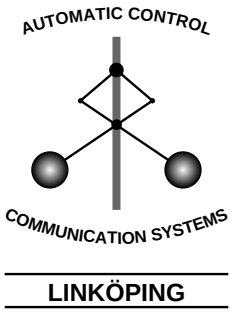
Variance-Bias Tradeoff in Finite Impulse Response Estimates Obtained by Correlation Analysis

Martin Enqvist

Division of Automatic Control
Department of Electrical Engineering
Linköpings universitet, SE-581 83 Linköping, Sweden
WWW: http://www.control.isy.liu.se
E-mail: maren@isy.liu.se

March 28, 2002



Report no.: LiTH-ISY-R-2416

Technical reports from the Control & Communication group in Linköping are available at http://www.control.isy.liu.se/publications.

Variance-Bias Tradeoff in Finite Impulse Response Estimates Obtained by Correlation Analysis

Martin Enqvist

March 28, 2002

Abstract

Correlation analysis can in some cases produce better identification results than an ordinary least squares approach. This is for example the case when a Finite Impulse Response system is estimated from ill-conditioned input-output measurements. In this report, the correlation analysis method is rewritten as a regularized least squares algorithm and the performance of the method is discussed in this context. It turns out that the fact that correlation analysis can be viewed as a kind of regularization explains why and in what sense this method sometimes produces more accurate estimates than the ordinary least squares approach.

Keywords: System identification, correlation analysis, regularization, least squares method.

1 Introduction

Correlation analysis is a classical system identification method (cf. e.g. [2]) that can be used to estimate the impulse response of an unknown linear system. There are several different variants of correlation analysis methods, e.g. the CRA algorithm from the System Identification Toolbox for MATLAB, (cf. Algorithm 2.1 for a detailed description). The underlying principle of this method, and of all similar algorithms as well, is the fact that the correlation between the input and output of a linear system is closely related to the impulse response of the system.

This relation between the input output correlation and the impulse response gives a motivation for the correlation analysis method. The CRA algorithm is however not derived from this relation in a mathematically rigorous way, but instead only heuristically stated. The correlation analysis approach thus differs from many other identification methods, namely from those which are derived from certain criteria and which thus are optimal in certain senses.

One common example of such a method is the least squares approach. This method can be used for estimating the impulse response of a system and it provides estimates that are optimal in the sense that they minimize the sum of squared prediction errors. In addition, the least squares estimator is the minimum variance unbiased linear estimator of the impulse response.

Despite these nice optimality properties of the least squares method there are circumstances in which other identification methods produce results that seem to be better. One example of this is the identification of Finite Impulse Response, FIR, systems from data where the input is very narrow-banded. In these cases the CRA algorithm often seems to outperform the least squares method.

This, perhaps somewhat surprising, observation about the identification of FIR systems is the main topic of this report. The key to understanding this phenomenon lies in the fact that the least squares method gives the minimum variance estimator only among all linear *unbiased* estimators, i.e. there might exist biased estimators that have a lower variance. It is not obvious that the CRA algorithm defines such an estimator. However, it turns out that the CRA algorithm can be viewed as a regularized least squares method and that it is easier to analyze its properties in this framework.

The notation used in this report is introduced in Section 2. This section also contains a detailed problem formulation and an example of what results the CRA and least squares methods can produce on a certain FIR estimation problem. In Section 3, the CRA method is reformulated as a regularized least squares method and an algorithm called CRALS, (of which the CRA and least squares methods are special cases), is introduced. The variance-bias tradeoff of this method is discussed in Section 4 and an extension of the method to cases with colored measurement noise is described in Section 5. Finally, some conclusions and open issues in this subject are presented in Section 6.

2 Notation and problem formulation

A discrete linear system that is time invariant and causal can be described by its impulse response $\{g(t)\}_{t=1}^{\infty}$. The relation between the input $u(t)$ and the output $y(t)$ of the system can be written as

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k) + v(t) \tag{1}$$

where $v(t)$ is a noise term with $E\{v(t)\} = 0$.

