Information-Preserving Transformations for Signal Parameter Estimation

Manuel Stein, Mario Castañeda, Amine Mezghani, and Josef A. Nossek

***Abstract*—The problem of parameter estimation from large noisy data is considered. If the observation size N is large, the calculation of efficient estimators is computationally expensive. Further, memory can be a limiting factor in technical systems where data is stored for later processing. Here we follow the idea of reducing the size of the observation by projecting the data onto a subspace of smaller dimension $M \ll N$, but with the highest possible informative value regarding the estimation problem. Under the assumption that a prior distribution of the parameter is available and the output size is fixed to M, we derive a characterization of the Pareto-optimal set of linear transformations by using a weighted form of the *Bayesian Cramér-Rao lower bound* (BCRLB) which stands in relation to the expected value of the *Fisher information measure*. Satellite-based positioning is discussed as a possible application. Here N must be chosen large in order to compensate for low signal-to-noise ratios (SNR). For different values of M, we visualize the information-loss and show by simulation of the MAP estimator the potential accuracy when operating on the reduced data.**

***Index Terms*—Dimensionality reduction, parameter estimation.**

I. Introduction

ESTIMATION of multiple signal parameters from large data sets is a problem encountered in different fields. For signal parameter estimation in low SNR scenarios, receivers must operate with a large number of samples in order to compensate the lack of receive power. The problem with a large number of samples is that optimum methods like *maximum likelihood* (ML), *maximum a posteriori* (MAP) or *conditional mean* (CM) estimation turn out to be computationally expensive due to the high number of observations N. Realizing that information contained in large data sets can usually be characterized in spaces with less dimensions, an option is to diminish the size of the data before performing the estimation step. Care must be taken during such a dimensionality reduction as its design should assert minimum loss of information with respect to the unknown parameters. Having available an information-preserving reduction technique, estimation can be performed at low complexity on the reduced data or the data can be stored efficiently for later processing.

Manuscript received December 04, 2013; revised February 13, 2014; accepted March 08, 2014. Date of publication April 03, 2014; date of current version May 01, 2014. This work was supported by the German Federal Ministry for Economic Affairs and Energy under Grant 50NA1110. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Azadeh Vosoughi.

The authors are with the Institute for Circuit Theory and Signal Processing, Technische Universität München, München, 80290 Germany (e-mail: manuel.stein@tum.de; mario.castaneda@tum.de; amine.mezghani@tum; josef.a.nossek@tum.de).

Digital Object Identifier 10.1109/LSP.2014.2315537

One possibility to reduce the data size N is to find a basis which spans the M-dimensional subspace with maximum signal energy. Popular approaches in this context include *Karhunen-Loeve transform* (KLT) [1], *principal component analysis* (PCA) [2], *canonical correlation analysis* (CCA) [3] and the *Gaussian information bottleneck* (GIB) [4]. For Gaussian signals, it can be shown that these approaches are equivalent [4]. In this case the data reduction can be achieved by projecting the original observation onto the subspace spanned by the $M \ll N$ eigenvectors corresponding to the M largest eigenvalues of the covariance matrix of the noiseless signal [5], [6]. Such energy-based approaches have been extensively discussed in the context of filtering and estimation [7]–[12]. However, as these works assume linear observation models, they do not cover the fact that in general a non-linear relation between the signal and its parameters exists. Other works consider non-linear models but focus on maximizing the *Fisher information measure* (FIM) on the reduced data set [13]–[15]. The problem with this technique is that in the case of *Fisher estimation theory*, the FIM is dependent on the parameter. As the parameter is unknown during the reduction step, estimation and the dimensionality reduction have to be performed iteratively.

Here, we address dimensionality reduction under a *Bayesian* interpretation. Therefore, parameters are *random variables* drawn according to a prior distribution. In this context, we formulate the problem of finding the optimum transform under a fixed output size M. In contrast to the Fisher approach, this reflects better practical situations as in general prior knowledge is available through a detection or tracking algorithm and has the advantage that the optimum transformation can be determined offline before estimation. Motivated by a simple implementation, we restrict the transformation to be linear.

