

AUTOMATIC ROBUST ADAPTIVE BEAMFORMING VIA RIDGE REGRESSION

Yngve Selén, Richard Abrahamsson and Petre Stoica

Dept. of Information Technology
Uppsala University
P.O. Box 337, SE-751 05 Uppsala, Sweden.
Email: ys@it.uu.se

ABSTRACT

In this paper we derive a class of new parameter free robust adaptive beamformers using the generalized sidelobe canceler reparameterization of the Capon beamformer. In this parameterization the minimum variance beamformer is obtained as the solution of a linear least squares problem. In the case of an inaccurate steering vector and/or few data snapshots this marginally overdetermined system gives an ill fit causing signal cancellation in the standard minimum variance solution. By regularizing the problem using ridge regression techniques we get a whole class of robust adaptive beamformers, none of which requires the choice of a user parameter. We also propose a novel empirical Bayes-based ridge regression technique. The performance is compared to other robust adaptive beamformers.

Index Terms— minimum variance beamforming, Capon beamforming, robust beamforming, ridge regression, regularization

1. INTRODUCTION

Data dependent beamformers [1, 2] have attracted lots of attention due to their potential to adaptively suppress interference and noise in an optimal manner. In particular, the standard Capon beamformer (SCB) appears to be fundamental in the sense that it can be derived from several different starting points (see, e.g., [2, 3]). It is well-known that the interference rejection capabilities of the SCB can in some cases lead to cancellation of the signal of interest. This happens, e.g., when the array is not perfectly calibrated. Also a limited number of data snapshots or correlated signal and interferences can lead to severe signal cancellation.

To mitigate this, robust Capon beamformers (RCBs) have been designed by assuming that the array steering vector belongs to an ellipsoidal uncertainty set [4, 5, 6]. These RCBs are dependent on the choice of a user parameter related to the size of the uncertainty set. However, there is no clear-cut solution to how one should choose this user parameter when the signal cancellation is due to signal correlated interferences or limited data.

The goal of this paper is to robustify the SCB against errors in the steering vectors and the data covariance matrix. Specifically, we are interested in completely automatic methods which do not require the choice of any user parameters. This is in contrast to most robust adaptive beamformers in the literature.

2. PROBLEM FORMULATION

Assume that a narrowband signal, $s(t)$, impinge on an m element array of sensors from a certain direction θ_s , possibly together with

directional interferences, $i_k(t)$, from other directions θ_{i_k} . The interferences are assumed to be uncorrelated with the signal of interest. The array data snapshots $\mathbf{x}(t) \in \mathbb{C}^m$ have the form

$$\mathbf{x}(t) = \mathbf{s}(t) + \mathbf{i}(t) + \mathbf{n}(t) = [\mathbf{a}_s \ \mathbf{a}_{i_1} \ \cdots \ \mathbf{a}_{i_{n-1}}] \begin{bmatrix} s(t) \\ i_1(t) \\ \vdots \\ i_{n-1}(t) \end{bmatrix} + \mathbf{n}(t)$$

for $t = 1, \dots, N$, where $\mathbf{s}(t)$, $\mathbf{i}(t)$ and $\mathbf{n}(t)$ are the contributions to the observed data from the signal of interest, the interferences and the noise (which may contain non-directional interferences), respectively. The steering vector for the signal is \mathbf{a}_s and for the interferences they are \mathbf{a}_{i_k} , where we have dropped the dependences on θ_s and θ_{i_k} for notational convenience. For an arbitrary steering vector in the array manifold we use \mathbf{a} . Without loss of generality we will assume that all steering vectors are normalized such that $\|\mathbf{a}\|^2 = m$.

For some complex array weight vector $\mathbf{w} \in \mathbb{C}^m$, the output of a beamformer is $y(t) = \mathbf{w}^H \mathbf{x}(t)$, where one ideally would like to null the interferences and noise while allowing the signal of interest to pass undistorted.

