COMPONENT-WISE CONDITIONALLY UNBIASED BAYESIAN PARAMETER ESTIMATION: GENERAL CONCEPT AND APPLICATIONS TO KALMAN FILTERING AND LMMSE CHANNEL ESTIMATION

Mahdi Triki[†], *and Dirk T.M. Slock*[*]

[†] CNRS, Communication Systems Laboratory
[*]Eurecom Institute
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Email: {triki,slock}@eurecom.fr

ABSTRACT

Bayesian parameter estimation techniques such as Linear Minimum Mean Squared Error (LMMSE) often lead to useful MSE reduction, but they also introduce a bias. In this paper, we introduce the concept of Component-Wise Conditionally Unbiased (CWCU) Bayesian parameter estimation, in which unbiasedness is forced for one parameter at a time. This concept had already been introduced in symbol detection a decade ago, where it led to unbiased LMMSE receivers and in which case global CU corresponds to Zero-Forcing. The more general introduction of the CWCU concept is motivated by LMMSE channel estimation, for which the implications of the concept are illustrated in various ways, including the effect on error probability in Maximum Likelihood Sequence Detection (MLSD), reprecussion for blind channel estimation etc. Motivated by the channel tracking application, we also introduce CWCU Kalman filtering.

1. INTRODUCTION

In most applications, estimator designs are subject to a tradeoff between bias and variance. Bias is due to 'mismatch' between the average value of the estimator and the true parameter (conditional bias); whereas variance arises from fluctuations in the estimator due to statistical sampling.

One way to fix this tradeoff is to develop bounds on the best achievable performance in estimating parameters of interest, as well as determine estimators that achieve these bounds. Using the Biased Cramer-Rao Bound B-CRB (which bounds the total MSE), we can bound the MSE of any estimator $\hat{\theta}$ with a given bias vector $b(\theta)$ by

$$\|b(\theta)\|^2 + \text{tr}\left(\left(I + D(\theta)\right) J^{-1}(\theta) \left(I + D(\theta)\right)^T\right) \tag{1}$$

where $J(\theta)$ is the Fisher Information Matrix, and $D(\theta) = \frac{\partial b^T(\theta)}{\partial \theta}$ is the bias gradient matrix.
Ideally, we would like to minimize (1) over all estimators and hence over all bias vectors $b(\theta)$. Unfortunately, if no limitations are imposed on $\widehat{\theta}$, an estimator can always be found that makes both the bias and variance zero at a given point θ. Thus, the minimal bound is the trivial (zero) bound.

Thus, instead of attempting to minimize the MSE over all possible estimators, we may restrict attention to estimator with a bias vector that lies in a suitable class. Traditionally, we consider the class of unbiased estimator. The MSE is bounded by the CRB. It can also be shown that for a Gaussian linear model, the Maximum Likelihood (ML) estimator achieves the CRB; and that is asymptotically unbiased for independent identically distributed (iid) measurement (under suitable regularity assumption).

For biased estimator, given a specified bias, the B-CRB serves as a bound on the smallest attainable variance. It turns out that the B-CRB does not depend directly on the bias but only on the bias gradient matrix. However, it may not be obvious how to choose a particular bias gradient. Hero et al. propose the Uniform CRB (U-CRB), which is a bound on the smallest attainable variance that can be achieved using any scalar estimator with a bounded bias gradient norm [1]. Reference [2] extends the U-CRB for vector parameter, and develops a class of estimators that asymptotically achieve the bound when estimating an unknown vector from iid vector measurements. However, Shahtalebi and Gasor show that for a linear model, the U-CRB is achievable by a class of linear estimators [3]. All estimators in this class have the same variance and the same gradient matrix. However, their performances (in terms of achievable MSE) are not the same. They conclude that the B-CRL is not a sufficient criterion to design optimal estimators.

