Correlation effect of transformed or corrected data inversion

László Balázs¹

Received: 10 August 2020 / Accepted: 28 January 2021 / Published online: 12 February 2021
© The Author(s) 2021

Abstract

Before performing the inversion process, the original measured data set is often transformed (corrected, smoothed, Fourier-transformed, interpolated etc.). These preliminary transformations may make the original (statistically independent) noisy measurement data correlated. The noise correlation on transformed data must be taken into account in the parameter fitting procedure (inversion) by proper derivation of likelihood function. The covariance matrix of transformed data system is no longer diagonal, so the likelihood based metrics, which determines the fitting process is also changed as well as the results of inversion. In the practice, these changes are often neglected using the “customary” estimation procedure (simple least square method) resulting wrong uncertainty estimation and sometimes biased results. In this article the consequence of neglected correlation is studied and discussed by decomposing the inversion functional to “customary” and additional part which represents the effect of correlation. The ratio of two components demonstrates the importance and justification of the inversion method modification.

Keywords Inversion · Correlated data · Likelihood · Covariance matrix · Sherman–Morrison theorem

1 Introduction—source of correlation

Transformation or correction of the original data system (y) may have an effect on the error model as the error component is also transformed. As a consequence of preliminary transformation the statistically independent measurement data may become more or less correlated. This correlation may have two basic reasons:

1. The transformation (e.g. correction) function may incorporate random variable(s), therefore all the transformed data will take in these variable(s) which will be the origin of the correlation. (E.g. Bouguer correction of gravity data embodies the Bouguer plate density or static correction of seismic data contains the equivalent upper layer velocity, shale effect correction in well logging interpretation requires the shale parameters etc.)

✉ László Balázs
balazs.laszlo@ttk.elte.hu

¹ Eötvös Loránd University, Budapest, Hungary

- During the transformation more original data point are used for the calculation of one transformed data (convolution, DFT, interpolation etc.), hence the elements of transformed system also has a statistically common parts which also results data correlation.

Eventually the statistically common part causes the correlation between the elements of transformed data ($\mathbf{z}$). The degree of this effect also depends on the applied error model. In this study the most widely-used error model was applied and examined: the zero mean, additive Gaussian noise ($\Delta\mathbf{y} \in N(0, \sigma_y)$). Hence the measured and transformed data vectors:

$$\begin{aligned}\mathbf{y} &= \mathbf{s} + \Delta\mathbf{y} \\ \mathbf{z} &= T(\mathbf{y})\end{aligned}\quad (1)$$

where $\mathbf{s}$ is the noiseless signal vector, and $T(\cdot)$ denotes the transformation. In general, this is a linear transformation in most of customary preliminary data processing steps.

$$\mathbf{z} = T(\mathbf{y}) = T(\mathbf{s}) + T(\Delta\mathbf{y}) \quad (2)$$

The purpose of this paper is the analysis of the correlation effect in the inversion, which often appears on the preprocessed data, changing the covariance structure.

2 Correlated data inversion

The maximum likelihood or the Bayesian approach (if used consistently) automatically answers the data correlation problem. The joint probability density function (f) cannot be written in a form of product, as can be done in the case of independent data system. (Taran-tola 1987; Szatmáry 2002) For the correlated zero mean Gaussian noise the well known likelihood-function is:

$$f(\mathbf{z} | \mathbf{C}_{\Delta\mathbf{z}}, \mathbf{z}_0(\mathbf{p})) = \frac{1}{(2\pi)^{k/2} |\mathbf{C}_{\Delta\mathbf{z}}|^{1/2}} e^{-\frac{1}{2} (\mathbf{z} - \mathbf{z}_0(\mathbf{p}))^T \mathbf{C}_{\Delta\mathbf{z}}^{-1} (\mathbf{z} - \mathbf{z}_0(\mathbf{p}))} \quad (3)$$

where the $\mathbf{C}_{\Delta\mathbf{z}}$ is the covariance matrix of correlated Gaussian random vector variable $\mathbf{z}$. The $(\mathbf{z} - \mathbf{z}_0)$ is the error term ($\Delta\mathbf{z}$) for vector $\mathbf{z}$ and $\mathbf{z}_0$ is the centre of distribution; k is the dimension of measurement space. Here, T denotes the transpose of vectors and $|\mathbf{C}_{\Delta\mathbf{z}}|$ is the determinant of covariance matrix. The function in Eq. (3) is a conditional probability density function according to the parameters ($\mathbf{p}$) and the error covariance matrix.