2.1. Adaptive Beamformers

The standard Capon beamformer (SCB) minimizes the output power subject to a unit gain constraint:

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a} = 1 \quad (1)$$

where $\mathbf{R} \in \mathbb{C}^{m \times m}$ is the data covariance matrix. In practice, \mathbf{R} is usually replaced by the sample covariance matrix:

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t) \mathbf{x}^H(t).$$

(In the numerical examples in Section 4, we will use $\hat{\mathbf{R}}$ in lieu of \mathbf{R} for all methods.) The solution to (1) is well-known to be [3]

$$\mathbf{w}_{\text{SCB}} = \frac{\mathbf{R}^{-1} \mathbf{a}}{\mathbf{a}^H \mathbf{R}^{-1} \mathbf{a}}. \quad (2)$$

One of the more recent RCBs [6] uses the alternative covariance fitting formulation of (1) with the added constraint that the true steering vector $\tilde{\mathbf{a}}$ should be within an uncertainty set centered on the assumed steering vector \mathbf{a} :

$$\left\{ \hat{\sigma}^2, \hat{\tilde{\mathbf{a}}} \right\} = \arg \max_{\sigma^2, \tilde{\mathbf{a}}} \sigma^2 \quad \text{s.t.} \quad \mathbf{R} - \sigma^2 \tilde{\mathbf{a}} \tilde{\mathbf{a}}^H \geq 0$$

$$\text{and} \quad \|\mathbf{a} - \tilde{\mathbf{a}}\|^2 \leq \epsilon. \quad (3)$$

Once the maximizing $\hat{\mathbf{a}}$ is found it is appropriate to rescale it so that $\|\hat{\mathbf{a}}\|^2 = m$, see [6, 7]. The weight vector is then found as

$$\hat{\mathbf{w}}_{\text{RCB}} = \frac{\mathbf{R}^{-1}\hat{\mathbf{a}}}{\hat{\mathbf{a}}^H \mathbf{R}^{-1}\hat{\mathbf{a}}}. \quad (4)$$

3. RIDGE REGRESSION-BASED BEAMFORMING

3.1. Beamforming in a Linear Regression Framework

Consider the SCB in (1). The weight vector \mathbf{w} can be reparameterized by a new parameter vector $\boldsymbol{\eta}$ according to

$$\mathbf{w} = \frac{\mathbf{a}}{m} - \mathbf{Q}\boldsymbol{\eta} \quad (5)$$

where $\mathbf{Q} \in \mathbb{C}^{m \times (m-1)}$ is a semi-unitary matrix so that $\mathbf{Q}^H \mathbf{a} = \mathbf{0}$, $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}$, and $\boldsymbol{\eta} \in \mathbb{C}^{m-1}$. This reparameterization is known from the generalized sidelobe canceler [2]. \mathbf{Q} can most efficiently be obtained by the QR-decomposition of \mathbf{a} where \mathbf{Q} consists of the last $m-1$ columns of the unitary matrix. Now, (1) can be rewritten as

$$\min_{\boldsymbol{\eta}} \left(\mathbf{Q}\boldsymbol{\eta} - \frac{\mathbf{a}}{m} \right)^H \mathbf{R} \left(\mathbf{Q}\boldsymbol{\eta} - \frac{\mathbf{a}}{m} \right) = \min_{\boldsymbol{\eta}} \left\| \underbrace{\mathbf{R}^{1/2} \mathbf{Q}}_{\triangleq \mathbf{X}} \boldsymbol{\eta} - \underbrace{\mathbf{R}^{1/2} \frac{\mathbf{a}}{m}}_{\triangleq \mathbf{b}} \right\|^2 \quad (6)$$

where $\mathbf{R}^{1/2}$ denotes the positive definite Hermitian square root of \mathbf{R} and $\mathbf{X} \in \mathbb{C}^{m \times (m-1)}$, $\mathbf{b} \in \mathbb{C}^m$. Now, (6) can be interpreted as the marginally overdetermined least squares (LS) formulation of the linear regression problem

$$\mathbf{b} = \mathbf{X}\boldsymbol{\eta} + \mathbf{e} \quad (7)$$

where \mathbf{e} is a residual term. The SCB (1) is obtained from (7) using the standard LS estimator (see (6)):

$$\hat{\boldsymbol{\eta}}_{\text{LS}} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{b}. \quad (8)$$

