Proceedings of the 4th International Symposium on Communications, Control and Signal Processing, ISCCSP 2010, Limassol, Cyprus, 3-5 March 2010

Optimal steering vector adaptation for linear filters leading to robust beamforming

Michal Natora, Felix Franke, Simon A. Broda, and Klaus Obermayer

***Abstract*—In many filter applications the exact steering vector is not know, and thus, robust beamforming methods have to be used. In this contribution, an algorithm which achieves robust beamforming via target tracking is proposed. In contrast to existing approaches, the algorithm works on sparse signals with arbitrary steering vector shapes, and the parameters of the algorithm are adapted in an optimal way. This is achieved by deriving and evaluating the probability of detection and false alarm for general steering vector mismatches. These probabilities are used to adjust the parameters, such that the number of false positive and false negative detections is minimal. Simulations confirm the theoretic results and show that the algorithm performs better than other approaches.**

***Index Terms*—STAP, array processing, target tracking, steering vector mismatch, ratio of random variables**

I. Introduction and problem statement

In many applications it is necessary to detect a specific signal within measurements of noisy data. Combined with thresholding the minimum variance distortionless response beamformer (MVDR), also called Capone beamformer, or the minimum power distortionless response beamformer (MPDR) are widely used as detectors for this task, due to their efficiency and low computational load [1].

Explicitly, the measured data $\boldsymbol{x}(t)$ is modeled as

$$\boldsymbol{x}(t) = \bar{\boldsymbol{q}} s(t) + \boldsymbol{n}(t), \tag{1}$$

where $\bar{\boldsymbol{q}}$ is the steering vector, s is the source signal and $\boldsymbol{n}$ is a noise vector sampled from a stationary, zero mean Gaussian distribution with covariance matrix $\boldsymbol{D}$, i.e $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$. For the sake of simplicity we assume that all quantities are real valued. The observation vector $\boldsymbol{x}(t)$ at time t, called snapshot, has dimension $N \times 1$. We are interested in filters in the form

$$\boldsymbol{f} = \frac{\boldsymbol{H}\bar{\boldsymbol{p}}}{\bar{\boldsymbol{p}}^\top \boldsymbol{H} \bar{\boldsymbol{p}}}, \tag{2}$$

and a signal is detected when the filter output exceeds a certain threshold γ, i.e. $\boldsymbol{f}^\top \boldsymbol{x}(t) \geq \gamma$.

The expression of Eq. 2 covers the MVDR, MPDR and general diagonal loading filters [2]. We refer to $\bar{\boldsymbol{q}}$ as the actual steering vector and to $\bar{\boldsymbol{p}}$ as the nominal steering vector. In order to maximize the detection performance it is desired that $\bar{\boldsymbol{p}} = \bar{\boldsymbol{q}}$, otherwise one has a, so called, steering vector mismatch. Since in general, however, the true steering vector is not known a priori, it has to be estimated from the data, which usually implies that $\bar{\boldsymbol{p}} \neq \bar{\boldsymbol{q}}$.

The field of robust beamforming emerged with the aim of designing filters which deliver acceptable detection performances even under steering vector mismatches [3]. Most of these approaches assume a model, either deterministic, or more recently, probabilistic, describing the steering vector mismatch and include this model into the optimization problem for the filter. This approach works well when the steering vector error is relatively small and stays constant over time. However, the actual steering vector might represent a target which is moving in space, as for example in the case of radar applications. Hence, it is $\bar{\boldsymbol{q}} = \bar{\boldsymbol{q}}(t)$ and, since the future trajectory of the target is not known, the mismatch between $\bar{\boldsymbol{p}}$ and $\bar{\boldsymbol{q}}$ can become arbitrarily large if $\bar{\boldsymbol{p}}$ is not adapted.

