On Discrete Least Squares Polynomial Fit, Linear Spaces and Data Classification

François Dubeau and Youness Mir
Département de mathématiques, Université de Sherbrooke, 2500 Boul. de l’Université, Sherbrooke (Qc), Canada, J1K2R1

Abstract: The best discrete least squares polynomial fit to a data set is revisited. We point out some properties related to the best polynomial and precise the dimension of vector spaces encountered to solve the problem. Finally, we suggest a basic classification of data sets based on their increasing or decreasing trend, and on their convexity or concavity form.

Keywords: Polynomial data fitting, weighted least squares, orthogonal polynomials, linear spaces, data classification.

INTRODUCTION

Let \( \{(\omega_i, t_i, f_i)\}_{i=1}^{m} \) be a set of \( m \) data points where the \( t_i \)'s represent the distinct values of the independent variable, the \( f_i \)'s are the values of the measured function, and each \( \omega_i \) is the weight associated to the data \((t_i, f_i)\). The problem we consider is to find a polynomial \( p_n \) of degree at most \( n \) to fit the data. To measure how well the polynomial fit the data we use the weighted least squares deviation given by

\[
F(p_n) = \sum_{i=1}^{m} \omega_i (f_i - p_n(t_i))^2. \tag{1}
\]

The best polynomial, called the weighted least squares estimate (WLSE), is given by

\[
p_n^* = \arg\min_{p_n \in P_n} F(p_n). \tag{2}
\]

where \( P_n \) is the set of polynomials of degree at most \( n \).

The motivation for this short note comes from a mistake in the proof of Theorem 1 in [5] and explained in the Remark 2 below. The goal of this paper is to clarify the dimension of some vector spaces encountered in solving this problem, establish a property useful for proving the existence of a WLSE for exponential models [2], and suggest a way to classify data using the best polynomial fits. For a standard presentation of the theory related to best (polynomial) least squares fit see [1, 3, 7, 8, 9].

POLYNOMIAL WEIGHTED LEAST SQUARES FITTING IN \( P_n \)

In the first part of this section we present the underlying subspaces of \( P = \text{Lin}\{t^j\mid j = 0,1,2,\ldots\} \) related to the polynomial weighted least squares problem. In the second part we solve the problem using a projection onto a subspace of \( P \).

Vector spaces: Let us recall that \( P_n = \text{Lin}\{t^j\mid j = 0,1,\ldots,n\} \). We consider also the following two other polynomial subspaces

\[
PV_k = \text{Lin}\{v_i^+(t) = (t + t_i)^k \mid i = 1, m \} \subseteq P_k. \tag{3}
\]

The best polynomial fit problem can be solved by considering an orthogonal projection onto \( P_n \) or, equivalently, by considering an orthogonal projection onto a subspace of \( IR^m \). In Section 2 we briefly review the solution of the problem in \( P_n \) and specify the dimension of subspaces of polynomials. In the first part of the Section 3 we consider the subspaces of \( IR^m \) that play a role in solving the problem in \( IR^m \). In the second part of this Section 3 we solve the problem using a projection onto a subspace of \( IR^m \). Finally in Section 4 we suggest a way to classify data which will be useful in the problem of finding existence results for weighted least squares estimator [2].
Polynomial weighted least squares fitting: Under the condition that \( n < m \), we introduce the scalar product on \( P_n \) defined by

\[
\langle p, q \rangle = \sum_{i=1}^{m} \omega_i p(t_i)q(t_i)
\]

for any pair of polynomials \( p \) and \( q \) in \( P_n \). In this case (1) becomes

\[
F(p_n) = \| f - p_n \|^2
\]

where \( \| \cdot \| \) is the norm on \( P_n \) induced by the scalar product. For the \( f_i \)'s we use the notation \( f_i = f(t_i) \) \( (i = 1, \ldots, m) \). It is well known that \( p_n^* \) is unique and is characterized by the normal equations

\[
\langle f - p_n^*, p_n \rangle = 0 \quad \text{for all} \quad p_n \in P_n.
\]

In this setting, to simplify the computation of \( p_n^* \), we can find a sequence of orthogonal polynomials by applying the Gram-Schmidt orthogonalization process to the standard basis \( \{1, t, t^2, \ldots, t^n \} \) of \( P_n \).

