Inside matrix inversion for TPA: how to prove if regularization is really needed!

Serafima Anisovich[1], Roland Sottek[2]

[1] HEAD acoustics GmbH, 52134 Herzogenrath, E-Mail: Serafima.Anisovich@head-acoustics.de

[2] HEAD acoustics GmbH, 52134 Herzogenrath, E-Mail: Roland.Sottek@head-acoustics.de

Introduction

The **M**atrix **I**nversion **M**ethod (MIM) as a part of transfer path analysis (TPA) is a method for estimating source characteristics, both for airborne and structure-borne cases. A big advantage of the method is that the resulting characteristic of the source is independent of its surrounding. A source acting in a complex way is represented by several simpler sources (components) and their characteristics (force for the structure-borne or volume velocity for the airborne case). This model of the source can be used to virtually put the source into any other environment and to estimate its effect there. This advantage is worthwhile even if difficulties such as the problem of a correct model estimation have to be dealt with and that accurate measurements require a lot of efforts [1]. After completing the measurements, an engineer faces yet another problem, namely ill-conditioning of the matrix, that makes the problem very sensitive to any measurement error. The most popular way of treating an ill-conditioned matrix is its regularization. However, it must be kept in mind that regularization can lead to a loss of physically important information and must be used with care. This paper provides structured information on the problem of ill-conditioning in the matrix inversion method and proposes a way of deciding if the regularization is really needed.

Motivation

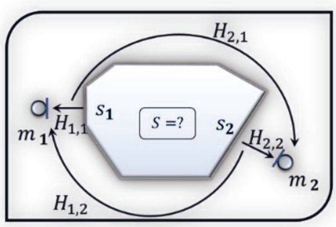

Figure 1: Simple example of measurements for the matrix inversion method. Source s consists of two components s_1 and s_2, $H_{k,l}$ – transfer function between k-th receiver and l-th component, m_1, m_2 – measured signals at receiver positions

As mentioned before, the MIM is aimed to source characterization. The method is illustrated through a simple example. A sound source consists of two main components (Figure 1). To find characteristics of these components (s_1, s_2, s – for "source"), two receiver positions are defined, a measurement under operational condition (m_1, m_2, m – for "measurement") and a measurement of all transfer functions $H_{k,l}$ is performed. The measured signals can be put into a system of equations (1)

$$\begin{cases} m_1 = H_{1,1}S_1 + H_{1,2}S_2 \\ m_2 = H_{2,1}S_1 + H_{2,2}S_2 \end{cases} \quad (1)$$

For a general case with a larger number of components and measurement points, this system can be written in a matrix notation as in equation (2).

$$\begin{pmatrix} m_1 \\ \vdots \\ m_L \end{pmatrix} = \begin{pmatrix} H_{1,1} & \cdots & H_{1,K} \\ \vdots & \ddots & \vdots \\ H_{L,1} & \cdots & H_{L,K} \end{pmatrix} \begin{pmatrix} S_1 \\ \vdots \\ S_K \end{pmatrix} \text{ or } m = Hs \quad (2)$$

The solution for this system provides the desired source characteristic (3), where † indicates pseudo-inversion.

$$s = H^{-1}m \text{ – for } K = L \text{ and small condition number}$$
$$s = H^{\dagger}m \text{ – for } K = L \text{ and large condition number, or for } K \neq L \quad (3)$$

The solution can be very sensitive to even small perturbations in the measurements that make the system challenging to solve. The problem is named ill-posed (or ill-conditioned) in this case. A standard recommendation for processing of such a problem would be to calculate a matrix condition number and, if it is large, to use a regularization technique. However, there is no rule that could provide the information, which condition number is large, and which is acceptably small. Moreover, regularization may lead to a loss of physically important information.

The goal was to clarify what regularization is really doing to the matrix, in order to decide if it is needed.

Ill-posed problem

An ill-conditioned or ill-posed problem is one that does not fulfil properties of a well-posed problem given by Jacques Hadamard [2], namely

- a solution exists,
- the solution is unique,
- the solution's behaviour changes continuously with the initial conditions.

