ORIGINAL ARTICLE

Least-squares prediction in linear models with integer unknowns

P. J. G. Teunissen

Abstract   The prediction of spatially and/or temporal varying variates based on observations of these variates at some locations in space and/or instances in time, is an important topic in the various spatial and Earth sciences disciplines. This topic has been extensively studied, albeit under different names. The underlying model used is often of the trend-signal-noise type. This model is quite general and it encompasses many of the conceivable measurements. However, the methods of prediction based on these models have only been developed for the case the trend parameters are real-valued. In the present contribution we generalize the theory of least-squares prediction by permitting some or all of the trend parameters to be integer valued. We derive the solution for least-squares prediction in linear models with integer unknowns and show how it compares to the solution of ordinary least-squares prediction. We also study the probabilistic properties of the associated estimation and prediction errors. The probability density functions of these errors are derived and it is shown how they are driven by the probability mass functions of the integer estimators. Finally, we show how these multimodal distributions can be used for constructing confidence regions and for cross-validation purposes aimed at testing the validity of the underlying model.

P. J. G. Teunissen (✉)
Delft Institute for Earth Observation and Space Systems (DEOS), Delft University of Technology,
Kluyverweg 1, 2629 HS Delft, The Netherlands
e-mail: P.J.G.Teunissen@tudelft.nl

1 Introduction

The topic of this contribution is the prediction of spatially and/or temporal varying variates based on observations of these variates (or functionals thereof) at some locations in space and/or instances in time. This topic has been extensively studied, albeit under different names, in the various spatial and Earth sciences disciplines. In physical geodesy it is known as least-squares collocation (LSC). Fundamental contributions to this field have been made by Krarup (1969) and Moritz (1973), see also Rummel (1976), Dermanis (1980), Sanso (1986), Grafarend and Rapp (1980). The underlying model of LSC is the so-called trend-signal-noise model. This model is quite general and it encompasses many of the conceivable geodetic measurements (Moritz 1980, p. 111). It also forms the basis of the concept of integrated geodesy as introduced in Eeg and Krarup (1973), see also Krarup (1980), Hein (1986).

Although LSC has been developed for spatially varying variates, it is closely connected with the fundamental work of Kolmogorov (1941) and Wiener (1949) on the interpolation, extrapolation and smoothing of stationary time-series. In the absence of a trend, LSC becomes the spatial analogue of Kolmogorov-Wiener prediction (Grafarend 1976; Moritz 1980). LSC also finds its analogue in Baarda's x^R-variates, which show how correlated, but free or constituent, variates are adjusted (Baarda 1968).

🖄 Springer

The trend-signal-noise model also forms the basis of prediction in geostatistics, where optimal linear prediction is called Kriging, named after Krige (1951) and further developed by Matheron (1970), see also e.g. Journel and Huijbregts (1991). When the trend is unknown it is referred to as universal Kriging and when the trend is absent or set to zero, it is called simple Kriging. In the statistical literature Kriging is called best linear unbiased prediction (Goldberger 1962). Least-squares prediction of spatially varying variates was also developed in meteorology, where it was originally referred to as objective analysis (Gandin 1963).

The above referred methods of prediction have been developed for models in which the trend parameters are real-valued. In the present contribution we will generalize the theory of least-squares prediction by permitting some or all of the trend parameters to be integer valued. Applications of such models can be found, for instance, in case of GNSS-based (GPS and/or Galileo) predictions of atmospheric fields (troposphere or ionosphere) or in case of InSAR-based predictions of deformation fields, see e.g. Odijk (2002), Hanssen et al. (2001). It is emphasized, in analogy with the trend-signal-noise model of least-squares collocation, that all trend parameters in this contribution, real-valued as well as integer-valued, are considered nonrandom. The Bayesian approach of integer estimation is treated in e.g. Betti et al. (1993), and Gundlich and Koch (2002), and the corresponding nonBayesian approach, giving identical estimates as the Bayesian approach, is based on the theory of integer equivariant estimation and is treated in Teunissen (2003).

This contribution is organized as follows. We start in Sect. 2 with a brief review of least-squares prediction. In this section, we first consider the linear model with an observable and an unobservable random vector, and then specialize to the familiar trend-signal-noise model on which LSC is based. The results of this section are used as reference for the sections following. In Sect. 3, we define and solve the problem of least-squares prediction in linear models with integer unknowns. We show that the solution has — apart from an additional computational step — the same structure as the solution of the standard least-squares prediction problem. The additional step is concerned with the estimation of the integer parameters. Apart from the integer least-squares estimator, we also consider the class of integer estimators of which the integer estimators of rounding and bootstrapping are members of. In Sect. 4, we study the probability distributions of the estimation and prediction errors. They are needed for constructing confidence regions or for validation purposes. In case of standard least-squares prediction one can do with the so-called error variances. Not so, however, in case integer parameters are involved. In that case, the second-order moments of the estimation and prediction errors can not be used to obtain an adequate quality description. We derive the density functions of the estimation and prediction errors and show how they depend on the probability mass function of the integer estimator. Finally, we show how the multimodal distribution of the prediction error can be used for cross-validation purposes. To illustrate the theory, various worked out examples are included. Two basic corollaries which are frequently used in the derivations, are given in the Appendix.

2 Least-squares prediction

2.1 Estimation and prediction

In this contribution, we speak of estimation if a function of an observable random vector y is used to guess the value of an unknown deterministic parameter vector x. If the function is given as f, then $f(y)$ is said to be the estimator of x (we call it an estimate of x if the function is taken of an outcome of y). We speak of prediction, if a function of an observable random vector y is used to guess the outcome of another random, but unobservable, vector y_0. If the function is given as g, then $g(y)$ is said to be the predictor of y_0 (we call it a prediction of y_0 if the function is taken of an outcome of y). In the following, we assume the dispersion of y and y_0 to be known, and their expectations (possibly unknown) to be linearly related to each other. Consider therefore the partitioned linear system of equations

$$\begin{bmatrix} y \\ y_0 \end{bmatrix} = \begin{bmatrix} A \\ A_0 \end{bmatrix} x + \begin{bmatrix} e \\ e_0 \end{bmatrix} \tag{1}$$

with matrices A and A_0 of order $m \times n$ and $m_0 \times n$, respectively, x a nonrandom parameter vector and $[e^T, e_0^T]^T$ a random vector, with expectation and dispersion given as,

$$E \begin{bmatrix} e \\ e_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad D \begin{bmatrix} e \\ e_0 \end{bmatrix}$$

$$= D \begin{bmatrix} y \\ y_0 \end{bmatrix} = \begin{bmatrix} Q_{yy} & Q_{yy_0} \\ Q_{y_0y} & Q_{y_0y_0} \end{bmatrix} \tag{2}$$

respectively. The matrices A and A_0 are assumed known, with A being of full column rank. Also the dispersion matrix is assumed known and to be positive definite. All the entries of the above vectors and matrices are assumed to be real-valued.

The objective function we will work with is given by the positive definite quadratic form,

$$F(y, y_0, x) = \begin{bmatrix} y - Ax \\ y_0 - A_0 x \end{bmatrix}^T \begin{bmatrix} Q_{yy} & Q_{yy_0} \\ Q_{y_0 y} & Q_{y_0 y_0} \end{bmatrix}^{-1} \begin{bmatrix} y - Ax \\ y_0 - A_0 x \end{bmatrix}$$

(3)

If y and y_0 are observable and x is unknown, then the unique $\hat{x}'$ satisfying $F(y, y_0, \hat{x}') \le F(y, y_0, x)$, for all $x \in R^n$, is said to be the least-squares estimator of x based on both y and y_0. If y is observable, x is known and y_0 is unobservable, then the unique $\hat{y}_0'$ satisfying $F(y, \hat{y}_0', x) \le F(y, y_0, x)$, for all $y_0 \in R^{m_0}$, is said to be the least-squares predictor of y_0. In the present contribution, we are interested in the case that is a combination of the previous two problems. We assume y to be observable, x to be unknown and y_0 to be unobservable. Then the unique pair $\hat{x}, \hat{y}_0$ satisfying $F(y, \hat{y}_0, \hat{x}) \le F(y, y_0, x)$, for all $x \in R^n$, $y_0 \in R^{m_0}$, is said to be the least-squares estimator-predictor pair of x, y_0.

To solve the latter problem, first note that the quadratic form in Eq. (3) can be written as a sum of two squares (see Corollary 1 of the Appendix),

$$F(y, y_0, x) = ||y - Ax||^2_{Q_{yy}} + ||y_0 - A_0 x$$
$$- Q_{y_0 y} Q_{yy}^{-1}(y - Ax)||^2_{Q_{y_0 y_0 | y}}$$

(4)

with $Q_{y_0 y_0 | y} = Q_{y_0 y_0} - Q_{y_0 y} Q_{yy}^{-1} Q_{yy_0}$ and where we used the shorthand notation $||.||^2_M = (.)^T M^{-1}(.)$. From Eq. (4) it follows that the estimator-predictor pair, $\hat{x}$ and $\hat{y}_0$, are given as,

$$\hat{x} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} y$$
$$\hat{y}_0 = A_0 \hat{x} + Q_{y_0 y} Q_{yy}^{-1}(y - A\hat{x})$$

(5)

Since $\hat{x}, \hat{y}_0$ set the second positive term of Eq. (4) equal to zero, while $\hat{x}$ minimizes the first positive term, it follows that Eq. (5) is indeed the solution to the minimization of the quadratic form of Eq. (4). It can be shown (see e.g. Koch 1980, p. 147; Teunissen et al. 2005, p. 197) that the simultaneously derived least-squares estimator-predictor pair $\hat{x}, \hat{y}_0$ constitute the BLUE (best linear unbiased estimator) and the BLUP (best linear unbiased predictor) of x and y_0, respectively. This is a consequence of having used the inverse variance matrix of $(y^T, y_0^T)^T$ as weight matrix in Eq. (3). This choice will also be used for the mixed linear model with integer parameters (see Sect. 3). As a result the derived integer estimator can be shown to have the largest possible probability of correct integer estimation.

The following additional remarks can be made with respect to Eq. (5). (i) if x would be known, then the least-squares predictor $\hat{y}_0'$ can be obtained by replacing $\hat{x}$ in the expression for $\hat{y}_0$ by the known x. (ii) Since

$F(y, \hat{y}_0, \hat{x}) \le F(y, \hat{y}_0, x)$, for all $x \in R^n$, it follows that the least-squares estimator of x remains unaffected when $\hat{y}_0$ would be taken *as if* it were the observed y_0. This is also what one would expect, since $\hat{y}_0$ should not contain information about x which is not already present in y. (iii) The first term in the expression of the least-squares predictor $\hat{y}_0$, $A_0 \hat{x}$, is the least-squares estimator of $E(y_0)$, the mean of y_0. Thus if y_0 and y are uncorrelated ($Q_{y_0 y} = 0$), then the predictor of y_0 coincides with the estimator of $E(y_0)$. (iv) If $y_0 = y$ and thus $A_0 = A$ and $Q_{y_0 y} = Q_{yy}$, then $\hat{y}_0 = y$. This shows that an observable is its own least-squares predictor.

