THE BENEFITS OF THE ORTHOGONAL LSM MODELS

Z. Mikulášek[1,2]

[1] Institute for Theoretical Physics and Astrophysics, Masaryk University
Kotlářská 2, CZ-611 37 Brno, Czech Republic, *mikulas@physics.muni.cz*

[2] Observatory and Planetarium of J. Palisa, VŠB-Technical University
Ostrava, Czech Republic

ABSTRACT. In the last few decades both the volume of high-quality observing data on variable stars and common access to them have boomed; however the standard used methods of data processing and interpretation have lagged behind this progress. The most popular method of data treatment remains for many decades Linear Regression (LR) based on the principles of Least Squares Method (LSM) or linearized LSM. Unfortunately, we have to state that the method of linear regression is not as a rule used accordingly namely in the evaluation of uncertainties of the LR parameters and estimates of the uncertainty of the LR predictions.

We present the matrix version of basic relations of LR and the true estimate of the uncertainty of the LR predictions. We define properties of the orthogonal LR models and show how to transform general LR models into orthogonal ones. We give relations for orthogonal models for common polynomial series.

Key words: variable stars, observation, data processing, LSM, linear regression, orthogonal LSM models

1. Introduction

The development in the field of variable stars research from Tsessevichs times is enormous. The number of known variable stars has arisen by at least two orders, as well as the number of their observers and interpreters. It has arisen both the volume and common access to high-quality variable stars observing data and computational techniques. The number of new efficient statistical techniques and methods that are available for everybody thanks to wide spread personal computers have been developed and published. Nevertheless, the methods used for processing of variable stars data mostly have remained the same as those used in Vladimir Platonovichs era.

Every astrophysicist likes large quantities and better quality of modern observational data, new methods of processing are not so popular. Majority of them needs a good knowledge of matrix calculus, what is in discordance with a frequent syndrome of variable stars observers, which could be named *Matrixphobia*. Very rarely we are encountering with the opposite syndrome of *Matrixphilia* which invades mathematically erudite theoreticians loving new methods and matrices so much that they do not use them for real observational data. Both extremes in the data processing are bad and we should find our golden mean.

The contemporary statistics shares inexhaustible quantity of methods. It is necessary to select several of the most versatile and diverse methods, master them and to learn to combine them. The method of processing must not be unique, but always must be made-to-measure of the set problem.

The majority of variable stars data processing tasks are solved using least square method, strictly speaking linear regression, where as models serve the most frequently common polynomials or sine/cosine series. It should be noted that there exist several other methods which are able to give the same or better results. One of them is for example the Advanced Principal Component Analysis, which is the combination of LSM and standard Principal Component Analysis (see Mikulášek, 2007). The method is optimal for solving of a lot astrophysics problems as a realistic fitting of multicolour light curves, the determination of the moments of extrema of multicolour light curves, modeling of light multicolour curves which is necessary for the process of improvement of ephemerides, diagnostics of light curve (LC) secular changes, and the classification of LCs. Other methods of modern data treatment are also mentioned in Andronov, I., these Proceedings.

In the following section we will pay attention to some details of linear regression procedure which is very likely the most frequently used tool of variable stars data processing.

2. The Least Squares Method

The very frequent astrophysical task is to fit a curve through a series of N observed points described by a triad $\{x_i, y_i, w_i\}$, where x_i is an independent (well measured) quantity like time or a phase, related to the i-th measurement y_i is the dependent, measured quantity like magnitude, O−C, and w_i is the weight of the measurement, as a rule inversely proportional to the square of the expected uncertainty of the value y_i. Hereafter we will use normalized weights w_i the mean value $\bar{w}$ of which is equal to 1.

