NOTE 1 Version one posted in the Internet on 22.01.2010*

NON-LINEAR ESTIMATION EMPLOYING INTERMEDIATE VARIABLES AND THE USE OF THE CRAMER-RAO BOUND

Luc Fraiture†

1. Introduction

The purpose of intermediate variables is to transform a system of non-linear equations in another non-linear equation system, where unknowns and stochastic input data (e.g. measurements) cannot be separated in the former system and can everywhere be separated in the transformed system. The non-separability in the original system assumes that the number of scalar unknowns and stochastic scalar input parameters remains unchanged in separation attempts. We call this 'conventional non-separability'. The proposed separability we address here, is achieved by an almost trivial artifice consisting in replacing a measurement m_i, mixed up with unknown(s) in an inseparable manner by a new unknown y_i, 'an intermediate variable', which is redefined by a supplementary trivial measurement equation $y_i = m_i$.

The price for the increase of the size of the system is largely compensated by the advantage of being able to separate measurements and unknowns into different equation members, because this yields the ability to perform an optimal Gauss-Markov estimation where measurements, equated to intermediate variables, are moved into a linear context, thereby decreasing the biasedness of the estimates. On top of this the error variance information of the stochastic input is disentangled, leading to an improved ability to obtain estimates whose accuracies statistically approach the Cramer-Rao bound more closely than competing estimation techniques. This has been exemplified by the author in a demanding satellite attitude determination problem[1]. By the way, intermediate variables can also be used, for instance, in the context of maximum likelihood estimation (MLE) and different forms of sequential estimation techniques. This is to say that intermediate variables are a support to estimation, but the basic theory of estimation itself is not touched.

---

* The author introduced intermediate variables for the first time in a numerical example added to his conference paper nr. 10 in 1991. Though he first employed the name 'intermediate variables' in a paper which had been submitted for publication 1997. This paper is largely reproduced in the second note.

† Private contribution; address: Luc Fraiture, Lucasweg 6, D-64287 Darmstadt, Germany, e-mail: lucfraiture@web.de

Our terminology combined with a mathematical representation, and the actual expansion of conventionally non-separable stochastic equation systems into separated intermediate variable systems is given in the next section.

We also address the Cramèr-Rao bound (CRB) in this note, because it yields an important performance reference in estimation. Moreover, in the limited amount of literature we consulted we did not find any presentation which explains the way to compute the CRB in practice when facing a conventionally non-separable overdetermined equation system: the application field of intermediate variables. This lack of readily accessible information was also observed for non-linear estimation involving constraint equations. Also this useful alternative has been included in section three.

The intermediate variables have no purpose if the stochastic equations are all linear. If there is just one unknown one will usually be able to separate the unknowns from the stochastic input. If this is not the case, one is normally able to find the solution by simple algorithms not involving linearization like, for instance, the method of the chord or the bisection method. This is thus an interesting case where accuracy performances can be compared between a 'direct' method and intermediate variable-use. Such a 'one unknown case' is presented in the last section to introduce the practical aspects and highlight the performance of intermediate variables.

2. THE INTERMEDIATE VARIABLE TRANSFORMATION

In this section we propose (without loss of generality) to keep non-linearity of the overdetermined non-separated systems considered to a minimum and therefore require that every equation is at most (in-homogeneously) bilinear, more specifically that each of them is separately linear in the p *basic or essential unknowns* $\mathbf{x_1}' = |x_1, \ldots, x_p|$, as well as in the q *indirect measurements or non-canonical random variables* $\mathbf{y}' = |y_1, \ldots, y_q|$ and the r *direct measurements or canonical random variables* $\mathbf{u}' = |u_1, \ldots, u_r|$, where primes denote transposition. The i-th generic equation of such a system then looks as follows:

$$\sum_{j=1}^{p} f_{ij}(y_1, \ldots, y_q)\, x_j \;+\; f_{i0}(y_1, \ldots, y_q) \;=\; g_i(u_1, \ldots, u_r) \qquad (i = 1, \ldots, N) \qquad (1)$$

