v. 4, n. 29, 2024

Journal of
Engineering
Research

ALGORITHM FOR THE GENERATION OF THE SYSTEM OF EQUATIONS USED IN THE LEAST SQUARES METHOD FOR POLYNOMIALS OF DEGREE 2 AND 4

Acceptance date: 27/12/2024

Oscar Leopoldo Pérez Castañeda
Tecnológico Nacional de México-Instituto Tecnológico de Tehuacán, Department of Electrical-Electronic Engineering
https://orcid.org/0000-0002-3271-5479

Jesús Daniel Pérez Castañeda
Tecnológico Nacional de México-Instituto Tecnológico de Tehuacán, Department of Electrical-Electronic Engineering
https://orcid.org/0000-0001-7774-8031

José Enrique Salinas Carrillo
Tecnológico Nacional de México-Instituto Tecnológico de Tehuacán Civil Engineering Career
http://orcid.org/0000-0002-4559-3132

Javier Trucíos Alonso
Tecnológico Nacional de México-Instituto Tecnológico de Tehuacán, Department of Electrical-Electronic Engineering
https://orcid.org/0009-0005-5113-9699

María Fernanda Tovany Salvador
Tecnológico Nacional de México-Instituto Tecnológico de Tehuacán, Department of Electrical-Electronic Engineering
https://orcid.org/0009-0001-3425-3116

Abstract: When fitting functions to a set of experimental data the least squares method (LSM) is used, considering that it will have the minimum error. This method is used to fit the function to a linear or not necessarily linear curve. This curve is associated with a polynomial of degree *n*, where the objective is to find each of the coefficients of the polynomial. The application of the CMM leads to a system of linear equations of the form $\boldsymbol{A\beta} = \boldsymbol{b}$, where finding the values of the column vector $\boldsymbol{\beta} = \boldsymbol{A}^{-1}$ b is the specific objective of the problem. The laboriousness of this problem does not lie in solving the system of linear equations but in generating it. In this work an algorithm is proposed which generates the equations that integrate the system of linear equations to obtain the coefficients of a polynomial of degree *n*, without using the formal method. To illustrate the use of this algorithm two examples are presented, the free fall of an object and the solar irradiance. The CMM for the generation of the equations that form the system of linear equations for obtaining the coefficients of the proposed polynomials is presented in detail. Then the proposed algorithm is presented. After that, the proposed algorithm is used to generate the equations of the system of linear equations. Finally, the systems of linear equations obtained by using the formal method and the proposed algorithm are compared. The results show the ease with which the system of linear equations is generated by using the proposed algorithm compared to the formal CMM.
Keywords: algorithm, coefficients of a polynomial, least squares method, linear regression, system of linear equations.

INTRODUCTION

In everyday problems, an analytical mathematical expression is rarely available to predict the quantitative value of the dependent variable of the problem or phenomenon to be dealt with. In the face of this real constraint, the collection, statistics and analysis of data obtained from the measurements made is necessary. From these data, a mathematical method can be applied to obtain an analytical expression that allows predicting the quantitative value of the dependent variable. For this purpose, the CMM or linear regression is originally available. This is a mathematical method used to predict the quantitative value of the dependent variable, leading to know the values of the coefficients or constants that are part of the polynomial that best fits the data.

In most of the examples found in the literature, the CMM is presented for polynomials of degree one, two and exceptionally degree three, due to the laboriousness of generating the system of linear equations, which leads to obtaining the values of the coefficients sought. The solution of the system of linear equations generated by the formal CMM is not complicated since there are different direct or mathematical methods that, with the help of the computer, simplify the task of solving the system and, with it, the obtaining of the values of the coefficients of the polynomial to be treated.

THEORETICAL FRAMEWORK

The CMM is based on simple regression. When working with two variables, only one independent variable and one dependent variable are considered. From the given changes in the independent variable the value of the dependent variable can be determined.

The use of CMM in modern statistics dates back to Greek mathematicians. However, the first modern precursor may be Galileo (Hervé, 2007). The essential idea of this method is to obtain the best least squares fit by minimizing the sum of quadratic residuals; where the residual is considered as the difference between the observed value and the value fitted by the model. The simplest model consists of fitting a straight line but this distribution of points can be better fitted to a higher degree polynomial.

The use of the CMM is privileged over other methods in modeling processes due to its effectiveness and completeness. In relation to obtaining a model, it is true that most of the times a linear function is used as a starting point, and sometimes this model is enough, but in other behaviors or phenomena of Science and Engineering the linear model is not enough. The use of data describing a behavior and the minimum amount of these, has made the CMM an interesting and powerful tool Neil (2007). The theoretical underpinning related to the CMM allows different types of statistical intervals easily interpretable in works such as calibration, optimization and prediction (Neil, 2007). The CMM provides information about a data set in a precise way and opens the way for the study of more complex phenomena.

Due to the nature of the CMM whose essence is the quadratic residual, the high sensitivity to extreme data or observations is one of the most considerable disadvantages of this method. This is a consequence of using the square of a number because the quadratic function "overstates", betrays or amplifies the departure from the reference value. For example, the squared difference between 20 and 10 is 100, but the squared difference between 202 and 102 is 10,000, so the CMM pays attention to extreme observations (Abdi et al., 2013). Despite this considerable disadvantage of the CMM different disciplines apply it.