It is quite difficult in real life to meet these points.

First, the assumption is made that a considered process can be described by a continuum model (refer to the flowchart in Figure 2), where $\mathcal{M}$ is a continuum measurement, $\mathcal{H}$ is a continuum system of transfer functions, S is a source and $\mathcal{E}$ is a possible distortion. At this point, it is assumed that the estimation of the model is correct, so it can be expected that a solution of this problem exists.

A computational model to describe this process is discrete and is supposed to operate with the exact measurements $\bar{m}$ that can help to find the exact solution $\bar{s}$. However, a real measurement can contain an error e and the system is transformed to $m = Hs$, where m and s include some error.

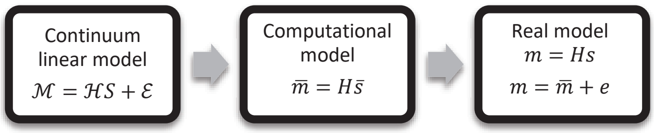

Figure 2: The transition from a continuum linear model to a real discrete model

Under these circumstances, it can be expected that the solution cannot be unique. To cope with that, the problem is solved in terms of a least square approach.

Least-square solution

A least square solution $\min\|Hs - m\|_2$ can be found as in equations (3). According to the Gauss-Markov theorem the least square solution s_{LSQ} is the best **unbiased** estimator of $\bar{s}$, in the sense that it has **minimum variance**. A visual representation of a direct problem as a starting point should show what this means using a transfer function matrix

$$H = \begin{pmatrix} H_{1,1} & H_{1,2} \\ H_{2,1} & H_{2,2} \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 5 & 7 \end{pmatrix} \tag{4}$$

with the exact solution $\bar{s} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then the exact measurement would be

$$\bar{m} = H\bar{s} = \begin{pmatrix} 5 \\ 12 \end{pmatrix} \tag{5}$$

In a real situation, of course, the exact solution is not known. Solving this problem means to find coordinates of m in a vector space defined by H (see Figure 3, left, equations (4) and (5)).

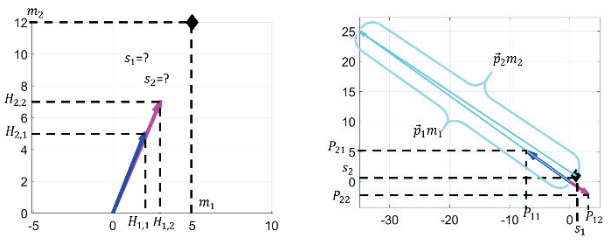

Figure 3: Visual representation of a sample problem, Left – direct problem, Right – indirect problem

To find the solution, the problem is formulated indirectly, which is described by the following equation

$$s = H^{\dagger}m = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} m = (\vec{p}_1 m_1, \vec{p}_2 m_2) \tag{6}$$

The elements of the matrix are named here P for "pseudo-inversion", and they do not have a physical meaning. A visual representation of the indirect problem is shown in Figure 3 (right). The vectors $\vec{p}_1 m_1$ and $\vec{p}_2 m_2$ point to almost opposite directions.

Having exact measurement, the exact solution can be found. It is worth mentioning that the exact measurement leads to the exact solution, and it does not matter if the problem is ill- or well-conditioned. However, if the measurement does contain some error, the least square solution of an ill-posed problem can differ dramatically. The exact solution is marked in Figure 4 by a diamond. The stars correspond to the solution of the system with an introduced error regarding m within ±5%. As can be seen, the resulting error is much larger than the measurement error. It breaks the third property of the well-posed problem. The unbiased estimator of the exact solution given by solving a least square problem has a minimum variance, even though the variance can still be large. To overcome this problem the regularization was introduced.

