Bias reduction in transfer function identification

Brian D. O. Anderson and Michel Gevers

***Abstract*— When one random variable is estimated from another measured random variable through a nonlinear mapping constituting the estimator, then any independent additive noise present in the measured variable creates a bias error in the estimated variable. This occurs even if the added noise has zero mean and symmetric density. This bias error can be computed approximately using the second derivative of the mapping when this mapping is available analytically, and hence a bias-corrected estimate can be constructed. We show that this idea can be extended to the case where the mapping is implicitly defined as the solution of a minimization problem, such as in Maximum Likelihood estimation. We also analyze the effect of this bias correction when applied to the estimation of a first order transfer function at one frequency on the basis of a noisy measurement of that transfer function at some other frequency.**

I. INTRODUCTION

The problem considered in this paper can be motivated by the following scenario. Suppose there is an underlying real time-invariant system with a strictly proper transfer function of known order. Suppose that at frequencies $\omega_1, \omega_2, \ldots, \omega_N$ noisy measurements of the transfer function are obtained; measurement noise is additive, bounded zero mean and of known variance and its values at different frequencies are independent. We seek to estimate the value of the transfer function at some other frequency, ω_0 say. How can we obtain a *bias-free*, or possibly *bias-corrected*, estimate of that value?

A very simple version of the problem capturing most aspects of the general problem is as follows: let $W(s)$ be an unknown first order transfer function and suppose there is available a measurement of $W(j\omega_1)$ with $\omega_1 \neq 0$ containing independent zero mean bounded variance additive noise. How can one obtain a bias-free estimate of $W(j\omega_0)$? More generally, if we had available noisy measurements of $W(j\omega_i), i = 1, 2, \ldots, n$, how could we obtain a bias-free estimate of $W(j\omega_0)$?

As we demonstrate, the most natural way to estimate the value of a first order transfer function at one frequency given its value at another frequency is to construct a mapping from its real and imaginary parts at the first frequency to its real and imaginary parts at the second frequency. The mapping is nonlinear; it is exact and unique when the measurements are noiseless. It makes sense to use it when the measurements are noisy. Now it is a crucial property of any nonlinear mapping of an unbiased random variable that the image is (generically) biased. Accordingly, carrying the mapping applicable to the noiseless case over to the noisy measurement case without any adjustment will introduce a bias in the estimate. One of our contributions is to show how to compute this bias and thus allow for its correction.

In the case of multiple measurements at different frequencies of the first order transfer function, in the noiseless case the data *overdetermines* the values of the real and imaginary part at frequency ω_0. In the noisy case, it makes sense to seek a maximum likelihood estimate; this estimate is, obviously, a value constructed as a result of mapping the given data, i.e. there is underlying the construction a mapping (which continues to be nonlinear). Accordingly, bias again is to be expected, and one of our contributions is to show how to compute it. Note that this bias is completely different to bias in linear systems estimation problems caused by correlation between a regressor vector and noise, see e.g. [3].

Our interest in removal of bias in this situation was sparked through the study by one of us of the removal of bias in *localization problems*, where noisy measurements related to a 'target' at an unknown position are available, and through the measurements one wishes to estimate the target position [5]. An example of this phenomenon dating from a long time ago [2] is the determination of the position of a target by a radar which has noisy measurements of the target's range and bearing. Assuming these measurements are both corrupted by additive noise, the estimates of the Cartesian coordinates of the target are biased. The bias arises because of the interaction between the presence of additive zero mean noise in the data and the application of a nonlinear transformation of that data to produce the desired estimate.

As summarized below, the bias itself, or an approximation to it, is obtained by evaluating at the desired estimate the second derivative of the function mapping the data to the estimate. When the mapping from data to estimate is easily constructed analytically, the computation of the bias is easy. This is true for the problem of estimating for a first order $W(s)$ its value at ω_0 given a measurement of its value of ω_1, as it is for the radar localization problem. However, when the mapping is embedded within a maximum likelihood procedure, the determination of the bias is a good deal more difficult. There is very frequently no analytic expression for the actual mapping. Part of our contribution is to explain how to deal with this difficulty. We comment that in the localization literature, the problem of determining bias when

This work is supported by the Australian Research Council Grant DP1092571, and the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office.

B. Anderson is with Research School of Engineering, ANU College of Engineering and Computer Science, RSISE Building 115, Australian National University, Canberra ACT 0200, Australia; and National ICT Australia Limited, Locked Bag 8001, Canberra ACT 2601, Australia. brian.anderson@anu.edu.au

M. Gevers is with ICTEAM, University of Louvain, B1348 Louvain la Neuve, Belgium, and with Dept ELEC, Vrije Universiteit Brussel, Pleinlaan 2, B1050 Brussels, Belgium Michel.Gevers@uclouvain.be

a maximum likelihood method is used has been considered in [4]; the method of that paper involves Taylor series truncations and several approximations, and is certainly more complicated than the method advanced here, even though it may be abstractly equivalent.