In practice, various applications are encountered where the functions have linear and/or k-power terms, where k $\in N$. In this particular case the expression that best fits the data is expressed by equation (1), where i is the i-th data.

$$y_i=\beta_0+\beta_1\cdot x_i+\beta_2\cdot x_i^2+\beta_3\cdot x_i^3+\cdots+\beta_k\cdot x_i^k+\epsilon \qquad (1)$$

ϵ is the error or residual and $\beta_0, \beta_1, \beta_2, \ldots \beta_k$ the regression coefficients. Now, considering that we have n one sample data (measurements taken), then from equation (1) we have equation (2).

$$y_i=f(x_i)=\beta_0+\beta_1\cdot x_i+\beta_2\cdot x_i^2+\beta_3\cdot x_i^3+\cdots+\beta_k\cdot x_i^k \qquad (2)$$

Where $i \in \{1,2,3, \ldots n\}$, and n the number of data in the given sample.

By subtracting the error ϵ from equation (1) and, expressing it in terms of equation (2), equation (3) is obtained.

$$\epsilon = y_i - f(x_i) \qquad (3)$$

Since some errors (residuals) can be positive or negative, depending on whether the data is above or below the curve, then it is preferable to use the squares of the errors to quantify the error relative to the curve. Hence the error ϵ is the sum of the squares of the differences (standard deviation) between the y values on the approximation line or curve and the given y values$_i$, then this generates equation (4):

$$\epsilon_i^2 = [y_i - f(x_i)]^2 \qquad (4)$$

Now it is necessary to find the coefficients or constants $\beta_0, \beta_1, \ldots \beta_k$ of equation (1) that minimize the least squares error. Therefore, considering the n data of the given sample we obtain equation (5).

$$\sum_{i=1}^{n}\epsilon_i^2 = \sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots \beta_k x_i^q)]^2 \qquad (5)$$

The general problem of fitting the optimal least squares line to a set of data pairs $\{(x_i, y_i)\}$, where $i \in \{1,2,3, \ldots n\}$, where n is the number of data in the sample, requires minimizing the error in equation (5) with respect to the parameters $\beta_0, \beta_1, \ldots \beta_k$, i.e., it is necessary to minimize the error function. To find the minimum value, first the partial derivative of ϵ_i^2 with respect to each coefficient must be applied. After that, each equation must be equated to zero, thus generating a system of equations. Finally, solve said system of equations to obtain the values of the coefficients or constants of the treated polynomial. This procedure is represented by the set of equations (I), Nakamura (1992), Burden (1998) and Samir (2019).

$$\frac{\partial \in^2}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots \beta_k x_i^m)]^2 = 0$$

$$\frac{\partial \in^2}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots \beta_k x_i^m)]^2 = 0 \quad \text{(I)}$$

$$\vdots$$

$$\frac{\partial \in^2}{\partial \beta_k} = \frac{\partial}{\partial \beta_k} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots \beta_k x_i^m)]^2 = 0$$

The system of equations (I) reduces to a system of m x m linear equations of the form: $\boldsymbol{A\beta=b}$ where $\boldsymbol{A}$ is the matrix of coefficients, $\boldsymbol{b}$ is the constant vector, and $\boldsymbol{\beta}$ is the vector formed by the coefficients β_0, β_1, β_2 ,... β_k of the polynomial.

There are scenarios in which when making physical measurements, the behavior of the data set conforms to a polynomial of order equal to or greater than two. While this may be simple to understand, it is not always obvious. Therefore, it has been determined to illustrate the use of the CMM for the generation of the system of linear equations that leads to obtaining the coefficients of the polynomial to be treated, two examples presented to fulfill this task.

DEVELOPMENT OF THE MMC

EXAMPLE 1: GRAVITATIONAL ACCELERATION CONSTANT

The height at which an object is released from an initial height d_0 with an initial velocity v_0 and with the value of the gravitational acceleration constant g, for a given time t, is given by equation (6).

$$d = d_0 + v_o t + \tfrac{1}{2} g t^2 \quad (6)$$

where:

d, is the height or distance traveled by the object in free fall.

d_0, is the initial height at which the object is released.

v_o, is the initial velocity at which the body or object begins its fall.

g, the gravitational acceleration constant.

t, the time taken to travel a given distance d.

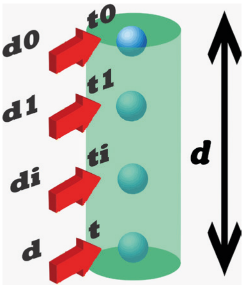

Figure 1. Object in free fall.

Generally in equation (6), the values of d_0, v_0 and g are considered as known constants, having as independent variable the time t and d the dependent variable, but in practice the coefficients of the polynomial are unknown and it is when the CMM is applied to know their values. In order to illustrate the use of the CMM the following scenario is proposed: suppose that from equation (6) we know the time t_i that the object has taken to travel a certain distance or height d_i and, we want to know the values of the coefficients d_0, v_0 and g from a set of data obtained as samples or tests performed see figure 1. In this example, the pairs of data considered in different measurements are: $\{(t_1, d_1), (t_2, d_2), (t_3, d_3), (t_4, d_4), (t_5, d_5)\}$, i.e., $n = 5$.

It was assumed that there is an observer located at a height d from the ground, and he sees how the object is falling; being $d_1 = 0$ and $v_1 = 0$, using equation (6), for which it is desired to make the adjustment. In equation (6) the pairs of known values are (t, d_{ii}), and the constants to be obtained are d_0, v_0 *and* g. According to the CMM, it is proposed that equation (6) be expressed as in equation (7):

$$x(t_i) = d_0 + v_0 t_i + \tfrac{1}{2} g {t_i}^2 \quad (7)$$

In the first member of equation (7)$x(t_i)$ corresponds to y_i of equation (5). The second member of equation (7), $d_0 + v_0 t_i + \frac{1}{2} g t_i{}^2$ corresponds to the general polynomial in equation (5). So, substituting equation (7) into equation (5) to obtain the sum of squares of the errors or residuals, we have equation (8).

