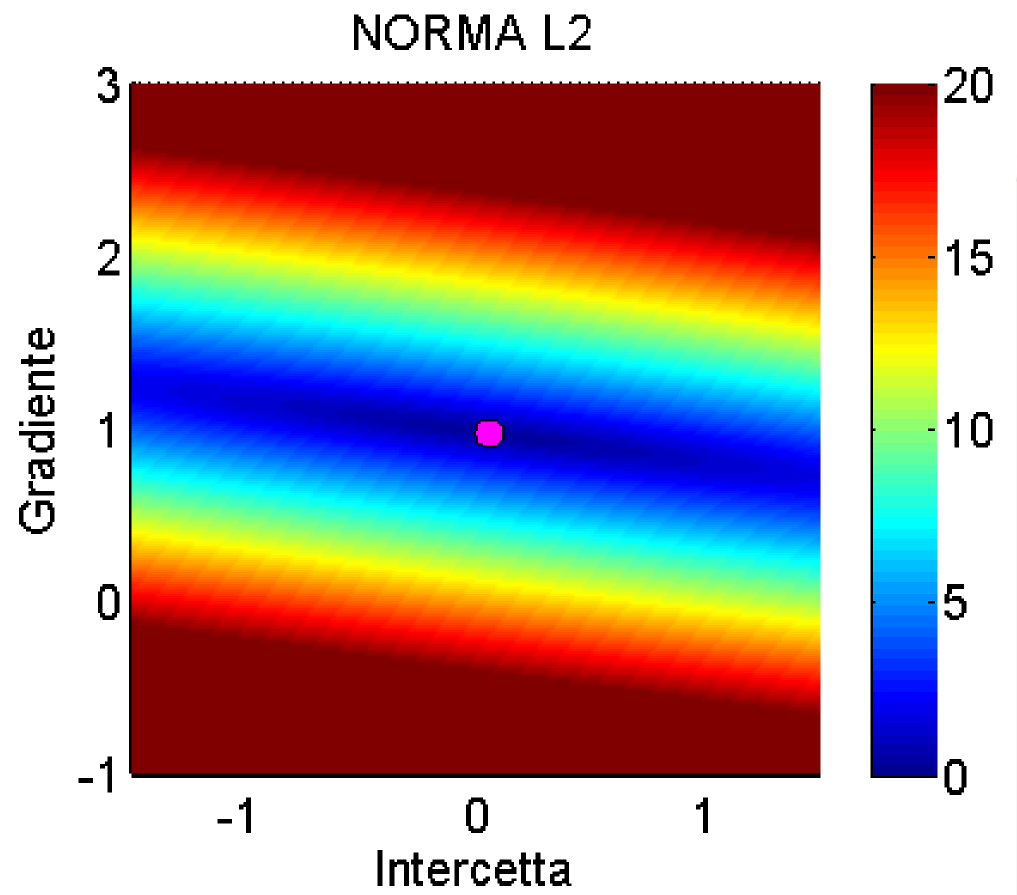
Finding the best-fitting line (i.e. combination of slope and intercept that minimizes the difference between **observed** and **predicted** data).

What is the meaning of best fitting??
For example finding the combination of slope and intercept that minimizes the L2 norm difference between predictions and observations (the Euclidean length of the residual data vector)

This translates into the minimization of an error function defined as follows (for N observations):

$$E(m) = \sum_{i=1}^{N} \left(d^{i}_{pred} - d^{i}_{obs}\right)^{2} = ||d - Gm||_{2}^{2}$$

Linear regression as an inverse problem



Example of an objective function associated with a linear regression problem. Our aim is to locate the global minimum of the Euclidean difference between predictions and observations (the positions of magenta dot!).

It can be demonstrated (see the following for details) that the model (or in other words the combination of slope and intercept) that minimizes the L2 norm error function for a linear problem can be found by solving this linear system of equation

Gauss normal equation → $m = (G^T G)^{-1} G^T d$

Gauss normal equations (demonstration)

Let us find the L2 norm solution:

$$\mathrm{E}(m) = (d - Gm)^T(d - Gm) = (d^T - (Gm)^T)(d - Gm) =$$

$$= (d^T d - (Gm)^T d - d^T(Gm) + (Gm)^T Gm) = (d^T d - 2d^T(Gm) + m^T G^T Gm)$$

To find the minimum of *r(m)* we must find the point in the model space where the derivative of r with respect to m is zero. $\nabla_m E = 0$

The gradient of r with respect to m is given by:

$$\nabla_m(d^T d) = 0; \nabla_m\left(-2d^T(Gm)\right) = -2d^T G;$$

$$\nabla_m(m^T G^T Gm) = m^T\left((G^T G)^T + G^T G\right) = 2m^T G^T G$$

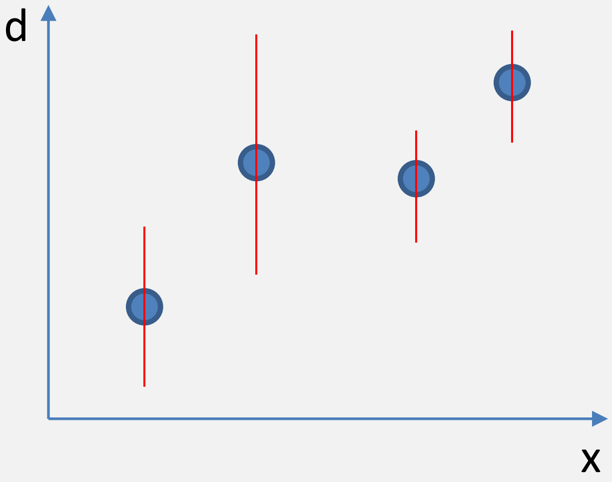
Then:

$$\nabla_m r(m) = 2m^T G^T G - 2d^T G = 0 \rightarrow m^T G^T G = d^T G \rightarrow G^T Gm = G^T d$$

$$\boxed{m = (G^T G)^{-1} G^T d} \longrightarrow$$

This solution exists only if $G^T G$ has full column rank. The rank of $G^T G$ is equal to the rank of G. Therefore, all the columns of G must be linearly independent vectors!!

Weighted L2 norm solution



Data affected by different uncertainties. We wish the linear regression be mainly driven by the data affected by low uncertainty. To this end we can weight each data residual differently:

$$\mathrm{E}(m) = ||W_d^{\frac{1}{2}}(d - Gm)||_2^2 = (d - Gm)^T W_d (d - Gm)$$

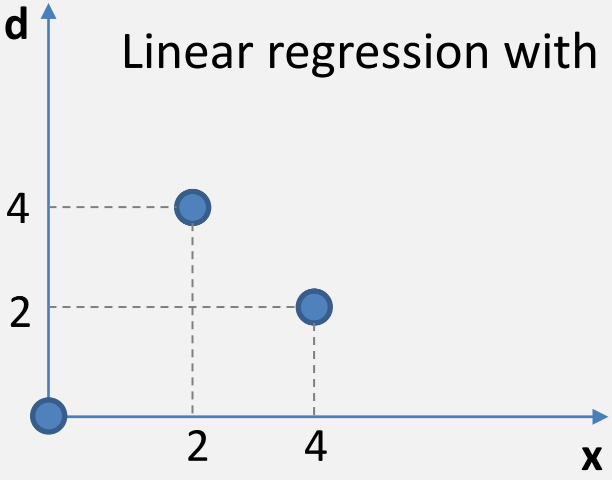
The most popular weighting matrix is the inverse of the so-called **data covariance matrix** (Cd):

$$W_d^{\ -1} = Cd \ = \begin{bmatrix} \sigma_1\sigma_1 & \sigma_1\sigma_2 & \dots & \sigma_1\sigma_M \\ \vdots & & \vdots & \vdots \\ \vdots & & & \vdots \\ \sigma_M\sigma_1 & \sigma_M\sigma_2 & \dots & \sigma_M\sigma_M \end{bmatrix}$$

The data covariance matrix quantifies the uncertainties affecting each data point. If the data are independent this matrix is diagonal, with diagonal entries expressing the noise variance of each data point. The weighted L2 norm solution becomes:

$$\boxed{m = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} d}$$

A Simple Example



Linear regression with three pionts

$d = q + px$ Analytical equation

Problem formulation:

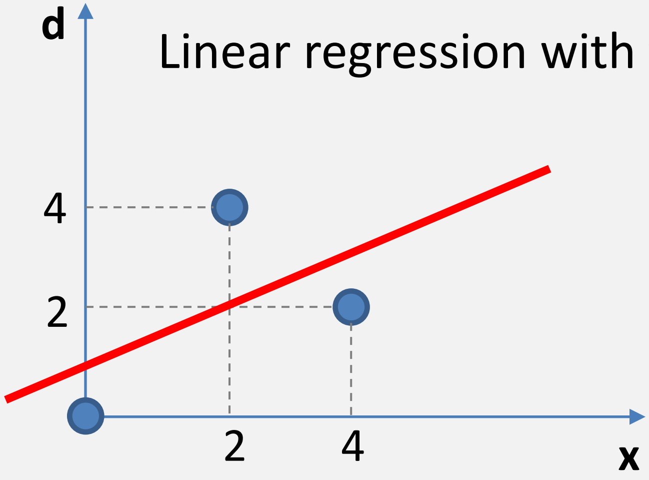
$$d = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} \qquad m = \begin{bmatrix} q \\ p \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}$$

Standard L2 norm solution: $m = (G^T G)^{-1} G^T d$

$$\Downarrow$$

$$m = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

A Simple Example



Linear regression with three pionts

$d = q + px$ Analytical equation

Problem formulation:

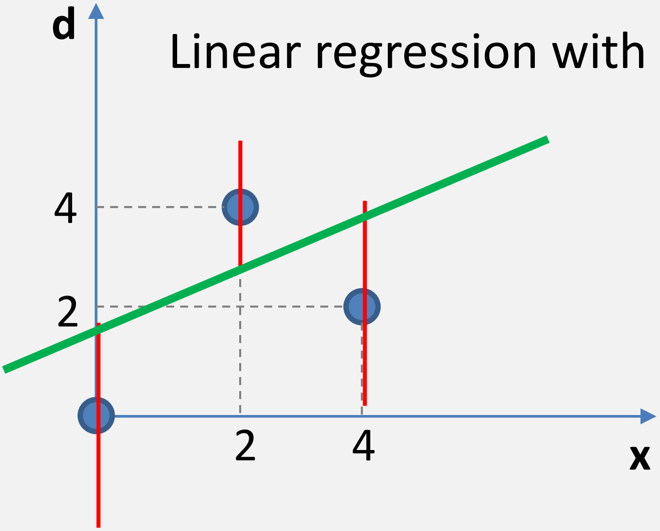
$$d = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} \quad m = \begin{bmatrix} q \\ p \end{bmatrix} \quad G = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}$$

Standard L2 norm solution: $m = (G^T G)^{-1} G^T d$

$$\Downarrow$$

$$m = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

A Simple Example



Linear regression with three pionts

$$d = q + px$$ Analytical equation

Problem formulation with data uncertainty:

$$d = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} \quad m = \begin{bmatrix} q \\ p \end{bmatrix} \quad G = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}$$

$$C_d = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

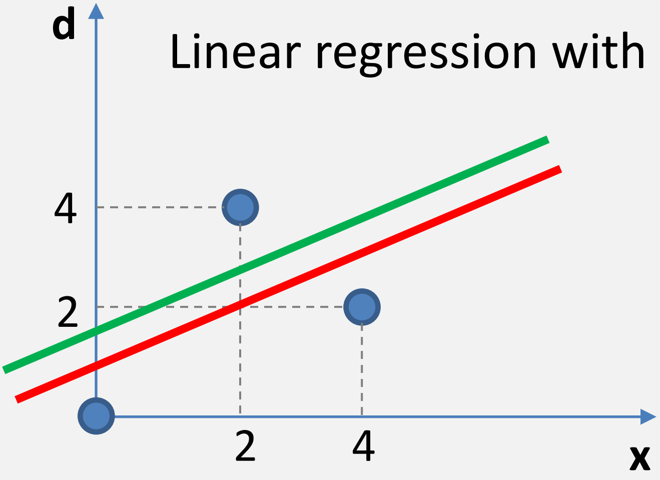
Weighted L2 norm solution:

$$m = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} d$$

$$\Downarrow$$

$$m = \begin{bmatrix} 3/2 \\ 0.5 \end{bmatrix}$$

A Simple Example



Linear regression with three pionts

$$d = q + px$$ Analytical equation

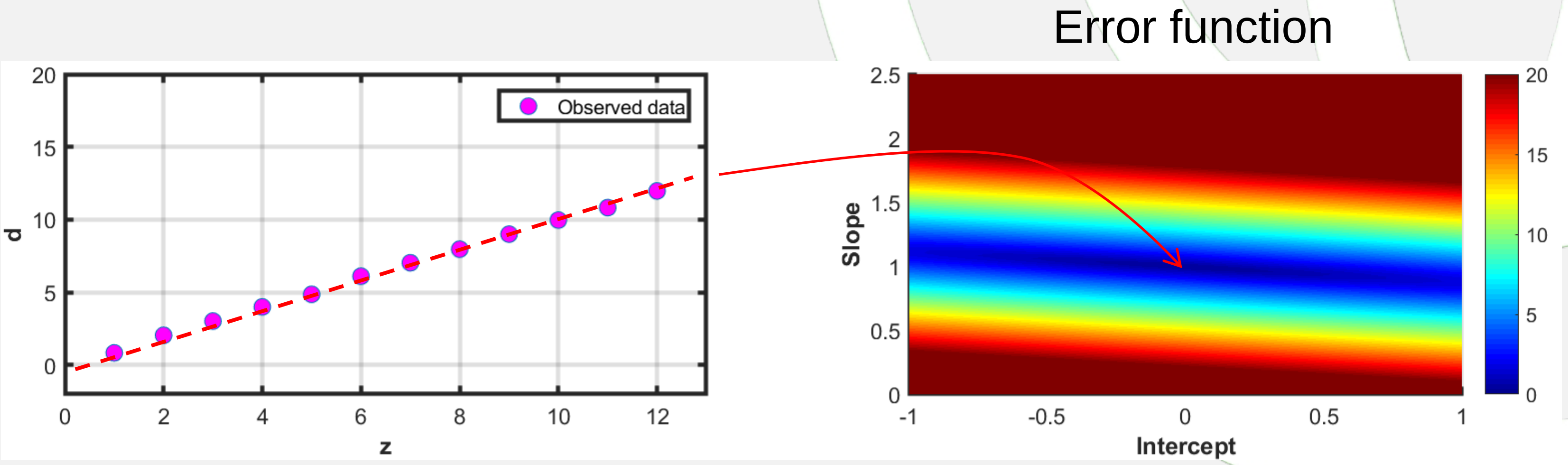
$$d = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} \quad m = \begin{bmatrix} q \\ p \end{bmatrix} \quad G = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}$$

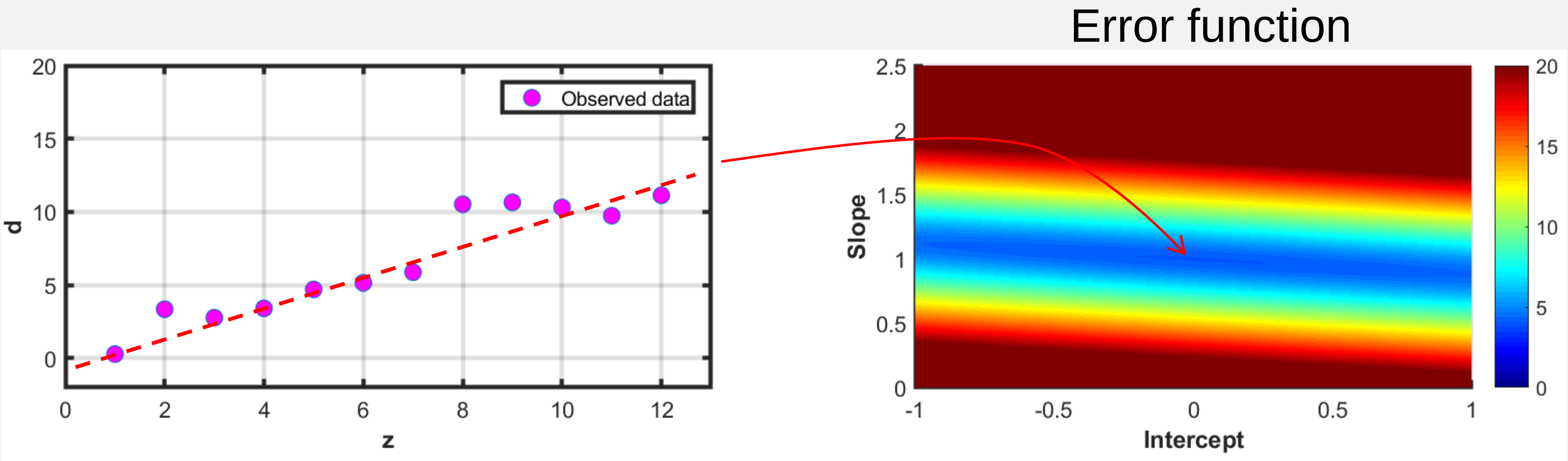
$$C_d = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$m = (G^T G)^{-1} G^T d \quad \Downarrow \quad m = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