where f and g are linear functions of the random variables. We derive the denomination *non-canonical variables* from the fact that if all $\partial f_{ij}/\partial y_k \equiv 0$ for all valid i, j and k except $j = 0$, the system is canonical (or separated by assumption) and allows the application of a Gauß-Markov estimation applicable to canonical equation systems in the sense defined by Bard[2]. To transform the previous system in a new but now canonical one, we introduce the *intermediate variables* $x_{p+1}, \ldots, x_{p+q}$ grouped in $\mathbf{x_2}$, which we define by the q trivial random equations $x_{p+s} = y_s$ with $s = 1, \ldots, q$. These new variables are introduced as auxiliary unknowns and substituted in all equations of shape (1). In this way, a new but

canonical system is created comprising $q + N = n$ equations and $p + q = m$ unknowns. The intended transformation is not always legitimate, because the system may become singular if, for instance, for a larger number of values of i, the functions g_i are identically zero, too many constraint equations are created in which all random input variables have disappeared. These conditions are studied hereafter.

The first trivial condition is that the matrix $F(N \times p)$ containing the scalar elements f_{ij} excluding f_{i0}, all filled in with the measurements, has rank N. Agreeing that constants in (1) are exclusively included in the functions f_{i0} we observe that by assumption

$$g_i(\mathbf{u}) = \sum_{j=1}^{r} b_{ij}\, u_j$$

with known coefficients b_{ij} for which we introduce the vector equation

$$\mathbf{G} = B\,\mathbf{u} \tag{2}$$

where $B(N \times r)$ is the matrix with the elements b_{ij}. If s is the rank of B and $s \leq r < N$, we can apply a linear transformation to (1) such that the last $N - s$ lines of B are nullified. At any rate, the system defined in (1) is not affected in its information by a full rank $(N \times N)$ linear transformation and remains bilinear as specified. Applying such a transformation is a prerequisite in order to be able to treat the system correctly. Without loss of generality we may simplify life by assuming that **G'** is transformed into $|\mathbf{u}', \mathbf{0'}_{N-s}|$, where $\mathbf{0_{N-s}}$ is a $N - s$ dimensional zero vector. After this step has been performed, we end up with the following system of equations:

$$\mathbf{x_2} = \mathbf{y} \tag{3}$$

$$F_1(\mathbf{x_2})\,\mathbf{x_1} + \mathbf{F_{01}}(\mathbf{x_2}) = \mathbf{u} \tag{4}$$

$$F_2(\mathbf{x_2})\,\mathbf{x_1} + \mathbf{F_{02}}(\mathbf{x_2}) = \mathbf{0}_{N-s} \tag{5}$$

where (3) and (4) are random equations and (5) represents $N - s = \ell$ constraints. Although almost self explanatory, we note that the matrices $F_1(s \times p)$ and $F_2(\ell \times p)$ contain the scalar functions f_{ij} after transformation excluding the functions f_{i0}, which on the other hand are packed into the vectors $\mathbf{F_{01}}(s \times 1)$ and $\mathbf{F_{02}}(\ell \times 1)$ after transformation.

From a pure mathematical standpoint it is necessary that the last ℓ equations are algebraically independent. Further, the number ℓ of independent constraints cannot be larger than n, because otherwise the system is overdetermined independent of the random input. Even if $\ell = m = p+q$, the system is still 'determined' independent of random input. If $N < \ell$ there must have been some conceptual mistake when setting up the system. To allow a Gauß-Markov estimation applied to the system(3) to (5) we finally also require the

availability of the non-singular covariance matrices $C_y(q \times q)$ for the random variables in $\mathbf{y}$ and $C_u(r \times r)$ for the random variables in $\mathbf{u}$.

If there are no canonical variables on top of the non-canonical random variables in (1), there is no vector equation (4) in the random overdetermined system and then Gauß-Markov estimation is theoretically equivalent to global least squares in the bilinear case (namely the case we will treat in Note 2). If however there are canonical random variables in parallel to non-canonical variables in a system with or without constraints, or the system non-linearity goes beyond bilinearity, the introduction of the intermediate variables untangles the covariance information of the non-canonical random variables. This yields probably the most favorable presentation of the statistical properties belonging to the stochastic input allowing an optimal weighing in the subsequent estimation. Hitherto this could not be achieved by global least squares or any other method known to the author.