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} [x(t_i) - (d_0 + v_0 t_i + \frac{1}{2} g\, t_i^2)]^2 \qquad (8)$$

The next step is to optimize the function, minimizing the error described by equation (8).

The partial derivative of the function described in equation (8) is applied with respect to d_0, v_0 and g, which are the coefficients of the polynomial to be obtained and then each equation is equaled to zero. The above is presented in equations (9), (10), and (11).

$$\frac{\partial}{\partial d_0} \sum_{i=1}^{n} \epsilon_i^2 = 2 \sum_{i=1}^{n} [x(t_i) - (d_0 + v_0 t_i + \frac{1}{2} g t_i^2)]\,(1) = 0 \qquad (9)$$

$$\frac{\partial}{\partial v_0} \sum_{i=1}^{n} \epsilon_i^2 = 2 \sum_{i=1}^{n} [x(t_i) - (d_0 + v_0 t_i + \frac{1}{2} g t_i^2)]\,(t_i) = 0 \qquad (10)$$

$$\frac{\partial}{\partial g} \sum_{i=1}^{n} \epsilon_i^2 = 2 \sum_{i=1}^{n} [x(t_i) - (d_0 + v_0 t_i + \frac{1}{2} g t_i^2)](\frac{1}{2} t_i^2) = 0 \qquad (11)$$

$x(t_i) = x_i$ and developing equations (9)-(11), equations (*9a*), (*10a*) and (*11a*) are obtained:

$$\sum_{i=1}^{n} d_0 + \sum_{i=1}^{n} v_0 t_i + \frac{1}{2} \sum_{i=1}^{n} g\, t_i^2 = \sum_{i=1}^{n} x_i \qquad (9a)$$

$$\sum_{i=1}^{n} d_0 t_i + \sum_{i=1}^{n} v_0 t_i^2 + \frac{1}{2} \sum_{i=1}^{n} g\, t_i^3 = \sum_{i=1}^{n} x_i\, t_i \qquad (10a)$$

$$\sum_{i=1}^{n} d_0 t_i^2 + \sum_{i=1}^{n} v_0 t_i^3 + \frac{1}{2} \sum_{i=1}^{n} g\, t_i^4 = \sum_{i=1}^{n} x_i t_i^2 \qquad (11a)$$

Since five pairs of data are available then the summations range from i = 1 to n = 5. Developing and rearranging the terms of equations (*9a*)-(*11a*), the system of linear equations formed by equations (12), (13) and (14) is obtained.

$$5 d_0 + \sum_{i=1}^{5} t_i v_0 + \frac{1}{2} \sum_{i=1}^{5} t_i^2 g = \sum_{i=1}^{5} x_i \qquad (12)$$

$$\sum_{i=1}^{5} t_i d_0 + \sum_{i=1}^{5} t_i^2 v_0 + \frac{1}{2} \sum_{i=1}^{5} t_i^3 g = \sum_{i=1}^{5} x_i t_i \qquad (13)$$

$$\sum_{i=1}^{5} t_i^2 d_0 + \sum_{i=1}^{5} t_i^3 v_0 + \frac{1}{2} \sum_{i=1}^{5} t_i^4 g = \sum_{i=1}^{5} x_i t_i^2 \qquad (14)$$

Equations (12)-(14) yield a matrix equation of the form ***Aβ=b*** where ***A*** is the 3 x 3 coefficient matrix represented in equation (15), ***β*** is the column vector to be obtained formed by the constants d_0, v_0 and g equation (16) and finally ***b*** is the 3 x 1 constant column vector equation (17).

$$\boldsymbol{A} = \begin{pmatrix} 5 & \sum_{i=1}^{5} t_i & \frac{1}{2}\sum_{i=1}^{5} t_i^2 \\ \sum_{i=0}^{4} t_i & \sum_{i=1}^{5} t_i^2 & \frac{1}{2}\sum_{i=1}^{5} t_i^3 \\ \sum_{i=1}^{5} t_i^2 & \sum_{i=1}^{5} t_i^3 & \frac{1}{2}\sum_{i=1}^{5} t_i^4 \end{pmatrix} \qquad (15)$$

$$\boldsymbol{\beta} = \begin{pmatrix} d_0 \\ v_0 \\ g \end{pmatrix} \qquad (16)$$

$$\boldsymbol{b} = \begin{pmatrix} \sum_{i=1}^{5} x_i \\ \sum_{i=1}^{5} x_i \cdot t_i \\ \sum_{i=1}^{5} x_i t_i^2 \end{pmatrix} \qquad (17)$$

Thus, the solution of the matrix equation ***Aβ=b*** to solve this matrix equation is described by equation (18), provided that $\mathbf{A}^{-1} \neq 0$.

$$\boldsymbol{\beta} = \boldsymbol{A}^{-1}\boldsymbol{b} \qquad (18)$$

So, for this example the solution can be expressed as in matrix equation (II).

$$\begin{pmatrix} d_0 \\ v_0 \\ g \end{pmatrix} = \begin{pmatrix} 5 & \sum_{i=1}^{5} t_i & \frac{1}{2}\sum_{i=1}^{5} t_i^2 \\ \sum_{i=1}^{5} t_i & \sum_{i=1}^{5} t_i^2 & \frac{1}{2}\sum_{i=1}^{5} t_i^3 \\ \sum_{i=1}^{5} t_i^2 & \sum_{i=1}^{5} t_i^3 & \frac{1}{2}\sum_{i=1}^{5} t_i^4 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{5} x_i \\ \sum_{i=1}^{5} x_i \cdot t_i \\ \sum_{i=1}^{5} x_i t_i^2 \end{pmatrix} \qquad (II)$$

Solving the 3 x 3 system of linear equations described by equation (II), the values of the coefficients d_0, v_0 and g are obtained. The solution of this system can be performed using different methods; for a detailed solution and verification of the results of this example consult Perez (2021); remember that the purpose of this work is to obtain the system of linear equations from the given polynomial.

