Inverse Problem

1 Inverse Problem Formulation

An inverse problem is the process of calculating from a set of observations d the causal factors m that produced them.

$$\mathbf{d} = G(\mathbf{m}) \mapsto \mathbf{m} = G^{-1}(\mathbf{d})$$

Where G is known as the forward model.

1.1 Classification

- **Linear vs Non-linear**: if the forward model is linear ($\mathbf{d} = \mathbf{Gm}$) or not.
- **Well-posed**: a problem is well posed when a solution (1) exist, (2) is unique and (3) change continuously with initial condition.
- **Well-conditioned**: When a small error in the initial data does not result in much larger errors in the answers, the problem is said to be well-conditioned. Even if a problem is well-posed, it may still be ill-conditioned. This is evaluated using the condition number (subsection 2.3)
- **Under vs. Overdetermined**: When more equations than unknowns, the system is overdetermined (more than one solution possible) and inversely is under-determined.

2 Tool

2.1 Probability

- The conditional probability measures the probability of an event A given that another event B has occurred:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Bayes' theorem relates an updated probability $P(A|B)$ to its prior probability $P(A)$ in relation to an event B and its relation to it $P(B|A)$

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

- The conditional probability density function of Y given the occurrence of the value x of X can be written as

$$f_Y(y \mid X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

where $f_{X,Y}(x, y)$ gives the joint density of X and Y, while $f_X(x)$ gives the marginal density for X.

2.2 Homogeneous Probability Distribution

The homogeneous probability distribution assigns to each region of the space a probability proportional to the volume of the region. With a volume density $v(\mathbf{x})$, the volume of a region A is define with

$$V(A) = \int_A v(\mathbf{x}) d\mathbf{x}$$

The homogenous probability density is define by the ratio of the volume by the total volume $V = \int_\Omega v(\mathbf{x}) d\mathbf{x}$:

$$\mu(\mathbf{x}) = v(\mathbf{x})/V$$

It is always constant in case of ???

2.3 Condition number

The condition number of a function measures how much the output value of the function can change for a small change in the input argument.

For a linear system, adding a error ϵ in $\mathbf{d}$

$$\mathbf{Gm} = \mathbf{d} + \epsilon \Rightarrow \mathbf{m} = \mathbf{G}^{-1}(\mathbf{d} + \epsilon)$$

result in a condition number:

$$\kappa(G) = \frac{\|\mathbf{G}^{-1}\epsilon\|/\|\mathbf{G}^{-1}\mathbf{d}\|}{\|\epsilon\|/\|\mathbf{d}\|} \leq \|\mathbf{G}^{-1}\| \cdot \|\mathbf{G}\|$$

For non-linear system, the condition number is

$$\frac{\mathbf{m}G'(\mathbf{m})}{G(\mathbf{m})}$$

2.4 Sensitivity coefficient

This measure how much the function change relative to its parameters:

$$\mathbf{J}_G(\mathbf{m}) = \mathbf{J}_{i,j} = \frac{\partial G_i(\mathbf{m})}{\partial m_j}$$

Three standard way to compute

1. Finite-difference: compute for each parameter m_j the difference when changed to $m_j + \Delta m_j$

$$\frac{\partial G_i(\mathbf{m})}{\partial m_j} \approx \frac{G_i(m_1, \cdots, m_j + \Delta m_j, \cdots, m_m) - G_i(\mathbf{m})}{\Delta m_j}$$

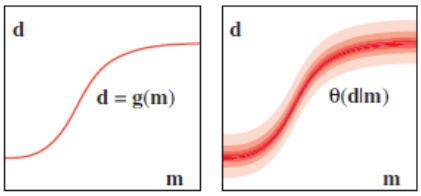
2. Sensitivity equation

3 General Formulation

3.1 Forward Model

The forward model is physical theories which predict the outcome $\mathbf{d}$ for a given parameter set $\mathbf{m}$. When uncertainty ($\mathbf{C_T}$, either from measurement or modelization) cannot be neglected, we can write down the conditional pdf:

$$\theta(\mathbf{d} \mid \mathbf{m}) \sim \mathcal{N}(G(\mathbf{m}), \mathbf{C_T}) \propto \exp\left(-\frac{1}{2}(\mathbf{d} - G(\mathbf{m}))^{\mathrm{T}} \mathbf{C_T}^{-1} (\mathbf{d} - G(\mathbf{m}))\right)$$



The joint pdf (conditional times the marginal) gives the correlations of the physical theory together with its uncertainty. The marginal distribution is the homogenous pdf which characterise...

$$\Theta(\mathbf{d}, \mathbf{m}) = \theta(\mathbf{d} \mid \mathbf{m})\mu(\mathbf{m})$$

3.2 A priori information

- Prior measurement pdf: $\rho_D(\mathbf{d})$.
- Prior parameter pdf: $\rho_M(\mathbf{m})$.

