Analysis of solution of the least squares problem*

RUDOLF SCITOVSKI† AND DRAGAN JUKIĆ‡

Abstract. *For the given data (p_i, t_i, f_i), $i = 1, \ldots, m$, we consider the existence problem of the best parameter approximation of the exponential model function in the sense of ordinary least squares and total least squares. Results related to that problem which have been obtained and published by the authors so far are given in the paper, as well as some new results on nonuniqueness of the best parameter approximation.*

Key words: *exponential growth model, ordinary least squares, total least squares, existence problem*

AMS subject classifications: 65D10, 62J02

1. Introduction

We consider the parameter estimation problem for the exponential model function

$$f(t; b, c) = b\, e^{ct}, \tag{1}$$

on the basis of experimental or empirical data (p_i, t_i, f_i), $i = 1, \ldots, m$, where t_i denote the values of the independent variable, f_i the respective function values and $p_i > 0$ are the data weights. Mathematical models described by an exponential function or a linear combination of such functions are very often used in different areas of applied research, e.g. biology, chemistry, electrical engineering, economy, nuclear physics, medicine, etc. (see [4], [12], [13], [17]).

If the errors in the measurements of the independent variables are negligible, and the errors in the measurements of the dependent variable are independent random variables following the normal distribution with expectancy zero, then in practical applications the unknown parameters b and c of the function (1) are usually estimated in the sense of the ordinary least squares (OLS) (see [3], [2]):

$$\min_{(b,c)\in\mathbb{R}^2} S(b,c), \qquad S(b,c) = \tfrac{1}{2}\sum_{i=1}^{m} p_i (b e^{ct_i} - f_i)^2. \tag{2}$$

*This paper was presented at the Special Section: Appl. Math. Comput. at the 7^{th} *International Conference on OR, Rovinj, 1998.*

†Faculty of Electrical Engineering, University of Osijek, Kneza Trpimira 2B, HR-31 000 Osijek, Croatia, e-mail:scitowsk@drava.etfos.hr

‡Faculty of Food Technology, University of Osijek, F. Kuhača 18, HR-31 000 Osijek, Croatia, e-mail:jukicd@oliver.efos.hr

If considerable errors occur in the measurements of independent variables as well as dependent variables, it is reasonable to estimate the parameters b and c by minimizing all errors. Geometrically, it means that we will minimize the weighted sum of squares of the distance d_i from the data points (t_i, f_i) to the curve $t \mapsto f(t; b, c)$ (see *Fig. 1*):

$$\min_{(b,c,\boldsymbol{\delta})\in\mathbb{R}^2\times\mathbb{R}^m} F(b,c,\boldsymbol{\delta}), \qquad F(b,c,\boldsymbol{\delta}) = \tfrac{1}{2}\sum_{i=1}^{m} p_i \left[(be^{c(t_i+\delta_i)} - f_i)^2 + \delta_i^2\right], \tag{3}$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)^T \in \mathbb{R}^m$. This approach is known in literature as the total least squares (TLS) problem (see [1], [7], [16]).