Besides many other fields in which linear filters are used, such as digital communications systems or speech enhancement, recently, they have been applied in biomedical engineering. In particular, they are used for processing electrophysiological recordings from electrodes, either as a spike detection [4], [5] or a spike sorting technique [6]. In this case, the spatial model in Eq. 1 is adapted to a temporal model $x(t) = \sum_\tau \bar{q}(\tau) s(t-\tau) + n(t)$, $\bar{\boldsymbol{q}}$ representing the waveform of the action potential, and $s(t)$ being the neuronal firing sequence. Due to tissue relaxation the distance between the electrode and the neuron is changing over time, which leads to an altered observed waveform of the action potentials [7]. The above mentioned spike detection and sorting methods, all of which are relying on filters shown in Eq. 2, will suffer from performance degradation since no robust beamforming methods were used.

To the knowledge of the authors there are only few methods which try to achieve robust beamforming by tracking, i.e. by adapting the nominal steering vector according to the changes of the actual steering vector. In [8] an adaptation scheme for the nominal steering vector based on the generalized sidelobe canceller algorithm is proposed. However, the noise is assumed to be white and an optimal adaptation rate is said to exist only in the case when the actual steering vector is not arbitrary but depends just on a single parameter, namely, the direction of arrival. This makes the algorithm unsuitable for applications where the steering vector cannot be represented by some simple underlying model.

A similar scheme is presented in [9]. In that study, the noise can be colored, but the adaptation rates of the filter and the nominal steering vector are fixed. An algorithm for

M. Natora and K. Obermayer are with the Institute for Software Engineering and Theoretical Computer Science, Berlin Institute of Technology, 10623 Berlin, Germany. Email: natora@cs.tu-berlin.de

F. Franke is with the Bernstein Center for Computational Neuroscience, 10115 Berlin, Germany.

S.A. Broda is with the Department of Quantitative Economics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands

978-1-4244-6287-2/10/$26.00

obtaining optimal adaptation, which would naturally depend on the system parameters, such as the target velocity, was not presented.

In [10] a very high signal-to-noise ratio is assumed, and therefore a noise model is not taken into account at all. Further, a specific physical model of the change of the actual steering is assumed, making the method unsuitable for more general changes.

In this contribution we consider measured data $\boldsymbol{x}$ in which the source signal s is present only at few specific times (sparse signal). This means that s is not a continuous process, but rather a sparse Bernoulli process. The former was assumed in [9] and implies that the steering vector is present in every measured snapshot. As a consequence of the sparseness, the nominal steering vector can only be adapted after a successful detection and not after every snapshot. To name a few examples, sparse signals are encountered in the aforementioned electrophysiological recordings or in geophysics [11].

In Sec. II-A the performance of a linear detector in case of steering vector mismatch is derived. This result is used in Sec. II-B to propose an optimal adaptation scheme of the nominal steering vector. This leads to robust beamforming via target tracking. Simulations in Sec. III show the effectiveness of this approach, and conclusive remarks are given in Sec. IV.

II. Method

A. Performance analysis under steering vector mismatch

In order to derive an *optimal* adaptation scheme, one has to understand how the performance of a detector depends on a steering vector mismatch. As performance measure we use the probability of detection (P_D) and false alarm (P_{FA}). These two probabilities translate directly into the number of true positive and the number of false positive detections, which is a meaningful performance measure for detectors of sparse signals. P_D is defined by

$$P_D := \text{Prob}[\boldsymbol{f}^\top \cdot \boldsymbol{q} \geq \gamma] = \text{Prob}\left[\frac{\boldsymbol{p}^\top \boldsymbol{H}^\top \cdot \boldsymbol{q}}{\boldsymbol{p}^\top \boldsymbol{H} \boldsymbol{p}} \geq \gamma\right]. \quad (3)$$

The random variable $\boldsymbol{q}$ is a noisy observation of the actual steering vector and, according to the model in Eq. 1, distributed as $\boldsymbol{q} \sim \mathcal{N}(\bar{\boldsymbol{q}}, \boldsymbol{D})$. Since the nominal steering vector will be based on an estimation from noisy data, $\boldsymbol{p}$ is also a random variable and is assumed to be distributed as $\boldsymbol{p} \sim \mathcal{N}(\bar{\boldsymbol{p}}, \boldsymbol{C})$ (in general $\boldsymbol{C} \neq \boldsymbol{D}$).