The joint prior information is form by the two independent prior information

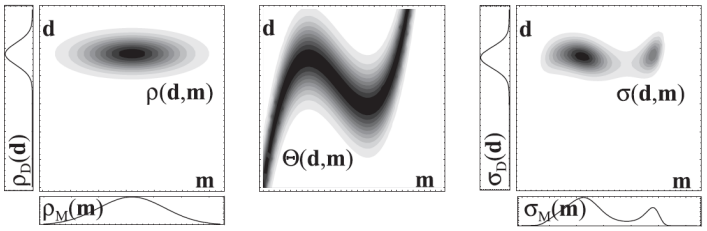
$$\rho(\mathbf{d}, \mathbf{m}) = \rho_D(\mathbf{d})\rho_M(\mathbf{m})$$

3.3 Inverse Problem

The posteriori states of information is given by the conjuction of the theorical pdf Θ and prior pdf ρ. This can be consider as a very general eqution of inversion problem which simplify in most common form.

$$\sigma(\mathbf{d}, \mathbf{m}) = k\frac{\rho(\mathbf{d}, \mathbf{m})\Theta(\mathbf{d}, \mathbf{m})}{\mu(\mathbf{d}, \mathbf{m})}$$

k represent a normalization constant.



The model parameter are updated with:

$$\boxed{\sigma_M(\mathbf{m}) = \int_D \sigma(\mathbf{d}, \mathbf{m}) = k\rho_M(\mathbf{m}) \int_D \frac{\rho_D(\mathbf{d})\theta(\mathbf{d} \mid \mathbf{m})}{\mu_D(\mathbf{d})}}$$

4 Least-Square Problem

The least squares method approximate an overdetermined systems by minimizing the sum of the squares of the errors (L-2 norm, distance...etc). Two following simplification are made (i) the data space is linear and therefore the homogeneous pdf $\mu_D(\mathbf{d})$ is constant and (ii) modelization uncertainty $\mathbf{C_T}$ are neglectable compare to observation uncertainty C_D. This second assumption imply that forward relation pdf is 0 every except in $\mathbf{d} = G(\mathbf{m})$ (the relation has no error):

$$\theta(\mathbf{d} \mid \mathbf{m}) = \delta(\mathbf{d} - G(\mathbf{m}))$$

The model paramter pdf can be simplify:

$$\sigma_M(\mathbf{m}) \propto \rho_M(\mathbf{m}) \int_D \rho_D(\mathbf{d})\delta(\mathbf{d}-G(\mathbf{m})) \propto \rho_M(\mathbf{m})\cdot\rho_D(G(\mathbf{m}))$$

Least-square assume a Gaussian distribution for the priori unknown model parameter $\mathbf{m}$ and the observed measurement $\mathbf{d}$

$$\rho_M(\mathbf{m}) \sim \mathcal{N}(\mathbf{m^{prior}}, \mathbf{C_M})$$
$$\rho_D(\mathbf{d}) = \rho_D(G(\mathbf{m})) \sim \mathcal{N}(\mathbf{d^{obs}}, \mathbf{C_D})$$

The posteriori model parameters pdf become:

$$\sigma_M(\mathbf{m}) \propto \exp\left\{-\frac{1}{2}\left((G(\mathbf{m}) - \mathbf{d^{obs}})^{\mathrm{T}}\mathbf{C_D}^{-1}(G(\mathbf{m}) - \mathbf{d^{obs}}) + (\mathbf{m} - \mathbf{m^{prior}})^{\mathrm{T}}\mathbf{C_M}^{-1}(\mathbf{m} - \mathbf{m^{prior}})\right)\right\}$$

The misfit function is defined

$$\boxed{2S(\mathbf{m}) = \|G(\mathbf{m}) - \mathbf{d^{obs}}\|_D^2 + \|\mathbf{m} - \mathbf{m^{prior}}\|_M^2}$$