$F(x,\vec{\beta})$ is so called *model function* of x described by the k free parameters $\beta_1,\beta_2,\ldots,\beta_k$ arranged into the vector $\vec{\beta}$. We define a function of this vector $S(\vec{\beta})$:

$$S(\vec{\beta}) = \sum_{i=1}^{N}\left[y_i - F(x_i,\vec{\beta})\right]^2 w_i. \quad (1)$$

The solution of the LSM minimalization procedure, is finding of the vector of parameters $\vec{\beta} = \mathbf{b}$, for which is the quantity $S(\vec{\beta})$ minimal. The success of the method in the given situation depends above all on our skill in the creating of the mathematical model expressed by the function $F(x,\vec{\beta})$. Then the finding of the best fit in the range of functions admissible by the pre-selected model is relatively simple and straightforward. In principle it is solution of k equations of k unknown parameters arranged in the vector $\mathbf{b}$:

$$\left.\frac{\partial S}{\partial\vec{\beta}}\right|_{\vec{\beta}=\mathbf{b}} = \mathbf{grad}\left[S(\vec{\beta}=\mathbf{b})\right] = \vec{0}. \;\Rightarrow \quad (2)$$

$$\sum_{i=1}^{N} y_i\,\frac{\partial F(x_i,\mathbf{b})}{\partial\beta_j}\,w_i = \sum_{i=1}^{N} F(x_i,\mathbf{b})\,\frac{\partial F(x_i,\mathbf{b})}{\partial\beta_j}\,w_i, \quad (3)$$

for $j = 1,2,\ldots,k$.

2.1. Linear regression

The LSM procedure of the determination of the solution will be considerably simplified if we use the linear model of the found function $F(x,\vec{\beta})$, assuming:

$$F(x,\vec{\beta}) = \sum_{j=1}^{k}\beta_j\,f_j(x), \quad (4)$$

where $f_j(x)$ are arbitrary functions of x. Eq. 1 then can be rewritten in the form:

$$S(\vec{\beta}) = \sum_{i=1}^{N}\left[y_i - \sum_{j=1}^{k}\beta_j\,f_j(x_i)\right]^2 w_i. \quad (5)$$

Eq. 3 then switches to:

$$\sum_{i=1}^{N} y_i\,f_j(x_i)\,w_i = \sum_{i=1}^{N}\left[\sum_{p=1}^{k} b_p\,f_p(x_i)\right] f_j(x_i)\,w_i, \quad (6)$$

It is advantageous to express all operations in matrix form. Then

$$\mathbf{X} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_N) & f_2(x_N) & \cdots & f_k(x_N) \end{pmatrix}, \quad (7)$$

$$\mathbf{Y} = (y_1\,y_2\,\cdots\,y_N)^{\mathrm{T}};\; \mathbf{W} = \mathbf{diag}\,(w_1\,w_2\,\cdots\,w_N), \quad (8)$$

$$\mathbf{H} = \left(\mathbf{X^T W\,X}\right)^{-1},\; \mathbf{b} = \mathbf{H\,X^T W\,Y},\; \mathbf{Y_p} = \mathbf{X\,b}, \quad (9)$$

$$R = \mathbf{Y^T W\,Y} - \mathbf{b^T X^T W\,Y},\; s = \sqrt{\frac{R}{(N-k)}}, \quad (10)$$

where $\mathbf{Y}_\mathrm{p}$ is the vector of the predictions, $\overline{w}$ is the mean value of weights w_i, R is the weighted sum of square deflections, s is the weighted standard deviation of the fit.

The procedure of linear regression with the explicit linear model is quick and its solution is unique. In the general case we may find several solutions although some of them could be physically unreal. The most common method of finding of local minima on the $S(\beta)$ plane is an iterative gradient method, where we use the above mentioned apparatus of linear regression applied on the linearized model function.