EXAMPLE 2: SOLAR IRRADIATION

The second example is related to solar irradiance.

One of the earth's energy sources is radiant energy from the sun. This radiated energy is designated solar irradiance. Solar irradiance designates the flux of electromagnetic energy emitted by the sun per unit area (Garner, 2017).

Proposing that the solar irradiance is expressed by a polynomial of degree 4 as in equation (19), find the function that best fits the data. Equation (20) presents the error, i.e., the difference between the fitted function and the data. Equation (21) presents the squared error and equation (22) presents the total error, seeking to minimize the error.

$$I_i = a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4 \qquad (19)$$

$$e_i = I_i - (a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4) \qquad (20)$$

$$e_i^2 = (-I_i + a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4)^2 \qquad (21)$$

$$e_{tot} = e^2 = \sum_{i=1}^{n}(-I_i + a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4)^2 \qquad (22)$$

To simplify obtaining the optimality of it is suggested to use equation (23).

$$f(\theta) = -I_i + a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4 \qquad (23)$$

Thus equation (23) is left as equation (*23a*),

$$e_{tot} = e^2 = \sum_{i=1}^{n} f(\theta)^2 \qquad (23a)$$

In equation (19) it can be seen that the irradiance I_i depends on θ_i but to make use of this equation it is necessary to know the coefficients (values of the coefficients *a, b, c, d* and *e*), which can be obtained by applying the CMM; for this the error in equation (22) is minimized. Now we have to obtain the values of the five coefficients (*a, b, c, d* and *e*) then the polynomial is a function of these five variables. So, to minimize the error in equation (22) we must apply the partial derivative of e_{tot} with respect to each of the now variables *a, b, c, d* and *e*, which is represented in equations 24-28.

$$\frac{\partial e_{tot}}{\partial a} = 2\sum_{i=1}^{n} f(\theta)\cdot(1) \qquad (24)$$

$$\frac{\partial e_{tot}}{\partial b} = 2\sum_{i=1}^{n} f(\theta)\cdot\theta_i \qquad (25)$$

$$\frac{\partial e_{tot}}{\partial c} = 2\sum_{i=1}^{n} f(\theta)\cdot\theta_i^2 \qquad (26)$$

$$\frac{\partial e_{tot}}{\partial d} = 2\sum_{i=1}^{n} f(\theta)\cdot\theta_i^3 \qquad (27)$$

$$\frac{\partial e_{tot}}{\partial e} = 2\sum_{i=1}^{n} f(\theta)\cdot\theta_i^4 \qquad (28)$$

After that, equations (24)-(28) are equaled to zero, and by rearranging, equations (29)-(33) are obtained.

$$2\sum_{i=1}^{n} f(\theta)\cdot 1 = 0 \rightarrow \sum_{i=1}^{n}(-Ii + a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4) = 0$$

$$\sum_{i=1}^{n} a + \sum_{i=1}^{n} b\theta_i + \sum_{i=1}^{n} c\theta_i^2 + \sum_{i=1}^{n} d\theta_i^3 + \sum_{i=1}^{n} d\theta_i^4 = \sum_{i=1}^{n} I_i$$

$$na + \sum_{i=1}^{n} b\theta_i + \sum_{i=1}^{n} c\theta_i^2 + \sum_{i=1}^{n} d\theta_i^3 + \sum_{i=1}^{n} d\theta_i^4 = \sum_{i=1}^{n} I_i \qquad (29)$$

$$2\sum_{i=1}^{n} f(\theta)\cdot\theta_i = 0 \rightarrow \sum_{i=1}^{n}(-Ii + a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4)\theta_i = 0$$

$$\sum_{i=1}^{n} a\theta_i + \sum_{i=1}^{n} b\theta_i^2 + \sum_{i=1}^{n} c\theta_i^3 + \sum_{i=1}^{n} d\theta_i^4 + \sum_{i=1}^{n} e\theta_i^5 = \sum_{i=1}^{n} I_i\,\theta_i \qquad (30)$$

$$2\sum_{i=1}^{n} f(\theta)\cdot\theta_i^2 = 0 \rightarrow \sum_{i=1}^{n}(-Ii + a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4)\theta_i^2 = 0$$

$$\sum_{i=1}^{n} a\theta_i^2 + \sum_{i=1}^{n} b\theta_i^3 + \sum_{i=1}^{n} c\theta_i^4 + \sum_{i=1}^{n} d\theta_i^5 + \sum_{i=1}^{n} e\theta_i^6 = \sum_{i=1}^{n} I_i\,\theta_i^2 \qquad (31)$$

$$2\sum_{i=1}^{n} f(\theta)\cdot\theta_i^3 = 0 \rightarrow \sum_{i=1}^{n}(-Ii + a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4)\theta_i^3 = 0$$

$$\sum_{i=1}^{n} a\theta_i^3 + \sum_{i=1}^{n} b\theta_i^4 + \sum_{i=}^{n} c\theta_i^5 + \sum_{i=1}^{n} d\theta_i^6 + \sum_{i=1}^{n} e\theta_i^7 = \sum_{i=1}^{n} I_i\,\theta_i^3 \qquad (32)$$

$$2\sum_{i=1}^{n} f(\theta)\cdot\theta_i^4 = 0 \rightarrow \sum_{i=1}^{n}(-Ii + a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4)\theta_i^4 = 0$$

$$\sum_{i=1}^{n} a\theta_i^4 + \sum_{i=1}^{n} b\theta_i^5 + \sum_{i=1}^{n} c\theta_i^6 + \sum_{i=1}^{n} d\theta_i^7 + \sum_{i=1}^{n} e\theta_i^8 = \sum_{i=1}^{n} I_i\,\theta_i^4 \qquad (33)$$