However, by using different approaches for estimating $\boldsymbol{\eta}$, other adaptive beamformers can be obtained. Specifically, we are interested in using regularization methods to obtain more stable estimators. This can be motivated by the fact that, in the linear regression framework, estimating $m-1$ parameters in $\boldsymbol{\eta}$ from only m samples in \mathbf{b} often leads to poor estimates which are very sensitive to noise. For the case of an uncalibrated array we can indeed expect noise to be a problem, since uncertainties in \mathbf{b} are not only due to additive noise in the data via the sample covariance matrix, but also from errors in \mathbf{a} . Furthermore, the benefits of an adaptive beamformer are usually most significant when the directional interferences are much larger than the noise level. For such a case, the regressor matrix $\mathbf{X} = \mathbf{R}^{1/2} \mathbf{Q}$ ‘‘inherits’’ a low condition number from \mathbf{R} which means that errors in (7) are amplified in $\hat{\boldsymbol{\eta}}_{\text{LS}}$.

3.2. Ridge Regression

Ridge regression (RR) is a common approach to regularize ill-conditioned linear regression problems. The so-called ordinary RR estimate of the problem (7) is defined as [8]

$$\hat{\boldsymbol{\eta}}_{\text{RR}}(\rho) = (\mathbf{X}^H \mathbf{X} + \rho \mathbf{I})^{-1} \mathbf{X}^H \mathbf{b} \quad (9)$$

where $\rho > 0$ is a user parameter. This is called the ordinary RR estimate because the same load ρ is applied to all diagonal elements in the matrix to be inverted in (9). In contrast, the generalized RR

(see (11) below) uses different loads for the corresponding diagonal elements. The problems of regression and RR are well studied in the statistical literature. Some methods for automatically choosing ρ will be discussed and evaluated below.

The RR methods to be presented here are, just as most RR methods in the literature, constructed for real-valued problems. However, our linear regression problem (7) is complex-valued. To cope with this, we use the following real-valued version of (7):

$$\bar{\mathbf{b}} = \bar{\mathbf{X}}\bar{\boldsymbol{\eta}} + \bar{\mathbf{e}} \quad (10)$$

where we use a bar ($\bar{\cdot}$) to denote the real-valued version of a vector, $\bar{\mathbf{z}} = \begin{bmatrix} \text{Re } \mathbf{z} \\ \text{Im } \mathbf{z} \end{bmatrix}$, and $\bar{\mathbf{X}} = \begin{bmatrix} \text{Re } \mathbf{X} & -\text{Im } \mathbf{X} \\ \text{Im } \mathbf{X} & \text{Re } \mathbf{X} \end{bmatrix}$. We present the

RR methods in a general framework where $\bar{\mathbf{b}} \in \mathbb{R}^{\bar{m}}$, $\bar{\mathbf{X}} \in \mathbb{R}^{\bar{m} \times \bar{p}}$, $\bar{\boldsymbol{\eta}} \in \mathbb{R}^{\bar{p}}$ and $\bar{\mathbf{e}} \in \mathbb{R}^{\bar{m}}$. In our specific application we have $\bar{p} \triangleq 2(m-1)$ and $\bar{m} \triangleq 2m$. The LS and RR estimates $\hat{\boldsymbol{\eta}}_{\text{LS}}$, $\hat{\boldsymbol{\eta}}_{\text{RR}}$ are defined analogously to (8), (9) (note that $\hat{\boldsymbol{\eta}}_{\text{LS}} \equiv \hat{\boldsymbol{\eta}}_{\text{LS}}$ and $\hat{\boldsymbol{\eta}}_{\text{RR}}(\rho) \equiv \hat{\boldsymbol{\eta}}_{\text{RR}}(\rho)$).