2.2. Linearized regression

The linearization of the general model function $F(x,\vec{\beta})$ consists in substitution of it by its Taylor expansion in respect of $\vec{\beta}$. We need to know as good as possible estimate $\mathbf{b_e}$ of the solution of LSM equations $\mathbf{b}$, $\mathbf{b_e}\to\mathbf{b}$. Then we can write:

$$F(x_i,\vec{\beta}) \cong F(x_i,\mathbf{b_e}) + \sum_{j=1}^{k}\frac{\partial F(x_i,\mathbf{b_e})}{\partial\beta_j}(\beta_j - b_{ej}). \quad (11)$$

$$S(\vec{\beta}) = \sum_{i=1}^{N}\left[\Delta y_i - \sum_{j=1}^{k} f_j(x_i)\,\Delta\beta_j\right]^2 w_i, \quad (12)$$

where

$$\Delta y_i = y_i - F(x_i,\mathbf{b_e}), \quad f_j(x) = \frac{\partial F(x,\mathbf{b_e})}{\partial\beta_j},$$
$$\Delta\vec{\beta} = \vec{\beta} - \mathbf{b_e}. \quad (13)$$

The equations Eq. 5 and Eq. 12 are formally identical, despite the meanings of particular terms in them are different. We define column vector $\mathbf{\Delta Y} = [\Delta y_1\,\Delta y_2\,\cdots\,y_N]$, and the column vector of the correction of the solution estimate $\mathbf{b_e}$, $\mathbf{\Delta b}$.

$$\mathbf{H} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{W\,X}\right)^{-1},\; \mathbf{\Delta b} = \mathbf{H\,X}^{\mathrm{T}}\mathbf{W\Delta Y},$$
$$R = \mathbf{\Delta Y}^{\mathrm{T}}\mathbf{W\Delta Y}, \quad s = \sqrt{\frac{R}{(N-k)}}. \quad (14)$$

Correcting $\mathbf{b_e}$ by $\mathbf{\Delta b}$ we get the next solution estimate $\mathbf{b_e}$ and we can repeat the whole procedure several times. The convergence of accordingly selected LSM model function is as a rule very swift: after a few steps we state that $\mathbf{\Delta b}\to\mathbf{0}$, hence $\mathbf{b} = \mathbf{b_e}$.

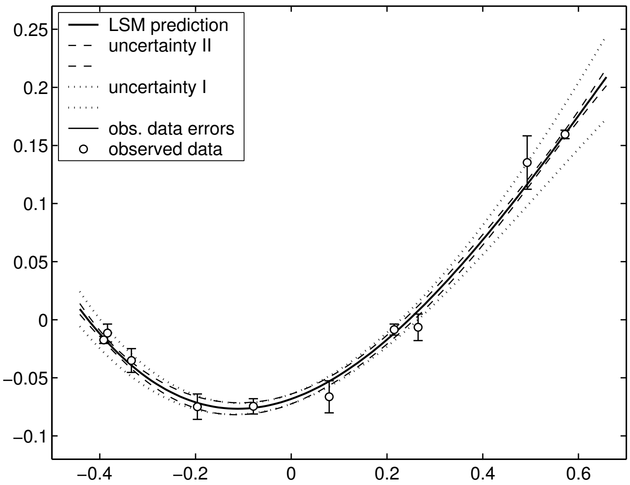

Figure 1: The illustrative figure displays the time dependence of an observed quantity measured with the accuracy denoted by the abscissa. The continuous line represents LSM fit by the polynomial of the 3-rd order (cubic parabola). Expected uncertainties of this prediction calculated by the formula Eq. 16 are denoted by dotted lines, true uncertainties given by Eq. 17 are signed by dashed lines.

2.3. Uncertainties of parameters and prediction

There are at least three reasons why we should estimate the measure of uncertainty of the found parameters. Firstly, errors of parameters tell us a lot about the reliability of our results, secondly uncertainties of parameters would enable to calculate the uncertainty of the prediction done on the basis of our LSM analysis, and last but not least above mentioned errors are strictly demanded by teachers, scientific editors and referees. All LSM instructions and codes congruently get for uncertainty of the j-the parameter δb_j the following relation:

$$\delta b_j = s\sqrt{H_{jj}}, \tag{15}$$

where H_{jj} is the j-th element in the diagonal of the matrix $\mathbf{H}$.