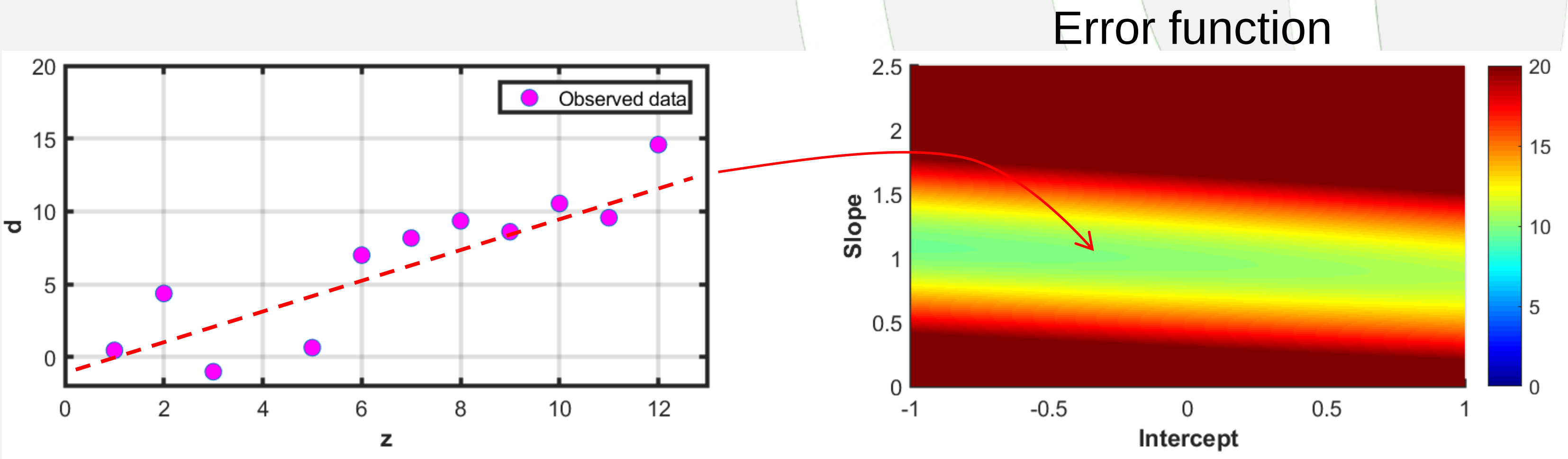
$$m = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} d \quad \Downarrow \quad m = \begin{bmatrix} 3/2 \\ 0.5 \end{bmatrix}$$

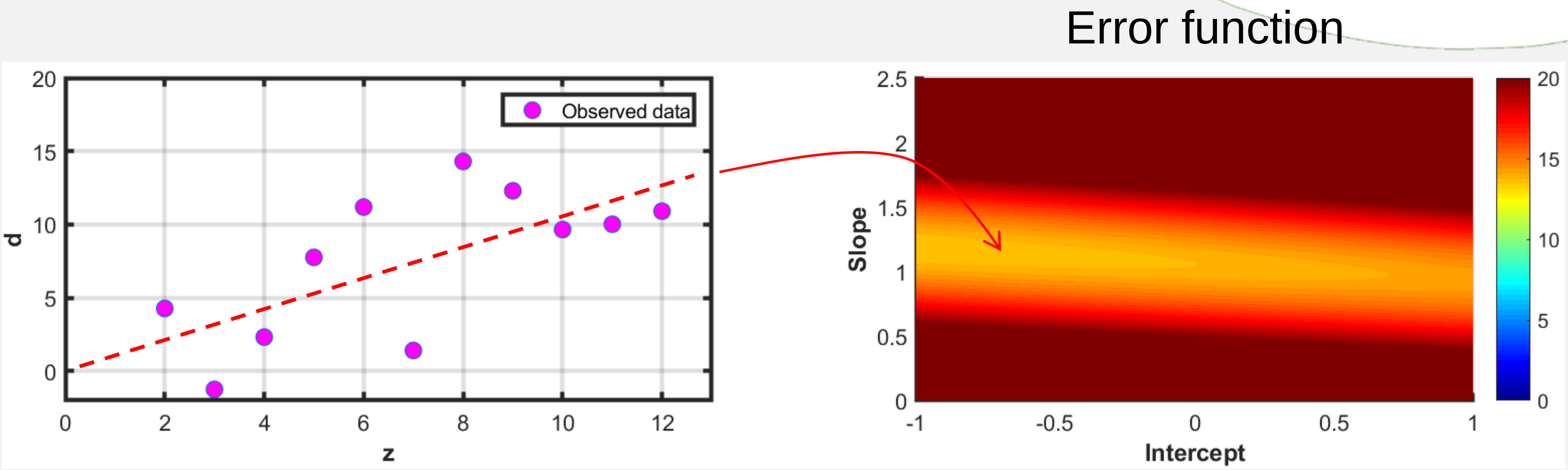
Error functions for different noise levels





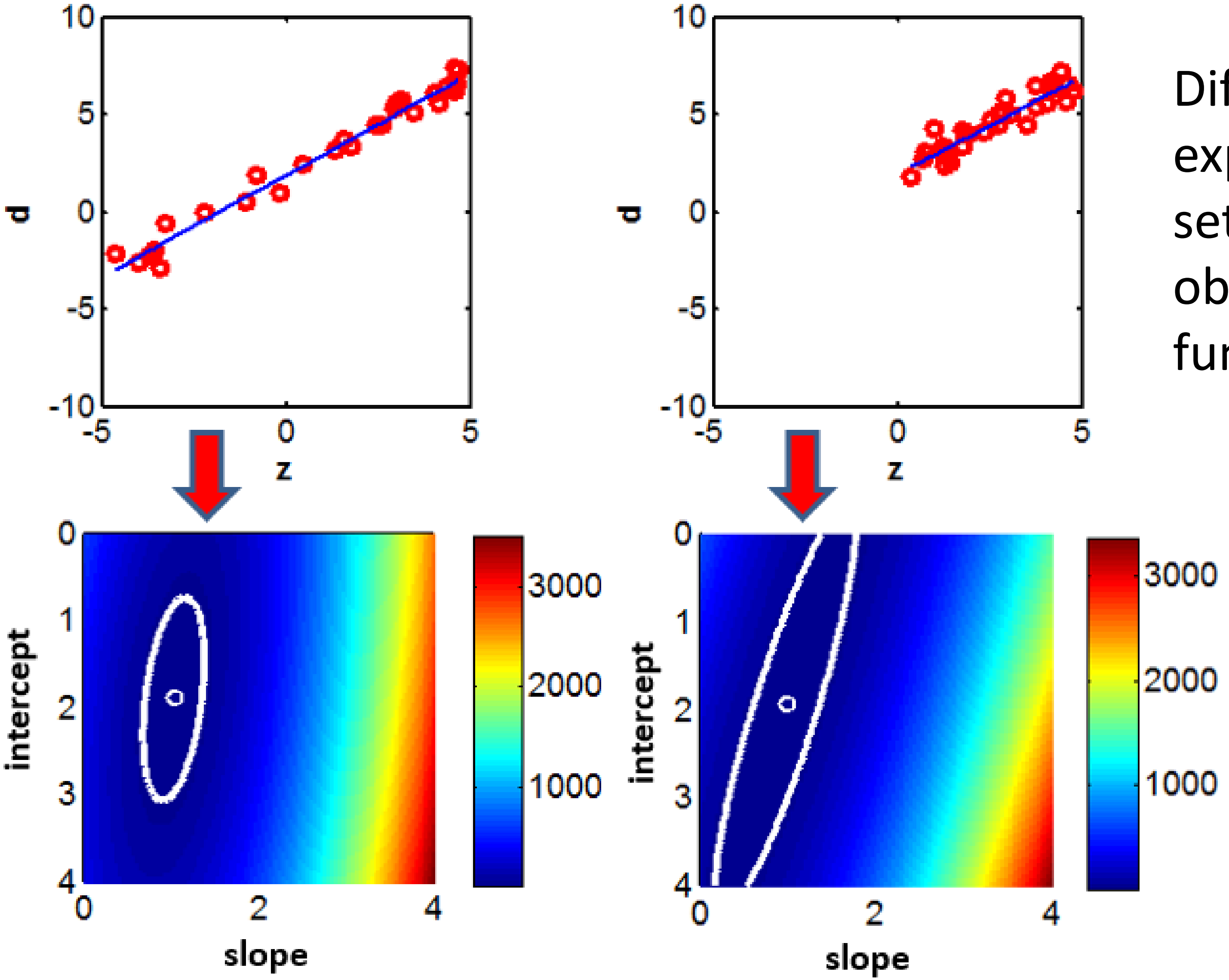
Error functions for different noise levels





The position of the minimum moves away from the theoretical (optimal) position (intercept=0 slope=1) as the noise contamination increases

Objective functions

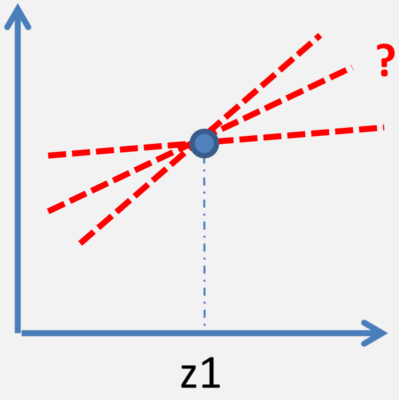


Different experimental settings: different objective functions!

Existence of the L2 norm solution

We have found the solution that offers the best fit (e.g. in the sense of the L2 norm) with the observed data. In doing that, we assume that the solution is **unique**. This least squares approach fails if this assumption is not verified. In this case **multiple solutions** (more models) give the same L2 norm fit with the observed data.

For example, what happens is we have a single observation and we want to find the best-fitting line in the L2 norm sense



$$M \rightarrow number\ of\ data\ points$$

$$(G^T G)^{-1} = \begin{bmatrix} M & \sum_{i=1}^{M} z_i \\ \sum_{i=1}^{M} z_i & \sum_{i=1}^{M} z_i^2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & z_1 \\ z_1 & z_1^2 \end{bmatrix}^{-1}$$

The inverse of a matrix is directly proportional to the determinant:

$$(G^T G)^{-1} \propto 1/(z_1^2 - z_1^2)$$

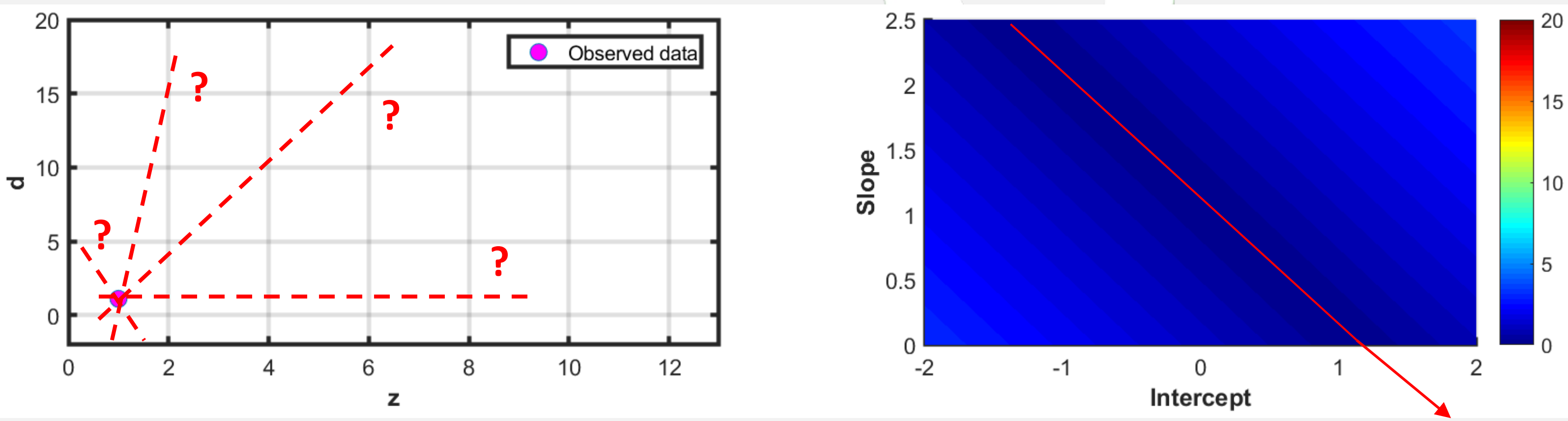
The least-squares formula fails. G is not full columns rank because is made of linearly dependent columns

Under-determined problems (i.e. more unknowns than observations)

Under-determined problems

What does it happen if the experimental data completely overlap each other or we have a single observed data?

Valley of minima. Infinite models correspond to a L2 norm data misfit equal to 0. We have infinite models that perfectly fit the data.



Models that perfectly fit the data

Infinite minima in the objective function. Infinite models **PERFECTLY** fit the data!! We need additional information (or a different objective function) to choose one model from this infinite ensemble

Under-determined problems

Let us consider an undetermined problem in which the number of unknowns is greater than the number of observed data. Therefore, there exists more than one solution that offers a null data prediction error (L2 norm). In this context, the problem has multiple solutions. In this case, to solve the problem additional information must be infused into the inversion framework to select a single model from the ensemble of infinite solutions. To do so, we add to the problem (Gm = d) some additional information that are usually called the **a priori information**. The a priori information can be derived from a previous knowledge about the solution or inferred from previously acquired data (but independent from the data **d**). For example, in the case of a straight line passing through a single point, this a priori information could be that the regression line passes through the origin of the axes. Adding this additional information enable us to solve the inverse problem and get a single solution. In other terms, this additional information allows for the extraction of a single model (line) from the ensemble of possible solutions (multiple lines).

Minimum L2 norm solution in the model space

The simplest a priori information that can be infused into the inversion framework of an under-determined problem, is that the final solution must be the “simplest” between the ensemble of possible solutions. This simplicity is usually quantified by the length of the vector m representing this solution. The most popular measure of simplicity is given by the Euclidean (L2) norm of the model vector:

$$L = m^T m = \sum_{i=1}^{N} m_i^2$$ N: number of unknwons

The Error function becomes:

Minimize $L = m^T m$ under the **constraint** that $d - Gm = 0$.

This minimization problem can be solved with Lagrange multipliers (see the following slides for details).

The method of Lagrange multipliers is a strategy for finding the local maxima and minima of a function subject to equality constraints (i.e., subject to the condition that one or more equations have to be satisfied exactly by the chosen values of the variables)

Minimum L2 norm solution

$$F(m) = m^T m + \lambda^T (d - Gm) \longrightarrow$$

The Lagrangian function

In matrix form and for all the model parameters, we obtain the following derivative with respect to m:

$$\nabla_m F(m) = 2m^T - \lambda^T G$$

The gradient must be equal to zero. If we additionally add the constraint d = Gm we get:

$$\begin{cases} 2m = G^T \lambda \\ d = Gm \end{cases}$$

We can derive m from the first equation and substitute it in the second equation:

$$d = G(G^T \lambda / 2)$$

Then:

$$\lambda = 2(GG^T)^{-1} d$$

If we substitute the so obtained λ value into the first equation, we get:

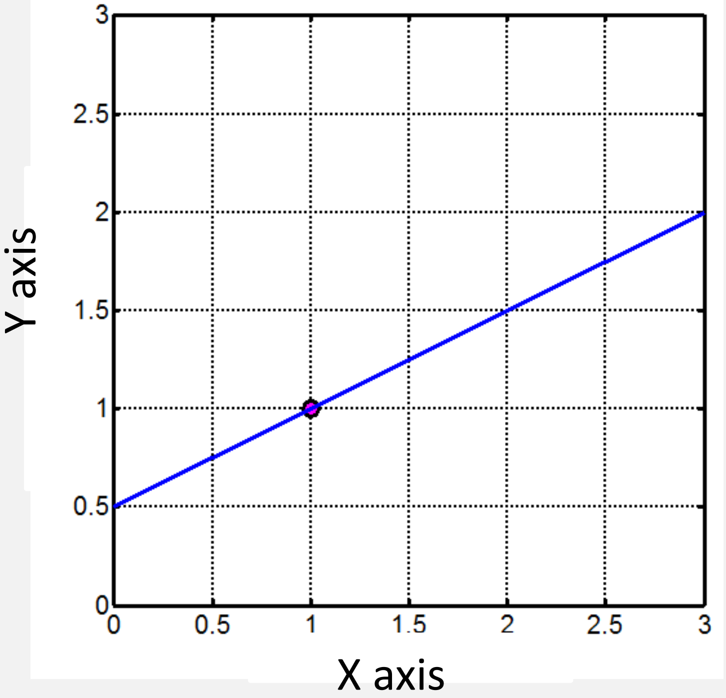
$$\boxed{m = G^T (GG^T)^{-1} d}$$

This equation represents the “simplest” solution for an underdetermined problem.

Minimum L2 norm solution in the model space

Finding a straight line when we have one single observation.

Final solution: Intercept=0.5; slope=0.5

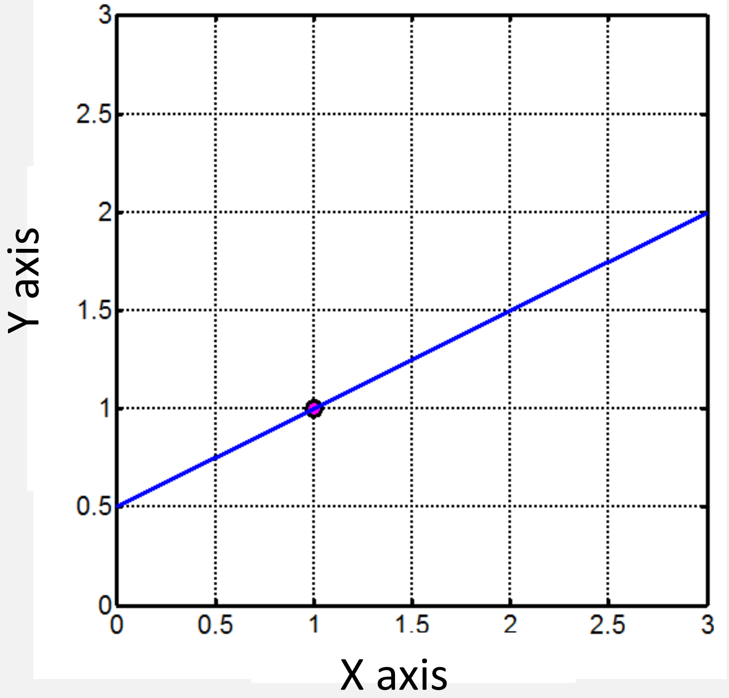


→ A straight line passing through a single point obtained by solving the minimum L2 norm solution in the model space.