The structure of the paper is as follows. In the next section, we recall the standard calculations for obtaining the bias in an estimate when it is derived by nonlinear transformation of data corrupted by additive zero mean noise. Section III explains the calculation of bias when a maximum likelihood procedure underpins the generation of an estimate. Following that, we illustrate in Section IV the concept with the simple case of estimating for a first order transfer function $W(s)$ its value at ω_0 given noisy measurements of its value at ω_1. Section V presents some simulation evidence in corroboration of the results, while Section VII contains conclusions and suggestions for future work.

II. A BIAS FORMULA

In this section, we recall the bias formula mentioned in the preceding section, see also [2]. Let $p \in \mathbb{R}^m$ denote an underlying variable whose value is to be estimated, and suppose that $q \in \mathbb{R}^n$ denotes a data vector, corresponding to measurements. Most commonly, $n \geq m$. In case q is noiseless, suppose that p can be obtained from q via a smooth function $g : \mathbb{R}^n \to \mathbb{R}^m$, thus $p = g(q)$. Suppose the underlying value of p is $\bar{p}$ and $\bar{q}$ denotes the corresponding value for the noiseless measurement vector. Then obviously, $\bar{p} = g(\bar{q})$.

Now suppose that noise is present, so that the measurement is $\breve{q} = \bar{q} + dq$, where dq denotes zero mean noise with covariance matrix Σ. It would seem logical to estimate p by continuing to use the same formula, but with the noisy argument replacing the noiseless argument; thus the estimate would be $\widehat{p} = g(\breve{q})$. It is not difficult to obtain an approximate expression for the bias; the approximation results from the fact that a Taylor series is truncated:

$$\begin{aligned} \widehat{p} &= g(\bar{q} + dq) && (1) \\ &= g(\bar{q}) + \sum_{j=1}^{n} \frac{\partial g}{\partial q_j}(\bar{q}) dq_j + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g}{\partial q_i \partial q_j}(\bar{q}) dq_i dq_j \\ &\quad + \text{higher order terms} && (2) \end{aligned}$$

To abbreviate the expressions we shall adopt the Einstein summation convention here and later in the paper. Thus (2) can be rewritten (with the understanding that there are summations over i and j) as

$$\begin{aligned} \widehat{p} &= g(\bar{q}) + \frac{\partial g}{\partial q_j}(\bar{q}) dq_j + \frac{1}{2} \frac{\partial^2 g}{\partial q_i \partial q_j}(\bar{q}) dq_i dq_j \\ &\quad + \text{higher order terms} \end{aligned}$$

Taking expectations, and neglecting the effect of the higher terms results in

$$E[\widehat{p}] - \bar{p} = \frac{1}{2} \frac{\partial^2 g}{\partial q_i \partial q_j}(\bar{q}) \Sigma_{ij} \quad (3)$$

where Σ_{ij} denotes the (i, j) component of Σ. If we assume that the two values of the Hessian $g_{qq}(\bar{q})$ and $g_{qq}(\bar{q}+dq)$ are suitably close, then an approximate expression for the bias becomes

$$E[\widehat{p}] - \bar{p} = \frac{1}{2} \frac{\partial^2 g}{\partial q_i \partial q_j}(\breve{q}) \Sigma_{ij} \quad (4)$$

This suggests that instead of using $\widehat{p}$ as an estimate of $\bar{p}$, we should use the bias-corrected estimate

$$\hat{p} = g(\breve{q}) - \frac{1}{2} \frac{\partial^2 g}{\partial q_i \partial q_j}(\breve{q}) \Sigma_{ij} \quad (5)$$

One should expect that use of the bias-corrected estimate will improve the mean square error. Assume for convenience that p, q are scalar. Then it is not hard to check that

$$E[\hat{p} - \bar{p}]^2 = E[\widehat{p} - \bar{p}]^2 - [\frac{1}{2} g''(\bar{q}) g''(\breve{q}) - \frac{1}{4} g''(\breve{q})^2] \sigma^4. \quad (6)$$

If the two values of the Hessian $g''(\bar{q})$ and $g''(\breve{q})$ are close, then the reduction in MSE is apparent.

We make no claim for the validity of this formula for all possible noise distributions. In particular, a gaussian distribution will give rise to particular noise values for which the Taylor series truncation may be extremely inaccurate, and in our simulations bounded noise has been used.

III. BIAS ASSOCIATED WITH MAXIMUM LIKELIHOOD ESTIMATES

In this section, we explain how to obtain, at least approximately, a bias associated with a maximum likelihood estimate. Suppose $p \in \mathbb{R}^m$ is the variable we are seeking to estimate, $q \in \mathbb{R}^n$ denotes the measurement vector, assumed to lie in some $Q \subset \mathbb{R}^n$ and in terms of a known smooth function $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, the following minimization problem is well defined for all $q \in Q$:

$$\widehat{p} = \arg\min_p F(p, q). \quad (7)$$

Requiring that the solution of the minimization problem is well defined is equivalent to postulating the existence of a function $g : Q \to \mathbb{R}^m$ such that $p = g(q)$ is the achieved minimum. Note that we are *not* postulating *analytic* knowledge of the function g. In many estimation problems, g is in effect evaluated for a noisy version $\breve{q}$ of q by a program which solves an optimization problem, yielding a biased estimate $\widehat{p} = \arg\min_p F(p, \breve{q})$.