It is a question whether δb_j really expresses the uncertainty in the common sense. The response is no, strictly speaking sometimes yes, but very rarely. It can be demonstrated on the error of the absolute term in the LSM fit by straight line, which evidently depends on the choice of the origin of x coordinate.

The suspicion that there is something incorrect in our comprehension of the true meaning of the quantity δb_j defined by Eq. 15 will be supported by our attempt use these errors for the evaluating of the expected uncertainty of the prediction by the model function for the arbitrarily selected value of x:

$$\delta y_p(x) = \sqrt{\sum_{j=1}^{k} \delta^2 b_j\, f_j^2(x)} = \sqrt{\mathbf{g}(x)\, \mathbf{H}_{\rm dg}\, \mathbf{g}^{\rm T}(x)}, \tag{16}$$

where $\mathbf{H}_{\rm dg}$ equals to the matrix $\mathbf{H}$, whose all non-diagonal elements has been put zero. $\mathbf{g}(x)$ is the row vector of the gradient of the solution model function $\mathbf{F}(x, \mathbf{b})$, $\mathbf{g}(x) = [f_1(x)\, f_2(x)\, \ldots\, f_k(x)]$

The instructive picture Fig. 1 will show you that this intuitive relation gives quite inadequate results. Nevertheless, it can be shown that it is valid formally rather similar relation:

$$\delta y_p(x) = \sqrt{\mathbf{g}(x)\, \mathbf{H}\, \mathbf{g}^{\rm T}(x)}. \tag{17}$$

The matrix $\mathbf{H}$ is by the definition (see Eq. 9 and 14) a symmetric square $k \times k$ matrix with $k(k+1)/2$ independent elements. If we want to enable to anybody to compute the uncertainty of the prediction, we should publish either the whole matrix $\mathbf{H}$ or its non-trivial part at least. Nevertheless, there is another (more illustrative) possibility: to transform the model function into the orthogonal one. Then the matrix $\mathbf{H}$ will change in the diagonal one and the uncertainties of parameters will acquire its standard meaning. It will help you among other things expertly examine importance of individual terms.

3. Orthogonal LSM models

Let us assume that the functional dependence of observed quantities y on x is well described by the model function which can be expressed in the form of the linear combination of k basic functions of $f_j(x)$ with coefficients b_j. The found solution does not change if we use another set of k functions $\vartheta_j(x)$, which are created as linear combinations of the basic functions $f_j(x)$. Let us combine them so that the new set of basic functions $\vartheta_j(x)$ is orthogonal. It means we find the set of coefficients $\{a_{pj}\}$:

$$\vartheta_p(x) = \sum_{j=1}^{k} a_{pj}\, f_j(x), \quad \text{so that,} \tag{18}$$

$$\overline{\vartheta_p\, \vartheta_q} = \sum_{i=1}^{N} \vartheta_p(x_i)\, \vartheta_q(x_i)\, w_i = 0 \quad \text{if } p \neq q \tag{19}$$

The calculation of linear regression parameters and their uncertainties is then very simple:

$$b_j = \frac{\sum_{i=1}^{N} y_i\, \vartheta_j(x_i)\, w_i}{\sum_{i=1}^{N} \vartheta_j^2(x_i)\, w_i}; \quad \delta b_j = \frac{s}{\sqrt{\sum_{i=1}^{N} \vartheta_j^2(x_i)\, w_i}};$$

$$\delta y_p(x) = \sqrt{\sum_{j=1}^{k} \delta^2 b_j\, \vartheta_j^2(x)}. \tag{20}$$