From equations (29)-(33) it can now be seen that *a, b, c, d* and *e* are the unknowns, whereas θ_i, θ_i^2, θ_i^3, θ_i^4 through θ_i^8 and I_i are known values, the latter generating constant values for each of the summations. So that equations (29)-(33) form a 5 x 5 system of linear equations, matrix equation (III).

$$\begin{pmatrix} n & \sum_{i=1}^{n}\theta_i & \sum_{i=1}^{n}\theta_i^2 & \sum_{i=1}^{n}\theta_i^3 & \sum_{i=1}^{n}\theta_i^4 \\ \sum_{i=1}^{n}\theta_i & \sum_{i=1}^{n}\theta_i^2 & \sum_{i=1}^{n}\theta_i^3 & \sum_{i=1}^{n}\theta_i^4 & \sum_{i=1}^{n}\theta_i^5 \\ \sum_{i=1}^{n}\theta_i^2 & \sum_{i=1}^{n}\theta_i^3 & \sum_{i=1}^{n}\theta_i^4 & \sum_{i=1}^{n}\theta_i^5 & \sum_{i=1}^{n}\theta_i^6 \\ \sum_{i=1}^{n}\theta_i^3 & \sum_{i=1}^{n}\theta_i^4 & \sum_{i=1}^{n}\theta_i^5 & \sum_{i=1}^{n}\theta_i^6 & \sum_{i=1}^{n}\theta_i^7 \\ \sum_{i=1}^{n}\theta_i^4 & \sum_{i=1}^{n}\theta_i^5 & \sum_{i=1}^{n}\theta_i^6 & \sum_{i=1}^{n}\theta_i^7 & \sum_{i=1}^{n}\theta_i^8 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} I_i \\ \sum_{i=1}^{n} I_i\,\theta_i \\ \sum_{i=1}^{n} I_i\,\theta_i^2 \\ \sum_{i=1}^{n} I_i\,\theta_i^3 \\ \sum_{i=1}^{n} I_i\,\theta_i^4 \end{pmatrix} \quad \text{(III)}$$

From the matrix equation (III) we have that the solution matrix is given by the matrix equation (IV). Therefore, by solving the matrix equation (IV), the values of *a, b, c, d* and *e* will be known and thus the coefficients of the polynomial of degree 4 of the solar irradiation will be obtained.

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n}\theta_i & \sum_{i=1}^{n}\theta_i^2 & \sum_{i=1}^{n}\theta_i^3 & \sum_{i=1}^{n}\theta_i^4 \\ \sum_{i=1}^{n}\theta_i & \sum_{i=1}^{n}\theta_i^2 & \sum_{i=1}^{n}\theta_i^3 & \sum_{i=1}^{n}\theta_i^4 & \sum_{i=1}^{n}\theta_i^5 \\ \sum_{i=1}^{n}\theta_i^2 & \sum_{i=1}^{n}\theta_i^3 & \sum_{i=1}^{n}\theta_i^4 & \sum_{i=1}^{n}\theta_i^5 & \sum_{i=1}^{n}\theta_i^6 \\ \sum_{i=1}^{n}\theta_i^3 & \sum_{i=1}^{n}\theta_i^4 & \sum_{i=1}^{n}\theta_i^5 & \sum_{i=1}^{n}\theta_i^6 & \sum_{i=1}^{n}\theta_i^7 \\ \sum_{i=1}^{n}\theta_i^4 & \sum_{i=1}^{n}\theta_i^5 & \sum_{i=1}^{n}\theta_i^6 & \sum_{i=1}^{n}\theta_i^7 & \sum_{i=1}^{n}\theta_i^8 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} I_i \\ \sum_{i=1}^{n} I_i\,\theta_i \\ \sum_{i=1}^{n} I_i\,\theta_i^2 \\ \sum_{i=1}^{n} I_i\,\theta_i^3 \\ \sum_{i=1}^{n} I_i\,\theta_i^4 \end{pmatrix} \quad \text{(IV)}$$

PROPOSED ALGORITHM

From the two preceding examples it can be seen that obtaining the system of linear equations to be solved from the original polynomials is more than complex, it is a laborious process with a high possibility of making a mistake. This has led to analyze the polynomial to be solved, the process for obtaining the system of equations as well as the system generated as such, leading to an algorithm.

It is recalled that the essential purpose of this article is that, from a polynomial of degree *n*, the system of linear equations generated by applying the CMM is obtained and not the values of the coefficients of the polynomial. These coefficients are obtained by solving the system of equations.

The algorithm is presented below in general terms and then applied to the two examples presented previously.

ALGORITHM

1. Express the polynomial to be treated as an equation.

i. Express the polynomial as an equation where in the first member of the equation will be all the terms of the independent variable with their respective coefficients or constants to be obtained.

ii. In the second member of the equation leave only the dependent variable.

2. Identify the number *n* of data pairs available, which were obtained by the measurements.

Each pair of data is formed by the duplicate of the i-th independent and dependent variables.