Now the bias in the estimate derived by this method in principle would be definable as before in terms of the second derivative of g. However if this derivative is not available, one is led to ask: can we express the second derivative of g evaluated at some $\breve{q}$ in terms of derivatives of F, evaluated at $\widehat{p}, \breve{q}$, where $\widehat{p} = g(\breve{q})$? The answer is yes, as we now show.

***Theorem** 1:* Let $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ be a smooth function for which a unique solution of the optimization problem (7) is available for all $q \in Q \subset \mathbb{R}^n$. Let $g(q)$ denote the minimizing value of p, i.e. $g(q) = \arg\min_p F(p, q)$. Let $w(p, q) = \frac{\partial F}{\partial p}$, defining a mapping from $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}^m$. Then the following hold at every point $(p, q) = (g(q), q)$ for $q \in Q$ and for all $i = 1, \ldots, m$ and $k = 1, \ldots, n$:

$$\sum_{j=1}^{m} \frac{\partial w^i}{\partial p_j} \frac{\partial g^j}{\partial q_k} + \frac{\partial w^i}{\partial q_k} = 0, \quad (8)$$

which we can write more succinctly, using the Einstein summation convention, as

$$w^i_{p_j} g^j_{q_k} + w^i_{q_k} = 0, \quad i = 1, \ldots, m; \; k = 1, \ldots, n. \tag{9}$$

Similarly, using the same convention, the following equations hold for $i = 1, \ldots, m$; $k = 1, \ldots, n$; $t = 1, \ldots, n$:

$$\begin{aligned} w^i_{p_j} g^j_{q_k q_t} &= -[w^i_{p_j p_s} g^j_{q_k} g^s_{q_t} + w^i_{p_j q_t} g^j_{q_k} + w^i_{p_s q_k} g^s_{q_t} + w^i_{q_k q_t}] \\ &= -[g^j_{q_k} \quad 1] \begin{bmatrix} w^i_{p_j p_s} & w^i_{p_j q_t} \\ w^i_{p_s q_k} & w^i_{q_k q_t} \end{bmatrix} \begin{bmatrix} g^s_{q_t} \\ 1 \end{bmatrix} \end{aligned} \tag{10}$$

Assuming invertibility for all $q \in Q$ of the Jacobian matrix with (i, j) entry $w^i_{p_j}$, first derivatives of g can be expressed in terms of first derivatives of w by (9), and second derivatives of g can be expressed in terms of first and second derivatives of w by (10).

Proof: Because $g(q)$ is a minimizer of $F(p, q)$, it is a zero of $\frac{\partial F}{\partial p} = w(p, q)$, i.e.

$$w(g(q), q) = 0 \; \forall q \tag{11}$$

Differentiating with respect to q_k yields (9). Equation (10) results from differentiation of (9) with respect to q_t. ■

The condition that the Jacobian matrix be nonsingular is not unreasonable; singularity of the Jacobian matrix generally implies that an estimation problem is not well posed. It is a sufficient, though admittedly not necessary, condition for the existence of a unique function g in the neighborhood of any point $q \in Q$. If Q is compact, this means there is a single global g (given the stated assumption on F).

IV. ESTIMATING A FIRST ORDER TRANSFER FUNCTION

In this section we apply the bias correction formulas of Section II to the estimation of a first order transfer function at some frequency ω_0 from a noisy measurement of that same transfer function at another frequency ω_1.

A. The problem set-up

Suppose there is an underlying real system with a transfer function $W(s) = \frac{\bar{b}}{s+\bar{a}}$, with $\bar{a}, \bar{b}$ real and $\bar{a} > 0$. Suppose we measure the transfer function at some frequency $s = j\omega_1, \omega_1 \neq 0$, with the aim of inferring the value of the transfer function at some other frequency ω_0. We are typically interested in the quality of the estimate (bias, variance) when there is zero mean noise of known variance perturbing the measurement of $W(j\omega_1)$.

B. Calculations with noise-free measurements

For future convenience , we shall set

$$x_i = \text{Re}[W(j\omega_i)], \quad y_i = \text{Im}[W(j\omega_i)]. \tag{12}$$

Observe that from the definition of $W(s)$, it follows easily that

$$x_1 = \frac{\bar{a}\bar{b}}{\omega_1^2 + \bar{a}^2}, \quad y_1 = \frac{-\omega_1 \bar{b}}{\omega_1^2 + \bar{a}^2} \tag{13}$$

and so, in particular,

$$\bar{a} = \frac{-\omega_1 x_1}{y_1} \tag{14}$$

$$\bar{b} = -\frac{\omega_1(x_1^2 + y_1^2)}{y_1} \tag{15}$$

We see also that

$$\begin{aligned} W(j\omega_0) &= \frac{j\omega_1 + \bar{a}}{j\omega_0 + \bar{a}} W(j\omega_1) \\ &= \frac{j\omega_1 + \frac{-\omega_1 x_1}{y_1}}{j\omega_0 + \frac{-\omega_1 x_1}{y_1}} W(j\omega_1) \\ &= \frac{\omega_1 |W(j\omega_1)|^2}{\omega_1 x_1 - j\omega_0 y_1} \end{aligned} \tag{16}$$

The mapping from (x_1, y_1) to (x_0, y_0) is thus:

$$x_0 + jy_0 = \frac{\omega_1(x_1^2 + y_1^2)(\omega_1 x_1 + j\omega_0 y_1)}{\omega_1^2 x_1^2 + \omega_0^2 y_1^2} \tag{17}$$

whence we see that

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \frac{\omega_1(x_1^2 + y_1^2)}{\omega_1^2 x_1^2 + \omega_0^2 y_1^2} \begin{bmatrix} \omega_1 x_1 \\ \omega_0 y_1 \end{bmatrix} \tag{18}$$

Of course, it is quite legitimate, and perhaps more traditional, to think of the mapping as the composition of two mappings, one from (x_1, y_1) to $(\bar{a}, \bar{b})$ and then from $(\bar{a}, \bar{b})$ to (x_0, y_0). For our purposes, it is more convenient to work with the single mapping.