II. System Model

For the discussion, we consider a discrete receive model

$$\boldsymbol{y} = \boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ and $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathbb{R}^N$ is a signal of known structure, modulated by the unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^K$. The parameter $\boldsymbol{\theta}$ is random and distributed according to $p(\boldsymbol{\theta})$. The additive noise $\boldsymbol{\eta} \in \mathbb{R}^N$ is assumed to be independent of $\boldsymbol{\theta}$, zero-mean and normally distributed with covariance matrix $\boldsymbol{R} \in \mathbb{R}^{N \times N}$. Given $\boldsymbol{y}$, here the goal is to calculate an estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$. The MSE matrix $\boldsymbol{R}_{\hat{\boldsymbol{\theta}}} \in \mathbb{R}^{K \times K}$ [16]

$$\boldsymbol{R}_{\hat{\boldsymbol{\theta}}} = \mathrm{E}_{\boldsymbol{y},\boldsymbol{\theta}}\left[(\hat{\boldsymbol{\theta}}(\boldsymbol{y}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\boldsymbol{y}) - \boldsymbol{\theta})^T\right], \tag{2}$$

can be used to characterize the performance of $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$.

III. Compressed Parameter Estimation

If N is large, calculation of $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ is in general computationally expensive. To simplify the problem, a mapping from the N-dimensional to the M-dimensional space, $f : \mathbb{R}^N \rightarrow \mathbb{R}^M$, $N \gg M$, is applied. Here we restrict to linear forms

$$\boldsymbol{y}_c = \boldsymbol{C}^T\boldsymbol{y} = \boldsymbol{C}^T\boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{C}^T\boldsymbol{\eta} = \boldsymbol{s}_c(\boldsymbol{\theta}) + \boldsymbol{\eta}_c, \tag{3}$$

with $\boldsymbol{y}_c \in \mathbb{R}^M, \boldsymbol{C} \in \mathbb{R}^{N\times M}$. It is now possible to substitute $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ by the estimate $\hat{\boldsymbol{\theta}}_c(\boldsymbol{y}_c)$, which is calculated exclusively with the compressed data $\boldsymbol{y}_c$ and exhibits the MSE matrix

$$\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{C}) = \mathrm{E}_{\boldsymbol{y},\boldsymbol{\theta}}\left[(\hat{\boldsymbol{\theta}}_c(\boldsymbol{C}^T\boldsymbol{y}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_c(\boldsymbol{C}^T\boldsymbol{y}) - \boldsymbol{\theta})^T\right]. \tag{4}$$

Having decided for an output configuration M, an appropriate transformation $\boldsymbol{C}$ has to be found and the performance-loss measure for each element $[\boldsymbol{\theta}]_k$ of the parameter vector $\boldsymbol{\theta}$

$$[\chi(\boldsymbol{C})]_k = \frac{[\boldsymbol{R}_{\hat{\boldsymbol{\theta}}}]_{kk}}{[\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{C})]_{kk}} = \frac{[\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{I}_N)]_{kk}}{[\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{C})]_{kk}}, \tag{5}$$

where $\boldsymbol{I}_N \in \mathbb{R}^{N\times N}$ denotes identity, has to be evaluated. In order to find the right design of $\boldsymbol{C}$ under a fixed output size M, we follow the approach of characterizing the set of projections $\mathcal{P} \subset \mathbb{R}^{N\times M}$ which leads to Pareto-optimal points in terms of estimation performance $\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}$. This is achieved by solving for the best design under a weighted MSE criterion

$$\boldsymbol{C}^\star = \arg\min_{\boldsymbol{C}\in\mathbb{R}^{N\times M}} \mathrm{MSE}(\boldsymbol{W},\boldsymbol{C}), \tag{6}$$

where

$$\mathrm{MSE}(\boldsymbol{W},\boldsymbol{C}) = \mathrm{Tr}\left(\boldsymbol{W}\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{C})\right). \tag{7}$$

Having available the solutions $\boldsymbol{C}^\star$ for all positive-semidefinite weightings $\boldsymbol{W}$, allows access to $\mathcal{P}$ such that the transformation can be chosen by selecting the matrix $\boldsymbol{C} \in \mathcal{P}$ which leads to the desired operational performance $\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{C})$ for the application of interest. The expressions (2) (4) (5) and (7) are in general difficult to evaluate due to the characterization of $\boldsymbol{R}_{\hat{\boldsymbol{\theta}}}$ and $\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}$. This can be circumvent by exploiting that (7) and the individual entries on the diagonal of (4) can be bounded through the *Bayesian Cramér-Rao lower bound* (BCRLB) [16]