If there exists an M such that $g(t) = 0$ for all $t > M$ the system is called a Finite Impulse Response system (i.e. an FIR system). In this case, the relation between the input and the output of the system is

$$y(t) = \sum_{k=1}^{M} g(k)u(t-k) + v(t). \tag{2}$$

An important class of input signals are quasi-stationary signals. The following definition comes from [4].

Definition 2.1 (Quasi-Stationary Signals) *A signal $\{s(t)\}$ is said to be quasi-stationary if it is subject to*

$$E\{s(t)\} = m_s(t), \quad |m_s(t)| \leq C, \quad \forall t$$
$$E\{s(t)s(r)\} = R_s(t,r), \quad |R_s(t,r)| \leq C, \quad \forall t, r$$
$$\overline{E}\{s(t)s(t-\tau)\} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} R_s(t, t-\tau) = R_s(\tau), \quad \forall t$$

The correlation analysis algorithm that is studied in this report is based on the fact that the cross covariance $R_{yu}(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E\{y(t)u(t-\tau)\}$ between the input and output of a linear system is closely related to the impulse response of the system. If u is a quasi-stationary sequence that is uncorrelated with v, (i.e. $\overline{E}\{v(t)u(t-\tau)\} = 0$), R_{yu} can be written as a sum consisting of products of g and R_u (cf. [4])

$$R_{yu}(\tau) = \sum_{k=1}^{\infty} g(k) R_u(\tau - k). \tag{3}$$

If u is white noise, (i.e. $R_u(\tau) = \alpha\delta(\tau)$), then

$$R_{yu}(\tau) = \alpha g(\tau). \tag{4}$$

Hence, given measurements $\{u(t), y(t)\}_{t=1}^{N}$ where u is white noise, estimates $\hat{R}_{yu}$ and $\hat{g}$ of the cross covariance and the impulse response, respectively, can be obtained as

$$\hat{R}_{yu}(\tau) = \frac{1}{N} \sum_{t=\tau+1}^{N} y(t)u(t-\tau) \tag{5}$$

$$\hat{g}(\tau) = \frac{\hat{R}_{yu}(\tau)}{\hat{\alpha}(0)} \tag{6}$$

where $\hat{\alpha}(\tau)$ is an estimate of the variance of u,

$$\hat{\alpha}(\tau) = \frac{1}{N} \sum_{t=1}^{N} u^2(t-\tau). \tag{7}$$

If the linear system is an FIR system with an impulse response of length M and we have experimental data $\{u(t), y(t)\}_{t=1-M}^{N}$ where u is white noise, we can use (5) and (6) to estimate its impulse response.

$$\hat{R}_{yu}(\tau) = \frac{1}{N} \sum_{t=1}^{N} y(t)u(t-\tau) \tag{8}$$

$$\hat{g}(\tau) = \frac{\hat{R}_{yu}(\tau)}{\hat{\alpha}(\tau)} = \frac{\frac{1}{N}\sum_{t=1}^{N} y(t)u(t-\tau)}{\frac{1}{N}\sum_{t=1}^{N} u^2(t-\tau)} = \\ = \frac{\sum_{t=1}^{N} y(t)u(t-\tau)}{\sum_{t=1}^{N} u^2(t-\tau)} \qquad \tau = 1, 2, \ldots, M \tag{9}$$

Equation (9) is derived under the assumption that u is white noise. If this is not the case, the input-output data has to be prefiltered in order to make u as white as possible before the estimator $\hat{g}$ in (9) can be used. This is done in the CRA algorithm. An alternative to prefiltering is to use estimates of R_u and R_{yu} together with (3) in order to get an estimate of g (cf. [4]). Here, we will however only discuss the version of correlation analysis used in the CRA algorithm.