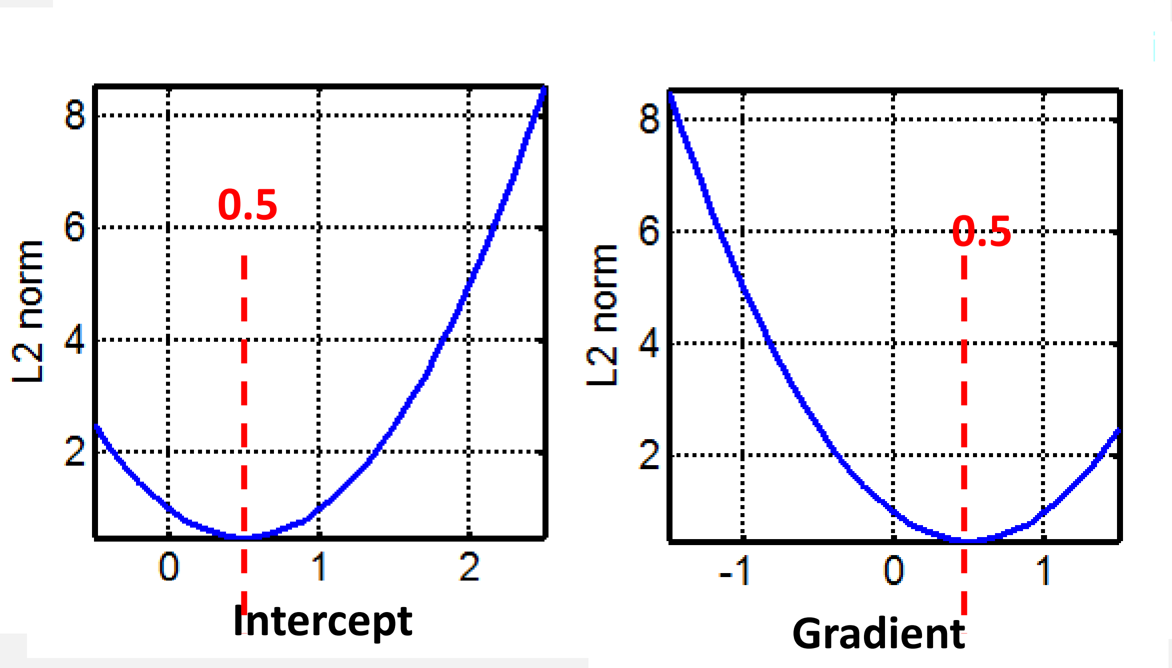
Minimum L2 norm solution in the model space

Finding a straight line when we have one single observation.

Final solution: Intercept=0.5; slope=0.5

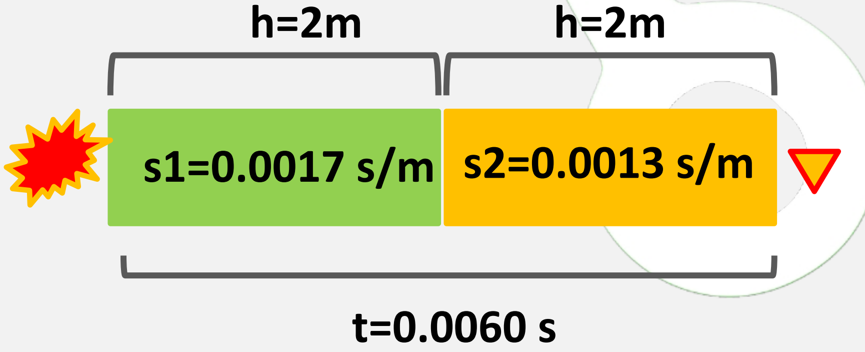


For all the combinations of intercept and slope that satisfies d=gm we compute the L2 norm. We get:



A geophysical example of an under-determined problem

A single block with length 2h, with two cells with different velocities. A single source and a single receiver that measures the traveltime t1 of a transmitted seismic ray. The goal is to retrieve the slowness (s) of the two blocks.



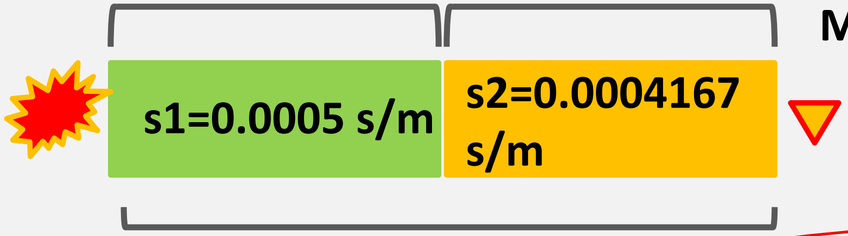
$$\boldsymbol{G} = [2\ 2]$$

Obviously, G has two linearly dependent columns and it has not full-column rank!! Therefore, the standard least squares solution does not exist !!!

On the contrary, there exists the minimum L2 norm solution in model space. This solution gives slowness for the 2 blocks equal to 0.0015 s/m. This is exactly the average value of the two actual slownesses!

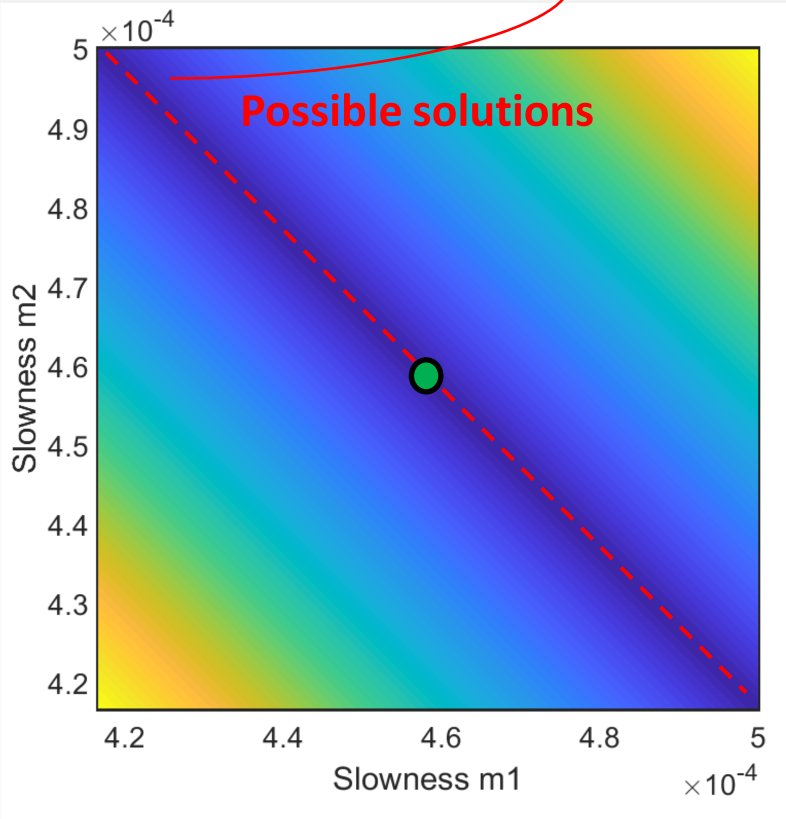
In this context, the only univocally determined parameter is the average slowness of the two blocks.

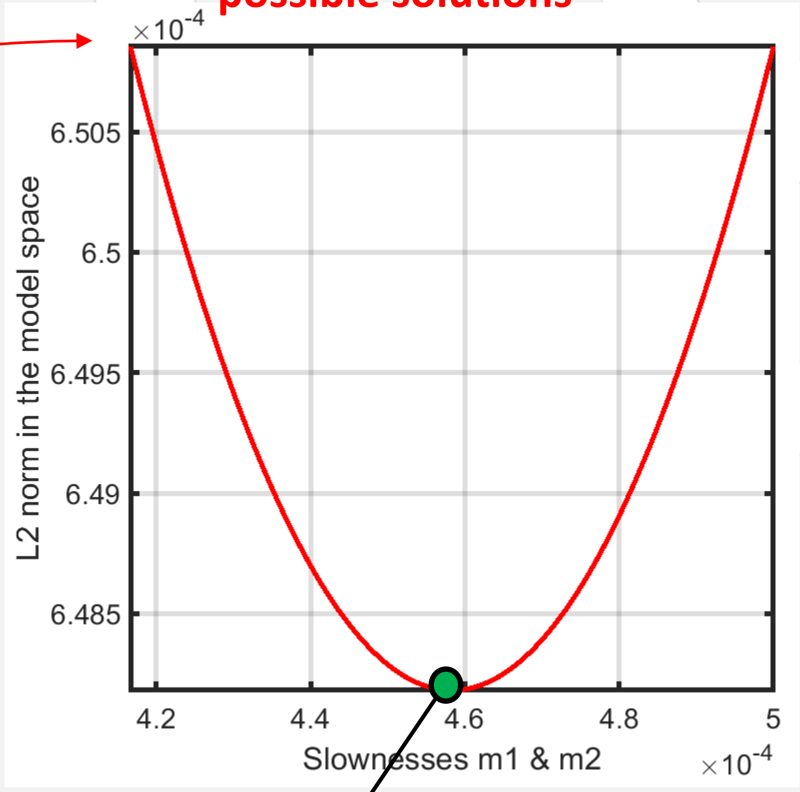
Objective function



Mean slowness = 4.5833e-04 s/m

The L2 length for the ensemble of possible solutions

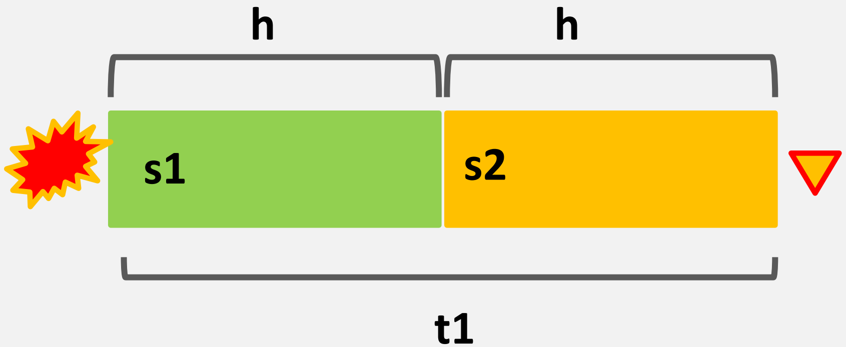




The minimum gives the **predicted model**. This minimum corresponds to the average slowness values of the two cells

Geometrical meaning of the minimum L2 norm solution

A single block with length 2h, with two cells with different velocities. A single source and a single receiver that measures the traveltime t1 of a transmitted seismic ray. The goal is to retrieve the slowness (s) of the two blocks.



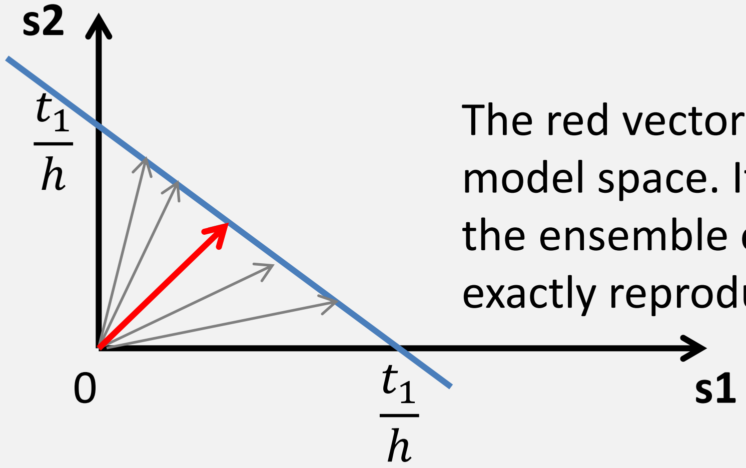
Gm=d

$$[h\ h]\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = t_1$$

G has 2 linearly dependent columns and for this reason it has not full-column rank!! The standard lest-square solution does not exists!!!

The analytical expression of this problem:

$$hs_1 + hs_2 = t_1 \ \rightarrow hs_2 = -hs_1 + t_1 \ \rightarrow \ s_2 = -s_1 + \frac{t_1}{h}$$

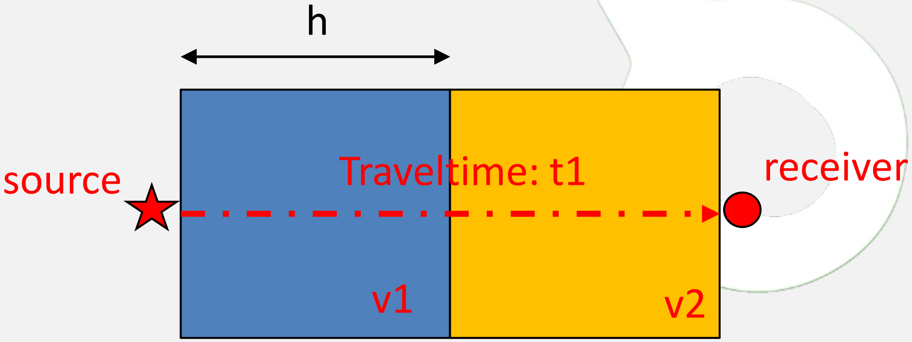


The red vector represents the L2 minimum norm solution in the model space. It is the vector with the minimum length between the ensemble of gray vectors (representing possible solutions that exactly reproduce the observations).

Some other examples

Our aim is to estimate the P velocity of the two blocks by measuring the traveltimes of the P-wave.

We consider straight ray-paths. For example:



$$t_1 = \frac{h}{v_1} + \frac{h}{v_2} = hs_1 + hs_2$$

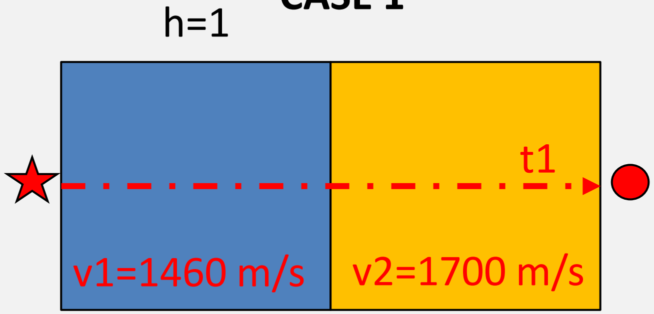
Analytical form

$$[t_1] = [h\ h]\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \longrightarrow \mathbf{d} = \mathbf{Gm}$$

Matrix form

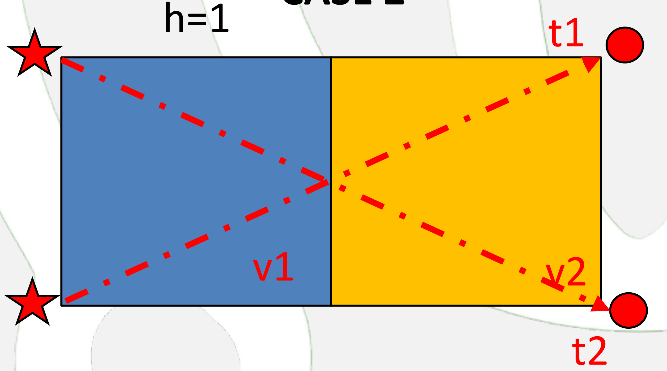
Some other examples

CASE 1



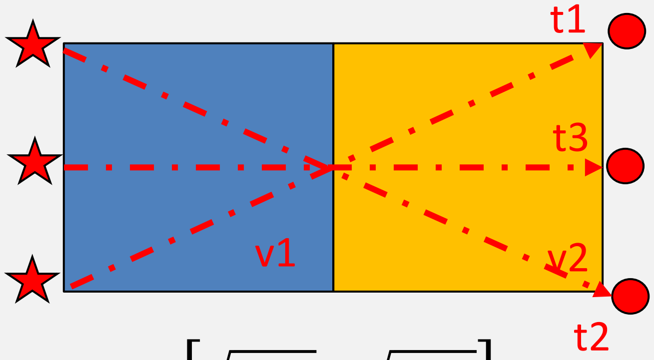
$$[t_1] = [1\ 1]\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

CASE 2



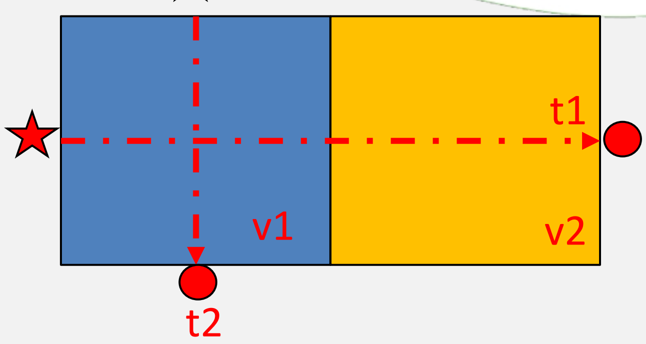
$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} \sqrt{1,25} & \sqrt{1,25} \\ \sqrt{1,25} & \sqrt{1,25} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