In the case of the MVDR beamformer, we have $\boldsymbol{H} = \boldsymbol{D}^{-1}$, whereas in the case of the MPDR beamformer, $\boldsymbol{H} = (\text{Cov}(\boldsymbol{x}))^{-1}$. Also for diagonal loading filters, $\boldsymbol{H}$ is in general positive semi-definite or positive definite, and symmetric. We suppose that $\boldsymbol{H}$ can be estimated on the basis of a large amount of samples. Therefore, it is justified to assume that the estimate of $\boldsymbol{H}$ is so accurate that there is no mismatch between the estimate and the true covariance, and its distribution can be neglected. Summarizing, the probability of detection is given by

$$P_D = 1 - \text{cdf}_\gamma\left(\frac{\boldsymbol{p}^\top \boldsymbol{H} \boldsymbol{q}}{\boldsymbol{p}^\top \boldsymbol{H} \boldsymbol{p}}\right), \quad (4)$$

where the notation $\text{cdf}_a(z) := \text{Prob}[z \leq a]$ is used. The probability of false alarm P_{FA} is obtained by setting $\bar{\boldsymbol{q}} = \boldsymbol{0}$ in Eq. 4. Hence, one has to obtain the cumulative distribution function of a ratio of quadratic forms of Gaussian random variables in order to analyze the performance of the desired detectors.

The expression in Eq. 4 can be simplified by defining $\boldsymbol{r} := \left(\boldsymbol{p}^\top, \boldsymbol{q}^\top\right)^\top$. Then, $\boldsymbol{r}$ is distributed as $\boldsymbol{r} \sim \mathcal{N}(\bar{\boldsymbol{r}}, \boldsymbol{G})$, where $\bar{\boldsymbol{r}} := \left(\bar{\boldsymbol{p}}^\top, \bar{\boldsymbol{q}}^\top\right)^\top$ and $\boldsymbol{G} := \begin{pmatrix} \boldsymbol{C} & \boldsymbol{E} \\ \boldsymbol{E} & \boldsymbol{D} \end{pmatrix}$, where $\boldsymbol{E}$ denotes the covariance between $\boldsymbol{p}$ and $\boldsymbol{q}$. This definition allows to rewrite the random variable in Eq. 4 as

$$\frac{\boldsymbol{p}^\top \boldsymbol{H} \boldsymbol{q}}{\boldsymbol{p}^\top \boldsymbol{H} \boldsymbol{p}} = \frac{\boldsymbol{r}^\top \begin{pmatrix} \boldsymbol{0} & \boldsymbol{H} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{r}}{\boldsymbol{r}^\top \begin{pmatrix} \boldsymbol{H} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{r}} = \frac{\tilde{\boldsymbol{r}}^\top \boldsymbol{A} \tilde{\boldsymbol{r}}}{\tilde{\boldsymbol{r}}^\top \boldsymbol{B} \tilde{\boldsymbol{r}}}, \quad (5)$$

where $\tilde{\boldsymbol{r}} \sim \mathcal{N}(\boldsymbol{G}^{-1/2} \cdot \bar{\boldsymbol{r}}, \boldsymbol{1})$, $\boldsymbol{A} := \boldsymbol{G}^{1/2^\top} \cdot \begin{pmatrix} \boldsymbol{0} & \boldsymbol{H} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \cdot \boldsymbol{G}^{1/2}$, and $\boldsymbol{B} := \boldsymbol{G}^{1/2^\top} \cdot \begin{pmatrix} \boldsymbol{H} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \cdot \boldsymbol{G}^{1/2}$, where $\boldsymbol{G}^{1/2}$ denotes the unique non-negative square root of $\boldsymbol{G}$.