Eldar consider the minimization of the MSE bound under a linear biased constraint[4]. In fact, bias vectors are allowed to be linear in θ, so that $b(\theta) = M\theta$ for some matrix M (which include unbiased estimation as a special case). An advantage of this class of estimators is that we can use results on unbiased estimating to find estimators that achieve the corresponding MSE bound. In fact, if $\widehat{\theta}_0$ is an efficient estimator, i.e., an unbiased estimator that achieves the CRB, the $\widehat{\theta} = (I + M)\widehat{\theta}_0$ achieves the MSE bound for estimators with bias is equal to $b(\theta) = M\theta$. The problem is that minimization cannot be solved in the general case. However, the author shows that there often exists linear biased vectors that results in an MSE bound that dominates the CRB. The dominating bound can be obtained by solving a certain minimax optimization problem.

If prior information on the parameter statistics is available, Bayesian estimation theory shows that under Bayesian unbiasedness constraint, the MSE is bounded by the Bayesian CRB. Bayesian unbiasedness for random parameters corresponds to unbiasedness on the average, which is a very weak requirement. So that, we

[*]Institut Eurécom's research is partially supported by its industrial members: Bouygues Télécom, France Télécom , Hitachi Europe, SFR, Sharp, STMicroelectronics, Swisscom, Texas Instruments, Thales.

can achieve better bias vs. variance tradeoff. In recent years, the Bayesian formulation of channel estimation has become popular, as it allows for instance the exploitation of the power delay profile. This allows to reduce the number of parameters to be estimated from an a priori delay spread range to the effective delay spread of the power delay profile. For SIMO, MISO or MIMO channels, the Bayesian formulation allows to exploit correlation between antennas and reduce the number of parameters from the physical number of antennas to an effective number of uncorrelated antennas. When the channel is fading in time, the Doppler spectrum and hence correlation in time can be exploited via Wiener or Kalman filtering to further reduce the MSE.

In all these cases, estimation lead to biased channel estimates. This bias is detrimental for a number of applications: MLSD in a SISO system using the Viterbi algorithm (the bias is as detrimental as in biased LMMSE symbol receivers), fitting a parametric (pathwise) model to the channel impulse response, or using the channel estimate for the design of the receiver or the transmitter. The type of unbiasedness that is required here is conditional unbiasedness (where unbiasedness for Bayesian estimation corresponds to unbiasedness on the average, which is very weak requirement). However, conditional unbiasedness for vectors of parameters is usually introduced globally, requiring all parameter components to be jointly unbiased. However, such a stringent requirement, which corresponds to zero-forcing when the parameters are multiple symbols, prevents the exploitation of correlations between the parameters, and hence leads to a significant reduction in the benefits brought about by the Bayesian framework, the prior knowledge.

This motivates us to introduce the Component-Wise Conditionally Unbiased (CWCU) Bayesian parameter estimation. Instead of constrainting the estimator to be globally unbiased, i.e., $E_{/\theta}\left(\widehat{\theta}-\theta\right) = 0$, we impose conditional unbiasedness on one parameter component at a time, i.e.,

$$E_{Y|\theta_k}\left(\widehat{\theta}_k-\theta_k\right)=0 \quad k=1:K \tag{2}$$

where $E_{Y|x}\left[Z(Y,X)\right] = E_Y\left[Z(Y,X)|x\right] = \int Z(Y,x) f_{Y|x}(y|x) dY$ denotes the expectation of $Z(X,Y)$ on Y conditionally to $X=x$; and $\theta = [\theta_1 \cdots \theta_K]^T$ is the parameter vector to be estimated.
In such way, the parameter of interest is constrained to be conditionally unbiased. Other parameters are treated as nuisance parameters. Note that the component-wise concept can be defined at different levels. For example, if we consider Multi-channel impulse response estimation; the component-wise concept can be defined at scalar lever (by considering conditional unbiasedness separately for different channels and time lags). It can also be defined at a block level (by considering conditional unbiasedness jointly for different channels, and separately for different time lags).

This paper is organized as follows. In section 2, we investigate lower bounds for the CWCU estimation. The CWCU-LMMSE estimation, and CWCU linear filtering are derived respectively in sections 3, and 4. Application of the concept for channel estimation for mobile localization is presented in section 5.

