NOTE 2 **Version one first posted in the Internet on 22.01.2010***

INTERMEDIATE VARIABLES COMPARED TO IQML IN A DETERMINISTIC SIGNAL ESTIMATION CASE

Luc Fraiture†

1. INTRODUCTION

The purpose of present considerations is the certification of non-linear estimation employing intermediate variables in an extreme case as we will explain hereafter. Intermediate variables (described in note 1) are employed in non-linear estimation problems involving overdetermined systems of equations to separate measurements from unknowns by introducing new unknowns, which are called 'intermediate variables'. This separation allows the grouping of measurements and unknowns in separate equation members. The transformed system then allows a Gauß-Markov estimation which leads to a minimum variance result. The prize to be paid is the creation of a number of new trivial equations and the fact that the original measurement equations may become constraints. In the example selected it will happen that **all** measurement equations are turned into constraints. This is a typical extreme case which one encounters in auto regressive filtering. We will Demonstrate that, also in this case, intermediate variables work at least equally effective as the most accurate known techniques both from a precision as from a numerical point of view.

The estimation of real sinusoids in white noise starting from small data samples where close frequencies can no longer be resolved by periodograms, represents a class of problems which has the properties just described and allows a comparison of intermediate variables on the one hand with the **i**terative **q**uadratic **m**aximum **l**ikelihood method introduced in [1] and [2] and commonly termed IQML on the other hand. The latter method determines the M unknown constant radian frequencies ω_m, phases ϕ_m and amplitudes A_m from the N equidistantly observed signals

$$y_n = y(n) = \sum_{m=1}^{M} A_m \cos(\omega_m n + \phi_m) + v_n \tag{1}$$

* Apart from the introduction, this note is almost equal to a correspondence submitted to the Signal Processing Journal in the summer of 1998, which was declared 'lost' by the technical editor in July 1999, requiring resubmission, which was not made. This had been preceded by two earlier versions which did not refer to IQML, the first of which was submitted in spring 1997. Apart from a few minor differences the present version, extended by section two presented in note 1, was resubmitted to the same journal in January 2008 without success.

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

The observed signal is subject to white noise v_n with $E(v_n) = 0$ and $E(v_k v_\ell) = \sigma^2 \delta_{k\ell}$. Direct exact maximum likelihood estimation (MLE) as presented in [3] is quite difficult, for the right global extremum is surrounded by secondary maxima, thus requiring a very accurate initial guess to be able to guarantee convergence to the right estimate. The IQML method provides an *exact MLE* result in the parameter space of the prediction polynomial where the multimodal environment does — according to experience — not require a very close initial guess and achieves unproblematic convergence to the right solution for modest **s**ignal to **n**oise **r**atios (SNRs); in the example we will give this is 20dB and higher. The introduction of IQML as total least squares problem [4] was preceeded by the publication of an intuitive iteration algorithm for the purpose of linear system identification by Steiglitz and McBride in [5]. This and also the estimation filter proposed in [6] have been shown in [7] to be *exactly equivalent* to IQML. It means that the numerical steps in the iterations of the different approaches can be reduced to each other. More recent investigations [8-9] have shown that classical IQML as introduced in [1-2] – which we have employed in the comparison – is not necessarily consistent (this means: does not converge in the limit to the true solution, even if one starts the limiting process close enough to the true solution). We have included a numerical test which addresses this problem at low SNR.

What we propose in this case realizes the separation of measurements from unknowns by simply transforming **all** the incriminated measurements into unknown intermediate variables. These new unknowns are then separately equated to the measurements they replace, creating a corresponding number of new trivial measurement equations. In the present problem, as said before, **all** N equations (1) become constraint equations. The constrained system obtained is solved by a Gauß-Markov estimation method also called *exact constrained least squares*. We will thus show empirically that intermediate variables, if treated correctly, competitively lead to the MLE result.