Although the correlation analysis approach is rather straightforward there is of course an easier way to estimate the impulse response, namely to apply the

least squares method to model (2). This model can also be written in regression form

$$y(t) = \varphi^T(t)G + v(t) \tag{10}$$

where

$$\varphi(t) = (u(t-1),\ u(t-2),\ \dots\ , u(t-M))^T$$

and

$$G = (g(1),\ g(2),\ \dots\ , g(M))^T.$$

If we let $Y = (y(1),\ y(2),\ \dots\ , y(N))^T$, $\Phi = (\varphi(1),\ \varphi(2),\ \dots\ , \varphi(N))^T$ and $V = (v(1),\ v(2),\ \dots\ , v(N))^T$ we can write the system (10) in matrix form

$$Y = \Phi G + V. \tag{11}$$

Given measurements $\{u(t), y(t)\}_{t=1-M}^{N}$, the least squares estimator $\hat{G}_{LS}$ based on the model (11) can be computed (cf. [1] or [4]).

$$\hat{G}_{LS} = (\Phi^T\Phi)^{-1}\Phi^T Y \tag{12}$$

It should be noted that the estimator (12) is derived under the assumption that the elements of V are uncorrelated stochastic variables with equal variances.

Under this assumption, the least squares estimator is better than all other linear estimators in the sense that it minimizes the prediction error $\|Y - \Phi G\|_2$. It is also the estimator among all linear unbiased estimators of G that has the lowest variance (cf. [1]).

As a matter of fact, the least squares method is used also in the CRA method. The prefiltering of the measurements that is done in order to make the input as white as possible is performed by assuming that the input u can be described as a Kth order AR process

$$u(t) = -\sum_{i=1}^{K} a_i u(t-i) + e(t) = \varphi^T(t)A + e(t) \tag{13}$$

where φ now is defined as

$$\varphi(t) = (-u(t-1),\ -u(t-2),\ \dots\ , -u(t-K))^T$$

and

$$A = (a_1,\ a_2,\ \dots\ , a_K)^T.$$

The coefficients a_i can be estimated by the least squares method in the same way as the impulse response is estimated in (12). The filtered input and output signals u_F and y_F can then be obtained by reversing the AR filter

$$u_F(t) = u(t) + \sum_{i=1}^{K} \hat{a}_i u(t-i) \tag{14}$$

$$y_F(t) = y(t) + \sum_{i=1}^{K} \hat{a}_i y(t-i) \tag{15}$$

where $\hat{a}_i$ are the estimated AR coefficients. As the filtering in (14) and (15) is linear, the model (2) will still hold with the same g_k even if u, y and v are replaced by u_F, y_F and v_F, respectively. Thus we can use the prefiltered measurements together with (9) in order to estimate the impulse response of the system, provided that we have measurements from $t = 1 - M - K$ to N. If the order of the AR model used in the prefilter is appropriately chosen the signal u_F will be more similar to a white stochastic process than u and the correlation approach will thus work better. The CRA method is summarized in Algorithm 2.1.

Algorithm 2.1 (The CRA method) *Assume that we have measurements* $\{u(t), y(t)\}_{t=1-M-K}^{N}$.

1 Form

$$\Phi = (\varphi(1-M),\ \varphi(2-M),\ \dots\ ,\varphi(N))^T$$

where

$$\varphi(t) = (-u(t-1),\ -u(t-2),\ \dots\ ,-u(t-K))^T$$

and

$$U = (u(1-M),\ u(2-M),\ \dots\ ,u(N))^T.$$

2 Compute the estimate $\hat{A} = (\hat{a}_1,\ \hat{a}_2,\ \dots\ ,\hat{a}_K)^T$*:*

$$\hat{A} = (\Phi^T\Phi)^{-1}\Phi^T U.$$

3 Form

$$u_F(t) = u(t) + \sum_{i=1}^{K} \hat{a}_i u(t-i)$$

$$y_F(t) = y(t) + \sum_{i=1}^{K} \hat{a}_i y(t-i)$$

for $t = (1-M),\ (2-M),\ \dots\ ,N$.

4 Compute the impulse response estimates:

$$\hat{g}(\tau) = \frac{\sum_{t=1}^{N} y_F(t)u_F(t-\tau)}{\sum_{t=1}^{N} u_F^2(t-\tau)}$$

for $\tau = 1, 2, \dots\ ,M$.