The quadratic form with inverse of the covariance matrix in the exponent will be the basis of metrics for the parameter fitting (the well-known Mahalanobis distance) (Mahalanobis 1936). The nonzero non diagonal elements in $\mathbf{C}_{\Delta\mathbf{z}}$ (normalized by standard deviations) indicate the measure of random dependence between the preprocessed random variables (elements of $\mathbf{z}$). This matrix can be transformed into correlation matrix to study the correlation.

3 Case of correction

In the data evaluation process, the model of whole direct problem is often divided into two parts and the effect of geologically uninterested part is separated as a correction (Bouguer correction, seismic static correction etc.) in order to enhance the essential

information. Sometimes, at the correction some random parameter is involved (like Bouguer plate density, upper zone velocity etc.) and after the correction, these random variables are built in the new corrected data set as a statistically common part (causing correlation on error term of corrected data).

3.1 Effect of subtractive correction

With a simple subtractive correction on a 1D noisy data system ($\mathbf{y}$) will be demonstrated the effect of correlation resulting from a correction. The corrected data points are:

$$z_i = y_i - c\mu h_i \quad (4)$$

The c constant represents all non random multipliers, μ is the equivalent petrophysical properties for the correction (like the Bouguer plate density, equivalent seismic velocity above the datum level etc.), $\mathbf{h}$ is the geometric parameter (e.g. thickness of the layer) appears in the correction.

The error part of corrected data set ($\Delta\mathbf{z}$) is composite, containing the independent measurement error vector ($\Delta\mathbf{y}$), the error of correction parameter ($\Delta\mu$) and the error of layer thicknesses ($\Delta\mathbf{h}$). The error model is additive for all variables with the associated centers of distributions ($s_i, h_{0,i}, \mu_0$):

$$\begin{aligned} y_i &= s_i + \Delta y_i \\ h_i &= h_{0,i} + \Delta h_i \\ \mu &= \mu_0 + \Delta\mu \end{aligned} \quad (4a)$$

Then the detailed error term ($\Delta\mathbf{z}$) of the corrected data vector ($\mathbf{z}$) element is the following:

$$z_i = s_i + \Delta y_i - c(\mu_0 + \Delta\mu)(h_{0,i} + \Delta h_i) = s_i + c\mu_0 h_{0,i} + \Delta y_i - c\Delta\mu h_{0,i} - c\mu_0 \Delta h_i - c\Delta\mu \Delta h_i \quad (5)$$

The last three error terms appear as a consequence of correction. To frame the covariance matrix with expectation value of $\Delta\mathbf{z}$ vector dyadics, the different error terms in Eq. (5) are considered independent and zero mean, so:

$$E(\Delta\mathbf{z}\Delta\mathbf{z}^T) = E(\Delta\mathbf{y}\Delta\mathbf{y}^T) + c^2\Delta\mu^2\mathbf{h}_0\mathbf{h}_0^T + c^2\Delta\mu^2\Delta\mathbf{h}\Delta\mathbf{h}^T + c^2\mu_0^2\Delta\mathbf{h}\Delta\mathbf{h}^T \quad (6)$$

After taking the expectation value we get the covariance matrix of corrected data:

$$\mathbf{C}_{\Delta\mathbf{z}} = \mathbf{C}_{\Delta\mathbf{y}} + c^2\sigma_\mu^2\mathbf{h}_0\mathbf{h}_0^T + c^2(\mu_0^2 + \sigma_\mu^2)\mathbf{C}_{\Delta\mathbf{h}} \quad (7)$$

Therefore the covariance matrix of corrected data is not diagonal because the second term of Eq. (7), which has dyadic form. The diagonal elements are also modified because the 3rd term of Eq. (7). To derive the functional to be minimized in the inversion, the inverse of $\mathbf{C}_{\Delta\mathbf{z}}$ is needed for the calculation of Mahalanobis distance.