$$\begin{aligned}\mathrm{Tr}\left(\boldsymbol{W}\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{C})\right) &\geq \mathrm{Tr}\left(\boldsymbol{W}\boldsymbol{J}_B^{-1}(\boldsymbol{C})\right)\\ [\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{C})]_{kk} &\geq [\boldsymbol{J}_B^{-1}(\boldsymbol{C})]_{kk},\end{aligned} \tag{8}$$

where $\boldsymbol{J}_B$ is the *Bayesian Fisher information matrix* (BFIM)

$$\boldsymbol{J}_B(\boldsymbol{C}) = \boldsymbol{J}_D(\boldsymbol{C}) + \boldsymbol{J}_P, \tag{9}$$

with the *expected Fisher information matrix* (EFIM)

$$\begin{aligned}\boldsymbol{J}_D(\boldsymbol{C}) &= \mathrm{E}_{\boldsymbol{y},\boldsymbol{\theta}}\left[\left(\frac{\partial \log p(\boldsymbol{C}^T\boldsymbol{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T \frac{\partial \log p(\boldsymbol{C}^T\boldsymbol{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\\ &= \mathrm{E}_{\boldsymbol{\theta}}\left[\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T \boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{R}\boldsymbol{C})^{-1}\boldsymbol{C}^T\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right],\end{aligned} \tag{10}$$

under Gaussian noise and the *prior information matix* (PIM)

$$\boldsymbol{J}_P = E_{\boldsymbol{\theta}}\left[\left(\frac{\partial \log p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T \frac{\partial \log p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]. \tag{11}$$

Estimators which attain equality in (8) are called *Bayesian efficient estimates* [16]. Therefore, under the assumption that an efficient estimate $\hat{\boldsymbol{\theta}}_c(\boldsymbol{C}^T\boldsymbol{y})$ can be used for the problem at hand, it can be assumed that the MSE equals the BCRLB

$$\begin{aligned}\mathrm{MSE}_c(\boldsymbol{W},\boldsymbol{C}) &= \mathrm{Tr}\left(\boldsymbol{W}\boldsymbol{J}_B^{-1}(\boldsymbol{C})\right)\\ [\boldsymbol{R}_{\hat{\boldsymbol{\theta}}_c}(\boldsymbol{C})]_{kk} &= [\boldsymbol{J}_B^{-1}(\boldsymbol{C})]_{kk},\end{aligned} \tag{12}$$

$\forall \boldsymbol{C} \in \mathbb{R}^{N\times M}$. This allows to replace problem (6) by

$$\boldsymbol{C}^\star = \arg\min_{\boldsymbol{C}\in\mathbb{R}^{N\times M}} \mathrm{Tr}\left(\boldsymbol{W}\boldsymbol{J}_B^{-1}(\boldsymbol{C})\right), \tag{13}$$

and the figure of merit (5) by

$$[\boldsymbol{\chi}(\boldsymbol{C})]_k = \frac{[\boldsymbol{J}_B^{-1}(\boldsymbol{I}_N)]_{kk}}{[\boldsymbol{J}_B^{-1}(\boldsymbol{C})]_{kk}}. \tag{14}$$

IV. Alternative Transformation Design

The solutions to problem (13) with positive-semidefinite matrices $\boldsymbol{W}$ provide the boundary of the achievable MSE region, leading to the set of Pareto-optimal transformations $\mathcal{P}$. However, we consider an alternative problem which turns to be solvable in closed form. This problem reads as

$$\begin{aligned}\boldsymbol{C}^\star &= \arg\max_{\boldsymbol{C}\in\mathbb{R}^{N\times M}} \mathrm{Tr}\left(\boldsymbol{W}'\boldsymbol{J}_B(\boldsymbol{C})\right)\\ &= \arg\max_{\boldsymbol{C}\in\mathbb{R}^{N\times M}} \mathrm{Tr}\left(\boldsymbol{W}'\boldsymbol{J}_D(\boldsymbol{C})\right),\end{aligned} \tag{15}$$

for all positive-semidefinite matrices $\boldsymbol{W}'$. In other words, instead of finding the boundary of the MSE region (13), we aim at finding the boundary of the achievable BFIM region (15). Both problems turn out to be equivalent, i.e. if $\boldsymbol{J}_B$ is a boundary point of the BFIM region, then $\boldsymbol{J}_B^{-1}$ is a boundary point of the achievable MSE region. This is due to the fact that the inverse $f(\boldsymbol{A}) = \boldsymbol{A}^{-1}$ is a matrix-decreasing function ($\boldsymbol{A} \succeq \boldsymbol{B} \Rightarrow \boldsymbol{A}^{-1} \preceq \boldsymbol{B}^{-1}$) and therefore a boundary preserving mapping [17]. Note that the second step in (15) follows from the fact that $\boldsymbol{J}_P$ is not affected by the transformation $\boldsymbol{C}$. Using the substitution $\boldsymbol{C} = \boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{C}'$ it is possible to write

$$\boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{R}\boldsymbol{C})^{-1}\boldsymbol{C}^T = \boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{C}'(\boldsymbol{C}'^T\boldsymbol{C}')^{-1}\boldsymbol{C}'^T\boldsymbol{R}^{-\frac{1}{2}}. \tag{16}$$