The set of coefficients $\{a_{pj}\}$ fulfilling constraints Eq. 19 is not unique as well as the procedures of its finding. We recommend to use the following procedure which seems to us the simplest one:

$$\vartheta_1 = f_1; \quad \vartheta_2 = f_2 - a_{21}\vartheta_1;$$
$$\vartheta_3 = f_3 - a_{32}\vartheta_2 - a_{31}\vartheta_1;$$
$$\vartheta_p(x) = f_p(x) - \sum_{q=1}^{p-1} a_{pq}\,\vartheta_q(x), \tag{21}$$

where

$$a_{pq} = \frac{\overline{f_p\,\vartheta_q}}{\overline{\vartheta_q^2}} = \frac{\sum_{i=1}^N f_p(x_i)\,\vartheta_q(x_i)\,w_i}{\sum_{i=1}^N \vartheta_q^2(x_i)\,w_i}. \tag{22}$$

The first three orthogonalized terms will be:

$$\vartheta_1(x) = f_1(x); \quad \vartheta_2(x) = f_2(x) - \frac{\overline{f_2 f_1}}{\overline{f_1^2}};$$
$$\vartheta_3(x) = f_3(x) - \frac{\overline{f_3 f_2} - \overline{f_3}\;\overline{f_2}}{\overline{f_2^2} - \overline{f_2}^2}\, f_2(x) -$$
$$-\left[\frac{\overline{f_3 f_1}}{\overline{f_1^2}} - \frac{\overline{f_2 f_1}\left(\overline{f_3 f_2} - \overline{f_3}\;\overline{f_2}\right)}{\overline{f_1^2}\left(\overline{f_2^2} - \overline{f_2}^2\right)}\right] f_1(x). \tag{23}$$

The explicit expression of successive terms of a set of the orthogonalized functions is more and more complex, however it is not very complicated to write an iterative PC code enabling to compute the formulae for arbitrary number of parameters.

3.1. Orthogonal polynomial model

The most popular linear regression model (not only in astrophysics) $F(x,\vec{\beta})$ is:

$$F(x,\vec{\beta}) = \sum_{j=1}^{k} \beta_j\, x^{j-1}. \tag{24}$$

The model is known to have a lot uncomfortable properties which complicate both the calculation and the interpretation of found results. We should never used it without orthogonalization.

We recommend to put the origin of x-coordinates into the center of gravity of observations: $x \;\to\; x - \bar{x}$ before the application of the orthogonalization procedure. It will result in the considerable simplification in the form of regression model. Assuming now $\bar{x} = 0$ the first four orthogonal polynomials are as follows:

$$\vartheta_1(x) = 1;\; \vartheta_2(x) = x;\; \vartheta_3(x) = x^2 - \frac{\overline{x^3}}{\overline{x^2}}\, x - \overline{x^2},$$
$$\vartheta_4(x) = x^3 - \frac{\overline{x^2}^2\,\overline{x^3} + \overline{x^3}\,\overline{x^4} - \overline{x^2}\,\overline{x^5}}{\overline{x^2}^3 + \overline{x^3}^2 - \overline{x^2}\,\overline{x^4}}\, x^2 -$$
$$\frac{\overline{x^3}\,\overline{x^5} + \overline{x^2}^2\,\overline{x^4} - \overline{x^4}^2 - \overline{x^3}^2\,\overline{x^2}}{\overline{x^2}^3 + \overline{x^3}^2 - \overline{x^2}\,\overline{x^4}}\, x - \frac{\overline{x^2}^2\,\overline{x^5} + \overline{x^3}^3 - 2\overline{x^3}\,\overline{x^4}}{\overline{x^2}^3 + \overline{x^3}^2 - \overline{x^2}\,\overline{x^4}},$$

