The Significance of Imaginary Points in Linear Least Square Approximation

Mikael Fridenfalk
Department of Game Design
Uppsala University
Visby, Sweden
mikael.fridenfalk@speldesign.uu.se

***Abstract*—The linear least square method constitutes one of the most useful statistical methods in mathematics. This paper shows that in this method, real, versus imaginary points may act as minimizers, versus maximizers of the error. The use of imaginary points provides thereby an additional degree of freedom in the design of methods based on this statistical method.**

Keywords-complex number; curve fitting; imaginary number; least square method

I. Introduction

Previous research on complex numbers in conjunction with the linear least square method has been restricted to applications, such as constrained phases [2], stochastic processes [5], and complex monomial neural networks [1]. In this paper, a new application is presented, using real, versus imaginary points as error minimizers, versus maximizers. The initial motivation for the development of the method presented in this paper was to accelerate the backpropagation process in the evaluation of the weights of a large-scale feedforward neural network. An analytic solution was however found along the way with the potential to replace backpropagation in large-scale neural networks [4]. As a brief overview, in Section II, the theory behind the linear least square is reiterated, along with a curve-fitting example. In Section III, a new and more generalized method is proposed for linear least square fitting, including imaginary points, followed by experimental results in Section IV for the verification of the proposed method.

II. State of the Art

To begin with, the theory behind the standard method is reiterated. To reproduce a textbook example on the subject [3], given the inconsistent linear equation system:

$$x_1 + x_2 = 4 \tag{1}$$
$$2x_1 + x_2 = 8 \tag{2}$$
$$x_1 + 2x_2 = 5 \tag{3}$$

or alternatively expressed, using a matrix $\mathbf{A}$ of size $M \times N$ (with $M = 3$ and $N = 2$), and the vectors $\mathbf{x}$ and $\mathbf{b}$:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{4}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} \tag{5}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{6}$$

$$\mathbf{b} = \begin{bmatrix} 4 \\ 8 \\ 5 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{7}$$

where $\mathbf{c}_1$ and $\mathbf{c}_2$ are the column vectors of $\mathbf{A}$, and $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ the row vectors of $\mathbf{A}$. It is possible to schematically represent such system by Fig. 1, with:

$$\mathbf{b} = \mathbf{p} + \mathbf{q} \tag{8}$$

where $\mathbf{b}$ denotes in this example a vector outside a plane V, spanned by $\mathbf{c}_1$ and $\mathbf{c}_2$, and $\mathbf{p}$ denotes the orthogonal projection of $\mathbf{b}$ on V, as shown in Fig. 1. Due to orthogonal projection, the smallest distance from $\mathbf{b}$ to V is $\epsilon = |\mathbf{p} - \mathbf{b}|$, and therefore $\epsilon^2 = |\mathbf{p} - \mathbf{b}|^2$ is minimal. The least square error may in this example be expressed as:

$$\epsilon^2 = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 \tag{9}$$

$$|\mathbf{p} - \mathbf{b}|^2 = (\mathbf{r}_1 \cdot \mathbf{x} - b_1)^2 + (\mathbf{r}_2 \cdot \mathbf{x} - b_2)^2 + (\mathbf{r}_3 \cdot \mathbf{x} - b_3)^2 \tag{10}$$

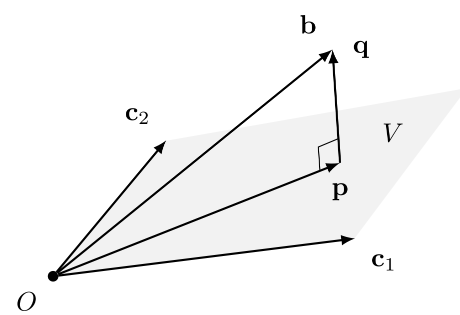

Figure 1. Orthogonal projection of $\mathbf{b}$ on plane V, spanned by $\mathbf{c}_1$ and $\mathbf{c}_2$.