C. Calculation of the bias

Our starting point is (18) where, in the notation of Section II, the measured quantity is $q = [x_1, y_1]$ and the quantity to be estimated is $p = [x_0, y_0]$. Let us suppose that the measured values of x_1, y_1 are $\breve{x}_1, \breve{y}_1$, due to perturbations of the true values $\bar{x}_1, \bar{y}_1$ by amounts dx_1, dy_1. With no bias correction, the estimates $\widehat{x}_0, \widehat{y}_0$ of the real and imaginary parts of $W(j\omega_0)$ would be given by

$$\begin{bmatrix} \widehat{x}_0 \\ \widehat{y}_0 \end{bmatrix} = \frac{\omega_1(\breve{x}_1^2 + \breve{y}_1^2)}{\omega_1^2 \breve{x}_1^2 + \omega_0^2 \breve{y}_1^2} \begin{bmatrix} \omega_1 \breve{x}_1 \\ \omega_0 \breve{y}_1 \end{bmatrix} \tag{19}$$

Note incidentally that if the noise can be large in relation to the magnitude of, say, $\bar{y}_1$, then $\breve{y}_1$ could be zero or even negative, even though $\bar{y}_1$ could never assume such a value. This indirectly suggests that there will be clear problems unless noise values are somehow limited, e.g. by assuming uniform rather than gaussian distributions.

Now for convenience in the following manipulations, define the functions

$$\begin{aligned} h(x_1, y_1) &= \frac{\omega_1(x_1^2 + y_1^2)}{\omega_1^2 x_1^2 + \omega_0^2 y_1^2} \\ k(x_1, y_1) &= \frac{2\omega_1(\omega_0^2 - \omega_1^2) x_1 y_1}{(\omega_1^2 x_1^2 + \omega_0^2 y_1^2)^2} \end{aligned} \tag{20}$$

It is easily checked that $\frac{\partial h}{\partial x_1} = k(x_1, y_1) y_1$ and $\frac{\partial h}{\partial y_1} = -k(x_1, y_1) x_1$. Also, $\frac{\partial k}{\partial x_1} x_1 + \frac{\partial k}{\partial y_1} y_1 + 2k = 0$. Evidently, from (18) it follows that the two functions associated with mapping the data to the desired estimates are $g_1(x_1, y_1) = h(x_1, y_1)\omega_1 x_1$ and $g_2(x_1, y_1) = h(x_1, y_1)\omega_0 y_1$.

Using this, one can then obtain the following equations for the second derivatives of g_1, g_2:

$$\begin{bmatrix} \frac{\partial^2 g_1}{\partial x_1^2} & \frac{\partial^2 g_1}{\partial x_1 \partial y_1} \\ \frac{\partial^2 g_1}{\partial x_1 \partial y_1} & \frac{\partial^2 g_1}{\partial y_1^2} \end{bmatrix} = -\omega_1 \frac{\partial k}{\partial y_1} \begin{bmatrix} y_1 \\ -x_1 \end{bmatrix} [y_1 \quad -x_1] \tag{21}$$

and

$$\begin{bmatrix} \frac{\partial^2 g_2}{\partial x_1^2} & \frac{\partial^2 g_2}{\partial x_1 \partial y_1} \\ \frac{\partial^2 g_2}{\partial x_1 \partial y_1} & \frac{\partial^2 g_2}{\partial y_1^2} \end{bmatrix} = \omega_0 \frac{\partial k}{\partial x_1} \begin{bmatrix} y_1 \\ -x_1 \end{bmatrix} [y_1 \quad -x_1] \tag{22}$$

The calculations lead to the following expressions for $\widehat{x}_0, \widehat{y}_0$ where in these expressions, we have $\bar{x}_i = \mathrm{Re}[W(j\omega_i)]$ and $\bar{y}_i = \mathrm{Im}[W(j\omega_i)]$ for $i = 0, 1$, and further, terms in the Taylor series involving derivatives higher than 2 are neglected:

$$\begin{aligned} \widehat{x}_0 &= \bar{x}_0 + [k(\bar{x}_1, \bar{y}_1)\bar{x}_1\bar{y}_1\omega_1 + h(\bar{x}_1, \bar{y}_1)\omega_1]dx_1 \\ &\quad -k(\bar{x}_1, \bar{y}_1)\bar{x}_1^2\omega_1 dy_1 \qquad (23) \\ &\quad -\frac{\omega_1}{2}\frac{\partial k}{\partial y_1}[\bar{y}_1^2(dx_1)^2 - 2\bar{x}_1\bar{y}_1 dx_1 dy_1 + \bar{x}_1^2(dy_1)^2] \\ \widehat{y}_0 &= \bar{y}_0 + k(\bar{x}_1, \bar{y}_1)\omega_0\bar{y}_1^2 dx_1 \\ &\quad +[h(\bar{x}_1, \bar{y}_1)\omega_0 - k(\bar{x}_1, \bar{y}_1)\omega_0\bar{x}_1\bar{y}_1]dy_1 \qquad (24) \\ &\quad +\frac{\omega_0}{2}\frac{\partial k}{\partial x_1}[\bar{y}_1^2(dx_1)^2 - 2\bar{x}_1\bar{y}_1 dx_1 dy_1 + \bar{x}_1^2(dy_1)^2] \end{aligned}$$

For notational convenience but with no real loss of generality, let us assume, as will often occur, that the noises dx_1, dy_1 perturbing the true values of $\bar{x}_1, \bar{y}_1$ in generating the measurements are independent. Then the associated biases are

$$E[\widehat{x}_0] - \mathrm{Re}[W(j\omega_0)] = -\frac{1}{2}\omega_1\frac{\partial k}{\partial y_1}[\bar{y}_1^2\sigma_{x_1}^2 + \bar{x}_1^2\sigma_{y_1}^2] \tag{25}$$

$$E[\widehat{y}_0] - \mathrm{Im}[W(j\omega_0)] = \frac{1}{2}\omega_0\frac{\partial k}{\partial x_1}[\bar{y}_1^2\sigma_{x_1}^2 + \bar{x}_1^2\sigma_{y_1}^2] \tag{26}$$

with the derivatives being evaluated at $(\bar{x}_1, \bar{y}_1)$.

In order to reduce the bias in the estimates $\widehat{x}_0, \widehat{y}_0$ obtained using the formula (19), we propose the following estimates, with derivatives computed not at $(\bar{x}_1, \bar{y}_1)$, but rather at the measured values, i.e. at $(\breve{x}_1, \breve{y}_1) = (\bar{x}_1 + dx_1, \bar{y}_1 + dy_1)$:

$$\hat{x}_0 = \widehat{x}_0 + \frac{1}{2}\omega_1\frac{\partial k}{\partial y_1}[(\breve{y}_1)^2\sigma_{x_1}^2 + (\breve{x}_1)^2\sigma_{y_1}^2] \tag{27}$$

$$\hat{y}_0 = \widehat{y}_0 - \frac{1}{2}\omega_0\frac{\partial k}{\partial x_1}[(\breve{y}_1)^2\sigma_{x_1}^2 + (\breve{x}_1)^2\sigma_{y_1}^2] \tag{28}$$

Remark: It is of interest to understand when the bias correction is likely to be large. Evidently, the values of $\frac{\partial k}{\partial x_1}, \frac{\partial k}{\partial y_1}$ are crucial. One can verify that

$$\frac{\partial k}{\partial x_1} = \frac{2\omega_1(\omega_0^2 - \omega_1^2)(\omega_0^2 y_1^2 - 3\omega_1^2 x_1^2)y_1}{(\omega_1^2 x_1^2 + \omega_0^2 y_1^2)^3} \tag{29}$$

$$\frac{\partial k}{\partial y_1} = \frac{2\omega_1(\omega_0^2 - \omega_1^2)(\omega_1^2 x_1^2 - 3\omega_0^2 y_1^2)x_1}{(\omega_1^2 x_1^2 + \omega_0^2 y_1^2)^3} \tag{30}$$

These expressions show, for example, that bias is less of a problem when ω_0, ω_1 are close. In Section V we shall compare the standard and bias-corrected estimates on a simulated example.

For the sake of completeness we shall also compare the estimates for a and b, and illustrate their behaviour in Section V. It follows from (14) that the standard estimates are:

$$\widehat{a} = \frac{-\omega_1\breve{x}_1}{\breve{y}_1} \tag{31}$$

$$\widehat{b} = -\frac{\omega_1(\breve{x}_1^2 + \breve{y}_1^2)}{\breve{y}_1} \tag{32}$$

As for the bias-corrected estimates, simple calculations based on the Taylor series approximations lead to the following expressions:

$$\hat{a} = \widehat{a} + \frac{\omega_1\breve{x}_1}{\breve{y}_1^3}\sigma_{y_1}^2 \tag{33}$$

$$\hat{b} = \widehat{b} + \frac{\omega_1}{\breve{y}_1}\sigma_{x_1}^2 + \frac{\omega_1\breve{x}_1^2}{\breve{y}_1^3}\sigma_{y_1}^2 \tag{34}$$

D. Higher order transfer functions

Suppose that the first order transfer function $W(s)$ is replaced by a strictly proper n-th order transfer function. Assume for definiteness that this transfer function has relative degree 1; thus it is parametrised by $2n$ real parameters. If noiseless measurements are available of $W(j\omega_i), i = 1, 2, \ldots, n, \omega_i \neq 0, \omega_i \neq \omega_j$ for $i \neq j$, then the coefficients of the numerator and denominator polynomials of $W(s)$ can be expressed in terms of these measurements. Following that, for any ω_0, the value of $W(j\omega_0)$ can be determined. With some extensive but no difficult calculations, one could analytically express the values of the real and imaginary parts as a function of the measurements.

