Why Is the Least Square Error Method Dangerous?

Vaclav Skala[1] Edward Kansa[2]

[1]University of West Bohemia, Faculty of Applied Sciences, Pilsen, Czech Republic
[2]Convergent Solutions, Livermore, CA, USA

[1]skala@kiv.zcu.cz [2]edwardjkansa@gmail.com

Abstract. This contribution briefly describes some "dangerous" features of the Least Square Error (LSE) methods, which are not generally known, but often used in applications, and researchers are not aware of those. The LSE is usually used in approximations of acquired data to find "the best fit" of the data, especially in financial economics and related fields. However, the LSE method is not invariant to some standard basic operations used within a solution of a linear system of equations.

Keywords. Least square error, system of linear equations, numerical mathematics, over determined system, invariant operations.

1 Introduction

The Least Square Error (LSE) is usually used for finding "the best fit" of measured data, which leads to a solution of an over-determined system of linear equations $\mathbf{Ax} = \mathbf{b}$. The LSE method is very often used in financially oriented applications using linear and non-linear regressions. In some specific cases, the total least square method is to be used, mostly related to the implicit representation [1][6][9][10].

However, the LSE method's result depends on the physical units of the data domain used in the polynomial regression case.

1.1 Linear System of equations

In the case of the linear system of equations $\mathbf{Ax} = \mathbf{b}$, when the matrix $\mathbf{A}$ $(n \times n)$ is non-singular, there are several standard methods for solving a linear system of equations [4]. However, solution of the linear system of equations $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{Ax} = 0$ is equivalent to the *outer product* (*extended cross product*)[8], and the modified Gauss elimination method can be used without division operation [7]. Some operations are used quite frequently, especially in connection with preconditioning or in a solution of the linear system, e.g. a row multiplication, a row swap, etc.

$$\mathbf{PAD}\,\mathbf{D}^{-1}\mathbf{x} = \mathbf{Pb} \qquad (1)$$

where: $\mathbf{P}$ and $\mathbf{D}$ are non-singular matrices $(n \times n)$. A simple preconditioning method for a large system of equations uses diagonal matrices $\mathbf{P}$ and $\mathbf{D}$ [15]. Multiplication of the i-th row of the extended matrix $[\mathbf{A}|\mathbf{b}]$ by $p_i \neq 0$ is invariant to the linear system's solution. The multiplication of the j-th column of the matrix $\mathbf{A}$ by $d_j \neq 0$ represents the unit change of the x_j, see Eq.1.

2 Over-determined systems

In the case of the over-determined linear system, the matrix $\mathbf{A}$ is $(n \times m)$, $n > m$, the vector $\mathbf{b}$ is $(m \times 1)$, the LSE is usually used to obtain an approximate solution. However, in many cases, users are not aware of the LSE properties [17]. It is well known, that a result of the LSE approximation depends on physical units used, if polynomial regression is used, e.g. in the estimation of processing time, etc.

Let us consider a regression function $\varphi(t)$:

$$\varphi(t) = a_0 + a_1 t + a_2 t \log(t) + a_3 t^2 + \dots \qquad (2)$$

If the time unit [s] is used, the results are different from the case, when the unit [ms] is used. Also, the element a_0, which represents a value for $t = 0$, causes some problems; detected also in interpolation and approximation using Radial Basis Function (RBF) [2][5] [13][14].

In the case of the ***linear regression***, the LSE method is usually applied directly to the data set using pseudo-inverse as follows:

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b} \quad ,i.e. \quad \mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \quad (3)$$

Let us consider the LSE formulation as in Eq.1, but modified for an over-determined system of linear equations. Then the LSE use leads to:

$$(\mathbf{PAD})^T\mathbf{PAD}\ \mathbf{D}^{-1}\mathbf{x} = (\mathbf{PAD})^T\mathbf{Pb} \quad (4)$$

where $\mathbf{P}(n \times n)$ and $\mathbf{D}(m \times m)$ are non-singular diagonal matrices. Using algebraic operations:

$$\mathbf{D}^T\mathbf{A}^T\mathbf{P}^T\ \mathbf{PAD}\ \mathbf{D}^{-1}\mathbf{x} = \mathbf{D}^T\mathbf{A}^T\mathbf{P}^T\ \mathbf{Pb} \quad (5)$$