3. DERIVING THE CRAMER-RAO BOUND

The well known Cramer-Rao bound (CRB) is the value of the theoretically absolute lowest possible variance(s) one can achieve in an estimation. This bound has a certain degree of generality as one can read in chapter 8 of the book by Wasan[3]. Here it is limited to overdetermined equation systems with and without constraint equations subjected to Gauss-Markov estimation. The ability to compute the CRB, is based on the availability of the covariance matrix of the stochastic input and the existence of an error free point satisfying exactly all equations – stochastic or not – involved in the system. These error free values of the error free point involve all measurements and unknowns. In practice these values are not available, otherwise we would not need an estimation. In a simulation context, on the contrary, we have them.

Hereafter, we will work out the numerical derivation of the CRB without constraints in a first step and with constraints in a second step. It should be clear, that there is no point to introduce the intermediate variables for the purpose of computing the CRB, because the CRB is a value at the ideal solution and there intermediate variables can rigorously be eliminated. Restricted to common cases, we will ignore all pathological situations one could imagine mathematically and propose the following heuristic approach.

3.1 Unconstrained Systems

Consider a system of N equations, of the generic shape:

$$\phi_i\left(x_1, \ldots, x_p, m_1, \ldots, m_q\right) = 0 \qquad (i = 1, \ldots, N) \tag{6}$$

without the introduction of any intermediate variables. The parameter x_j is an unknown with $0 < j \leq p$ and $p < N$, while m_k represents a given (stochastic) measurement with $0 < k \leq q$ and $N \leq q$ and the corresponding known full rank $(q \times q)$ covariance matrix Q_m. The inequality $N \leq q$ is a necessary condition to prevent the presence of implicit or

explicit constraint equations in the system. Unfortunately, this is not a sufficient condition. In order not to be led astray into the discussion of a number of side issues which are meaningful in a rigorous formulation, we just assume that such a constraint is not implied in the system. The potential presence of a finite number of ambiguous solutions, is another theme which we do not address. We assume that that the 'error free values', satisfying the equations represented by (6), are $(x_{1*}, \ldots, x_{p*}, m_{1*}, \ldots, m_{q*})$ and that the equations display a continuous and differentiable behavior in a reasonable interval around this error free point. For these error free values we may further assume that $m_{k*} = \mathrm{E}(m_k)$. Due to the non-linearity we can, for whatever (optimal) estimation approach, not guaranty that x_{j*} is equal to $x_{j0} = \mathrm{E}(x_j)$, where x_j is an estimation result obtained by the selected estimation method. In other words, it is likely that estimates are biased. Nevertheless, we can expand (6) around the error free point, and this expansion to the first order reads:

$$\begin{aligned} \phi_i\,(x_1, \ldots, x_p, m_1, \ldots, m_q) = \phi_{i*} &+ \sum_{j=1}^{p} (x_j - x_{j*}) \Big(\frac{\partial \phi_i}{\partial x_j}\Big)_* \\ &+ \sum_{k=1}^{q} (m_k - m_{k*}) \Big(\frac{\partial \phi_i}{\partial m_k}\Big)_* \end{aligned} \tag{7}$$

where an asterisk is added as subscript to functions to indicate that their value is taken at the error free point. Thus, ϕ_* is zero by definition. When the left hand side of (7) is considered in an estimation context, it should also be equal to zero, but not being at the error free point there is an error caused by the measurements. We can group these errors into the q-dimensional vector $\mathbf{\Delta m}$, whose components are $\delta m_k = m_k - m_{k*}$. To achieve zero for all ϕ_i as closely as possible, these measurement errors have to be compensated in the best possible manner by a deviation away from the error free unknowns, which we represent by the p-dimensional vector $\mathbf{\Delta x}$ with the components $\delta x_j = x_j - x_{j*}$. Consequently, we can write down the corresponding generic equation for the unknown $\mathbf{\Delta x}$ in the presence of known errors $\mathbf{\Delta m}$, namely:

$$-\sum_{j=1}^{p} \delta x_j \, \Big(\frac{\partial \phi_i}{\partial x_j}\Big)_* \; = \; \sum_{k=1}^{q} \delta m_k \, \Big(\frac{\partial \phi_i}{\partial m_k}\Big)_*$$