2. LOWER BOUNDS FOR CWCU-MMSE ESTIMATION

We consider the estimation of a random parameter vector $\theta = [\theta_1 \cdots \theta_K]^T$ given a set the measurements collected in Y. The MMSE parameter estimation under the component-wise conditionally unbiasedness constraint can be formulated as:

$$\begin{cases} \min_{\widehat{\theta}} E\left\|\widehat{\theta}-\theta\right\|^2 = \sum_k E\left\|\widehat{\theta}_k-\theta_k\right\|^2 \\ E_{Y|\theta_k}\left(\widehat{\theta}_k-\theta_k\right)=0 \quad k=1:K \end{cases}$$

It is easy to see that the minimization problem is separable, and $\widehat{\theta}_k$ is a solution of

$$\begin{cases} \min_{\widehat{\theta}_k} E\left\|\widehat{\theta}_k-\theta_k\right\|^2 \\ E_{Y|\theta_k}\left(\widehat{\theta}_k-\theta_k\right)=0 \end{cases}$$

Without loss of generality, we can assume that θ can be decomposed as $\theta^T = [\theta_k^T \;\; \bar{\theta}_k^T]$. A Bayesian lower bound on the error variance is given by,

$$E\left\|\widehat{\theta}_k-\theta_k\right\|^2 \geq E_{\theta_k}\left(J^{-1}_{CRB,k}\right) \tag{3}$$

where $J_{CRB,k} = E_{Y|\theta_k}\Big(\frac{\partial \ln f(Y|\theta_k)}{\partial\theta_k}\Big)\Big(\frac{\partial \ln f(Y|\theta_k)}{\partial\theta_k}\Big)^T$ is the Fisher Information Matrix (FIM) where θ_k is considered as deterministic, and $\bar{\theta}_k$ as nuisance parameters.
To evaluate the above expression, we should evaluate the conditional pdf with respect to θ_k:

$$f\left(Y|\theta_k\right) = \int f\left(Y,\bar{\theta}_k|\theta_k\right) d\bar{\theta}_k = \int f\left(Y|\theta\right) f\left(\bar{\theta}_k|\theta_k\right) d\bar{\theta}_k$$

Usually, the above bound is difficult to compute because either the above integration is not solvable, or the resulting expectation is not analytically tractable. This difficulty motivates the use of the modified Cramer-Rao bound (MCRB) (introduced by D'Andrea et al. in [5]),

$$E\left\|\widehat{\theta}_k-\theta_k\right\|^2 \geq E_{\theta_k}\left(J^{-1}_{MCRB,k}\right) \tag{4}$$

where $J_{MCRB,k} = E_{Y,\bar{\theta}_k|\theta_k}\Big(\frac{\partial \ln f(Y|\theta)}{\partial\theta_k}\Big)\Big(\frac{\partial \ln f(Y|\theta)}{\partial\theta_k}\Big)^T$.
With respect to the classical CRB, The MCRB is much easier to compute, but generally lower. The problem is that the MCRB can be not tight enough for use in paractical applications.

Another approach to facilitate the calculation of the CRB is to resort the CRB of the joint estimation of the desired parameter together with the nuisance terms [6]. Under the CWCU constraints, the estimation problem becomes

$$\begin{cases} \min_{\widehat{\theta}} E\left\|\widehat{\theta}-\theta\right\|^2 \\ E_{Y,\bar{\theta}_k|\theta_k}\left(\widehat{\theta}-\theta\right)=0 \end{cases}$$

Note that only $\widehat{\theta}_k$ is of interest; $\widehat{\bar{\theta}}_k$ is only estimated to reduce the interference. In such way, the component of interest gets treated as deterministic whereas the other (correlated) parameter components continue to be treated as Bayesian. So that, other components are estimated better (taking into account prior information); as well as the component of interest (due to the coupling through prior and/or data).
As in the Bayesian and deterministic case, a performance bound on CWCU estimation can be defined based on the Fisher Information

Matrix (FIM). The FIM for component-wise conditional estimation problem with respect to the parameter θ_k can be defined as:

$$J_{/k}(\theta_k)=E_{Y,\bar{\theta}_k|\theta_k}\left(\frac{\partial \ln f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial\theta}\right)\left(\frac{\partial \ln f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial\theta}\right)^T \quad (5)$$