Many of the methods below are based on assuming a normally distributed $\bar{\mathbf{e}}$ and sometimes a normally distributed $\bar{\boldsymbol{\eta}}$. It is difficult to say how well such assumptions will hold in practice. However, they are often used for computational convenience and tractability. Also, as will be shown below, these assumptions lead to well performing estimators in the current application.

In [8], a Generalized RR (GRR) estimate was derived based on the following orthogonal version of (10):

$$\bar{\mathbf{b}} = \mathbf{Z}\boldsymbol{\alpha} + \bar{\mathbf{e}}$$

where $\mathbf{Z} = \bar{\mathbf{X}}\mathbf{P}$, $\boldsymbol{\alpha} = \mathbf{P}^T \bar{\boldsymbol{\eta}}$ and $\bar{\mathbf{X}}^T \bar{\mathbf{X}} = \mathbf{P} \text{diag}\{d_1, \dots, d_{\bar{p}}\} \mathbf{P}^T$, $\mathbf{P}^T \mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$ is the eigenvalue decomposition of $\bar{\mathbf{X}}^T \bar{\mathbf{X}}$. The GRR estimator is defined by

$$\begin{aligned} \hat{\boldsymbol{\eta}}_{\text{GRR}} &= \mathbf{P}\hat{\boldsymbol{\alpha}}_{\text{GRR}} = \mathbf{P}(\mathbf{Z}^T \mathbf{Z} + \text{diag}\{\rho_1, \dots, \rho_{\bar{p}}\})^{-1} \mathbf{Z}^T \bar{\mathbf{b}} \\ &= \mathbf{P} \text{diag}\{(d_1 + \rho_1)^{-1}, \dots, (d_{\bar{p}} + \rho_{\bar{p}})^{-1}\} \mathbf{Z}^T \bar{\mathbf{b}}. \end{aligned} \quad (11)$$

If $\bar{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, the values of ρ_i that minimize the mean square error (MSE) of $\hat{\boldsymbol{\eta}}_{\text{GRR}}$ are given by $\rho_i = \sigma^2 / \alpha_i^2$ [8]. Since both σ^2 and $\boldsymbol{\alpha}$ are unknown in practice, they are replaced by their estimates: $\hat{\rho}_i = \hat{\sigma}^2 / \hat{\alpha}_{i,\text{LS}}^2$ where $\hat{\boldsymbol{\alpha}}_{\text{LS}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \bar{\mathbf{b}}$ and

$$\hat{\sigma}^2 = \frac{\|\bar{\mathbf{b}} - \bar{\mathbf{X}}\hat{\boldsymbol{\eta}}_{\text{LS}}\|^2}{\bar{m} - \bar{p}} \quad (12)$$

which will be used throughout.

In [9] Hoerl, Kennard and Baldwin suggested using the harmonic mean of the $\{\hat{\rho}_i\}_{i=1}^{\bar{p}}$ above as a value for ρ in the ordinary RR estimate (9):

$$\hat{\rho}_{\text{HKB}} = \frac{\bar{p}\hat{\sigma}^2}{\hat{\boldsymbol{\eta}}_{\text{LS}}^T \hat{\boldsymbol{\eta}}_{\text{LS}}}. \quad (13)$$

Based on Bayesian arguments, assuming Gaussian priors for both $\bar{\mathbf{e}}$ and $\bar{\boldsymbol{\eta}}$, Lawless and Wang [10] proposed

$$\hat{\rho}_{\text{LW}} = \frac{\bar{p}\hat{\sigma}^2}{\hat{\boldsymbol{\eta}}_{\text{LS}}^T \bar{\mathbf{X}}^T \bar{\mathbf{X}} \hat{\boldsymbol{\eta}}_{\text{LS}}}. \quad (14)$$