These orthogonal polynomials are given by

\[
q_0(t) = 1, \quad q_1(t) = t - \alpha_1,
\]

and for \( j = 2, \ldots, n, \)

\[
q_j(t) = (t - \alpha_j)q_{j-1}(t) - \beta_jq_{j-2}(t)
\]

where

\[
\alpha_j = \frac{\langle q_{j-1}, q_{j-1} \rangle}{\langle q_{j-1}, q_{j-1} \rangle}, \quad (j = 1, 2, \ldots, n),
\]

and

\[
\beta_j = \frac{\langle q_{j-2}, q_{j-2} \rangle}{\langle q_{j-1}, q_{j-1} \rangle}, \quad (j = 2, 3, \ldots, n).
\]

Hence the best \( n \)-degree least squares polynomial \( p_n^* \) can be written as

\[
p_n^*(t) = \sum_{j=0}^{n} \gamma_j^* q_j(t) \quad (5)
\]

where

\[
\gamma_j^* = \frac{\langle f, q_j \rangle}{\langle q_j, q_j \rangle}, \quad (j = 0, 1, \ldots, n).
\]

The next two results will be useful for finding sufficient conditions for the existence of the WLSE for a 3-parametric exponential model [2].

**Theorem 3:** \( \langle f - p_{n-1}^*, t^n \rangle = \gamma_n^* \| q_n \|^2 \) for \( n = 0, \ldots, m-1 \).

**Proof.** For \( n = 0 \) it is obvious because \( p_{n-1}^* = 0 \). For \( n > 0 \), since \( q_n(t) = t^n + p_{n-1}(t) \) where \( p_{n-1}(t) \) is a polynomial of degree \( \leq n-1 \), and \( p_n^*(t) = \gamma_n^* q_n(t) + p_{n-1}^*(t) \), we have

\[
\gamma_n^* \| q_n \|^2 = \langle \gamma_n^* q_n, q_n \rangle = \langle p_n^* - p_{n-1}^*, q_n \rangle
\]

\[
= \langle p_n^* - f, q_n \rangle + \langle f - p_{n-1}^*, q_n \rangle = \langle f - p_{n-1}^*, t^n + p_{n-1} \rangle = \langle f - p_{n-1}^*, t^n \rangle.
\]
**Theorem 4:** If the \( q_j \)'s are the orthogonal polynomials associated to \( \{ \langle a_i, t_i \rangle \}_{i=1}^m \), the orthogonal polynomials \( \tilde{q}_j \)'s associated to \( \{ \langle a_i, \tilde{t}_i = -t_i \rangle \}_{i=1}^m \) are given by \( \tilde{q}_j(t) = (-1)^j q_j(-t) \).

**POLYNOMIAL WEIGHTED LEAST SQUARES FITTING IN IR^m**

In the first part of this section we present the underlying subspaces of \( IR^m \) related to the polynomial weighted least squares problem. In the second part we solve the problem using a projection onto a subspace of \( IR^m \).

**Vector spaces:** Let \( \{ t_i \}_{i=1}^m \) be a set of \( m \) distinct real numbers. For any positive integer \( j \) let us define the vectors \( \tilde{t}_j \in IR^m \) by

\[
\tilde{t}_j = \begin{pmatrix} t_{j1}^1 \\ t_{j1}^2 \\ \vdots \\ t_{j1}^m \end{pmatrix} \in IR^m.
\]

For any positive integer \( k \), we also define the vectors

\[
\tilde{v}_{k,i}^+ = (\tilde{t} + t_i) = \sum_{j=0}^{k} \binom{k}{j} t_j^i \tilde{t}^{k-j}
\]

for \( i = 1, \ldots, m \), and

\[
\tilde{v}_{k,i}^- = (\tilde{t} - t_i) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} t_j^i \tilde{t}^{k-j}
\]

for \( i = 1, \ldots, m \).

