ISTITUTO
GEOGRAFICO
MILITARE
FIRENZE

Bollettino di Geodesia e Scienze Affini

Unbiasedness vs. Unboundedness: an alternative perspective on the principle of least-squares estimation

CHRISTOPHER KOTSAKIS

ESTRATTO DAL «*BOLLETTINO DI GEODESIA E SCIENZE AFFINI*»
RIVISTA DELL'ISTITUTO GEOGRAFICO MILITARE
ANNO LXIV - N. 3 - LUGLIO-AGOSTO-SETTEMBRE 2005

ANNO LXIV – BOLLETTINO DI GEODESIA E SCIENZE AFFINI – N. 3, 2005

Unbiasedness vs. Unboundedness: an alternative perspective on the principle of least-squares estimation

CHRISTOPHER KOTSAKIS
Department of Geodesy and Surveying, Aristotle University of Thessaloniki, Greece

Summary. – An alternative route for developing the logic of least-squares estimation is presented in this paper. In particular, the standard property of unbiasedness for the least-squares estimators is replaced with a different, yet equivalent, constraint. It is shown that the exact same optimal results are obtained when we switch the a priori requirement of having unbiased estimates with a condition which implies that the numerical range of the unknown parameters is unbounded. The theoretical and practical consequences of this strange dualism are discussed and some critique on the logical foundations of the classic least-squares method is also made.

STIMA NON DEVIATA E STIMA NON LIMITATA: UN PUNTO DI VISTA ALTERNATIVO SUL PRINCIPIO DELLA STIMA AI MINIMI QUADRATI.

Sommario. – L'articolo presenta una strada alternativa di sviluppo della logica della stima ai minimi quadrati. Si dimostra che si ottengono gli stessi esatti risultati ottimali quando si sostituisce il requisito a priori di stime non deviate con una condizione che implica la non limitatezza dell'escursione numerica dei parametri incogniti. Vengono discusse le conseguenze teoriche e pratiche di questo singolare dualismo e vengono inoltre esaminate da un punto di vista critico i fondamenti logici del metodo classico ai minimi quadrati.

Keywords: Least-squares, unbiased estimation, unbounded parameters
Parole Chiave: minimi quadrati, parametri non limitati, stima non deviata.

1. – INTRODUCTION

The method of least-squares (LS) has a history that spans almost two centuries of human intellectual work. Having been conceived in the nineteenth century when new and revolutionary ideas were systematically emerging, least-squares theory marks the starting point of modern data analysis for the applied sciences. Over the years, LS methods managed to acquire a status of scientific universality and they still provide the first tool that comes to mind when one deals with optimal prediction or estimation problems (this is particularly true in the geodetic world). Unfortunately, the prevailing use of LS methods has not been complemented by a widespread common understanding of their rudiments. Although there is complete agreement on how to form the 'normal equations' from the 'observation equations' and everyone can obtain the very same values for their solution, the reasons for employing LS-based estimation techniques, the perception of their objectives and the conditions under which these are achieved, as well as the interpretation of their final results, may be quite different among researchers. This somewhat controversial situation is probably due to the fact that the LS method was originally developed from three distinctively separate viewpoints: (i) *least sum of squared residuals* (Legendre, 1805), (ii) *maximum probability of zero error of estimation* (Gauss, 1809), and (iii) *least mean squared error of estimation* (Gauss, 1821, 1823, 1826); see also Plackett (1972). The previous realization, however, seems to be of little concern in today's pragmatic world where often practical results overshadow theoretical questions. Nevertheless, the aim of this paper is to present an alternative view of the basic theoretical principles hidden behind LS estimation. As the title of the paper unveils, the focus is put on switching the well known property of unbiasedness for the LS estimators with a different, yet equivalent, constraint. In particular, it will be shown that the same LS-type estimators can be obtained if we replace the a priori requirement of unbiasedness with a condition which implies that the numerical range of the unknown parameters is unbounded. It is hoped that such an alternative formulation can contribute to a stronger comprehension of the LS estimation process and its possible weaknesses.