or in the general case, given row m in $\mathbf{A}$ and element m in $\mathbf{b}$, with $\mathbf{p} = \mathbf{A}\mathbf{x}$, as:

$$\epsilon_m = |\mathbf{r}_m \cdot \mathbf{x} - b_m| \tag{11}$$

Given an arbitrary vector $\mathbf{y}$ in $\mathbb{R}^N$:

$$\mathbf{A}\mathbf{y} = y_1\mathbf{c}_1 + y_2\mathbf{c}_2 + \ldots + y_N\mathbf{c}_N \tag{12}$$

Since $\mathbf{A}\mathbf{y}$ will always lie in V, and is, therefore, orthogonal to $\mathbf{q}$, the linear least square method can be derived by the following equations:

$$(\mathbf{A}\mathbf{y}) \cdot \mathbf{q} = (\mathbf{A}\mathbf{y})^T\mathbf{q} = 0 \tag{13}$$
$$(\mathbf{A}\mathbf{y})^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0 \tag{14}$$
$$\mathbf{y}^T\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0 \tag{15}$$
$$\mathbf{y}^T(\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{A}^T\mathbf{b}) = 0 \tag{16}$$

Thus:

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b} \tag{17}$$
$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{18}$$

 ISBN: 978-1-61208-354-4

The estimated solution $\mathbf{x}$, is for maximum evaluation speed, preferably solved by the direct solution of (17). As an example of a curve fitting application, given M points (a_m, b_m):

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix} \tag{19}$$

and a second degree polynomial:

$$b = x_1 + x_2 a + x_3 a^2 \tag{20}$$

$$\mathbf{A} = \begin{bmatrix} 1 & a_1 & a_1^2 \\ 1 & a_2 & a_2^2 \\ \vdots & \vdots & \vdots \\ 1 & a_N & a_N^2 \end{bmatrix} \tag{21}$$

Equations (19) and (21) with (17) give thus the least square solution $\mathbf{x}$ for the polynomial in (20), that best approximates the distribution of the sample points in (19).

III. Proposal

As an example of the extension of the standard method described above for least square curve fitting, we present here a similar method, but with the addition of imaginary points. The definition of an imaginary point is in this paper any row m in $\mathbf{A}$ and element m in $\mathbf{b}$, that has been multiplied with i, as in $\sqrt{-1}$. In the following example, $m = 2$ is designated as an imaginary point:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \cdot i \\ \vdots \\ a_M \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \cdot i \\ \vdots \\ b_M \end{bmatrix} \tag{22}$$

Given the polynomial equation in (20):

$$\mathbf{A} = \begin{bmatrix} 1 & a_1 & a_1^2 \\ i & a_2 \cdot i & a_2^2 \cdot i \\ \vdots & \vdots & \vdots \\ 1 & a_N & a_N^2 \end{bmatrix} \tag{23}$$

and the definitions of the square matrix $\mathbf{U} = \mathbf{A}^T\mathbf{A}$ and the vector $\mathbf{v} = \mathbf{A}^T\mathbf{b}$:

$$\mathbf{U} = \begin{bmatrix} 1 & i & \dots & 1 \\ a_1 & a_2 \cdot i & \dots & a_N \\ a_1^2 & a_2^2 \cdot i & \dots & a_N^2 \end{bmatrix} \begin{bmatrix} 1 & a_1 & a_1^2 \\ i & a_2 \cdot i & a_2^2 \cdot i \\ \vdots & \vdots & \vdots \\ 1 & a_N & a_N^2 \end{bmatrix} \tag{24}$$

$$\mathbf{v} = \begin{bmatrix} 1 & i & \dots & 1 \\ a_1 & a_2 \cdot i & \dots & a_N \\ a_1^2 & a_2^2 \cdot i & \dots & a_N^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \cdot i \\ \vdots \\ b_N \end{bmatrix} \tag{25}$$

