GEODEZJA I KARTOGRAFIA
GEODESY AND CARTOGRAPHY
Vol. 55, No 1, 2006, pp. 3-22

Revisiting least squares: A discussion on the leading estimation principle in geodesy

Christopher Kotsakis[1], Michael G. Sideris[2]

[1]Department of Geodesy and Surveying, Aristotle University of Thessaloniki
University Box 440, Thessaloniki, GR-54124, Greece
e-mail: kotsaki@topo.auth.gr
[2]Department of Geomatics Engineering, University of Calgary
2500 University Drive N.W., Calgary, Alberta T2N 1N4, Canada
e-mail: sideris@ucalgary.ca

Received: 5 October 2005/Accepted: 29 March 2006

Abstract: Least squares (LS) estimation is one of the most important tools in geodetic data analysis. However, its prevailing use is not often complemented by an objective view of its rudiments. Within the standard formalism of LS estimation theory there are actually several paradoxical and curious issues which are seldom explicitly formulated. The aim of this expository paper is to present some of these issues and to discuss their implications for geodetic data analysis and parameter estimation problems. In the first part of the paper, an alternative view of the statistical principles that are traditionally linked to LS estimation is given. Particularly, we show that the property of unbiasedness for the ordinary LS estimators can be replaced with a different, yet equivalent, constraint which implies that the numerical range of the unknown parameters is boundless. In the second part of the paper, the shortcomings of the LS method are exposed from a purely algebraic perspective, without employing any concepts from the probabilistic/statistical framework of estimation theory. In particular, it is explained that what is 'least' in least squares is certainly not the errors in the estimated model parameters, and that in every LS-based inversion of a linear model there exists a critical trade-off between the Euclidean norms of the parameter estimation errors and the adjusted residuals.

Keywords: Least squares, unbiasedness, unboundedness, linear model, ill-posed problems, uncertainty principle, estimation error

1. Introduction

The method of least squares (LS) estimation is one of the most popular and important tools for geodetic data analysis (Dermanis and Rummel, 2000). Since its formal development by Gauss and Legendre, scientific inference based on LS techniques has overwhelmingly dominated the work of geodesists in both theoretical and practical aspects. Moreover, the underlying principles of LS estimation have acquired a status of universal consent over the years, which occasionally discourages the pursue of alternative routes for data analysis and parameter estimation problems. For example,

geodesists are often guided by an unenthusiastic (if not negative) attitude towards the use of *biased* estimation techniques, thus conforming to the widespread belief that unbiasedness is a natural property that should always accompany the process of optimal parameter estimation. A characteristic paradigm is the use of ridge-type estimators which is commonly justified on the basis of improving the numerical stability of the ordinary LS solution at the expense of a small deliberately induced bias, and rarely on the fact that ridge regression is an independent estimation technique which has its own compound theoretical background that strongly competes with the traditional LS logic (Efron, 1975; Rao and Toutenburg, 1995).

The objective of this paper is to re-visit the LS "logic" by discussing some of its problematic aspects for parameter estimation problems. In the first part of the paper an alternative view of the statistical principles that are traditionally linked to LS estimation is given. Specifically, it is shown that the property of *unbiasedness* for the ordinary LS estimators can be replaced with a different, yet equivalent, constraint which implies that the numerical range of the unknown parameters is boundless. The consequences that arise from this dualism in the context of geodetic parameter estimation are discussed, and some comments on the statistical foundations of least-squares are also made. In the second part of the paper, the shortcomings of the LS logic are exposed from a purely algebraic perspective without making any reference to quantities commonly used within a probabilistic framework (e.g. variance, bias, mean squared error). In particular, it is explained that what is 'least' in least squares is certainly not the errors in the estimated parameters, and that in every LS-based inversion of a linear model there exists a critical trade-off between the Euclidean norms of the parameter estimation errors and the adjusted residuals. A type of *uncertainty principle* for inverse problems with linear models is also introduced in the paper, which gives further insight into the limitations of classic LS estimation.

It must be noted that the ideas presented herein should not be viewed as a disagreement with the practice of applying the LS methodology for geodetic data processing, but merely as a suggestion to adopt a more objective and unbiased appreciation of its optimality principles. Also, let us add that there is nothing mathematically new about the considerations presented in this paper (although some well-known issues have been elaborated more deeply), and what we only attempt to convey is that the logic associated with LS estimation receives a more objective treatment than it is often given by geodesists. An interesting appraisal of the LS theory for geodetic adjustment problems can also be found in Moritz (1989), which however puts emphasis on rather different issues from the ones discussed herein.

2. Unbiased or biased estimation?

The criterion of unbiasedness is intrinsically related to the *frequentist* approach in probability theory. The latter provides the backbone of what is known as the classic view for statistical inference problems, in contrast to the more controversial *Bayesian* methods which represent the movement of subjectivists in estimation theory (Jaynes,

2003). Note that for Bayesians there is usually no contemplation about possible biases in their estimators, since this notion is totally irrelevant within the Bayesian vision of statistical inference (Barnett, 1982; pp. 16–19).

Without trying to diminish the value of the unbiasedness criterion, a warning must be raised to indicate that we may have been caught in a semantic trap of our own making. The very name *unbiased* suggests the appeal of this concept to notions of scientific objectivity. However, let us take a closer look at the general form of the *mean squared error* (MSE) of an arbitrary parameter estimation process

$$MSE_{\hat{\theta}} = E\{(\hat{\theta} - \theta)^2\} = [E\{\hat{\theta}\} - \theta]^2 + var(\hat{\theta}) \quad (1)$$

where $\hat{\theta}$ is some estimator of an unknown deterministic parameter θ, $E\{\}$ denotes the ensemble averaging operator in the sense of probability theory, and *var*() stands for the variance of an underlying random variable. The difference $E\{\hat{\theta}\} - \theta$ is commonly termed the 'bias' of the estimate $\hat{\theta}$, a name which implies something that we must remove at all costs. If this difference had been called instead the "component of the MSE that is orthogonal to the variance of $\hat{\theta}$", as suggested by the Pythagorean expression of (1), it would have been more clear that the two contributions to the total MSE are on equal footing. It is foolish to decrease one at the expense of increasing the other, yet this is exactly what unbiased estimators normally do when they set a priori the difference $E\{\hat{\theta}\} - \theta$ equal to zero.

Various arguments can be found to support and/or to oppose the use of the unbiasedness criterion in parameter estimation problems. Statisticians often dispose the significance of unbiased estimates because they do not remain invariant under a simple transformation in the underlying parameters, i.e. the square of an unbiased estimate $\hat{\theta}$ of θ is not necessarily an unbiased estimate of θ^2. On the other hand, a similar reasoning can be adopted against biased estimates, since a linear transformation of the input data does not automatically implies a corresponding linear transformation to a biased estimator (Efron, 1975). An appraisal of the unbiasedness property is given in (Savage, 1972; pp. 244–245), where it is claimed that a serious reason to prefer unbiased estimates in statistics seems never to have been proposed, while Tukey (1960) supports the abandoning of the unbiasedness property in data analysis problems from a modeller's point of view. Note that, in spite of its apparent persuasive logic, an unbiased estimator may lead to absurd results in practice. An example is given in (Barnard, 1963) where it is shown that the condition of unbiasedness can produce an estimate $\hat{\theta}$ of all values which lie outside the range of possible θ.