By the SVD $\boldsymbol{C}' = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T$, we have $\boldsymbol{C}'(\boldsymbol{C}'^T\boldsymbol{C}')^{-1}\boldsymbol{C}'^T = \boldsymbol{U}\boldsymbol{U}^T$, such that (15) can be solved

$$\boldsymbol{C}'^\star = \arg\max_{\boldsymbol{C}'\in\mathbb{R}^{N\times M}} \mathrm{Tr}\left(\boldsymbol{W}'\boldsymbol{J}_D(\boldsymbol{C}')\right), \tag{17}$$

while, without loss of generality, restricting to matrices $\boldsymbol{C}' \in \mathcal{C}$ with orthonormal columns, i.e.

$$\boldsymbol{C}'^T\boldsymbol{C}' = \boldsymbol{I}_M. \tag{18}$$

By using (16), (18) the matrix (10) is

$$\boldsymbol{J}_D(\boldsymbol{C}') = \mathrm{E}_{\boldsymbol{\theta}}\left[\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \boldsymbol{R}^{-\frac{1}{2}} \boldsymbol{C}' \boldsymbol{C}'^T \boldsymbol{R}^{-\frac{1}{2}} \frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]. \quad (19)$$

Defining the matrix

$$\boldsymbol{\Omega} = \boldsymbol{R}^{-\frac{1}{2}} \mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{W}' \left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right] \boldsymbol{R}^{-\frac{1}{2}}, \quad (20)$$

problem (17) can be written

$$\begin{aligned} \boldsymbol{C}'^{\star} &= \arg\max_{\boldsymbol{C}' \in \mathcal{C}} \operatorname{Tr}\left(\boldsymbol{W}' \boldsymbol{J}_D(\boldsymbol{C}')\right) \\ &= \arg\max_{\boldsymbol{C}' \in \mathcal{C}} \operatorname{Tr}\left(\boldsymbol{C}'^T \boldsymbol{\Omega} \boldsymbol{C}'\right) \\ &= \arg\max_{\boldsymbol{C}' \in \mathcal{C}} \left(\sum_{m=1}^{M} \boldsymbol{c}'^T_m \boldsymbol{\Omega} \boldsymbol{c}'_m\right), \end{aligned} \quad (21)$$

where $\boldsymbol{c}'_m$ is the m-th column of $\boldsymbol{C}'$. We relax condition (18) and use instead $\|\boldsymbol{c}_m\|_2^2 = 1, \forall m$, with the restriction $\boldsymbol{c}_m \neq \boldsymbol{c}_n$ for $m \neq n$, such that the Lagrangian function is

$$L(\boldsymbol{C}', \boldsymbol{\lambda}) = \sum_{m=1}^{M} \boldsymbol{c}'^T_m \boldsymbol{\Omega} \boldsymbol{c}'_m - \lambda_m (\boldsymbol{c}'^T_m \boldsymbol{c}'_m - 1), \quad (22)$$

where $\boldsymbol{\lambda} = [\,\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_M\,]^T$ are Lagrangian multipliers. Differentiating with respect to $\boldsymbol{c}'_m$ for $m = 1, \dots, M$, a solution for (21) must satisfy $\boldsymbol{\Omega} \boldsymbol{c}'_m = \lambda^{\star}_m \boldsymbol{c}'_m, \forall m$, such that

$$\operatorname{Tr}\left(\boldsymbol{C}'^{\star T} \boldsymbol{\Omega} \boldsymbol{C}'^{\star}\right) = \sum_{m=1}^{M} \lambda^{\star}_m. \quad (23)$$

The maximum value for (23) is achieved by choosing the M columns of $\boldsymbol{C}'^{\star}$ according to the M orthogonal eigenvectors of $\boldsymbol{\Omega}$ corresponding to the largest eigenvalues. Note, that taking into account the derivatives in (20) (sensitivity-based) is the major difference to a KLT-based approach, where only the covariance of the signal itself is used (energy-based). The derivatives deliver the space of maximum sensitivity with respect to changes in the parameter $\boldsymbol{\theta}$.