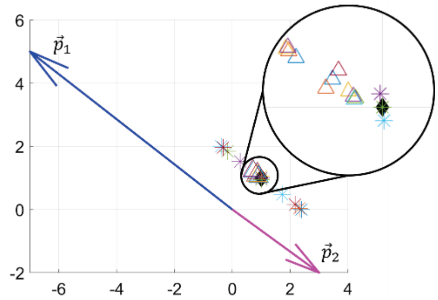

Figure 4: Visual representation of an indirect sample problem. Vectors – columns of the inverted matrix: $\vec{p}_1$ and $\vec{p}_2$. Diamond - exact solution, stars – least square solution, assuming a measurement error of ±5%, triangles – regularized solution, assuming the same measurement error (for better visibility see the zoomed area)

Regularization and matrix condition number

Regularization aims to stabilize the problem regarding the error, but the penalty, that must be paid is the agreement on a **biased** solution. It means that a regularized solution will never be exact, even if the measurements are correct. Regularized solutions to the problem discussed above are shown in Figure 4 by triangles. The variance of the solutions is not so large anymore, but none of the solutions match the exact one, even for the case of no error.

It is important to bear in mind that in more complex cases the regularized solution can be misleading, even being stable with regard to the measurement error.

Therefore, some instrument is needed that can provide information on how sensitive the solution is to an error. A well-known indicator is a condition number of a matrix.

The matrix condition number is a ratio of the largest and the smallest non-zero singular value of the matrix. Theorems of singular value decomposition (SVD), principal component decomposition (PCD) and their relationship can help to understand why the condition number is important, but at the same time why it is not so useful.

The theorem of singular value decomposition states that any matrix can be represented as in (7). SVD is a powerful tool and there is a lot of useful information behind it, but here only its relationship with PCD is considered:

$$H = U\Sigma V^H = \sum_{i=1}^{n} u_i \sigma_i v_i^T,$$
$$\text{where } \Sigma = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{pmatrix}, \tag{7}$$
$$\text{and } \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

According to PCD a matrix can be represented as a sum of its principal components. The components can be found as in equation 8 [3].

$$H = C_1 + C_2 + \cdots + C_r,$$
$$C_i = \sigma_i \sum_j u_{i,j} v_{j,i} = \sigma_i E_i, \quad (8)$$

The principal components C_i are decreasing in terms of their variance. The singular values are the weights of each principal component $\sigma_k = \|C_k\|$. In the direct problem, the larger singular values correspond to more significant principal components. However, in the inversion the least significant principal components get the largest weight as can be seen from equation 9.

$$H = E_1\sigma_1 + E_2\sigma_2 + \cdots + E_r\sigma_r,$$
$$H^\dagger = E_1^{-1} \cdot \frac{1}{\sigma_1} + E_2^{-1} \cdot \frac{1}{\sigma_2} + \cdots + E_r^{-1} \cdot \frac{1}{\sigma_r} \quad (9)$$

Thus, any small error multiplied by the last principal components will be emphasized by an inverted small singular value. It is hard to predict whether they dominate on the more significant components or not. In theory, the condition number as described in equation 10 is used to estimate it:

$$\mathrm{cond}(H) = \|H\|_2 \left\|H^\dagger\right\|_2 = {}^{\sigma_1}/_{\sigma_r} \quad (10)$$

It is worth mentioning that there are no tips in literature regarding what can be considered a critically large condition number for each specific matrix. Each problem must be considered individually.

However, if somehow a conclusion can be made that the measurements contain too many uncertainties, regularization is useful to make the result more stable. The purpose of regularization is to dampen or to filter out the small singular values. Thus, it reduces the condition number. A critical example of regularization is truncated SVD (TSVD): components that correspond to the small singular values are simply discarded. It is clear that this can change the matrix strongly, which means it changes the model of the problem.

A gentler and probably more popular regularization technique is Tikhonov regularization, which was used in the illustrated example as well. The topic of application of regularization and choice of the regularization parameter is very comprehensive, and for more details on its application, the reader can refer to the specific literature [4, 5].

In practice, there are other techniques useable as well for decreasing the matrix condition number, for instance, usage of overdetermination points or reduction of the matrix by discarding some transfer functions. It is important to keep in mind that decreasing the condition number does not necessarily yield an improvement of the results. The rule is rather "the larger – the possibly worse" and not "the smaller – the better". Moreover, this is another point about why the regularization must be used with care.