In this section we clarify the properties of the following vector spaces, in particular the dimension of the vector spaces,

\[
T^n = \text{Lin}\{\tilde{t}^j | j = 0, \ldots, n\} \quad \text{(6)}
\]

\[
V_k^+ = \text{Lin}\{\tilde{v}_{k,i}^+ | i = 1, \ldots, m\} \quad \text{(7)}
\]

\[
V_k^- = \text{Lin}\{\tilde{v}_{k,i}^- | i = 1, \ldots, m\} \quad \text{(8)}
\]

for any integers \( n \) and \( k \) such that \( n \geq 0 \) and \( 0 \leq k \leq m-1 \).

**Theorem 5:** Let \( T^n = \text{Lin}\{\tilde{t}^j | j = 0, \ldots, n\} \subseteq IR^m \)

(a) If \( n < m \), the set \( \{ \tilde{t}^j \}_{j=0}^n \) is linearly independent and \( \dim T^n = n+1 \).

(b) If \( n \geq m \), the set \( \{ \tilde{t}^j \}_{j=0}^n \) is linearly dependent and \( \dim T^n = m \).

**Proof:** We consider \( \sum_{j=0}^{n} \lambda_j \tilde{t}^j = 0 \). But the Vandermonde matrix \( A_{m,n} = \begin{pmatrix} t^0 & t^1 & \cdots & t^n \end{pmatrix} \) is of rank \( n+1 \) as long as \( n < m \), and hence \( \lambda_j = 0 \) for \( j = 0, \ldots, n \). If \( n \geq m \) its rank is \( m \) and there exits non zero solutions to the system. Hence the result follows because \( T^n \subseteq IR^m \).

**Remark 1:** For any positive integer \( l \), since \( \tilde{t}^{m+l} \in T^{m-1} = IR^m \), we have

\[
\tilde{t}^{m+l} = \sum_{j=0}^{m-1} \lambda_j(\tilde{t})^j,
\]

where

\[
\begin{pmatrix} \lambda_0(l) \\ \lambda_1(l) \\ \vdots \\ \lambda_{m-1}(l) \end{pmatrix} = A_{m,m}^{-1} \tilde{t}^{m+l} = A_{m,m}^{-1} \text{diag}(\tilde{t}^m) \tilde{t}^l,
\]

and

\[
\text{diag}(\tilde{t}^m) = \begin{pmatrix} t_1^m & 0 & \cdots & 0 \\ 0 & t_2^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_m^m \end{pmatrix}
\]

**Theorem 6:** Let \( k \) be any integer such that \( 0 \leq k \leq m-1 \), and let \( V_k^+ \) and \( V_k^- \) be defined by (7) and (8), then

\[
V_k^+ = T^k = V_k^-,
\]

and \( \dim V_k^+ = k+1 = \dim V_k^- \).

**Proof.** We prove the result for \( V_k^+ \) only, the proof for \( V_k^- \) is identical. Since

\[
\sum_{i=1}^{m} u_i v_{k,i}^+ = \sum_{i=1}^{m} \mu_i \left( \sum_{j=0}^{k} \binom{k}{j} t_j^i \tilde{t}^{k-j} \right) = \sum_{j=0}^{k} \left( \sum_{i=1}^{m} \mu_i t_j^i \right) \tilde{t}^{k-j},
\]
then \( \sum_{i=1}^{m} \mu_i \tilde{v}_{k,i}^+ = 0 \) if and only if \( \sum_{j=0}^{k} \binom{k}{j} \sum_{i=1}^{m} \mu_i t_i^j \tilde{v}_{k-j} = 0 \). From Theorem 5, the set \( \{ \tilde{v}_{k-j} \}_{j=0}^{k} \) is linearly independent for \( k < m \). It follows that \( \sum_{i=1}^{m} \mu_i t_i^j = 0 \) for \( j = 0, \ldots, k \). But this system of \( k+1 \) equations and \( m \) unknowns has a unique solution only for \( k = m-1 \). Moreover, the matrix associated to this system, \( A'_{m,k+1} \), is of rank \( k+1 \) for \( k < m \). Hence \( \dim V_k^+ = k+1 \).