Finally we remark that the general formulation of the model in Eqs. (1) and (2) also allows one to cover other simultaneous estimation-prediction problems. (i) Let $y = Ax' + e$, in which x' is a random vector with known variance matrix $Q_{x'x'}$ and unknown mean x, and e is a zero-mean random vector, uncorrelated with x', with known variance matrix Q_{ee}. To determine the least-squares predictor of the random vector x' with unknown mean x, we set $e \to A(x' - x) + e$, $y_0 \to x'$, $A_0 \to I$, $e_0 \to x' - x$, followed by an application of Eq. (5). As a result, the predictor of x' follows as $\hat{x}' = \hat{x} + Q_{x'x'} A^T (A Q_{x'x'} A^T + Q_{ee})^{-1}(y - A\hat{x})$. Would the unknown mean of the random vector x' be integer valued, then, as our results of Sect. 3 show, the least-squares predictor of x' becomes $\check{x}' = \check{x} + Q_{x'x'} A^T (A Q_{x'x'} A^T + Q_{ee})^{-1}(y - A\check{x})$, where $\check{x} = \arg\min_{z \in Z^n} ||\hat{x} - z||^2_{Q_{\hat{x}\hat{x}}}$, with $Q_{\hat{x}\hat{x}} = Q_{x'x'} + (A^T Q_{ee}^{-1} A)^{-1}$. (ii) Let e in $y = Ax + e$ be given as $e = Bd$, with matrix B known and where d is a zero-mean random vector with known variance matrix Q_{dd}. As an application of this formulation, the entries of d can be thought of as being the individual error components that contribute to the overall error vector e. To predict d, we set $e \to Bd$, $y_0 \to d$, $A_0 \to 0$, $e_0 \to d$, followed by an application of Eq. (5). As a result, the predictor of d follows as $\hat{d} = Q_{dd} B^T (B Q_{dd} B^T)^{-1}(y - A\hat{x})$. Note that for the special case $B = I$, we obtain the least-squares predictor of e as $y - A\hat{x}$, which is the least-squares residual. (iii) Finally we note that the so-called trend-signal-noise model is another important case for which the general formulation of Eqs. (1) and (2) with solution Eq. (5) applies. This case is worked out in more detail in the next section.

2.2 Trend, signal and noise model

An important case of the model in Eqs. (1) and (2), and one which has found wide-spread application in the spatial and Earth sciences, is the so-called trend-signal-noise model, see e.g. Moritz (1980), Stark (1987), Journel and Huijbregts (1991), Cressie (1991), Wackernagel

(1995), Torge (2001). This model is applicable to a wide range of applications for which heterogeneous, over- as well as under-determined data need to be combined, see e.g. Dermanis (1980), Rummel (1976), Sanso (1986), Grafarend and Rapp (1980). Such applications can be found in e.g. physical geodesy, mining engineering, hydrogeology, spatial data analysis, environmental engineering and digital image processing. The model also forms the basis of the concept of integrated geodesy as introduced in Eeg and Krarup (1973), see also Krarup (1980) and Hein (1986). Further examples are the determination of the geoid from gravity anomalies, see e.g. Hofmann-Wellenhof and Moritz (2005), or the determination of spherical harmonics from satellite data see e.g. Tscherning (1978), Moritz and Suenkel (1978), Sanso and Tscherning (2003).

In the trend-signal-noise model the observable vector y is written as a sum of three terms, $y = t + s + n$, with t a deterministic, but unknown trend, s a zero-mean random signal vector, and n a zero-mean random noise vector. The trend is usually further parametrized in terms of an unknown parameter vector x as $t = Ax$. The signal and noise vector are assumed to be uncorrelated and their variance matrices are given as Q_{ss} and Q_{nn}, respectively. Thus we have $y = Ax + s + n$, with $Q_{yy} = Q_{ss} + Q_{nn}$ and where the sum of signal and noise, $s + n$, plays the role of the zero-mean random vector e of Eq. (1). We may now apply Eq. (5) to separate trend, signal and noise. This gives

$$\hat{x} = \left(A^T(Q_{ss} + Q_{nn})^{-1}A\right)^{-1} A^T(Q_{ss} + Q_{nn})^{-1}y$$
$$\hat{s} = Q_{ss}(Q_{ss} + Q_{nn})^{-1}(y - A\hat{x})$$
$$\hat{n} = Q_{nn}(Q_{ss} + Q_{nn})^{-1}(y - A\hat{x})$$

The first equation follows directly from applying the first equation of Eq. (5). The second and third equation are obtained from applying the second equation of Eq. (5), by interpreting s and n, respectively, as the unobservable y_0. Thus $y_0 \rightarrow (s^T, n^T)^T$, $A_0 \rightarrow 0$, $e_0 \rightarrow (s^T, n^T)^T$ and $e \rightarrow s + n$. Note that $y = A\hat{x} + \hat{s} + \hat{n}$, which reflects the property that the observable is its own predictor. Also note that $\hat{e} = \hat{s} + \hat{n} = y - A\hat{x}$ is the predictor of e.

Often one can extend the trend-signal-noise model so as to hold true for an unobservable vector y_0 as well. This gives $y_0 = A_0x + s_0 + n_0$, in which s_0 and n_0 are uncorrelated zero-mean random vectors, with variance matrices $Q_{s_0s_0}$ and $Q_{n_0n_0}$, respectively. The two signal vectors, s_0 and s, are assumed correlated ($Q_{s_0s} \neq 0$), whereas the two noise vectors, n_0 and n, are (usually) assumed to be uncorrelated ($Q_{n_0n} = 0$). The signal plus noise vector of the unobservable vector plays now the role of the zero-mean random vector e_0 of Eq. (1). We

may now apply Eq. (5) again to predict y_0, s_0 and n_0. In this case we set $y_0 \rightarrow (y_0^T, s_0^T, n_0^T)^T$, $A_0 \rightarrow (A_0^T, 0, 0)^T$, $e_0 \rightarrow \left((s_0 + n_0)^T, s_0^T, n_0^T\right)^T$ and $e \rightarrow s + n$. This gives

$$\hat{y}_0 = A_0\hat{x} + Q_{s_0s}(Q_{ss} + Q_{nn})^{-1}(y - A\hat{x})$$
$$\hat{s}_0 = Q_{s_0s}(Q_{ss} + Q_{nn})^{-1}(y - A\hat{x})$$
$$\hat{n}_0 = 0$$

Note that the predictor of the trend plus signal, $A_0x + s_0$, is identical to the predictor of y_0. Both are given as $A_0\hat{x} + \hat{s}_0$. In general, this is not the case. In the present situation, the two predictors coincide since the noise vector n_0 was assumed to be uncorrelated with s and n. For the same reason, the predictor of n_0 is identically zero.

2.3 Estimation and prediction error

The least-squares estimator of $E(y_0)$ is $A_0\hat{x}$ and the least-squares predictor of y_0 is $\hat{y}_0$. In order to judge the quality of the estimator and of the predictor, we need to consider their errors. The estimation error of $A_0\hat{x}$ and the prediction error of $\hat{y}_0$ are defined as $\hat{e}_0 = E(y_0) - A_0\hat{x}$ and $\hat{\epsilon}_0 = y_0 - \hat{y}_0$, respectively. Both error vectors are zero-mean random vectors, $E(\hat{e}_0) = 0$ and $E(\hat{\epsilon}_0) = 0$. Thus the least-squares estimator $A_0\hat{x}$ and the least-squares predictor $\hat{y}_0$ are both unbiased. The variance matrix of $\hat{e}_0$ is called the *error* variance matrix of $A_0\hat{x}$ and the variance matrix of $\hat{\epsilon}_0$ is called the error variance matrix of $\hat{y}_0$. Since $E(y_0)$ is a nonrandom vector, the error variance matrix of $A_0\hat{x}$ is equal to the variance matrix of $A_0\hat{x}$. We have

$$Q_{\hat{e}_0\hat{e}_0} = A_0 Q_{\hat{x}\hat{x}} A_0^T \tag{6}$$

with $Q_{\hat{x}\hat{x}} = (A^T Q_{yy}^{-1} A)^{-1}$ being the variance matrix of $\hat{x}$. When the observable vector y is normally distributed, also the estimation error will be normally distributed, $\hat{e}_0 \sim N(0, Q_{\hat{e}_0, \hat{e}_0})$.

In case of the predictor $\hat{y}_0$, the error variance matrix and the variance matrix are not the same. This is due to the fact that both y_0 and $\hat{y}_0$ in $\hat{\epsilon}_0 = y_0 - \hat{y}_0$ are random vectors. In order to judge the quality of the predictor, it is the variance matrix of the prediction error that counts, rather than the variance matrix of the predictor itself. To determine the error variance matrix $Q_{\hat{\epsilon}_0\hat{\epsilon}_0}$, we first write $\hat{\epsilon}_0 = y_0 - \hat{y}_0$ as $\hat{\epsilon}_0 = (y_0 - Q_{y_0y}Q_{yy}^{-1}y) - (A_0 - Q_{y_0y}Q_{yy}^{-1}A)\hat{x}$. Note that the first bracketed term is uncorrelated with y. Since $\hat{x}$ is a linear function of y, it follows that the first bracketed term is also uncorrelated with $\hat{x}$ and therefore with the second bracketed term. Application of the variance propagation law gives

therefore

$$Q_{\hat{\epsilon}_0\hat{\epsilon}_0} = Q_{y_0y_0} - Q_{y_0y}Q_{yy}^{-1}Q_{yy_0}$$
$$+ (A_0 - Q_{y_0y}Q_{yy}^{-1}A)Q_{\hat{x}\hat{x}}(A_0 - Q_{y_0y}Q_{yy}^{-1}A)^T \quad (7)$$

The three terms on the right-hand side of this expression can be understood as follows. Should x be known and y be absent, the error variance matrix would be given as $Q_{\hat{\epsilon}_0\hat{\epsilon}_0} = Q_{y_0y_0}$. In this case, the uncertainty is completely due to the uncertainty of y_0. When the observable vector y is present and x is still known, then the error variance matrix gets reduced to $Q_{\hat{\epsilon}_0\hat{\epsilon}_0} = Q_{y_0y_0} - Q_{y_0y}Q_{yy}^{-1}Q_{yy_0}$. The uncertainty reduces due to the contribution of y. In our case however, x is unknown and has to be estimated. This implies that the error variance matrix gets enlarged by the third term in Eq. (7). With $Q_{y_0y_0|y} = Q_{y_0y_0} - Q_{y_0y}Q_{yy}^{-1}Q_{yy_0}$ and $A_{0|y} = A_0 - Q_{y_0y}Q_{yy}^{-1}A$, we can write the error variance matrix in compact form as

$$Q_{\hat{\epsilon}_0\hat{\epsilon}_0} = Q_{y_0y_0|y} + A_{0|y}Q_{\hat{x}\hat{x}}A_{0|y}^T \quad (8)$$

A complete probabilistic description of the prediction error can be given once its probability distribution is known. If we assume, in addition to Eqs. (1) and (2), that y and y_0 are normally distributed, then also the prediction error is normally distributed. Its distribution is then given as $\hat{\epsilon}_0 \sim N(0, Q_{\hat{\epsilon}_0\hat{\epsilon}_0})$.

3 Integer-based least-squares prediction

3.1 The integer-based least-squares predictor

We now extend the model of the previous section so as to include the option that some or all the parameters of x are integer valued. We therefore assume $x = (x_1^T, x_2^T)^T \in Z^p \times R^{n-p}$. Thus the first p entries of x are assumed to be unknown integers and the last $n - p$ entries are assumed to be unknown real-valued parameters. The matrices A, A_0 and $A_{0|y}$ are partitioned accordingly, $A = [A_1, A_2]$, $A_0 = [A_{01}, A_{02}]$ and $A_{0|y} = [A_{01|y}, A_{02|y}]$. We consider the same objective function $F(y, y_0, x)$, cf. Eq. (3), but with the stipulation that $x \in Z^p \times R^{n-p}$. Then the unique pair $\check{x}, \check{y}_0$ satisfying $F(y, \check{y}_0, \check{x}) \leq F(y, y_0, x)$, for all $x \in Z^p \times R^{n-p}$, $y_0 \in R^{m_0}$, is said to be the integer-based least-squares estimator-predictor pair of x, y_0.