"Least square" comes from this this sum of square of the misfit function. Solving least square method will involved minimizing this function. This become an optimization problem

$$\arg\min_{\mathbf{m}} S(\mathbf{m}) := \{\mathbf{m}^* \mid \forall \mathbf{m} : f(\mathbf{m}^*) \leq f(\mathbf{m})\}.$$

This is equivalent to maximized the posteriori model parameter pdf (see maximum posteriori likelihood **??**)

4.1 Linear least squares

For linear problem $\mathbf{d} = \mathbf{Gm}$, the misfit became quadratic (polynomial of degree 2).

$$2S(\mathbf{m}) = \|\mathbf{Gm} - \mathbf{d^{obs}}\|_D^2 + \|\mathbf{m} - \mathbf{m^{prior}}\|_M^2$$

It can be show (see **??**) that this equation is equivalent to:

$$S(\mathbf{m}) = (\mathbf{m} - \tilde{\mathbf{m}})^{\mathrm{T}}\tilde{\mathbf{C}}_{\mathbf{M}}^{-\mathbf{1}}(\mathbf{m} - \tilde{\mathbf{m}})$$

with:

$$\tilde{\mathbf{m}} = (\mathbf{G}^{\mathrm{T}}\mathbf{C_D^{-1}}\mathbf{G} + \mathbf{C_M^{-1}})^{-\mathbf{1}}(\mathbf{G}^{\mathrm{T}}\mathbf{C_D^{-1}}\mathbf{d^{obs}} + \mathbf{C_M^{-1}}\mathbf{m^{prior}})$$

$$\tilde{\mathbf{C}}_{\mathbf{M}} = (\mathbf{G}^{\mathrm{T}}\mathbf{C_D^{-1}}\mathbf{G} + \mathbf{C_M^{-1}})^{-\mathbf{1}}$$

And the misfit function is minimized for $\mathbf{m} = \tilde{\mathbf{m}}$.

4.2 Ordinary least squares

OLS assume zero or negligible errors in the independent variable $\mathbf{m} = \mathbf{m^{prior}}$ and the misfit equation simplify to

$$S(\mathbf{m}) = \|\mathbf{Gm} - \mathbf{d^{obs}}\|_D^2$$

The most common approach is by setting the derivative of the misfit function to zero

$$\frac{\partial S(\mathbf{m})}{\partial \mathbf{m}} = 0$$

$$\begin{aligned}\nabla_{\mathbf{m}}(S) &= 2\nabla_{\mathbf{m}}(\mathbf{d^{obs}} - \mathbf{Gm^{est}})^{\mathrm{T}}\mathbf{C_D}(\mathbf{d^{obs}} - \mathbf{Gm^{est}}) \\ 0 &= 2\mathbf{G}^{\mathrm{T}}\mathbf{C_D}(\mathbf{d^{obs}} - \mathbf{Gm^{est}}) \\ \mathbf{G}^{\mathrm{T}}\mathbf{C_D}\mathbf{d^{obs}} &= \mathbf{G}^{\mathrm{T}}\mathbf{C_D}\mathbf{Gm^{est}}\end{aligned}$$

The predicted value will be:

$$\mathbf{m^{est}} = (\mathbf{G}^{\mathrm{T}}\mathbf{C_D}\mathbf{G})^{-\mathbf{1}}\mathbf{G}^{\mathrm{T}}\mathbf{C_D}\mathbf{d^{obs}}$$

$$\mathbf{d^{pred}} = \mathbf{Gm^{est}} = \mathbf{G}(\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-\mathbf{1}}\mathbf{G}^{\mathrm{T}}\mathbf{d}$$

This can be view as a projection where $\mathbf{G}(\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-\mathbf{1}}\mathbf{G}^{\mathrm{T}}$ is the projection matrix and $\mathbf{G}$ is a matrix whose column vector are a (not necessarily orthonormal) basis.