V. Application - Satellite-based Positioning

As an application, we discuss satellite-based positioning with a GNSS receiver. In order to determine a position in time and space the receiver measures the distance to four or more synchronized in-view satellites. This is achieved by measuring the time-delay due to signal propagation from the satellites to the receiver. This is possible due to the knowledge of the satellite signal structure. With the unknown parameters $\boldsymbol{\theta} = [\,\gamma \quad \tau\,]^T$, the analog receive signal is

$$y(t) = s(t; \boldsymbol{\theta}) + \eta(t) = \gamma x(t - \tau) + \eta(t), \quad (24)$$

where γ is the signal strength, τ denotes a delay and

$$x(t) = \sum_{l=-\infty}^{+\infty} [\boldsymbol{b}]_l \cdot g(t - lT_c), \quad (25)$$

is the satellite signal, consisting of a binary sequence $\boldsymbol{b}$ and a transmit pulse $g(t)$. Here the signal GPS L1 C/A [18] is used, where $\boldsymbol{b}$ is periodic with 1023 chips, each with a duration of $T_c = 977.52$ ns. The pulse $g(t)$ is assumed to be rectangular. Band-limiting the analog receive signal by an ideal low-pass filter with one-sided bandwidth $B = \frac{1}{T_c} = 1.023$ MHz and sampling at a rate of $f_s = 2B$ for $T_o = 1$ ms, the digital receive signal can be written as

$$\boldsymbol{y} = \boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{\eta} = \gamma \boldsymbol{x}(\tau) + \boldsymbol{\eta}, \quad (26)$$

with $\boldsymbol{x}(\tau) \in \mathbb{R}^N$, $[\boldsymbol{x}(\tau)]_n = x(nT_s - \tau)$, $N = 2046$ and $\boldsymbol{R} = BN_0 \boldsymbol{I}_N$, where $\frac{N_0}{2}$ is the power spectral density of the analog noise $\eta(t)$ before low-pass filtering. For simplicity we consider independent parameters $p(\boldsymbol{\theta}) = p(\gamma, \tau) = p(\gamma)p(\tau)$ and priors $p(\gamma) \sim \mathcal{N}(\mu_\gamma, \sigma^2_\gamma)$, $p(\tau) \sim \mathcal{N}(\mu_\tau, \sigma^2_\tau)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. This results in a diagonal PIM

$$\boldsymbol{J}_P = \begin{bmatrix} \frac{1}{\sigma^2_\gamma} & 0 \\ 0 & \frac{1}{\sigma^2_\tau} \end{bmatrix}. \quad (27)$$

The matrix with the derivatives for all entries of $\boldsymbol{\theta}$ is

$$\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{x}(\tau) & \gamma \frac{\partial \boldsymbol{x}(\tau)}{\partial \tau} \end{bmatrix}, \quad (28)$$

such that the matrix $\boldsymbol{\Omega} \in \mathbb{R}^{N \times N}$ can be written as

$$\begin{aligned} \boldsymbol{\Omega} &= \frac{1}{BN_0} \mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{W}' \left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right] \\ &= \frac{w'_{11}}{BN_0} \boldsymbol{R}_{xx} + \frac{w'_{12}\mu_\gamma}{BN_0} \boldsymbol{R}_{x\partial x} \\ &\quad + \frac{w'_{21}\mu_\gamma}{BN_0} \boldsymbol{R}_{\partial x x} + \frac{w'_{22}(\sigma^2_\gamma + \mu^2_\gamma)}{BN_0} \boldsymbol{R}_{\partial x \partial x}, \end{aligned} \quad (29)$$

where

$$\begin{aligned} \boldsymbol{R}_{xx} &= \int_{-\infty}^{\infty} p(\tau) \boldsymbol{x}(\tau) \boldsymbol{x}^T(\tau) d\tau \\ \boldsymbol{R}_{\partial x \partial x} &= \int_{-\infty}^{\infty} p(\tau) \frac{\partial \boldsymbol{x}(\tau)}{\partial \tau} \left(\frac{\partial \boldsymbol{x}(\tau)}{\partial \tau}\right)^T d\tau \\ \boldsymbol{R}_{x \partial x} &= \boldsymbol{R}^T_{\partial x x} = \int_{-\infty}^{\infty} p(\tau) \boldsymbol{x}(\tau) \left(\frac{\partial \boldsymbol{x}(\tau)}{\partial \tau}\right)^T d\tau. \end{aligned} \quad (30)$$