An elegant argument in support of unbiased estimators is that they can recover the true model parameters in the presence of errorless data (Liebelt, 1967; p. 137), whereas biased estimators generally fail to achieve this property. In fact, Gauss had incorporated this reasoning into the requirements for an optimal linear estimator when he established the LS method in a rigorous statistical context and proved its equivalency with Legendre's original formulation (Eisenhart, 1964). However, one may provocatively argue that it is not reasonable to adopt an optimal estimation criterion on the

hypothetical basis of using it with errorless observations. After all, geodetic data sets are always contaminated by external random noise.

In a probabilistic context, the unbiasedness condition for parameter estimation problems ensures the exact (errorless) recovery of an unknown quantity after infinitely many repetitions of the data collection process and their subsequent adjustment: $E\{\hat{\theta}\} = \theta$. This is, however, a rather peculiar principle to base the optimal processing of a *single* data set, since all that we have available in practice is a finite group of measurements! If asymptotic properties need to be imposed on the statistical behaviour of a parameter estimator, why not choose *consistency* instead of unbiasedness? A consistent estimator $\hat{\theta}$ approaches the true unknown value as the number of the available data increases (Barnett, 1982; p. 134). It can be argued that, in view of the increasing volume of geodetic data that becomes available nowadays, such an asymptotic property seems a more sensible requirement than the unbiasedness condition. Note that consistency is quite different in spirit from unbiasedness, and the two properties are generally unrelated since there exist unbiased inconsistent estimators as well as biased consistent estimators (Savage, 1972).

Taking a more practical standpoint, it is instructive (and fair) to recognize the merit of unbiasedness in the sense that it sets to zero one of the two factors contributing to the total MSE of any parameter estimation procedure. LS estimators, in particular, combine this property with the additional minimization of the second term that contributes to the total MSE, namely $var(\hat{\theta})$; see (1). However, the fact that an estimator has minimum variance in the class of unbiased estimators does not guarantee that its variance is numerically small. In linear models with an ill-conditioned design matrix, in particular, the variances of the LS estimated parameters may be too large for practical purposes. A simple example can be mentioned for the case of the standard Gauss-Markov model, $\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{v}$, $E\{\mathbf{v}\} = \mathbf{0}$, $E\{\mathbf{v}\mathbf{v}^{\mathrm{T}}\} = \sigma^2\mathbf{I}$, where it holds that (Hoerl and Kennard, 1970)

$$trace\ \mathbf{C}_{\hat{\mathbf{x}}} > \frac{\sigma^2}{\lambda_{\min}} \tag{2}$$

where $\mathbf{C}_{\hat{\mathbf{x}}}$ is the covariance (CV) matrix of the LS estimated parameters, and $\lambda_{\min}$ is the smallest eigenvalue of the normal matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. If some of the columns of $\mathbf{A}$ are nearly collinear, then $\lambda_{\min}$ is close to zero, which in turn causes a significant deterioration in the statistical accuracy of the LS solution.

3. Two equivalent statistical views of LS estimation

According to the Gauss-Markov theorem, the LS methodology provides a linear and uniformly unbiased estimator that has the minimum MSE among any other linear unbiased solution; such estimators will be abbreviated in the sequel as BLUE (Best Linear Unbiased Estimators). In this section, after a short review of the BLUE perspective, we present an alternative viewpoint for the statistical properties that can be associated with the ordinary LS estimators.

3.1. The general linear model

Let us begin with a system of linear (or linearized) observation equations of the following form

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{v} \tag{3}$$

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

$$E\{\mathbf{v}\} = 0, \quad E\{\mathbf{v}\,\mathbf{v}^{\mathrm{T}}\} = \mathbf{C} \tag{4}$$

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

Despite its simplistic linear character and its inherent restriction for additive data errors, the model of (3) is overwhelmingly used in all areas of modern geodetic research. In principle, in all such cases we generally seek to determine an unknown quantity θ which depends on the archetypal parameter vector $\mathbf{x}$. For convenience, we consider here only the case where θ is a linear function of the unknown parameters

$$\theta = \mathbf{q}^{\mathrm{T}}\,\mathbf{x} \tag{5}$$

with $\mathbf{q}$ being an arbitrary known vector. Based on the knowledge of the data vector $\mathbf{y}$, various types of estimators $\hat{\theta} = \hat{\theta}(\mathbf{y})$ can be considered, each of which complies to specific optimal criteria and assumptions. In the following section, we describe the statistical characteristics of the usual LS estimators that can be associated with the linear models of (3) and (5).

3.2. The BLUE approach

A linear estimator of the unknown scalar quantity θ has the general form

$$\hat{\theta} = \mathbf{b}^{\mathrm{T}}\,\mathbf{y} + c \tag{6}$$

where the vector $\mathbf{b}$ and the scalar c need to be determined according to some optimal criteria. The statistical formulation of LS estimation is based on two fundamental properties that the estimator of (6) should satisfy simultaneously, namely

1) *Uniform unbiasedness*: $E\{\hat{\theta}\} = \theta = \mathbf{q}^{\mathrm{T}}\,\mathbf{x}$ for any parameter vector $\mathbf{x}$,

2) *Minimum MSE*: $E\{(\hat{\theta} - \theta)^2\}$ = minimum.

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

$$\mathbf{b}^T \mathbf{A} = \mathbf{q}^T \tag{7}$$

$$c = 0 \tag{8}$$

The MSE of the linear estimator $\hat{\theta}$ has the general form

$$E\{(\hat{\theta} - \theta)^2\} = \mathbf{b}^T \mathbf{C}\, \mathbf{b} + [(\mathbf{b}^T\mathbf{A} - \mathbf{q}^T)\, \mathbf{x} + c]^2 \tag{9}$$

and its minimization, subject to the constraints of (7) and (8), leads to a unique solution for $\mathbf{b}$ through the method of Lagrange multipliers. The result is given by the following equation

$$\mathbf{b} = \mathbf{C}^{-1}\mathbf{A}\, (\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}\, \mathbf{q} \tag{10}$$

Based on (6), (8) and (10), the optimal estimate of $\theta = \mathbf{q}^T \mathbf{x}$ is thus given by

$$\hat{\theta} = \mathbf{q}^T\, (\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\, \mathbf{C}^{-1}\, \mathbf{y} \tag{11}$$

which implies the following, well known LS estimate for the parameter vector $\mathbf{x}$

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\, \mathbf{C}^{-1}\, \mathbf{y} \tag{12}$$

In the next section, the BLUE estimators of (11) and (12) are derived through a different approach, without departing from the broad context of optimal statistical inference. This alternative perspective represents only an attempt to elucidate the logic of the unbiasedness condition that is associated with LS estimation.

3.3. Unboundedness vs. unbiasedness: an alternative approach for LS estimators

Taking into account (9), let us point out the critical fact that the MSE of the linear estimator $\hat{\theta}$ depends directly on the vector of the unknown parameters $\mathbf{x}$. This is something that we should generally expect from any type of estimator, regardless of the algorithmic relationship between the parametric quantities $\mathbf{x}$ and θ. Consequently, if the range of $\mathbf{x}$ is unbounded, the second term in the MSE expression of (9) becomes unbounded too.

In order to ensure that the MSE of $\hat{\theta}$ remains finite, regardless of the numerical range of the unknown parameters, the following condition should thus be satisfied (see Eq. (9))

$$\mathbf{b}^T \mathbf{A} - \mathbf{q}^T = \mathbf{0}^T \tag{13}$$

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

$$\mathbf{b} = \mathbf{C}^{-1}\mathbf{A}\,(\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\,\mathbf{q} \quad \text{and} \quad c = 0 \tag{14}$$

which, in turn, gives rise to the same optimal estimates for θ and $\mathbf{x}$ that were derived in the last section.