The derivation of the cost function based on intermediate variables is made in section 2. To avoid digression into side issues the simplest application has been selected, namely the sum of real cosines without damping. Generalization to damped signal problems is straight forward. In section 3 we compare IQML with the intermediate variable approach. As introduction to the numerical example we cursorily explain how we have implemented the IQML algorithm. The algorithm we have employed to minimize the cost function of the new method, although not part of the method as such, is described in the appendix

3. EXACT CONSTRAINED LEAST SQUARES

Concerning (1) we only know M and the measurements y_n, while all other parameters are unknown. In this deterministic context the idealized model for $y_n - v_n = x_n$ can be taken as a linear combination of complex exponentials, or

$$x_n = 0.5 \sum_{m=-M}^{+M} \alpha_m z_m^n$$

with the shorthands:

$$z_m = z_{-m}^{-1} = e^{i\omega_m}, \qquad \alpha_m = \overline{\alpha_{-m}} = A_m e^{i\phi_m}, \qquad \alpha_0 = 0$$

where overbar means complex conjugation. There are thus $2M$ different values of z_m which are considered to be the collection of roots of a polynomial of degree $2M$ with coefficients a_m, namely

$$0 = \prod_{k=-M}^{-1} (x - z_k) \prod_{\ell=1}^{M} (x - z_\ell) = \sum_{m=-M}^{+M} a_m x^{M+m} \tag{2}$$

Then, x_n is an asymmetric homogeneous power sum of the roots, and as such

$$\sum_{m=-M}^{+M} a_m x_{r+m} = 0 \tag{3}$$

must strictly hold for any integer value of r. We know more about the polynomial coefficients, because if z_m is a root, then also $\overline{z_m}$ as well as z_m^{-1} must be roots. Hence, the coefficients a_m are real and $a_m = a_{-m}$. This leaves only $M + 1$ different coefficients one of which can to be set to an arbitrary non-zero value to get rid of the scaling indetermination. We select $a_0 = 1$. In analogy to auto-regressive filtering and the Prony approach we call this the prediction polynomial.

If we wish to determine the internally consistent and deterministic model just presented, we must simultaneously estimate both the N values of x_n – which in this problem are *the intermediate variables* – grouped in the vector $\mathbf{x}$, as well as the M polynomial coefficients which we combine in the vector $\mathbf{a}$. Thereby the global vector of unknowns $\mathbf{\Theta}' = |\mathbf{x}'|\mathbf{a}'|$ in our alternative method has dimension $N + M$. Accents are employed to denote transposition. To perform an optimal estimation we thus dispose of the N trivial linear random equations for the intermediate variables, namely

$$x_n = y_n, \qquad \text{or} \qquad \mathbf{x} = \mathbf{y} \tag{4}$$

and the bilinear constraint equations resulting from the application of (3), or

$$\begin{aligned} 0 = f_r(\mathbf{a}, \mathbf{x}) = x_{r+M} &+ a_1(x_{r+M+1} + x_{r+M-1}) \\ &+ a_2(x_{r+M+2} + x_{r+M-2}) + \ldots + a_M(x_{r+2M} + x_r) \end{aligned} \tag{5}$$

for r running from 1 to $N - 2M$. This means that the estimation method can be applied only if $3M < N$.

The equations (4) and (5) are in a shape compatible with the Gauß-Markov theorem, allowing an optimal least squares solution achieving minimum variance estimates subject to the **C**ramèr-**R**ao **b**ound (CRB) with all types of bias free white noise. Setting up the *exact* least squares or Gauss-Markov estimation as it is normally called, must thus lead to the equivalent of the maximum likelihood of (1) in the context of Gaussian noise, provided the parameterization of the unknowns in (1) achieved by $\mathbf{\Theta}$ is one-to-one. Unfortunately, not all constrained values of Θ correspond to values fitting (1). More precisely, the condition $a_m = a_{-m}$ given earlier is necessary but not sufficient to force all roots to be on the complex unit circle. Couples of real roots which are each other inverse are possible as well as quadruples of complex roots, for if u is any complex root with $||u|| \neq 1$, then also

$\overline{u}$ and the inverses of these two values are roots of a polynomial with $a_m = a_{-m}$. The one-to-one correspondence would exist if (1) contained hyperbolic instead of trigonometric cosines. Nevertheless, we can expect good performances as long as roots are not attracted towards the real axis. The fact that the real axis is not possible for a parameterization of undamped signals by a symmetric prediction polynomial, suggests that the bench mark test for damped signals found in [1], [2], [11] and [12] and consequently having one of its roots spot on the real axis does not provide the right performance picture for the case of undamped signals.