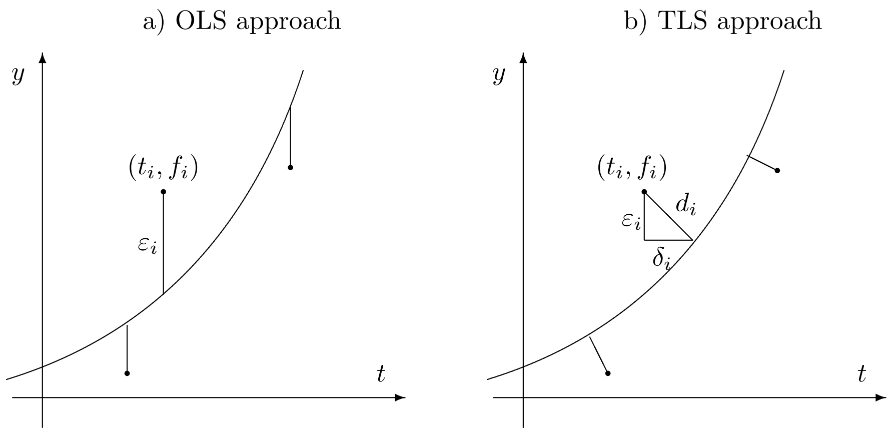

Figure 1. *Ordinary and total least squares approach*

Remark 1. *A 3-parametric exponential regression model $f(t; a, b, c) = a + b\,e^{ct}$ is very popular in applied research. For example, in [12] it is used in analysis of long–term selection experiments in biology. Also, this model is frequently used as a test function for testing numerical algorithms for function minimization. In [8] we consider the existence problem for the best least squares approximation of parameters for this model function.*

If among the data strongly deviating data can appear, so-called "outliers", then instead of the least squares approximation of parameters the so-called robust approximation can be used. In general, L_p, $p > 0$ aproximation can be considered. In [5] the total L_p-norm approximation problem for the exponential function is considered.

The following example shows that neither the OLS problem (2) nor the TLS problem (3) has always a solution.

Example 1. *Let the given data $(1, i, f_i)$, $i = 1, \ldots, m$, satisfy*

$$f_1 = f_2 = \ldots = f_{m-2} = 0, \quad f_{m-1} = -1 \quad \text{and} \quad f_m = 1.$$

If (b_n, c_n) is the sequence in $\mathbb{R}^2$, such that

$$b_n = e^{-mc_n} \quad \text{and} \quad c_n \to \infty,$$

then $S(b_n, c_n) \to \frac{1}{2}$*. It is easy to show that* $S(b,c) > \frac{1}{2}$ *for all* $(b,c) \in \mathbb{R}^2$*, which means that in this example the OLS problem (2) does not have a solution.*

In this example also the TLS problem (3) does not have a solution. Namely, $F(b_n, c_n, \mathbf{0}) \to \frac{1}{2}$*. On the other hand,* $F(b,c,\boldsymbol{\delta}) > \frac{1}{2}$ *for all* $(b,c,\boldsymbol{\delta}) \in \mathbb{R}^2 \times \mathbb{R}^m$*.*

It can be shown that both the OLS problem (2) and the TLS problem (3) have the solution, provided the data satisfy only the natural conditions:

Theorem 1. *Let the data* (p_i, t_i, f_i)*,* $i = 1, \ldots, m$*, be given, such that*

$$t_1 < t_2 < \ldots < t_m \quad \& \quad f_i > 0 \ (\forall i = 1, \ldots, m),$$

and let $\mathcal{B} = \{(b,c) \in \mathbb{R}^2 : b > 0\}$*.*

(i) Then there exists a point $(b^\star, c^\star) \in \mathcal{B}$*, at which the functional* S *defined by (2) attains the global minimum on the set* $\mathbb{R}^2$*.*

(ii) Then there exists a point $(b^\star, c^\star, \boldsymbol{\delta}^\star) \in \mathcal{B} \times \mathbb{R}^m$*, at which the functional* F *defined by (3) attains the global minimum on the set* $\mathbb{R}^2 \times \mathbb{R}^m$*.*

The proof of the statement *(i)* can be found in [7], and the proof of the statement *(ii)* can be found in [4].

The results of *Theorem 1.* will be specified further depending on the data characteristic.

2. Data classification

Under natural conditions on the data *Theorem 1.* ensures the existence of optimal parameters $b^\star, c^\star$ both in the sense of OLS and in the sense of TLS. However, it does not say anything whether the model function (1) should be sought in the class of increasing or decreasing exponential functions. It is natural to describe the increasing data by an increasing, and the decreasing data by a decreasing exponential function. For that purpose the data will be classified into three main groups and for each of them the existence problem will be analysed separately. In that way the choice of the initial approximation of the parameters will be used. Therefore, we introduce the following definition (see also [6], [10], [15]).