In order to determine this pair, we again decompose the objective function $F(y, y_0, x)$ into a sum of squares. This time however, it will be decomposed into a constant term and three variable terms. We have

$$F(y, y_0, x) = ||y - Ax||^2_{Q_{yy}}$$
$$+ ||y_0 - A_0x - Q_{y_0y}Q_{yy}^{-1}(y - Ax)||^2_{Q_{y_0y_0|y}}$$
$$= ||y - A\hat{x}||^2_{Q_{yy}} + ||\hat{x} - x||^2_{Q_{\hat{x}\hat{x}}}$$
$$+ ||y_0 - A_0x - Q_{y_0y}Q_{yy}^{-1}(y - Ax)||^2_{Q_{y_0y_0|y}}$$
$$= ||y - A\hat{x}||^2_{Q_{yy}} + ||\hat{x}_1 - x_1||^2_{Q_{\hat{x}_1\hat{x}_1}}$$
$$+ ||\hat{x}_2 - x_2 - Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - x_1)||^2_{Q_{\hat{x}_2\hat{x}_2|\hat{x}_1}}$$
$$+ ||y_0 - A_0x - Q_{y_0y}Q_{yy}^{-1}(y - Ax)||^2_{Q_{y_0y_0|y}}$$
$$(9)$$

with $\hat{x} = (\hat{x}_1^T, \hat{x}_2^T)^T$ given by the first equation of Eq. (5) and $Q_{\hat{x}_2\hat{x}_2|\hat{x}_1} = Q_{\hat{x}_2\hat{x}_2} - Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}Q_{\hat{x}_1\hat{x}_2}$. The first equality of Eq. (9) is Eq. (4). With $||y - Ax||^2_{Q_{yy}} = ||y - A\hat{x}||^2_{Q_{yy}} + ||\hat{x} - x||^2_{Q_{\hat{x}\hat{x}}}$, which holds true since $A^TQ_{yy}^{-1}(y - A\hat{x}) = 0$ and $Q_{\hat{x}\hat{x}} = (A^TQ_{yy}^{-1}A)^{-1}$, the second equality of Eq. (9) follows from its first. The third equality follows then from the second since $||\hat{x} - x||^2_{Q_{\hat{x}\hat{x}}} = ||\hat{x}_1 - x_1||^2_{Q_{\hat{x}_1\hat{x}_1}} + ||\hat{x}_2 - x_2 - Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - x_1)||^2_{Q_{\hat{x}_2\hat{x}_2|\hat{x}_1}}$, the proof of which follows again from an application of Corollary 1.

Note that the last term in the third decomposition of Eq. (9) can be made zero for any $x \in Z^p \times R^{n-p}$ and that the before last term can be made zero for any $x_1 \in Z^p$. Hence, the solution for $\check{x}, \check{y}_0$ follows as

$$\check{x}_1 = \arg\min_{z \in Z^p} ||\hat{x}_1 - z||^2_{Q_{\hat{x}_1\hat{x}_1}}$$
$$\check{x}_2 = \hat{x}_2 - Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - \check{x}_1) \quad (10)$$
$$\check{y}_0 = A_0\check{x} + Q_{y_0y}Q_{yy}^{-1}(y - A\check{x})$$

Note that the structure of the predictor $\check{y}_0$ is identical to that of the predictor $\hat{y}_0$, cf. Eq. (5). That is, $\check{y}_0$ can be obtained from the expression of $\hat{y}_0$ by replacing $\hat{x}$ by $\check{x}$. Also note that $\check{x}_2$ can alternatively be expressed as $\check{x}_2 = (A_2^TQ_{yy}^{-1}A_2)^{-1}A_2^TQ_{yy}^{-1}(y - A_1\check{x}_1)$. The steps in computing $\check{y}_0$ can now be summarized as follows. First one computes the least-squares estimate $\hat{x} = [\hat{x}_1^T, \hat{x}_2^T]^T$ based on the first expression of Eq. (5). Then $\hat{x}_1$ is used to determine the integer least-squares estimate $\check{x}_1$, which is the integer minimizer of $||\hat{x}_1 - z||^2_{Q_{\hat{x}_1\hat{x}_1}}$. Finally, both $\check{x}_1$ and $\check{x}_2$ are used to compte $\check{y}_0$. Note, in case all entries of x are integer valued, that the second line of Eq. (10) is absent and the first line gets replaced by $\check{x} = \arg\min_{z \in Z^n} ||\hat{x} - z||^2_{Q_{\hat{x}\hat{x}}}$.

To see Eq. (10) at work, we consider the following examples.

Example 1 (prediction of error components) Consider the single equation

$$y = ax + e_1 + \cdots + e_q$$

with scalar a given, x an unknown integer and the e_i being q uncorrelated, zero-mean random variables. If the sum $e = \sum_{i=1}^{q} e_i$ constitutes the total measurement error, then the e_i may be considered the individual error components. The variance of e_i will be denoted as σ_i^2. We will now predict such an individual error component using Eq. (10). Note that no real-valued parameters occur in the above model. Hence, only the first and the third expression of Eq. (10) need to be applied. Setting $x_1 \to x$, $A \to a$, $y_0 \to (e_1, \ldots, e_q)^T$, $A_0 \to 0$, $e \to e_1 + \cdots + e_q$ and $e_0 \to (e_1, \ldots, e_q)^T$, yields $Q_{yy} = Q_{ee} = \sum_{j=1}^{q} \sigma_j^2$, $Q_{y_0 y} = Q_{e_0 e} = (\sigma_1^2, \ldots, \sigma_q^2)^T$, $(\check{e}_1, \ldots, \check{e}_q)^T = Q_{y_0 y} Q_{yy}^{-1}(y - a\check{x}) = (\sigma_1^2, \ldots, \sigma_q^2)^T \left(\sum_{j=1}^{q} \sigma_j^2 \right)^{-1} (y - a\check{x})$, and thus

$$\check{e}_i = \frac{\sigma_i^2}{\sum_{j=1}^{q} \sigma_j^2}(y - a\check{x})$$

with the integer least-squares estimator of x given as $\check{x} = [y/a]$, in which '$[.]$' denotes rounding to the nearest integer. This result shows that fractions of the residual $y - a\check{x}$ are assigned as predictors of the individual error components. The fractions are the ratios of the variance of the individual error component and the total variance. The predictors get an equal share of the residual when all variances are equal.

Example 2 (ionospheric prediction) Consider as a trend-signal-noise model, the single-frequency, single epoch, geometry-free GPS equations, based on double-differenced (DD) carrier phase and pseudorange,

$$y_1 = \lambda x_1 + x_2 + s + n_1$$
$$y_2 = \qquad\ + x_2 - s + n_2$$

with x_1 the unknown integer DD carrier phase ambiguity, λ the known wavelength of the carrier phase, x_2 the unknown DD range, s the residual ionospheric signal, and n_1 and n_2 the noise of the carrier phase and the pseudorange, respectively. Let σ_1^2 and σ_2^2 denote the variances of the DD carrier phase and pseudorange, respectively, and let σ_s^2 denote the variance of the ionospheric signal. Then

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} (y_1 - y_2)/\lambda \\ y_2 \end{bmatrix} \text{ and } Q_{\hat{x}\hat{x}}$$
$$= \frac{1}{\lambda^2} \begin{bmatrix} 4\sigma_s^2 + \sigma_1^2 + \sigma_2^2 & -\lambda(2\sigma_s^2 + \sigma_2^2) \\ -\lambda(2\sigma_s^2 + \sigma_2^2) & \lambda^2(\sigma_s^2 + \sigma_2^2) \end{bmatrix}$$

from which the integer-based least-squares parameter solution follows as,

$$\check{x}_1 = [\hat{x}_1]$$
$$\check{x}_2 = \hat{x}_2 + \lambda \frac{2\sigma_s^2 + \sigma_2^2}{4\sigma_s^2 + \sigma_1^2 + \sigma_2^2}(\hat{x}_1 - \check{x}_1)$$

If we want to predict the signal s_0 (e.g. the residual ionospheric delay at another time instant), then s_0 plays the role of y_0 and the integer-based least-squares predictor $\check{s}_0 = Q_{s_0 y} Q_{yy}^{-1}(y - A\check{x})$ works out as

$$\check{s}_0 = \frac{\sigma_{s_0 s}/\sigma_1^2}{1 + \sigma_s^2/\sigma_1^2 + \sigma_s^2/\sigma_2^2}\left[(y_1 - \lambda\check{x}_1 - \check{x}_2) - \frac{\sigma_1^2}{\sigma_2^2}(y_2 - \check{x}_2) \right]$$

3.2 Integer estimators

In order to study the properties of the integer-based predictor $\check{y}_0$, we need to understand the role played by the integer least-squares estimator. The integer least-squares estimator $\check{x}_1 = \arg\min_{z \in Z^p} ||\hat{x}_1 - z||^2_{Q_{\hat{x}_1 \hat{x}_1}}$ is a member of the class of integer estimators as introduced in Teunissen (1999a). For the ease of our discussion, we restrict our attention in this and the next section to the all integer case and therefore consider the class of integer estimators of which $\check{x} = \arg\min_{z \in Z^n} ||\hat{x} - z||^2_{Q_{\hat{x}\hat{x}}}$ is a member of.

Let S be a mapping from the n-dimensional space of real numbers to the n-dimensional space of integers, $S : R^n \mapsto Z^n$. Then S is a many-to-one map, implying that different real-valued vectors may be mapped to the same integer vector. Hence, we can assign a subset $S_z \subset R^n$ to each integer vector $z \in Z^n$ such that

$$S_z = \{x \in R^n \mid z = S(x)\} \tag{11}$$

Thus the subset S_z contains all real-valued vectors that are mapped by S to the same integer vector z. This subset is referred to as the *pull-in region* of z. It is the region in which all vectors are pulled to the same z. We say that S is an integer estimator if its pull-in regions satisfy the following three conditions,

1. $\underset{z \in Z^n}{\cup} S_z = R^n$
2. $\text{Int}S_{z_1} \cap \text{Int}S_{z_2} = \emptyset, \quad \forall z_1, z_2 \in Z^n, \quad z_1 \neq z_2$
3. $S_z = z + S_0, \quad \forall z \in Z^n \tag{12}$

The first condition guarantees that the union of all pull-in regions covers the n-dimensional space completely. Without this condition, gaps could occur, in which case not every real-valued vector would be assigned to a corresponding integer vector. The second condition ensures that the interiors of the pull-in regions do not overlap. Without this condition, it would not be possible to assign a real-valued vector uniquely to a single integer vector. Note that we allow the pull-in regions to have common boundaries. This is premitted if we assume zero probability that the outcome of a random vector lies on one of these boundaries. This will be the case for continuous random vectors. Finally, the third condition states that the pull-in regions are translated copies of one another.

Also this 'linearity' property is a reasonable one to ask of an integer estimator. It states that when the real-valued input is perturbed by an integer amount z, the corresponding integer solution is perturbed by the same amount. This property allows one to use the integer remove-restore technique: $S(x - z) + z = S(x)$. It therefore allows one to work with the fractional parts of a real-valued vector, instead of with its complete entries.