Finally, in [11] an *information complexity*-based regularization method was presented. Assuming $\bar{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$,

$$\hat{\rho}_{\text{ICOMP}} = \arg \min_{\rho} \frac{1}{\hat{\sigma}^2} \|\bar{\mathbf{b}} - \bar{\mathbf{X}}\hat{\boldsymbol{\eta}}_{\text{RR}}(\rho)\|^2 + 2\text{Tr}\{\mathbf{H}\} + 2C_1(\mathbf{F}^{-1}) \quad (15)$$

where $\mathbf{H} = \bar{\mathbf{X}}(\bar{\mathbf{X}}^T \bar{\mathbf{X}} + \rho \mathbf{I})^{-1} \bar{\mathbf{X}}^T$, $\text{Tr}\{\cdot\}$ denotes the trace operator, $C_1(\mathbf{F}^{-1}) = \frac{p}{2} [\log(\sum_{j=1}^p \nu_j / \bar{p}) - \log(\prod_{j=1}^p \nu_j) / \bar{p}]$, where $\{\nu_j\}_{j=1}^p$ are the singular values of $\mathbf{F}^{-1} = (\bar{\mathbf{X}}^T \bar{\mathbf{X}} + \rho \mathbf{I})^{-1}$. Also, Mallows' C_L [12] is obtained using [11]

$$\hat{\rho}_{\text{CL}} = \arg \min_{\rho} \|\bar{\mathbf{b}} - \bar{\mathbf{X}} \hat{\boldsymbol{\eta}}_{\text{RR}}(\rho)\|^2 + 2\hat{\sigma}^2 \text{Tr}\{\mathbf{H}\}. \quad (16)$$

3.3. Empirical Bayesian Ridge Regression

In this subsection we consider an empirical Bayesian approach for estimating ρ . Although the estimator is based on quite natural assumptions, we have not been able to find it in the literature. Our approach is presented in the complex-valued setting. Similarly to the previous section, for generality of our approach, we use p to denote the length of $\boldsymbol{\eta}$ (in our application $p \triangleq m - 1$).

Assume $\boldsymbol{\eta}$ and e are independent random variables with the following distributions:

$$\boldsymbol{\eta} \sim \mathcal{CN}(0, \alpha^2 \mathbf{I}), \quad e \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}). \quad (17)$$

It is easy to show that (17) leads to a Gaussian prior on the weight vector: $\mathbf{w} \sim \mathcal{CN}(\mathbf{a}/m, \alpha^2 \mathbf{Q} \mathbf{Q}^H)$, where $E(\mathbf{w}) = \mathbf{a}/m$ is nothing but the standard beamformer (BF) [3].

Under the assumptions (17), the MMSE estimate of $\boldsymbol{\eta}$ is well-known to be [13]

$$\hat{\boldsymbol{\eta}}_{\text{MMSE}} = \alpha^2 \mathbf{X}^H (\alpha^2 \mathbf{X} \mathbf{X}^H + \sigma^2 \mathbf{I})^{-1} \mathbf{b} = (\mathbf{X}^H \mathbf{X} + \rho \mathbf{I})^{-1} \mathbf{X}^H \mathbf{b} \quad (18)$$

where $\rho = \sigma^2 / \alpha^2$ was used in the second equality. Note that (18) is exactly the RR estimate (9), but with a special interpretation of the ridge parameter ρ . As before, the expression (18) can not be used directly, unless ρ , i.e. σ^2, α^2 , is known. We suggest replacing σ^2, α^2 by their ML estimates.

Under the current model, $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \alpha^2 \mathbf{X} \mathbf{X}^H + \sigma^2 \mathbf{I})$, so

$$p(\mathbf{b} | \alpha^2, \sigma^2) = \frac{\exp\{-\mathbf{b}^H (\alpha^2 \mathbf{X} \mathbf{X}^H + \sigma^2 \mathbf{I})^{-1} \mathbf{b}\}}{\pi^m |\alpha^2 \mathbf{X} \mathbf{X}^H + \sigma^2 \mathbf{I}|}. \quad (19)$$