The distribution of the ratio on the right hand side of Eq. 5 has been analyzed for a long time, since it is of importance in econometrics and statistics [12]. The analysis is mainly focused on cases where $\boldsymbol{A}$ is symmetric and $\boldsymbol{B}$ is positive semi-definite or positive definite. In the definition above, $\boldsymbol{A}$ is not symmetric. However, we can replace $\boldsymbol{A}$ by the symmetrized version of it $\tilde{\boldsymbol{A}} := 1/2 \cdot \left(\boldsymbol{A} + \boldsymbol{A}^\top\right)$, since it is $\boldsymbol{z}^\top \boldsymbol{A} \boldsymbol{z} = \boldsymbol{z}^\top \boldsymbol{A}^\top \boldsymbol{z}$ for any square matrix $\boldsymbol{A}$ and any vector $\boldsymbol{z}$. Since $\boldsymbol{G}$ is positive definite (it is a covariance matrix) it follows that $\boldsymbol{B}$ is positive semi-definite if $\boldsymbol{H}$ is positive semi-definite. In the case of the beamformers considered herein, $\boldsymbol{H}$ is indeed positive semi-definite or positive definite as discussed above. Finally, it is

$$P_D = 1 - \text{cdf}_\gamma\left(\frac{\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{A}} \tilde{\boldsymbol{r}}}{\tilde{\boldsymbol{r}}^\top \boldsymbol{B} \tilde{\boldsymbol{r}}}\right) = 1 - \text{cdf}_0\left(\tilde{\boldsymbol{r}}^\top \left(\tilde{\boldsymbol{A}} - \gamma \boldsymbol{B}\right) \tilde{\boldsymbol{r}}\right). \quad (6)$$

There exists a closed form expression for the cdf in Eq. 6, however, it involves an infinite series of top order polynomials with a very slow convergence rate [12]. Techniques for fast evaluation by means of saddlepoint approximations were developed in [13], [14]. These approximations, which are based on asymptotic expansions of the corresponding inversion integrals, however, are not accurate enough for the problem at hand.

Instead, the basis for our algorithm is a result from [15], where it was shown that the inversion integral can be reduced to integrating a real function over an infinite range, namely

$$\text{cdf}_0\left(\tilde{\boldsymbol{r}}^\top \left(\tilde{\boldsymbol{A}} - \gamma \boldsymbol{B}\right) \tilde{\boldsymbol{r}}\right) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin(\beta(u))}{u \cdot \rho(u)} \mathrm{d}u, \quad (7)$$

where β and ρ mainly depend on the eigenvalues λ_i and eigenvectors $\mathbf{e}_i$ of $\tilde{\boldsymbol{A}} - \gamma \boldsymbol{B}$, i.e. $\beta(u) := \frac{1}{2} \sum_i \arctan(a_i) + \frac{\bar{\theta}_i a_i}{c_i}$,

$\rho(u) := \exp\left\{\frac{1}{2}\sum_i \frac{\theta_i b_i}{c_i} + \frac{1}{4}\ln(c_i)\right\}$, $a_i := \lambda_i u$, $b_i := {a_i}^2$, $c_i := 1 + b_i$, and $\theta_i := (\boldsymbol{e}_i^\top \boldsymbol{G}^{-1/2}\bar{\boldsymbol{r}})^2$.

We evaluate this integral by mapping it onto the finite range $[0, 1]$ via the substitution $u = (1 - v)/v$ and replacing the integrand at $v = 0$ and $v = 1$ with its limits, which are zero and $1/2\sum_i(1 + \theta_i)\lambda_i$, respectively. The resulting integral can then be straightforwardly evaluated using standard numeric integration routines.

B. Adaptation scheme

To account for the time varying actual steering vector $\boldsymbol{q}(t)$, the nominal steering vector is adapted after constant time steps of length T. Notably, after every period T, the nominal steering vector is estimated as the sample mean of the K last detections $\boldsymbol{x}^i$ ($\boldsymbol{x}^i := \boldsymbol{x}(t(i))$ such that $\boldsymbol{f}^\top \cdot \boldsymbol{x}(t(i)) \geq \gamma$), i.e. $\bar{\boldsymbol{p}} = 1/K \cdot \sum_{i=K_{\max}-K+1}^{K_{\max}} \boldsymbol{x}^i$, where $K_{\max}$ denotes the total number of detections at time t.