Using the pull-in regions, one can give an explicit expression for an integer estimator. It is given as

$$\check{x} = \sum_{z \in Z^n} z s_z(\hat{x}) \tag{13}$$

with the indicator function $s_z(\hat{x}) = 1$ if $\hat{x} \in S_z$ and $s_z(\hat{x}) = 0$ otherwise. Equation (13) shows how an integer estimator is defined through its pull-in regions. Important members of the class of integer estimators are the integer least-squares estimator, the integer bootstrapping estimator and the integer rounding estimator (Teunissen 1999b).

Integer rounding From a computational point of view, the integer rounding estimator is the simplest. It follows from rounding each of the entries of $\hat{x}$ to its nearest integer. It is given as

$$\check{x}_R = ([\hat{x}_1], \dots, [\hat{x}_n])^T \tag{14}$$

The pull-in region of this integer estimator equals the multivariate version of the unit-square,

$$S_{R,z} = \left\{ x \in R^n \mid | c_i^T(x - z) | \le \frac{1}{2}, i = 1, \dots, n \right\},$$
$$\forall z \in Z^n \tag{15}$$

where c_i is the canonical unit vector having a 1 as its ith entry and zeros otherwise, see Fig. 1a.

Integer bootstrapping Another relatively simple integer estimator is the bootstrapped estimator. The bootstrapped estimator can be seen as a generalization of integer rounding. It still makes use of integer rounding, but it also takes some of the correlation between the entries of $\hat{x}$ into account. The bootstrapped estimator follows from a sequential conditional least-squares adjustment and it is given as

$$\check{x}_B = ([\hat{x}_1], [\hat{x}_{2|1}], \dots, [\hat{x}_{n|1,\dots,n-1}])^T \tag{16}$$

where $\hat{x}_{i|1,\dots,i-1} = \hat{x}_i - \sum_{j=1}^{i-1} \sigma_{n,j|1,\dots,j-1} \sigma_{j|1,\dots,j-1}^{-2} (\hat{x}_{j|1,\dots,j-1} - \check{x}_{B,j})$ and where $\sigma_{j|1,\dots,j-1}^2$ denotes the variance of $\hat{x}_{j|1,\dots,j-1}$ and $\sigma_{i,j|1,\dots,j-1}$ denotes the covariance between $\hat{x}_i$ and $\hat{x}_{j|1,\dots,j-1}$. Note that the bootstrapped estimator reduces to the rounding estimator in case the variance matrix of $\hat{x}$ is diagonal. The pull-in region of the bootstrapped estimator equals the multivariate version of a parallellogram, see Fig. 1b. It is given as

$$S_{B,z} = \left\{ x \in R^n \mid | c_i^T L^{-1}(x - z) | \le \frac{1}{2}, i = 1, \dots, n \right\},$$
$$\forall z \in Z^n \tag{17}$$

where L denotes the unique unit lower triangular matrix of the triangular decomposition $Q_{\hat{x}\hat{x}} = LDL^T$.

Integer least-squares For the integer least-squares estimator

$$\check{x} = \arg \min_{z \in Z^n} ||\hat{x} - z||_{Q_{\hat{x}\hat{x}}}^2 \tag{18}$$

the pull-in region is given as $S_{LS,z} = \{ x \in R^n \mid ||x - z||_{Q_{\hat{x}\hat{x}}}^2 \le ||x - u||_{Q_{\hat{x}\hat{x}}}^2, \forall u \in Z^n \}$. It consists of all vectors which are closer to z than to any other integer vector of Z^n. One can give the integer least-squares pull-in region a representation that resembles the representation of the bootstrapped pull-in region. It is given as

$$S_{LS,z} = \bigcap_{c_i \in Z^n} \left\{ x \in R^n \mid | c_i^T Q_{\hat{x}\hat{x}}^{-1}(x - z) | \le \frac{1}{2} ||c_i||_{Q_{\hat{x}\hat{x}}}^2 \right\},$$
$$\forall z \in Z^n \tag{19}$$
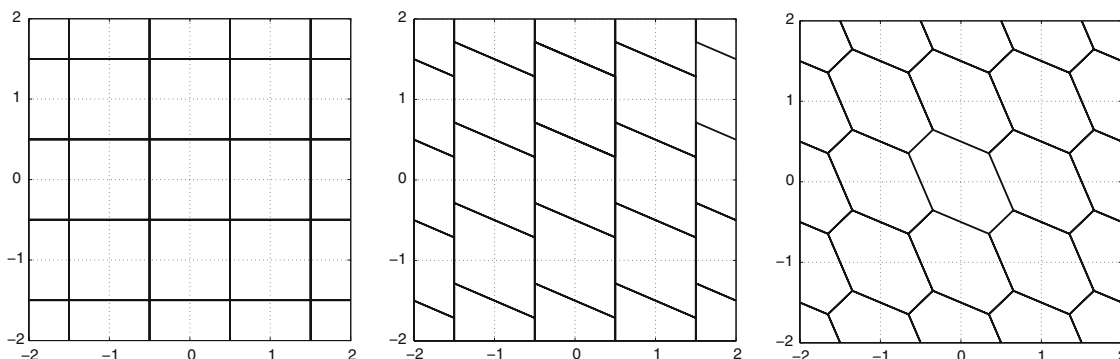
Fig. 1 The 2D pull-in regions of rounding, bootstrapping and least-squares: **a** unit square for rounding; **b** parallellogram for bootstrapping; **c** hexagon for least-squares

This shows that the pull-in region consists of intersecting half-spaces of which the planes are orthogonal to the c_i and pass through $z \pm \frac{1}{2}c_i$. They are hexagons in the two-dimensional case, see Fig. 1c.

3.3 On the computation of the integer least-squares estimator

Note that the two integer estimators $\check{x}_R$ and $\check{x}_B$ are easy to compute. That is, no search is needed to compute their integer outcomes. This is generally not true, however, for the integer least-squares estimator. The integer least-squares estimator is only easy to compute when the variance matrix of $\hat{x}$ is diagonal. In that case the integer least-squares estimator becomes identical to $\check{x}_R$ and $\check{x}_B$. When the variance matrix of $\hat{x}$ is nondiagonal, however, an integer search is needed to find the integer least-squares solution. Although we refrain from discussing the computational intricacies, the conceptual steps for computing the outcomes of the integer least-squares estimator will be briefly described. For more information we refer to e.g. Teunissen (1993, 1995) and de Jonge and Tiberius (1996a) or to the textbooks Hofmann-Wellenhof et al. (1997), Strang and Borre (1977), Teunissen and Kleusberg (1998), Leick (2004) and Misra and Enge (2001) in which the LAMBDA-method for solving the integer least-squares problem is described. Practical results obtained with the method can be found, for example, in Boon and Ambrosius (1997), Boon et al. (1997), Chang et al. (2005), Cox and Brading (1999), Dai et al. (2005), de Jonge and Tiberius (1996b), de Jonge and Tiberius (1996), Han (1995), Moenikes et al. (2005), Peng et al. (1999), Svendsen (2005), Tiberius and de Jonge (1995), Tiberius et al. (1997).

The first step for finding the integer least-squares solution is to define the integer search space

$$\Omega_z = \{z \in Z^n \mid (\hat{x} - z)^T Q_{\hat{x}\hat{x}}^{-1}(\hat{x} - z) \leq \chi^2\} \quad (20)$$

with χ^2 a chosen positive constant. The boundary of this search space is ellipsoidal. It is centred at $\hat{x}$, its shape is governed by the variance matrix $Q_{\hat{x}\hat{x}}$ and its size is determined by χ^2. The constant χ^2 is chosen such that $\check{x} \in \Omega_z$. Since an elongated shape of the search space usually hinders the computational efficiency of the search, the search space is transformed to a more spherical shape by means of an integer transformation $\hat{z} = Z^T\hat{x}$ and $Q_{\hat{z}\hat{z}} = Z^T Q_{\hat{x}\hat{x}} Z$. In order to preserve the integer nature in this transformation, matrix Z and its inverse need to have integer entries. The transformed search space is then given as $\Phi_z = \{z \in Z^n \mid (\hat{z}-z)^T Q_{\hat{z}\hat{z}}^{-1}(\hat{z}-z) \leq \chi^2\}$.

The two search spaces Ω_z and Φ_z contain the same number of integer vectors.

To search for the integer vectors inside Φ_z, the triangular decomposition of $Q_{\hat{z}\hat{z}}$ is used to write the search space defining quadratic inequality as a sum-of-squares:

$$\sum_{i=1}^{n} \frac{(\hat{z}_{i|I} - z_i)^2}{\sigma_{i|I}^2} \leq \chi^2 \quad (21)$$

with $I = \{1, \ldots, i-1\}$. On the left-hand side one recognizes the conditional least-squares estimator $\hat{z}_{i|I}$, which follows when the conditioning takes place on the integers $z_1, z_2, \ldots, z_{i-1}$. Using the sum-of-squares structure, one can set up the n scalar intervals which are used for the search. These sequential intervals are given as

$$(\hat{z}_1 - z_1)^2 \leq \sigma_1^2 \chi^2, \ (\hat{z}_{2|1} - z_2)^2 \leq \sigma_{2|1}^2 \left(\chi^2 - \frac{(\hat{z}_1 - z_1)^2}{\sigma_1^2}\right),$$

etc. These intervals are used to collect the integer vectors which lie inside the search space. From this set the solution $\check{z}$ is identified as the one which returns the smallest value for $||\hat{z} - z||_{Q_{\hat{z}\hat{z}}}^2$. The sought for integer least-squares solution follows then finally as $\check{x} = Z^{-T}\check{z}$.

4 Distribution of estimation and prediction error

4.1 The probability mass function of integer estimators

In this section, we will derive the probability distributions of the estimation and prediction errors for the case some or all of the entries of the parameter vector x are estimated as integers. Similar to before, we define the estimation error of $A_0\check{x}$ as $\check{e}_0 = E(y_0) - A_0\check{x}$ and the prediction error of $\check{y}_0$ as $\check{\epsilon}_0 = y_0 - \check{y}_0$. In order to determine their distributions, we first need to determine the distribution of the integer estimator $\check{x}_1$. We allow $\check{x}_1$ to be any integer estimator of $x_1 \in Z^p$. Here and in the following we will assume that y and y_0 are normally distributed. Then the input of the integer estimator is normally distributed too, $\hat{x}_1 \sim N(x_1, Q_{\hat{x}_1\hat{x}_1}), x_1 \in Z^p$.

