Application of least-squares fitting of ellipse and hyperbola for two dimensional data

M P Lawiyuniarti1, E Rahmadiantri1, I M Alamsyah1 and G Rachmaputri1,2

1Algebra Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha No. 10 Bandung 40132

E-mail: gantina@math.itb.ac.id

Abstract. Application of the least-square method of ellipse and hyperbola for two-dimensional data has been applied to analyze the spatial continuity of coal deposits in the mining field, by using the fitting method introduced by Fitzgibbon, Pilu, and Fisher in 1996. This method uses $4a_0a_2 - a_1^2 = 1$ as a constrain function. Meanwhile, in 1994, Gander, Golub and Strebel have introduced ellipse and hyperbola fitting methods using the singular value decomposition approach. This SVD approach can be generalized into a three-dimensional fitting. In this research we, will discuss about those two fitting methods and apply it to four data content of coal that is in the form of ash, calorific value, sulfur and thickness of seam so as to produce form of ellipse or hyperbola. In addition, we compute the error difference resulting from each method and from that calculation, we conclude that although the errors are not much different, the error of the method introduced by Fitzgibbon et al is smaller than the fitting method that introduced by Golub et al.

1. Introduction

As we know, least square fitting method is a way to get curves (functions) through a number of data points $(x_1, y_1)$ to $(x_n, y_n)$ using the polynomial function $f(x, y)$. At [2] Fitzgibbon, et al describes the method of elliptical fittings to minimize the distance to the distribution of data by using a constrain of $4a_0a_2 - a_1^2 = 1$. This fitting technique with constrain has been applied in the field of mining by Fikri et al on [1].

In Golub et al [4] the method of fitting of ellipse and hyperbola to a distribution of data has been introduced. They did not use the same restriction as before but rather using the SVD approach. Recently this approach has been generalized into three dimensional fitting of ellipsoids and hyperboloids by Rahmadiantri et al on [6].

In this research, we will discuss both methods mentioned above and furthermore, we will give the application of the calculation of both methods to the data of coal deposits contained in [1] which consists of the ash, calorific value, sulfur and steam thickness by using the two methods. In addition, we will also compute the difference errors generated from each method used.

2. Methodology

In this section, we will explain the two methods underlying the fitting technique in two-dimensional data with different functions, namely the “Direct Least Square Method” by Fitzgibbon et al in 1996 and the “SVD method” by Gander, et al in 1994.
Initially given a form of presentation of the general equation two-dimensional space as follows

\[ f(a, x) = a_0 x_1^2 + a_1 x_1 x_2 + a_2 x_2^2 + a_3 x_1 + a_4 x_2 + a_5 = a \cdot x \]

With:

\[ a = (a_0 \ a_1 \ a_2 \ a_3 \ a_4 \ a_5)^T \]

\[ x = (x_1^2 \ x_1 x_2 \ x_2^2 \ x_1 \ x_2 \ 1)^T, \]

when \( f(a, x) = 0 \) the equation is the general equation of a conic. We will determine \( a \) with Direct Least-Square Method and Singular Value Decomposition.

2.1. Direct Least-Square Method

The steps of fitting data in this method are as follow (see [2]):

Step 1. Build Matrix \( D = \begin{pmatrix} x_1^2 & x_1 x_2 & x_2^2 & x_1 & x_2 & 1 \\ x_1^2 & x_1 x_2 & x_2^2 & x_1 & x_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1}^2 & x_{m1} x_{m2} & x_{m2}^2 & x_{m1} & x_{m2} & 1 \end{pmatrix} \)

and define \( S = D^T D \).

Step 2. Use the equation \( 4a_0 a_2 - a_1^2 = 1 \) as constrain and build constrain matrix \( C \) so that \( a^T C a = 1 \).

Step 3. Use Lagrange method to change the problem minimizing \( d(a) = \sum_{i=1}^{N} f^2(a, x_i) \) with constrain \( a^T C a = 1 \) into finding \( \lambda \) and \( a \) satisfying \( S a = \lambda C a \) and \( a^T C a = 1 \).