4.3 Constrainted Least-square

The KKT (Kharush-Kuhn and Tucker) equations is used.The objective function is $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2$, and the constraint function $\mathbf{Cx} = \mathbf{D}$:

$$\Lambda(\mathbf{x}, \lambda) = \|\mathbf{Ax} - \mathbf{b}\|^2 + \sum \lambda_i(\mathbf{c}_i^T\mathbf{x} - d_i)$$

By searching the stationary point of the Lagrangian function $\frac{\partial \Lambda}{\partial x_i} = \frac{\partial \Lambda}{\partial \lambda_i} = 0$ we find the following equations:

$$2\mathbf{A}^T(\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}) + \mathbf{C}^T\lambda = 0 \qquad \mathbf{C}\hat{\mathbf{x}} + \mathbf{d} = 0$$

which can be re-written:

$$\begin{bmatrix} 2\mathbf{A}^T\mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & 0 \end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}} \\ \lambda\end{bmatrix} = \begin{bmatrix}2\mathbf{A}^T\mathbf{b} \\ \mathbf{d}\end{bmatrix}$$

4.4 Non-linear least square

Two linearisation of $\mathbf{g}(\mathbf{m})$ are possible

1. The weakest case is when it can be linearised around m^{prior}

$$\mathbf{g}(\mathbf{m}) \approx \mathbf{g}(\mathbf{m^{prior}}) + \mathbf{G}(\mathbf{m} - \mathbf{m^{prior}})$$

2. The second one is using maximum likelihood with an iterative method (eg. quasi-newton)

5 Maximum Likelihood

The likelihood of a set of parameter values, , given outcomes x, is equal to the probability of those observed outcomes given those parameter values, that is

$$\mathcal{L}(\mathbf{m}|\mathbf{d}) = P(\mathbf{d}|\mathbf{m}).$$

We will show that with a Gaussian prior and a linear forward model, the posteriori pdf is also Gaussian. Using [1] and [2]

$$\begin{aligned}\sigma_M(\mathbf{m}) &\propto \rho_M(\mathbf{m}) \cdot \rho_D(\mathbf{Gm}) \\ &\propto \mathcal{N}(\mathbf{m^{prior}}, \mathbf{C_M}) \cdot \mathcal{N}(\mathbf{G^{-1}d^{obs}}, (\mathbf{G^{-1}})^{\mathrm{T}}\mathbf{C_D}\mathbf{G^{-1}}) \\ &\sim \mathcal{N}(\tilde{\mathbf{m}}, \tilde{\mathbf{C}}_\mathbf{M})\end{aligned}$$

with

$$\tilde{\mathbf{m}} = (\mathbf{G}^{\mathrm{T}}\mathbf{C_D^{-1}}\mathbf{G} + \mathbf{C_M^{-1}})^{-1}(\mathbf{G}^{\mathrm{T}}\mathbf{C_D^{-1}}\mathbf{d^{obs}} + \mathbf{C_M^{-1}}\mathbf{m^{prior}})$$

$$\tilde{\mathbf{C}}_\mathbf{M} = (\mathbf{G}^{\mathrm{T}}\mathbf{C_D^{-1}}\mathbf{G} + \mathbf{C_M^{-1}})^{-1}$$

5.1 Maximum Likelihood Estimate (MLE)

The Maximum Likelihood Estimate (MLE) selects the parameters which maximized (the mode) the likelihood function a given statistical model.

$$\begin{aligned}\hat{\mathbf{m}}_\mathbf{MLE} &= \arg\max_{\mathbf{m}} p(\mathbf{d} \mid \mathbf{m}) \\ &= \arg\min_{\mathbf{m}} (\mathbf{d^{obs}} - G(\mathbf{m}))^{\mathrm{T}}\mathbf{C_D}^{-1}(\mathbf{d^{obs}} - G(\mathbf{m}))\end{aligned}$$

This is nothing more than the weighted least-squared method with the covariance matrix corresponding to the weighted matrix.

5.2 Maximum A Posteriori estimation (MAP)

MAP assume that the parameter are drown from a random process (prior) and therefore it can be seen as a regularization (addition information) of MLE. This avoid over fitting of the data as the assumed (or known) prior knowledge bring back the estimate to its value.

$$\hat{\mathbf{m}}_\mathbf{MAP} = \arg\max_{\mathbf{m}} p(\mathbf{m} \mid \mathbf{d}) = p(\mathbf{d} \mid \mathbf{m})p(\mathbf{d})$$

[1] $\mathbf{x} \sim \mathcal{N}(\mathbf{a}, \mathbf{A}) \Rightarrow \mathbf{Ga} \sim \mathcal{N}(\mathbf{Ga}, \mathbf{G}^{\mathrm{T}}\mathbf{AG})$

[2] $\mathcal{N}(\mathbf{a}, \mathbf{A}) \cdot \mathcal{N}(\mathbf{b}, \mathbf{B}) \sim \mathcal{N}(\mathbf{c}, \mathbf{C})$ with $\mathbf{C} = (\mathbf{A^{-1}} + \mathbf{B^{-1}})^{-1}$ and $\mathbf{c} = \mathbf{C}(\mathbf{A^{-1}a} + \mathbf{B^{-1}b})$