As the matrix $\mathbf{D}$ is diagonal and non-singular, it is possible to multiply Eq.5 from the left by $(\mathbf{D}^T)^{-1}$. It results to:

$$\mathbf{A}^T\mathbf{QA}\ \mathbf{D}\ \mathbf{D}^{-1}\mathbf{x} = \mathbf{A}^T\mathbf{Q}\ \mathbf{b} \quad (6)$$

where $\mathbf{Q} = \mathbf{P}^T\mathbf{P}$ is a diagonal matrix of p_i^2 row multipliers.

If $\boldsymbol{\xi} = \mathbf{D}^{-1}\mathbf{x}$, then Eq.6 can be rewritten as:

$$\mathbf{A}^T\mathbf{QA}\ \mathbf{D}\ \boldsymbol{\xi} = \mathbf{A}^T\mathbf{Q}\ \mathbf{b} \quad (7)$$

then the solution of Eq.7 using LSE method:

$$\begin{aligned} \boldsymbol{\xi} &= (\mathbf{A}^T\mathbf{QA}\ \mathbf{D})^{-1}\ \mathbf{A}^T\mathbf{Q}\ \mathbf{b} \\ &= \mathbf{D}^{-1}(\mathbf{A}^T\mathbf{QA})^{-1}\ \mathbf{A}^T\mathbf{Q}\ \mathbf{b} \quad (8) \\ \mathbf{x} &= \mathbf{D}\boldsymbol{\xi} \quad (9) \end{aligned}$$

Therefore in the case of linear regression, the LSE method, Eq.3:

— ***is invariant*** to physical units used, if the transformation $\mathbf{x} = \mathbf{D}\boldsymbol{\xi}$ is used,

— ***is not invariant*** to row multiplications due to dependency on the matrix $\mathbf{P}$, resp. $\mathbf{Q}$, which represents multipliers of rows.

3 Example

Let us consider two simple examples of the LSE use for two different simple cases with a modification, when the first row of the extended matrix $[\mathbf{A}|\mathbf{b}]$ is multiplied by the value 10:

— the first case - a function is given as $z = a_1x + a_2y$, i.e. a plane passing the origin, and values of (x, y, z) are given as $(1, 2, 1), (2, 2, 2), (3, 7, 7)$

$$\begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} \qquad \begin{bmatrix} 10 & 30 \\ 2 & 2 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 7 \end{bmatrix} \quad (10)$$

The solutions are $\mathbf{x} = [11/21, 2/3]^T$ and $\mathbf{x} = [275/129, -46/129]^T$.

— the second case - a function is given as $y = kx + q$, i.e. a line in E^2 not passing the origin, and values of (x, y) are given as $(1, 1), (2, 2), (3, 7)$.

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} k \\ q \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} \qquad \begin{bmatrix} 10 & 10 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} k \\ q \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 7 \end{bmatrix} \quad (11)$$

The solutions are $\mathbf{x} = [3, -3/4]^T$ and $\mathbf{x} = [435/167, -808/501]^T$.

These elementary examples serve to understand the limitations of the LSE use. The results are valid for d-dimensional space, in general. In the first case, usually, users are not aware of that. The second case can be easily understood as the k represents a normal vector generally in a higher dimension, while q is related to a distance from the origin. The solution of Eq.3 might be unstable, as the matrix $\mathbf{A}^T\mathbf{A}$ is generally numerically ill-conditioned [3][11][13][14].

It should be noted, that in many cases the Total Least Square Error (TLSE) should be used instead. However, it leads to more complicated computation [9][16].

A simple preconditioning [12][15] should be considered. Also, the modified Gauss elimination method [7][8] can be used as a solution of a linear system is equivalent to the outer product use.