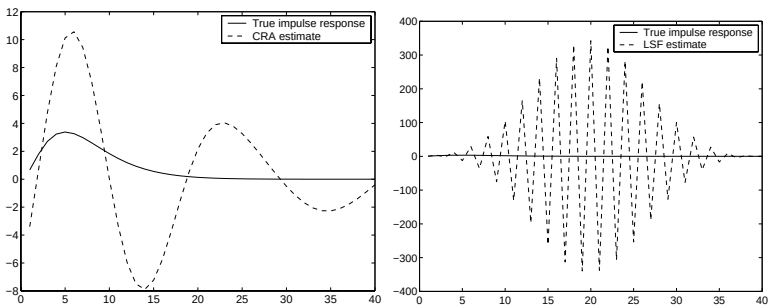
As mentioned previously, the CRA algorithm relies on the assumption that u_F is a white noise process. This is however not always the case. The least squares method, on the other hand, has optimality properties, the minimum prediction error and minimum variance mentioned earlier, that are independent of the distribution of the input signal. One might therefore assume that the CRA method should produce results not equally good as those from the least squares method under all circumstances but this is not the case. In Example 2.1,

the impulse response of an FIR system is estimated with the CRA and the least squares methods and, from the results in Figure 1, it seems as if the first of these algorithms is somewhat better than the second. At least it produces smoother estimates. The main objective of this report is to explain this phenomenon. In the next section we will rewrite the CRA method as a regularized least squares algorithm and the results in Example 2.1 will be easy to explain from this new point of view.

Example 2.1 *Assume that we want to estimate an* FIR *system G with impulse response $g(\tau) = \frac{1}{\tau^2}e^{-0.4\tau}$, $\tau = 1, \ \ldots \ , 40$ from input and output measurements. The input signal u is generated as the output of a 10th order butterworth band pass filter when the input is white Gaussian noise. The passband of this filter is* $[0.1, \ 0.14]$.

1045 samples of u and the corresponding samples of the output signal y are collected and a AR(5) *prewhitening filter is estimated from u. The prefiltered measurements u_F and y_F are constructed as in (14) and (15) with the difference that a white Gaussian noise term with variance 10^{-10} is added to y_F. The measurements of u_F and y_F are then used to obtain the least squares and the* CRA *impulse response estimates.*

The CRA *and the least squares impulse response estimates from a certain realization of u_F and y_F are shown in Figure 1. It is easy to see that the* CRA *estimate is smoother than the least squares estimate. As a matter of fact, the root mean square error, (*RMSE*), of the* CRA *estimate is only 2.5 per cent of the* RMSE *of the least squares estimate.*



(a) (b)

Figure 1: The (a) CRA and (b) least squares impulse response estimates compared with the true impulse response. (Notice that the scalings for the vertical axes are quite different).

---

3 Correlation analysis viewed as a regularized least squares method

The fact that the CRA estimator (9) is written as a sum over time makes it harder to compare it with the matrix version of the least squares estimator (12). It is, however, easy to construct a matrix version of the CRA algorithm as well. From now on, we will only use the prefiltered measurements $\{u_F(t), y_F(t)\}_{t=1-M}^{N}$ with zero mean that can be obtained from the original data set $\{u(t), y(t)\}_{t=1-M-K}^{N}$ as in (14) and (15). We will also assume that the following model holds for the prefiltered data

$$Y_F = \Phi_F G + V_F \tag{16}$$

where the elements of V_F are uncorrelated stochastic variables with the same variance σ^2 and mean zero. (Actually, the assumption that the elements of V_F are uncorrelated is not necessary as we will see in Section 5).