or in matrix form

$$-A_* \, \mathbf{\Delta x} \; = \; B_* \, \mathbf{\Delta m} \tag{8}$$

where A is the $N \times p$ Jacobian with the elements $a_{ij} = \partial \phi_i / \partial x_j$ and B is the $N \times q$ Jacobian with the elements $b_{ik} = \partial \phi_i / \partial m_j$. This is a canonical overdetermined equation system for which the Gauß-Markov solution is known to be 'minimum variance'. The $N \times N$ covariance matrix Q_N applicable to the right hand side of this system is obviously equal to

$$Q_N \; = \; \mathrm{E}\,(B_* \mathbf{\Delta m} \mathbf{\Delta m}' B_*') \; = \; B_* \, Q_m \, B_*' \tag{9}$$

The well known minimum variance values matrix Q_X applicable to the estimate of $\mathbf{\Delta x}$ is equal to

$$Q_X \;=\; (A'_*\, Q_N^{-1}\, A_*)^{-1} \;=\; (A'_*\, (B_*\, Q_m\, B'_*)^{-1}\, A_*)^{-1} \tag{10}$$

To be able to perform the required matrix inversions it is necessary that A and B have full rank. Equation (10) thus yields the minimum variance (the minimum of Trace (Q_X)) obtained at the error free point, which is thus the minimum looked for if $x_{j0} = x_{j*}$ for all j . In practice the natural non-linearity biases will lead to often substantially larger variances than those of the CRB. Equation (10) adapted to a single unknown will be used for computing the CRB in the example presented later.

In the light of the previous considerations, the CRB values given in our paper number 11 are incorrect even as approximation, because the attempt to compute the CRB, explicitly included the intermediate variables. This, however, completely ignores the ability of trivial elimination of intermediate variables at the error free point.

3.2 Constrained Systems

To the system of N stochastic equations displayed in (6) we now add the following K constraint equations:

$$\psi_\ell\,(x_1, \ldots, x_p) \;=\; 0 \qquad (\ell \;=\; 1, \ldots, K < p) \tag{11}$$

We still assume the existence of 'error free values' $(x_{1*}, \ldots, x_{p*}, m_{1*}, \ldots, m_{q*})$, satisfying (6) and (11) simultaneously and further that the equations display a continuous and differentiable behavior in a reasonable interval around this error free point. Expanding (11) to the first order around this point yields:

$$\psi_\ell \;=\; \psi_{\ell *} \;+\; \sum_{j=1}^{K} \left(\frac{\partial \psi_\ell}{\partial x_j}\right)_* \delta x_j \;=\; C_*\; \mathbf{\Delta X} \;=\; 0 \tag{12}$$

where $C(K \times p)$ is the Jacobian of the constraints with the elements $c_{\ell j} = \partial \psi_\ell / \partial x_j$. The constraints are not stochastic by their very nature and, consequently, the increments δx_j have to exactly satisfy them. In other words, $\mathbf{\Delta X}$ must be orthogonal to the K row vectors $\mathbf{c}_\ell$ of C which can be considered to be an non-orthonormal basis of the vector subspace $\mathcal{V}_K \subset \mathcal{V}_p$ of dimension K, while $\mathcal{V}_p$ is the p-dimensional vector space in which $\mathbf{\Delta X}$ has been defined so far. Starting from C we have to derive an incomplete orthogonal matrix $R_X(L \times p)$ with $L = p - K$ such that $R'_X\, R_X \;=\; 1_L$ and $C\, R'_X \;=\; 0_{K \times L}$ where $0_{K \times L}$ is a null matrix.

There obviously are an infinite number of matrices R_X which satisfy these constraints and any of them is acceptable to support the present task. The computation of the Householder transformation represented by the $p \times p$ rotation matrix R, which transforms

C' is an upper triangular $p \times K$ matrix is most probably the simplest *general* approach. One observes that the rows $k+1$ to p of R together correspond to an eligible form of R_X. This technique is employed in estimation as well in a slightly different way, because then the linearization of constraints in a differential correction leads to inhomogeneous equations. For more information we refer to section 7.2 of the practice oriented book by W. Menke[4].