Since $i^2 = -1$, both $\mathbf{U}$ and $\mathbf{v}$ will only consist of real numbers, yielding the least square equation:

$$\mathbf{U}\mathbf{x} = \mathbf{v} \tag{26}$$

Given the standard least square method, based on M sample points, the squared error may for any sample m be expressed as:

$$\epsilon_m^2 = (\mathbf{r}_m\mathbf{x} - b_m)^2 \tag{27}$$

or more explicitly:

$$\epsilon_m^2 = \left(\begin{bmatrix} a_{m1} & a_{m2} & \dots & a_{mN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} - b_m \right)^2 \tag{28}$$

Therefore, the total squared error is equal to:

$$\epsilon^2 = \sum_{m=1}^{M} (\mathbf{r}_m\mathbf{x} - b_m)^2 \tag{29}$$

The division of the M sample points into J real points (a_j, b_j), versus K imaginary points (a_k, b_k), yields the following equations for the squared errors:

$$\epsilon_j^2 = \left(\begin{bmatrix} 1 & a_j & a_j^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - b_j \right)^2 \tag{30}$$

$$\epsilon_k^2 = \left(\begin{bmatrix} i & a_k \cdot i & a_k^2 \cdot i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - b_k \cdot i \right)^2 \tag{31}$$

By extraction of i and the relation $i^2 = -1$:

$$\epsilon_k^2 = (\mathbf{r}_k\mathbf{x} \cdot i - b_k \cdot i)^2 \tag{32}$$

$$\epsilon_k^2 = i^2(\mathbf{r}_k\mathbf{x} - b_k)^2 \tag{33}$$

$$\epsilon_k^2 = -(\mathbf{r}_k\mathbf{x} - b_k)^2 \tag{34}$$

$$\epsilon^2 = \sum_{j=1}^{J} (\mathbf{r}_j\mathbf{x} - b_j)^2 - \sum_{k=1}^{K} (\mathbf{r}_k\mathbf{x} - b_k)^2 \tag{35}$$

Thus, imaginary points seem to reverse the direction or “polarity” of the least square method. However, since this method is based on projection, and the parameter that is minimized is $|\epsilon|$, large error components may reverse the expected polarities of sample points.

IV. Experimental Results

An equation solver was developed in C++ for the solution of linear equation systems of the form $\mathbf{U}\mathbf{x} = \mathbf{v}$. To optimize the evaluation speed of $\mathbf{U}$ and $\mathbf{v}$, a specialized matrix multiplication method was implemented, by the reversal of the signs of the products that correspond to imaginary points. Using this equation solver, Figs. 2-9 show the results of curve fitting of a third degree polynomial, based on real ($\bullet$), versus imaginary points ($\times$).

To comment these figures, Fig. 2 presents a curve fitting experiment using a third degree polynomial, with four real points, and an imaginary point, placed slightly above the line formed by the real points, yielding due to symmetry, a parabolic curve. As shown here, while the distance to the real points is minimized, the distance to the imaginary point is maximized. Figure 3 presents the same case as in previous figure, but with an imaginary point placed on the same line as the one formed by the real points, yielding as expected a flat line. Figure 4 presents the same case as in previous figure, but here with an imaginary point placed slightly below the line formed by the real points. Figure 5 presents the inversion of real versus imaginary points with respect to previous figure. Since the least square method minimizes $|\epsilon|$, even if the system is reversed (with $\epsilon^2 < 0$, where ϵ is an imaginary number), as expected by the derived theory, the result remains the same. Figure 6 presents a least square curve fitting of a third degree polynomial based on seven real points. Figures 7-8 present the same case as in Fig. 6, except for the replacement of a real point with an imaginary. Regarding Fig. 9, in our experiments, a borderline point such as this showed to nominally reverse the polarity of an imaginary point. A borderline point seems thus to be able to affect the error by a significant amount.

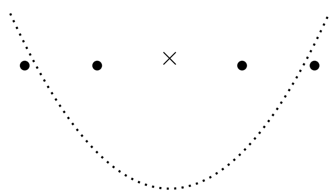

Figure 2. Curve fitting using a third degree polynomial, with four real points (●), and an imaginary point ($\times$).