The question arises, how many detections should be used for this estimate. If a large K is chosen, the estimate of the mean of $\boldsymbol{p}$ will be robust ($\boldsymbol{C}$ small), but the deviation from the true mean steering vector might be large ($\bar{\boldsymbol{p}} \neq \bar{\boldsymbol{q}}$). On the other hand, if a smaller K is chosen, then on average the estimated mean of the nominal steering vector will be closer to the mean of the actual steering vector ($\bar{\boldsymbol{p}} \approx \bar{\boldsymbol{q}}$), but the estimate will be more noisy ($\boldsymbol{C}$ large).

As an optimal trade-off, the value for K should be chosen such that the performance M of the detector is maximized. Usually, it is desirable that the number of total errors, which is the sum of false positive (FP) and false negative (FN) detections, is minimal. Therefore, we define our performance as

$$M := p_s \cdot P_D + p_n \cdot (1 - P_{FA}), \tag{8}$$

where p_s rsp. p_n is the probability that a randomly picked data snapshot contains a signal rsp. is a pure noise segment, and P_D rsp. P_{FA} are given by the expression in Eq. 6. Consequently, the optimal value for K is given by

$$K_{\text{opt}} = \underset{K}{\operatorname{argmax}} \{M(K)\}. \tag{9}$$

In order to use the expression of P_D in Eq. 6 for this adaptation scheme, the assumptions made in its derivation have to be verified. Firstly, it was assumed that $\boldsymbol{p}$ is Gauss distributed. This would be entirely correct only if all detections $\boldsymbol{x}^i$ were true positive, there were not any false negative detections, and the target was stationary ($\boldsymbol{q}(t) = \text{const.}$). For reasonable threshold values γ, most detections will indeed be true positive and only few signal occurrences will be missed, and if the change of the actual steering vector is not too rapid, then $\boldsymbol{p}$ will be approximatively Gauss distributed as $\boldsymbol{p} \sim \mathcal{N}(1/K \cdot \sum_{i=K_{\max}-K+1}^{K_{\max}} \boldsymbol{x}^i, \boldsymbol{C})$, where $\boldsymbol{C} = 1/K \cdot \boldsymbol{D}$. Secondly, the covariance matrix $\boldsymbol{H}$ was assumed to be known. The validity of this assumption depends on the amount of available data. In the case of continuous data recordings, as in biomedical recordings or wireless communications, a large amount of observed samples $\boldsymbol{x}$ are indeed available, and hence, $\boldsymbol{H}$ can be estimated very reliably.

The evaluation of P_D requires the knowledge of the true actual steering vector at all time intervals, i.e. $\bar{\boldsymbol{q}}(k \cdot T)$, where k is an integer. Generally, this information is not available and the actual steering vector has to be estimated from the data itself as well. For this estimation, again the sample mean of Q last detections is used, i.e. $\bar{\boldsymbol{q}} = 1/Q \cdot \sum_{i=K_{\max}-Q+1}^{K_{\max}} \boldsymbol{x}^i$. Assuming a linear change of the shape of the actual steering vector [1] this will give an estimate of $\bar{\boldsymbol{q}}$ for a time t_e, $t_e < kT$. A large value of Q will lead to a more robust estimate, but also $t_e \ll kT$, which means that there is a large "lag" in the adaptation. A small value of Q might give a more noisy estimate in the specific setting (but still a correct estimate *on average*), as an advantage, however, the adaptation follows the change of the actual steering vector more rapidly. Despite this estimation, the covariance of the true actual steering vector is still given by $\boldsymbol{D}$ (and not $1/Q \cdot \boldsymbol{D}$), hence, $\boldsymbol{q} \sim \mathcal{N}(1/Q \cdot \sum_{i=K_{\max}-Q+1}^{K_{\max}} \boldsymbol{x}^i, \boldsymbol{D})$.

We assumed complete blindness about the temporal structure of $\bar{\boldsymbol{q}}(t)$. Therefore, an optimal value of Q cannot be provided. In Sec. III-B, however, we will show that the adaptation scheme works reliably for a wide range of Q values, thus the algorithm exhibits a robust behavior with respect to this parameter.