Given prefiltered measurements from a system that is described by (16), the least squares estimator $\hat{G}_{LSF}$ of the impulse response G is

$$\hat{G}_{LSF} = (\Phi_F^T \Phi_F)^{-1} \Phi_F^T Y_F \tag{17}$$

where $Y_F = (y_F(1),\ y_F(2),\ \dots\ , y_F(N))^T$, $\Phi_F = (\varphi_F(1),\ \varphi_F(2),\ \dots\ , \varphi_F(N))^T$ and $\varphi_F(t) = (u_F(t-1),\ u_F(t-2),\ \dots\ , u_F(t-M))^T$. Furthermore, the CRA estimator $\hat{G}_{CRA}$ of the impulse response can be written as

$$\hat{G}_{CRA} = (\mathrm{diag}(\Phi_F^T \Phi_F))^{-1} \Phi_F^T Y_F. \tag{18}$$

We can also construct the estimator $\hat{G}(\delta)$

$$\hat{G}(\delta) = ((1-\delta)\Phi_F^T \Phi_F + \delta \mathrm{diag}(\Phi_F^T \Phi_F))^{-1} \Phi_F^T Y_F = T(\delta) \hat{G}_{LSF} \tag{19}$$

where $T(\delta) = ((1-\delta)I + \delta(\Phi_F^T \Phi_F)^{-1} \mathrm{diag}(\Phi_F^T \Phi_F))^{-1}$ and $\delta \in [0,1]$ is a user-defined parameter. It is easy to see that $\hat{G}(1) = \hat{G}_{CRA}$ and $\hat{G}(0) = \hat{G}_{LSF}$. When $0 < \delta < 1$, $\hat{G}(\delta)$ is a sort of mixture between the CRA and least squares estimators and we will thus call it the CRALS estimator.

The CRALS estimator (19) is related to a version of the least squares method called regularized least squares or ridge regression (cf. [1] or [3]). This method is sometimes used when the $\Phi_F^T \Phi_F$ matrix is ill-conditioned.

The basic idea in the regularized least squares method is to add a small diagonal matrix to the $\Phi_F^T \Phi_F$ matrix before the least squares estimate is computed. More precisely, a regularized least squares estimator $\hat{G}_{LSR}(\delta)$ of our impulse response can be defined as

$$\hat{G}_{LSR}(\delta) = (\Phi_F^T \Phi_F + \delta I)^{-1} \Phi_F^T Y_F \tag{20}$$

where δ is a user-defined small positive number, usually in the range $[0,1]$.

The regularized least squares estimator and the CRALS estimator both have the property that $\delta = 0$ gives the ordinary least squares estimator. The matrices that must be inverted in these two methods are also both made more and more diagonally dominant as δ is increased from 0. These similarities motivate why the CRALS method, and in particular the CRA method, can be viewed as a kind of a regularized least squares method in the sense that variance is traded against bias.

4 Variance-bias tradeoff

The δ parameter in the CRALS method can be used to control the tradeoff between variance and bias of the CRALS estimator. This can be seen from the mean square error MSE(δ) of the estimator $\hat{G}(\delta)$.

$$\begin{aligned}
\text{MSE}(\delta) &= E\{(\hat{G}(\delta) - G)^T(\hat{G}(\delta) - G)\} = \\
&= E\{(T(\delta)\hat{G}_{LSF} - G)^T(T(\delta)\hat{G}_{LSF} - G)\} = \\
&= E\{(\hat{G}_{LSF} - G)^T T(\delta)^T T(\delta)(\hat{G}_{LSF} - G)\} + \\
&\quad + \underbrace{G^T(T(\delta) - I)^T(T(\delta) - I)G}_{=b(\delta)} = \\
&= \sigma^2 \text{trace}\{T(\delta)(\Phi_F^T\Phi_F)^{-1}T(\delta)^T\} + b(\delta) = \\
&= w(\delta) + b(\delta) \qquad (21)
\end{aligned}$$

The first term in (21), $w(\delta)$, is the sum of the variances of the elements of the estimator $\hat{G}(\delta)$ while the second term, $b(\delta)$, is a bias term. When $\delta = 0$ the bias term is 0 and the mean square error is just the sum of the variances of the components of the estimator. By increasing δ from 0, it is sometimes possible to lower the variance term without increasing the bias term too much. In this way a lower mean square error can be obtained. It turns out that the existence of a δ_0 that results in a lower MSE(δ_0) can be checked by a criterion which only includes the matrix $\Phi_F^T\Phi_F$. This is shown in Theorem 4.1.