The Hessian of $\ln f(\theta|Y)$ can be formulated as:

$$\frac{\partial}{\partial\theta}\left(\frac{\partial \ln f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial\theta}\right)^T=\frac{1}{f\left(Y,\bar{\theta}_k|\theta_k\right)}\frac{\partial}{\partial\theta}\left(\frac{\partial f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial\theta}\right)^T - \left(\frac{\partial \ln f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial\theta}\right)\left(\frac{\partial \ln f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial\theta}\right)^T \quad (6)$$

For the expectation of the first term, we get

$$E_{Y,\bar{\theta}_k|\theta_k}\frac{1}{f\left(Y,\bar{\theta}_k|\theta_k\right)}\frac{\partial}{\partial\theta}\left(\frac{\partial f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial\theta}\right)^T=0 \quad (7)$$

Thus, using Bayes' rule, (6), and (7), the CWCU-FIM can be decomposed onto:

$$\begin{aligned} J_{/k}(\theta_k) &= \frac{\partial}{\partial\theta}\left(\frac{\partial \ln f_{\theta_k}(\theta_k)}{\partial\theta}\right)^T - E_{\theta|\theta_k}\frac{\partial}{\partial\theta}\left(\frac{\partial \ln f_{\theta}(\theta)}{\partial\theta}\right)^T \\ &- E_{Y,\theta|\theta_k}\frac{\partial}{\partial\theta}\left(\frac{\partial \ln f_{Y|\theta}(Y|\theta)}{\partial\theta}\right)^T \\ &= -J_{/k}^{cw} + J_{/k}^{prior} + J_{/k}^{data} \end{aligned} \quad (8)$$

As expected, we see that $J_{/k}^{prior}+J_{/k}^{data} \geq J_{/k} \geq J_{/k}^{data}$, since the CWCU estimation exploits the correlation between the parameters, and imposes an unbiasedness constraint for the parameter of interest.
Using the Schur components lemma, and the block matrix inversion formula, one can show that the error variance is bounded by:

$$E\left\|\widehat{\theta}_k-\theta_k\right\|^2 \geq E_{\theta_k}\left(J_{HCRB,k}^{-1}\right) \quad (9)$$

where $J_{HCRB,k}=J_{/k}(\theta_k,\theta_k)-J_{/k}^T(\theta_k,\bar{\theta}_k)J_{/k}^{-1}(\bar{\theta}_k,\bar{\theta}_k)J_{/k}(\theta_k,\bar{\theta}_k)$, and $J_{/k}(x,z)=E_{Y,\theta|\theta_k}\left(\frac{\partial \ln f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial x}\right)\left(\frac{\partial \ln f\left(Y,\bar{\theta}_k|\theta_k\right)}{\partial z}\right)^T$.
Remark that if θ_k and $\bar{\theta}_k$ are independent $J_{/k}(\theta_k,\theta_k)=J_{MCRB,k}$. $J_{HCRB,k}$, and $J_{MCRB,k}$ overlap if and only if there is no coupling between the different parameters (through prior nor data).

Now, we consider a linear Gaussian model

$$y=H\theta+v \quad (10)$$

where y is the received signal, $\theta \sim N(0,C_{\theta\theta})$ is the parameters to be estimated, and $v \sim N(0,C_{vv})$ is a Gaussian noise independent with θ.
One can show that,

$$\begin{aligned} J_{CRB,k} &= C_{\theta_k\theta_k}^{-1}C_{\theta_k\theta}H^T \\ &\times\left(C_{vv}H\left(C_{\theta\theta}-C_{\theta\theta_k}C_{\theta_k\theta_k}^{-1}C_{\theta_k\theta}\right)H^T\right)^{-1}C_{\theta\theta_k}C_{\theta_k\theta_k}^{-1} \\ J_{MCRB,k} &= h_k^T C_{vv}^{-1}h_k \\ J_{/k} &= C_{\theta\theta}^{-1}-C_{\theta_k\theta_k}^{-1}e_k^Te_k+H^TC_{vv}^{-1}H \end{aligned}$$