For \( k \geq m \) we have no clear result about the dimension of \( V_k^- \) and \( V_k^+ \) as illustrated by the following example for \( m = 3 \).

**Example:** Let \( m = 3 \).

(a) For \( V_k^- \), since we have

\[
\text{Det}(v_{1,1}, v_{2,1}, v_{3,1}) = \begin{vmatrix}
0 & (t_1 - t_2)^j & (t_1 - t_3)^j \\
(t_1 - t_2)^j & 0 & (t_2 - t_3)^j \\
(t_1 - t_3)^j & (t_2 - t_3)^j & 0
\end{vmatrix}
\]

\[
= 2 \begin{cases}
1 + (-1)^j (t_1 - t_2)^j & \text{if } k \text{ is odd}, \\
0 & \text{if } k \text{ is even},
\end{cases}
\]

it follows that

\[ \dim V_k^- = \begin{cases}
2 & \text{if } k \text{ is odd}, \\
3 & \text{if } k \text{ is even}.
\end{cases} \]

(b) For \( V_k^+ \), we have

\[
\text{Det}(v_{1,1}, v_{2,1}, v_{3,1}) = \begin{vmatrix}
2(t_1 - t_2)^j & (t_1 + t_2)^j & (t_1 + t_3)^j \\
t_1 - t_2 & t_1 - t_3 & 0 \\
t_1 - t_2 & 0 & t_2 - t_3
\end{vmatrix}
\]

\[= 8(t_1 - t_2)^j (t_1 + t_3)^j + 2(t_1 + t_2)^j (t_2 - t_3)^j + (t_1 - t_2)^j (t_2 - t_3)^j.
\]

This determinant can be 0. Indeed for \( t_1 + t_2 = 0 \) and \( t_2 = 0 \) the determinant is 0 for odd \( k \). It follows that \( \dim V_k^+ = 2 \) or 3 depending on the values of \( t_1, t_2 \) and \( t_3 \).

**Remark 2:** In \( [5] \) it is asserted that \( V_k^- \) is of dimension \( m \) which is clearly false except for \( m = 3 \). As a consequence the proof given in \( [5] \) for the existence of a WLSE for a 3-parametric exponential function is not correct. There are also errors in the proof of the existence of a WLSE in \( [6] \).

**Polynomial weighted least squares fitting:** We introduce the scalar product on \( IR^m \) defined by

\[
\langle \vec{u}, \vec{v} \rangle = \sum_{i=1}^{m} \theta_i u_i v_i,
\]

for any pair of vectors \( \vec{u} \) and \( \vec{v} \) in \( IR^m \)

\[
\vec{u} = \begin{pmatrix}
u_1 \\
u_2 \\
\vdots \\
u_m
\end{pmatrix} \quad \text{and} \quad \vec{v} = \begin{pmatrix} v_1 \\
v_2 \\
\vdots \\
v_m
\end{pmatrix}.
\]

The norm on \( IR^m \) induced by the scalar product is \( ||\vec{u}|| = (\langle \vec{u}, \vec{u} \rangle)^{1/2} \). Then (1) becomes

\[ F(p_n) = ||\vec{f} - \vec{p}_n||^2, \]

where

\[ \vec{p}_n = \sum_{j=0}^{n} \alpha_j \vec{t}^j, \quad \vec{t}^j = \begin{pmatrix} t_1^j \\
t_2^j \\
t_3^j \\
\vdots \\
t_m^j
\end{pmatrix}, \quad \text{and} \quad \vec{f} = \begin{pmatrix} f_1 \\
f_2 \\
\vdots \\
f_m
\end{pmatrix} \]

The problem is to find the orthogonal projection of \( \vec{f} \) on \( T^n \). This projection is completely characterized by the normal equations \( \langle \vec{f} - \vec{p}_n, \vec{p}_n \rangle = 0 \) for all \( \vec{p}_n \in T^n \).