Hence, we can obtain the same BLUE-type optimal solution without using the requirement of unbiasedness for the estimated parameters. An equivalent statistical formulation of the LS estimation process can thus emerge which is articulated as follows: "among all linear estimators that can provide finite MSE for a set of unknown parameters with unbounded (or unspecified) range, LS estimators yield results with minimum MSE".

It may appear that we have removed the requirement of unbiasedness at the expense of a more restricted version for the LS estimation method. Obviously, the property of unbiasedness for the estimated parameters has not been lost in this case, since it will now be a direct consequence of (14). Furthermore, the resulting estimators are not restricted in any way, and they can be implemented regardless of the actual range of the true parameter vector $\mathbf{x}$. In fact, what the previous alternative formulation should make us sceptical about is the following question: *will the ordinary LS estimation algorithm give optimal results (in the MSE sense) when* $\mathbf{x}$ *is a vector of bounded parameters within a known specified range?*

4. Discussion

An instructive way to look at LS estimation is to recognize the fact that its statistical optimality is closely associated with the assumption that the range of the unknown parameters $\mathbf{x}$ is unbounded. Such a perception is not new and it has already appeared in the statistical literature (e.g. Barnard, 1963; Hoerl and Kennard, 1970). Geodesists should be aware of this connection, since in all geodetic adjustment problems the parameters that need to be determined surely lie within a finite range. It should be acknowledged that this type of information is not integrated in the ordinary LS estimation process.

The statistical logic of the LS principle, as this is depicted in terms of a linear unbiased estimator with minimum MSE, ignores the fact that the unknown parameters are finite in magnitude. A manifestation of this fact is that LS solutions tend to give numerical answers that are 'longer', when measured by some Euclidean-type norm, than the actual true parameter vector. As an example, we can recall the following well known result from the LS adjustment of a simple Gauss-Markov model

$$E\{\hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}}\} = \mathbf{x}^{\mathrm{T}}\mathbf{x} + \sigma^2\, trace(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1} \tag{15}$$

which shows that the estimated parameter vector is indeed expected to be longer than the corresponding true vector $\mathbf{x}$.

From the analysis given in section 3, it can be deduced that the property of unbiasedness is responsible for causing the BLUE estimators to be blind on the bounded

nature of the unknown parameters. This dualism brings up a fairly strong argument in favor of biased estimation methods, although in geodesy we have often kept a pessimistic mindset for such techniques. Nevertheless, shouldn't we prefer an estimation algorithm which respects the fact that quantities such as the geodetic coordinates in control networks, or the spherical harmonic coefficients of the gravitational potential, cannot exceed some physically reasonable limits? We must also not forget that the linear(-ized) model of observation equations in (3) gives a realistic picture of physical reality when the parameters $\mathbf{x}$ are confined within certain intervals around their initial approximate values.

At this point, one can argue that the finite range of the unknown parameters can be taken into account without giving up the standard LS principle that was originally introduced by Legendre. That is, we can always seek optimal estimators $\hat{\mathbf{x}}$ which minimize the sum of the squared values in the residual vector $\hat{\mathbf{v}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$, subject to appropriate inequality constraints that bound the size of the unknown parameters $\mathbf{x}$. Such problems are very common in various disciplines of geosciences and engineering, and they can be handled through well known techniques of convex optimisation and non-linear programming (e.g. Björck, 1996). However, it is important to point out that such constrained LS estimators, using either linear or quadratic inequality constraints, do not generally produce unbiased results for the unknown parameters and their exact statistical properties are quite difficult to be determined (Rao and Toutenburg, 1995; pp. 75–82).

As an example, we can mention the problem where an upper bound $b_{\max}$ is imposed on the Euclidean length of the parameter vector in the linear model of (3). If we seek the least squares estimator in the Legendre's sense, i.e. $(\mathbf{y} - \mathbf{A}\mathbf{x})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}) = \textit{minimum}$, subject to the restriction $\mathbf{x}^{\mathrm{T}}\mathbf{x} \leqslant b_{\max}$, the final result takes the *ridge regression* form

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{A} + k\mathbf{I})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y} \tag{16}$$

where the regularization parameter k is uniquely determined in this case from the given bound $b_{\max}$ as discussed, for example in (Björck, 1996; p. 205). The biased ridge estimator of (16) is a common competitor of the LS solution given in (12). The ordinary LS estimators are actually known to have worse MSE performance than ridge estimators, when the Euclidean length of the parameter vector $\mathbf{x}$ is bounded (Hoerl and Kennard, 1970). However, it should be pointed out that in geodetic applications with linear models having a proper (well-conditioned) algebraic structure, the differences that occur in the estimation performance between classic LS techniques and other regularization (biased estimation) methods are usually insignificant.

Closing this discussion, we suggest that potential users of ridge regression or other similar techniques that compete with the standard LS estimators should not think only in terms of having to choose between a biased or unbiased solution, but they should also consider the concurrent knowledge about physical reality that each of these methods brings to our quest for understanding.

5. An algebraic perspective for linear model inversion

The logic of LS estimation will now be examined from a different perspective. In particular, the properties of ordinary LS estimators will be studied from an algebraic point of view, without utilizing any probabilistic concepts such as variance, MSE, bias, etc. Actually, the rationale of the original formulation of LS estimation by Legendre, i.e. minimization of the sum of the squared residuals, refers to the specific data set that is available and needs to be analysed. After all, the asymptotic behaviour of an estimator is not really relevant for practical purposes. The real problem has always been to optimise the inference that we can obtain from a particular and finite set of data values. Adopting a frequentist's approach that optimises the average estimation performance over an infinite number of repeated data sets obtained under 'identical conditions' can be considered irrelevant, since it corresponds to an imaginary scenario that is never attained in practice.

5.1. Input/output error flow in linear models

The linear model in (3) provides again the basic framework for our analysis. A crucial point in the algebraic approach for optimal parameter estimation is that both $\mathbf{x}$ and $\mathbf{v}$ have specific and fixed values. In the statistical approach the residual vector $\mathbf{v}$ is considered as a realization of a set of random variables, whereas here it is treated as a fixed error vector that is directly linked with the particular data set $\mathbf{y}$.

If we denote by $\hat{\mathbf{x}}$ an estimate for the model parameters and by $\hat{\mathbf{v}}$ the corresponding estimate of the observation errors, the following equation can be written

$$\mathbf{y} = \mathbf{A}\hat{\mathbf{x}} + \hat{\mathbf{v}} \tag{17}$$

which, in conjunction with (3), leads to the decomposition formula

$$\mathbf{v} = \hat{\mathbf{v}} + \mathbf{A}\hat{\mathbf{e}} \tag{18}$$

The quantity $\hat{\mathbf{e}}$ corresponds to the error vector for the model parameters

$$\hat{\mathbf{e}} = \hat{\mathbf{x}} - \mathbf{x} \tag{19}$$

It is worth noting that the left-hand side of (18) is independent of the estimation procedure. The values in $\mathbf{v}$ are solely dictated by: (i) the quality of the specific measurements in our disposal, and (ii) the true values of the unknown parameters. One can claim that the residuals $\mathbf{v}$ are also a function of our parameterisation choice (i.e. $\mathbf{Ax}$) and a part of them will reflect possible modelling errors in the description of physical reality through a finite set of parameters. The most common example is the linearization errors that occur when we neglect higher order terms in non-linear models. Even so, the fact remains that $\mathbf{v}$ is not affected by the method used to invert the linear system in (3).

A key aspect in linear model inversion is to determine how the original errors $\mathbf{v}$ will be distributed between the two output error 'channels', namely $\hat{\mathbf{v}}$ and $\hat{\mathbf{e}}$. It is logical to claim that any notion of optimality for the parameter estimator $\hat{\mathbf{x}}$ should take into account (at least to some degree) the mutual dependence that exists between the adjusted residuals and the parameter estimation errors according to (18).