where $e_k=[0\cdots 0\ 1\ 0\cdots 0]^T$ is the k^{th} element of the standard R^K basis; and $h_k=He_k$ is the k^{th} column of H. Using Monte Carlo simulations with the linear model, we obtain:

$$J_{CRB,k}^{-1} \geq J_{HCRB,k}^{-1} \geq J_{MCRB,k}^{-1}$$

3. CWCU-LMMSE ESTIMATION FOR LINEAR GAUSSIAN MODEL

We consider a linear Gaussian model (as in (10)); and we assume that the additive noise is white $\left(C_{vv}=\sigma_v^2 I\right)$.
As θ and y are jointly Gaussian, minimizing the MSE leads to the LMMSE estimator:

$$\begin{aligned} \widehat{\theta}_{LMMSE} &= \arg\min_{\widehat{\theta}=Fy} E\left\|\widehat{\theta}-\theta\right\|^2 \\ &= \left(\sigma_v^2 C_{\theta\theta}^{-1}+H^HH\right)^{-1}H^Hy \end{aligned}$$

Under the unbiasedness constraints, minimizing the MMSE leads to the BLUE estimator

$$\widehat{\theta}_{BLUE}=\begin{cases} \arg\min_{\widehat{\theta}=Fy} E\left\|\widehat{\theta}-\theta\right\|^2 \\ E_{Y|\theta}\left(\widehat{\theta}-\theta\right)=0 \end{cases}$$

LMMSE and BLUE estimators are related by

$$\widehat{\theta}_{LMMSE}=\underbrace{\left(\sigma_v^2\left(H^HH\right)^{-1}C_{\theta\theta}^{-1}+I\right)^{-1}}_{B}\widehat{\theta}_{BLUE} \quad (11)$$

where B represents the bias of the LMMSE estimation.
Under the CWCU constraints, one can show that

$$\begin{aligned} \widehat{\theta}_{CWCULMMSE,k} &= \begin{cases} \arg\min_{\widehat{\theta}_k=F_ky} E\left\|\widehat{\theta}_k-\theta_k\right\|^2 \\ E_{Y|\theta_k}\left(\widehat{\theta}_k-\theta_k\right)=0 \end{cases} \forall k \\ &= \begin{cases} \arg\min_{\widehat{\theta}_k=F_ky} E_{Y|\theta_k}\left\|\widehat{\theta}_k-\theta_k\right\|^2 \\ E_{Y|\theta_k}\left(\widehat{\theta}_k-\theta_k\right)=0 \end{cases} \forall k \end{aligned}$$

Thus, in the CWCU-LMMSE estimation, the component of interest θ_k is treated as deterministic, whereas the other (correlated) parameter components $\bar{\theta}_k$ continue to be treated as Bayesian. From (11), we have

$$\widehat{\theta}_{CWCULMMSE,k} \propto \widehat{\theta}_{LMMSE,k}$$

Thus

$$\widehat{\theta}_{CWCULMMSE} = D\,\widehat{\theta}_{LMMSE} \quad (12)$$

where $D=diag(B)^{-1}$ is a diagonal matrix that ensures the component-wise unbiasedness constraint.
If the parameters are decoupled (B diagonal), the CWCU-LMMSE corresponds to the BLUE estimation. The CWCU-LMMSE estimation is of interest if there is a coupling through prior ($C_{\theta\theta}$ is not diagonal), and/or data ($\left(H^HH\right)$ is not diagonal).
Remark that from linear multi-user detection ($C_{\theta\theta}$ is diagonal), the CWCU-LMMSE estimation corresponds to the Unbiased LMMSE; whereas, the (jointly) conditionally unbiased estimator (BLUE) corresponds the MMSE-ZF.