CASE 3



$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} \sqrt{1,25} & \sqrt{1,25} \\ \sqrt{1,25} & \sqrt{1,25} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

CASE 4

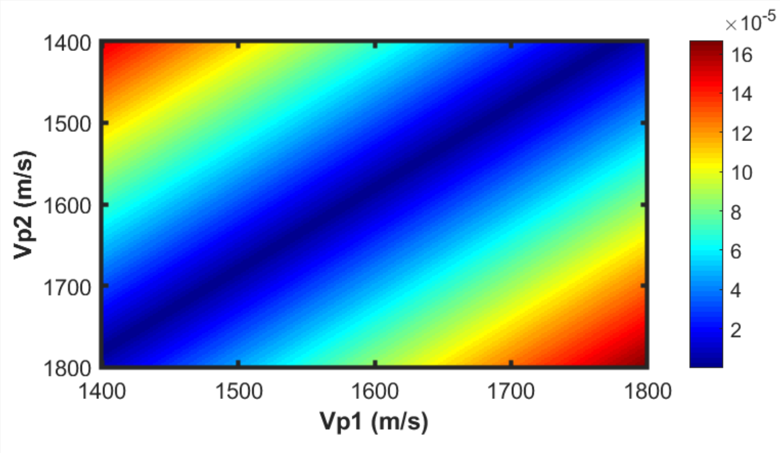


$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

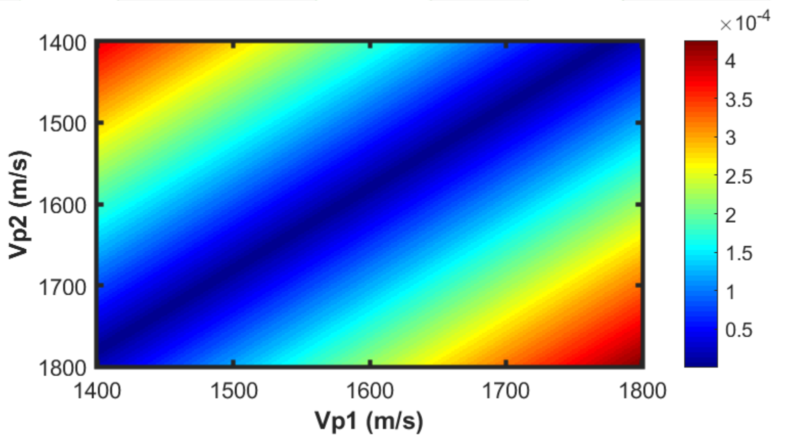
Objective functions

$$\mathrm{E}(\mathbf{m}) = \left|\left|\mathbf{d} - \mathbf{Gm}\right|\right|_2^2 \longrightarrow \mathrm{E}(\mathbf{m}) = ||obs.traveltime - pred.traveltimes||_2$$

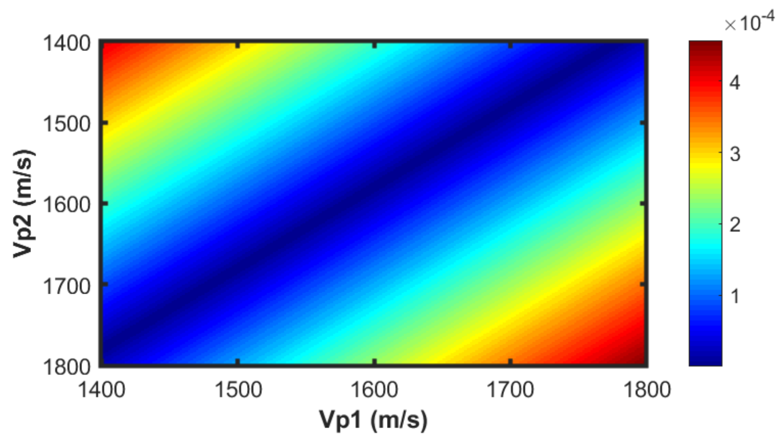
CASE 1



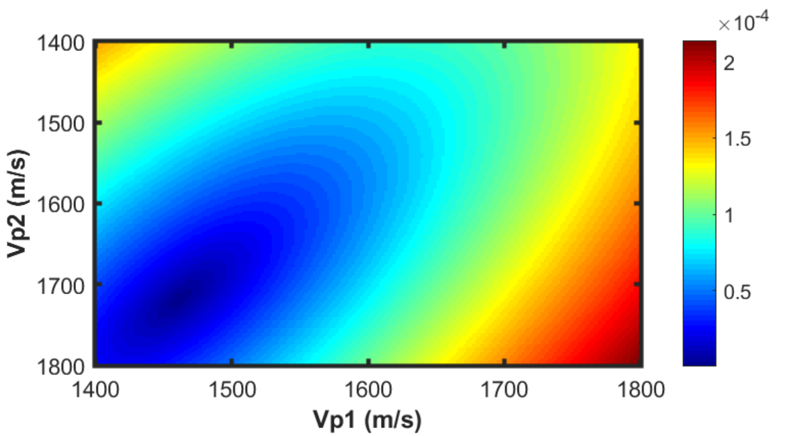
CASE 2



CASE 3



CASE 4



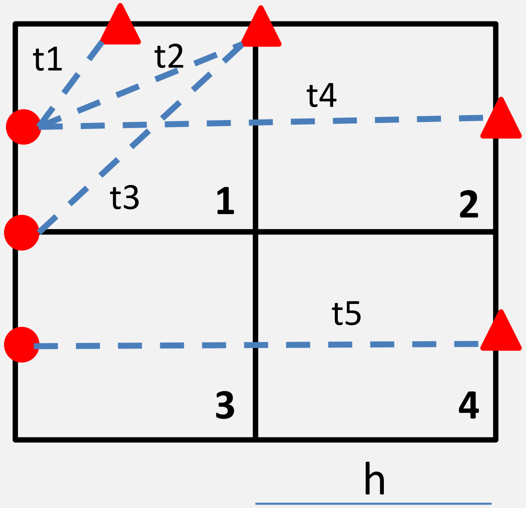
Over- or under-parameterization?

In an inverse problem we can contemporarily have model parameters well-constrained, poorly, or even not constrained by the data.

Example of a seismic tomography:

$$d = Gm$$

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{bmatrix} = \begin{bmatrix} h/\sqrt{2} & 0 & 0 & 0 \\ \sqrt{5}h/2 & 0 & 0 & 0 \\ \sqrt{2}h & 0 & 0 & 0 \\ h & h & 0 & 0 \\ 0 & 0 & h & h \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix}$$



G has 3 linearly dependent rows. Indeed, the traveltimes t1, t2 and t3 are all referred to the same block, and the associated equations are linearly correlated. The rank of G (which corresponds to the number of linearly independent columns) is equal to 3. Block 1 is overdetermined (constrained by the traveltimes t1, t2 and t3); block 2 is exactly determined (constrained by traveltime t4), while blocks 3 and 4 are under-determined (constrained by a single observation, t5).

Mixed-determined problems

Very often in solving a geophysical inverse problem we deal with a mixed determined problem. In this context, we have more experimental data than unknown parameters, but the experimental setting is not able to univocally constraint the solution. In other terms, we have multiple solutions that equally fit the observed data. Differently from under-determined problems the match between observed and predicted data is not perfect, and the minimum L2 norm misfit value is not zero. In other words some model parameters are overdetermined and others underdetermined. A typical example in this case is the seismic tomography in which some parameters are well constrained by the data while others are not. Ideally in this case we would like to subdivide the parameters of the model into 2 groups: the overdetermined and the underdetermined ones and solve them in some way separately: however, this strategy is usually not applicable because of the experimental set up.

A usually employed strategy imposes a compromise between the minimization of the L2 norm data misfit and the L2 norm length of the model vector:

$$E(m) = (d - Gm)^T (d - Gm) + \varphi m^T m \longrightarrow \min \left\| \begin{bmatrix} G \\ \varphi I \end{bmatrix} m - \begin{bmatrix} d \\ 0 \end{bmatrix} \right\|_2^2$$

where φ represents an arbitrary weight that establishes the importance of a solution strategy (minimum L2 norm data misfit) with respect to the other (minimum norm in the model space).

Mixed-determined problems

As usual the solution that minimizes the error function E(m) can be found by setting the gradient of the objective function to zero.

$$E(m) = d^T d - 2d^T Gm + m^T G^T Gm + \varphi m^T m$$

$$\downarrow$$

$$\nabla_m E(m) = -2d^T G + m^T\left((G^T G)^T + G^T G\right) + 2\varphi m^T \rightarrow m^T(G^T G + \varphi I) = d^T G \rightarrow$$

$$m^T = d^T G(G^T G + \varphi I)^{-1} \rightarrow \boxed{m = (G^T G + \varphi I)^{-1} G^T d}$$

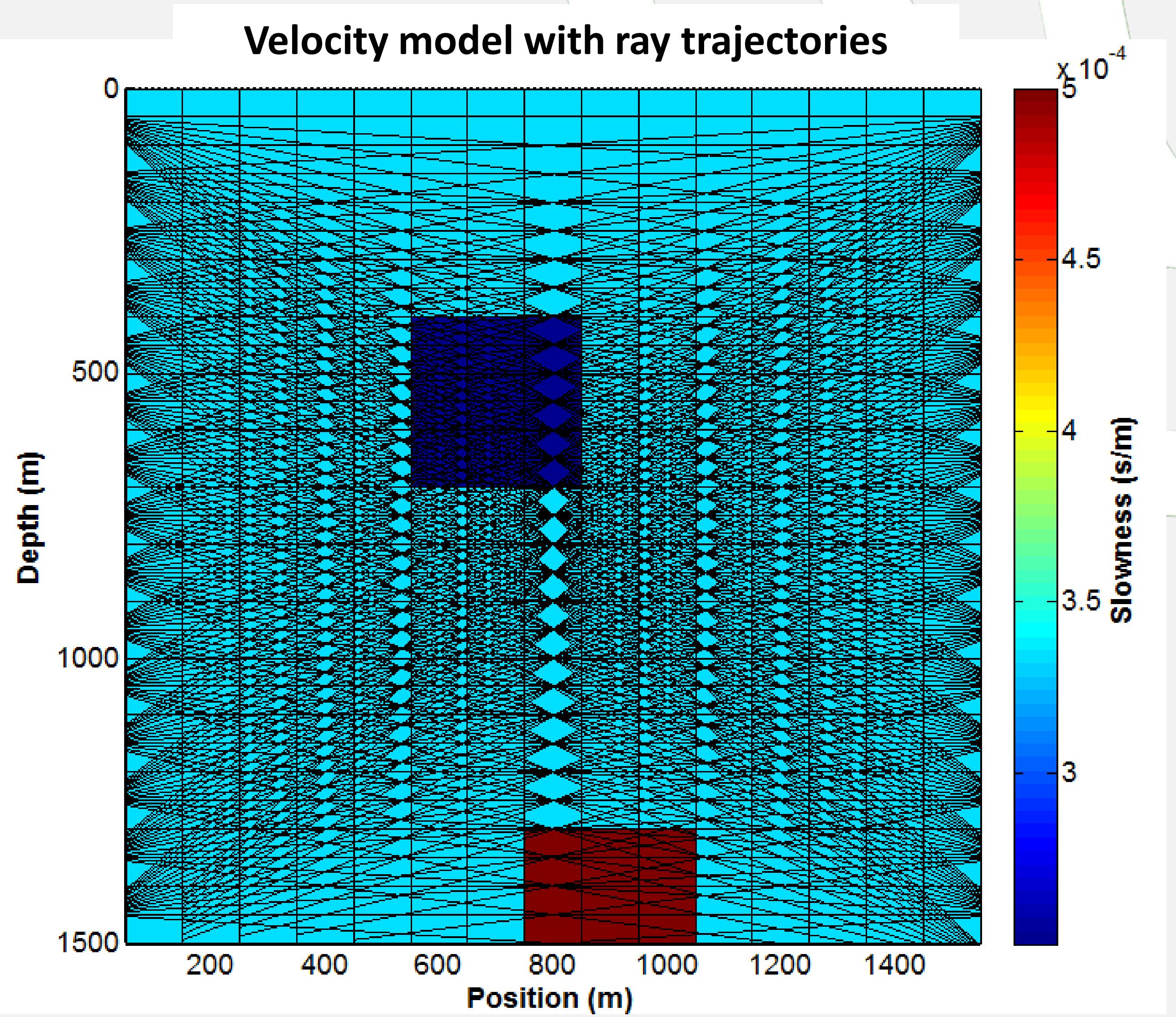
$$\downarrow$$

I -> *NxN* indentity matrix (*N*: number of unknowns)

This is the **damped least-squares** solution for a mixed-determined inverse problem.

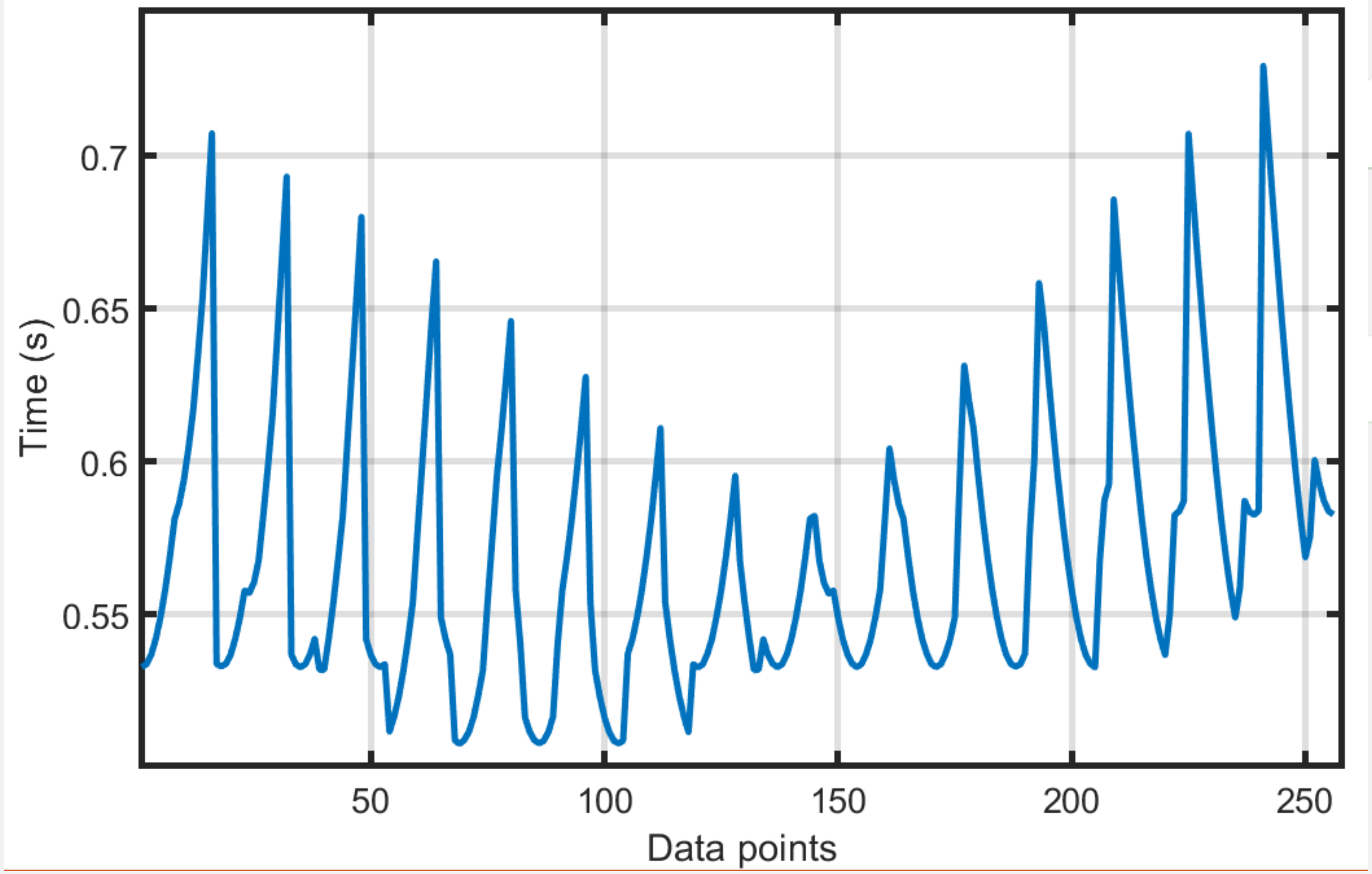
Note: the final solution depends on the selected φ value. We do not enter into details here.