III. Results

A. Comparison of cdf evaluation techniques

In this section we briefly show the accuracy of our cdf evaluation technique proposed at the end of Sec. II-A by presenting two examples. The expression in Eq. 7 was evaluated via MATLAB® using the standard commands `eig` for finding eigenvalues and eigenvectors, and `quadl` for numerical integration based on an adaptive Lobatto rule. The integration error was set to 10^{-8}. For comparison, both examples were also evaluated by an first and second order saddlepoint approximation (see [14] and [16] respectively).

The first example is according to [12]. Namely, in the special case when $\tilde{\boldsymbol{A}} = \boldsymbol{l} \cdot \boldsymbol{l}^\top/(\boldsymbol{l}^\top \cdot \boldsymbol{l})$, where $\boldsymbol{l}$ is a N-dimensional vector of ones, $\boldsymbol{B} = (\boldsymbol{1}_{N\times N} - \boldsymbol{A})/(N-1)$, and $\tilde{\boldsymbol{r}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{1}_{N\times N})$, then, the cdf in Eq. 6 is given by the cdf of the F-distribution $F(1, N-1)$.

As an example, the value $N = 10$ was chosen, and the threshold γ was varied in steps of 0.05 in the interval $[0, 11]$. The exact value of the cdf was assumed to be given by the MATLAB function `fcdf(γ,1,N-1)`. The three techniques were compared by means of the maximum absolute error between the exact cdf value and the value given by the corresponding technique for all thresholds. The results are shown in Table I.

In the second example, we considered $\tilde{\boldsymbol{r}} = \left(\left(\boldsymbol{C_1}^{-1/2}\boldsymbol{r_1}\right)^\top, \boldsymbol{r_2}^\top\right)^\top$, where $\boldsymbol{r_1}$ is an arbitrary N_1-dimensional vector, $\boldsymbol{r_2}$ is the N_2-dimensional zero vector, and

[1] This is true for small time intervals considering a first order Taylor approximation.

TABLE I
MAXIMUM ABSOLUTE ERROR FOR DIFFERENT CDF EVALUATION TECHNIQUES IN CASE OF AN F-DISTRIBUTION. SDPA1 AND SDPA2 DENOTE THE FIRST AND SECOND ORDER SADDLEPOINT APPROXIMATION TECHNIQUE

Our method	SdpA1	SdpA2
$1.6 \cdot 10^{-8}$	0.0121	0.0089

$\boldsymbol{C_1}$ is a diagonal matrix, $\boldsymbol{A} = \begin{pmatrix} N_2 \cdot \mathbf{1}_{N_1 \times N_1} & \mathbf{0}_{N_1 \times N_2} \\ \mathbf{0}_{N_2 \times N_1} & \mathbf{0}_{N_2 \times N_2} \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} \mathbf{0}_{N_1 \times N_1} & \mathbf{0}_{N_1 \times N_2} \\ \mathbf{0}_{N_2 \times N_1} & N_1 \cdot \mathbf{1}_{N_2 \times N_2} \end{pmatrix}$. In this case, the cdf in Eq. 6 is given by the cdf of the noncentral F-distribution $ncF(N_1, N_2, \lambda)$ with $\lambda = \sum_{i=1}^{N_1} (\boldsymbol{r_1}_i)^2 / \boldsymbol{C}_{i,i}$, which can be evaluated very accurately for example with the MATLAB function `ncfcdf(`γ`,` N_1`,` N_2`,` λ`)`.

We chose $\boldsymbol{r_1} = (1, 2, ..., 6)^\top$, $N_2 = 8$, $\boldsymbol{C} = \text{diag}(7, 6, ..., 2)$, and the threshold γ was varied in the interval $[1, 50]$ in steps of 0.05. The maximum absolute error to the exact value over all thresholds is reported in Table II.

TABLE II
MAXIMUM ABSOLUTE ERROR FOR DIFFERENT CDF EVALUATION TECHNIQUES IN CASE OF AN NONCENTRAL F-DISTRIBUTION. THE SAME NOTATION AS IN TABLE I IS USED

Our method	SdpA1	SdpA2
$2.5 \cdot 10^{-6}$	0.0785	0.0785

Clearly, in both examples our method is more accurate than both of the saddlepoint approximations.