4. THE CWCU LINEAR FILTERING

Consider two stochastic processes $\{x_k\}_{k\in Z}$ and $\{y_k\}_{k\in Z}$ that are correlated. We observe the process y_k but we are interested in the process x_k that we cannot observe. The linear estimation of x_k form y_k can be formulated as a filtering operation, i.e.,

$$\widehat{x}_k = f(q)y_k = \sum_i f_j y_{k-i} \tag{13}$$

where $f(q) = \sum_i f_i q^{-i}$ is a given linear filter; and q^{-1} is the one sample time delay operator.
We assume that

$$E_{|x_k}\widehat{x}_k = \sum_i f_i \; E_{|x_k} y_k = B x_k \tag{14}$$

where B represents the filtering Bias. Note that if $\{y_k, x_k\}_k$ are jointly Gaussian, or if y_k is a linear mixture of independent parameters x_k, the previous assumption is valid.
From (14), one can show that the bias B is given by

$$\begin{aligned} B &= E\left[\widehat{x}_k x_k^H\right] \left(E\left[x_k x_k^H\right]\right)^{-1} \\ &= \sum_i f_i E\left[y_{i-k} x_k^H\right] R_{xx}^{-1} = \oint f(z) S_{yx}(z) \frac{dz}{z} R_{xx}^{-1} \end{aligned} \tag{15}$$

If x_k is a stochastic vector process, the notion of "component-wise" can be defined on different levels:
- per vector sample (removing bias using B^{-1}).
- per scalar sample (removing bias using diag $(B)^{-1}$).
In the following, we will derive the bias update for the Kalman, and Wiener filtering.

4.1. CWCU Kalman filtering

Consider the signal process model

$$\begin{cases} x_{k+1} = F_k x_k + G_k u_k \\ y_k \;\; = H_k x_k + v_k \end{cases} \tag{16}$$

where F_k,G_k, and H_k are given matrices. The initial state x_0, the driving disturbance u_k, and the measurement disturbance v_k are unknown complex vectors. The output y_k is assumed known for all k. We assume also that $E\left[x_0 x_0^H\right] = R_{x,0}$, $E\left[u_k u_l^H\right] = Q\delta_{kl}$, $E\left[v_k v_l^H\right] = R\delta_{kl}$, and $E\left[u_k v_l^H\right] = 0$.

Using Kalman update equations [7], one can show that the Kalman Bias can be updated using:

$$\begin{aligned} R_{x,k+1} &= F_k R_{x,k} F_k^H + G_k Q G_k^H \\ B_{k+1}^{pred} &= \left(F_k B_k^{filt} R_{x,k} F_k^H\right) R_{x,k+1}^{-1} \\ B_{k+1}^{filt} &= \left(I - K_{f,k+1} H_{k+1}\right) B_{k+1}^{pred} + K_{f,k+1} H_{k+1} \end{aligned} \tag{17}$$

where $R_{x,k} = E\left(x_k x_k^H\right)$ is the corelation matrix of x_k, B_{k+1}^{pred}, B_{k+1}^{filt} represent respectively the time and the measurement update of the Kalman bias matrices, and $K_{f,k}$ denotes the Kalman filtering gain.

4.2. CWCU Wiener filtering

Consider the signal process model

$$y_k = x_k + v_k \tag{18}$$

where x_k is the desired signal to be estimated, and v_k represents an additive noise (assumed to have zero mean, and to be uncorrelated with the signal of interest x_k).

Using (15), one can show that the bias of the Wiener, and the causal Wiener filters are given by

$$\begin{aligned} B_{wiener} &= \left(\oint S_{xx}(z) S_{yy}^{-1}(z) S_{xx}(z) \frac{dz}{z}\right) R_x^{-1} \\ B_{wiener}^{causal} &= \left(\oint \left(1 - S_{vv}(z)\Sigma^{-1}A(z)\right) S_{xx}(z) \frac{dz}{z}\right) R_x^{-1} \end{aligned}$$

where $S_{yy}(z)$, $S_{xx}(z)$, and $S_{vv}(z)$ represent respectively the spectrum of the observed, desired, and noise signals, where $A(z)$ denotes the optimal prediction filter for the observed signal y_k, and Σ is the associate prediction error variance.