Theorem 4.1 *Let* $H = \Phi_F^T\Phi_F$. *If* $c(H) = \text{trace}\{(H^{-1} - H^{-2}\text{diag}(H))\} < 0$ *then there exists a* $\delta_0 \in (0, 1]$ *such that* $\text{MSE}(\delta_0) < \text{MSE}(0)$.

Proof: It is easy to verify that

$$T(0) = I \qquad (22)$$

$$T'(\delta) = T(\delta)(I - H^{-1}\text{diag}(H))T(\delta) \qquad (23)$$

$$\lim_{\delta \to 0^+} T'(\delta) = (I - H^{-1}\text{diag}(H)) \qquad (24)$$

This gives

$$\begin{aligned}
w'(\delta) &= \sigma^2 \text{trace}\{T'(\delta)H^{-1}T(\delta)^T + T(\delta)H^{-1}T'(\delta)^T\} \\
&= \sigma^2 \text{trace}\{T(\delta)(I - H^{-1}\text{diag}(H))T(\delta)H^{-1}T(\delta)^T + \\
&\quad + T(\delta)H^{-1}T(\delta)^T(I - \text{diag}(H)H^{-1})T(\delta)^T\} \qquad (25) \\
b'(\delta) &= G^T T'(\delta)^T(T(\delta) - I)G + G^T(T(\delta) - I)^T T'(\delta)G \qquad (26)
\end{aligned}$$

Furthermore

$$\begin{aligned}
\lim_{\delta \to 0^+} w'(\delta) &= \sigma^2 \text{trace}\{(I - H^{-1}\text{diag}(H))H^{-1} + H^{-1}(I - \text{diag}(H)H^{-1})\} = \\
&= 2\sigma^2 \text{trace}\{(H^{-1} - H^{-2}\text{diag}(H))\} \qquad (27) \\
\lim_{\delta \to 0^+} b'(\delta) &= 0 \qquad (28)
\end{aligned}$$

where we have used that $\text{trace}\{AB\} = \text{trace}\{BA\}$ and that $\text{trace}\{A+B\} = \text{trace}\{A\} + \text{trace}\{B\}$ for square matrices A and B. Thus

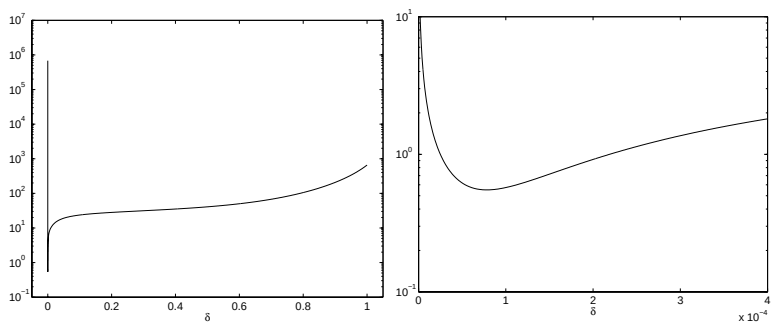
$$\lim_{\delta \to 0^+} \text{MSE}'(\delta) = 2\sigma^2 \text{trace}\{(H^{-1} - H^{-2}\text{diag}(H))\} < 0 \tag{29}$$

This shows that there exists a $\delta_0 \in (0, 1]$ such that $\text{MSE}(\delta)$ is strictly decreasing for $\delta < \delta_0$, i.e. $\text{MSE}(\delta_0) < \text{MSE}(0)$. Q.E.D.

Theorem 4.1 says that if $c(H) = \text{trace}\{(H^{-1} - H^{-2}\text{diag}(H))\} < 0$ then there is a δ_0 such that $\hat{G}(\delta_0)$ is a better estimator than $\hat{G}_{LSF}$ in the sense that it has a lower mean square error. However, there is no easy way to find an optimal δ_0 and in practice a number of different values of δ have to be tested in order to find the best estimator. An example of how well the CRALS method works for different δ values is given in Example 4.1.