An instructive way to view (18) is to perceive it as an error conservation formula. In any estimation process that can be designed to invert a linear model, the amount of the input errors is always balanced out by the amount of the output errors. In our case, the input errors are related to the uncertainty of the given measurements (including also possible modelling errors), whereas the output errors contain both the uncertainty in the estimated model parameters ($\hat{\mathbf{e}}$) and the misfit between the estimated model and the given data set ($\hat{\mathbf{v}}$). This kind of error flow that takes place during the inversion of every linear model is illustrated in Figure 1.

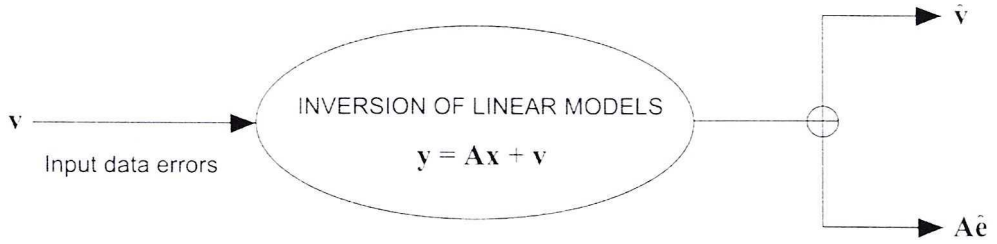

Fig. 1. Error flow during the inversion of linear model

5.2. *The Gauss-Legendre uncertainty principle*

In this section, we present an uncertainty principle that underlies every algorithm used in linear model inversion. For this purpose, we need to adopt a metric or a norm that can measure the size of vectors in the so-called observation space. Such vectors with common dimensionality are: the data vector $\mathbf{y}$, the vector of the true data errors $\mathbf{v}$, the vector of the adjusted data errors $\hat{\mathbf{v}}$, and the vector $\mathbf{A}\hat{\mathbf{e}}$ that represents the propagated effect of the parameter estimation errors into the observation space.

A Euclidean norm will be adopted to define the geometric structure in the observation space. In particular, the length of the data error vector is given by the quadratic form

$$\|\mathbf{v}\|_{\mathbf{P}} = (\mathbf{v}^{\mathrm{T}}\mathbf{P}\mathbf{v})^{1/2} \tag{20}$$

where $\mathbf{P}$ denotes some positive-definite symmetric matrix that specifies the Euclidean metric properties of the observation space. From (18) we have

$$\|\mathbf{v}\|_{\mathbf{P}} = \|\hat{\mathbf{v}} + \mathbf{A}\hat{\mathbf{e}}\|_{\mathbf{P}} \tag{21}$$

Applying the well known *triangle inequality* to the above formula, yields the following relationship

$$\|\mathbf{v}\|_{\mathbf{P}} \leqslant \|\hat{\mathbf{v}}\|_{\mathbf{P}} + \|\mathbf{A}\hat{\mathbf{e}}\|_{\mathbf{P}} \tag{22}$$

or equivalently

$$\|\mathbf{v}\|_{\mathbf{P}} \leqslant \|\hat{\mathbf{v}}\|_{\mathbf{P}} + \|\hat{\mathbf{e}}\|_{\mathbf{Q}} \tag{23}$$

where $\mathbf{Q} = \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A}$. Note that the metric change between (22) and (23) is permissible when quadratic norms are used, since

$$\begin{aligned} \|\mathbf{A}\hat{\mathbf{e}}\|_{\mathbf{P}} &= ([\mathbf{A}\hat{\mathbf{e}}]^{\mathrm{T}}\mathbf{P}[\mathbf{A}\hat{\mathbf{e}}])^{1/2} \\ &= (\hat{\mathbf{e}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A}\hat{\mathbf{e}})^{1/2} \\ &= (\hat{\mathbf{e}}^{\mathrm{T}}\mathbf{Q}\hat{\mathbf{e}})^{1/2} \\ &= \|\hat{\mathbf{e}}\|_{\mathbf{Q}} \end{aligned} \tag{24}$$

The inequality in (23) identifies an important bound for the performance of any parameter estimation method in the linear model of (3). It actually demonstrates the fact that, regardless of the specific type of estimator $\hat{\mathbf{x}}$ that we choose to use, we are always limited by the quality of our own given data.

The basic meaning of (23) is that, whenever we try to invert a linear system in the presence of additive observation errors, we cannot simultaneously achieve an arbitrarily good estimate for the model parameters *and* an arbitrarily good fitting to the given data. If, for example, an estimation algorithm is used such that the error in the recovered parameters becomes very small in the sense that $\|\hat{\mathbf{e}}\|_{\mathbf{Q}} \rightarrow 0$, then the misfit $\hat{\mathbf{v}}$ between the actual data and the estimated model will be *at least* as large as the original observation errors.

The above result must be kept in mind by LS users who should not blindly get the impression that small adjusted residuals are always associated with good results for their estimated parameters. A relevant example in geodetic practice is provided by the numerous studies on gravimetric geoid evaluation via comparisons with GPS and levelling data. The cm-level residuals that are often obtained after the common adjustment of ellipsoidal, orthometric and geoid heights at a network of co-located GPS/levelling benchmarks do not necessarily guarantee that the gravimetric geoid accuracy is at the same level, since a significant part of the data errors can be absorbed by the parameters of the corrector surface model that is traditionally employed for this type of adjustment problems.

The inequality in (23) can be identified as the *Gauss-Legendre uncertainty principle* since these two scientists were the first who made the most important contributions for dealing with experimental uncertainty in data analysis problems. For the purpose of this paper, equation (23) provides a useful guideline that should remind us of the inherent limitations that accompany all criteria, including LS, for the optimal inversion of linear models.

5.3. *Input/output error relationships in linear models*

The Gauss-Legendre uncertainty principle indicates that in every linear model inversion there is a trade-off between the accuracy of the estimated model parameters and the fit to the original data. The purpose of this section is to present some alternative formulae among the various error norms emerging from the inversion of (3), that can give further insight on the trade-off between $\hat{\mathbf{v}}$ and $\hat{\mathbf{e}}$.

Starting from the basic decomposition in (18), we can obtain the quadratic formula

$$\mathbf{v}^{\mathrm{T}}\mathbf{P}\mathbf{v} = \hat{\mathbf{v}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{v}} + \hat{\mathbf{e}}^{\mathrm{T}}\mathbf{Q}\hat{\mathbf{e}} + 2\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{P}\mathbf{A}\hat{\mathbf{e}} \tag{25}$$

where $\mathbf{P}$ is a positive-definite symmetric matrix whose role is to provide a metric structure for the observation space. In contrast to the choice of $\mathbf{P}$, which is rather arbitrary, the metric matrix that is used to quantify the size of the parameter error vector $\hat{\mathbf{e}}$ in (25) is uniquely determined as $\mathbf{Q} = \mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A}$.