Definition 1. *The data* (p_i, t_i, f_i)*,* $i = 1, \ldots, m$*, are said to have the* preponderant increase *(resp.* preponderant decrease*)* property *if the slope of the associated linear regression is positive (resp. negative). If this coefficient is equal to zero, then the data are said to be* preponderantly stationary*.*

Remark 2. *The coefficients* k *and* l *of the corresponding linear regression* $\varphi(t) = k\,t + l$ *can be found by minimizing the functional*

$$G(k,l) = \tfrac{1}{2} \sum_{i=1}^{m} p_i \left(k\, t_i + l - f_i\right)^2.$$

By equating the gradient of the functional G with zero, it is easy to show that $k = \frac{D_1}{D}$*, where*

$$D = \begin{vmatrix} \sum p_i t_i^2 & \sum p_i t_i \\ \sum p_i t_i & \sum p_i \end{vmatrix}, \qquad D_1 = \sum_{i=1}^{m} p_i t_i f_i \sum_{i=1}^{m} p_i - \sum_{i=1}^{m} p_i t_i \sum_{i=1}^{m} p_i f_i.$$

Because of the Cauchy-Schwarz inequality, $D > 0$*. This means that the data* (p_i, t_i, f_i), $i = 1, \ldots, m$*, have the preponderant increase property if and only if* $D_1 > 0$*, and they have the preponderant decrease property if and only if* $D_1 < 0$*. The data are preponderantly stationary if and only if* $D_1 = 0$.

The condition of preponderant increase [resp. preponderant decrease] of data is weaker than the condition of increase [resp. decrease] of data, as stated in the following proposition (for the proof see [15]).

Proposition 1. *Let the data* (p_i, t_i, f_i), $i = 1, \ldots, m$*, be given. If*

$$(f_1 \le \ldots \le f_m) \,\&\, (f_1 < f_m),$$

then the data have the property of preponderant increase. If

$$(f_1 \ge \ldots \ge f_m) \,\&\, (f_1 > f_m),$$

then the data have the property of preponderant decrease.

Remark 3. *Note that conditions* $D_1 > 0$*, resp.* $D_1 < 0$*, correspond to well-known Chebyschev's inequalities:*

$$\sum_{i=1}^{m} p_i t_i f_i \sum_{i=1}^{m} p_i - \sum_{i=1}^{m} p_i t_i \sum_{i=1}^{m} p_i f_i > 0, \tag{4}$$

resp.

$$\sum_{i=1}^{m} p_i t_i f_i \sum_{i=1}^{m} p_i - \sum_{i=1}^{m} p_i t_i \sum_{i=1}^{m} p_i f_i < 0. \tag{5}$$

(see [11], [14], [16]).

Note also if the data are centered around the origin, then conditions (4), resp. (5), become simpler:

$$\sum_{i=1}^{m} p_i t_i f_i > 0, \qquad \textit{resp.} \qquad \sum_{i=1}^{m} p_i t_i f_i < 0, \tag{6}$$

The following theorem shows that preponderantly increasing [resp. preponderantly decreasing] data can be described by an increasing [resp. decreasing] exponential model (for the proof see [4], [15], [16]). The existence problem for preponderantly stationary data will be considered in the next section.

Theorem 2. *Let the data* (p_i, t_i, f_i), $i = 1, \ldots, m$*, be given and suppose that* $f_i > 0$, $i = 1, \ldots, m$*. Then* *(i) If the data have the property of preponderant increase (4),*

a) then there exists a pair $(b^\star, c^\star) \in \operatorname{int}\mathcal{U}$,

$$\mathcal{U} = \left\{ (b,c) \in \mathbb{R}^2 \,:\, b \geq 0, c \geq 0 \right\},$$

which minimizes on $\mathcal{U}$ *the functional* S *defined by (2).*

b) then there exists an $(m+2)$*- tuple* $(b^\star, c^\star, \boldsymbol{\delta}^\star) \in \operatorname{int}\mathcal{U} \times \mathbb{R}^m$*, which minimizes on* $\mathcal{U} \times \mathbb{R}^m$ *the functional* F *defined by (3).*