Step 4. Breaking the equation \( S a = \lambda C a \) into:

\[ (S_{11} \ S_{12}) (a_1) = \lambda \begin{pmatrix} c_{11} & 0 \\ 0 & 0 \end{pmatrix} (a_1) \]

Step 5. Solve these equations: \( S_{11} a_1 + S_{12} a_2 = \lambda c_{11} a_1 \) and \( S_{21} a_1 + S_{22} a_2 = 0 \) by solving \((I - E) a_1 = 0\) with \( I \) is identity matriks \( 3 \times 3 \), \( E = c_{11}^{-1} (S_{11} + S_{12} S_{22}^{-1} S_{12}) \). Clearly \( \lambda \) is an eigenvalue of \( E \), in this case choose positive value. Find \( \lambda \) and corresponding eigenvector, namely \( v \).

Step 6. Choose \( k = \frac{1}{\sqrt{v^T C_1 v}} \).

Step 7. Compute \( a_1 = k v \) and \( a_2 = S_{22}^{-1} S_{21} a_1 \). Then we get \( a = (a_1, a_2) \).

2.2. Singular Value Decomposition Method

The steps of fitting data in this method as follows (see [4]):

Step 1. From data, build matrix \( S = \begin{pmatrix} x_{11} & x_{21} & 1 & x_{11}^2 & \sqrt{2} x_{11} x_{21} & x_{21}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & 1 & x_{m1}^2 & \sqrt{2} x_{m1} x_{m2} & x_{m2}^2 \end{pmatrix} \).

Step 2. Use QR fartonization to get \( S = QR \) with \( Q \) an orthogonal matrix and \( R \) an upper triangular matrix.

Step 3. Write \( R \) as \( R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix} \) and as \( \| S (v) \| = \| QR (v) \| = \| R (w) \| \), then the problem now is to solve \( \min \| R_{22} w \| \) with constrain \( \| w \| = 1 \).

Step 4. Use SVD on \( R_{22} \) to get \( R_{22} = U \Sigma V^T \).

Step 5. By Theorem SVD [3] the solution of \( \min \| R_{22} w \| \) with constrain \( \| w \| = 1 \) is \( w = v_n \) with \( v_n \) is the last column of matriks \( V \).

Step 6. Get \( v = -R_{11}^{-1} R_{12} w \). Then get \( a = (v) \).

2.3. Method for Computing Error

In this research will compute error value for each data point \((x_i, y_i)\) by using Lagrange Multiplier (see also [5]). In this case, error is minimum average distance between each data points that belongs to the ellipse or hyperbola curve. The steps to calculate the error value for each method are as follows:
Step 1. Take the equation of ellipse or hyperbola \( f(x, y) \) that we get as constrain and define \( d(x, y) = (x - x_1)^2 + (y - y_1)^2 \) with \( d^2(x, y) \) is the function that we will minimize.

Step 2. Calculate the first partial derivatives of function \( f(x, y) \) and \( d^2(x, y) \) on variable \( x \) and \( y \).

Step 3. Find \( x \) and \( y \) that satisfy: \( f(x, y) = 0 \) and \( \nabla d^2(x, y) = \lambda \nabla f(x, y) \).

Step 4. Calculate the value \( d(x, y) \) by substituting \( x \) and \( y \) value that have been obtained.

3. Result and Discussion

The data used to apply this method is the data of coal mining used in [1] consisting of ash content, calorific value, sulfur content and steam thickness for each point drill.

The conic equation is as follows:

\[
f(a, x) = a_0x^2 + a_1xy + a_2y^2 + a_3x + a_4y + a_5 = a \cdot x = 0
\]

The ash content data is as follows:

| Direction                  | X    | Y    |
|----------------------------|------|------|
| N-S (North)                | 0.00 | 105.00 |
| N-S (South)                | 0.00 | -105.00 |
| E-W (East)                 | 75.00 | 0.00 |
| E-W (West)                 | -75.00 | 0.00 |
| NE – SW (North – East)     | 68.94 | 68.94 |
| NE – SW (South – West)     | -68.94 | -68.94 |
| NW – SE (South – East)     | 42.43 | -42.43 |
| NW – SE (North - West)     | -42.43 | 42.43 |