If we now express the Euclidean inner product between the vectors $\hat{\mathbf{v}}$ and $\mathbf{A}\hat{\mathbf{e}}$ as

$$\langle \hat{\mathbf{v}}, \mathbf{A}\hat{\mathbf{e}} \rangle_{\mathbf{P}} = \hat{\mathbf{v}}^{\mathrm{T}}\mathbf{P}\mathbf{A}\hat{\mathbf{e}} \tag{26}$$

then (25) may equivalently be written as

$$\|\mathbf{v}\|_{\mathbf{P}}^2 = \|\hat{\mathbf{v}}\|_{\mathbf{P}}^2 + \|\hat{\mathbf{e}}\|_{\mathbf{Q}}^2 + 2\langle \hat{\mathbf{v}},\ \mathbf{A}\hat{\mathbf{e}} \rangle_{\mathbf{P}} \tag{27}$$

An alternative expression can also be obtained if we employ the usual geometric interpretation of the inner product between two vectors. In such case the previous equation takes the form

$$\|\mathbf{v}\|_{\mathbf{P}}^2 = \|\hat{\mathbf{v}}\|_{\mathbf{P}}^2 + \|\hat{\mathbf{e}}\|_{\mathbf{Q}}^2 + 2\,\|\hat{\mathbf{v}}\|_{\mathbf{P}}\,\|\hat{\mathbf{e}}\|_{\mathbf{Q}} \cos\theta \tag{28}$$

where θ is the 'angle' between the vectors $\hat{\mathbf{v}}$ and $\mathbf{A}\hat{\mathbf{e}}$ such that

$$\cos\theta = \frac{\langle \hat{\mathbf{v}},\ \mathbf{A}\hat{\mathbf{e}} \rangle_{\mathbf{P}}}{\|\hat{\mathbf{v}}\|_{\mathbf{P}}\,\|\mathbf{A}\hat{\mathbf{e}}\|_{\mathbf{P}}} = \frac{\langle \hat{\mathbf{v}},\ \mathbf{A}\hat{\mathbf{e}} \rangle_{\mathbf{P}}}{\|\hat{\mathbf{v}}\|_{\mathbf{P}}\,\|\hat{\mathbf{e}}\|_{\mathbf{Q}}} \tag{29}$$

Another interesting expression can be derived by using the identity

$$\|\mathbf{a} + \mathbf{b}\|^2 + \|\mathbf{a} - \mathbf{b}\|^2 = 2\,\|\mathbf{a}\|^2 + 2\,\|\mathbf{b}\|^2 \tag{30}$$

that is valid for any vectors $\mathbf{a}$ and $\mathbf{b}$ in a vector space whose metric structure is derived from an inner product (Cloud and Drachman, 1998; pp. 61–62). By setting $\mathbf{a} = \hat{\mathbf{v}}$ and $\mathbf{b} = \mathbf{A}\hat{\mathbf{e}}$, and taking into account the error decomposition from (18), the following equation is obtained

$$\|\mathbf{v}\|_{\mathbf{P}}^2 = 2\,\|\hat{\mathbf{v}}\|_{\mathbf{P}}^2 + 2\,\|\hat{\mathbf{e}}\|_{\mathbf{Q}}^2 - \|\hat{\mathbf{r}}\|_{\mathbf{P}}^2 \tag{31}$$

where the vector $\hat{\mathbf{r}}$ denotes the difference

$$\hat{\mathbf{r}} = \hat{\mathbf{v}} - \mathbf{A}\hat{\mathbf{e}} \tag{32}$$

All the previous formulae relating the quadratic error norms that emerge from the inversion of (3) are summarized in Table 1. Note that these equations are valid for *any* type of estimator $\hat{\mathbf{x}}$ that leads to the particular adjusted residuals $\hat{\mathbf{v}}$ and parameter estimation errors $\hat{\mathbf{e}}$.

Keeping in mind that $\|\mathbf{v}\|_{\mathbf{P}}$ is fixed and independent of the chosen estimation method, the three equations shown in Table 1 should be viewed as constraints on the way that the measurement noise is spread among the output errors in linear model inversion. Clearly, if one wants to follow a pure LS inversion approach by minimizing only the norm of the adjusted residuals $\|\hat{\mathbf{v}}\|_{\mathbf{P}}$, he takes the risk that a large part of the data error energy may leak into $\|\hat{\mathbf{e}}\|_{\mathbf{Q}}$ thus causing bad parameter estimates.

Table 1. Summary of the various relationships that exist between the input and output quadratic error norms during the inversion of a linear model

Input Data Error Norm (*fixed*)	**Output Error Contributions** (*depend on the chosen inversion principle*)		
$\|\mathbf{v}\|_{\mathbf{P}}^2$	$\|\hat{\mathbf{v}}\|_{\mathbf{P}}^2$	see (27) $\|\hat{\mathbf{e}}\|_{\mathbf{Q}}^2$	$2\langle\hat{\mathbf{v}}, \mathbf{A}\hat{\mathbf{e}}\rangle_{\mathbf{P}}$
	$\|\hat{\mathbf{v}}\|_{\mathbf{P}}^2$	see (28) $\|\hat{\mathbf{e}}\|_{\mathbf{Q}}^2$	$2\|\hat{\mathbf{v}}\|_{\mathbf{P}}\|\hat{\mathbf{e}}\|_{\mathbf{Q}}\cos\theta_{\hat{\mathbf{v}},\mathbf{A}\hat{\mathbf{e}}}$
	$2\|\hat{\mathbf{v}}\|_{\mathbf{P}}^2$	see (31) $2\|\hat{\mathbf{e}}\|_{\mathbf{Q}}^2$	$-\|\hat{\mathbf{v}} - \mathbf{A}\hat{\mathbf{e}}\|_{\mathbf{P}}^2$

5.4. Is the LS method a paradoxical choice?

The merit of the algebraic LS principle is based on the minimization of a Euclidean norm for the adjusted residuals. Indeed, if we minimize $\|\hat{\mathbf{v}}\|_{\mathbf{P}}$ subject to the quadratic constraint given in (25), we obtain the familiar LS estimator

$$\hat{\mathbf{x}}^{\mathrm{LS}} = (\mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{y} \tag{33}$$

In this particular case, the input and output error norms are connected through a simple Pythagorean relationship

$$\|\mathbf{v}\|_{\mathbf{P}}^2 = \left\|\hat{\mathbf{v}}^{\mathrm{LS}}\right\|_{\mathbf{P}}^2 + \left\|\hat{\mathbf{e}}^{\mathrm{LS}}\right\|_{\mathbf{Q}}^2 \tag{34}$$

where $\hat{\mathbf{v}}^{\mathrm{LS}}$ and $\hat{\mathbf{e}}^{\mathrm{LS}}$ denote the adjusted residuals and the parameter errors that correspond to the least-squares solution $\hat{\mathbf{x}}^{\mathrm{LS}}$.

When viewed under this perspective, the LS logic can be questioned on the basis that it places no emphasis on the actual errors of the estimated parameters. Moreover, the sole minimization of $\|\hat{\mathbf{v}}\|_{\mathbf{P}}$ carries the risk that most of the data error energy is propagated into the model parameters. A more sensible approach would be to control

the relative importance of the adjusted residuals and the parameter errors through a regularized regression of the data noise $\|\mathbf{v}\|_{\mathbf{P}}$ between $\|\hat{\mathbf{v}}\|_{\mathbf{P}}$ and $\|\hat{\mathbf{e}}\|_{\mathbf{Q}}$. Actually, the regularization techniques that are commonly employed in geodesy for the solution of ill-posed inverse problems, such as the *Philips – Tikhonov* inversion scheme or the truncated singular value decomposition method (Schwarz, 1979; Xu and Rummel, 1994; Bouman and Koop, 1997; Xu, 1998), have the potential to produce a more balanced dissemination of the data noise $\mathbf{v}$ between the adjusted residuals $\hat{\mathbf{v}}$ and the parameter estimation errors $\hat{\mathbf{e}}$.