B. Simulations

The proposed adaptation scheme was tested on data generated by Monte Carlo simulations. A single dataset consisted of 1400 snapshots, half of them containing a signal. White noise with a variance of 0.25, i.e. $\boldsymbol{D} = 0.25 \cdot \mathbf{1}_{N \times N}$, was used for the noise source. In the first 400 and last 400 snapshots a constant actual steering vector $\bar{\boldsymbol{q}}_1$ rsp. $\bar{\boldsymbol{q}}_2$ was simulated having the dimension $N = 7$. The signal-to-noise ratio (SNR) of both steering vectors was identical (6.0 db), however, they were orthogonal to each other. In between, a normalized linear mixture was simulated. To sum up

$$\bar{\boldsymbol{q}}(t) = \begin{cases} \bar{\boldsymbol{q}}_1, \forall t \leq 400 \\ \alpha(t) \cdot \bar{\boldsymbol{q}}_3(t), \forall t \in [400, 1000] \\ \bar{\boldsymbol{q}}_2, \forall t \geq 1000 \end{cases} \tag{10}$$

where $\bar{\boldsymbol{q}}_3(t) := \frac{\bar{\boldsymbol{q}}_2 - \bar{\boldsymbol{q}}_1}{600} \cdot t + \frac{1000 \cdot \bar{\boldsymbol{q}}_1 - 400 \cdot \bar{\boldsymbol{q}}_2}{600}$. In this setting every snapshot corresponds to one time unit. The value of $\alpha(t)$ was set such that $\text{SNR}(\alpha(t) \cdot \bar{\boldsymbol{q}}_3) = 6.0$ db $\forall t$. This guaranteed that any performance loss of the detector was caused by the change in shape of the actual steering vector, and not due to a simple decrease in SNR.

The first 400 snapshots served as initialization, and were used to estimate the initial nominal steering vector. The adaptation scheme, hence, was applied on the snapshots $401 - 1400$. For this, the performance given by Eq. 8 was calculated for all K values (see Fig. 1), whereas the values for p_s and p_n were assumed to be known [2]. The expression in Eq. 7 was evaluated in the same way as described in Sec. III-A. Finally,

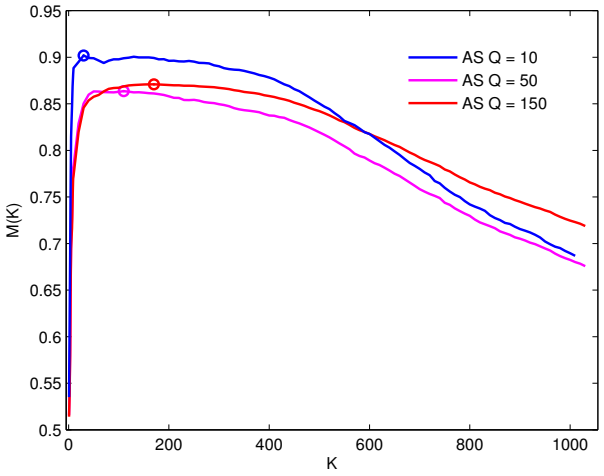

Fig. 1. Exemplarily, the performance M (see Eq. 8) as a function of K ($K = 1, 2, ..., 9, 10, 20, 30, ..., K_{\text{max}}$, where K_{max} was the number of total detections at the corresponding time) is shown for $t = 700$ and for a threshold of $\gamma = 0.525$. The maximal value of M is marked with a circle, which determined K_{opt}. The different curves correspond to adaptation schemes (AS) with different Q values.

the K value for which M was the largest was chosen as K_{opt} (corresponds to Eq. 9), and the filter was re-calculated (using Eq. 2) based on the adapted nominal steering vector.