Let us thus assume that we have the orthogonal matrix $R'(p \times p) = |R'_C \vdots R'_X|$ at our disposition. This implies that the matrix $R_C(K \times p)$ contains K orthonormal vectors belonging to $\mathcal{V}_K$, which is the case if R stems from the Householder transformation just mentioned. Hence the left hand side of (8) can be modified as follows

$$\begin{aligned} A\,\mathbf{\Delta x} &= A\,R'R\,\mathbf{\Delta x} = A\,|R'_C \vdots R'_X| \left| \begin{matrix} R_C \\ \dots \\ R_X \end{matrix} \right| \mathbf{\Delta x} \\ &= A\,|R'_C \vdots R'_X| \left| \begin{matrix} 0_{K\times p} \\ \dots \\ R_X\mathbf{\Delta x} \end{matrix} \right| = A\,R'_X(R_X\mathbf{\Delta x}) \end{aligned} \tag{13}$$

But $\mathbf{\Delta y} = R_X\mathbf{\Delta x}$ is a projection of $\mathbf{\Delta x}$ onto the subspace $\mathcal{V}_L$ in which the constrained solution is confined. In analogy to (10), the CRB_Y is now

$$\begin{aligned} CRB_Y &= Q_{Y*} = \mathrm{E}\,(\mathbf{\Delta y}_*\,\mathbf{\Delta y}'_*) \\ &= (R_X\,A'_*\,Q_N^{-1}\,A_*\,R'_X)^{-1} = [R_X\,A'_*\,(B_*\,Q_m\,B'_*)^{-1}\,A_*\,R'_X]^{-1} \end{aligned} \tag{14}$$

which is a full rank $L \times L$ matrix. Although $\mathrm{Tr}(\mathrm{Q_Y})$ yields the overall accuracy minimum which the CRB mathematically guarantees, one is normally interested in the variances and correlations of $\mathbf{X}$ in the $\mathcal{V}_p$ space. From (14) we derive that

$$\begin{aligned} CRB_X = Q_{X*} &= \mathrm{E}\,(\mathbf{\Delta x}\,\mathbf{\Delta x}') \\ &= R'_X\mathrm{E}\,(\mathbf{\Delta y}\mathbf{\Delta y}')R_X \\ &= R'_X\,[R_X\,A'_*\,Q_N^{-1}\,A_*\,R'_X]^{-1}R_X \end{aligned} \tag{15}$$

which is a $p \times p$ matrix of rank $L < p$ and whose diagonal elements represent the breakdown of the CRB for the expected variances of the components of $\mathbf{X}$.

4. SOME PRACTICAL REMARKS

Being in the context of constraint equations in (over)determined stochastic equation systems, we are used to not consider a constraint in whatever form if there is any reason to doubt about its absolute applicability. Mathematically this may seem to be correct. But what about constraints prescribed to be within well defined limits by technical specifications? Normally, these limits are specified as stochastic values and, moreover, in many

cases, costly tests are performed to guarantee these limits. If these tests are performed for a particular device separately (practice for complex medical instrumentation, artificial satellites and others) they are equivalent to one or more measurements assorted with their stochastic quality. Let us consider an example inspired by one of our papers[1].

Consider three GPS antennae located at the points A,B, and C represented by their co-ordinated in the vectors $\mathbf{a} = |x_1, y_1, z_1|'$, $\mathbf{b} = |x_2, y_2, z_2|'$ and $\mathbf{c} = |x_3, y_3, z_3|'$ with their stochastic precision. Assume that we would like to calibrate these points based on the GPS data themselves. Except if collinear, these points define a plane by the nature of plane geometry but they do not give rise to any constraint, because the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are not coplanar. Only the affine vectors (see note 3) $w_1 = \mathbf{a} - \mathbf{b}$, $w_2 = \mathbf{b} - \mathbf{c}$, $w_3 = \mathbf{c} - \mathbf{a}$ are coplanar by construction. Things change by adding a fourth point $\mathbf{d} = |x_4, y_4, z_4|'$, which should also be in the same plane. In that case the determinant containing the components of w_1, w_2 and either $w_{da} = \mathbf{d} - \mathbf{a}$, $w_{db} = \mathbf{d} - \mathbf{b}$ or $w_{cd} = \mathbf{d} - \mathbf{c}$ in its rows, should be zero. In practice, there are legitimate doubts that the co-ordinates of A,B,C and D are really exactly in the same plane, but it is equally legitimate to assume that they are quite close to it. It is thus equally reasonable to accept the zero determinant as a measurement equation which means that ZERO CAN ALSO BE CONSIDERED TO BE A MEASUREMENT. This is an unusual step, but it is with certainty a meaningful usage of an available information, provided one can compute a corresponding variance and correlation for this supplementary equation. Further, such downgraded constraints may be very useful to prevent divergence in a non-linear estimation problem.