Example of a mixed-determined problem

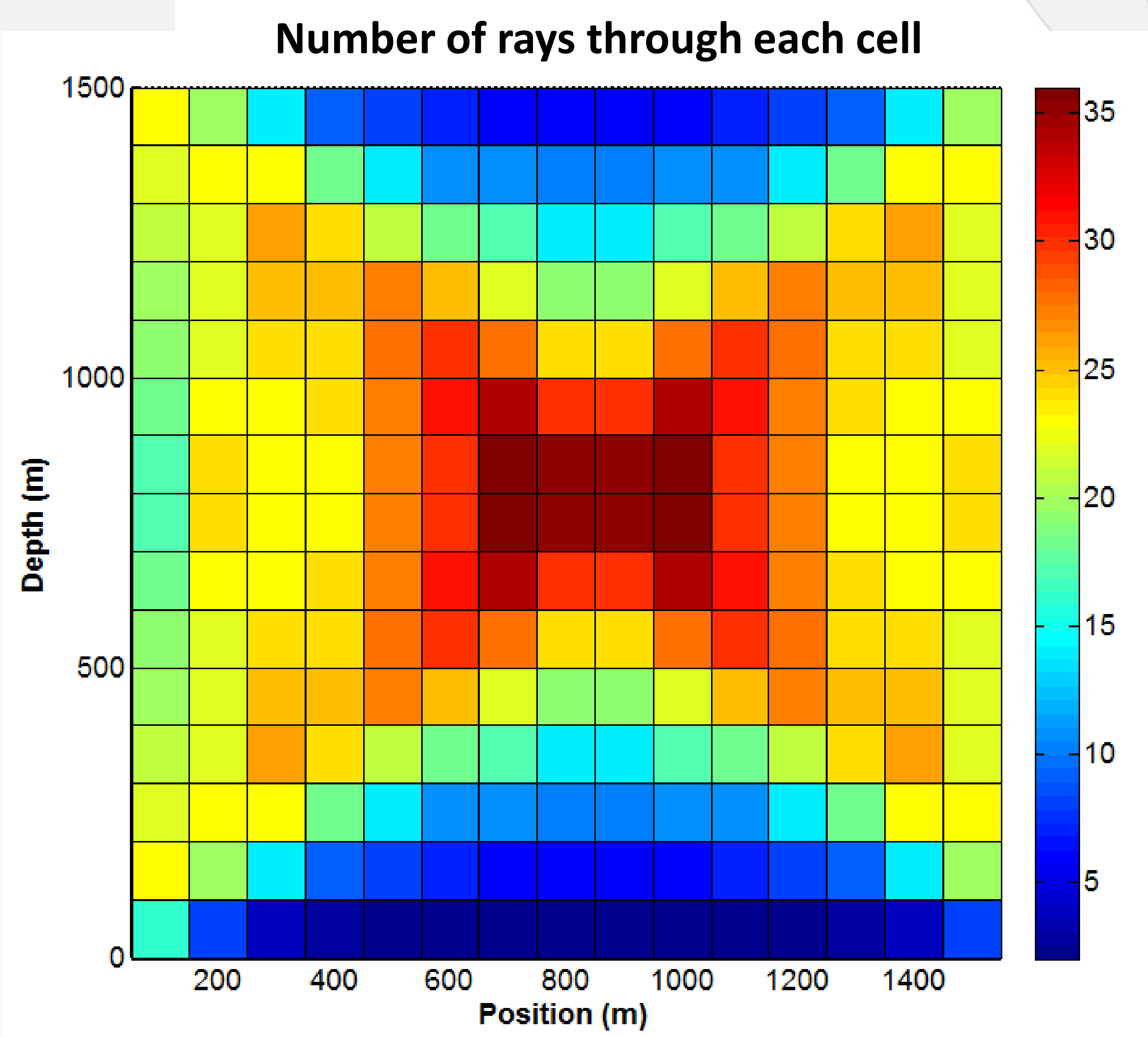


Example of a mixed-determined problem

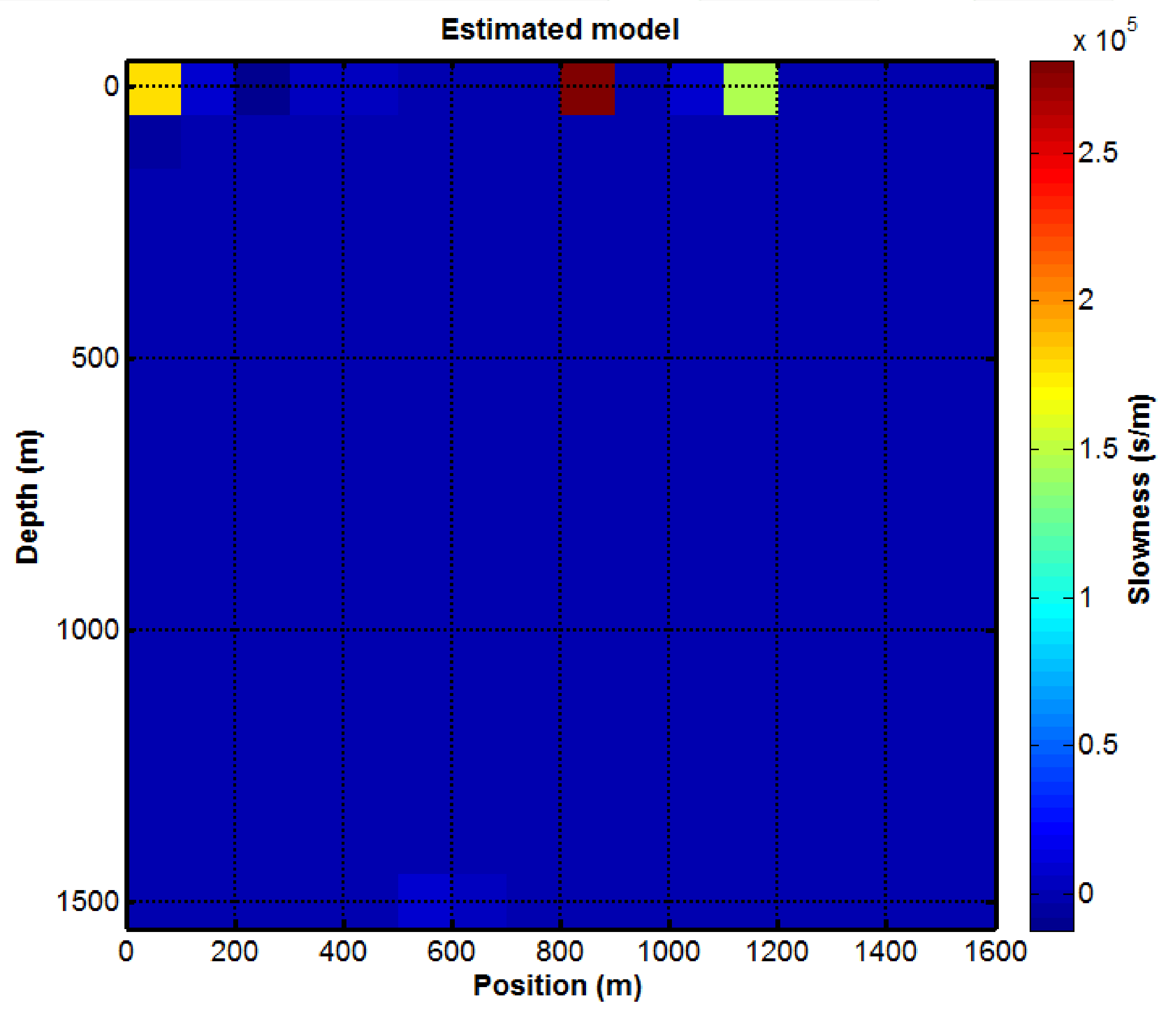
Observed traveltimes



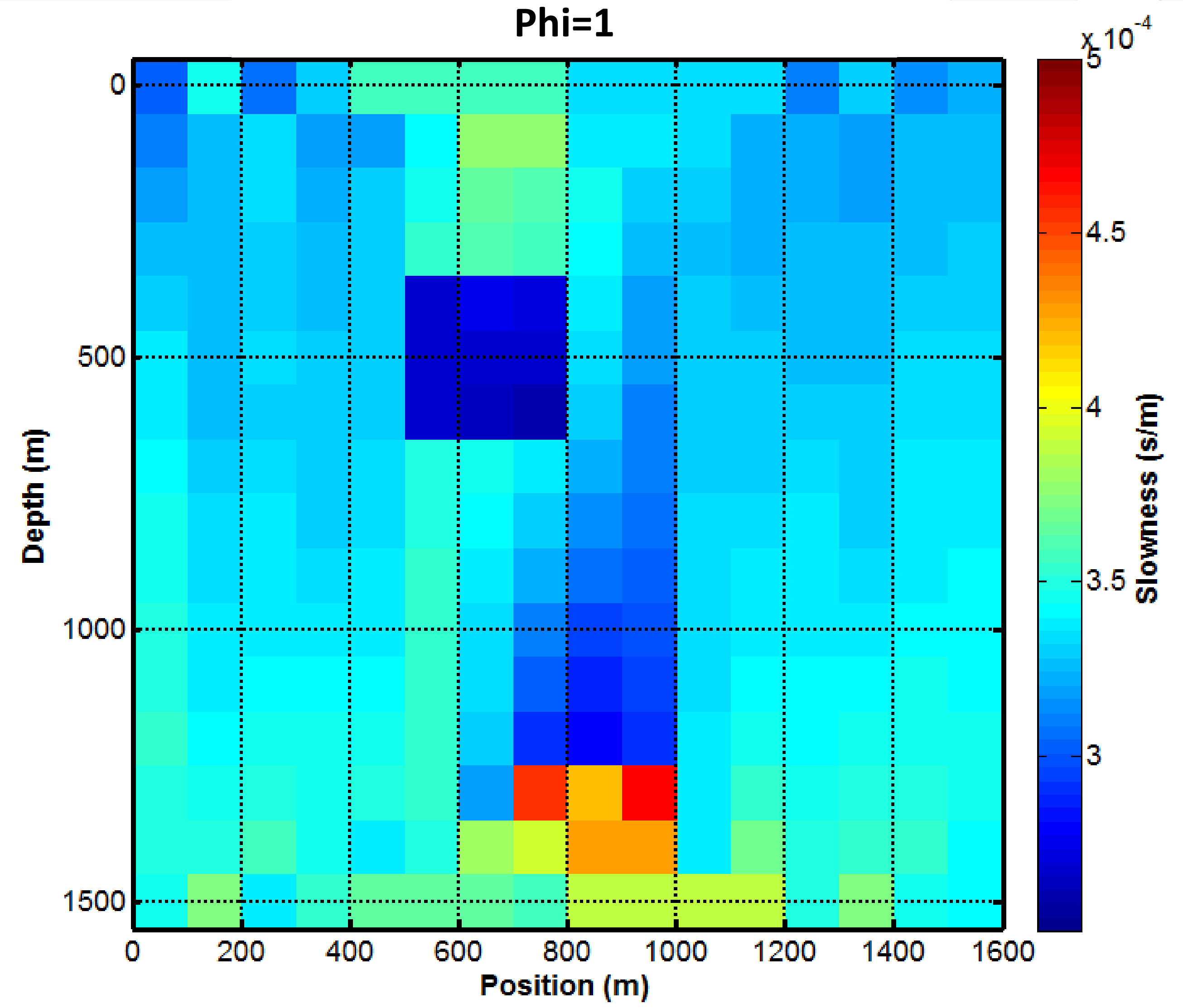
Example of a mixed-determined problem



Standard least-squares solution



Damped least-squares solution



… and now what happens in non linear inverse problems (just the basic!)…

Non linearity

The following properties do not apply to non-linear problems:

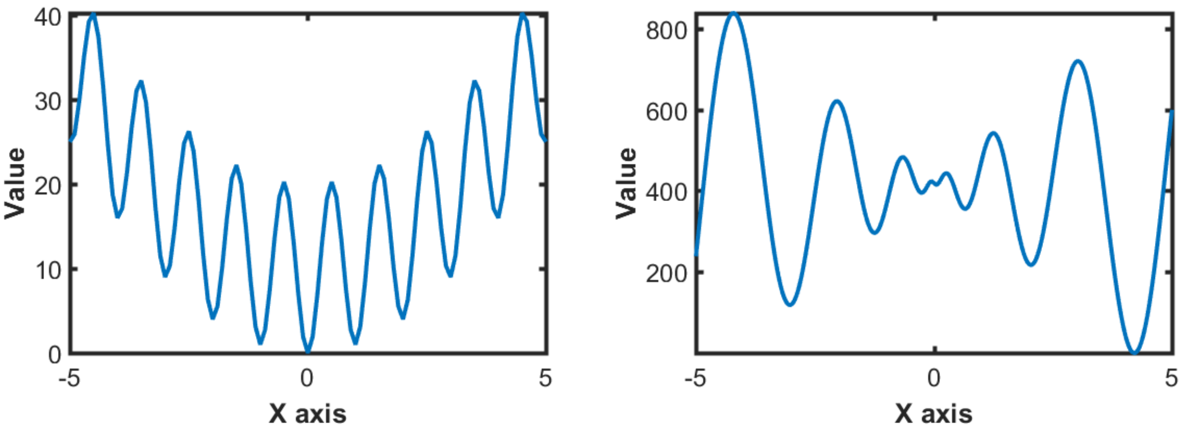
$$G(m_1 + m_2) = G(m_1) + G(m_2) \longrightarrow \text{Superposition law}$$

$$G(am) = aG(m) \longrightarrow \text{Scaling law}$$

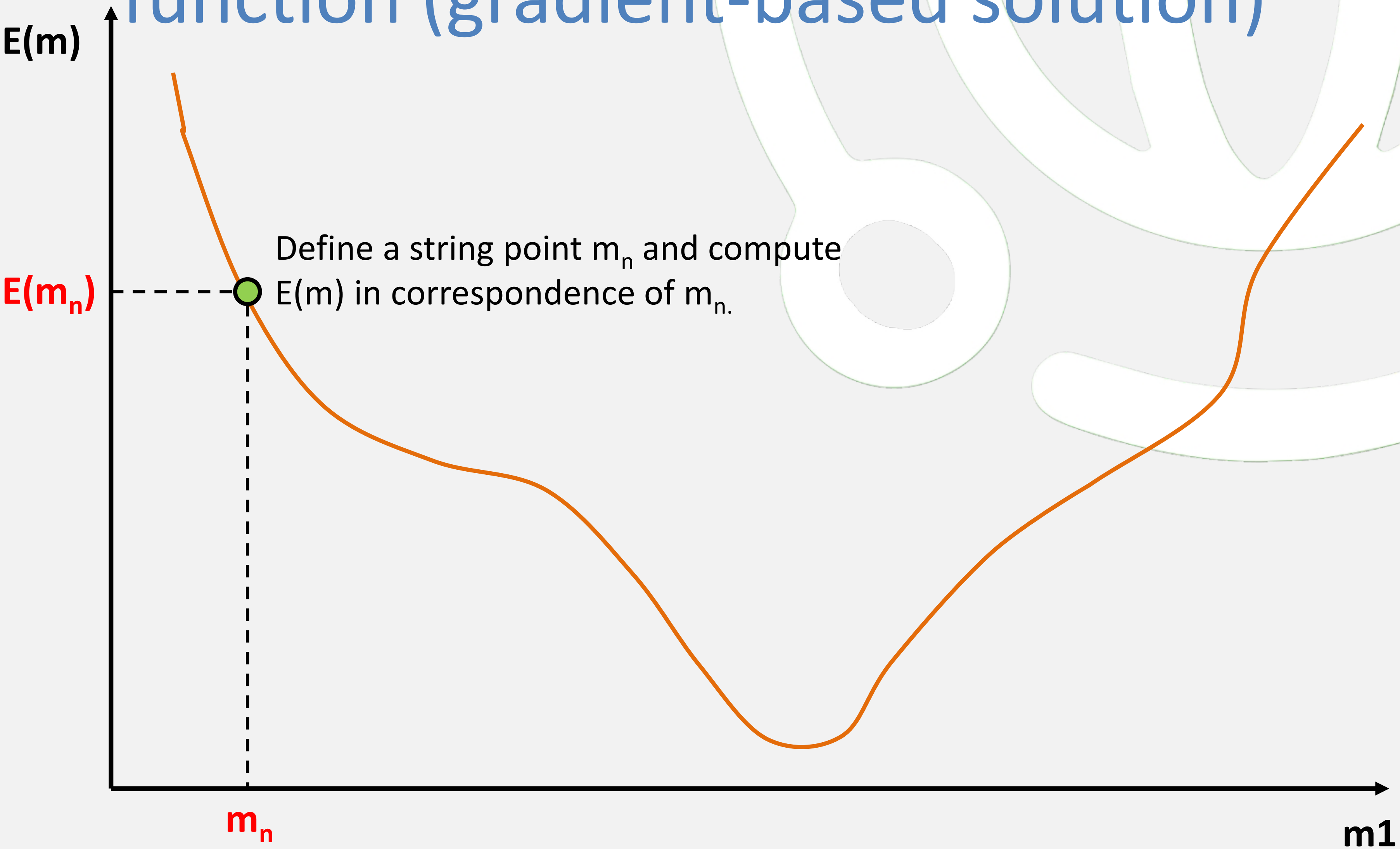
To solve non-linear problems there exists two main strategies:

- Translate the inverse problem into an optimization problem and use a global search algorithm to search the optimal solution in the model space
- Linearize the problem and use a gradient-based method to converge toward the local minimum of the objective function

Differently from linear problems in which the error function is convex with a unique minimum, non-linear problems are often characterized by more complex topologies of the objective function:

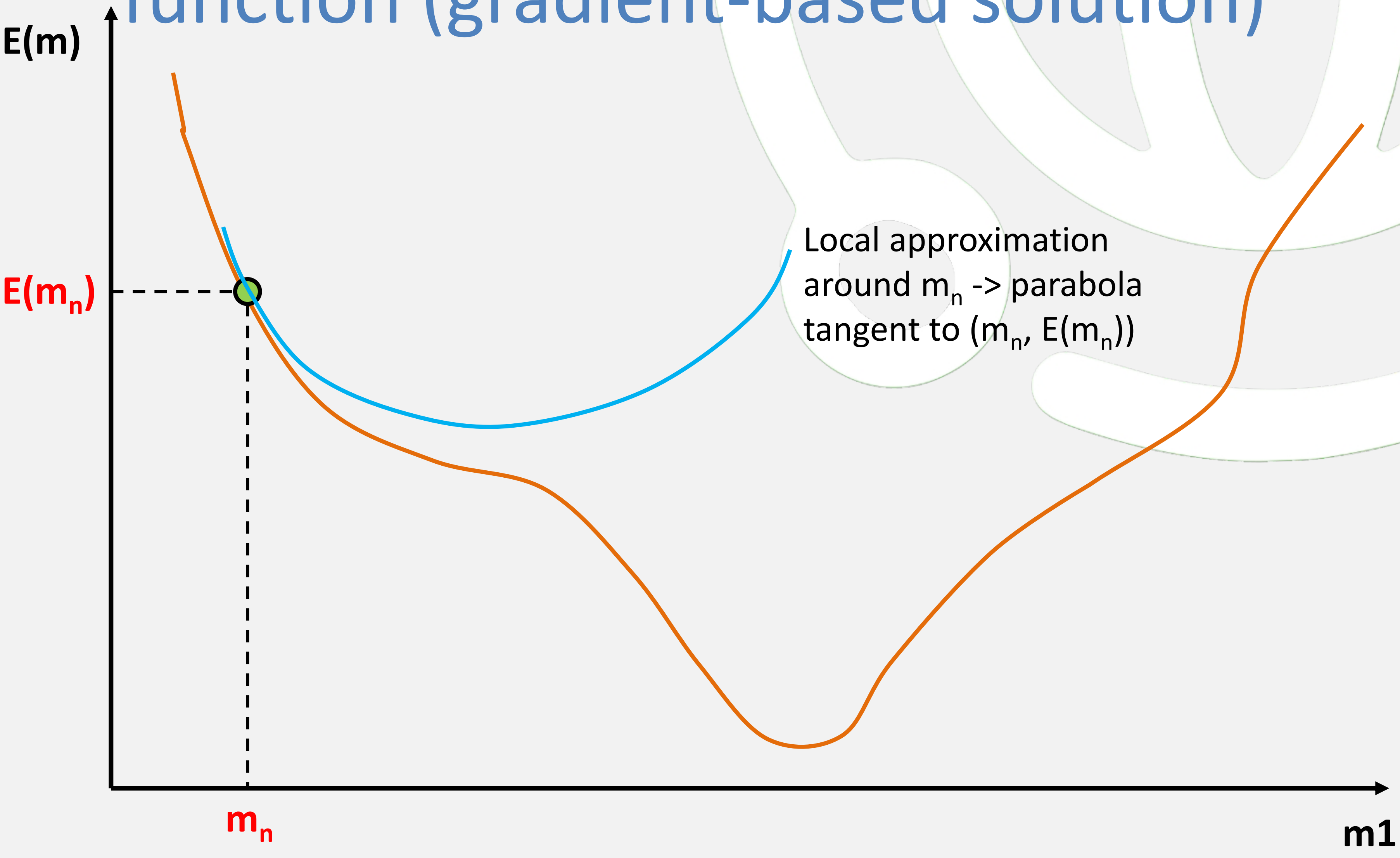


Approximation of the objective function (gradient-based solution)



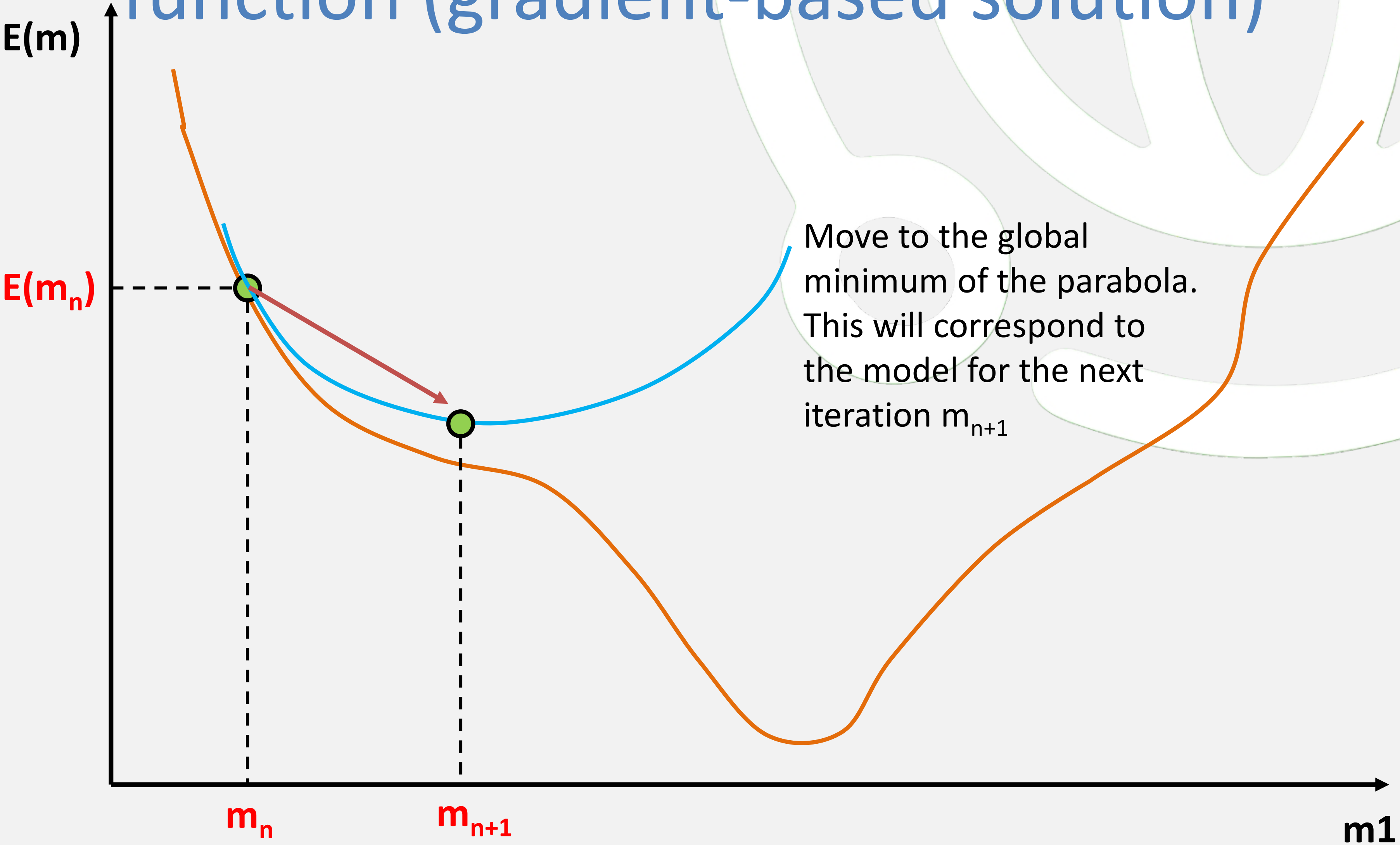
Define a string point m_n and compute E(m) in correspondence of m_n.

Approximation of the objective function (gradient-based solution)



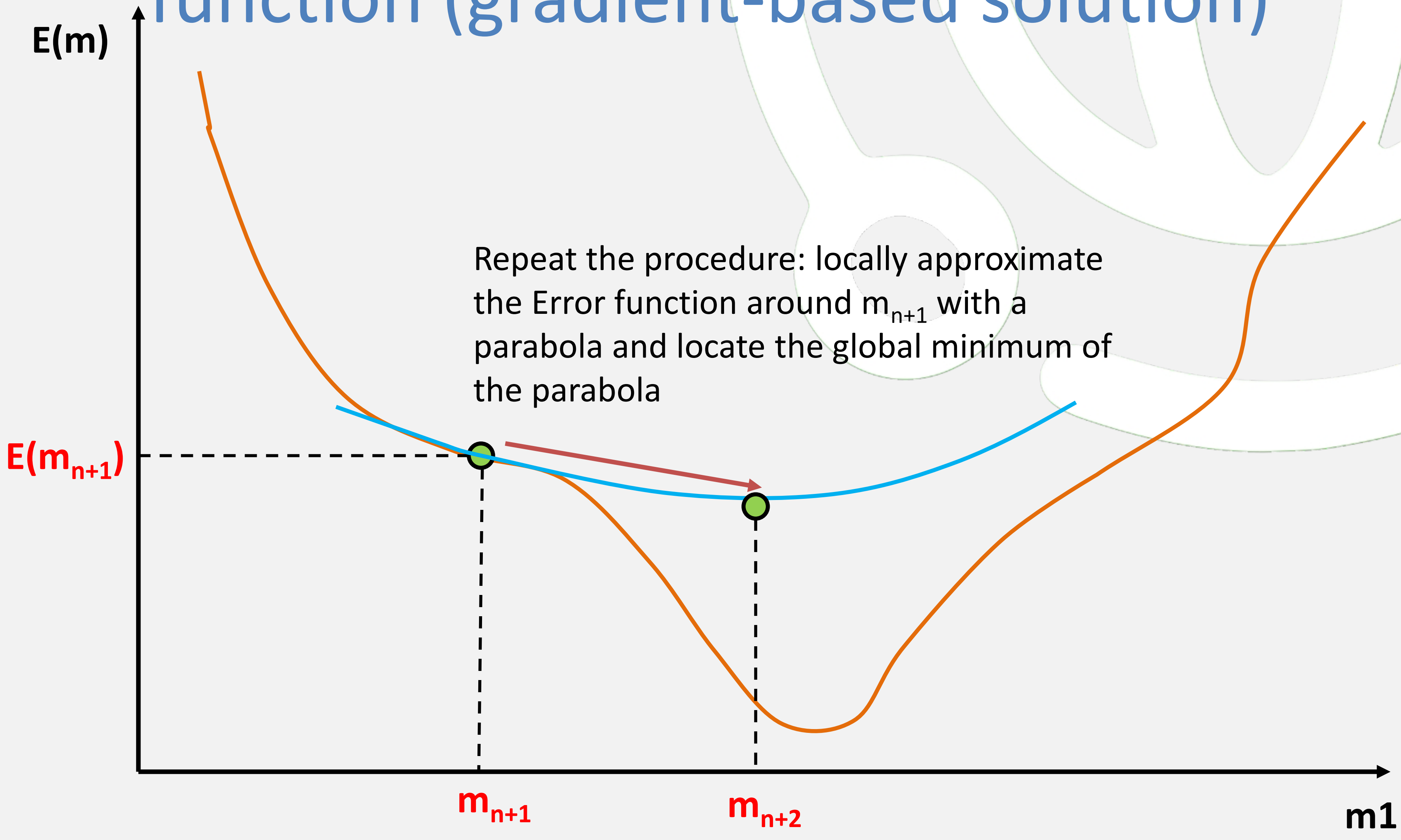
Local approximation around m_n -> parabola tangent to (m_n, $E(m_n)$)

Approximation of the objective function (gradient-based solution)



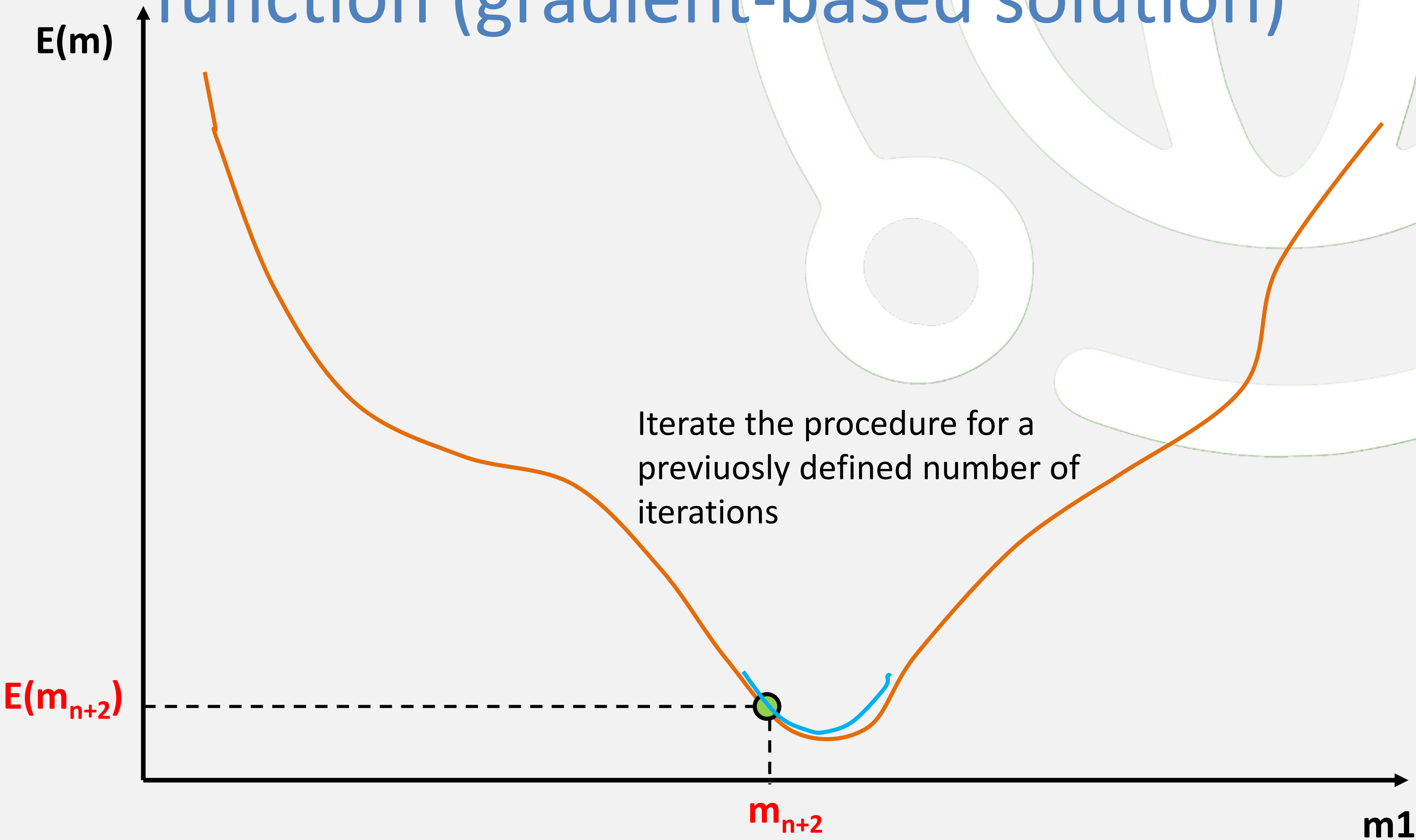
Move to the global minimum of the parabola. This will correspond to the model for the next iteration m_{n+1}

Approximation of the objective function (gradient-based solution)



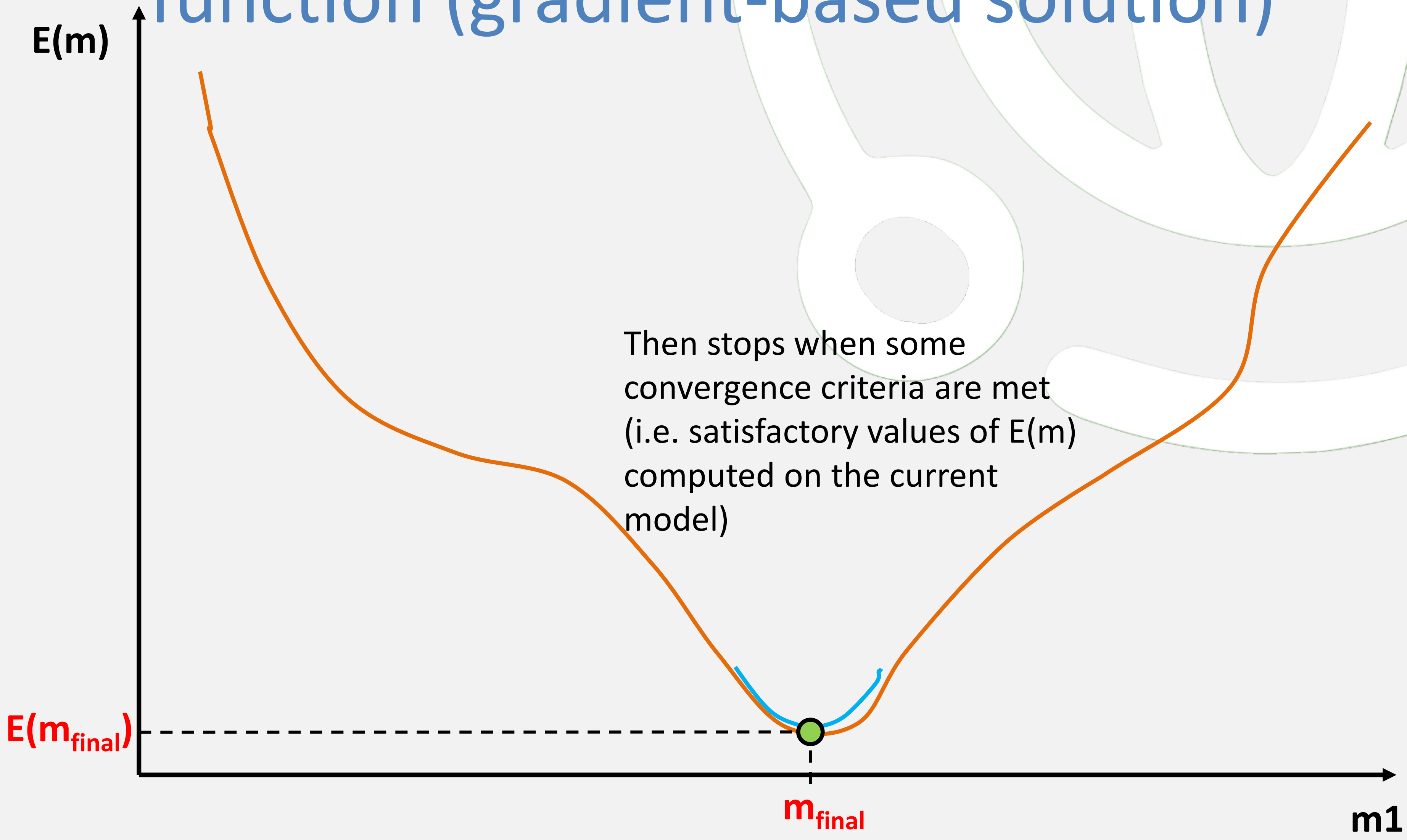
Repeat the procedure: locally approximate the Error function around m_{n+1} with a parabola and locate the global minimum of the parabola