3. Obtain the value of *m*, which is the size of the system of linear equations to be generated.

The value of *m* is obtained from the degree of the polynomial to be solved; $m = q + 1$, where *q* is the degree of the polynomial in question.

4. Generate the first equation of the system of linear equations.

Apply the summation from 1 to *n* to the equation in step 1; where *n* is the number of data pairs available.

5. Generate the second equation of the system of linear equations.

For the first member of the first equation of the system of linear equations generated in step 4:

i. Multiply the first term by the summation as well as by the independent variable and then divide by *n*, the number of data pairs.

ii. Multiply from the second to the last term by the independent variable.

iii. Multiply the second member of the first equation by the independent variable.

iv. Form the new equation.

Generate the new equation with all the terms (from the first to the last) generated previously, i.e., the two members of the equation.

6. Generate the remaining equations of the system of equations.

Multiply the second equation (the two members) generated in step 5 by the independent variable. Repeat this step until the last equation of the system of equations is obtained.

7. Express the generated system of linear equations as a system of linear equations in matrix form.

APPLICATION OF THE ALGORITHM TO EXAMPLE 1

The polynomial to be worked out is expressed in equation (6), which for convenience is rewritten here.

$$d = d_0 + v_o t + \frac{1}{2} g t^2 \quad (6)$$

Obtain the system of linear equations for the polynomial of equation (6) using the proposed algorithm.

APPLICATION OF THE ALGORITHM

1. Express the polynomial as an equation.

i. Express the polynomial as an equation where in the first member of the equation are all the terms of the independent variable with their respective constants or coefficients to be obtained.

ii. In the second member of the equation will be exclusively the dependent variable.

By performing the two previous steps on equation (6), equation (*6a*) is obtained.

$$d_0 + v_o t + \frac{1}{2} g t^2 = d \quad (6a)$$

2. Identify the number n of data pairs available. Each data pair consists of the duplicate of the i-th independent and dependent variables.

In this example the number n of data pairs (d_i , t_i) is 5, i.e., $n = 5$. The value of n is not obtained from the equation or polynomial but from the number of measured data pairs. For this example the independent variable is t and the dependent variable is d.

3. Obtain the value of m, which is the size of the system of linear equations to be generated.

This value of m is obtained from the degree of the polynomial to be solved; $m = q + 1$, where q is the degree of the polynomial to be solved.

In this example the degree of the polynomial is $q = 2$, so the system of linear equations generated will be of size $m = q + 1 = 3 \rightarrow m = 3$; remember that the system of linear equations must be square; in this example it is m x m = 3 x 3.

4. Generate the first equation of the system of linear equations.

Apply the summation from 1 to n to the equation generated in step 1. In this example the equation generated in step 1 is (*6a*). So, applying the summation to it, the equation is expressed by equation (*6b*).

$$\sum_{i=1}^{n=5} \left(d_0 + v_0 t_i + \frac{1}{2} g {t_i}^2 \right) = \sum_{i=1}^{n=5} d_i \quad (6b)$$

Distributing the sum in each term of equation (*6b*) and in both members of the equation, equation (*6c*) is obtained.

$$\sum_{i=1}^{5} d_0 + v_0 \sum_{i=1}^{5} t_i + \frac{1}{2} g \sum_{i=1}^{5} t_i^2 = \sum_{i=1}^{5} d_i \quad (6c)$$

Developing the summation over the first term of the first member of equation (*6c*) gives equation (*6d*). Equation (*6d*) is the first equation of the system of linear equations.

$$5d_0 + v_0 \sum_{i=1}^{5} t_i + \frac{1}{2} g \sum_{i=1}^{5} t_i^2 = \sum_{i=1}^{5} d_i \qquad (6d)$$

5. Generate the second equation of the system of linear equations.

From the equation generated in step 4, equation (*6d*) in this example, for the first member of this equation perform the following:

i. Multiplying the first term by the summation as well as by the independent variable and then dividing by *n*, the number of pairs of data. So, considering equation (*6d*), the first term is $5d_0$, and the independent variable is t_i; applying the summation over the first term and dividing by 5, which is the value of *n*, we have that the first term of the first member of the second equation is expressed by expression (A).

$$d_0 \sum_{i=1}^{5} t_i \qquad (A)$$

ii. Multiply from the second to the last term of the first member by the independent variable. Multiplying by t_i the independent variable the above terms of the first member gives the expression (B).

$$t_i \left(v_0 \sum_{i=1}^{5} t_i + \frac{1}{2} g \sum_{i=1}^{5} t_i^2 \right) = t_i v_0 \sum_{i=1}^{5} t_i + t_i \frac{1}{2} g \sum_{i=1}^{5} t_i^2 = v_0 \sum_{i=1}^{5} t_i^2 + \frac{1}{2} g \sum_{i=1}^{5} t_i^3 \qquad (B)$$

iii. Multiplying the second member of the first equation generated in step 4 by the independent variable, we have the expression (C).

$$t_i \left(\sum_{i=1}^{5} d_i \right) = \sum_{i=1}^{5} d_i t_i \qquad (C)$$

iv. Form the second equation.

With all the terms of the first member generated above, expressions (A), (B) as well as the terms of the second member, expression (C), generate the new equation, equation (*6e*), which is the second equation of the system.

$$d_0 \sum_{i=1}^{5} t_i + v_0 \sum_{i=1}^{5} t_i^2 + \frac{1}{2} g \sum_{i=1}^{5} t_i^3 = \sum_{i=1}^{5} d_i t_i \qquad (6e)$$

6. Generate the remaining equations of the system of equations.

For this example *m* = 3 so the system of equations is formed by three equations and the first two equations have already been generated, then it only remains to generate the last equation, i.e., the third equation.