Then the pattern of the previous calculations can be followed, to determine the bias at least approximately in case there is measurement noise, and a formula obtained for an estimate in terms of the data with a bias correction.

To illustrate further, suppose that $W(s) = (\bar{a}s + \bar{b})(s^2 + \bar{c}s + \bar{d})^{-1}$ and that measurements are taken at nonzero frequencies ω_1, ω_2. Let x_i, y_i denote the real and imaginary part of $W(j\omega_i)$. Then in the absence of noise, there holds

$$\begin{bmatrix} \omega_1 & 0 & -\omega_1 x_1 & -y_1 \\ 0 & 1 & \omega_1 y_1 & -x_1 \\ \omega_2 & 0 & -\omega_2 x_2 & -y_2 \\ 0 & 1 & \omega_2 y_2 & -x_2 \end{bmatrix} \begin{bmatrix} \bar{a} \\ \bar{b} \\ \bar{c} \\ \bar{d} \end{bmatrix} = \begin{bmatrix} 0 \\ -\omega_1^2 x_1 \\ 0 \\ -\omega_2^2 x_2 \end{bmatrix} \tag{35}$$

and then

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \frac{(\bar{a}\bar{c} - \bar{b})\omega_0^2 + \bar{b}\bar{d}}{(\omega_0^2 - \bar{d})^2 + \omega_0^2\bar{c}^2} \\ \frac{\omega_0(-\omega_0^2\bar{a} + \bar{a}\bar{d} - \bar{b}\bar{c})}{(\omega_0^2 - \bar{d})^2 + \omega_0^2\bar{c}^2} \end{bmatrix} \tag{36}$$

Evidently, provided the matrix in (35) is nonsingular (and it is provably so when $\omega_1 \neq \omega_2, \omega_i \neq 0, W(s) \not\equiv 0$) the mapping from the data (x_1, y_1, x_2, y_2) to (x_0, y_0) is analytically defined, and so a formula can be found for approximating the bias in the event that the measurements are contaminated by noise.

Generalization of the idea is obviously possible. There is no conceptual difficulty; bookkeeping and formulas will be more complicated.

V. ILLUSTRATIVE EXAMPLES

In this section we illustrate the behaviour of the bias reduction method for the application described in Section IV, where the task is to estimate the frequency response of a first order transfer function at some frequency ω_0 on the basis of noisy measurements made at some other frequency ω_1. We shall illustrate the effect on the achievable bias reduction of the noise level, and of the location of the frequency of measurement, ω_1, with respect to the frequency of the estimate, ω_0.

We shall assume, as stated earlier, that the real and imaginary parts of $W(j\omega_1)$ are measured with two independent noises, i.e. $\breve{x}_1 = \bar{x}_1 + dx_1$ and $\breve{y}_1 = \bar{y}_1 + dy_1$, where dx_1 and dy_1 are independent (see Section IV). Gaussian noises may cause unacceptable behaviour, due to signals that may become zero; note in particular that the standard estimate of $\bar{a}$ would be $\widehat{a} = \frac{-\omega_1 \breve{x}_1}{\breve{y}_1}$, which would cause a problem if $\breve{y}_1 = 0$. As a result, in our simulations we have chosen additive noises dx_1 and dy_1 that are independent and that are uniformly distributed over the interval $[-0.95|\bar{x}_1|, +0.95|\bar{x}_1|]$ for the noise dx_1, and over the interval $[-0.95|\bar{y}_1|, +0.95|\bar{y}_1|]$ for the noise dy_1. As a result, the noisy measurements $\breve{x}_1$ and $\breve{y}_1$ are never closer to zero than 5% of $|x_1|$ and $|y_1|$, respectively.

The simulations have been performed with the transfer function $W(s) = \frac{10}{s+1}$. The first simulation examines the effect of the noise level on the bias reduction achieved by the bias-reduced estimates $\hat{x}_0$ and $\hat{y}_0$ of (27)-(28). The bias of the standard estimates $\widehat{x}_0, \widehat{y}_0$ and of the bias-reduced estimates $\hat{x}_0, \hat{y}_0$ have been computed at frequency $\omega_0 = 5$ from noisy measurements made at frequency $\omega_1 = 1$, over the range of standard deviations corresponding to the uniform distributions defined above. The bias has been computed for the two estimates by averaging over $50,000$ Monte Carlo runs. Figure 1 shows the reduction of the bias obtained with the bias-reduced estimate $\hat{x}_0$, i.e. the ratio of the absolute value of the average bias of $\hat{x}_0$ over the absolute value of the average bias of $\widehat{x}_0$ as a function of the standard deviation σ of the noises. Figure 2 shows the same results for $\hat{y}_0$, while Figures 3 and 4 present the same results for the biases on $\widehat{a}$ and $\hat{a}$, and $\widehat{b}$ and $\hat{b}$, respectively.

The figures show that a significant bias reduction is obtained for small to medium noise levels, but that this bias reduction disappears above a certain threshold in the noise. This is due to the fact that for large noise levels the Taylor series approximations performed in (23)-(24) and in the derivations of (33)-(34) cease to be valid.