Estimation of ill-posed problems

To estimate how far the matrix is from being well-posed, the dependency of the resulting error regarding the measurement error is observed again but in a more general way. The condition number of a given matrix is changed by adjusting its singular values and observing the reaction of the new matrix to the measurement error. The simplified algorithm is shown in Figure 5.

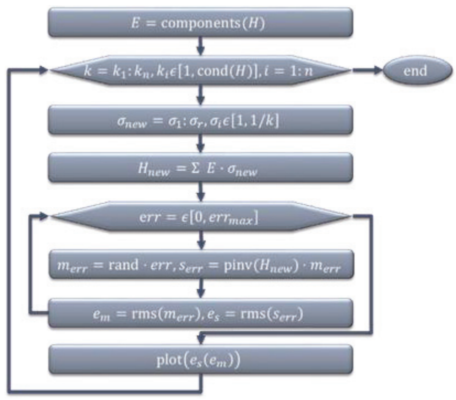

Figure 5: Observation of the dependency of resulting error over measurement error

In an ideal case, when the matrix condition number equals 1 (all the principal components have equal weights, e.g., all singular values are equal), the resulting error is supposed to be proportional to the measurement error: the larger the non-proportionality, the more ill-posed the problem. An example of such an observation for the sample problem is shown in Figure 6. The direction of the measurement error was randomly chosen for each inner loop of the algorithm (see Figure 5). As can be seen from Figure 6 (left), the ideal line (cond=1) is straight and the resulting error e_s matches the measurement error. Slope and fluctuations increase with increasing condition number. Some directions of the measurement error show much larger influence on the resulting error.

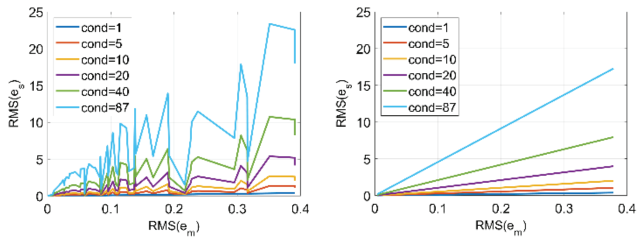

Figure 6: Resulting error e_s for different condition numbers, Left: Direction of measurement error e_m randomly chosen for each point, Right: Fixed direction of measurement error e_m for each point

Figure 6 (right) shows the resulting error e_s for a fixed direction of the measurement error e_m, leading to lines for any condition number, as expected. The influence of the direction of the measurement error e_m can be illustrated well as a two-dimensional graph in Figure 7. The resulting error is low for the 45° direction because of opposite directions of the vectors $\vec{p}_1 m_1$ and $\vec{p}_2 m_2$ (see Figure 3, right).

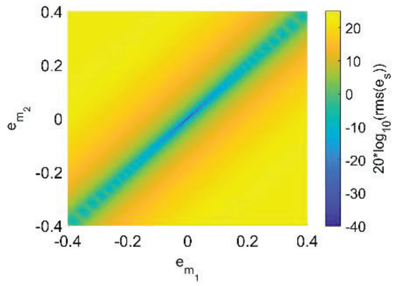

Figure 7: Resulting error e_s as a function of the direction of measurement error, i.e., its two components m_1 and m_2

By randomly changing the components of the measurement error, a statistical analysis of the resulting error can be given, as shown in Figure 8.

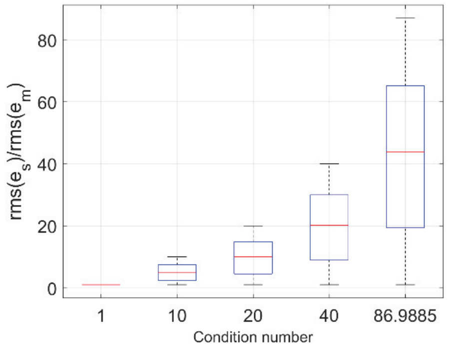

Figure 8: Box plot of the ratio of the resulting error to the measurement error for different condition numbers. The direction of the measurement error was changed randomly.