6 More simplier way

6.1 Probabilistic framework

We can rephrase the problem in a simpler probabilistic manner: We observed some data which can be explain by the model and some white error $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{C_D})$

$$\mathbf{d^{obs}} = G(\mathbf{m^{real}}) + \boldsymbol{\epsilon}$$

And therefore,

$$\begin{aligned}p(\mathbf{d^{obs}} \mid \mathbf{m}) &\sim \mathcal{N}(G(\mathbf{m}), \mathbf{C_D}) \\ &\propto \exp\left(-\frac{1}{2}(\mathbf{d^{obs}} - G(\mathbf{m}))^{\mathrm{T}}\mathbf{C_D}^{-1}(\mathbf{d^{obs}} - G(\mathbf{m}))\right)\end{aligned}$$

Using Bayes probability, the probability density of the parameter are updated with the observations data.

$$p(\mathbf{m} \mid \mathbf{d^{obs}}) \propto p(\mathbf{d^{obs}} \mid \mathbf{m})p(\mathbf{m})$$

6.2 Optimization

We can modify further the problem to transform it into a optimization question. The objective function measure the mismatch between observed and simulated value and is minimized by optimization algorithm

$$\arg\min_{\mathbf{m}} OF(\mathbf{d^{obs}} - G(\mathbf{m}))$$

6.3 Regularization

Regularization introduces additional information in order to solve an ill-posed problem or to prevent overfitting. It is similar to adding a prior information in Bayesian point of view.

$$\arg\min_{\mathbf{m}} OF(\mathbf{d^{obs}} - G(\mathbf{m}) + R)$$

One common assumptions used as regulariation include:

- Minimizing euclidean length of the solution $R = \mathbf{m}^{\mathrm{T}}\mathbf{m}$.
- Minimizing distance to a known value (eg. mean): $R = (\mathbf{m} - \overline{\mathbf{m}})^{\mathrm{T}}(\mathbf{m} - \overline{\mathbf{m}})$.
- Minimizing a linear function of the value (such as a flatness measure): $R = (D\mathbf{m})^{\mathrm{T}}(D\mathbf{m}) = \mathbf{m}^{\mathrm{T}}D^{\mathrm{T}}D\mathbf{m}$. Where $D^{\mathrm{T}}D$ might be interpreted as a weighted factor. It could also be the difference matrix (measuring $\mathbf{x_i} - \mathbf{x_{i-1}}$), in which case, we smooth the solution.
- Combining to last approach: $R = (\mathbf{m} - \overline{\mathbf{m}})^{\mathrm{T}}D^{\mathrm{T}}D(\mathbf{m} - \overline{\mathbf{m}})$
-

For under-determined problem, regularisation methods are employed and the problem is solved with Lagrange where the regularisation is viewed as a constraint on the solution.

6.4 Resolution Matrix

6.4.1 Data Resolution Matrix

The Data Resolution Matrix $\mathbf{R_d}$ describe how well the the observed data describe the predicted data.

$$\mathbf{d^{pred}} = \mathbf{Gm^{est}} = \mathbf{GG^{-1}d^{obs}} = \mathbf{R_d d^{obs}}$$

If $\mathbf{R_d} = \mathbf{I}$, the prediction error are zero. The ith row of the matrix describes how each observed data influence the ith predicted data (as a weighted sum).

6.4.2 Model Resolution Matrix

While the data resolution matrix characterizes whether the data can be independently predicted, the Model Resolution Matrix $\mathbf{R_m}$ related the true model parameter to the estimated one. If $\mathbf{R_m} = \mathbf{I}$, the model parameter is uniquely determined. The row of the matrix describe how well neighbouring data can be independently resolved.

$$\mathbf{m^{est}} = \mathbf{G^{-1}d} = \mathbf{G^{-1}Gm^{true}} = \mathbf{R_m m^{true}}$$

6.4.3 Covariance Resolution Matrix

The model covariance matrix is define as:

$$\mathrm{Cov}(\mathbf{m^{est}}) = \mathbf{G^{-1}}\mathrm{Cov}(\mathbf{d})\mathbf{G^{-1T}}$$

7 Solution

7.1 Linear Problem: Invertible

In case of an linear problem with an invertible matrix ($\exists \mathbf{A^{-1}} \mid \mathbf{A^{-1}A} = \mathbf{I}$), the problem become a matrix inversion problem ($\mathbf{m} = \mathbf{G^{-1}d}$) which has a unique solution.