With the eigenvalue decomposition $\boldsymbol{\Omega} = \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^H$, where $\boldsymbol{\Lambda}$ contains the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M \geq \dots \geq \lambda_N$ on its diagonal, the M principal eigenvectors

$$\boldsymbol{Q}_M = [\,\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \dots \quad \boldsymbol{q}_M\,] = \boldsymbol{C}'^{\star}, \quad (31)$$

lead to the desired transformation

$$\boldsymbol{C}^{\star} = \boldsymbol{R}^{-\frac{1}{2}} \boldsymbol{C}'^{\star} = \sqrt{\frac{1}{BN_0}} \boldsymbol{Q}_M. \quad (32)$$

Note that for $\boldsymbol{C}^{\star}$ the entries of the EFIM can be written as

$$\begin{aligned} [\boldsymbol{J}_D(\boldsymbol{C}^{\star})]_{11} &= \frac{1}{BN_0} \mathrm{E}_\tau \left[\boldsymbol{x}^T(\tau) \boldsymbol{C}'^{\star} \boldsymbol{C}'^{\star T} \boldsymbol{x}(\tau)\right] \\ &= \frac{1}{BN_0} \mathrm{E}_\tau \left[\operatorname{Tr}\left(\boldsymbol{C}'^{\star T} \boldsymbol{x}(\tau) \boldsymbol{x}^T(\tau) \boldsymbol{C}'^{\star}\right)\right] \\ &= \frac{1}{BN_0} \operatorname{Tr}\left(\boldsymbol{C}'^{\star T} \boldsymbol{R}_{xx} \boldsymbol{C}'^{\star}\right) \\ [\boldsymbol{J}_D(\boldsymbol{C}^{\star})]_{12/21} &= \frac{\mu_\gamma}{BN_0} \operatorname{Tr}\left(\boldsymbol{C}'^{\star T} \boldsymbol{R}_{\partial x x} \boldsymbol{C}'^{\star}\right) \\ [\boldsymbol{J}_D(\boldsymbol{C}^{\star})]_{22} &= \frac{(\sigma^2_\gamma + \mu^2_\gamma)}{BN_0} \operatorname{Tr}\left(\boldsymbol{C}'^{\star T} \boldsymbol{R}_{\partial x \partial x} \boldsymbol{C}'^{\star}\right). \end{aligned} \quad (33)$$

VI. Performance - Information-Loss

To visualize the quality of $\mathcal{P}$ (Fig. 1), we evaluate

$$\boldsymbol{\chi}_{\gamma/\tau}(\boldsymbol{C}^{\star}) = \frac{[\boldsymbol{J}_B^{-1}(\boldsymbol{I}_N)]_{11/22}}{[\boldsymbol{J}_B^{-1}(\boldsymbol{C}^{\star})]_{11/22}}, \tag{34}$$

with priors $10\log\mu_\gamma^2 = 50$ dB, $10\log\sigma_\gamma^2 = 10$ dB, $\mu_\tau = 0$ ns and $\sigma_\tau = \frac{T_c}{3}$ for the solutions (32) attained by sampling over all positive-semidefinite matrices

$$\boldsymbol{W'} = \begin{bmatrix} \alpha & \beta \\ \beta & (1-\alpha) \end{bmatrix}, \tag{35}$$

with $0 \leq \alpha \leq 1$ and $|\beta| \leq \sqrt{\alpha(1-\alpha)}$. It can be observed (Fig. 1), that $M = 4$ is sufficient to perform estimation in the considered scenario with a negligible performance-loss. Results for $M < 4$ show a significant loss in the MSE region.

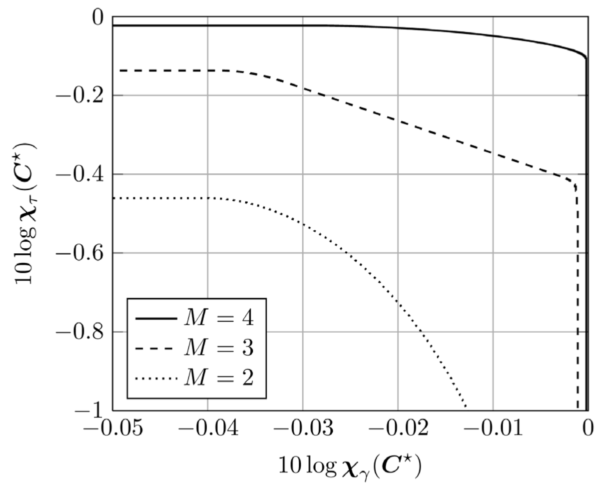

Fig. 1. MSE-Loss Region ($\mu_\gamma^2 = 50$ dB, $\sigma_\gamma^2 = 10$ dB, $\mu_\tau = 0, \sigma_\tau = \frac{T_c}{3}$).