If we consider the signal process model given in (16). If we assume the F_k, G_k, and H_k are time invariant, one can show that in steady state we have

$$\begin{aligned} S_{xx}(z) &= H\left(zI - F\right)^{-1} GQG^H \left(zI - F^H\right)^{-1} H^H \\ S_{vv}(z) &= R \\ S_{xx}(z) &= HS_{xx}(z)H^H + R \end{aligned} \tag{19}$$

Anderson and Moore show that the causal Wienner solution of the prediction problem is [7]

$$H_{wiener}^{causal}(z) = \left(zI - F\left(I - K_f H\right)\right)^{-1} KH \tag{20}$$

The bias matrix is then given by

$$B^{pred} R_x = \sum_{k=0}^{\infty} \left(F - FK_f H\right)^k kH \left(F^H\right)^{k+1} \tag{21}$$

which correspond to the steady state Kalman Bias (that we can derive using (17)).

5. CWCULMMSE CHANNEL ESTIMATION

In this section, we will focus on one particular problem setting, in which the channels from different Base Stations to a mobile station need to be estimated jointly, for a mobile localization application.
The estimation of the transmission channel has a crucial role in communications systems (for mobile positioning applications, multi-user detection...). Channel parameters are observed indirectly by the received data : convolved with a known training sequence and embedded in a white Gaussian noise.

$$y = \sum_{k=1}^{M} X_k h_k + v \tag{22}$$

where

- $y = [y_1 \cdots y_N]^T$ denotes received data on a given idle period. N is the idle period length.
- $v = [v_1 \cdots v_N]^T$ represents the additive white gaussian noise.
- M is the number of detected base stations.
- $h_k = [h_{k,1} \cdots h_{k,L_k}]^T$ denotes the channel impulse response between the MS and the k^{th} base station. L_k is k^{th} CIR length.
- $X_k = \begin{bmatrix} x_1 & \cdots & x_L \\ \vdots & & \vdots \\ x_N & \cdots & x_{N+L-1} \end{bmatrix}$ is an $N \times L$ matrix characterizing the training sequence of the k^{th} base station.

Using a compact notation, received data can be written as:

$$y = Xh + v \tag{23}$$

where $X = [X_1 \cdots X_M]$, and $h = \left[h_1^T \cdots h_M^T\right]^T$.

Whereas the direct use of the Bayesian channel estimate for an interfering signal allows to better suppress of interference, its use for the user of interest may lead to bias problem. The fact that can undesirable for Mobile localization applications (as TOA is estimated by fitting a parametric model of the channel impulse response). That motivate the use of the CWCU-LMMSE channel estimation, in order to exploit the prior knowledge about the power delay profile, while ensuring the estimator unbiasness (in the component-wise sense [8]).

The linear Minimum Mean Square Error (LMMSE) estimator is given by:

$$\begin{aligned}\widehat{h}_{LMMSE} &= \left(C_{hh}^{-1} + \frac{1}{\sigma_v^2} X^H X\right)^{-1} \frac{1}{\sigma_v^2} X^H y \\ &= \underbrace{\left(\sigma_v^2 \left(X^H X\right)^{-1} C_{hh}^{-1} + I\right)^{-1}}_{B} \left(X^H X\right)^{-1} X^H y \\ &= B\ \widehat{h}_{BLUE} \end{aligned} \tag{24}$$

The component-wise Unbiasness constraint can be formulated as

$$E\left[\widehat{h}_{k,j}/h_{k,j}\right] = h_{k,j} \qquad k = 1:M\ ,\ j = 1:L_k \tag{25}$$

By minimizing the MSE, we have

$$\widehat{h}_{CCULMMSE} = (diag(B))^{-1}\, B\ \widehat{h}_{BLUE} \tag{26}$$

5.1. SIC implementation of the CCULMMSE estimator

The inherent complexity of the CCULMMSE scheme is cubic (as the technique requires the inversion of a, non necessarily Toeplitz, matrix). For practical implementation, Successive Interference Cancellation (SIC) approach can been used to approximate the CCULMMSE estimator; with, only a linear complexity.