The distribution of $\check{x}_1 \in Z^p$ will be a probability mass function (PMF). Let the S_{z_1}, with $z_1 \in Z^p$, be the pull-in regions of the integer estimator $\check{x}_1$ (e.g. one of the three types as given in Sect. 3.2). Then $\check{x}_1 = z_1 \Leftrightarrow \hat{x}_1 \in S_{z_1}$ and therefore $P[\check{x}_1 = z_1] = P[\hat{x}_1 \in S_{z_1}]$, where the notation $P[E]$ is used to denote the probability of the event E. Hence, the PMF of $\check{x}_1$ follows as

$$P[\check{x}_1 = z_1] = \int_{S_{z_1}} f_{\hat{x}_1}(u)du, \quad z_1 \in Z^p \quad (22)$$

with the probability density function (PDF) of $\hat{x}_1$ given as

$$f_{\hat{x}_1}(u) = \frac{1}{(2\pi)^{p/2}\sqrt{\det Q_{\hat{x}_1\hat{x}_1}}} \exp\left\{-\frac{1}{2}||u - x_1||^2_{Q_{\hat{x}_1\hat{x}_1}}\right\}$$

Thus the PMF of the integer estimator follows from integrating the PDF of $\hat{x}_1$ over the translational invariant pull-in regions. The PMF can be used to study the properties of the integer estimator $\check{x}_1$. Since the PMF is symmetric with respect to x_1 ($P[\check{x}_1 = x_1 + z] = P[\check{x}_1 = x_1 - z]$ for all $z \in Z^p$), it follows that the integer estimator is unbiased, $E(\check{x}_1) = x_1$. This property holds true for all integer estimators and thus also for integer rounding, integer bootstrapping and integer least-squares. It can also be shown that of all integer estimators, the integer least-squares estimator has the largest possible probability of correct integer estimation (Teunissen 1999b). Thus the probability $P[\check{x}_1 = x_1]$ is maximized when the integer least-squares estimator is used. This estimator is therefore the best estimator of its class. This optimality property has its price however. As we have seen, the computation of the integer least-squares solution is not as straightforward as the computation of the integer rounding solution or the integer bootstrapped solution. In some applications one might therefore still decide to opt for integer rounding or integer bootstrapping, in particular if their PMF's are not too different from the integer least-squares' PMF.

4.2 The PDF of the estimation error

The estimation error is defined as $\check{e}_0 = A_0(x - \check{x})$. If we substitute the second expression of Eq. (10) into $\check{e}_0 = A_0(x - \hat{x}) + A_0(\hat{x} - \check{x}) = \hat{e}_0 + A_0(\hat{x} - \check{x})$, we obtain

$$\check{e}_0 = \hat{e}_0 + (A_{01} + A_{02}Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1})(\hat{x}_1 - \check{x}_1) \quad (23)$$

This expression shows how the difference between the two estimation errors is driven by the residual vector $\hat{x}_1 - \check{x}_1$. Note that this residual vector always lies inside the origin centred pull-in region of the integer estimator. Thus $\hat{x}_1 - \check{x}_1 \in S_0 \subset R^p$. This implies that the difference between the two estimation errors, $\check{e}_0$ and $\hat{e}_0$, is bounded. Also note that the covariance between $\hat{e}_0 = A_0(x - \hat{x}) = A_{01}(x_1 - \hat{x}_1) + A_{02}(x_2 - \hat{x}_2)$ and $\hat{x}_1$ is given as $Q_{\hat{e}_0\hat{x}_1} = -(A_{01}Q_{\hat{x}_1\hat{x}_1} + A_{02}Q_{\hat{x}_2\hat{x}_1})$. Hence, the relation between the two estimation errors can alternatively be expressed as

$$\check{e}_0 = \hat{e}_0 - Q_{\hat{e}_0\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - \check{x}_1) \quad (24)$$

In order to determine the PDF of $\check{e}_0$, we need to take the distributions of $\hat{e}_0$, $\hat{x}_1$ and $\check{x}_1$ into account. Since y is nomally distributed, also $\hat{e}_0$ and $\hat{x}_1$ are normally

distributed. But since the integer estimator $\check{x}_1$ is not normally distributed, also the estimation error $\check{e}_0$ will not be normally distributed. Note, however, if we replace the random integer vector $\check{x}_1$ in Eq. (24) by the nonrandom integer vector z_1, that the result will be normally distributed. We denote this random vector as

$$\check{e}_0|_{\hat{x}_1 = z_1} = \hat{e}_0 - Q_{\hat{e}_0\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - z_1) \quad (25)$$

The distribution of this conditional random vector is given as

$$\hat{e}_0|_{\hat{x}_1 = z_1} \sim N\left(E(\hat{e}_0|_{\hat{x}_1 = z_1}), Q_{\hat{e}_0\hat{e}_0|\hat{x}_1}\right) \quad (26)$$

with mean $E(\hat{e}_0|_{\hat{x}_1 = z_1}) = (A_{01} + A_{02}Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1})(x_1 - z_1)$ and variance matrix $Q_{\hat{e}_0\hat{e}_0|\hat{x}_1} = A_{02}Q_{\hat{x}_2\hat{x}_2|\hat{x}_1}A_{02}^T$, as is easily verified by means of Corollary 2 of the Appendix. Note, upon comparing Eq. (24) with Eq. (25), that one can expect the PDF of $\check{e}_0$ to approach the normal distribution of $\hat{e}_0|_{\hat{x}_1 = z_1}$ if the probability $P[\check{x}_1 = z_1]$ approaches one.

We will use the distribution of $\hat{e}_0|_{\hat{x}_1 = z_1}$ to determine the PDF of $\check{e}_0$. We start from the probability $P[\check{e}_0 \in \Omega]$, in which Ω is an arbitrary subset of R^{m_0}. Since the pull-in regions $S_{z_1} \subset R^p$ form a partition of R^p, we may use the total probability rule to write

$$P[\check{e}_0 \in \Omega] = \sum_{z_1 \in Z^p} P[\check{e}_0 \in \Omega|\hat{x}_1 \in S_{z_1}]P[\hat{x}_1 \in S_{z_1}] \quad (27)$$

With Eq. (24) and $\hat{x}_1 \in S_{z_1} \Leftrightarrow \check{x}_1 = z_1$, we have $P[\check{e}_0 \in \Omega|\hat{x}_1 \in S_{z_1}] = P[\hat{e}_0 - Q_{\hat{e}_0\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - \check{x}_1) \in \Omega|\check{x}_1 = z_1] = P[\hat{e}_0 - Q_{\hat{e}_0\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - z_1) \in \Omega|\check{x}_1 = z_1] = P[\hat{e}_0|_{\hat{x}_1 = z_1} \in \Omega|\check{x}_1 = z_1]$. Note that $\hat{e}_0|_{\hat{x}_1 = z_1}$ is uncorrelated with $\hat{x}_1$ (cf. Corollary 2). Hence, since it is normally distributed, it is independent of $\hat{x}_1$ and thus also independent of $\check{x}_1$. We therefore have $P[\check{e}_0 \in \Omega|\hat{x}_1 \in S_{z_1}] = P[\hat{e}_0|_{\hat{x}_1 = z_1} \in \Omega]$. Substituting this into Eq. (27) gives, together with $P[\hat{x}_1 \in S_{z_1}] = P[\check{x}_1 = z_1]$, $P[\check{e}_0 \in \Omega] = \sum_{z_1 \in Z^p} P[\hat{e}_0|_{\hat{x}_1 = z_1} \in \Omega]P[\check{x}_1 = z_1]$. Since this result holds true for any Ω, the PDF of the estimation error follows as

$$f_{\check{e}_0}(v) = \sum_{z_1 \in Z^p} f_{\hat{e}_0|\hat{x}_1 = z_1}(v)P[\check{x}_1 = z_1] \quad (28)$$

in which $f_{\hat{e}_0|\hat{x}_1 = z_1}(v)$ is the PDF of $\hat{e}_0|_{\hat{x}_1 = z_1}$. Note, although the PDF $f_{\hat{e}_0|\hat{x}_1 = z_1}(v)$ and the PMF $P[\check{x}_1 = z_1]$ both depend on the unknown parameter x_1, that the PDF $f_{\check{e}_0}(v)$ is independent of x_1. This is due to the fact that x_1 is integer and the summation in Eq. (28) covers the whole of Z^p.

The result Eq. (28) shows that the PDF of the estimation error is a multimodal distribution. It is an infinite sum of weighted and shifted versions of the PDF of $\hat{e}_0|_{\hat{x}_1 = z_1}$. The weights are given by the probability masses of the PMF of $\check{x}_1$. Note that the PDF is symmetric with

respect to the origin. Hence, the estimation error has a zero mean, $E(\check{e}_0) = 0$. Also note, if the probability of correct integer estimation approaches one, that the PDF approaches the normal distribution of $\hat{e}_0|_{\hat{x}_1 = x_1}$. Finally note, in case all entries of x are integers and estimated as such, that the PDF of the estimation error reduces to a PMF. In that case we have $\check{e}_0 = A_0(x - \check{x})$, with the PMF $P[\check{e}_0 = A_0 z] = P[\check{x} = x - z]$ for all $z \in Z^n$.

The PDF Eq. (28) can be used to describe the quality of any estimator $A_0\check{x}$. Let us illustrate its application with a simple example.

Example 3 (estimation of carrier phase) Consider again the single-frequency, single epoch, geometry-free GPS-model (see Example 2), but now with the ionospheric signal absent (that is, the two GPS receivers are assumed to be close enough to neglect the ionospheric delay). Then $\sigma_{\hat{x}_1}^2 = (\sigma_1^2 + \sigma_2^2)/\lambda^2$ and the PMF $P[\check{x}_1 = x_1 - z_1]$ follows as function of z_1 as $F(z_1) = \Phi\left(\frac{(1+2z_1)}{2\sigma_{\hat{x}_1}}\right) + \Phi\left(\frac{(1-2z_1)}{2\sigma_{\hat{x}_1}}\right) - 1$, with $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}w^2\}dw$. If we want to consider the estimation error of the carrier phase, we set $A_0 = [\lambda, 1]$. Then the mean and variance of the conditional PDF $f_{\hat{e}_0|\hat{x}_1 = x_1 - z_1}(v)$ are given as $m(z_1) = (A_{01} + A_{02}Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1})z_1 = \frac{\lambda\sigma_1^2}{\sigma_1^2+\sigma_2^2}z_1$ and $\sigma^2 = A_{02}Q_{\hat{x}_2\hat{x}_2|\hat{x}_1}A_{02}^T = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$, respectively. Hence, the PDF of the carrier phase estimation error follows then as

$$f_{\check{e}_0}(v) = \sum_{z_1 \in Z} \frac{F(z_1)}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{v - m(z_1)}{\sigma}\right)^2\right\}$$

4.3 The PDF of the prediction error

For the prediction error $\check{\epsilon}_0 = y_0 - \check{y}_0$ we can derive similar results. First we will derive the counterpart of Eq. (24) for the prediction error. If we substitute the second expression of Eq. (10) into $\check{\epsilon}_0 = y_0 - \hat{y}_0 + \hat{y}_0 - \check{y}_0 = \hat{\epsilon}_0 + (A_0 - Q_{y_0y}Q_{yy}^{-1}A)(\hat{x} - \check{x}) = \hat{\epsilon}_0 + A_{0|y}(\hat{x} - \check{x})$, we obtain $\check{\epsilon}_0 = \hat{\epsilon}_0 + (A_{01|y} + A_{02|y}Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1})(\hat{x}_1 - \check{x}_1)$. And since the covariance between $\hat{\epsilon}_0$ and $\hat{x}_1$ is given as $Q_{\hat{\epsilon}_0\hat{x}_1} = -(A_{01|y}Q_{\hat{x}_1\hat{x}_1} + A_{02|y}Q_{\hat{x}_2\hat{x}_1})$, we find that

$$\check{\epsilon}_0 = \hat{\epsilon}_0 - Q_{\hat{\epsilon}_0\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - \check{x}_1) \tag{29}$$

Compare with Eq. (24). That $Q_{\hat{\epsilon}_0\hat{x}_1} = -(A_{01|y}Q_{\hat{x}_1\hat{x}_1} + A_{02|y}Q_{\hat{x}_2\hat{x}_1})$ can be seen as follows. We have $\hat{\epsilon}_0 = y_0 - \hat{y}_0 = y_0|_{y=0} - A_{0|y}\hat{x}$, with $y_0|_{y=0} = y_0 - Q_{y_0y}Q_{yy}^{-1}y$ and $A_{0|y} = A_0 - Q_{y_0y}Q_{yy}^{-1}A$. And since $y_0|_{y=0}$ and $\hat{x}$ are uncorrelated, it follows that $Q_{\hat{\epsilon}_0\hat{x}_1} = -A_{0|y}Q_{\hat{x}\hat{x}_1}$.