Following the previous section:

\[
S = D^T D = \begin{bmatrix}
1.15 \cdot 10^8 & 3.87 \cdot 10^7 & 5.17 \cdot 10^7 & 0 & 0 & 24356057 \\
3.87 \cdot 10^7 & 5.17 \cdot 10^7 & 3.87 \cdot 10^7 & 0 & 0 & 59048374 \\
5.17 \cdot 10^7 & 3.87 \cdot 10^7 & 2.95 \cdot 10^8 & 0 & 0 & 35156057 \\
0 & 0 & 0 & 24356057 & 59048374 & 0 \\
0 & 0 & 0 & 59048374 & 35156057 & 0 \\
24356057 & 59048374 & 35156057 & 0 & 0 & 0
\end{bmatrix}
\]

By using Maple, we get:

\[
E = C_{11}^{-1}(S_{11} + S_{12} S_{22}^{-1} S_{21}) = \begin{bmatrix}
-2.78 \cdot 10^7 & 6.373 \cdot 10^6 & 7.01 \cdot 10^7 \\
-2.072 \cdot 10^7 & -4.73 \cdot 10^7 & -1.28 \cdot 10^7 \\
2.04 \cdot 10^7 & 1.04 \cdot 10^7 & -2.77 \cdot 10^7
\end{bmatrix},
\]

the eigenvalue \( \lambda = 2.92 \cdot 10^6 \) and \( a_1 = \begin{bmatrix}-0.78 \\ 0.42 \\ -0.38\end{bmatrix} \), \( a_2 = \begin{bmatrix}0 \\ 0 \\ 3718.06\end{bmatrix} \), so that we get

\[
a = \begin{bmatrix}-0.78 \\ 0.42 \\ -0.38 \\ 0 \\ 0 \end{bmatrix}.
\]
From that calculation we get the equation for ellipse curve as follows:

\begin{align*}
a) & \quad \frac{x^2}{a_0^2} + \frac{y^2}{a_1^2} = 1 \\
b) & \quad \frac{x^2}{b_0^2} + \frac{y^2}{b_1^2} = 1 \\
c) & \quad \frac{x^2}{c_0^2} + \frac{y^2}{c_1^2} = 1 \\
d) & \quad \frac{x^2}{d_0^2} + \frac{y^2}{d_1^2} = 1
\end{align*}

\textbf{Figure 1. Fitting of ellipse with constrain } 4a_0a_2 - a_1^2 = 1

Figure 1 describes fitting of ellipse with constrain \(4a_0a_2 - a_1^2 = 1\): (a) Fitting ash content \(f(x) = -0.78x^2 + 0.42xy - 0.38y^2 + 3718.06\), (b) Fitting calorific \(f(x) = -0.52x^2 + 0.37xy - 0.55y^2 - 0.98x - 3.13y + 3654.63\), (c) Fitting sulfur \(f(x) = -0.44x^2 + 0.87xy - 0.57y^2 + 4998.12\), and (d) Fitting steam \(f(x) = 0.78x^2 + 0.057xy + 0.32y^2 + 10679.97\).

Following the previous section and also by using Maple, we get the result as follows:

\[
S = \begin{pmatrix}
0 & 105 & 1 & 0 & 0 & 11025 \\
0 & -105 & 1 & 0 & 0 & 11025 \\
75 & 0 & 1 & 5625 & 0 & 0 \\
-75 & 0 & 1 & 5625 & 0 & 0 \\
68.94 & 68.94 & 1 & 4752.72 & 4752.72\sqrt{2} & 4752.72 \\
-68.94 & -68.94 & 1 & 4752.72 & 4752.72\sqrt{2} & 4752.72 \\
42.43 & -42.43 & 1 & 1800.31 & -1800.31\sqrt{2} & 1800.31 \\
-42.43 & 42.43 & 1 & 1800.31 & -1800.31\sqrt{2} & 1800.31
\end{pmatrix}
\]

Then the QR decomposition on the matrix \(S\) to produces the upper triangular matrix \(R\) with \(R_{22}\) as follows:

\[
R_{22} = \begin{pmatrix}
6386.55 & 3243.89\sqrt{2} & -8670.39 \\
0 & 8576.45 & 4765.55\sqrt{2} \\
0 & 0 & 4435.09 \\
0 & 0 & 0
\end{pmatrix}
\]