Multiply the second equation generated in step 5 by the independent variable. In this case the second equation is given by the expression (*6e*). By performing such multiplication, equation (*6e*) gives way to equation (*6f*), which is the third equation.

$$t_i \left(d_0 \sum_{i=1}^{5} t_i + v_0 \sum_{i=1}^{5} t_i^2 + \frac{1}{2} g \sum_{i=1}^{5} t_i^3 \right) = t_i \sum_{i=1}^{5} d_i t_i$$

$$d_0 \sum_{i=1}^{5} t_i^2 + v_0 \sum_{i=1}^{5} t_i^3 + \frac{1}{2} g \sum_{i=1}^{5} t_i^4 = \sum_{i=1}^{5} d_i t_i^2 \qquad (6f)$$

7. Express the generated system of linear equations as a matrix equation.

Thus equations (*6d*), (*6e*) and (*6f*) form the 3 x 3 system of linear equations, expressed in the matrix equation (*6g*).

$$\begin{pmatrix} 5 & \sum_{i=1}^{5} t_i & \frac{1}{2}\sum_{i=1}^{5} t_i^2 \\ \sum_{i=1}^{5} t_i & \sum_{i=1}^{5} t_i^2 & \frac{1}{2}\sum_{i=1}^{5} t_i^3 \\ \sum_{i=1}^{5} t_i^2 & \sum_{i=1}^{5} t_i^3 & \frac{1}{2}\sum_{i=1}^{5} t_i^4 \end{pmatrix} \begin{pmatrix} d_0 \\ v_0 \\ g \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{5} x_i \\ \sum_{i=1}^{5} x_i \cdot t_i \\ \sum_{i=1}^{5} x_i t_i^2 \end{pmatrix} \qquad (6g)$$

APPLICATION OF THE ALGORITHM TO EXAMPLE 2

In this example, from the polynomial of equation (19) we want to obtain the coefficients *a*, *b*, *c*, *d* and *e*. Recall that is the irradiance angle while I_i is the irradiance value, is the independent variable and I_i the dependent variable.

APPLICATION OF THE ALGORITHM

1. Express the polynomial as an equation.

i. Express the polynomial as an equation where in the first member of the equation will be all the terms of the independent variable with their respective constants or coefficients to be obtained.

ii. In the second member of the equation will be exclusively the dependent variable.

By performing the two previous steps on equation (19), equation (*19a*) is obtained.

$$a + b\theta_i + c\theta_i^2 + d\theta_i^3 + e\theta_i^4 = I_i \qquad (19a)$$

2. Identify the number *n* of data pairs available. Each data pair consists of the duplicate of the i-th independent and dependent variables.

In this example the number *n* of data pairs (,) is 14, i.e., *n* = 14. This value was proposed to obtain the values of the coefficients, consult Perez (2021).

3. Obtain the value of *m*, which is the size of the system of linear equations to be generated. This value of *m* is obtained from the degree of the polynomial to be solved; *m* = *q* +1, where *q* is the degree of the polynomial to work with.

In this example the degree of the polynomial is *q* = 4, so the system of linear equations generated will be of size *m* = *q* + 1 = 5→ *m* = 5; remember that the system of linear equations must be square; in this example *m* x *m* = 5 x 5.

4. Generate the first equation of the system of linear equations.

Apply the summation from 1 to *n* to the equation generated in step 1. In this example the equation generated in step 1 is (*19a*). To simplify the writing of the summations these will be expressed without the initial and final indices and also the numerical value of *n* will not be used. So, applying the summation equation (*19a*) is expressed by equation (*19b*).

$$\sum_{i=1}^{n}\left(a + b\theta_i + c\,\theta_i^2 + d\,\theta_i^3 + e\theta_i^4 = \sum_{i=1}^{n} I_i\right)$$

$$14a + b\sum\theta_i + c\sum\theta_i^2 + d\sum\theta_i^3 + e\sum\theta_i^4 = \sum I_i \qquad (19b)$$

5. Generate the second equation of the system of linear equations.

From the equation generated in step 4, equation (*19b*) in this example, for the first member of this equation perform the following:

i. Multiply the first term by the summation as well as by the independent variable and then divide by *n*. So, considering equation (*19b*), the said term is and the independent variable is ; applying the summation over the first term and dividing by *n*, we have that the first term of the first member of the second equation is expressed by the expression (*Aa*).

$$a\sum\theta_i \quad (Aa)$$

ii. Multiplying from the second to the last term of the first member by the independent variable. Multiplying by the above terms of the first member we have the expression (*Ab*).

$$b\sum\theta_i^2 + c\sum\theta_i^3 + d\sum\theta_i^4 + e\sum\theta_i^5 \quad (Ab)$$

iii. Multiply the second member of the first equation generated in step 4 by the independent variable, which is expressed by (*Ac*).

$$\sum_{i=1}^{n} I_i \theta_i \qquad (Ac)$$

iv. Form the second equation.

With all the terms of the first member generated above, expressions (*Aa*), (*Ab*) as well as the terms of the second member, expression (*Ac*), generate the new equation, equation (*19c*), which is the second equation.

$$a\sum \theta_i + b\sum \theta_i^2 + c\sum \theta_i^3 + d\sum \theta_i^4 + e\sum \theta_i^5 = \sum I_i\theta_i \qquad (19c)$$

6. Generate the remaining equations of the system of equations.

Since $q = 4$ then there are three equations left to generate; the first and second equations were already generated in the previous steps.

i. Generate the third equation.