In order to illustrate further the shortcomings of the algebraic LS principle, a simple instructive example can be set up as follows. Let us assume that a vector of observations $\mathbf{y}$ is given which is affected by additive unknown errors $\mathbf{v}$. We can generally write

$$\mathbf{y} = \mathbf{b} + \mathbf{v} \tag{35}$$

where $\mathbf{b}$ denotes the unknown theoretical values of the observable quantities. Let us also assume that various possible parameterisations $\mathbf{b} = f(\mathbf{x})$ exist for the observable vector. For the sake of argument, we consider only two linear schemes

$$\mathbf{y} = \mathbf{A}_1\mathbf{x}_1 + \mathbf{v} \quad (\textit{Model 1}) \tag{36}$$

and

$$\mathbf{y} = \mathbf{A}_2\mathbf{x}_2 + \mathbf{v} \quad (\textit{Model 2}) \tag{37}$$

Note that the data errors are assumed equal in both cases, which implies that each parameterisation is theoretically compatible with the underlying physical phenomenon, i.e. adopting one or the other for the representation of $\mathbf{b}$ does not introduce extra *modelling errors* into our formulation. The question now is which of the above parametric models should we prefer for the optimal analysis of the data set $\mathbf{y}$?

If we adopt the logic of the algebraic LS principle, we are inclined to choose the configuration that provides the best fit to the given data. Let us assume that *model 1* performs better than *model 2* in the LS sense. In this way, we have

$$\left\|\hat{\mathbf{v}}_1^{\mathrm{LS}}\right\|_{\mathbf{P}}^2 < \left\|\hat{\mathbf{v}}_2^{\mathrm{LS}}\right\|_{\mathbf{P}}^2 \tag{38}$$

where $\hat{\mathbf{v}}_1^{\mathrm{LS}}$ and $\hat{\mathbf{v}}_2^{\mathrm{LS}}$ denote the adjusted residuals that are obtained from the LS solutions of the two models. Note that both quadratic norms in (38) employ the same weight matrix $\mathbf{P}$, since the same set of observations is used in both cases.

Taking into account (34), the previous inequality yields

$$\|\mathbf{v}\|_{\mathbf{P}}^2 - \left\|\hat{\mathbf{e}}_1^{\mathrm{LS}}\right\|_{\mathbf{Q}_1}^2 < \|\mathbf{v}\|_{\mathbf{P}}^2 - \left\|\hat{\mathbf{e}}_2^{\mathrm{LS}}\right\|_{\mathbf{Q}_2}^2 \tag{39}$$

where $\|\mathbf{v}\|_{\mathbf{P}}$ is the Euclidean norm of the data noise which is common in both models.

From (39), we finally obtain

$$\left\|\hat{\mathbf{e}}_1^{LS}\right\|_{\mathbf{Q}_1}^2 > \left\|\hat{\mathbf{e}}_2^{LS}\right\|_{\mathbf{Q}_2}^2 \tag{40}$$

The vectors $\hat{\mathbf{e}}_1^{LS}$ and $\hat{\mathbf{e}}_2^{LS}$ correspond to the parameter estimation errors from the LS solutions of *models 1* and *2*, whereas their corresponding metrics are $\mathbf{Q}_1 = \mathbf{A}_1^T\mathbf{P}\mathbf{A}_1$ and $\mathbf{Q}_2 = \mathbf{A}_2^T\mathbf{P}\mathbf{A}_2$, respectively. According to the algebraic LS principle, we are thus faced with the paradoxical situation where the optimal choice between the two models is the one that provides us with the largest quadratic parameter error!

The above seemingly absurd result is only a manifestation of the well known fact that overparameterisation can cause serious problems in LS inversion. Indeed, an improved fit to a given data set is normally achieved by employing a more detailed model ($\mathbf{A}_1\mathbf{x}_1$) which consists of a larger number of parameters than a simpler model ($\mathbf{A}_2\mathbf{x}_2$). However, as it is seen from the preceding analysis, the more we try to improve the LS fit by expanding the parametric model, the larger become the errors in its estimated parameters.

The previous LS 'paradox' does not imply that the fitting performance of a parametric model to a given set of measurements should be excluded from the criteria that characterize an optimal data analysis procedure. Let us not forget that one of the most important aspects for performing parameter estimation is to facilitate the interpolation/extrapolation of physical processes and signals based on their observation at scattered points. Nevertheless, the trade-off between the fitting performance of a model and the accuracy of its estimated parameters suggests that a more balanced approach than LS is needed if one wants to reduce the effect of the *particular* data errors in the estimated model parameters (by 'particular' errors we mean the errors that have affected the specific data set with which we will perform the estimation).

6. A numerical example

A simple example is presented to demonstrate how a biased estimator can produce more accurate results in practice than the ordinary (unbiased) LS solution. The evaluation is performed by comparing the adjusted residuals and the parameter estimation errors that are produced from a given set of simulated data. Our example refers to a common application in local gravity field modelling, namely the representation of a gravity anomaly field within a $1^\circ \times 1^\circ$ local area in terms of an analytic functional model. In particular, the following second-order polynomial model is employed

$$\Delta g(\varphi_i, \lambda_i) = a_0 + a_1\varphi_i + a_2\lambda_i + a_3\varphi_i\lambda_i + a_4\varphi_i^2 + a_5\lambda_i^2 \tag{41}$$

Using a set of simulated observed values Δg_i at 14 points (φ_i, λ_i), we seek optimal estimates for the six polynomial coefficients $(a_0, a_1, a_2, a_3, a_4, a_5)$. The true values for the model parameters and the coordinates of the data points are given in Table 2. The zero-mean observation noise was simulated according to the error variances given in Table 2 (all observations are assumed uncorrelated).

Table 2. The coordinates of the data points, the simulated measurement noise level, and the true values of the model parameters for the local polynomial representation of the free-air gravity anomaly field

Simulated data points			Simulated noise level	True model parameters $\Delta g\ [\text{mGal}] = a_0 + a_1\varphi_i + a_2\lambda_i + a_3\varphi_i\lambda_i + a_4\varphi_i^2 + a_5\lambda_i^2$
i	$\varphi_i[°]$	$\lambda_i[°]$	$\sigma_i[\mu\text{Gal}]$	
1	20.3323	20.3551	8.2	
2	20.1236	20.1452	7.1	$a_0 = 300$
3	20.2023	20.1025	8.8	
4	20.7654	20.7695	9.3	$a_1 = -1.7$
5	20.4290	20.4262	7.5	
6	20.6532	20.6993	8.2	$a_2 = -2.5$
7	20.5656	20.5604	6.1	
8	20.6321	20.6319	8.8	$a_3 = -0.3$
9	20.6435	20.7099	6.3	
10	20.4103	20.4187	7.5	$a_4 = -0.15$
11	20.8656	20.8504	8.8	
12	20.9321	20.9419	10.3	$a_5 = 0.6$
13	20.9435	20.9399	7.5	
14	20.3103	20.3187	9.8	

The determination of the polynomial coefficients from the simulated data is performed using: (i) the ordinary LS adjustment algorithm, and (ii) the ridge regression method according to the biased estimation formula in (16). The value for the regularization parameter in the latter case was set equal to $k = 0.01$. The results obtained from the two schemes are summarized in Table 3.

For the sake of economy, we present here only the results from two simulations, yet similar results have also been obtained in numerous other test repetitions. From the values given in Table 3, it is interesting to note the significant reduction (by two orders of magnitude!) which is achieved in the parameter estimation errors $\hat{\mathbf{e}}$ when the biased estimation formula of ridge regression is used. On the other hand, the adjusted residuals $\hat{\mathbf{v}}$ remain within the same order of magnitude in both cases. Moreover, in the case of the ridge regression method, the Euclidean norm $\|\hat{\mathbf{v}}\|_{\mathbf{P}}$ of the adjusted residuals is closer to the norm of the actual data errors $\|\mathbf{v}\|_{\mathbf{P}}$, thus making the LS estimate for $\mathbf{v}$ less accurate than the corresponding ridge estimate.