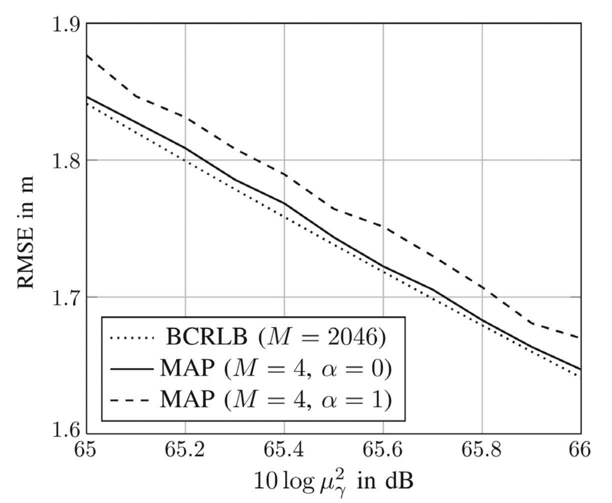

Fig. 2. Performance - MAP ($\sigma_\gamma^2 = 10$ dB, $\mu_\tau = 0, \sigma_\tau = \frac{T_c}{3}$).

VII. Performance - Compressed MAP

In order to analyze the performance of a real estimator with compressed data, we simulate the MAP estimator

$$\hat{\boldsymbol{\theta}}_{c,\mathrm{MAP}} = \arg\max_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_c|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta}). \tag{36}$$

For the signal (26) and Gaussian priors, the MAP estimator operating on $\boldsymbol{y}_c = \boldsymbol{C'}^T\boldsymbol{y}$, can be calculated by

$$\begin{aligned} \hat{\tau}_c &= \arg\max_{\tau} \frac{\left(\boldsymbol{y}_c^T\boldsymbol{R}_c^{-1}\boldsymbol{s}_c(\tau) + \frac{\mu_\gamma}{\sigma_\gamma^2}\right)^2}{\left(\boldsymbol{s}_c^T(\tau)\boldsymbol{R}_c^{-1}\boldsymbol{s}_c(\tau) + \frac{1}{\sigma_\gamma^2}\right)} - \frac{(\tau-\mu_\tau)^2}{\sigma_\tau^2} \\ \hat{\gamma}_c &= \frac{\left(\boldsymbol{y}_c^T\boldsymbol{R}_c^{-1}\boldsymbol{s}_c(\hat{\tau}_c) + \frac{\mu_\gamma}{\sigma_\gamma^2}\right)}{\left(\boldsymbol{s}_c^T(\hat{\tau}_c)\boldsymbol{R}_c^{-1}\boldsymbol{s}_c(\hat{\tau}_c) + \frac{1}{\sigma_\gamma^2}\right)}. \end{aligned} \tag{37}$$

Fig. 2 shows the estimation performance of the MAP under priors $10\log\sigma_\gamma^2 = 10$ dB, $\mu_\tau = 0$ ns and $\sigma_\tau = \frac{T_c}{3}$, while $10\log\mu_\gamma^2$ is varied between $65.0 - 66.0$ dB. The MAP estimators working on the data $\boldsymbol{y}_c$, which results from the two projections $\alpha = 1, \beta = 0$ and $\alpha = \beta = 0$ with $M = 4$, are compared to the achievable performance (BCRLB) with the original data $\boldsymbol{y}$. It is observed that by using the optimized transformation which favors the estimation of τ, i.e. $\alpha = 0$, it is possible to achieve estimation accuracy close to the estimator which is using the original data $\boldsymbol{y}$. The estimator using the transformation with $\alpha = 1$ favors the parameter γ and is therefore not able to achieve the full performance with respect to τ. Note that here the special case $\alpha = 1$ coincides with the classical energy-based approaches such as KLT, PCA, CCA and GIB, while $\alpha = 0$ is a pure sensitivity-based approach with full weight on τ.