Again, to simplify the computation of \( \vec{p}_n^* \), we can determine an orthogonal basis \( \{ \vec{q}_j \}_{j=0}^{n} \) for \( T^n \) by applying the Gram-Schmidt process to its basis \( \{ \vec{t}^j \}_{j=0}^{n} \). We obtain

\[ \vec{q}_0 = \vec{1}, \quad \vec{q}_1 = \vec{t} - \alpha_1 \vec{1}, \]

and for \( j = 2, \ldots, n, \)

\[ \vec{q}_j = (\vec{t} - \alpha_j \vec{1}) \cdot \vec{q}_{j-1} - \beta_j \vec{q}_{j-2}, \]

where

\[ \alpha_j = \frac{\langle \vec{t} \cdot \vec{q}_{j-1} \cdot \vec{q}_{j-1} \rangle}{\langle \vec{q}_{j-1} \cdot \vec{q}_{j-1} \rangle} \quad (j = 1, 2, 3, \ldots), \]

and

\[ \beta_j = \frac{\langle \vec{t} \cdot \vec{q}_{j-1} \cdot \vec{q}_{j-2} \rangle}{\langle \vec{q}_{j-2} \cdot \vec{q}_{j-2} \rangle} \quad (j = 2, 3, 4, \ldots). \]
In these identities, \( \bar{u}, \bar{v} \) is the coordinatewise multiplication of two vectors of \( \mathbb{R}^m \) defined by
\[
\bar{u} \cdot \bar{v} = \begin{bmatrix} u_1 v_1 \\ u_2 v_2 \\ \vdots \\ u_m v_m \end{bmatrix}.
\]
Let us observe that \( \bar{q}_j \in T^j \) for \( j = 0, \ldots, n \).

It follows that the projection is given by
\[
\bar{p}_n^* = \sum_{j=0}^n \gamma_j^* \bar{q}_j,
\]
where
\[
\gamma_j^* = \frac{\langle \bar{f}, \bar{q}_j \rangle}{\langle \bar{q}_j, \bar{q}_j \rangle} (j = 0, 1, \ldots, n)
\]
The next theorem is equivalent to Theorem 2.3.

**Theorem 7:** \( \langle \bar{f} - \bar{p}^*_n, \bar{i}^n \rangle = \gamma_n^* \| \bar{q}_n \| \) for \( n = 0, \ldots, m-1 \).

**Proof.** For \( n = 0 \) we have \( \bar{p}^*_{n-1} = \bar{0} \) and the result follows. For \( n > 0 \), since \( \bar{q}_n = \bar{i}^n + \bar{p}^*_{n-1} \) where \( \bar{p}^*_{n-1} \) is a vector in \( T^{n-1} \), and \( \bar{p}_n^* = \gamma_n^* \bar{q}_n + \bar{p}^*_{n-1} \), we have
\[
\gamma_n^* \| \bar{q}_n \|^2 = \langle \bar{p}_n^* - \bar{p}^*_{n-1}, \bar{q}_n \rangle
= \langle \bar{p}_n^* - \bar{f}, \bar{q}_n \rangle + \langle \bar{f} - \bar{p}^*_{n-1}, \bar{q}_n \rangle
= \langle \bar{f} - \bar{p}^*_{n-1}, \bar{i}^n + \bar{p}^*_{n-1} \rangle
= \langle \bar{f} - \bar{p}^*_{n-1}, \bar{i}^n \rangle.
\]

**CLASSIFICATION OF DATA**

Let \( \{\omega_i, t_i, f_i\}_{i=1}^m \) be a set of \( m \) data points. If we use a discrete least squares polynomial to fit the data with the orthogonal basis \( \{\bar{q}_j\}_{j=0}^m \), the coefficients of \( \bar{p}_n^* \) with respect to its expansion (5) or (9) suggest the following classification of the data.