The constrained cost function Q, which has to be minimized to obtain an optimal least squares solution for our problem, becomes:

$$Q = \frac{1}{\sigma^2} \sum_{n=1}^{N} (y_n - x_n)^2 + 2 \sum_{r=1}^{N-2M} \lambda_r f_r \tag{6}$$

where λ_r are Lagrange multipliers. The algorithm employed in the numerical examples for minimizing (6) is given in the appendix without claiming that there may not be better numerical alternatives. There are $N + M$ unknowns wherefrom $N - 2M$ are fixed by the constraints. Consequently, the vector $\mathbf{\Theta}$ to be estimated contains only 3M independent components. The others can be reconstructed by applying (5). The vector $\mathbf{a}$ uses up M of these independent components by yielding the estimates for the ω_m via the polynomial roots and hence, an arbitrary selection of $2M$ estimated intermediate variables combined with ω_m uniquely determines the estimates for A_m and ϕ_m. This later step is straight forward and does not need to be detailed.

4. COMPARISON OF IQML AND LEAST SQUARES ESTIMATES

The comparison presented in this section is made against the CRB applicable to (1) which has a canonical shape. The computation of the CRB for the estimates of frequencies appearing in sums of sinusoids is well known and a good account about it can be found in section 13.4 of the book by Kay[13].

The IQML application we have implemented is a non-complex formulation where the vector $\mathbf{a}$ of the coefficients of the symmetric prediction polynomial is selected so as to minimize

$$\mathbf{b}' B' (J_x \, J_x')^{-1} B \mathbf{b}$$

We have taken $\mathbf{b}' = |1, \mathbf{a}'|$, where $\mathbf{a}$ is as in previous section. The $(N - 2M) \times N$ matrix J_x is defined in (9) in the appendix. It is a banded matrix which is a function of $\mathbf{a}$ only. The $(N - 2M) \times (M + 1)$ matrix B is equivalent to $|\mathbf{g}(\mathbf{y}), J_a(\mathbf{y})|$ also defined in eq. (9) and (10), but here it is only a function of the measurements $\mathbf{y}$ whereas the corresponding matrices in the appendix are functions of the intermediate variables. At each iteration of the IQML method we determine a new eigenvector $\mathbf{b}$ corresponding to the smallest eigenvalue of $B'(J_X \, J_x')^{-1} B$ based on the previous value for $\mathbf{b} = \mathbf{b_0}$ substituted in J_x. Iterations are terminated if $1 - \epsilon < (\mathbf{b} \cdot \mathbf{b_0}) / (||\mathbf{b}|| \, ||\mathbf{b_0}||) < 1 + \epsilon$. The initialization of the IQML and intermediate variable algorithms is performed by means of the Prony estimate of the symmetric prediction polynomial. To avoid unnecessary iterations it appears that ϵ set to 1.10^{-9} is sufficient. A different convergence criterion has been applied to the algorithm described in the appendix. This criterion consists in checking the goodness of the realized constraint equations as described there.

The example selected for the comparison of the new algorithm with IQML is

$$y_t = \cos 2.1t + \cos 2t + v_t, \qquad (t = 1, \ldots, 25) \tag{7}$$

where v_t is a bias free Gaussian noise with variance σ_v^2. We have to show that both methods give similar results and therefore we have scanned the performances in a range going from 5 to 30dB in steps of 5dB. Each time a Monte-Carlo simulation of 100 realizations has been performed. The theoretically expected similarity of both methods is very pronounced for 25dB ($\sigma_v = 0.0562$) and higher onwards and for these SNRs the CRB is approached by

para-meter	exact value	bias IQML	bias Int. Var.	std. dev. IQML	std. dev. Int. Var.	CRB
ω_1	2.1	-.257E-3	-.922E-3	.7146E-2	.7116E-2	.7111E-2
ω_2	2.0	-.734E-3	-.660E-4	.7850E-2	.7836E-2	.6925E-2
A_1	1.0	.982E-2	-.165E-1	.6889E-1	.6906E-1	.6581E-1
A_2	1.0	.975E-2	-.383E-2	.7449E-1	.7444E-1	.6555E-1
ϕ_1	0.0	.141E-1	.173E-1	.1221	.1218	.1116
ϕ_2	0.0	-.154E-1	-.156E-1	.1312	.1323	.1142