2. – BEST LINEAR UNBIASED ESTIMATION (BLUE) – THE USUAL APPROACH FOR LEAST-SQUARES PROBLEMS

In the context of optimal statistical estimation, the LS methodology provides a linear and uniformly unbiased estimator that has minimum mean squared error among any other linear unbiased estimator. This is the standard perspective that is typically followed to describe LS estimation techniques, and it has been the basis upon which their choice for the solution of practical problems is usually justified. In a historical context, this probabilistic viewpoint is due to Gauss's second formulation for the LS method. The rigorous link between this approach, which will be identified thereafter by the acronym BLUE, and Legendre's (1805) original deterministic conception is provided by the well known *Gauss-Markov theorem* (see, e.g., Dermanis and Rummel, 2000, pp. 48-49). In this section, a short exposition of the BLUE version for the LS estimation procedure is given. Although our presentation does not follow the most general setting, it is nevertheless sufficient for the purpose of this paper.

A system of linear(-ized) observation equations is given as follows:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v} \tag{1}$$

where $\boldsymbol{y}$ is a known observation vector, $\boldsymbol{x}$ is an unknown parameter vector and $\boldsymbol{A}$ is a design matrix of known coefficients with full column rank. The residual vector $\boldsymbol{v}$ contains unknown random errors (data noise) whose statistical characteristics are typically given in terms of their first and second order moments, i.e.

$$E\{\boldsymbol{v}\} = \boldsymbol{0}, \qquad E\{\boldsymbol{v}\boldsymbol{v}\}^{\mathrm{T}} = \boldsymbol{C} \tag{2}$$

In practice, the error covariance (CV) matrix $\boldsymbol{C}$ is often considered partially known and its uncertainty is controlled by one or more unknown scaling factors (variance components). Since the knowledge of the error CV matrix does not play a crucial role in the rest of this paper, we assume that $\boldsymbol{C}$ is a fully known symmetric and positive-definite matrix.

The model of eq. (1) is suitable for the study of a variety of physical systems, including most areas of modern geodetic research. In principle, in all such cases we generally seek to estimate an unknown quantity θ which depends, directly or indirectly, on the parameter vector $\boldsymbol{x}$. For convenience, we consider only the case where θ is a linear function of the unknown parameters

$$\theta = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x} \tag{3}$$

with $\boldsymbol{q}$ being an arbitrary known vector. A general *linear estimator* of θ will have the form

$$\hat{\theta} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} + c \tag{4}$$

where the vector $\boldsymbol{b}$ and the scalar c need to be determined according to some optimality criteria.

The classic statistical formulation of LS estimation is based on two fundamental properties that should be satisfied simultaneously by the linear estimator of eq. (4), namely

(i) *Uniform unbiasedness* – $E\{\hat{\theta}\} = \theta = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x}$, for any parameter vector $\boldsymbol{x}$

and

(ii) *Minimum mean squared error* – $E\{(\hat{\theta} - \theta)^2\}$ = minimum

It is easily shown that the first property leads to the following constraints for $\boldsymbol{b}$ and c

$$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{q}^{\mathrm{T}} \tag{5}$$

$$c = 0 \tag{6}$$

Using eqs. (3) and (4) we can also establish that the mean squared error (MSE) of the linear estimator $\hat{\theta}$ has the general form

$$E\{(\hat{\theta}-\theta)^2\} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{b} + [(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A} - \boldsymbol{q}^{\mathrm{T}})\boldsymbol{x} + c]^2. \tag{7}$$

The minimization of the above quantity, in conjunction with the linear constraints of eqs. (5) and (6), lead to a unique optimal solution for $\boldsymbol{b}$ through the method of Lagrange multipliers. The result is given by the following equation

$$\boldsymbol{b} = \boldsymbol{C}^{-1}\boldsymbol{A}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A})^{-1}\boldsymbol{q} \tag{8}$$

Based on eqs. (4), (6) and (8), the LS estimate of $\theta = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x}$ is thus given by the well known expression

$$\hat{\theta} = \boldsymbol{q}^{\mathrm{T}}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{y} \tag{9}$$

which, in turn, implies the following LS estimate for the parameter vector $\boldsymbol{x}$

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{y} \tag{10}$$

Remark. Normally, the statistical optimality of the LS method is attributed to the fact that it provides minimum *error variance* among all other linear unbiased estimation algorithms. However, since LS estimators are unbiased, the variance and the mean squared value of their estimation error are exactly equal. Therefore, the BLUE formulation can be equivalently based either on the minimization of the *error variance* or on the minimization of the *mean squared error* of a linear unbiased estimator. Here we have chosen to follow the latter approach since it provides a more direct connection with the discussion given in the following section.