We now examine the performance of the bias-correction as a function of the frequency difference between the measured and estimated transfer functions. Thus, we have computed the standard estimates, $\widehat{x}_0$, $\widehat{y}_0$, $\widehat{a}$, $\widehat{b}$, and the corresponding bias-corrected estimates in the frequency range $\omega_0 \in [0.05\ \ 5]$ using noisy measurements $\breve{x}_1$ and $\breve{y}_1$ made at frequency $\omega_1 = 1$. The averages were again computed, for each estimated frequency point, over $50,000$ noisy measurements. The standard deviations of the noises on the measurements

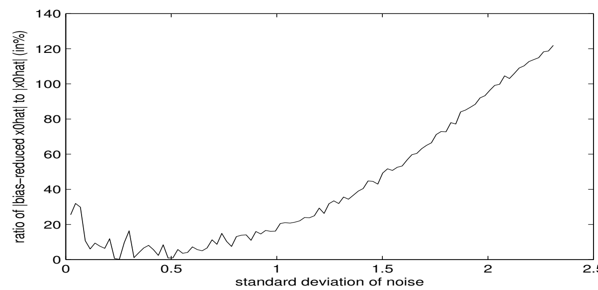

Fig. 1. Bias($|\hat{x}_0|$)/Bias($|\widehat{x}_0|$) versus noise level (in %)

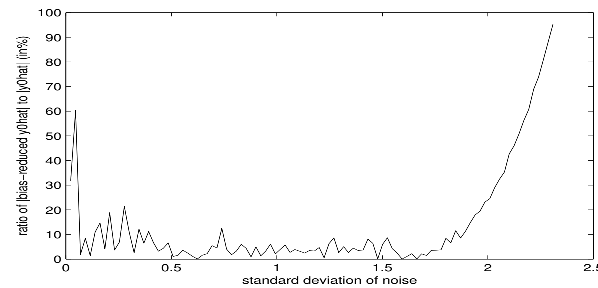

Fig. 2. Bias($|\hat{y}_0|$)/Bias($|\widehat{y}_0|$) versus noise level (in %)

were $\sigma_{x_1} = \sigma_{y_1} = 0.58$. In order to avoid dividing by zero (which happens when ω_0 passes through 1), we present the bias errors (rather than their ratios) as a function of the frequency at estimation, in Figure 5 for the estimation of x_0 and in Figure 6 for the estimation of y_0. As for the average bias errors on the estimates of a and b, they are

$$\text{bias } \widehat{a} = 0.0140, \ \text{bias } \hat{a} = -0.0005$$
$$\text{bias } \widehat{b} = 0.1364, \ \text{bias } \hat{b} = 0.0104$$

The bias-corrected estimates $\hat{a}$ and $\hat{b}$ are considerably better than the standard estimates. In addition, figures 5 and 6 show that the bias-corrected estimates $\hat{x}_0$ and $\hat{y}_0$ of the transfer functions are uniformly better than the standard estimates, except in the immediate vicinity of the measurement frequency. This benefit increases with the distance between the frequency at which the estimate is computed and the frequency where the measurement is made. This follows directly from the fact that the gains of the bias correction terms $\frac{\partial k}{\partial y_1}$ and $\frac{\partial k}{\partial x_1}$, which appear in (27)-(28) are proportional to $\omega_0^2 - \omega_1^2$, as shown by (29)-(30).

VI. EXTENSIONS

The problem addressed in Section IV can of course be extended to the estimation of $W(j\omega_0)$ for some ω_0 using noisy estimates of $W(j\omega_i), i = 1, 2, \ldots, n$. The mapping from the noisy data $\{Re[\breve{W}(\omega_i), Im[\breve{W}(\omega_i)], i = 1, \ldots, n\}$, collected into a real vector $\breve{q}$, to the 2-vector comprising the real and imaginary parts of $W(j\omega_0)$ can be regarded as the composition of two functions. The first function, g, maps $\breve{q}$ into estimates $(\widehat{a}, \widehat{b})$ of $(\bar{a}, \bar{b})$ and is defined by a maximum

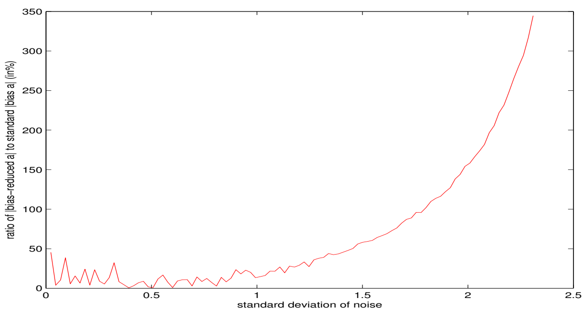

Fig. 3. Bias($|\hat{a}|$)/Bias($|\widehat{a}|$) versus noise level (in %)