This inverse—which is the metric tensor in the inversion functional—can be calculated by using Sherman–Morrison theorem. (Sherman and Morrison 1949) This theorem is for the derivation of inverse of special matrices which can be decomposed to the diagonal and dyadic part. By this theorem the metric tensor for Mahalanobis distance will be:

$$\mathbf{C}_{\Delta z}^{-1} = \left(\frac{1}{\sigma_y^2 + c^2(\mu_0^2 + \sigma_\mu^2)\sigma_h^2} \right) \mathbf{I} - \frac{1}{1 + \left(\frac{c^2\sigma_\mu^2}{\sigma_y^2 + c^2(\mu_0^2 + \sigma_\mu^2)\sigma_h^2} \right) \mathbf{h}_0^T \mathbf{h}_0} \left(\frac{c^2\sigma_\mu^2}{(\sigma_y^2 + c^2(\mu_0^2 + \sigma_\mu^2)\sigma_h^2)^2} \right) \mathbf{h}_0 \mathbf{h}_0^T \quad (8)$$

where $\mathbf{I}$ is the identity matrix. The inverse of original $\mathbf{y}$ vector covariance was:

$$\mathbf{C}_{\Delta y}^{-1} = \frac{1}{\sigma_y^2} \mathbf{I} \quad (9)$$

The importance of correlation correction [second part in Eq. (8)] is characterized by the following weight which depends on the ratio of error term variances:

$$\frac{c^2\sigma_\mu^2}{\sigma_y^2 + c^2(\mu_0^2 + \sigma_\mu^2)\sigma_h^2} \quad (10)$$

The decomposition derived above [Eq. (8)], shows how the diagonal variances is blurred, and how coupled the different measuring points in the exponent of formula 3.

3.2 Effect of correction, which is nonlinear function of random parameter

If the form of correction is not linear, the modified covariance matrix ($\mathbf{C}_{\Delta z}$) can be approximated by linearization. The corrected data vector elements in this case:

$$z_i = y_i + g(h_i, \mu) = y_i + g_i \quad (11)$$

where $g(h_i, \mu)$ is the correction function. (E.g. the static correction in seismic preprocessing has this form, because it depends on the reciprocal of the upper zone velocity) Expressing Eq. (11) with the associated error parts:

$$\begin{aligned} z_i &= s_i + \Delta y_i + g(h_{0,i} + \Delta h_i, \mu_0 + \Delta \mu) \approx s_i + g(h_{0,i}, \mu_0) + \Delta y_i + \frac{\partial g(h_{0,i}, \mu_0)}{\partial h} \Delta h_i + \frac{\partial g(h_{0,i}, \mu_0)}{\partial \mu} \Delta \mu \\ \Delta z_i &= \Delta y_i + \frac{\partial g(h_{0,i}, \mu_0)}{\partial h} \Delta h_i + \frac{\partial g(h_{0,i}, \mu_0)}{\partial \mu} \Delta \mu \end{aligned} \quad (12)$$

After taking the expectation value of the $\Delta \mathbf{z} \Delta \mathbf{z}^T$ dyadics in the first order approximation [Eq. (12)], the $\mathbf{C}_{\Delta z}$ will be:

$$\mathbf{C}_{\Delta z} \approx \mathbf{C}_{\Delta y} + \sigma_h^2 \mathbf{C}_1 + \sigma_\mu^2 \mathbf{u} \mathbf{u}^T \quad (13)$$

where the elements of $\mathbf{C}_1$ diagonal matrix and $\mathbf{u}$ vector are the following:

$$\begin{aligned} C_{1,i,j} &= \delta_{i,j} \left(\frac{\partial g(h_{0,i}, \mu_0)}{\partial h} \right)^2 \\ u_i &= \frac{\partial g(h_{0,i}, \mu_0)}{\partial \mu} \end{aligned} \quad (14)$$

where δ is the Kronecker symbol. Using the Sherman–Morrison equation again, the inverse of $\mathbf{C}_{\Delta z}$:

$$\mathbf{C}_{\Delta z}^{-1} = (\mathbf{C}_{\Delta y} + \sigma_h^2 \mathbf{C}_1)^{-1} - \frac{(\mathbf{C}_{\Delta y} + \sigma_h^2 \mathbf{C}_1)^{-1} \mathbf{u} \mathbf{u}^T (\mathbf{C}_{\Delta y} + \sigma_h^2 \mathbf{C}_1)^{-1} \sigma_\mu^2}{1 + \mathbf{u}^T (\mathbf{C}_{\Delta y} + \sigma_h^2 \mathbf{C}_1)^{-1} \mathbf{u} \sigma_\mu^2} \quad (15)$$

The second part of Eq. (15) expresses the random dependence of corrected variables which may modify the results of inversion.

3.3 Effect of product type correction

Sometimes the correction has a product form, such as the seismic amplitude correction or the multiplication by sonde coefficient to transform the raw measured data into apparent values etc.

In the case of product type correction with random parameter, the form of corrected data vector element is the following:

$$\begin{aligned} z_i &= g(\mu_0 + \Delta\mu)(s_i + \Delta y_i) \approx \left(g(\mu_0) + \frac{\partial g(\mu_0)}{\partial \mu} \Delta\mu \right) (s_i + \Delta y_i) \\ &= g(\mu_0) s_i + g(\mu_0) \Delta y_i + \frac{\partial g(\mu_0)}{\partial \mu} s_i \Delta\mu + \frac{\partial g(\mu_0)}{\partial \mu} \Delta y_i \Delta\mu \end{aligned} \quad (16)$$

where $(g(\mu))$ is now the correction function (with random parameter). The covariance matrix of the error part:

$$\mathbf{C}_{\Delta z} = (g(\mu_0))^2 \mathbf{C}_{\Delta y} + \sigma_\mu^2 \left(\frac{\partial g(\mu_0)}{\partial \mu} \right)^2 \mathbf{s} \mathbf{s}^T + \left(\frac{\partial g(\mu_0)}{\partial \mu} \right)^2 \sigma_\mu^2 \mathbf{C}_{\Delta y} \quad (17)$$

Using the Sherman-Morrison formula to determine the inverse matrix:

$$\mathbf{C}_{\Delta z}^{-1} = \frac{1}{(g(\mu_0))^2 + \left(\frac{\partial g(\mu_0)}{\partial \mu} \right)^2 \sigma_\mu^2} \mathbf{C}_{\Delta y}^{-1} - \frac{\frac{\sigma_\mu^2 \left(\frac{\partial g(\mu_0)}{\partial \mu} \right)^2}{\left((g(\mu_0))^2 + \left(\frac{\partial g(\mu_0)}{\partial \mu} \right)^2 \sigma_\mu^2 \right)^2} \mathbf{s} \mathbf{s}^T}{1 + \frac{\sigma_\mu^2 \left(\frac{\partial g(\mu_0)}{\partial \mu} \right)^2}{\left((g(\mu_0))^2 + \left(\frac{\partial g(\mu_0)}{\partial \mu} \right)^2 \sigma_\mu^2 \right)} \mathbf{s}^T \mathbf{s}} \quad (18)$$

This inverse matrix is also separated into diagonal and dyadic part like before. During the estimation procedure $\mathbf{s}$ vector can be approximated by $\mathbf{y}$ vector.