3. – LINEAR ESTIMATION WITH MINIMUM MEAN SQUARED ERROR FOR UNBOUNDED PARAMETERS – AN EQUIVALENT VIEW OF THE LEAST-SQUARES PRINCIPLE

The LS estimators given in eqs. (9) and (10) of the previous section can be obtained through a different formulation, without departing from the broad context of optimal statistical estimation. The alternative approach that is presented here represents only an attempt to explain the logic of the unbiasedness condition which is associated with LS estimators.

Keeping on the same setting of the last section and starting again from a typical linear estimator $\hat{\theta} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} + c$ for an unknown scalar quantity $\theta = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x}$, we seek optimal values for $\boldsymbol{b}$ and c. As it was mentioned already, the mean squared estimation error in such a case has the general form

$$E\{(\hat{\theta}-\theta)^2\} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{b} + [(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A} - \boldsymbol{q}^{\mathrm{T}})\boldsymbol{x} + c]^2 .$$

Let us point out the fact that the MSE of $\hat{\theta}$ depends, in general, on the vector of the original unknown parameters. Now, *if the range of* $\boldsymbol{x}$ *is unbounded*, the second term in the above expression becomes also unbounded. In order to ensure that the MSE of the linear estimate $\hat{\theta}$ remains always finite, regardless of the numerical range of the unknown parameters, the following condition should be satisfied

$$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A} - \boldsymbol{q}^{\mathrm{T}} = \boldsymbol{0}^{\mathrm{T}} \tag{11}$$

where $\boldsymbol{0}^{\mathrm{T}}$ corresponds to a row vector of zeros. Subject to this condition and given the fact that c is only a constant scalar, the MSE minimization for the linear estimator $\hat{\theta}$ yields the result

$$\boldsymbol{b} = \boldsymbol{C}^{-1}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A})^{-1}\boldsymbol{q} \quad \text{and} \quad c = 0 \tag{12}$$

which, in turn, gives rise to the same LS estimates for θ and $\boldsymbol{x}$ that were derived in the previous section.

It is thus seen that we are able to obtain the same optimal estimators as in the BLUE case, without invoking a priori the requirement of unbiased estimates for the unknown parameters. Hence, an equivalent formulation of the LS estimation process can emerge which is articulated in the following words: *in the class of all linear estimators with finite mean squared error for a set of unknown parameters with unbounded range, least-squares estimators provide results with minimum mean squared error.*

Under the preceding perspective it may appear that we have removed the requirement of unbiasedness at the expense of a more «restricted» version for the LS method. Obviously, the property of unbiasedness for the optimal estimators has not been lost in this case since it will now be a direct consequence of eq. (12). On the other hand, the resulting estimators are not restrictive in any way because they can always be implemented, regardless of the actual range of the parameter vector $\boldsymbol{x}$ and/or the values of the data vector $\boldsymbol{y}$. In fact, what the previous alternative formulation should make us skeptical about is the following question: *will LS estimation give optimal results, in the MSE sense, when $\boldsymbol{x}$ is a vector of bounded parameters?*

4. – DISCUSSION - CONCLUSIONS

An instructive way to look at LS estimation is to recognize the fact that its statistical optimality is associated with the inherent assumption that the range of the unknown parameter vector $\boldsymbol{x}$ is unbounded. Clearly, in all application areas where LS techniques are used the values of the parameters that need to be estimated always lie within a finite range. Nevertheless, what should be acknowledged here is that this important piece of *information* (or even a *fact* for most physical systems under study) is not integrated at all in the ordinary LS estimation process. The logic of the LS principle, which we routinely use in almost every geodetic estimation or adjustment problem, ignores the fact that the unknown parameters (e.g., network coordinates, harmonic potential coefficients, orbital parameters, etc.) have always finite magnitude. That is probably the reason why LS solutions tend to give numerical answers that are usually «longer» (when measured by some Euclidean-type norm) than the actual true parameter vector.