In the case just considered it would be an alternative to consider each separate co-ordinate of the three or four points to be stochastic unknowns for which we have a direct field measurements at our disposition. We further equate each of these co-ordinates to an intermediate variable. Adding this information to the estimation problem is fully in the sense of Bayes' original motivation, namely employing that what we know already with its given statistical properties. Even then the inclusion of downgraded constraints is a prerequisite to a complete information usage.

5. AN EXAMPLE WITH ONE UNKNOWN

The details of the example given hereafter, are sufficient to verify our results within statistically reasonable but nevertheless close limits. Let m_1, m_2 and m_3 be some measurements, then it will be difficult to analytically extract the unknown x from

$$\sin^{-1}(\sqrt{x})\,[\cos(m_1\,x) + \sin(m_2\,x)] = m_3 \tag{16}$$

to prepare an optimal estimation. In fact, there is not much to 'estimate' as long as (16) is the only equation of the system. In that case the result is 'deterministically' or 'directly determined' by the triplet m_1, m_2, m_3 and remains the same with or without intermediate

variables. Also if there are $1 < p$ essential unknowns and the system has but $p = N$ equations as well, the (non-singular) equation system will be called directly determined, in contrast to *overdetermined* systems.

Assume now that we have $1 < N$ measurement triplets at our disposition for determining x, or equivalently we have N equations in the subsystem represented by (4) with the generic shape

$$\sin^{-1}(\sqrt{x})\,[\cos(y_{i1}\,x) + \sin(y_{i2}\,x)] = u_i \quad (i = 1, \ldots, N) \tag{17}$$

When setting up the equations fitting into (3) we have to differentiate between two essentially different situations. The simplest case, which we give the label (A), requires only two intermediate variables, because

$$\text{(A)} \qquad x_1 = y_{2i-1}, \qquad x_2 = y_{2i}, \qquad (i = 1, \ldots, N) \tag{18}$$

which, in this case, also implies that $E(u_i) = u$ for all i. In other words, the mathematical expectation for any triplet of the system is the same. Hence, it will be sufficient to compute, if necessary, the weighted mean of the triplet and this is the input to (17). Consequently, we again face a deterministic case. In other words, the system is only seemingly overdetermined, but becomes determined after having made use of the information in (18). Therefore, we call case (A) an indirectly determined system. On the other hand, Case (B) corresponds to the situation, where some or all measurement triplets available for the determination of x on the basis of (17), have different expectations. Then the required intermediate variables would become

$$\text{(B)} \qquad x_{i1,j} = y_{i1,j}, \qquad x_{i2,j} = y_{i2,j}, \qquad (j = 1, \ldots, N) \tag{19}$$

This is the case we will consider in our numerical example.

We assume the error free value $x = \pi/4$, and have selected three triplets or $N = 3$ with the error free measurements :

$$\begin{aligned} y_{i1,1*} &= 1.000, \quad y_{i1,2*} = -1.50, \quad y_{i1,3*} = 2.8 \\ y_{i2,1*} &= 2.100, \quad y_{i2,2*} = 2.860, \quad y_{i2,3*} = -1.10 \end{aligned}$$

The k-th measurement triplet then consists of $(y_{i1,k*} + \epsilon_1, y_{i2,k*} + \epsilon_2, u_{k*} + \epsilon_3)$, where the error free value of u_k is obtained by substituting the error free values of x, y_{i1} and y_{i2} into (17). The value ϵ_i represents a uniformly distributed uncorrelated random error in intervals which are mutually different for the measurements y_{i1}, y_{i2} and u_i. We will consider three cases, one being a very accurate or *precise* case with the error intervals

$$\max / \min \epsilon_1 = \pm 0.075 \quad \max / \min \epsilon_2 = \pm 0.030, \quad \max / \min \epsilon_3 = \pm 0.024$$