4 Conclusion

This contribution describes selected mostly unknown properties of the Least Square Error for the approximation of acquired data. The LSE method is non-invariant to the multiplication of a row of the extended matrix $[\mathbf{A}|\mathbf{b}]$. Also, in the case of non-existent metric between parameters, like a distance and a normal vector, the LSE based approximation should not be used.

Acknowledgements

The author would like to thank colleagues and students at the University of West Bohemia for hints and suggestions. Thanks also belong to colleagues at Shandong University and Zhejiang University (China) for their critical comments and constructive suggestions and to anonymous reviewers for their valuable comments and hints provided.

This research was partially supported by the Czech Science Foundation (GACR), project No. GA 17-05534S.

References

1. **Alciatore, D. G. & Miranda, H. P. (1995).** The best least-squares line fit. In **Paeth, A. W.**, editor, *Graphics Gems V*. Academic Press, pp. 91–97.
2. **Cervenka, M. & Skala, V. (2020).** Conditionality analysis of the radial basis function matrix. *LNCS*, Vol. 12250, pp. 30–43.
3. **Kansa, E. & Holoborodko, P. (2017).** On the ill-conditioned nature of c^{∞} RBF strong collocation. *Engineering Analysis with Boundary Elements*, Vol. 78, pp. 26–30.
4. **Lay, D. (2006).** *Linear Algebra and Its Applications*. Pearson international edition. Pearson/Addison-Wesley.
5. **Majdisova, Z. & Skala, V. (2017).** Big geo data surface approximation using radial basis functions: A comparative study. *Computers & Geosciences*, Vol. 109, pp. 51 – 58.
6. **Markovsky, I. & Van Huffel, S. (2007).** Overview of total least-squares methods. *Signal Processing*, Vol. 87, No. 10, pp. 2283–2302.
7. **Skala, V. (2013).** Modified Gaussian elimination without division operations. *ICNAAM 2013, AIP Conference Proceedings*, Vol. 1558, pp. 1936–1939.
8. **Skala, V. (2016).** ”Extended cross-product” and solution of a linear system of equations. *Lecture Notes in Computer Science*, Vol. 9786, pp. 18–35.
9. **Skala, V. (2016).** A new formulation for total least square error method in d-dimensional space with mapping to a parametric line. *Proceedings of the ICNAAM 2016*, volume 1738, pp. 480106–1 – 480106–4.
10. **Skala, V. (2016).** Total least square error computation in E2: A new simple, fast and robust algorithm. *Proceedings of the 33rd Computer Graphics International*, CGI ’16, Association for Computing Machinery, New York, NY, USA, pp. 1=4.
11. **Skala, V. (2017).** High Dimensional and Large Span Data Least Square Error: Numerical Stability and Conditionality. *Int. J. Appl. Phys. Math.-IJAPM*, Vol. 7, pp. 148–156.
12. **Skala, V. (2017).** Least square method robustness of computations: What is not usually considered and taught. *FedCSIS 2017*, pp. 537–541.
13. **Skala, V. (2017).** RBF Interpolation with CSRBF of Large Data Sets. *Proceedings of the ICCS 2017*, volume 108, pp. 2433–2437.
14. **Skala, V. (2018).** RBF Approximation of Big Data Sets with Large Span of Data. *Proceedings of the MCSI 2017*, volume 2018-January, pp. 212–218.
15. **Skala, V. (2020).** Conditionality of linear systems of equations and matrices using projective geometric algebra. *Computational Science and Its Applications – ICCSA 2020*, Springer, pp. 3–17.
16. **Smolik, M., Skala, V., & Majdisova, Z. (2020).** A new simple, fast and robust total least square error computation in E2: Experimental comparison. *Lecture Notes in Electrical Engineering*, Vol. 554, pp. 325–334.
17. **Smolik, M., Skala, V., & Nedved, O. (2016).** A comparative study of LOWESS and RBF approximations for visualization. *LNCS*, Vol. 9787, pp. 405–419.

Article received on 08/09/2020; accepted on xx/xx/xxxx.
Corresponding author is Vaclav Skala