Table 1. Results of Monte-Carlo simulation comparing IQML and intermediate variables for $y_n = \cos 2.1\,t + \cos 2.0\,t + v_t, \quad \sigma_v = 0.0562$ or SNR = 25dB.

both methods equally well. Concerning the numerical performance we mention that we have not optimized the numerical aspects of neither of the methods, and thus the IQML implementation advocated by Hua [14] has not been considered. In practice in one iteration step of both methods we require then one inversion of a $N-2M$ banded symmetric matrix. In the same iteration step in the new algorithm we further need a second inversion of a $N-2M$ matrix but as a counterpart the IQML method requires a single eigenvalue and eigenvector determination for a $M+1$ matrix. By counting the iterations we note that IQML required 34,3 iterations in the mean for 100 realizations, while this number was 33,2 when applying intermediate variables to obtain table 1. Also the execution times of the tests were the same within 10%.

At 20dB the similarity of both results is still good. At lower SNRs convergence gets rapidly problematic and fails systematically at 10dB when initializing with the Prony estimates. The amount of secondary minima getting very close to the minimum variance solution makes that it is only meaningful to empirically check consistency of both methods by initializing them with the error free coefficients. By the way, this approach was taken as well in reference [9] to check estimation consistency. The results at 5dB limited to frequencies, are given in table 2. From this table we see that IQML yields a substantially

para-meter	exact value	bias IQML	bias Int. Var.	std. dev. IQML	std. dev. Int. Var.	CRB
ω_1	2.1	.260	.061	.2859	.1397	.0711
ω_2	2.0	.015	.018	.0844	.0757	.0692

Table 2. Results of Monte-Carlo simulation comparing IQML and intermediate variables for $y_n = \cos 2.1\,t + \cos 2.0\,t + v_t, \quad \sigma_v = 0.562$ or SNR = 5dB.

larger error for the larger frequency. This is in line with the theoretical considerations [8] on the consistency of IQML estimates and empirical analysis made some years later [9] which reveals that the classical IQML as used here is not consistent. This is particularly conspicuous when operating very close to the resolution limit. That we are working close to this limit is suggested by the results displayed in table 3, which deals with the same case as in table 2 except that both frequencies are now spaced by 0.3. This time the CRB

para-meter	exact value	bias IQML	bias Int. Var.	std. dev. IQML	std. dev. Int. Var.	CRB
ω_1	2.3	.928E-2	-.111E-2	.3320E-2	.2946E-2	.2628E-2
ω_2	2.0	.429E-2	-.312E-2	.4420E-1	.3044E-1	.2896E-1

Table 3. Results of Monte-Carlo simulation comparing IQML and intermediate variables for $y_n = \cos 2.3\,t + \cos 2.0\,t + v_t$, $\sigma_v = 0.0562$ or SNR = 5dB.

is approached but clearly closer by the new method. To complete the picture we report another case where IQML (and not the intermediate variable approach) converged in the mean for 100 realizations to a secondary minimum, namely at 5dB with $A_1 = A_2 = 1$, $\phi_1 = \phi_2 = 0$, $\omega_1 = 2.5$ and $\omega_2 = 1.5$ starting with perfect initial conditions.

Notwithstanding this slight superiority of the intermediate variable based algorithm, IQML and exact least squares using the intermediate variable approach have a similar complexity and a similar performance when employed as estimators in this case. The more important conclusion is that the intermediate variables have thereby passed some kind of a supplementary qualification test. Together with the favorable results obtained in an other context, see [10], and in our note 1 before, this convincingly opens up the ability to truly achieve optimal accuracies in a well feasible numerical context for a large class of problems, where this had not been practicable hitherto.