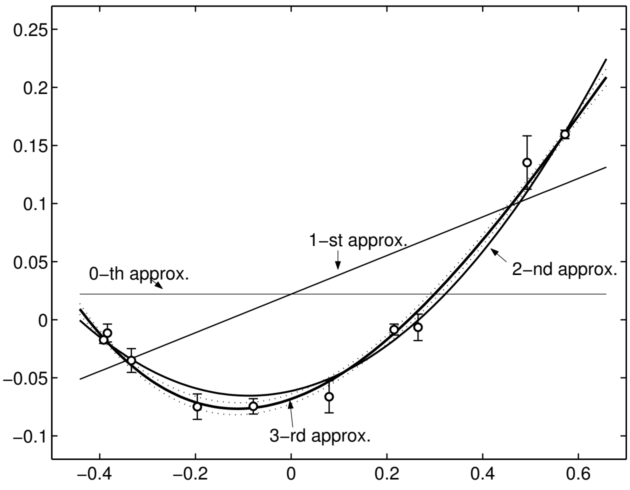

Figure 2: The subsequent approximations of the fit of observed data by orthogonal polynomial regression.

where,

$$\overline{x^p} = \frac{\sum_{i=1}^N x_i^p\, w_i}{\sum_{i=1}^N w_i}. \tag{25}$$

Fig. 2 displays the results of subsequent fitting of the model situation by constant, linear, quadratic and cubic orthogonal polynomials.

If the data are distributed uniformly in the interval $x_i \in \langle -\Delta;\, \Delta\rangle$, we can use the transformed Legendre polynomials (orthogonal on the interval $\langle -1;\, 1\rangle$) as the orthogonal (or quasiorthogonal) LSM model:

$$\vartheta_1 = 1;\; \vartheta_2 = x;\; \vartheta_3 = x^2 - \frac{\Delta^2}{3};\; \vartheta_4 = x^3 - \frac{3\,\Delta^2}{5}\,x;$$
$$\vartheta_5 = x^4 - \frac{6\,\Delta^2}{7}\,x^2 + \frac{3\,\Delta^4}{35};\; \cdots \tag{26}$$

3.2. Orthogonal sine, cosine model

The basic tool for the analysis of cyclic and periodic processes in astrophysics is the linear regression with the model consisting of simple periodic functions, the most commonly:

$$F(\varphi,\vec{\beta}) = \beta_1 + \sum_{j=1}^{q} \beta_{2j}\cos(2\pi j\varphi) + \beta_{2j+1}\sin(2\pi j\varphi), \tag{27}$$

where φ is the phase as an independent variable, q is the order of set of harmonic functions. The model need not contain all of functions, it might be limited e.g. only to even functions etc.

In the case that the observations are spread over the whole cycle more or less uniformly, it is not needed to do any orthogonalization, because all functions are orthogonal itself. In the opposite case we should do orthogonalization e.g. by the procedure described by Eq. 21 and Eq. 22.

4. Conclusions

We displayed the benefits of consequential usage of orthogonal LSM model functions with the emphasis on the polynomial regression as the chief tool of astrophysical data processing. Orthogonal models enable to give the true sense to errors of found parameters and easily compute estimates for uncertainties of the prediction. The orthogonality of the models removes the bad conditioning of the solved systems of equations and help us to obtain results not affected by computational errors. We recommend to use them always, compulsorily in the case of polynomial regression.

It is demanding to use new methods of variable stars data processing which enable us better exploit information hidden in observations. Endeavor connected with mastering of them will return in new subtle discoveries and revealing.

Matrix calculus, true using of weights, advanced principal component analysis, factor analysis, robust regression, creation and usage of orthogonal models and several other processing techniques should appertain to compulsory outfit of each variable stars observer of the 21st century.

Acknowledgements. This work was supported by grants GA ČR 205/06/0217, and MVTS ČRSR 10/15. The author is indebted to prof. Izold Pustylnik and Dr. Miloslav Zejda for careful and critical reading of the manuscript and suggestions which considerable improved the article.

References

Mikulášek Z.: 2007, *Astron. Astrophys. Trans.,* **26**, 63.