(ii) If the data have the property of preponderant decrease (5),

a) then there exists a pair $(b^\star, c^\star) \in \operatorname{int}\mathcal{V}$,

$$\mathcal{V} = \left\{ (b,c) \in \mathbb{R}^2 \,:\, b \geq 0, c \leq 0 \right\},$$

which minimizes on $\mathcal{V}$ *the functional* S *defined by (2).*

b) then there exists an $(m+2)$*- tuple* $(b^\star, c^\star, \boldsymbol{\delta}^\star) \in \operatorname{int}\mathcal{V} \times \mathbb{R}^m$*, which minimizes on* $\mathcal{V} \times \mathbb{R}^m$ *the functional* F *defined by (3).*

2.1. Preponderantly stationary data. Nonuniqueness of the best LS approximation

Theorem 1. and *Theorem 2.* assure the existence of the best least squares approximations, but do not tell us anything about either the uniqueness or about a method for finding such a best approximation. By [2] the probability that the sum of squares for any nonlinear regression has at least two local minima is positive. The following example confirms this statement.

Example 2. *Consider the data* (p_i, t_i, f_i), $i = 1, \ldots, 9$, *given by the table*

t_i	-4	-3	-2	-1	0	1	2	3	4
f_i	10	0.1	0.1	0.1	0.1	0.1	0.1	9	9

where $p_1 = \ldots = p_9 = 1$ *(see* Fig. 2.a*)*

One can easily check that the corresponding functional S *has two local minima (See Fig. 2.b):*

$$b^\star = 0.794763, \quad c^\star = 0.632571, \quad S(b^\star, c^\star) = 123.307$$
$$\hat{b} = 0.00045, \quad \hat{c} = -2.50019, \quad S(\hat{b}, \hat{c}) = 162.556$$

The graphs of the corresponding regressions are shown in Fig. 2.a.

Figure 2.a. *Exponential regressions*

Figure 2.b. *Two local minima of the functional S*

On the basis of *Definition 1.* and *Remark 2.* and *Remark 3.* it is easy to show that the data are preponderantly stationary if and only if they satisfy the condition:

$$\sum_{i=1}^{m} p_i(t_i - t_p)(f_i - f_p) = 0, \tag{7}$$

where

$$t_p := \tfrac{1}{\kappa}\sum_{i=1}^{m} p_i t_i, \qquad f_p := \tfrac{1}{\kappa}\sum_{i=1}^{m} p_i f_i, \qquad \kappa := \sum_{i=1}^{m} p_i.$$

In that case the graph of the associated linear regression is parallel with the t-axis. It directs us to the conclusion that in this case the exponential model function could degenerate into a linear function $t \mapsto b$, that is, that the local minimum of the OLS problem (2) could be attained at the point $P_S(f_p, 0)$, and the local minimum of the TLS problem (3) could be attained at the point $P_F(f_p, 0, \mathbf{0})$.

Using the condition (7) it can easily be shown that

$$\operatorname{grad} S(f_p, 0) = 0, \qquad \text{and} \qquad \operatorname{grad} F(f_p, 0, \mathbf{0}) = 0,$$

which means that $P_S(f_p,0)$ is the critical point of the functional S, and $P_F(f_p,0,\mathbf{0})$ is the critical point of the functional F. It remains to examine the positive definiteness of the Hessian $H_S(f_p,0)$ of the functional S at the point $P_S(f_p,0)$,

$$\begin{bmatrix} \sum p_i & f_p \sum p_i t_i \\ f_p \sum p_i t_i & f_p \left(2 f_p \sum p_i t_i^2 - \sum p_i t_i^2 f_i\right) \end{bmatrix},$$

resp. Hessian $H_F(f_p,0,\mathbf{0})$ of the functional F at the point $P_F(f_p,0,\mathbf{0})$,

$$\begin{bmatrix} \sum p_i & f_p \sum p_i t_i & 0 & \cdots & 0 \\ f_p \sum p_i t_i & f_p \left(2 f_p \sum p_i t_i^2 - \sum p_i t_i^2 f_i\right) & p_1 f_p (f_p - f_1) & \cdots & p_m f_p (f_p - f_m) \\ 0 & p_1 f_p (f_p - f_1) & p_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & p_m f_p (f_p - f_m) & 0 & \cdots & p_m \end{bmatrix}.$$

Let us note that $H_S(f_p,0)$ and $H_F(f_p,0,\mathbf{0})$ will be indefinite matrices if $t_p = 0$ and

$$2 f_p < \frac{\sum_{i=1}^m p_i t_i^2 f_i}{\sum_{i=1}^m p_i t_i^2}. \tag{8}$$

In this case the functional S does not have the local minimum at the point $P_S(f_p,0)$, and the functional F does not have the local minimum at the point $P_F(f_p,0,\mathbf{0})$.