7.1.1 Back (Forward) substitution

For upper (lower) triangular matrice, each coefficient can be computed from previous coefficient up to the last row ($U_{n,n}x_n = b_n$)

$$x_i = \frac{b_n - \sum_{j=1}^{n-1} U_{i,j}x_j}{U_{i,i}}$$

For none-triangular matrix, the QR factorisation can be employ :

7.1.2 Factor and Solve methods

This set of method refere to those which use first a decomposition in easily invertible matrix and then a solve the inverted matrix. The factor step is usually more computer expensive

- LU decomposition $\mathbf{A} = \mathbf{LU}$
  (1) forward substitution $\mathbf{Ly} = \mathbf{b}$, with $\mathbf{y} = \mathbf{Ux}$
  (2) a backward substitution $\mathbf{Ux} = \mathbf{y}$.
- QR decomposition $\mathbf{A} = \mathbf{QR} \Rightarrow \mathbf{Rx} = \mathbf{Q}^{\mathrm{T}}\mathbf{b}$
  (1) compute y $\mathbf{y} = \mathbf{Q}^{\mathrm{T}}\mathbf{b}$
  (2) backward substitution $\mathbf{Rx} = \mathbf{y}$
- Cholesky decomposition $\mathbf{A} = \mathbf{LL}^*$,
  (1) Forward substitution $\mathbf{Ly} = \mathbf{b}$ with $\mathbf{y} = \mathbf{L}^*\mathbf{x}$
  (2) Back substitution $\mathbf{L}^*\mathbf{x} = \mathbf{y}$

7.1.3 Gauss elimination

7.1.4 Cramer's rule

For a nonzero determinant matrice A, the theorem state that the unique solution is

$$x_i = \frac{\det(A_i)}{\det(A)} \qquad i = 1, \ldots, n$$

where A_i is the matrix formed by replacing the i-th column of A by the column vector b.

7.2 Linear Problem : not Invertible

This become a linear regression which can be view as an optimization problem.

7.2.1 Pseudo-Inverse

The pseudo-inverse (or Moore-Penrose Inverse) provides a least squares solution to a system of linear equations. For full column rank (left invertible $A^+A = I$), pseudo-inverse is computed by: $A^+ = (A^*A)^{-1}A^*$ while for full row rank (right inverse): $A^+ = A^*(AA^*)^{-1}$

7.2.2 Newton's method

The model fonction $G(\mathbf{m})$ need to be 2 order differentiable (∇_G and $\mathbf{H}_G$)

7.3 Non-linear Problem

7.3.1 Non-linear least square - Newton's Method

The minimization of the general least-square problem with Newton's Method involved computing the Jacobien and Hessian of $S(\mathbf{m})$ Recall:

$$2S(\mathbf{m}) = \|G(\mathbf{m}) - \mathbf{d^{obs}}\|_D^2 + \|\mathbf{m} - \mathbf{m^{prior}}\|_M^2$$

$$\mathbf{J}_S(\mathbf{m}) = \mathbf{J}_G\mathbf{C}_D^{-1}(G(\mathbf{m}) - \mathbf{d^{obs}}) + \mathbf{C}_M^{-1}(\mathbf{m} - \mathbf{m^{prior}})$$

$$\mathbf{H}_S(\mathbf{m}) = \mathbf{J}_G^{\mathrm{T}}\mathbf{C}_D^{-1}\mathbf{J}_G + \mathbf{H}_G\mathbf{C}_D^{-1}(G(\mathbf{m}) - \mathbf{d^{obs}}) + \mathbf{C}_M^{-1}$$

where $\mathbf{J}_G(\mathbf{m}) = \mathbf{J}_{i,j} = \frac{\partial G(\mathbf{m})_i}{\partial m_j}$ and $\mathbf{H}_G(\mathbf{m}) = \mathbf{H}_{i,j} = \frac{\partial G(\mathbf{m})}{\partial m_i \partial m_j}$ (?)

In the Gaussian-Newton Method, the second term of the Hessian (involving the Hessian of G) can be neglected.

The interative process become:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \frac{\mathbf{J}_S}{\mathbf{H}_S}$$