Approximation of the objective function (gradient-based solution)



Iterate the procedure for a previuosly defined number of iterations

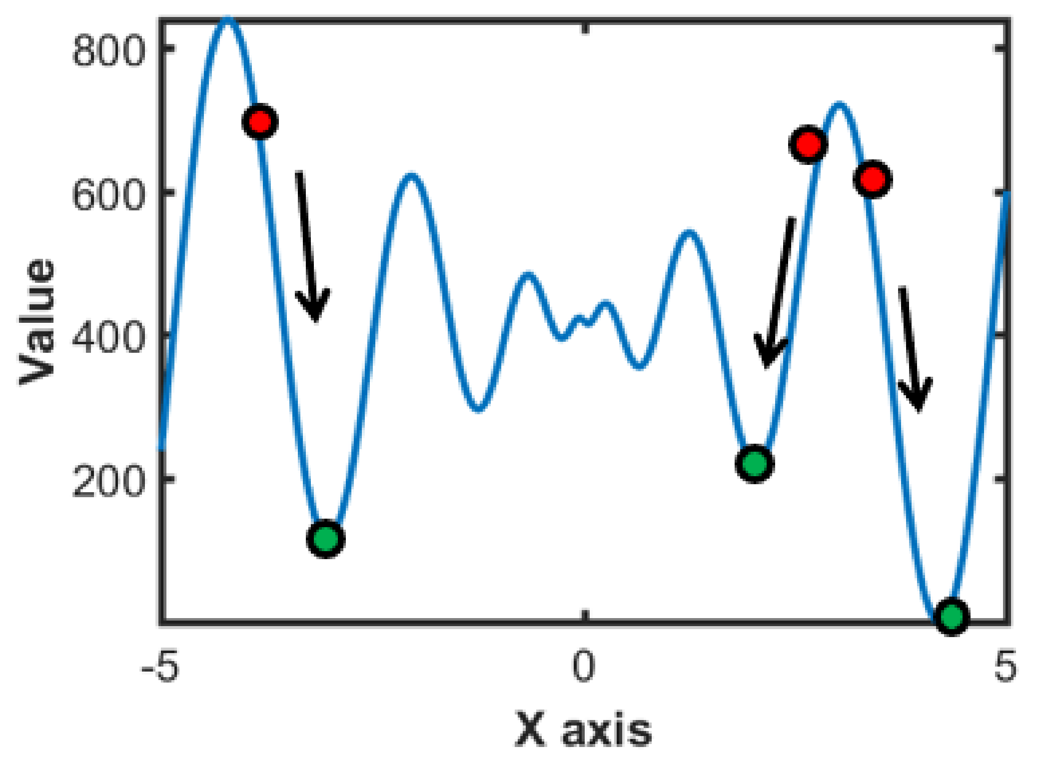
Approximation of the objective function (gradient-based solution)



Then stops when some convergence criteria are met (i.e. satisfactory values of E(m) computed on the current model)

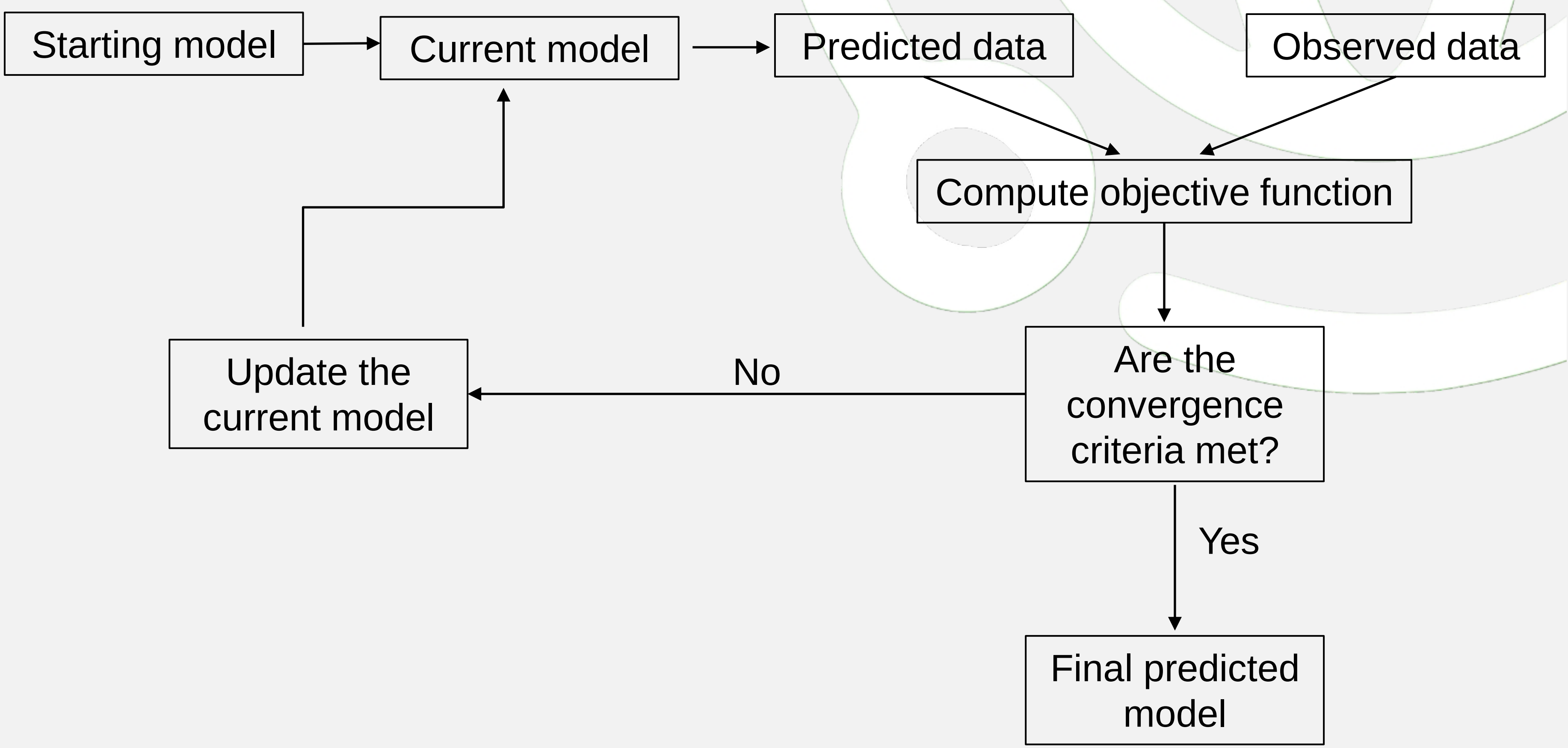
The main issue of this approach

The final solution depends on the starting point because the finally predicted model is located in the minimum valley that contains the starting point -> different starting points could result in different final predictions



- 🔴 Starting model
- 🟢 Final model

Iterative gradient-based inversion for non linear problems


Starting model

Current model

Predicted data

Observed data

Compute objective function

Are the convergence criteria met?

No

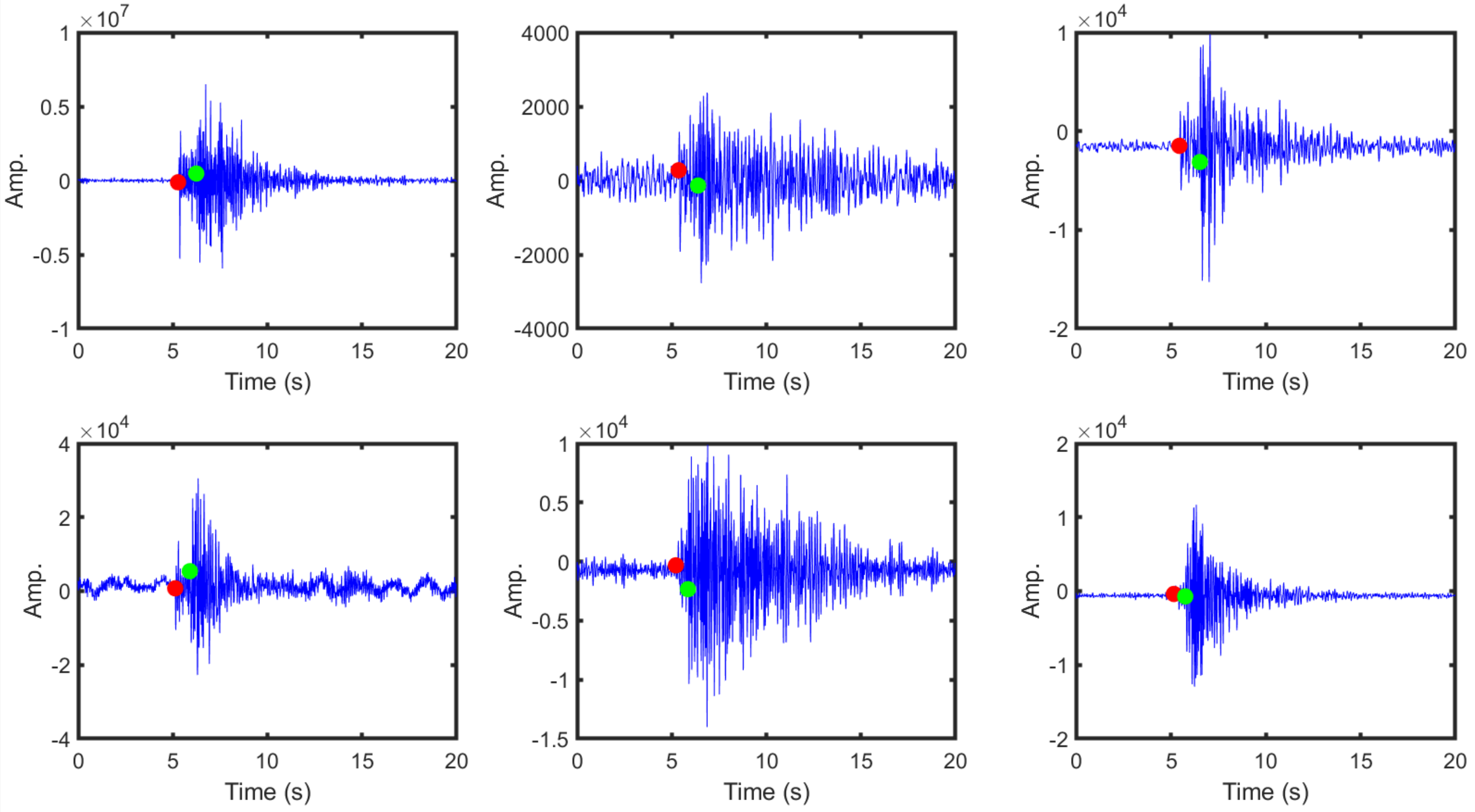
Update the current model

Yes

Final predicted model

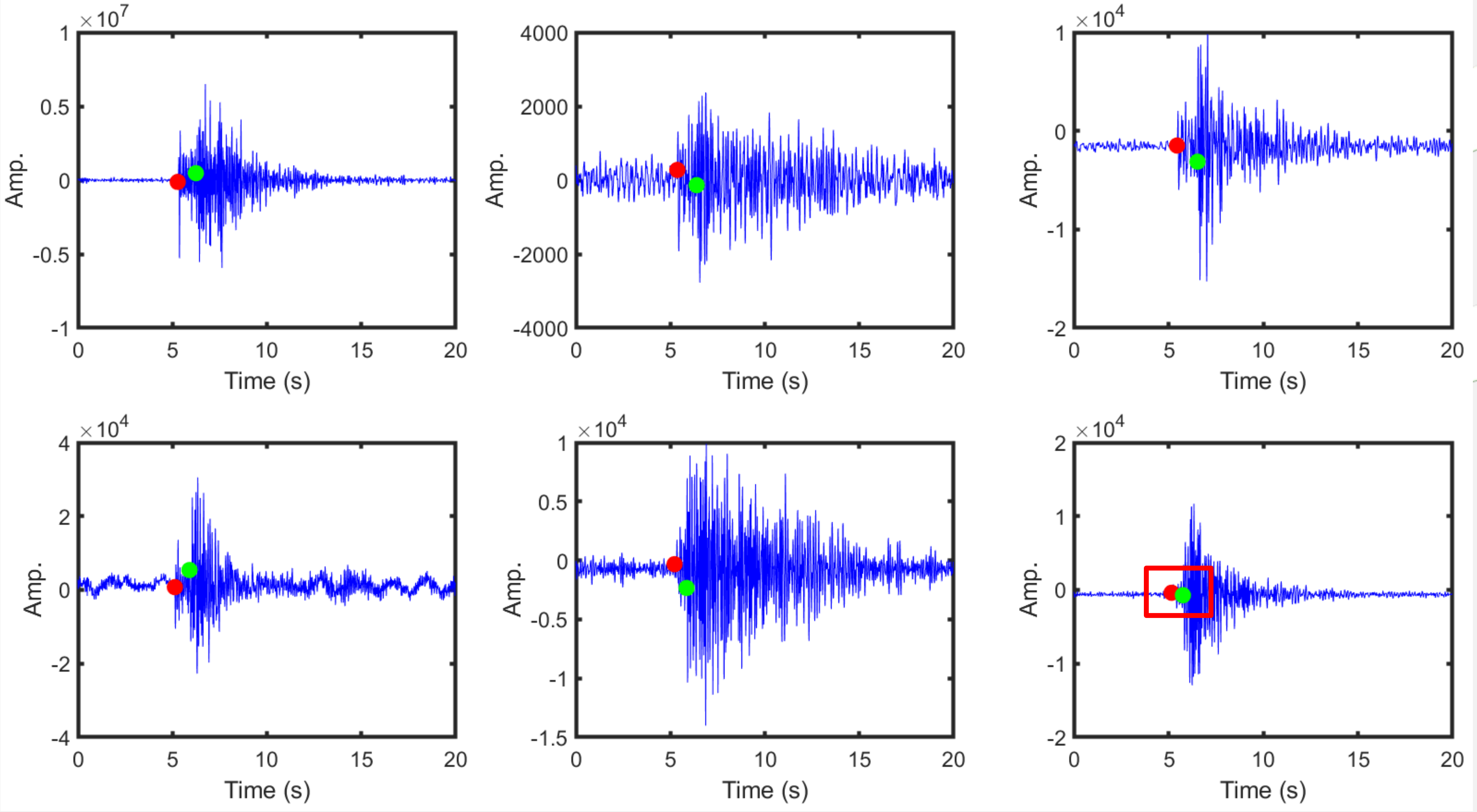
Example: Hypocenter location and T0 estimation

Observed data: P & S arrival times

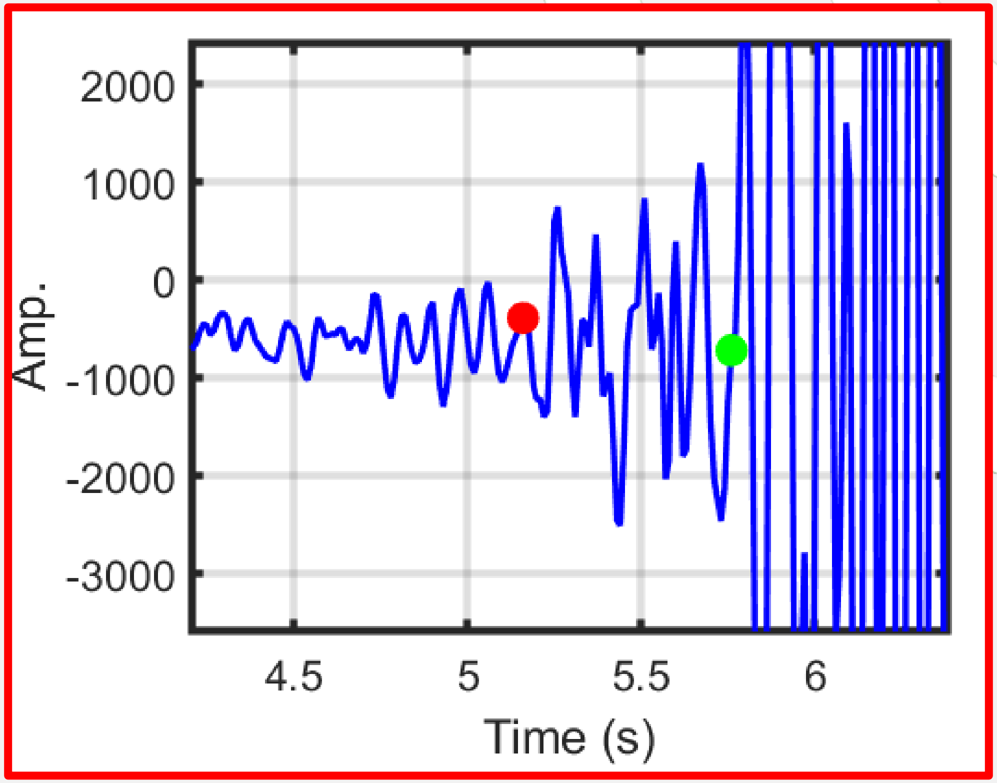


Example: Hypocenter location and T0 estimation

P-wave arrivals S-wave arrivals



Example: Hypocenter location and T0 estimation



Objective functions

Objective function 1:

$$E(\mathbf{m}) = ||\mathbf{d} - \mathbf{Gm}||_2 \rightarrow \left\| \begin{bmatrix} t_p^1 \\ t_p^2 \\ \vdots \\ t_p^6 \end{bmatrix} - \overbrace{\mathbf{G}\begin{bmatrix} x_I \\ y_I \\ z_I \\ T_0 \end{bmatrix}}^{\text{Predicted data}} \right\|_2$$