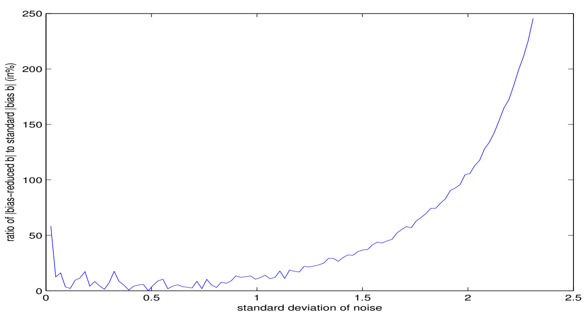

Fig. 4. Bias($|\hat{b}|$)/Bias($|\widehat{b}|$) versus noise level (in %)

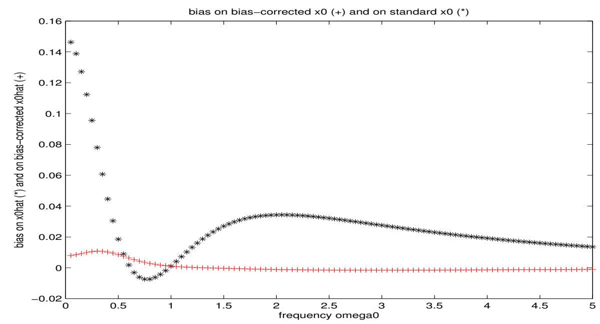

Fig. 5. Bias($\widehat{x}_0$) (black $*$) and Bias($\hat{x}_0$) (red +) versus frequency

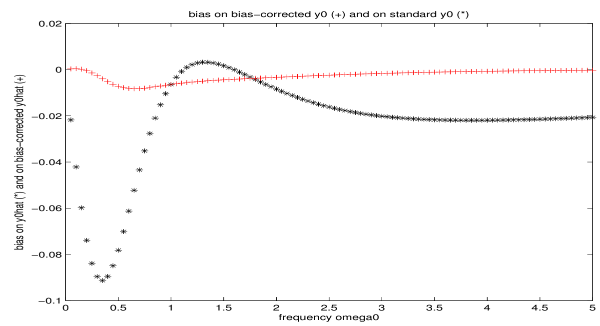

Fig. 6. Bias($\widehat{y}_0$) (black $*$) and Bias($\hat{y}_0$) (red +) versus frequency

likelihood criterion, such as

$$(\widehat{a},\widehat{b}) = g(\breve{q}) = \arg\min_{a,b} \sum_{i=1}^{n} \Big|\Big|\frac{b}{j\omega_i + a} - \breve{W}(j\omega_i)\Big|\Big|^2. \quad (37)$$

Note that the estimates of $\bar{a},\bar{b}$ in principle may be biased or unbiased. If the end goal were simply to obtain bias-free or approximately bias-free estimates of $(\bar{a},\bar{b})$, then naturally one could contemplate introducing a bias correction to form new estimates $(\hat{a},\hat{b})$. The technique for obtaining that bias correction has been explained in Section III.

However, if our interest is in obtaining estimates (indeed, preferably bias-free estimates) of the real and imaginary parts of $W(j\omega_0)$, then this requires a second function, call it h, with components h^r, h^i which maps $(\widehat{a},\widehat{b})$ to the estimate of the real and imaginary part of $W(j\omega_0)$, call them $\widehat{W_0^r},\widehat{W_0^i}$, according to

$$\widehat{W_0^r} + j\widehat{W_0^i} = \frac{\widehat{b}}{j\omega_0 + \widehat{a}}. \quad (38)$$

In order to compute a bias-corrected estimate $\hat{W}_0^r, \hat{W}_0^i$ of $W(j\omega_0)$ directly from the data $\breve{q}$, one then needs to consider the *composite function* $h(g)$ taking $\breve{q}$ to $(\widehat{W_0^r},\widehat{W_0^i})$. This requires the computation of the first and second derivatives of g and h. The full expressions for these derivatives as well as an application to the computation of bias-corrected estimates $\hat{W}_0^r, \hat{W}_0^i$ based on noisy measurements at two frequencies are given in [1].

VII. CONCLUSIONS

A bias error occurs when a variable is estimated through a nonlinear mapping of a related noisy variable. This bias error can be reduced by the use of a correction term that involves the second derivative of the nonlinear map. We have applied this procedure to the estimation of a transfer function at one frequency directly from the noisy measurement of that transfer function at some other frequency. We have also shown that this bias-correction procedure can be extended to the case where the mapping from measurement to estimated quantity is not known, but results from the minimization of a differentiable criterion, as is typical in maximum likelihood estimation.

REFERENCES

[1] B.D.O. Anderson and M. Gevers. Correction for bias in transfer function identification. *in preparation*, 2012.

[2] S.P. Drake and K. Dogancay. Some applications of tensor algebra to estimation theory. In *3rd International Symposium on Wireless Pervasive Computing ISWPC 2008*, pages 106–110, Santorini, Greece, 2000.

[3] C.B. Feng and W.X. Zheng. Robust identification of stochastic linear systems with correlated noise. *IEE Proceedings, Part D*, pages 484–492, 1991.

[4] M. Gavish and A. Weiss. Performance analysis of bearing-only target location algorithms. *IEEE Trans on Aerospace and Electronic Systems*, 28(3):817–827, 1992.

[5] Y. Ji, C. Yu, and B. D. O. Anderson. Localization Bias Correction in n-Dimensional Space. In *2010 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2854–2857, Dallas, TX, March 2010, 2010.