APPENDIX

Minimizing constrained cost functions similar to (6) is a well known problem for which working approaches can for instance be found in [15]. Specific in the present case is the bilinear nature of f_r. To explain the adapted calculation details, we reshape (6) in matrix form and introduce the supplementary vectors $\mathbf{\Lambda}' = |\lambda_1, \ldots, \lambda_{N-2M}|$ and $\mathbf{f}' = |f_1, \ldots, f_{N-2M}|$. Equation (6) can then be written as

$$Q = \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{x})'(\mathbf{y} - \mathbf{x}) + \mathbf{2\Lambda' f} \tag{8}$$

We further need the Jacobian of the constraint vector function, namely

$$(J_x)_{k\ell} = \frac{\partial f_k}{\partial x_\ell}, \qquad (J_a)_{k\ell} = \frac{\partial f_k}{\partial a_\ell} \tag{9}$$

where J_x is a $(N-2M) \times N$ and J_a is a $(N-2M) \times M$ matrix. The bilinear nature of $\mathbf{f}$ makes that

$$\mathbf{f} \equiv J_x \mathbf{x} \equiv J_a \mathbf{a} + \mathbf{g}(\mathbf{x}) \tag{10}$$

with $\mathbf{g}' = |x_{M+1}, x_{M+2}, \ldots, x_{N-M}|$ which corresponds to the linear terms in (5).

The constrained minimum is determined by taking the partial derivative of Q equated to zero, or

$$\frac{\partial Q}{\partial \mathbf{x}} = 2\{\frac{1}{\sigma^2}(\mathbf{y} - \mathbf{x}) + J_x' \boldsymbol{\Lambda}\} = \mathbf{0} \tag{11}$$

$$\frac{\partial Q}{\partial \mathbf{a}} = J_a' \boldsymbol{\Lambda} = \mathbf{0} \tag{12}$$

combined with the constraint equations themselves, namely $\mathbf{f}(\mathbf{x}, \mathbf{a}) = \mathbf{0}$ already specified in (10). The iterative solution of this non-linear system requires an initial value for $\mathbf{x}$ and $\mathbf{a}$, say $\mathbf{a_0}$ and $\mathbf{x_0}$. The subscript zero will further be used to identify vectors and functions taken at the initial value of an iteration step. The first value can be obtained fixing $\mathbf{x_0}$ equal to $\mathbf{y}$ substituted in (15) and solving for $\mathbf{a}$, or

$$\mathbf{a} = -(J_{a0}' J_{a0})^{-1} J_{a0}' \mathbf{g_0} \tag{13}$$

in which one recognizes the extended Prony method as described in [16].

In all subsequent steps one achieves linearization by selecting the Jacobians at the initial value (of the step) in (11) and (12), while for (10) we assume

$$\mathbf{f}(\mathbf{x}, \mathbf{a}) = J_{a0}\mathbf{a} + \mathbf{g_0} = J_{x0}\mathbf{x} \tag{14}$$

From (11) we get

$$\mathbf{x} = \mathbf{y} + \sigma^2 J_{x0}' \boldsymbol{\Lambda} \tag{15}$$

Left multiplying this equation with J_{x0} and applying (14) yields

$$J_{a0}\mathbf{a} + \mathbf{g_0} = J_{x0}\mathbf{y} + \sigma^2(J_{x0} J_{x0}') \boldsymbol{\Lambda} \tag{16}$$

By introducing the shorthand $U = J_{a0}' J_{a0}$ the previous equation reduces to

$$\sigma^2 \Lambda = U^{-1}(J_{a0}\mathbf{a} + \mathbf{g_0} - J_{x0}\mathbf{y}) \tag{17}$$