Now we are going to construct the preponderantly stationary data, such that the functional S does not have the local minimum at the point $P_S(f_p,0)$, and functional F does not have the local minimum at the point $P_F(f_p,0,\mathbf{0})$. Moreover, we will show that for such defined data functionals S and F attain their global minimum at at least two points.

Example 3. *Let $T, A > 0$ and $m = 2k$, where $3 \le k \in \mathbb{N}$. Choose $\varepsilon > 0$ such that*

$$2\frac{(k-1)\varepsilon + A}{k} < \frac{\varepsilon^3 \left[\frac{1}{2}k(k-1)\right]^2 + T^2 A}{\varepsilon^2 \left[\frac{1}{2}k(k-1)\right]^2 + T^2} \tag{9}$$

and define the data:

$$(1,-T,A),\ (1,-(k-1)\varepsilon,\varepsilon), (1,-(k-2)\varepsilon,\varepsilon), \ldots, (1,-2\varepsilon,\varepsilon), (1,-\varepsilon,\varepsilon),$$
$$(1,\varepsilon,\varepsilon), (1,2\varepsilon,\varepsilon), \ldots, (1,(k-2)\varepsilon,\varepsilon), (1,(k-1)\varepsilon,\varepsilon), (1,T,A).$$

Note that these data are preponderantly stationary, $t_p = 0$ and the inequality (9) represents the inequality (8). According to the above mentioned, the functional S does not have the local minimum at the point $P_S(f_p,0)$, and the functional F does not have the local minimum at the point $P_F(f_p,0,\mathbf{0})$.

Let $(b^\star,c^\star) \in \mathcal{B}$ be a point such that $\inf_{(b,c)\in\mathbb{R}^2} S(b,c) = S(b^\star,c^\star)$ (see Theorem 1*). Since $S(b,0) \ge S(f_p,0)$ for all $b \in \mathbb{R}$ and the functional S does not have the local minimum at the point $P_S(f_p,0)$, then $c^\star \ne 0$. Furthermore, since $S(b,c) = S(b,-c)$, we have $\inf_{(b,c)\in\mathbb{R}^2} S(b,c) = S(b^\star,c^\star) = S(b^\star,-c^\star)$, i.e. the functional S attains its global minimum at at least two points.*

Similarly, because $F(b,c,\boldsymbol{\delta}) = F(b,-c,-\boldsymbol{\delta})$ *for all* $(b,c,\boldsymbol{\delta}) \in \mathbb{R}^{2+m}$, *if* $(b^\star, c^\star, \boldsymbol{\delta}^\star) \in \mathcal{B}\times\mathbb{R}^m$ *such that* $\inf\limits_{(b,c,\boldsymbol{\delta})\in\mathbb{R}^2\times\mathbb{R}^m} F(b,c,\boldsymbol{\delta}) = F(b^\star,c^\star,\boldsymbol{\delta}^\star)$, *then* $\inf\limits_{(b,c,\boldsymbol{\delta})\in\mathbb{R}^2\times\mathbb{R}^m} F(b,c,\boldsymbol{\delta}) = F(b^\star,c^\star,\boldsymbol{\delta}^\star) = F(b^\star,-c^\star,-\boldsymbol{\delta}^\star)$, *i.e. and the functional* F *attains its global minimum at at least two points.*

3. Parameter estimation

Optimal parameters of the functional S defined by (2) can be estimated by using classical methods: Gauss–Newton's method or Levenberg–Marquardt's method (see [2], [3], [20]), but that requires a good initial approximation. The problem of choice of a good initial approximation is considered in [4], [6], [15]. Localization of the region of a good initial approximation is obtained from the proof of the existence theorem.