In Figure 2 one can see the change in the Euclidean norms for the parameter estimation errors $\hat{\mathbf{e}}$ and the adjusted residuals $\hat{\mathbf{v}}$, as a function of the regularization parameter k used in the ridge regression formula. The particular graphs refer to the first of the two simulations shown in Table 3. The values at the origin ($k = 0$) correspond to the results obtained from the ordinary (unbiased) LS solution. Apart from the decrease in the parameter estimation errors that occurs with a biased solution for certain non-zero values of k, it is also interesting to note that the optimal behaviour for the *total uncertainty* of the results ($\|\hat{\mathbf{v}}\|_{\mathbf{P}} + \|\hat{\mathbf{e}}\|_{\mathbf{Q}}$) is not obtained from the classic LS solution; see graph (c) in Figure 2.

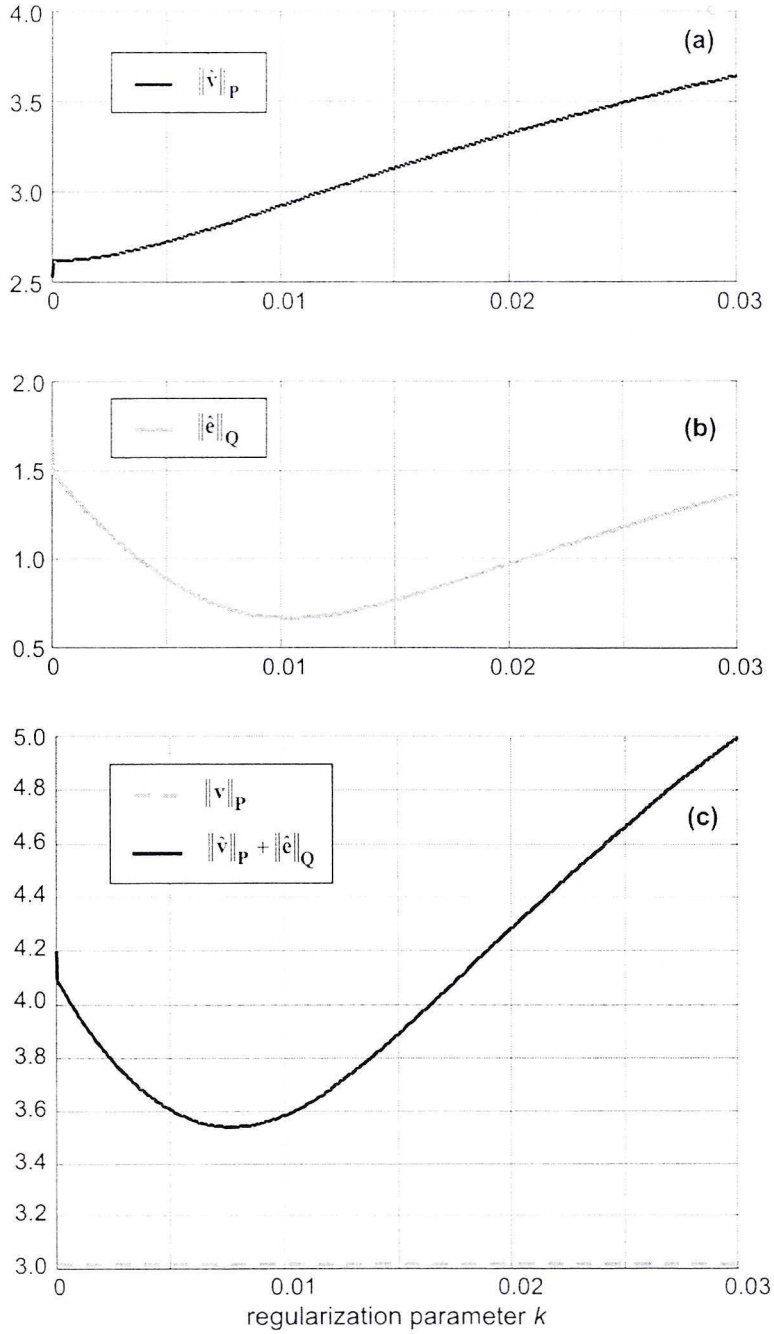

Fig. 2. Change in the Euclidean norms for the parameter estimation errors and the adjusted residuals, as a function of the regularization parameter k used in the ridge regression formula

Table 3. Results for the parameter estimation errors and the adjusted residuals that are obtained using (i) an ordinary LS solution, and (ii) a biased solution according to the simple ridge regression algorithm