Successive Interference Cancellation is a nonlinear type of multi-channel estimation scheme in which CIR's are estimated successively. The approach successively cancels strongest channels. Assume that channels have been ordered in order of decreasing $SNR_i = \frac{\sigma_x^2 \|h_i\|}{\sigma_v^2}$ at the channel estimator input. The first (strongest) channel impulse response estimation is produced by a simple matched filter,

$$\widehat{h}_1 = X_1^H y \tag{27}$$

After making an unbiased estimate of the CIR, the LMMSE estimator is derived, the interfering signal is recreated at the receiver, and subtracted from the received waveform.

$$\begin{aligned}\widehat{h}_1^{LMMSE} &= \left(\sigma_v^2 C_{hh}^{-1} + diag\left(X_1^H X_1\right)\right)^{-1} \widehat{h}_1 \\ \widehat{y}_1 &= X_1 \widehat{h}_1^{LMMSE} \\ y &\leftarrow y - \widehat{y}_1\end{aligned}$$

Note that, even if $X^H X$ can not be approximated as diagonal, non-diagonal elements of $X_1^H X_1$ can be neglected (as the number of unkown is M times less).

In this manner successive base stations does not have to encounter interference caused by initial base stations. SIC leads to good performance for all channel estimates: initial CIR estimates improve because the later channels are given less power which means less interference for the initial channels, and later CIR estimate improve because early BS's interference have been cancelled out. Figure 1 shows that the the SIC well approximate the CCULMMSE estimator (specially for low SNR).

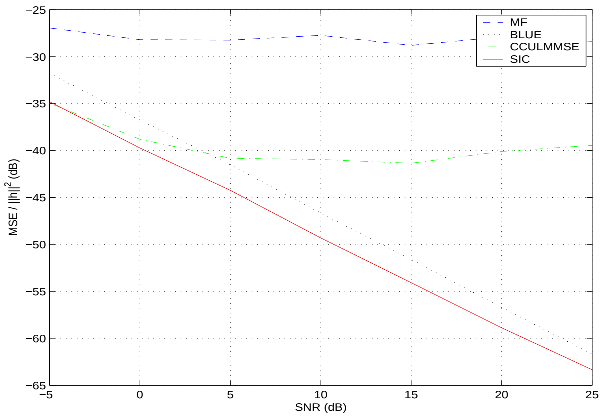

Fig. 1. CIR estimation accuracy using MF, BLUE, CCULMMSE, and SIC estimators

6. REFERENCES

[1] A.O. Hero, J.A. Fessler, and M. Usman. "Exploring Estimator Bias-Variance Tradeoffs Using the Uniform CR Bound," *IEEE Trans. on Signal Processing*, Vol.44, pp.2026-2041, August 1996.

[2] Y.C. Eldar. "Minimum Variance in Biased Estimation: Bounds and Asymptotically Optimal Estimators," *IEEE Trans. on Signal Processing*, Vol.52, pp.1915-1930, July 2004.

[3] K. Shahtalebi, and S. Gazor. "Adaptive Linear Estimators Using Biased Cramèr-Rao Bound," *In proc. of the SSP*, July 2005.

[4] Y.C. Eldar. "MSE Bounds Dominating the Cramèr-Rao Bound," *In proc. of the SSP*, July 2005.

[5] A.N. D'Andrea, U. Mengali, and R. Reggiannini. "The modified Cramèr-Rao bound and its application to synchronization problems," *IEEE Trans. on Communications*, Vol.42, pp.1391-1399, Feb./Mar./Apr. 1994.

[6] I. Reuven, and H. Messer. "A Barankin-type lower bound on the estimation error of a hybrid parameter vector," *IEEE Trans. on Information Theory*, Vol.43, pp.1084 - 1093, May 1997.

[7] B. Anderson and J. Moore. "Optimal Filtering," *Prentice Hall*, 1979.

[8] Mahdi Triki, Salah Abdellatif, and Dirk T.M. Slock. "Interference Cancellation with Bayesian Channel Models and Application to TDOA/IPDL," *In proc. of the ISSPA*, August 2005.