Example 4.1 *Consider once again the problem of identifying the impulse response of the* FIR *system in Example 2.1 from the same realizations of u_F and y_F. In this case, $c(H) = -2.2 \cdot 10^{26}$, and Theorem 4.1 can be applied. We can thus guarantee that there is a δ_0 that makes the* CRALS *estimator better than the least squares estimator.*

The mean square errors of the CRALS *estimators are shown in Figure 2. This figure shows that there is an optimal value of δ around $7.8 \cdot 10^{-5}$. The impulse response estimate that is obtained for this value of δ is shown in Figure 3. If this figure is compared to Figure 1 it is obvious that the* CRALS *method can produce very good results for certain values of δ even from very ill-conditioned input output measurements. It is however hard to find such good δ values when the true impulse response and measurement noise variance are unknown.*



(a) (b)

Figure 2: The mean square errors of the impulse response estimators for different values of δ. ((b) is a zoomed part of the lower left corner of (a)).

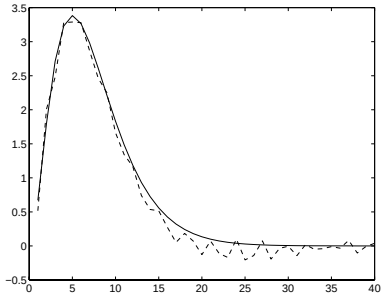


Figure 3: The CRALS impulse response estimate (dashed) with optimal δ compared with the true impulse response (solid).

---

5 Colored noise

In the last two sections we have assumed that the covariance matrix of the prefiltered measurement vector V_F in (16) equals $\sigma^2 I$. This restriction is however not necessary and has merely been made in order to simplify the comparison between the CRALS and the CRA methods. It is namely not possible to adjust the original CRA method so that it compensates for a noise covariance matrix $\mathrm{Cov}\{V_F\} = C \neq \sigma^2 I$.

In many cases it is appropriate to do such a compensation. For example, under the commonly used assumption that the original measurement noise vector V is defined by a white noise process we will have $\mathrm{Cov}\{V\} = \sigma^2 I$ for some σ. However, V_F will not have this property but instead have a non-diagonal covariance matrix that depends on the prewhitening filter.

The CRALS, and thus also the CRA, method can easily be modified so that it compensates for a covariance matrix $\mathrm{Cov}\{V_F\} = C$ just like the least squares method can be modified into a generalized least squares method [1]. The compensation for C is included in the following generalized version of the CRALS estimator

$$\hat{G}(\delta) = ((1-\delta)\Phi_F^T C^{-1}\Phi_F + \delta \mathrm{diag}(\Phi_F^T C^{-1}\Phi_F))^{-1}\Phi_F^T C^{-1} Y_F. \qquad (30)$$

From (30) we see that $\delta = 0$ gives the generalized least squares estimator and that $\delta = 1$ gives what could be called a generalized CRA estimator.

The conclusions that have been drawn in the previous sections are valid also for the estimator (30). In particular, Theorem 4.1 holds if $H = \Phi_F^T \Phi_F$ is replaced with $H = \Phi_F^T C^{-1} \Phi_F$.

6 Discussion

In this report, we have shown that the CRA method sometimes produces better impulse response estimates than the least squares method and that this can be explained by the fact that the CRA method can be viewed as a regularized least squares method.

We have also introduced the CRALS method, which is a method that includes both the CRA and least squares methods. Furthermore, we have presented a theorem that shows that the CRALS method produces impulse response estimates with a lower mean square error than the least squares method in some cases.

The CRALS, and consequently also the CRA, method uses prefiltering in order to make the input signal as white as possible. This kind of prefiltering is useful also for other reasons, (e.g. to get an unbiased truncated impulse response estimate for a system with infinite impulse response). Hence, the CRALS method can be a convenient alternative to the ordinary or the regularized least squares methods in cases where prewhitening is also necessary for other reasons.

References

[1] N. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, 3rd ed., 1998.

[2] K. R. Godfrey, *Correlation methods*, Automatica, 16 (1980), pp. 527–534.

[3] A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, 12 (1970), pp. 55–67.

[4] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Upper Saddle River, N.J. USA, 2nd ed., 1999.