Very few papers deal with the problem of the optimal parameter estimation by nonlinear TLS problems (see e.g. [1]). In [9] a special class of nonlinear TLS problem is considered, where the model function is of the form $f(x;a,b) = \phi^{-1}(ax+b)$, where $a \neq 0$ and b are some real parameters, and the function $\phi : I \to \mathbb{R}$, $I \subseteq \mathbb{R}$ is continuous and one-to-one. Specially, the exponential model function (1) belongs to this class.

Since the symmetry preserves the distances, instead of the TLS problem for the model function f we can consider the TLS problem for the inverse model function f^{-1}. In this way for this class of model functions we obtain a linear TLS problem, which can be solved much simpler (see [18], [19]).

For this class of model functions in the paper [9] we proved the existence theorem and proposed an efficient algorithm for searching optimal parameters.

References

[1] P. T. Boggs, R. H. Byrd, R. B. Schnabel, *A stable and efficient algorithm for nonlinear orthogonal distance regression*, SIAM J. Sci. Statist. Comput. **8**(1987), 1052–1078.

[2] E Z. Demidenko, *Optimization and Regression* (in Russian), Nauka, Moscow, 1989.

[3] J. E. Dennis, Jr, R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 1996.

[4] D. Jukić, *Parameter Identification Problem in Some Special Mathematical Models*, Ph.D. thesis, Department of Mathematics, University of Zagreb, Zagreb, 1996 (in Croatian).

[5] D. Jukić, T. Marošević, R. Scitovski, *Discrete total L_p-norm approximation problem for exponential function*, Appl. Math. Comput. **94**(1998), 137–143.

[6] D. Jukić, R. Scitovski, *Existence of optimal solution for exponential model by least squares*, J. Comput. Appl. Math. **78**(1997), 317–328.

[7] D. JUKIĆ, R. SCITOVSKI, *The exponential growth model* in: Proceedings of the 6th International Conference on Operational Research (T. Hunjak, Lj. Martić and L. Neralić, Eds.), Zagreb, 1996, 107–112.

[8] D. JUKIĆ, R. SCITOVSKI, *The best least squares approximation problem for a 3-parametric exponential regression model*, Manuscript

[9] D. JUKIĆ, R. SCITOVSKI, H. SPÄTH, *Partial linearization of one class of the nonlinear total least squares problem by using the inverse model function*, Computing, in press.

[10] D. JUKIĆ, R. SCITOVSKI, *Existence results for special nonlinear total least squares problem*, J. Math. Anal. Appl. **226**(1998) 348–363.

[11] D. S. MITRINOVIĆ, J. PEČARIĆ, A. M. FINK, *Classical and New Inequalities in Analysis*, Cluwer, Dordrecht, 1993.

[12] P. E. RUDOLPH, G. HERRENDÖRFER, *Optimal experimental design and accuracy of parameter estimation for nonlinear regression models used in long-term selection*, Biom. J. **37**(1995), 183

[13] R. SCITOVSKI, S. KOSANOVIĆ, *Rate of change in economics research*, Economics analysis and workers management **19**(1985), 65–75.

[14] R. SCITOVSKI, *Condition of preponderant increase and Tchebycheffs inequality*, in: Proceedings of VI. Conference on Applied Mathematics, (B. S. Jovanović, Ed.), Belgrade, 1988. 189–194.

[15] R. SCITOVSKI, *Some special nonlinear least squares problems*, Radovi matematički **4**(1988), 279–298.

[16] R. SCITOVSKI, D. JUKIĆ, *Total least squares problem for exponential function*, Inverse Problems **12**(1996), 341–349.

[17] R. SCITOVSKI, *Analysis of parameter identification problem*, Appl. Math. Comput. **82**(1997), 39–55.

[18] H. SPÄTH, *On discrete linear orthogonal L_p-approximation*, ZAMM **62**(1982), T354–T355.

[19] G. A. WATSON, *Numerical methods for Linear Orthogonal L_p approximation*, IMA J. Numer. Anal. **2**(1982), 275–287.

[20] S. WOLFRAM, *The Mathematica Book*, Wolfram Media, Inc., Champaign, 1998.