Observed P first arrivals

Hypocenter coordinates
and origin time

Objective function 2:

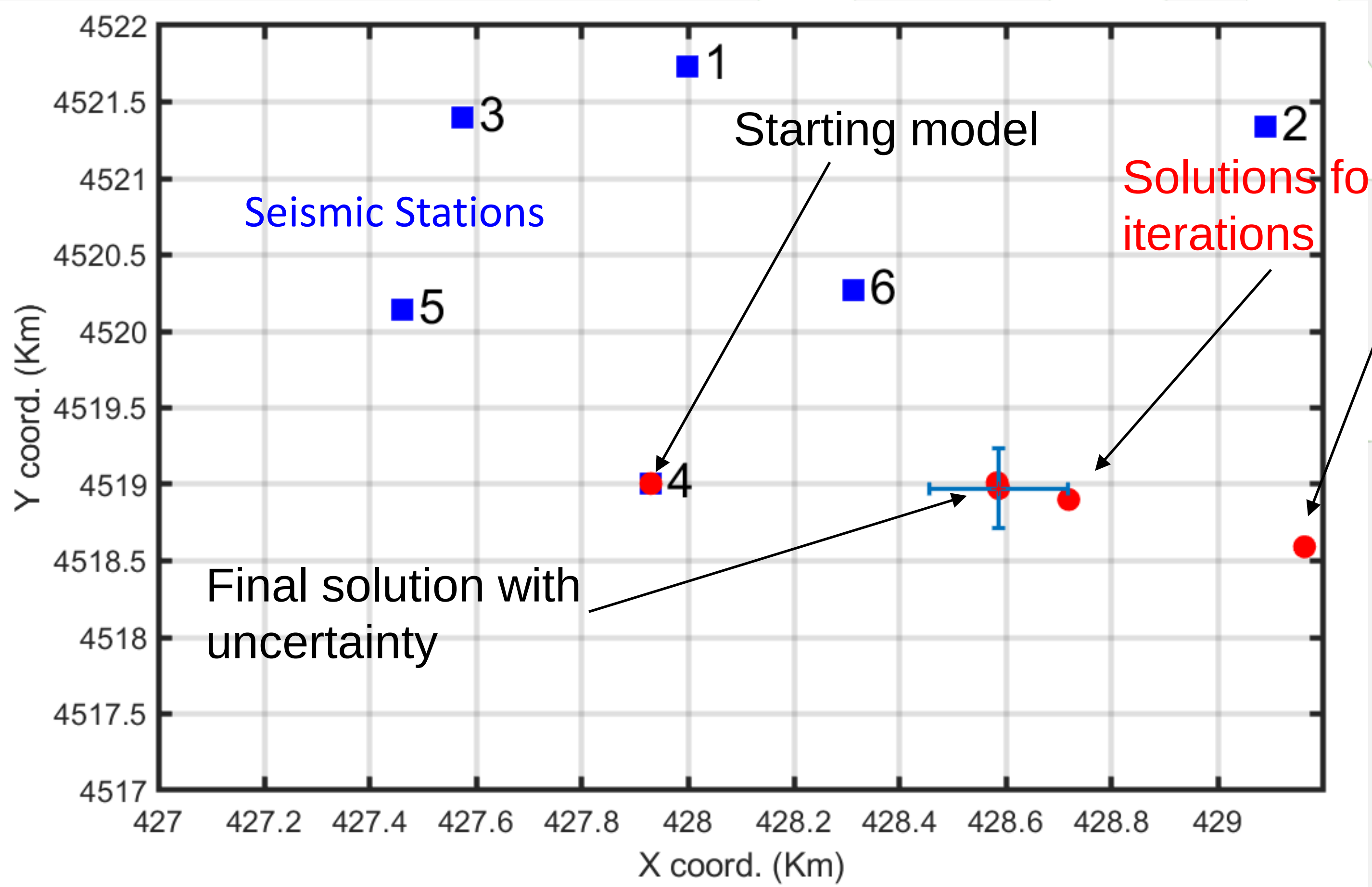
$$E(\mathbf{m}) = ||\mathbf{d} - \mathbf{Gm}||_2 \rightarrow \left\| \begin{bmatrix} t_p^1 \\ t_p^2 \\ \vdots \\ t_p^6 \\ t_s^1 \\ t_s^2 \\ \vdots \\ t_s^6 \end{bmatrix} - \mathbf{G}\begin{bmatrix} x_I \\ y_I \\ z_I \\ T_0 \end{bmatrix} \right\|_2$$

Hypocenter coordinates
and origin time

Observed P and S first arrivals

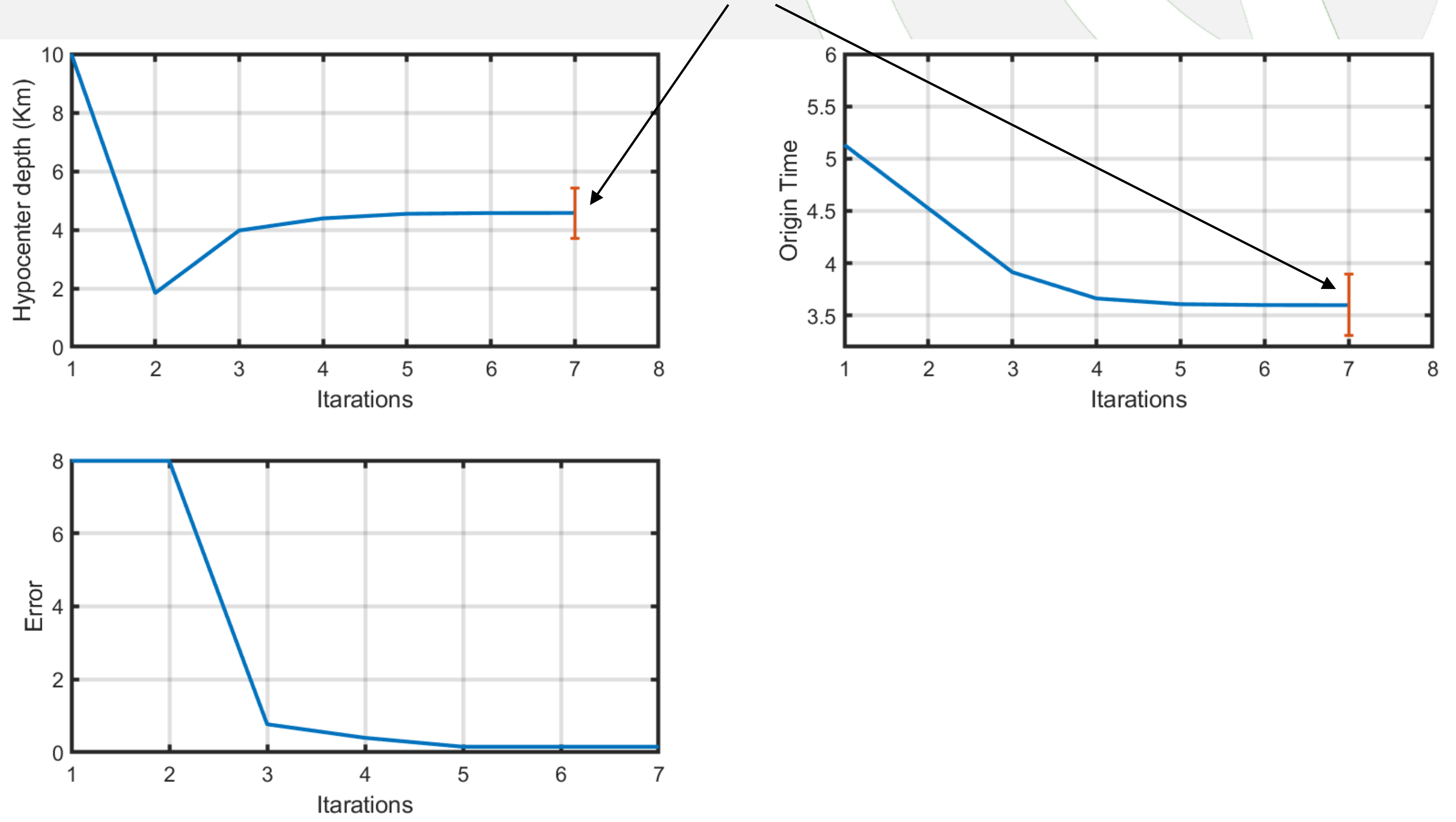
Example: Hypocenter location and T0 estimation

Iterative solution (Gauss-Newton) using P wave arrival times and assuming known **the Vp field**



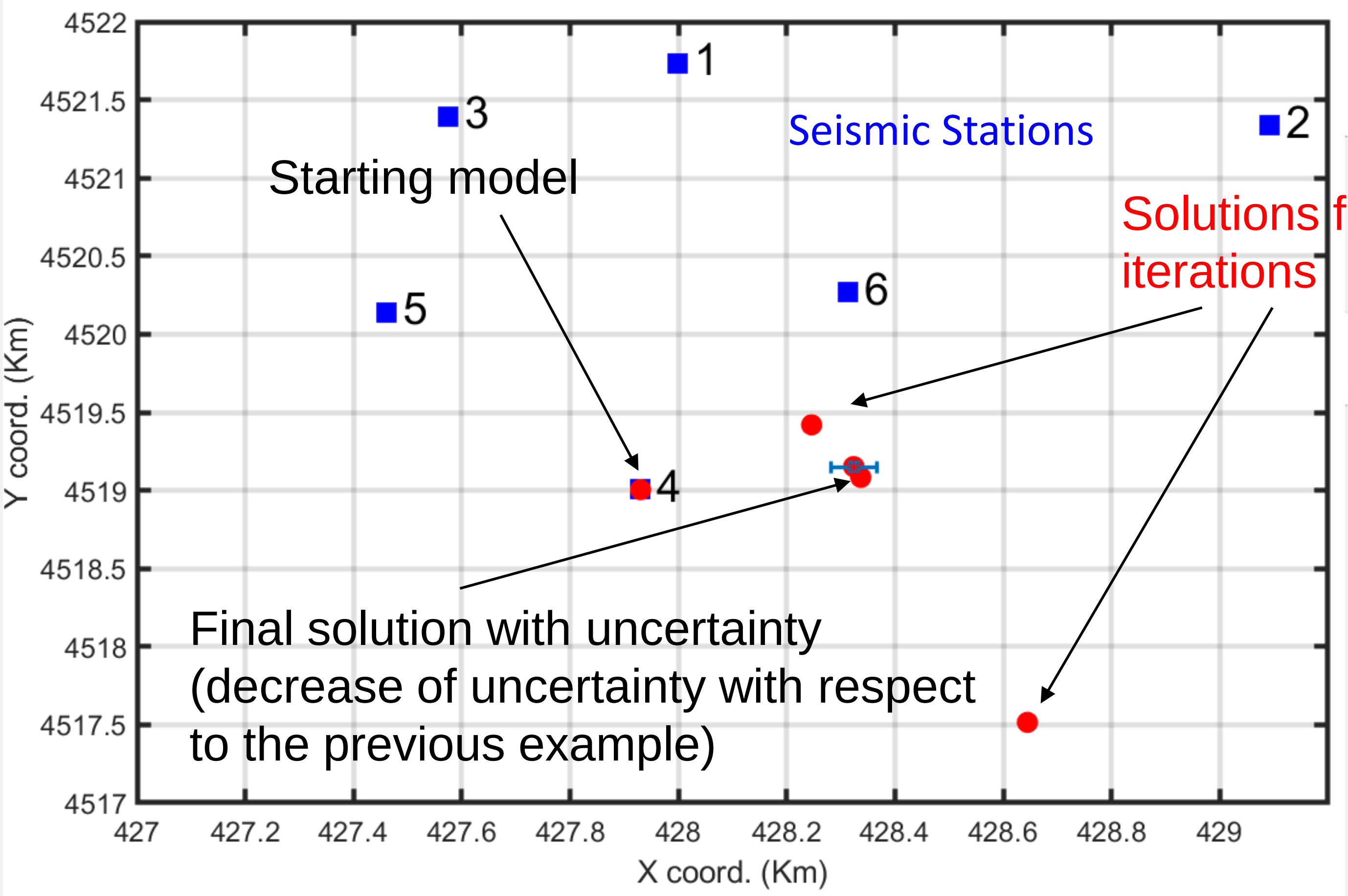
Example: Hypocenter location and T0 estimation

Evolution of the estimated model over iterations. Final results with uncertainty



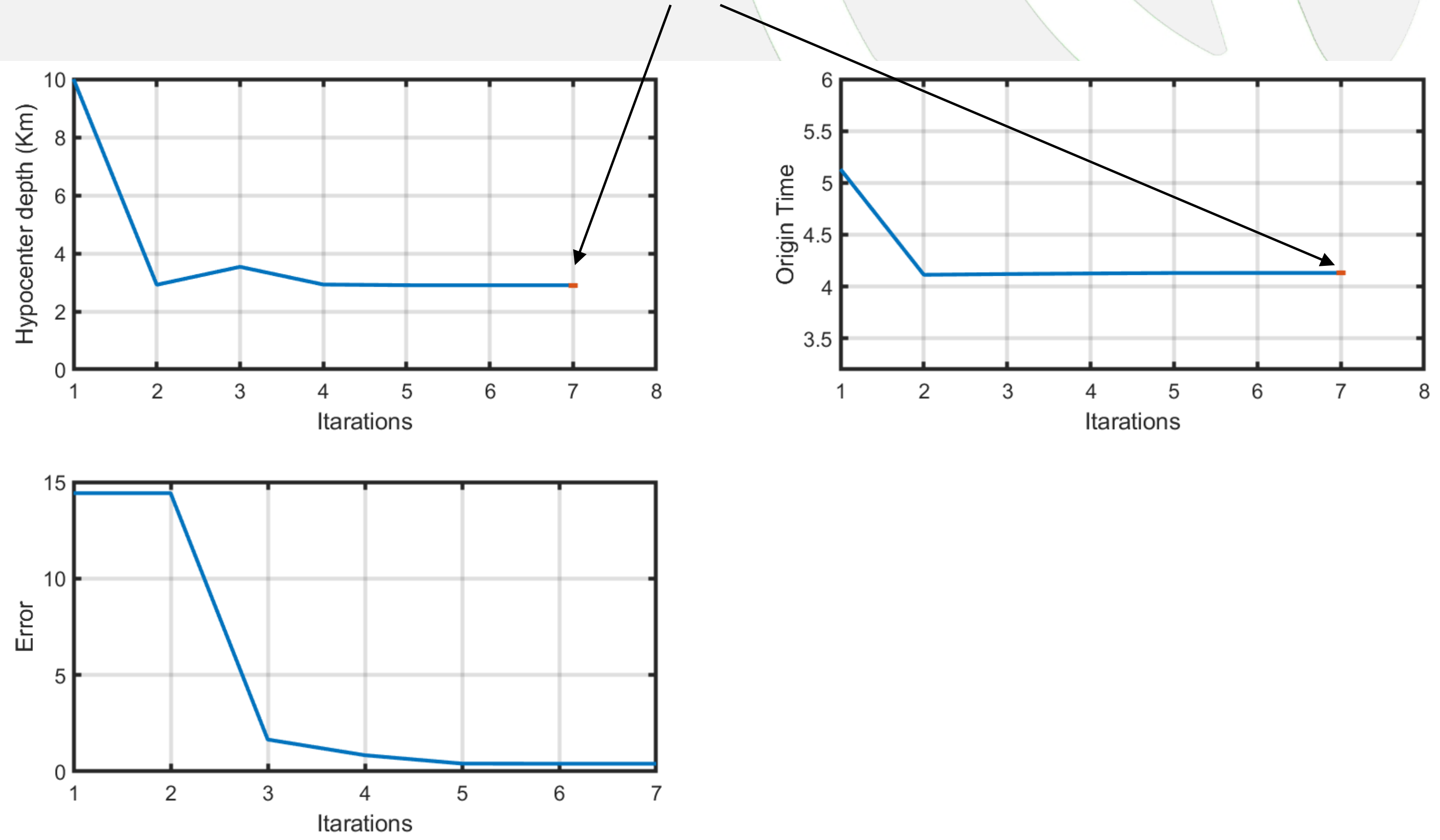
Example: Hypocenter location and T0 estimation

Iterative solution using P and S waves arrival times and assuming known the **Vp and Vs fields**



Example: Hypocenter location and T0 estimation

Evolution of the estimated model over iterations. Final results with uncertainty



ITINERIS

THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System (D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell’Università e della Ricerca

Italiadomani