4 Case of linear transformation

Typical preliminary phase operation in the data processing is the convolutional filtering or filtering in spectral range after Discrete Fourier Transformation (DFT). In general, the linear transformation can be represented with a transformation matrix ($\mathbf{F}$). The transformed data vector:

$$\mathbf{z} = \mathbf{F}\mathbf{y} = \mathbf{F}\mathbf{s} + \mathbf{F}\Delta\mathbf{y} = \mathbf{z}_0 + \Delta\mathbf{z} \quad (19)$$

The covariance matrix of data vector $\mathbf{z}$ which was obtained by linear transformation of originally independent (and uniform variance) data ($\mathbf{y}$):

$$\mathbf{C}_{\Delta\mathbf{z}} = \mathbf{E}(\Delta\mathbf{z}\Delta\mathbf{z}^T) = \mathbf{E}(\mathbf{F}\Delta\mathbf{y}\Delta\mathbf{y}^T\mathbf{F}^T) = \mathbf{F}\mathbf{C}_{\Delta\mathbf{y}}\mathbf{F}^T = \sigma_{\Delta\mathbf{y}}^2\mathbf{F}\mathbf{F}^T \quad (20)$$

The structure of matrix $\mathbf{F}\mathbf{F}^T$ determines the covariance structure which reflects the possible statistical dependence of vector $\mathbf{z}$ elements. Substituting the $\mathbf{C}_{\Delta\mathbf{z}}$ [Eq. (20)] into the Mahalanobis distance formula:

$$(\mathbf{z} - \mathbf{z}_0(\mathbf{p}))^T \mathbf{C}_{\Delta\mathbf{z}}^{-1} (\mathbf{z} - \mathbf{z}_0(\mathbf{p})) = \Delta\mathbf{z}^T \mathbf{C}_{\Delta\mathbf{z}}^{-1} \Delta\mathbf{z} = \Delta\mathbf{y}^T \mathbf{F}^T (\mathbf{F}\mathbf{C}_{\Delta\mathbf{y}}\mathbf{F}^T)^{-1} \mathbf{F}\Delta\mathbf{y} = \Delta\mathbf{y}^T \mathbf{C}_{\Delta\mathbf{y}}^{-1} \Delta\mathbf{y} \quad (21)$$

It can be seen [Eq. (21)], that using the proper Mahalanobis metrics (taking into account the possible data correlation), then the criteria of parameter fitting doesn't changed. So the parameter variances (estimated in the inversion) also remain unchanged. It shows, that the applied linear transformation doesn't increase the information content of dataset and may not reduce the Cramér–Rao bound for the estimated parameter variance (Cramér 1946; Rao 1945).

If the “traditional” simple least square estimation had been used—neglecting the correlation effect—we would get a lower parameter variance violating the Cramér–Rao bound. (While, the proper maximum likelihood estimation is known to be asymptotically efficient, that is the variance of estimated parameter reaches the Cramér–Rao bound asymptotically).

4.1 Covariance of filtered data

During the preliminary data processing the convolutional filtering is widely applied for noise suppression or extraction of low frequency trend or high frequency local effect etc. This procedure may also be the source of noise correlation since the filtered data elements are the composition of more raw measurement data, hence their error are inherited. In the case of additive noise, the filtered data ($\mathbf{z}$) and their decomposition can be expressed:

$$z_k = \sum_{i=-N}^N w_i y_{i-k} = \sum_{i=-N}^N w_i s_{i-k} + \sum_{i=-N}^N w_i \Delta y_{i-k} \quad (22)$$

were $\mathbf{w}$ is the vector of convolution weight in the $2N + 1$ points length filtering window. In the simpler matrix form (matrix $\mathbf{W}$ was made up of filter weights $\mathbf{w}$):

$$\mathbf{z} = \mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{s} + \mathbf{W}\Delta\mathbf{y} \quad (23)$$

The noise covariance matrix [in a same form as in Eq. (20)]:

$$\mathbf{C}_{\Delta\mathbf{z}} = \mathbf{E}[\mathbf{W}\Delta\mathbf{y}\Delta\mathbf{y}^T\mathbf{W}^T] = \mathbf{W}\mathbf{C}_{\Delta\mathbf{y}}\mathbf{W}^T \quad (24)$$

In this case the $\mathbf{C}_{\Delta\mathbf{z}}$ is a band matrix, which may have nonzero non-diagonal elements. The diagonal elements of $\mathbf{C}_{\Delta\mathbf{y}}$ are partially distributed among the non-diagonal elements according to the $\mathbf{w}$ vector. Therefore the trace of covariance matrix which represents the overall variances of data vector decreases as a consequence of filtering (compared to the raw data system variance).