**Definition 1:** The data \( \{\omega_i, t_i, f_i\}_{i=1}^m \) are said to be:
(i) essentially stationary if \( \gamma_1^* = 0 \);
(ii) essentially increasing, respectively decreasing, if \( \gamma_1^* > 0 \), respectively \( \gamma_1^* < 0 \);
(iii) essentially linear if \( \gamma_2^* = 0 \);
(iv) essentially convex, respectively concave, if \( \gamma_2^* > 0 \), respectively \( \gamma_2^* < 0 \).

Let us note that we could continue the classification with the higher order coefficients \( \gamma_m^* \) for \( n = 3, \ldots, m-1 \). This basic classification could help to find more realistic or complex fitting to the data with nonlinear function (see [4, 5, 6, 2] for an exponential functions).

Finally if we apply symmetric transformations to the data we obtain the following result.

**Theorem 8:** Effect of symmetric transformations on the data.
(a) If the \( \{\omega_i, t_i, f_i\}_{i=1}^m \) are essentially increasing, resp. decreasing, then the data \( \{\omega_i, -t_i, f_i\}_{i=1}^m \) are essentially decreasing, resp. increasing. The stationarity, linearity, and concavity or convexity properties are not modified by this transform.
(b) If the data \( \{\omega_i, t_i, f_i\}_{i=1}^m \) are essentially increasing, resp. decreasing, and essentially convex, resp. concave, then the data \( \{\omega_i, -t_i, -f_i\}_{i=1}^m \) are essentially decreasing, resp. increasing, and essentially concave, resp. convex. The stationarity and linearity properties are not modified by this transform.

**CONCLUSION**

We have revisited the polynomial weighted least squares analysis. Doing so we have specified the dimension of three vector subspaces of \( P \) (Theorem 1 and Theorem 2) and of \( IR^m \) (Theorem 5 and Theorem 6) used for solving this problem. We also have established a property (Theorem 3 and Theorem 7) and suggested a classification of data (Definition 1) which will play a role in finding sufficient conditions for the existence of a WLSE for a 3-parametric exponential model [3].

**ACKNOWLEDGMENTS**

This work has been supported by a NSERC (Natural Sciences and Engineering Research Council of Canada) individual discovery grant for the first author.
REFERENCES

1. S.D. Conte and C. de Boor, 1980. Elementary Numerical Analysis: An Algorithmic Approach, McGraw-Hill, New York.

2. F. Dubeau and Y. Mir, 2007. Existence of optimal weighted least squares estimate for 3-parametric exponential model, Communications in Statistics – Theory and Methods, to appear.

3. G.E. Forsythe, 1956. Generation and use of orthogonal polynomials for data-fitting with digital computer, J. SIAM, 5, 74-78.

4. D. Jukic and R. Scitovski, 1997. Existence of optimal solution for exponential model by least squares, Journal of Computational and Applied Mathematics, 78, 317-328.

5. D. Jukic and R. Scitovski, 2000. The best least squares approximation problem for a 3-parametric exponential regression model, ANZIAM J., 42, 254-266.

6. D. Jukic, 2004 A necessary and sufficient criteria for the existence of the least squares estimate for a 3-parametric exponential function, Applied Mathematics and Computation, 147, 1-17.

7. C.L. Lawson and R.J. Hanson, 1995. Solving Least Squares Problems, SIAM, Philadelphia.

8. L.F. Shampine, 1975. Discrete least squares polynomial fits, Communications of the ACM, 18, 179-180.

9. L.F. Shampine and R.C. Allen, 1973. Numerical Computing: An introduction, Saunders, Philadelphia.