Simulated errors $\mathbf{v}$		$k = 0$ (ordinary LS estimation) $\hat{\mathbf{e}} = \hat{\mathbf{x}} - \mathbf{x}$	$\hat{\mathbf{v}}$	$k = 0.01$ (ridge regression) $\hat{\mathbf{e}} = \hat{\mathbf{x}} - \mathbf{x}$	$\hat{\mathbf{v}}$
Simulation #1	$v_1 = 4.7564$		$\hat{v}_1 = -0.8359$		$\hat{v}_1 = 3.1224$
	$v_2 = 1.7023$	$e_1 = -10^4 \times 1.0060$	$\hat{v}_2 = -2.2753$	$e_1 = -10^2 \times 2.9855$	$\hat{v}_2 = 0.8366$
	$v_3 = -3.0878$	$e_2 = -10^4 \times 0.3076$	$\hat{v}_3 = -1.7532$	$e_2 = -10^2 \times 0.1657$	$\hat{v}_3 = -0.9991$
	$v_4 = 8.2965$	$e_3 = 10^4 \times 0.4066$	$\hat{v}_4 = 9.6868$	$e_3 = 10^2 \times 0.1247$	$\hat{v}_4 = 7.7370$
	$v_5 = 11.8372$	$e_4 = -10^4 \times 0.0657$	$\hat{v}_5 = 8.9627$	$e_4 = 10^2 \times 0.0012$	$\hat{v}_5 = 11.6808$
	$v_6 = -9.0870$	$e_5 = 10^4 \times 0.0402$	$\hat{v}_6 = -10.7198$	$e_5 = -10^2 \times 0.0190$	$\hat{v}_6 = -10.9749$
	$v_7 = -0.1582$	$e_6 = 10^4 \times 0.0231$	$\hat{v}_7 = -1.8330$	$e_6 = 10^2 \times 0.0108$	$\hat{v}_7 = -0.1879$
	$v_8 = -9.7735$		$\hat{v}_8 = -10.7846$		$\hat{v}_8 = -10.0981$
	$v_9 = 4.7302$	$\|\hat{\mathbf{e}}\|_Q = 1.6756$	$\hat{v}_9 = 2.4078$	$\|\hat{\mathbf{e}}\|_Q = 0.6640$	$\hat{v}_9 = 1.6334$
	$v_{10} = 3.7512$		$\hat{v}_{10} = -0.0856$		$\hat{v}_{10} = 2.9390$
	$v_{11} = -4.5519$		$\hat{v}_{11} = -0.9412$		$\hat{v}_{11} = -3.9224$
	$v_{12} = -5.7599$		$\hat{v}_{12} = 0.1618$		$\hat{v}_{12} = -6.6011$
	$v_{13} = -5.6503$		$\hat{v}_{13} = 0.2983$		$\hat{v}_{13} = -5.6801$
	$v_{14} = 9.0730$		$\hat{v}_{14} = 4.8864$		$\hat{v}_{14} = 8.2940$
	$\|\mathbf{v}\|_P = 3.0141$		$\|\hat{\mathbf{v}}\|_P = 2.5249$		$\|\hat{\mathbf{v}}\|_P = 2.9187$
Simulation #2	$v_1 = -7.1350$		$\hat{v}_1 = -8.7614$		$\hat{v}_1 = -4.3834$
	$v_2 = 5.7335$	$e_1 = -10^4 \times 1.1865$	$\hat{v}_2 = 4.5342$	$e_1 = -10^2 \times 2.9874$	$\hat{v}_2 = 8.2325$
	$v_3 = -4.4936$	$e_2 = -10^4 \times 0.4608$	$\hat{v}_3 = -2.1164$	$e_2 = 10^2 \times 0.1470$	$\hat{v}_3 = -7.5366$
	$v_4 = 6.9147$	$e_3 = 10^4 \times 0.5770$	$\hat{v}_4 = 6.4690$	$e_3 = 10^2 \times 0.1041$	$\hat{v}_4 = 6.0339$
	$v_5 = 6.3592$	$e_4 = 10^4 \times 0.5404$	$\hat{v}_5 = 2.1339$	$e_4 = 10^2 \times 0.0019$	$\hat{v}_5 = 6.0932$
	$v_6 = -6.8052$	$e_5 = -10^4 \times 0.2588$	$\hat{v}_6 = -4.5574$	$e_5 = 10^2 \times 0.0215$	$\hat{v}_6 = -4.9733$
	$v_7 = 3.2513$	$e_6 = -10^4 \times 0.2845$	$\hat{v}_7 = -0.8254$	$e_6 = -10^2 \times 0.0286$	$\hat{v}_7 = 2.1814$
	$v_8 = 9.0891$		$\hat{v}_8 = 6.0022$		$\hat{v}_8 = 8.2726$
	$v_9 = -6.6277$	$\|\hat{\mathbf{e}}\|_Q = 1.9679$	$\hat{v}_9 = 2.5066$	$\|\hat{\mathbf{e}}\|_Q = 1.0581$	$\hat{v}_9 = -2.7228$
	$v_{10} = 2.7158$		$\hat{v}_{10} = -0.7242$		$\hat{v}_{10} = 3.6747$
	$v_{11} = -0.3237$		$\hat{v}_{11} = -1.4003$		$\hat{v}_{11} = -3.6004$
	$v_{12} = -12.6446$		$\hat{v}_{12} = -7.6765$		$\hat{v}_{12} = -13.5371$
	$v_{13} = -2.0632$		$\hat{v}_{13} = 0.3038$		$\hat{v}_{13} = -4.4082$
	$v_{14} = -1.5723$		$\hat{v}_{14} = -4.7899$		$\hat{v}_{14} = -0.1984$
	$\|\mathbf{v}\|_P = 2.7980$		$\|\hat{\mathbf{v}}\|_P = 1.9883$		$\|\hat{\mathbf{v}}\|_P = 2.7371$

Acknowledgements

The authors would like to acknowledge the two reviewers and the editor-in-chief Jan Krynski for their helpful comments and suggestions.

References

Barnard G.A., (1963): *The logic of least squares*, Journal of the Royal Statistical Society, Vol. 25, No 1, pp. 124-127.

Barnett V., (1982): *Comparative Statistical Inference*, 2nd edition, John Wiley & Sons Ltd, New York.

Björck A., (1996): *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

Bouman J., Koop R., (1997): *Quality differences between Tikhonov regularization and generalized biased estimation in gradiometric analysis*, DEOS Progress Letter, Vol. 97, No 1, pp. 42-48.

Cloud M.J., Drachman B.C., (1998): *Inequalities with Applications to Engineering*, Springer Verlag, Berlin.

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

Efron B., (1975): *Biased versus unbiased estimation*, Advances in Mathematics, Vol. 16, pp. 259-277.

Eisenhart C., (1964): *The meaning of "least" in least squares*, Journal of the Washington Academy of Sciences, Vol. 54, No 2, pp. 24-33.

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

Jaynes E.T., (2003): *Probability Theory – The Logic of Science*, Cambridge University Press, Cambridge.

Liebelt P.B., (1967): *An Introduction to Optimal Estimation*, Addison-Wesley, New York.

Moritz H., (1989): *Paradoxes and curiosities in least squares*, in: Festschrift to Torben Krarup, Meddelelse No 58, Danish Geodetic Institute, Copenhagen, pp. 241-253.

Rao C.R., Toutenburg H., (1995): *Linear Models: – Least squares and alternatives*, Springer-Verlag, Berlin.

Savage L.J., (1972): *The Foundations of Statistics*, Dover, New York.

Schwarz K.P., (1979): *Geodetic improperly posed problems and their regularization*, Bollettino di Geodesia e Scienze Affini, Vol. 38, pp. 389-416.

Tukey J.W., (1960): *Where do we go from here?*, Journal of the American Statistical Association, Vol. 55, pp. 80-93.

Xu P., Rummel R., (1994): *Generalized ridge regression with applications in determination of potential fields*, Manuscripta Geodaetica, Vol. 20, pp. 8-20.

Xu P., (1998): *Truncated SVD methods for discrete ill-posed problems*, Geophysical Journal International, Vol. 135, pp. 505-514.

Rozważania dotyczące metody najmniejszych kwadratów: Dyskusja na temat podstawowej zasady estymacji w geodezji

Christopher Kotsakis[1], Michael G. Sideris[2]

[1]Wydział Geodezji i Miernictwa, Uniwersytet Arystotelesa w Salonikach
University Box 440, Thessaloniki, GR-54124, Grecja
e-mail: kotsaki@topo.auth.gr

[2]Wydział Inżynierii Geomatycznej, Uniwersytet w Calgary
2500 University Drive N.W., Calgary, Alberta T2N 1N4, Kanada
e-mail: sideris@ucalgary.ca

Streszczenie

Estymacja metodą najmniejszych kwadratów (LS) jest jednym z najważniejszych narzędzi w analizowaniu danych geodezyjnych. Jednakże powszechne korzystanie z tej metody nie zawsze idzie w parze z pełnym uświadomieniem sobie jej podstaw. W standardowym formalizmie teorii estymacji LS w rzeczywistości istnieje kilka paradoksalnych i osobliwych zagadnień rzadko formułowanych wprost. Celem niniejszej pracy jest przedstawienie niektórych z tych zagadnień i przedyskutowanie ich konsekwencji w analizie danych geodezyjnych oraz problematyce estymacji parametrów. W pierwszej części pracy przedstawiony jest alternatywny pogląd na podstawy statystyczne, które są tradycyjnie łączone z estymacją LS. W szczególności pokazano, że właściwość nieobciążoności dla zwykłych estymatorów LS może być zastąpiona

przez inne, równoważne jej uwarunkowanie, które powoduje, że zakres numeryczny nieznanych parametrów jest nieograniczony. W drugiej części pracy przedstawiono wady metody LS z czysto algebraicznego punktu widzenia, bez uwzględnienia pojęć z zakresu probabilistycznego/statystycznego teorii estymacji. W szczególności wyjaśnione zostało, do czego odnosi się 'najmniejszy' (least) w metodzie najmniejszych kwadratów. Z pewnością nie odnosi się do błędów wyznaczanych parametrów modelu. Ponadto stwierdzono, że w każdej inwersji modelu liniowego opartej na metodzie LS istnieje krytyczna zamiana pomiędzy normami euklidesowymi błędów wyznaczanych parametrów i wyrównanych residuów.