To make use of (12) we left multiply (17) by J_{a0}' and obtain an update of $\mathbf{a}$, namely

$$\hat{\mathbf{a}} = V^{-1} J_{a0} U^{-1} (J_{x0}\mathbf{y} - \mathbf{g_0})$$

with the abbreviation $V = J_{a0}' U^{-1} J_{a0}$. We introduce $\hat{\mathbf{a}}$ into (17) to get $\hat{\boldsymbol{\Lambda}}$ which in turn is introduced in (15) to provide $\hat{\mathbf{x}}$. One checks whether $\mathbf{f}'(\hat{\mathbf{x}}, \hat{\mathbf{a}})\, \mathbf{f}(\hat{\mathbf{x}}, \hat{\mathbf{a}}) < \epsilon$ where ϵ is preset to 10^{-9}. If the inequality is satisfied, the values for $\hat{\mathbf{a}}$ and $\hat{\mathbf{x}}$ are taken as the solution, otherwise we set $\mathbf{x_0} = \hat{\mathbf{x}}$ as well as $\mathbf{a_0} = \hat{\mathbf{a}}$ and repeat the iterations until convergence which was always reached in all cases treated also if not reported.

REFERENCES

[1] R. Kumerasan, L.L. Sharf and A.K. Shaw, An Algorithm for Pole-Zero Modeling and Spectral Analysis, IEEE Trans. Acoustic Speech Signal processing, vol. 34, June (1986) 634-640.

[2] Y. Bresler and A. Macovski, Exact Maximum Likelihood Estimation of Superimposed Signals in Noise, IEEE Trans. Acoustic Speech Signal processing, vol. 34, pp. 634-640, June 1986, vol. 34, Oct. (1986) 1081-1089.

[3] P. Stoica, R.L. Moses, B. Friedlander and T. Sderström, Maximum Likelihood estimation of the parameters of Multiple Sinusoids from Noisy Measurements, IEEE Trans. Acoustic Speech Signal processing, vol. 37, (1989) 378-392.

[4] G.H. Golub and V. Pereira, The Differentation of of Pseudo-Inverses and Non-Linar Least Squares Problems whose Variables Separate, SIAM J. Numerical Analysis, vol. 10, Apr. (1973) 413-432.

[5] K. Steiglitz and L.E. McBride, A Technique for the Identification of Linear Systems, IEEE Trans. Automatic Control, vol. 10, Oct. (1965) 461-464.

[6] A. Evans and R. Fischl, Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters, IEEE Audio Electroacoustics, vol. 21, Feb. (1973) 61-65.

[7] J.H. McClellan and D. Lee, Exact Equivalence of the Steiglitz-McBride Iteration and IQML, IEEE Trans. Signal processing, vol. 39, Feb.(1991) 509-512.

[8] P. Stoica, J. Li and T. Söderström, On the Inconsitency of IQML, Signal Processing. 56, (1997) 185-190.

[9] P. Lemmerling, L. Vanhamme, S. Van Huffel and B. De Moor, IQML-like Algorithms for Solving Structured Total Least Squares Problems: a unified view, Signal Processing. 81 (2001) 1935-1945.

[10] L. Fraiture, Attitude Estimation with GPS-Like Measurements, J. Astronautical Sciences, vol. 54, July (2006) 595-617.

[11] D.W. Tufts and R. Kumaresan, Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform like Maximum Likelihood, Proc. IEEE, vol. 70, Sept (1982) 975-989.

[12] R. Kumaresan, D.W. Tufts and L.L.sharf, A Prony Method for Noisy Data: Choosing the Signal Components and Selecting the Order in Exponenetial Signal Models, Proc. IEEE, vol 72, Sept. (1984) 975-985.

[13] S.M. Kay, *Modern Spectral Estimation.* Prentice-Hall, Englewoo d-Cliffs, 1988.

[14] Y. Hua, The most Efficient Implementation of the IQML Algortihm, IEEE Trans. Signal processing, vol. 42, August (1994) 2203-2204.

[15] S. Brandt, *Statistical and computational Methods in Data Analysis.* North-Holland Publishing Company, Amsterdam, (1970) 173-181.

[16] S.L. Marple Jr., *Digital Spectral Analysis with applications.* Prentice-Hall, Englewood-Cliffs, (1987) 308-315 .