In this example the complexity of the MAP without compression consists of SN multiplications, where S is the number of evaluations of the the MAP function during the maximization (36). For the proposed approach we require MN multiplications for the linear transformation and SM for the compressed MAP. Note that the transformation matrix $\boldsymbol{C}$ has only to be calculated once, which can be done offline. For instance, here we have used $S = 100$, such that $\frac{MN+SM}{SN} \approx 0.042$, i.e. the runtime complexity of the estimation process is reduced to four percent at the cost of a small performance-loss.

VIII. Conclusion

An approach to the problem of signal parameter estimation from large data sets has been presented, resulting in a Bayesian data reduction method through linear transforms. Establishing a tractable formulation for the problem of finding the optimum transformation with respect to the weighted MSE criterium, the projection matrix is found by an eigenvalue decomposition of a weighted sum of covariance matrices. These belong to the derivatives of the noiseless signal with respect to all signal parameters. The presented results show that with an optimized linear transform the size of the data can be reduced significantly without loosing the possibility of high-accuracy parameter estimation. Simulation of the MAP estimator for a GNSS scenario, show that the predicted theoretical estimation performance can be achieved with the transformed data. The general formulation of our sensitivity-based approach allows to adjust the transformation such that the parameters of interest are hardly affected by the information-loss due to dimensionality reduction.

REFERENCES

[1] K. K. Karhunen, “Über lineare Methoden in der Wahrscheinlichkeitsrechnung,” *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys*, no. 37, pp. 1–79, 1947.

[2] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *J. Educ. Psychol.*, vol. 24, pp. 417–441, Sep. 1933.

[3] T. R. Knapp, “Canonical correlation analysis: A general parametric significance-testing system,” *Psychol. Bull.*, vol. 85, no. 2, pp. 410–416, 1978.

[4] A. Emad and O. Milenkovic, “Compression of noisy signals with information bottlenecks,” in *IEEE Inf. Theory Workshop (ITW)*, Sep. 2013.

[5] J. Selva, “Complexity reduction in the parametric estimation of superimposed signal replicas,” *Signal Process.*, vol. 84, no. 12, pp. 2325–2343, Dec. 2004.

[6] I. Groh and W. Utschick, “High resolution reduced complexity synchronization in spread spectrum systems,” in *Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2008.

[7] L. L. Scharf and D. W. Tufts, “Rank reduction for modeling stationary signals,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 350–355, Mar. 1987.

[8] Y. Hua, M. Nikpour, and P. Stoica, “Optimal reduced-rank estimation and filtering,” *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 457–469, Mar. 2001.

[9] M. L. Honig and J. S. Goldstein, “Adaptive reduced-rank interference suppression based on the multistage Wiener filter,” *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 986–994, Jun. 2002.

[10] F. Dietrich, M. T. Ivrlac, and J. A. Nossek, “On performance limits of optimum reduced rank channel estimation,” in *IEEE Global Telecommun. Conf. GLOBECOM, 2002*, Nov. 2002, vol. 1, pp. 345–349.

[11] R. C. De Lamare and R. Sampaio-Neto, “Reduced-rank adaptive filtering based on joint iterative optimization of adaptive filters,” *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 980–983, Dec. 2007.

[12] L. Wang, R. C. De Lamare, and M. Haardt, “Reduced-rank DOA estimation based on joint iterative subspace recursive optimization and grid search,” in *IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP), 2010*, Mar. 2010, pp. 2626–2629.

[13] S. W. Provencher and R. H. Vogel, “Information loss with transform methods in system identification: A new set of transforms with high information content,” *Math. Biosci.*, vol. 50, pp. 251–262, 1980.

[14] M. Tegmark, A. Taylor, and A. Heavens, “Karhunen-Loeve eigenvalue problems in cosmology: How should we tackle large data sets?,” *Astrophys. J.*, vol. 480, pp. 22–35, May 1997.

[15] M. Chen and M. L. Fowler, “Data compression for multi-parameter estimation for emitter location,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, Jan. 2010.

[16] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Piscataway, NJ, USA: Wiley-IEEE Press, 2007.

[17] G. A. Swarup, “Boundary preserving maps of 3-manifolds,” *Proc. Amer. Math. Soc.*, vol. 78, no. 2, Feb. 1980.

[18] NAVSTAR GPS Space Segment/Navigation User Interfaces, IS-GPS-200, Rev. D, ARINC Engineering Services. El Segundo, CA, USA, 2004.