Figure 3. Same as previous figure, but with an imaginary point placed on the same line as the one formed by the real points.

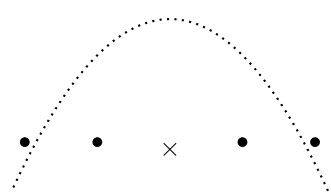

Figure 4. Same as previous figure, but here with an imaginary point placed slightly below the line formed by the real points.

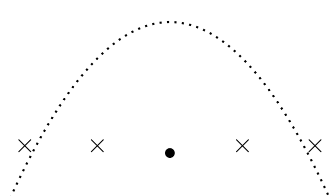

Figure 5. Inversion of real versus imaginary points.

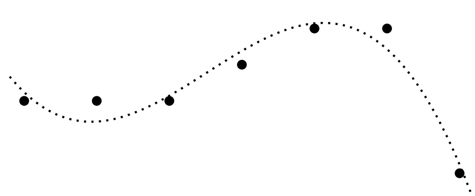

Figure 6. Least square curve fitting of a third degree polynomial based on seven real points.

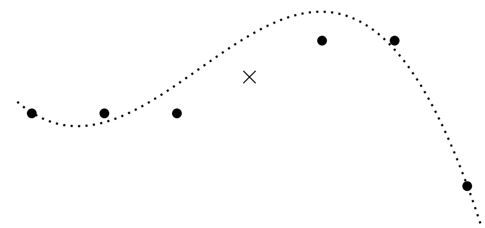

Figure 7. Same as previous figure, except for the replacement of a real point with an imaginary.

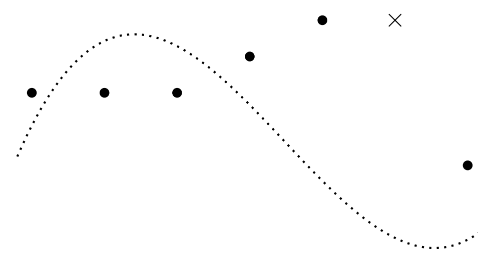

Figure 8. Same as previous figure.

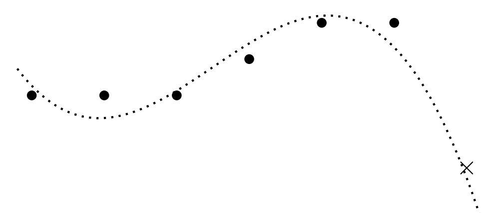

Figure 9. A demonstration of the effect of a borderline point.

V. CONCLUSION

The linear least square method is associated with the evaluation of coefficients of linear equation systems, minimizing errors. This paper shows that pure imaginary points (as defined in this paper) in a such system, tend nominally to act as squared error maximizers instead of minimizers. According to experimental results, the new method needs to be used with caution, since crossover between real and imaginary errors affects the behavior of the system.

REFERENCES

[1] M. F. Amin, R. Savitha, M. I. Amin, and K. Murase, “Orthogonal Least Squares Based Complex-Valued Functional Link Network”, Neural Networks, Elsevier, vol. 32, 2012, pp. 257-266.

[2] M. Bydder, “Solution of a Complex Least Squares Problem with Constrained Phase”, Linear Algebra Appl., vol. 433, no. 10-11, 2010, pp. 1719-1721.

[3] C. H. Edwards and D. E. Penney, Elementary Linear Algebra, Prentice Hall, 1988.

[4] M. Fridenfalk, “The Development and Analysis of Analytic Method as Alternative for Backpropagation in Large-Scale Multilayer Neural Networks”, accepted for publication in the proceedings of The Eighth International Conference on Advanced Engineering Computing and Applications in Sciences, ADVCOMP 2014, Rome, Italy, August, 2014.

[5] K. S. Miller, “Complex Linear Least Squares”, SIAM Review, Society for Industrial and Applied Mathematics, vol. 15, no. 4, 1973, pp. 706-726.