As was the case with the estimation error $\check{e}_0$, also the distribution of the prediction error $\check{\epsilon}_0$ will not be normal

due to the presence of $\check{x}_1$ in Eq. (29). But if we replace, as before, the random integer vector $\check{x}_1$ by the nonrandom integer vector z_1, we obtain a result that is normally distributed. We denote this random vector as

$$\hat{\epsilon}_0|_{\hat{x}_1 = z_1} = \hat{\epsilon}_0 - Q_{\hat{\epsilon}_0\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(\hat{x}_1 - z_1) \tag{30}$$

Its distribution is given as

$$\hat{\epsilon}_0|_{\hat{x}_1 = z_1} \sim N\left(E(\hat{\epsilon}_0|_{\hat{x}_1 = z_1}), Q_{\hat{\epsilon}_0\hat{\epsilon}_0|\hat{x}_1}\right) \tag{31}$$

with mean $E(\hat{\epsilon}_0|_{\hat{x}_1 = z_1}) = -Q_{\hat{\epsilon}_0\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}(x_1 - z_1) = (A_{01|y} + A_{02|y}Q_{\hat{x}_2\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1})(x_1 - z_1)$ and variance matrix $Q_{\hat{\epsilon}_0\hat{\epsilon}_0|\hat{x}_1} = Q_{\hat{\epsilon}_0\hat{\epsilon}_0} - Q_{\hat{\epsilon}_0\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}Q_{\hat{x}_1\hat{\epsilon}_0} = Q_{y_0y_0|y} + A_{0|y}(Q_{\hat{x}\hat{x}} - Q_{\hat{x}\hat{x}_1}Q_{\hat{x}_1\hat{x}_1}^{-1}Q_{\hat{x}_1\hat{x}})A_{0|y}^T = Q_{y_0y_0|y} + A_{02|y}Q_{\hat{x}_2\hat{x}_2|\hat{x}_1}A_{02|y}^T$. In the expression of the variance matrix we made use of Eq. (8) and of $Q_{\hat{\epsilon}_0\hat{x}_1} = -A_{0|y}Q_{\hat{x}\hat{x}_1}$. Compare Eq. (31) with Eq. (26). Note that $\hat{\epsilon}_0|_{\hat{x}_1 = z_1}$ is independent of $\hat{x}_1$. This property will be used to obtain the PDF of $\check{\epsilon}_0$.

To obtain the PDF of the prediction error $\hat{\epsilon}_0$, we can follow a similar derivation as used for the PDF of the estimation error. We start with

$$P[\check{\epsilon}_0 \in \Omega] = \sum_{z_1 \in Z^p} P[\check{\epsilon}_0 \in \Omega|\hat{x}_1 \in S_{z_1}]P[\hat{x}_1 \in S_{z_1}] \tag{32}$$

Using the same arguments as before, we have $P[\check{\epsilon}_0 \in \Omega|\hat{x}_1 \in S_{z_1}] = P[\hat{\epsilon}_0|_{\hat{x}_1 = z_1} \in \Omega]$ and $P[\hat{x}_1 \in S_{z_1}] = P[\check{x}_1 = z_1]$, and therefore $P[\check{\epsilon}_0 \in \Omega] = \sum_{z_1 \in Z^p} P[\hat{\epsilon}_0|_{\hat{x}_1 = z_1} \in \Omega]P[\check{x}_1 = z_1]$. Since this result holds true for any Ω, the PDF of the prediction error follows as

$$f_{\check{\epsilon}_0}(v) = \sum_{z_1 \in Z^p} f_{\hat{\epsilon}_0|\hat{x}_1 = z_1}(v)P[\check{x}_1 = z_1] \tag{33}$$

in which $f_{\hat{\epsilon}_0|\hat{x}_1 = z_1}(v)$ is the PDF of $\hat{\epsilon}_0|_{\hat{x}_1 = z_1}$. Thus also the PDF of the prediction error is a multimodal distribution. And as the PDF of the estimation error, also the PDF of the prediction error is symmetric with respect to the origin. Thus $E(\check{\epsilon}_0) = 0$. The PDF approaches the normal distribution of $\hat{\epsilon}_0|_{\hat{x}_1 = x_1}$, if the probability of correct integer estimation approaches one. To illustrate the use of Eq. (33), we consider two examples.

Example 4 (prediction in all integer case) As a special case of Eq. (33), assume that all entries of x are integers and estimated as such. We then need to consider the distribution of $\check{\epsilon}_0|_{\hat{x} = z}$, instead of the distribution of $\check{\epsilon}_0|_{\hat{x}_1 = z_1}$. Since $\check{\epsilon}_0|_{\hat{x} = z} \sim N\left(A_{0|y}(x - z), Q_{y_0y_0|y}\right)$ (compare with Eq. (31)), the PDF of the prediction error

follows as

$$f_{\check{\epsilon}_0}(v) = \sum_{z \in Z^n} \frac{P[\check{x} = x - z]}{(2\pi)^{m_0/2}\sqrt{\det Q_{y_0y_0|y}}}$$
$$\times \exp\left\{-\frac{1}{2}||v - A_{0|y}z||^2_{Q_{y_0y_0|y}}\right\} \qquad (34)$$

Note, since x is integer, that the PDF is independent of this unknown mean of $\hat{x}$.

Example 5 (prediction of deformation rate) Another example of a trend-signal-noise model $y = Ax + s + n$ with integer parameters is given by the application of stacked radar interferometry for deformation monitoring, see e.g. (Hanssen et al. 2001). In this case, y is the vector of observed phase differences over time between two permanent scatterers (expressed in units of range), x is the unknown vector of integer ambiguities, $s = av$ is the signal vector with known coefficient vector $a = [\Delta t_1/2, \ldots, \Delta t_m/2]^T$ and zero-mean deformation rate v, and n is the noise vector. The Δt_i are time intervals (expressed in years) between the current time and a reference time and matrix A is given as $A = \lambda I_m$, with λ the known wavelength. With the variance of v given as σ_v^2 and the noise variance matrix given as $Q_{nn} = \sigma_\phi^2 I_m$, we have $Q_{yy} = Q_{ss} + Q_{nn} = \sigma_v^2 aa^T + \sigma_\phi^2 I_m$.

When the goal is to predict the deformation rate v, the vector y_0 takes the form $y_0 = (a^Ta)^{-1}a^Ts = v$ and its predictor is given as $\check{v} = \check{y}_0 = Q_{y_0y}Q_{yy}^{-1}(y - A\check{x}) = (a^Ta)^{-1}a^TQ_{ss}(Q_{ss} + Q_{nn})^{-1}(y - A\check{x})$, which gives

$$\check{v} = \frac{\sigma_v^2/\sigma_\phi^2}{1 + a^Ta\sigma_v^2/\sigma_\phi^2}a^T(y - \lambda\check{x})$$

The PDF of the prediction error of the deformation rate is then given by Eq. (34), with $A_{0|y} = -\lambda\frac{\sigma_v^2/\sigma_\phi^2 a^T}{1+a^Ta\sigma_v^2/\sigma_\phi^2}$ and $Q_{y_0y_0|y} = \frac{\sigma_v^2}{1+a^Ta\sigma_v^2/\sigma_\phi^2}$.

4.4 Cross-validation

The PDF of the estimation error $\check{e}_0$ and the PDF of the prediction error $\check{\epsilon}_0$ can be used to obtain confidence regions. Note that in contrast with the results of Sect. 2, one can now not rely anymore on the error variance matrices. That is, information on the error variance matrices of $A_0\check{x}$ and $\check{y}_0$ is not sufficient to determine the confidence regions. To determine the confidence regions, one will need the complete multimodal PDFs.

The prediction error and its distribution can also be used for validation purposes, i.e. to validate the assumptions underlying the model of integer-based estimation and prediction. However, in order to be able to execute the validation, a sample of the prediction error is needed. This implies, since the prediction $\check{y}_0$ is given, that a sample of y_0 is needed. This is the concept of cross-validation, see e.g. Stone (1974) and Geisser (1975). That is, the outcome of the model, based on one part of the data, is used to predict the outcome of the other part. The underlying model is then considered suspect, if the difference between the predicted and observed value should be considered a rare event in light of the distribution of the prediction error. Note that this is very similar to Baarda's (1968) concept of datasnooping for blunder detection. In that case, one observation at a time is excluded from the model and confronted with its prediction based on the remaining observations.

Assuming that a sample of the prediction error is available, the question of determining the acceptance and rejection regions should be answered. Answering this question is made difficult by the multimodality of the distribution of the prediction error. Let $\Omega \subset R^{m_0}$ be the acceptance region with coverage probability $P[\check{\epsilon}_0 \in \Omega] = 1 - \alpha$. Thus the test leads to rejection if $\check{\epsilon}_0 \notin \Omega$. Since we want the rejection to be rare when the underlying model is correct, the false alarm probability α is chosen as a small value. But since there are an infinite number of subsets that can produce this false alarm probability, we still need to determine a way of defining a proper Ω. It seems reasonable to define the optimal subset as the one which has the smallest volume. In that case the probability $1 - \alpha$ would be the most concentrated. To determine the optimal Ω we thus need to solve the minimization problem,

$$\min_{\Omega \subset R^{m_0}} V_\Omega \quad \text{subject to} \quad P[\check{\epsilon}_0 \in \Omega] = 1 - \alpha \qquad (35)$$

where V_Ω denotes the volume of Ω. The solution to this problem is given by the subset

$$\Phi = \{v \in R^{m_0} \mid f_{\check{\epsilon}_0}(v) \geq \lambda\} \qquad (36)$$

where λ is chosen so as to satisfy the given probability constraint of Eq. (35). The proof goes as follows. Starting form the probability constraint $\int_\Phi f_{\check{\epsilon}_0}(v)dv = \int_\Omega f_{\check{\epsilon}_0}(v)dv$, we have
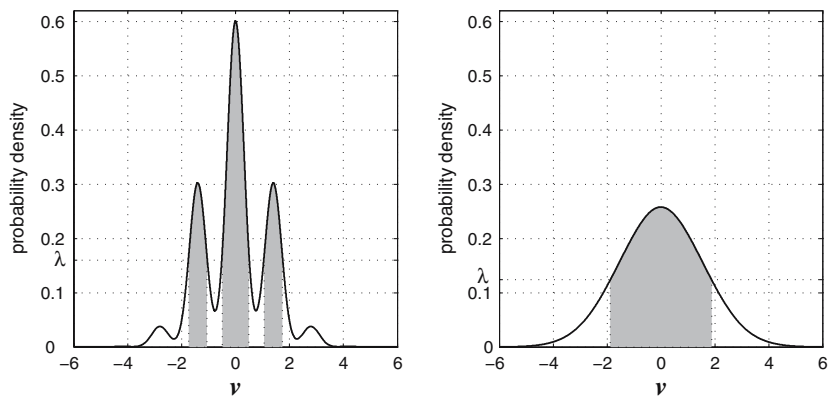
$$0 = \int_\Phi f_{\check{\epsilon}_0}(v)dv - \int_\Omega f_{\check{\epsilon}_0}(v)dv = \int_{\Phi/\Omega} f_{\check{\epsilon}_0}(v)dv - \int_{\Omega/\Phi} f_{\check{\epsilon}_0}(v)dv$$

$$\geq \lambda \int_{\Phi/\Omega} dv - \lambda \int_{\Omega/\Phi} dv$$

$$\geq \lambda\left(\int_\Phi dv - \int_\Omega dv\right)$$

and therefore $V_\Phi \leq V_\Omega$ for any $\Omega \subset R^{m_0}$. Note, that a similar derivation can be used to show that Φ is the solution to the maximization problem $\max_\Omega P[\check{\epsilon}_0 \in \Omega]$ subject to V_Ω = constant. Thus of all subsets with the same volume, Φ captures the largest possible probability mass.