An interesting conclusion that can be drawn from this paper is that the property of unbiasedness is responsible for causing the classic LS (or equivalently BLUE) estimators to be blind on the bounded nature of the unknown parameters. This rather strange dualism brings up a fairly strong argument in favor of biased estimation methods (Mayer and Willke, 1973). In fact, it is well known that if we assume an upper bound for the Euclidean length of the unknown parameter vector and then seek the LS estimator $\hat{\boldsymbol{x}}$ (in the Legendre's sense - $(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}})^{\mathrm{T}} \boldsymbol{C}^{-1} (\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}) =$ *minimum*) subject to this restriction, the final result will be similar to a *ridge regression* solution (Draper and Smith, 1998; Björck, 1996). Such a solution corresponds to a popular biased estimation technique which is known to outperform LS estimators, in the MSE sense, when $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ is bounded; see *theorem 7* in Marquardt (1970) and Hoerl and Kennard (1970). When viewed in this manner it is difficult to argue against the use of ridge regression or other biased estimation methods that confine reasonably the size of the unknown parameters, although in reality $\boldsymbol{x}$ is always unknown.

As a final remark, it should be noted that in geodesy we have often been guided by a non-enthusiastic (if not negative) attitude towards the use of optimal biased estimation techniques. Without trying to diminish the value of LS methods, and keeping also in mind the practical difficulties that are often associated with the implementation of biased estimation algorithms, what is claimed in this paper is that the basic statistical principles associated with LS estimation should receive a more objective and careful treatment than is usually given to them.

REFERENCES

A. Björck (1996), *Numerical methods for least squares problems.* Society for industrial and applied mathematics (SIAM), Philadelphia.

A. Dermanis, R. Rummel (2000), *Data analysis methods in geodesy.* In: A. Dermanis, A. Gruen and F. Sanso (eds.) Geomatic Methods for the Analysis of Data in Earth Sciences. Lecture Notes in Earth Sciences Series, vol. 95, pp. 17-92, Springer Verlag, Berlin Heidelberg.

N.R. Draper, H. Smith (1981), *Applied regression analysis.* 2nd edition, Wiley, New York.

C.F. Gauss (1809), *Theoria motus corporum coelestium in sectionibus conicis solem ambientium.* Perthes and Besser, Hamburg. English translation (originally in 1857 by C.H. Davis) reprinted as Theory of the Motions of the Heavenly Bodies Moving about the Sun in Conic Sections, Dover, New York, 1963.

C.F. Gauss (1821) *Theoria combinationis observationum erroribus minimis obnoxiae: Pars prior.* Göttingische gelehrte Anzeigen, no. 33, pp. 321-327.

C.F. Gauss (1823) *Theoria combinationis observationum erroribus minimis obnoxiae: Pars posterior.* Göttingische gelehrte Anzeigen, no. 52, pp. 313-318.

C.F. Gauss (1826) *Supplementum theoriae combinationis observationum erroribus minimis obnoxiae.* Göttingische gelehrte Anzeigen, no. 153, pp. 1521-1527.

A.E. Hoerl, R.W. Kennard (1970), *Ridge regression: biased estimation for nonorthogonal problems.* Technometrics, no. 12 (1), pp. 55-67.

A.M. Legendre (1805), *Nouvelles méthodes pour la détermination des orbites des comètes (appendix: sur la méthode des moindres quarrés).* Courcier, Paris. Cited in Stigler (1986) where the appendix on least squares is reproduced.

D.W. Marquardt (1970), *Generalized inverses, ridge regression, biased linear estimation and non-linear estimation.* Technometrics, no. 12 (3), pp. 591-612.

L.S. Mayer, T.A. Willke (1973) *On biased estimation in linear models.* Technometrics, no. 15 (3), pp. 497-508.

R.L. Plackett (1972), *The discovery of the method of least squares.* Biometrika, no. 59 (3), pp. 239-251.

S.M. Stigler (1986), *The history of statistics.* Belknap Press, Cambridge, Massachusetts.