The ML estimates of σ^2, α^2 (or ρ) are given by

$$\begin{aligned} \arg \max_{\alpha^2, \sigma^2} p(\mathbf{b} | \alpha^2, \sigma^2) &= \arg \min_{\alpha^2, \sigma^2} -\ln p(\mathbf{b} | \alpha^2, \sigma^2) \\ &= \arg \min_{\alpha^2, \sigma^2} \mathbf{b}^H (\alpha^2 \mathbf{X} \mathbf{X}^H + \sigma^2 \mathbf{I})^{-1} \mathbf{b} + \ln |\alpha^2 \mathbf{X} \mathbf{X}^H + \sigma^2 \mathbf{I}| \\ &= \arg \min_{\rho, \sigma^2} \frac{\mathbf{b}^H (\rho^{-1} \mathbf{X} \mathbf{X}^H + \mathbf{I})^{-1} \mathbf{b}}{\sigma^2} + m \ln \sigma^2 + \ln |\rho^{-1} \mathbf{X} \mathbf{X}^H + \mathbf{I}|. \end{aligned} \quad (20)$$

For any value of ρ , the minimizing σ^2 is available in closed form:

$$\hat{\sigma}_{\text{ML}}^2(\rho) = \frac{\mathbf{b}^H (\rho^{-1} \mathbf{X} \mathbf{X}^H + \mathbf{I})^{-1} \mathbf{b}}{m}. \quad (21)$$

Inserting (21) in (20) the function left for minimization becomes

$$\hat{\rho}_{\text{ML}} = \arg \min_{\rho > 0} m \ln [\mathbf{b}^H (\rho^{-1} \mathbf{X} \mathbf{X}^H + \mathbf{I})^{-1} \mathbf{b}] + \ln |\rho^{-1} \mathbf{X} \mathbf{X}^H + \mathbf{I}|. \quad (22)$$

In other words, we only need to perform the 1D minimization over ρ in (22). One might consider computing (22) over a grid for ρ , or resort to some more refined numerical optimization method. (Note that $\hat{\rho}_{\text{ML}}$ can be found as one of the roots of a polynomial of degree $2m - 1$. The proof is omitted for brevity.)

4. NUMERICAL EXAMPLES

We evaluate the methods using Monte-Carlo simulations considering $N = 20$ data samples from a ULA with $m = 10$ omnidirectional sensors with half wavelength spacing. The assumed steering vectors \mathbf{a} (both of the source of interest and of the interferences) are perturbed by white Gaussian noise such that the actual (unknown) steering vectors are $\tilde{\mathbf{a}} = \mathbf{a} + \boldsymbol{\delta}$ where $\boldsymbol{\delta} \sim \mathcal{CN}(\mathbf{0}, \gamma^2 \mathbf{I})$. We set $\gamma^2 = 0.01$ (i.e., the standard deviation of the error corresponds to 10% of the complex gain of each array element). We have three independent and temporally white complex Gaussian farfield signals, with the powers $\{1, 1, s^2\}$ (s^2 will be varied), impinging from the directions of arrival (DOAs) $\{-30^\circ, 0^\circ, 10^\circ\}$. The noise, $\mathbf{n}(t)$, is spatially and temporally white and it has a complex Gaussian zero-mean distribution with variance $\sigma_n^2 = 0.02$. We consider the first and second source as interferences and the third as our source of interest.

As performance measure we use the mean signal-to-interference-plus-noise ratio (SINR), when looking in the true direction of the source of interest (i.e., 10°):

$$\text{mean SINR} = \frac{1}{M} \sum_{j=1}^M \frac{s^2 |\hat{\mathbf{w}}_j^H \tilde{\mathbf{a}}_j|^2}{\hat{\mathbf{w}}_j^H \mathbf{R}_j^{i+n} \hat{\mathbf{w}}_j}$$

where $\mathbf{R}_j^{i+n} = E\{(\mathbf{i}_j(t) + \mathbf{n}(t))(\mathbf{i}_j(t) + \mathbf{n}(t))^H\} \in \mathbb{C}^{m \times m}$ is the (true) interference-plus-noise covariance matrix, $\hat{\mathbf{w}}_j$ is the array weight vector (obtained using $\hat{\mathbf{R}}_j$ and the assumed steering vectors \mathbf{a}), the sub-index j denotes the simulation number, and $M = 1000$ denotes the total number of Monte-Carlo simulations. (Note that \mathbf{R}_j^{i+n} varies with j , because $\mathbf{i}_j(t)$ is subject to different steering vector perturbations at each iteration.) For comparison, we include the mean optimal SINR [4]:

$$\text{mean SINR}_{\text{opt}} = \frac{1}{M} \sum_{j=1}^M s^2 \tilde{\mathbf{a}}_j^H (\mathbf{R}_j^{i+n})^{-1} \tilde{\mathbf{a}}_j. \quad (23)$$

We also consider the detection frequency versus the false alarm rate in an energy detector. This is further described in Section 4.2.

We evaluate the performance of the following methods: (a) The standard beamformer [3]; (b) the SCB (1); (c) the RCB (4) with $\epsilon = E\{\|\boldsymbol{\delta}\|^2\} = 0.1$ (i.e., the expected squared norm of the perturbation); (d) the RCB (4) with $\epsilon = 0.175$ set such that $\text{Prob}(\epsilon < \|\boldsymbol{\delta}\|^2) = 0.98$; (e) GRR (11); (f) HKB-RR (13); (g) LW-RR (14); (h) ICOMP-RR (15); (i) C_L -RR (16); and (j) ML-RR (22). For increased readability of our plots, however, we do not include the curves of (e) and (i). Of the omitted methods, GRR generally gives slightly higher SINRs than BF and SCB. C_L performs better and similarly to ICOMP. However, neither of these methods are among the methods delivering the best performance.

For the methods for which ρ is not available in closed form, i.e. for ICOMP (15), C_L (16) and ML (22), we compute the relevant expressions over a fine grid for ρ , from 0 up to the largest eigenvalue of $\hat{\mathbf{R}}$. (We also tried substantially larger values in the grid, but no significant difference was observed in the plots.)

4.1. Example 1: SINR performance for varying input SNR

In this example we vary s^2 between 0.002 and 0.2 such that the input SNR (signal-to-noise ratio) goes from -10 dB to 10 dB. The mean SINRs can be seen in Figure 1. We note that HKB, LW and ML show the best performance. SCB does not perform well, and in fact

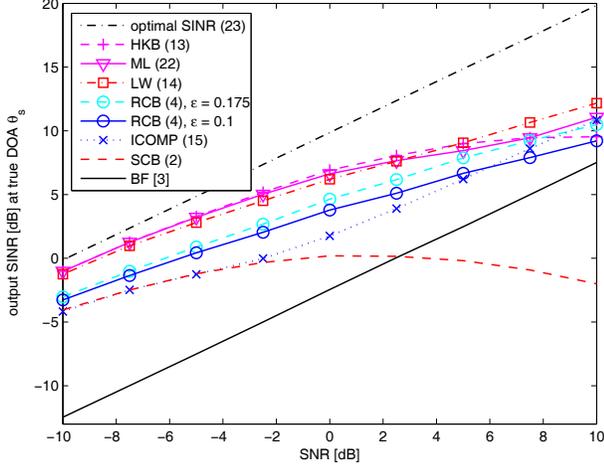


Fig. 1. Example 1: Mean output SINRs for varying input SNR.

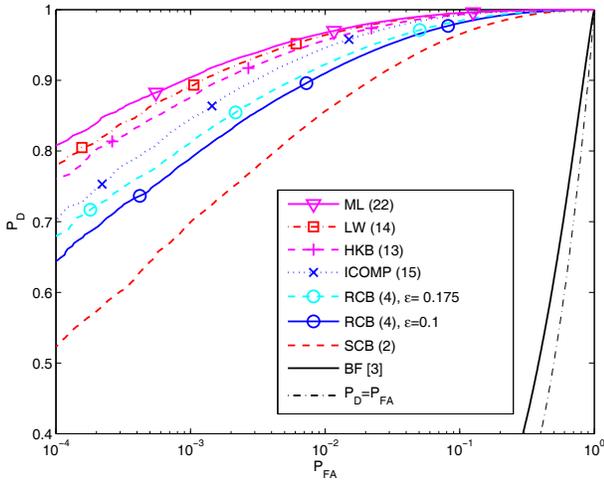


Fig. 2. Detection rate (P_D) versus false alarm frequencies (P_{FA}) for the energy detector for varying thresholds on the detection energy.