Due to the multimodality of the PDF, it is a nontrivial task to decide for a certain given α whether or not an observed sample of $\check{\epsilon}_0$ leads to rejection. The complication resides in the direct determination of λ from α. This complication can be avoided, however, if we make use of a Monte Carlo based approach. The computational steps for executing the test are then as follows. Given the observed sample of the prediction error, say $\check{\epsilon}_0^*$, one first computes $\lambda^* = f_{\check{\epsilon}_0}(\check{\epsilon}_0^*)$. This implies that the sample would lie on the boundary of the acceptance region if λ in Eq. (36) would be set equal to λ^*. Hence, this subset is given as $\Phi^* = \{v \in R^{m_0} \mid f_{\check{\epsilon}_0}(v) \geq \lambda^*\}$. The next step is then to compute the value of α that would correspond with λ^*: $\alpha^* = 1 - P[\check{\epsilon}_0 \in \Phi^*]$. Here the simulation enters. Let N be the number of times a sample is generated from $f_{\check{\epsilon}_0}(v)$ and let N_i be the number of times a generated sample lies in Φ. Then α^* can be approximated as $\alpha^* = 1 - N_i/N$. The decision to accept or reject the observed sample is then based on the difference between α and α^*. If $\alpha^* < \alpha$, then the test leads to rejection, otherwise it leads to acceptance.

Example 6 (continuation of Example 2) We determine the PDF of the prediction error of the predicted signal $\check{s}_0$ of Example 2. The PMF $P[\check{x}_1 = z]$ of integer rounding is given by the function $F(z) = \Phi\left(\frac{(1+2(x_1-z))}{2\sigma_{\hat{x}_1}}\right) + \Phi\left(\frac{(1-2(x_1-z))}{2\sigma_{\hat{x}_1}}\right) - 1$, with $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}w^2\}dw$ and $\sigma_{\hat{x}_1}^2 = (4\sigma_s^2 + \sigma_1^2 + \sigma_2^2)/\lambda^2$. The function $F(z)$ is symmetric with respect to x_1 and its shape is governed by $\sigma_{\hat{x}_1}$. The smaller this standard deviation is, the more peaked the PMF is. For $\sigma_{\hat{x}_1} < 0.10$, one will have $P[\check{x}_1 = x_1] \approx 1$.

The PDF of the prediction error follows as

$$f_{\check{e}_0}(v) = \sum_{z \in Z} \frac{F(z)}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{v - m(z)}{\sigma}\right)^2\right\}$$

with $m(z) = -\frac{2\lambda\sigma_{s_0 s}}{4\sigma_s^2 + \sigma_1^2 + \sigma_2^2}(x_1 - z)$ and $\sigma^2 = \sigma_{s_0}^2 - \frac{4\sigma_{s_0 s}^2}{4\sigma_s^2 + \sigma_1^2 + \sigma_2^2}$. Again note that the error PDF is independent of the unknown integer mean x_1. Figure 2 shows for two different values of σ the error PDF, together with their corresponding $1 - \alpha$ acceptance regions.

5 Summary and conclusions

In this contribution we have generalized the theory of least-squares prediction so as to be able to deal with integer-real mixed linear models. Starting from the model

$$\begin{bmatrix} y \\ y_0 \end{bmatrix} = \begin{bmatrix} A \\ A_0 \end{bmatrix} x + \begin{bmatrix} e \\ e_0 \end{bmatrix} \tag{37}$$

in which y is observable, y_0 is unobservable and x is unknown, the least-squares estimator–predictor pair of x, y_0 was derived as

$$\begin{aligned} \hat{x} &= (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} y \\ \hat{y}_0 &= A_0 \hat{x} + Q_{y_0 y} Q_{yy}^{-1}(y - A\hat{x}) \end{aligned} \tag{38}$$

for the case all entries of x are real-valued ($x \in R^n$), and as

$$\begin{aligned} \check{x}_1 &= \arg\min_{z \in Z^p} ||\hat{x}_1 - z||_{Q_{\hat{x}_1 \hat{x}_1}}^2 \\ \check{x}_2 &= \hat{x}_2 - Q_{\hat{x}_2 \hat{x}_1} Q_{\hat{x}_1 \hat{x}_1}^{-1}(\hat{x}_1 - \check{x}_1) \\ \check{y}_0 &= A_0 \check{x} + Q_{y_0 y} Q_{yy}^{-1}(y - A\check{x}) \end{aligned} \tag{39}$$

for the case the first p entries of x are integer-valued, while the remaining are real-valued ($x = (x_1^T, x_2^T)^T \in Z^p \times R^{n-p}$). It was shown that the general model formulation of Eq. (37) and its two sets of solutions

Fig. 2 PDF of the prediction error and corresponding $1 - \alpha$ acceptance region for $\sigma = 0.31$ **a** and $\sigma = 1.00$ **b**

Eq. (38) and Eq. (39), can be used to tackle a wide range of different estimation-prediction problems, one of which is the celebrated trend-signal-noise model that forms the basis of least-squares collocation. For instance, by setting $e \rightarrow s + n$, $y_0 \rightarrow (y_0^T, s_0^T, n_0^T)^T$, $A_0 \rightarrow (A_0^T, 0, 0)^T$ and $e_0 \rightarrow (e_0^T, s_0^T, n_0^T)^T$, followed by an application of Eq. (38), one obtains the well-known results of least-squares collocation. Application of Eq. (39) now extends least-squares collocation to the real-integer mixed parameter case.

In order to judge the quality of estimation and prediction, we need to consider the probabilistic properties of the estimation and prediction errors. For the all-real case, Eq. (38), these errors are defined as $\hat{e}_0 = A_0(x - \hat{x})$ and $\hat{\epsilon}_0 = y_0 - \hat{y}_0$, respectively, and for the mixed case, Eq. (39), they are defined as $\check{e}_0 = A_0(x - \check{x})$ and $\check{\epsilon}_0 = y_0 - \check{y}_0$, respectively. For the all-real case, the errors are normally distributed if y and y_0 are normally distributed. Their distributions are then given as

$$\hat{e}_0 \sim N(0, Q_{\hat{e}_0\hat{e}_0} = A_0 Q_{\hat{x}\hat{x}} A_0^T)$$
$$\hat{\epsilon}_0 \sim N(0, Q_{\hat{\epsilon}_0,\hat{\epsilon}_0} = Q_{y_0y_0|y} + A_{0|y} Q_{\hat{x}\hat{x}} A_{0|y}^T) \quad (40)$$

This shows that it are the error variance matrices $Q_{\hat{e}_0\hat{e}_0}$ and $Q_{\hat{\epsilon}_0,\hat{\epsilon}_0}$ which completely drive the probabilistic properties of the errors. Thus in the case of standard least-squares collocation, for instance, it indeed suffices to work with the error variances. This is, however, not allowed anymore in case some of the parameters are integer-valued. That is, the error variances will not suffice for describing the quality of integer-based least-squares collocation.

In the integer-real mixed parameter case, the estimation and prediction errors will not be normally distributed, even if y and y_0 are normally distributed. It was shown that the distributions of these errors are given by the multimodal PDFs

$$f_{\check{e}_0}(v) = \sum_{z_1 \in \mathbb{Z}^p} f_{\hat{e}_0|\hat{x}_1=z_1}(v) P[\check{x}_1 = z_1] \quad \text{and} \quad f_{\check{\epsilon}_0}(v)$$
$$= \sum_{z_1 \in \mathbb{Z}^p} f_{\hat{\epsilon}_0|\hat{x}_1=z_1}(v) P[\check{x}_1 = z_1] \quad (41)$$

in which $P[\check{x}_1 = z_1]$ is the PMF of $\check{x}_1$, and $f_{\hat{e}_0|\hat{x}_1=z_1}(v)$ and $f_{\hat{\epsilon}_0|\hat{x}_1=z_1}(v)$ are the PDFs of the normally distributed vectors

$$\hat{e}_0|_{\hat{x}_1=z_1} \sim N\left(\left(A_{01} + A_{02} Q_{\hat{x}_2\hat{x}_1} Q_{\hat{x}_1\hat{x}_1}^{-1}\right)(x_1 - z_1),\right.$$
$$\left. Q_{\hat{e}_0\hat{e}_0|\hat{x}_1} = A_{02} Q_{\hat{x}_2\hat{x}_2|\hat{x}_1} A_{02}^T\right)$$

$$\hat{\epsilon}_0|_{\hat{x}_1=z_1} \sim N\left(\left(A_{01|y} + A_{02|y} Q_{\hat{x}_2\hat{x}_1} Q_{\hat{x}_1\hat{x}_1}^{-1}\right)(x_1 - z_1),\right.$$
$$\left. Q_{\hat{\epsilon}_0\hat{\epsilon}_0|\hat{x}_1} = Q_{y_0y_0|y} + A_{02|y} Q_{\hat{x}_2\hat{x}_2|\hat{x}_1} A_{02|y}^T\right)$$

The multimodal PDFs of Eq. (41) are infinite sums of weighted and shifted versions of the PDFs of $\hat{e}_0|_{\hat{x}_1=z_1}$ and $\hat{\epsilon}_0|_{\hat{x}_1=z_1}$, respectively. The weights are given by the probability masses of the PMF of $\check{x}_1$. Since the multimodal PDFs are symmetric with respect to the origin, the estimation and prediction errors are zero-mean random vectors, $E(\check{e}_0) = 0$ and $E(\check{\epsilon}_0) = 0$.

The above multimodal PDFs can be used to construct confidence regions and to cross-validate the assumptions underlying the model with integer parameters. The procedure for cross-validation was described for a given false alarm probability and an acceptance region which captures the largest probability mass.