Multiplying the second equation, (*19c*), by the independent variable gives rise to equation (*19d*), which is the third equation.

$$a\sum \theta_i^2 + b\sum \theta_i^3 + c\sum \theta_i^4 + d\sum \theta_i^5 + e\sum \theta_i^6 = \sum I_i\theta_i^2 \qquad (19d)$$

ii. Generate the fourth equation.

Multiply the third equation, (*19d*), generated in the previous step, by the independent variable. By performing such multiplication, equation (*19d*) gives way to equation (*19e*), which is the fourth equation.

$$a\sum \theta_i^3 + b\sum \theta_i^4 + c\sum \theta_i^5 + d\sum \theta_i^6 + e\sum \theta_i^7 = \sum I_i\theta_i^3 \qquad (19e)$$

iii. Generate the fifth equation.

Multiply the fourth equation, (*19e*), generated in the previous step, by the independent variable, leading to equation (*19f*), which is the fifth equation.

$$a\sum \theta_i^4 + b\sum \theta_i^5 + c\sum \theta_i^6 + d\sum \theta_i^7 + e\sum \theta_i^8 = \sum I_i\theta_i^4 \qquad (19f)$$

7. Express the generated system of linear equations as a matrix equation.

Once the five equations have been generated, the system of linear equations formed by equations (*19b*), (*19c*), (*19d*), (*19e*) and (*19f*) is available and, expressing it as a matrix equation, equation (*19g*) is obtained.

$$\begin{pmatrix} 14 & \sum_{i=1}^{14}\theta_i & \sum_{i=1}^{14}\theta_i^2 & \sum_{i=1}^{14}\theta_i^3 & \sum_{i=1}^{14}\theta_i^4 \\ \sum_{i=1}^{14}\theta_i & \sum_{i=1}^{14}\theta_i^2 & \sum_{i=1}^{14}\theta_i^3 & \sum_{i=1}^{14}\theta_i^4 & \sum_{i=1}^{14}\theta_i^5 \\ \sum_{i=1}^{14}\theta_i^2 & \sum_{i=1}^{14}\theta_i^3 & \sum_{i=1}^{14}\theta_i^4 & \sum_{i=1}^{14}\theta_i^5 & \sum_{i=1}^{14}\theta_i^6 \\ \sum_{i=1}^{14}\theta_i^3 & \sum_{i=1}^{14}\theta_i^4 & \sum_{i=1}^{14}\theta_i^5 & \sum_{i=1}^{14}\theta_i^6 & \sum_{i=1}^{14}\theta_i^7 \\ \sum_{i=1}^{14}\theta_i^4 & \sum_{i=1}^{14}\theta_i^5 & \sum_{i=1}^{14}\theta_i^6 & \sum_{i=1}^{14}\theta_i^7 & \sum_{i=1}^{14}\theta_i^8 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{14} I_i \\ \sum_{i=1}^{14} I_i\theta_i \\ \sum_{i=1}^{14} I_i\theta_i^2 \\ \sum_{i=1}^{14} I_i\theta_i^3 \\ \sum_{i=1}^{14} I_i\theta_i^4 \end{pmatrix} \qquad (19g)$$

RESULTS

Comparing matrix equations (II) and (*6g*) of example 1, it is observed that both are the same, only that matrix equation (II) was obtained using the formal CMM while the second equation, equation (*6g*), was obtained using the proposed algorithm.

Comparing the matrix equation (IV) and (*19g*) of example 2, they are also the same. The matrix equation (IV) was obtained using the formal CMM while that of equation (*19g*) was obtained with the proposed algorithm.

CONCLUSIONS

Generating the system of linear equations to obtain the coefficients of a polynomial of degree 2 and 4, following the formal CMM is too laborious and time consuming. The higher the degree of the polynomial the more laborious the procedure to obtain such a system of equations becomes.

The algorithm proposed in this paper not only works correctly for the examples presented but is also simple to use.

The algorithm does not make use of partial derivatives and the system of equations sought is obtained directly.

Given the polynomial to be treated, to generate the system of equations using the proposed algorithm only five simple things need to be identified: the coefficients of the polynomial, the degree of the polynomial, the independent variable, the dependent variable and, the number of data pairs n; from this information the generation of the system of equations is simple and straightforward.

Although the proposed algorithm presents a certain mathematical informality, it works correctly.

Work is currently underway on a formal and rigorously mathematical demonstration for the use of this algorithm for use on a polynomial of degree n.

REFERENCES

Abdi, H. et al. **New Perspectives in Partial Least Squares and Related Methods. Controlling for common method variance in PLS analysis: the measured latent marker variable approach**, 231-235, (2013).

Burden, R. L., Faires, J. D. **Análisis Numérico. Aproximación discreta de mínimos cuadrados**, 394-406, (1998).

Garner, R. **Solar irradiance.** https://www.nasa.gov/mission_pages/sdo/science/solar-irradiance.html. 06 feb 2020, (2017).

Hervé, A. **Encyclopedia of Measurement and Statistics. Method of least squares**, (2007).

Nakamura, S. **Métodos numéricos aplicados con software. Ajuste de curvas**, 274-280, (1992).

Neil, J. S. **Encyclopedia of Measurement and Statistics.**

Pérez, O. et al., (2021). **Obtención de los coeficientes de un polinomio de grado 2 y 4 por mínimos cuadrados**. Revista de la Alta Tecnología y Sociedad Vol. 13, No. 1, 2021. ISSN 1940-2171, (2007).

Samir K. **Methodes de moindres carrés**. Polycopié de cours. Anné Universitaire Pag. 58-62, (2019).