Next, a singular value decomposition is performed on the block matrix \(R_{22}\) and we obtain \(R_{22} = U\Sigma V^T\) with

\[
U = \begin{pmatrix}
-0.79 & -0.57 & 0.24 & 0 & 0 \\
0.53 & -0.82 & -0.20 & 0 & 0 \\
0.32 & -0.03 & 0.95 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}; \quad \Sigma = \begin{pmatrix}
12856.4 & 0 & 0 \\
0 & 10345.7 & 0 \\
0 & 0 & 1826.41
\end{pmatrix}; \quad V = \begin{pmatrix}
-0.39 & -0.35 & 0.85 \\
0.07 & -0.93 & -0.35 \\
0.92 & -0.07 & 0.39
\end{pmatrix}
\]
Next, we get $\mathbf{w} = v_3 = \begin{pmatrix} 0.85 \\ -0.35 \\ 0.39 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ -3941.81 \end{pmatrix}$. Substituting the values of the vectors $\mathbf{v}$ and $\mathbf{w}$ into $\mathbf{a}$ so we get the equation as follows:

$$f(x, y) = 0.85x^2 - 0.35x + 0.39y^2 - 3941.81$$

$$f(x, y) = -0.78x^2 + 0.42xy - 0.38y^2 + 3718.06$$

With the same steps, we get equation of ellipses for all coal deposits that give by figure as follow:

![Figure 2](image)

**Figure 2.** Fitting with SVD approach.

Figure 2 describes fitting with SVD approach: (a) Fitting ash content $f(x) = 0.85x^2 - 0.35\sqrt{2}xy + 0.39y^2 - 3941.81$, (b) Fitting calorific value $f(x) = -0.86x^2 - 0.13\sqrt{2}xy + 0.49y^2 + 1126.96$, (c) Fitting sulfur $f(x) = 0.28x^2 + 0.85\sqrt{2}xy + 0.45y^2 - 2122.91$, and (d) Fitting steam $f(x) = 0.95x^2 - 0.06\sqrt{2}xy + 0.31y^2 - 11859.68$.

By the Lagrange Multiplier method explained in previous section, we get the following comparison of error value between the two methods:

| Number | Type of Coal Content | Comparison Least Squares Method with Constraining $4a_0a_2 - a_1^2 = 1$ | Least Squares Method with SVD Approach |
|--------|----------------------|-------------------------------------------------|---------------------------------------|
| 1.     | Ash                  | 5.42                                            | 5.52                                  |
| 2.     | Calorie              | 17.95                                           | 21.24                                 |
| 3.     | Sulfur               | 17.64                                           | 28.84                                 |
| 4.     | Seam Thickness       | 13.95                                           | 16.59                                 |
By the above comparison, we conclude that although the errors are not much different, the error of the method introduced by Fitzgibbon et al is smaller than the fitting method that introduced by Golub et al.

4. Conclusion

The results of this research indicate that the application of the least squares method with constraining of $4a_0a_2 - a_1^2 = 1$ to 4 data of coal content of ash, calorific value, sulfur and seam thickness results in fitting results that are all elliptical, while the fittings using the least squares method SVD approach produces ellipse shape for ash data and seam thickness data, whereas for fitting calorific and sulfur data produce a form of hyperbola. Furthermore, the error calculation results for each method in fitting the new data yields that the errors are not much different.

Acknowledgments

This research is supported by Hibah P3MI Institut Teknologi Bandung

References

[1] Fikri H, Heriawan M N, Alamsyah I M and Rachmaputri G 2016 Analysis on The Fitting Method of Ellipsoid Anisotropy Model and Its Influence for Resource Estimation TPT XXV Association of Indonesian Mining Professionals

[2] Fitzgibbon A W, Pillu M and Fisher R B 1996 Direct Least Square Fitting of Ellipses Technical Report DAIRP-794, Departement of Artificial Intelligence, The University of Edinburgh.

[3] Gander W 2008 The Singular Value Decomposition, Lecture Notes ETH Zurich.

[4] Gander W, Golub G H and Strebel R 1994 Least Square Fitting of Circle and Ellipses 34 558-578.

[5] Lawiyuniarti M P 2017 Penggunaan Metode Kuadrat Terkecil Elips dan Hiperbola untuk Data Dua Dimensi, Master Program Thesis Institut Teknologi Bandung

[6] Rahmadiantri E, Lawiyuniarti M P, Alamsyah I M and Rachmaputri G 2017 Proc. 1st Int. Conf. on Applied and Industrial Mathematics and Statistics (Kuantan: -)