Next we include a moderately accurate case where the previous error intervals are quadruplated, and finally we select a coarse case, where the values of the accurate case are multiplied by seven. Convergence problems in the differential correction method used, when trying the factor eight, are the reason to end with seven. The technical scenario one could imagine behind this example is the attempt to determine the not directly measurable system parameter x by means of an existing relation with parameters measurable in very different configurations, explaining the difference between the triplets. Then the 9×9 covariance matrix Q_m mentioned in (9) has the non-zero elements:

$$\begin{aligned} q_m(1,1) &= q_m(2,2) = q_m(3,3) = \epsilon_1^2/3 \\ q_m(4,4) &= q_m(5,5) = q_m(6,6) = \epsilon_2^2/3 \\ q_m(7,7) &= q_m(8,8) = q_m(9,9) = \epsilon_3^2/3 \end{aligned}$$

The set up of a conventional differential correction solved by means of steepest descent is straight forward. The initial value given to x to start the required iterations has been set to $x_* + 0.1$. The iteration is stopped when the size of the differential correction as a whole (thus also comprising the increments of the intermediate variables) is smaller than 1.0 E-11. Mean Values (MN) and Root Mean Squares (RMSQ) values are derived from samples containing 10.000 trials. The values obtained are compared with 'direct weighted estimates' computed in parallel for each trial.

A direct weighted estimate is based on the solution of (17) for each individual triplet. These solutions are obtained by means of the method of the chord applied until zero has been approached within $\pm 1.\,\mathrm{E}-11$. To weigh the three solutions inside the trial, we need the (first order approximation of the) variances of each solution. This computation theoretically requires the error free values of m_1, m_2, m_3 and x. These are not available in practice. To realistically derive the variances we employ the measurements as they are for the three first input variables and for x we have taken the mean of the three solutions. Further, we have verified that a computation of the variances employing error free data has a favorable effect on the mean error except for the mean bias of the coarse case. The influence on the RMSQ errors, however, is always negligible. If z_i is the solution for x of the i-th measurement triplet and σ_i^2 is the corresponding variance for that estimate, the conventional and optimal weighted mean we have used, is given by

$$x = \frac{\sum_i^3 z_i \, \sigma_i^{-2}}{\sum_i^3 \sigma_i^{-2}}$$

to avoid any misunderstanding.

error statistics	precise	medium accuracies	coarse
Direct MN	.574E-4	.160E-3	-.487E-3
Int. Var. MN	.352E-4	-.204E-3	-.103E-2
Direct RMSQ	.429E-2	.172E-1	.301E-1
Int. Var. RMSQ	.429E-2	.171E-1	.299E-1
CRB	.197E-2	.788E-2	.138E-2

Table 1. Results of Monte-Carlo simulation comparing intermediate variables and statistical weighted mean estimates for a single unknown

The results are summarized in table 1. They confirm that intermediate variables compete with the best optimal estimation one can conceive for this case. In the context of accurate measurements the low estimation bias of the new approach is conspicuous. This advantage disappears when error sizes increase. This obviously happens when we move away from the error free point. The close similarity of the RMSQ values for both approaches is remarkable, although in all cases studied (also those not in table 1) there is quite minute difference in favor of intermediate variables. We believe that this can be attributed to the more effective covariance input arrangement which is typical for the intermediate variables approach. Last but not least, in the special case of a single unknown, the direct estimation described before is very robust and remains operational long beyond the point where methods based on linearization have ceased to work.

References

1 L. Fraiture,'Attitude Estimation with GPS-like Measurements,' J. of the Astronautical Sciences, Vol.54, Nos 3 & 4,pp. 595-615, 2006.

2 Y. Bard, Nonlinear Estimation, Academic Press, 1974, New York,

3 M.T. Wasan, Parametric Estimation, Series in Probability and Statistics, McGraw-Hill Book Company, 1970, New York.

4 W. Menke, Geophysical Data Analysis: Discrete Inverse Theory, Rev. Ed., International Geophysics Series, Vol. 45, Academic Press, 1989, London.