Take the case of 3-step moving average (MA) filter as an example, the transformation matrix ($\mathbf{W}$) element is the following:

$$W_{ik} = \frac{1}{3} (\delta_{i,k-1} + \delta_{i,k} + \delta_{i,k+1}) \quad (25)$$

Using Eq. (24), the covariance matrix element of MA filtered data:

$$C_{\Delta z, ik} = \frac{\sigma_{\Delta y}^2}{9} (\delta_{i,k-2} + 2\delta_{i,k-1} + 3\delta_{i,k} + 2\delta_{i,k+1} + \delta_{i,k+2}) \quad (26)$$

It can be seen, that the trace of filtered data covariance matrix is one-third of original covariance matrix trace. Although the overall data variance decreases after filtering, the functional for the parameter estimation (the Mahalanobis distance) is not changed, as it was shown in Eq. (21), if the possible data correlation is also considered.

4.2 Covariance matrix of Fourier transformed data

The Discrete Fourier Transformation (DFT) is a frequently used tool in the preprocessing. It also can be described in matrix form, so the transformed data vector:

$$\mathbf{z} = \mathbf{F}_{DFT} \mathbf{y} = \mathbf{F}_{DFT} \mathbf{s} + \mathbf{F}_{DFT} \Delta \mathbf{y} \quad (27)$$

The element of $\mathbf{F}_{DFT}$ is:

$$F_{DFT, mn} = \exp \left(\frac{-i2\pi mn}{N} \right) \quad (28)$$

where N is the number of data. Since the results of transformation is complex, therefore the derivation of covariance matrix should be modified accordingly, using the conjugate of DFT matrix ($\overline{\mathbf{F}_{DFT}}$):

$$\mathbf{C}_{\Delta z} = \mathbf{E} \left(\mathbf{F}_{DFT} \Delta \mathbf{y} \Delta \mathbf{y}^T (\overline{\mathbf{F}_{DFT}})^T \right) \quad (29)$$

The DFT matrix has unitary symmetry, that is, the transposed its conjugate is also its inverse. Hence the transformed data covariance matrix remains diagonal, so the conventional least square method is still applicable in spectral range.

5 Summary

The covariance structure of initial data system is essential for the proper implementation of inversion. Although, previously various preprocessing steps are often performed, which may change the statistical properties of data; this is sometimes ignored during the inversion. In this paper the noise covariance of transformed data was studied.

In the first part the different type of correction is examined (subtractive, multiplicative, nonlinear), where the correction function contains random variables. It was shown, that in this case, the basic functional of inversion must be modified, because of the changes in covariance structure. For the mentioned cases, the covariance matrix of corrected data was generated and its inverse too which is necessary for the likelihood function.

In the second part, the covariance structure of linearly transformed data was examined, studying the consequence of transformation.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Cramér H (1946) Mathematical methods of statistics. Princeton University Press, Princeton, pp 473–497. ISBN 0-691-08004-6
- Mahalanobis PC (1936) On the generalized distance in statistics. *Proc Natl Inst Sci India* 2(1):49–55
- Rao CR (1945) Information and the accuracy attainable in the estimation of statistical parameters. *Bull Calcutta Math Soc* 37:81–89
- Sherman J, Morrison WJ (1949) Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. *Ann Math Stat* 20:621. <https://doi.org/10.1214/aoms/1177729959>
- Szatmáry Z (2002) The measurement evaluation, (in Hungarian), lecture notes, pp 102–112. <https://mek.oszk.hu/00800/00862>
- Tarantola A (1987) Inverse problem theory. Elsevier, Amsterdam, pp 22–25