As can be seen, the measurement error grows proportionally to the condition number, in terms of mean value as well as in terms of the deviation. So making such an analysis of the original matrix, it can be estimated how ill-posed the problem is and how strongly it needs to be "improved" to meet any constraints of each specific problem.

Real application example

The usage of MIM was studied for the characterization of a distributed sound source in [6]. It was shown, how a model can be changed regarding the matrix condition number. For more details, please refer to [6]. At that moment, an optimum condition number was intuitively defined. Now, it can be shown how effective the improvement of the problem is due to the reduction of the condition number. Additionally, a threshold parameter of an ill-conditioned problem can be estimated.

In [6] two models for estimating the volume velocity of a vibrating plate at a frequency of 260 Hz were presented. One corresponds to a 20x20 transfer function matrix (case 1) and another to a 59x59 matrix (case 2). For such a low frequency, case 1 shows much better results. An error analysis similar to the one shown in Figure 8 is given in Figure 9.

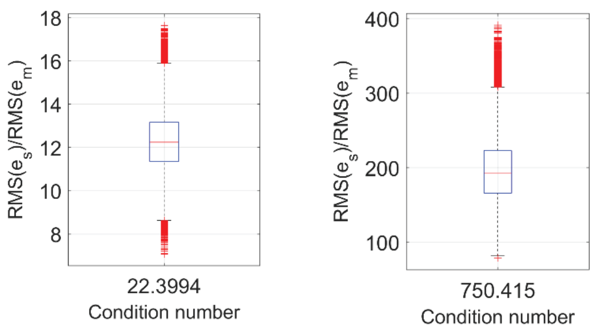

Figure 9: Observation of the dependency of the resulting error with respect to the measurement error for case 1 (left) and case 2 (right)

Besides the fact of a strong error amplification for case 2 by a factor of around 200 (median) compared to a factor of about 12 for case 1, the error distribution for case 2 is asymmetric and shows much more outliers at higher error values, whereas the distribution for case 1 is almost symmetric.

Conclusion

Ill-conditioning often becomes a problem of the matrix inversion method. First, it is difficult to determine if the system (the matrix) is ill-posed. The common indicator of an ill-conditioned matrix is a large condition number.

Mathematically, a large condition number can be caused by different reasons; some of them are a large size of the matrix, a wrong rank estimation, a large difference between the maximum and minimum value, a high correlation, etc. These features are observed in many practical applications, although they are difficult – or perhaps impossible – to prove in general.

However, there is no "threshold" value and each case has to be considered individually. For a "user" of the MIM it is important to know what kind of resulting error can be expected, knowing an approximate range of possible measurement errors. It can only be tested by trial, as was shown in Figure 4.

The common way of treating ill-conditioned matrices is regularization. However, it is important to know that the regularized solution is a biased estimator and can never be exact even for the case of no error. Regularization based on damping singular values of a matrix changes the model. Often a controllable change of the model by reconsidering it may provide results that are more reliable. There can be too many or too few measurement points, etc.

If regularization is needed, it is worth testing how strongly the matrix must be "improved". This can be achieved by observing the sensitivity of the MIM result to a measurement error. The results of such a sensitivity analysis should be compared for the cases of the original and the regularized matrix in order to see the effect of regularization, and if regularization is really needed.

Literature

[1] Velizhanina S., Sottek R.: Characterization of airborne sound sources using matrix inversion, DAGA 2018.

[2] Kirsch A.: An Introduction to the Mathematical Theory of Inverse Problems, 1996.

[3] Moler C. B.: Eigenvalues and Singular Values, Numerical Computing with MATLAB, Chapter 10, 2004, URL: https://bit.ly/2QYMXQR.

[4] Hansen P. C.: P. C. Hansen, Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems, Numerical Algorithms, 6 (1994).

[5] Hansen P.C.: The L-curve and its use in the numerical treatment of inverse problems, Computational Inverse Problems in Electrocardiology, 2000.

[6] Anisovich S., Sottek R., Wegerhoff M.: Analysis of a vibrating structure as an airborne sound source by means of matrix inversion, ICA 2019.