shows decreasing SINR as the SNR increases. This type of behavior has been observed before [4]. The reason why it occurs is because SCB tends to see the signal as an interference, and when the signal gets stronger, SCB uses more of its degrees of freedom to cancel it.

4.2. Example 2: Detection

Here we consider the performance of an energy detector applied to the beamformed data $y(t)$, i.e., $\sum_{t=1}^N |y(t)|^2 \underset{H_0}{\overset{H_1}{\geq}} \tau$, where H_1 (the signal present hypothesis) is chosen over H_0 (the interference plus noise only hypothesis) if the threshold τ is exceeded. By varying the threshold for several realizations a receiver operation characteristics (ROC) curve is obtained, where the detection rate is plotted versus the frequency of false alarm.

We use a relatively weak signal of interest (in order to make detection non-trivial): $s^2 = 0.0035$, which is approximately 4.6 dB below the noise and about 27.6 dB below the interference. Monte-Carlo simulations were performed comprising $M = 5 \cdot 10^5$ realizations with the signal present and equally many with noise and interferences only. Figure 2 shows the ROC curves using the different robust beamformers. For this particular scenario, the BF [3] is only marginally better than the coin-flip detector (dash-dotted). Here all

the parameter-free RR methods outperform the RCB. Also the methods excluded from the plots for clarity of presentation gave higher detection probability than the RCB.

5. CONCLUDING REMARKS

We have presented a class of fully automatic (i.e., free from user parameters) robust adaptive beamformers based on a reparameterization of the standard Capon beamformer and on different ridge regression (RR) techniques. In this context we have also proposed an RR technique from an empirical Bayes perspective, that to the best of our knowledge is novel.

The proposed methods have been shown, by means of numerical simulations, to perform, in many cases, better than the robust Capon beamformer (RCB) of [4, 5, 6] even when knowledge of the array perturbation has been exploited in the RCB method. Especially ML, HKB and LW show good overall performance and should be considered attractive methods for the problem under study.

6. REFERENCES

- [1] J. Capon, "High resolution frequency-wavenumber spectrum analysis," in *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, August 1969.
- [2] H. L. Van Trees, *Optimum Array Processing*. No. IV in Detection, Estimation, and Modulation Theory, New York, USA: John Wiley & Sons, 2002.
- [3] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice Hall, 2005.
- [4] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Transactions on Signal Processing*, vol. 51, pp. 313–324, February 2003.
- [5] R. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1684–1696, May 2005.
- [6] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1702–1715, July 2003.
- [7] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Processing Letters*, vol. 10, pp. 172–175, June 2003.
- [8] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, February 1970.
- [9] A. E. Hoerl, R. W. Kennard, and K. F. Baldwin, "Ridge regression: some simulations," *Communications in Statistics: Theory and Methods*, vol. 4, pp. 105–123, 1975.
- [10] J. F. Lawless and P. Wang, "A simulation study of ridge and other regression estimators," *Communications in Statistics: Theory and Methods*, vol. A5, no. 4, pp. 307–323, 1976.
- [11] A. M. Urmanov, A. V. Gribok, H. Bozdogan, J. W. Hines, and R. E. Uhrig, "Information complexity-based regularization parameter selection for solution of ill conditioned inverse problems," *Inverse Problems*, vol. 18, pp. L1–L9, April 2002.
- [12] C. L. Mallows, "Some comments on C_P ," *Technometrics*, vol. 15, pp. 661–675, November 1973.
- [13] S. M. Kay, *Estimation theory*, vol. I of *Fundamentals of statistical signal processing*. Upper Saddle River, NJ: Prentice Hall, 1993.