Appendix

Corollary 1 *A quadratic form*

$$F = \begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} Q_{aa} & Q_{ab} \\ Q_{ba} & Q_{bb} \end{bmatrix}^{-1} \begin{bmatrix} a \\ b \end{bmatrix} \quad (42)$$

can be decomposed into a sum of two quadratic forms as

$$F = a^T Q_{aa}^{-1} a + (b - Q_{ba} Q_{aa}^{-1} a)^T$$
$$\times (Q_{bb} - Q_{ba} Q_{aa}^{-1} Q_{ab})^{-1} (b - Q_{ba} Q_{aa}^{-1} a) \quad (43)$$

Proof The following block triangular factorization is easily verified,

$$\begin{bmatrix} Q_{aa} & Q_{ab} \\ Q_{ba} & Q_{bb} \end{bmatrix} = \begin{bmatrix} I & 0 \\ Q_{ba}Q_{aa}^{-1} & I \end{bmatrix} \begin{bmatrix} Q_{aa} & 0 \\ 0 & Q_{bb} - Q_{ba}Q_{aa}^{-1}Q_{ab} \end{bmatrix}$$
$$\begin{bmatrix} I & 0 \\ Q_{ba}Q_{aa}^{-1} & I \end{bmatrix}^T \quad (44)$$

The inverse of this matrix follows, with

$$\begin{bmatrix} I & 0 \\ Q_{ba}Q_{aa}^{-1} & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -Q_{ba}Q_{aa}^{-1} & I \end{bmatrix}, \quad (45)$$

as

$$\begin{bmatrix} Q_{aa} & Q_{ab} \\ Q_{ba} & Q_{bb} \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -Q_{ba}Q_{aa}^{-1} & I \end{bmatrix}^T$$
$$\times \begin{bmatrix} Q_{aa}^{-1} & 0 \\ 0 & (Q_{bb} - Q_{ba}Q_{aa}^{-1}Q_{ab})^{-1} \end{bmatrix}$$
$$\times \begin{bmatrix} I & 0 \\ -Q_{ba}Q_{aa}^{-1} & I \end{bmatrix} \quad (46)$$

Substitution of this expression into the quadratic form F of Eq. (42) gives the desired decomposition. $\square$

Corollary 2 *Let a and b be two random vectors with joint variance matrix*

$$D\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} Q_{aa} & Q_{ab} \\ Q_{ba} & Q_{bb} \end{bmatrix} \quad (47)$$

and define the random vector

$$b|_{a=c} = b - Q_{ba}Q_{aa}^{-1}(a - c) \qquad (48)$$

in which c is a constant vector. Then

(i) $\quad E(b|_{a=c}) = E(b) - Q_{ba}Q_{aa}^{-1}(E(a) - c)$

(ii) $\quad Q_{b|_{a=c}b|_{a=c}} = Q_{bb} - Q_{ba}Q_{aa}^{-1}Q_{ab}$

(iii) $\quad Q_{ab|_{a=c}} = 0$ (*a and $b|_{a=c}$ are uncorrelated*)

Proof The expectation (*i*) follows directly from applying the mean propagation law to Eq. (48). To prove (*ii*) and (*iii*), consider the transformation

$$\begin{bmatrix} a - c \\ b|_{a=c} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -Q_{ba}Q_{aa}^{-1} & I \end{bmatrix} \begin{bmatrix} a - c \\ b \end{bmatrix}$$

Application of the variance propagation law gives

$$\begin{bmatrix} Q_{aa} & Q_{ab|_{a=c}} \\ Q_{b|_{a=c}a} & Q_{b|_{a=c}b|_{a=c}} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -Q_{ba}Q_{aa}^{-1} & I \end{bmatrix} D \begin{bmatrix} a \\ b \end{bmatrix} \\ \times \begin{bmatrix} I & 0 \\ -Q_{ba}Q_{aa}^{-1} & I \end{bmatrix}^T$$

Substitution of Eq. (47), gives with the use of Eqs. (44) and (45),

$$\begin{bmatrix} Q_{aa} & Q_{ab|_{a=c}} \\ Q_{b|_{a=c}a} & Q_{b|_{a=c}b|_{a=c}} \end{bmatrix} = \begin{bmatrix} Q_{aa} & 0 \\ 0 & Q_{bb} - Q_{ba}Q_{aa}^{-1}Q_{ab} \end{bmatrix}$$

This proves (*ii*) and (*iii*). In the sequel we will use the shorthand notation $Q_{bb|_{a=c}}$ for $Q_{b|_{a=c}b|_{a=c}}$. □

References

Baarda W (1968) A testing procedure for use in Geodetic Networks. Netherlands Geodetic Commission, Publications on Geodesy, New Series, vol. 2, No. 5

Betti B, Crespi M, Sanso F (1993) A geometric illustration of ambiguity resolution in GPS theory and a Bayesian approach. Manuscr Geod 18:317–330

Boon F, Ambrosius B (1997) Results of real-time applications of the LAMBDA method in GPS based aircraft landings. In: Proceedings KIS97, pp 339–345

Boon F, de Jonge PJ, Tiberius CCJM (1997) Precise aircraft positioning by fast ambiguity resolution using improved troposphere modelling. In: Proceedings ION GPS-97. vol. 2, pp 1877–1884

Chang XW, Yang X, Zhou T (2005) MLAMBDA: a modified LAMBDA method for integer ambiguity determination. Proceedings ION GNSS2005, Long Beach

Cox DB, Brading JDW (1999) Integration of LAMBDA ambiguity resolution with Kalman filter for relative navigation of spacecraft. In: Proceedings ION NTM 99, pp 739–745

Cressie N (1991) Statistics for spatial data. Wiley, New York

Dai L, Nagarajan N, Hu G, Ling K (2005) Real-time attitude determination for micro satellites by LAMBDA method combined with Kalman filtering. In: AIAA proceedings 22nd ICSSC, Monterey

Dermanis A (1980) Adjustment of geodetic observations in the presence of signals. In: Proceedings of the international school of advanced geodesy. Bollettino di Geodesia e Scienze Affini. Vol. 38, pp 419–445

de Jonge PJ, Tiberius CCJM (1996a) The LAMBDA method for integer ambiguity estimation: implementation aspects. Publications of the Delft Computing Centre, LGR-Series No. 12

de Jonge PJ, Tiberius CCJM (1996b) Integer estimation with the LAMBDA method. In: Beutler G et al. (eds) Proceedings IAG symposium No. 115, 'GPS trends in terrestrial, airborne and spaceborne applications'. Springer, Heidelberg, pp 280–284

de Jonge PJ, Tiberius CCJM, Teunissen PJG (1996) Computational aspects of the LAMBDA method for GPS ambiguity resolution. In: Proceedings ION GPS-96, pp 935–944

Eeg J, Krarup T (1973) Integrated geodesy. Dan Geod Inst, Int Rep 7, Copenhagen

Gandin LS (1963) Objective analysis of meteorological fields. Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad

Geisser S (1975) The predictive sample reuse method with applications. J Math Statist 28:385–394

Goldberger AS (1962) Best linear unbiased prediction in the generalized linear regression model. J Amer Statist Assoc 57:369–375

Grafarend EW (1976) Geodetic applications of stochastic processes. Phys Earth Planet Interiors 12:151–179

Grafarend EW, Rapp RH (eds) (1980) Advances in geodesy. Selected papers from Rev. Geophys. Space Phys., Richmond, Virg., Am Geophys Union, Washington

Gundlich B, Koch KR (2002) Confidence regions for GPS baselines by Bayesian statistics. J Geod 76:55–62

Han S (1995) Ambiguity resolution techniques using integer least-squares estimation for rapid static or kinematic positioning. In: Symposium Satellite Navigation Technology: 1995 and beyond, Brisbane, Australia, 10 p

Hanssen RF, Teunissen PJG, Joosten P (2001) Phase ambiguity resolution for stacked radar interferometric data. In: Proc KIS2001, international symposium on kinematic systems in geodesy, Geomatics and Navigation, Banff, pp 317–320

Hein GW (1986) Integrated geodesy. In: Suenkel H (ed), Mathematical and numerical techniques in physical geodesy. Lecture Notes in Earth Sciences vol. 7. Springer, Heidelberg, pp 505–548

Hofmann-Wellenhof B, Moritz H (2005) physical geodesy. Springer Wien, New York

Hofmann-Wellenhof B, Lichtenegger H, Collins J (1997) Global positioning system: theory and practice, 4th edn. Springer, Berlin

Journel AG, Huijbregts ChJ (1991) Mining geostatistics. Academic, New York

Koch KR (1980) Parameterschaetzung und Hypothesentests in linearen Modellen. Dummler, Bonn

Kolmogorov AN (1941) Interpolation and extrapolation of stationary random sequences. Izvestiia Akademii Nauk SSSR, Seriia Matematicheskiia 5:3–14

Krarup T (1969) A contribution to the mathematical foundation of physical geodesy. Publ. Danish Geod. Inst. 44, Copenhagen

Krarup T (1980) Integrated geodesy. In: Proceedings of the international school of advanced geodesy. Bollettino di Geodesia e Scienze Affini 38:480–496

Krige D (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. J Chem Metall Mining Soc South Afr 52:119–139

Leick A (2004) GPS satellite surveying, 3rd edn. Wiley, New York

Matheron G (1970) The theory of regionalized variables and its applications. Fascicule 5, Les Cahiers du Centre de Morphologie Mathematique. Ecole des Mines de Paris, Fontainebleau, 211 p

Misra P, Enge P (2001) Global positioning system: signals, measurements, and performance. Ganga-Jamuna Press

Moenikes R, Wendel J, Trommer GF (2005) A modified LAMBDA method for ambiguity resolution in the presence of position domain constraints. In: Proceedings ION GNSS2005. Long Beach

Moritz H (1973) Least-squares collocation. DGK, A 59, Muenchen

Moritz H (1980) Advanced physical geodesy. Herbert Wichmann Verlag Karlsruhe

Moritz H, Suenkel H (eds) (1978) Approximation methods in geodesy. Sammlung Wichmann Neue Folge, Band 10, Herbert Wichmann Verlag

Odijk D (2002) Fast precise GPS positioning in the presence of ionospheric delays. Publications on Geodesy, vol. 52, Netherlands Geodetic Commission, Delft

Peng HM, Chang FR, Wang LS (1999) Attitude determination using GPS carrier phase and compass data. In: Proceedings of ION NTM 99, pp 727–732

Rummel R (1976) A model comparison in least-squares collocation. Bull Geod 50:181–192

Sanso F (1986) Statistical methods in physical geodesy. In: Suenkel H (ed), Mathematical and numerical techniques in physical geodesy, Lecture Notes in Earth Sciences, vol. 7 Springer, pp 49–156

Sanso F, Tscherning CC (2003) Fast spherical collocation: theory and examples. J Geod 77(1-2):101–112

Stark H (ed) (1987) Image recovery: theory and application. Academic, New York

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J R Statist Soc B 36:111–133

Strang G, Borre K (1997) Linear algebra, geodesy, and GPS. Wellesley-Cambridge Press

Svendsen JGG (2005) Some properties of decorrelation techniques in the ambiguity space. GPS Solutions. doi:10.1007/s10291-005-0004-6

Teunissen PJG (1993) Least-squares estimation of the integer GPS ambiguities. Invited Lecture, Section IV Theory and Methodology, IAG General Meeting, Beijing, China, August 1993. Also in: LGR Series. No. 6, Delft Geodetic Computing Centre

Teunissen PJG (1995) The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. J Geod 70:65–82

Teunissen PJG (1999a) The probability distribution of the GPS baseline for a class of integer ambiguity estimators. J Geod 73:275–284

Teunissen PJG (1999) An optimality property of the integer least-squares estimator. J Geod 73:587–593

Teunissen PJG (2003) Theory of integer equivariant estimation with application to GNSS. J Geod 77:402–410

Teunissen PJG, Kleusberg A (eds) (1998) GPS for geodesy, 2nd enlarged edition. Springer, Heidelberg

Teunissen PJG, Simons DG, Tiberius CCJM (2005) Probability and observation theory. Lecture Notes Delft University of Technology, 364 p

Tiberius CCJM, de Jonge PJ (1995) Fast positioning using the LAMBDA method. In: Proceedings DSNS-95, paper 30, 8 p

Tiberius CCJM, Teunissen PJG, de Jonge PJ (1997) Kinematic GPS: performance and quality control. In: Proceedings KIS97, pp 289–299

Torge W (2001) Geodesy, 3rd edn. Walter de Gruyter

Tscherning CC (1978) Collocation and least-squares methods as a tool for handling gravity field dependent data obtained through space research techniques. Bull Geod 52:199–212

Wackernagel H (1995) Multivariate geostatistics. Springer, Heidelberg

Wiener N (1949) Extrapolation, interpolation, and smoothing of stationary time series. MIT Press, Cambridge