The MVDR beamformer was used as a filter, which implied $\boldsymbol{H} = \boldsymbol{D}^{-1}$. The steering vector was adapted after every 10th snapshot, i.e. $T = 10$, see Fig. 2. Three different values for Q were tested, namely $Q = \{10, 50, 150\}$. For simplicity, the covariance between $\bar{\boldsymbol{p}}$ and $\bar{\boldsymbol{q}}$ was ignored, i.e. $\boldsymbol{E} = \mathbf{0}$ in the calculation of $\boldsymbol{G}$ (see Sec. II-A). In total, 10 datasets were simulated, over which the results were averaged. The proposed adaptation scheme was compared to approaches in which a fixed, pre-defined number of detections was used for the computation of the nominal steering vector. The results are shown in Fig. 3.

C. Evaluation and discussion

From Fig. 1 one can observe that $M(k)$ exhibits a kind of plateau region, on which the optimal maximal value is located, meaning that the performance does not change a lot when K is varied. This results in large variations of K_{opt} across the simulations (visible in Fig. 2), since the maximum might occur at different values of K due to fluctuations. Although the standard deviation of K_{opt} is large, this has small influence on the total performance (i.e. small standard deviations in Fig. 3), which is again consistent with the theoretic prediction in Fig. 1. Fig. 2 also demonstrates the tradeoff between

[2] Since a setting in which the target is stationary at the beginning (snapshots $1 - 400$) was used, it can be assumed that p_s and p_n can be estimated reliably, or there is a priori knowledge about them.

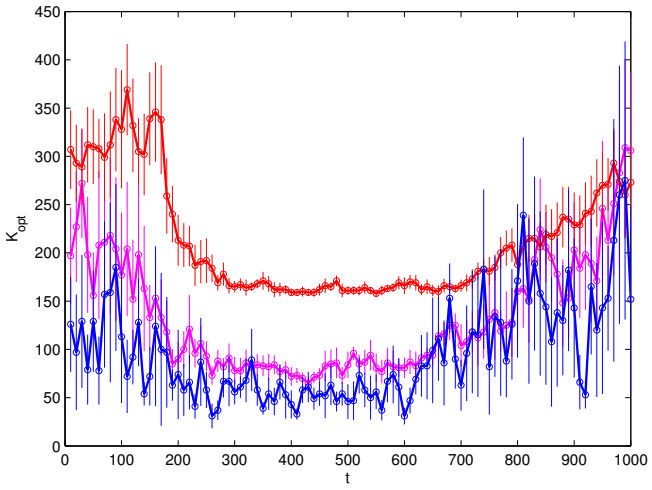

Fig. 2. Average value of K_{opt} as a function of time. The same threshold and color coding as in Fig. 1 was used. The vertical lines indicate the standard deviation across the 10 simulations.

minimization of the mismatch between the means of the steering vectors and the minimization of the variance of the nominal steering vector estimation. Notably, when the steering vector is close to stationary (at the beginning and the end of the simulations), the variance is minimized by taking a large number of detections for estimation, i.e. $K \gg Q$. On the other hand, when the actual steering vector varies rapidly, it is better to use a number closer to the number of detections used for the estimation of the actual steering vector, i.e. $K \approx Q$, in order to reduce the mismatch between the means of the steering vectors.

Fig. 3 illustrates that despite a rather rapid variation and extreme thresholds, both of which lead to a violation of the assumptions under which the adaptation scheme was derived, the proposed algorithm performs close to the theoretically optimal approach for a wide range of thresholds.

A non adaptive scheme, which assumes that the steering vector does change significantly, would use all detections for the steering vector estimation. Even if a change in the steering vector is detected, a fixed number of detections for estimation would be used in such a non-optimal adaptation scheme. Both approaches perform worse than the derived optimal adaptation scheme, see Table III.

The simulations also revealed that the proposed adaptation scheme works well for a wide range of Q values. In particular for the case when the total error is minimized, i.e. the threshold is around 0.5, the performance of the three considered cases is very similar (Fig. 3).

TABLE III
AVERAGE AREA UNDER THE CURVE (AUC) IN DESCENDING ORDER FOR VARIOUS ADAPTATION SCHEMES. THE AUC WAS CALCULATED BASED ON THE RESULTS IN FIG. 3 WITH RANGE OF FP $\in [0.002, 0.73]$

Opt.	$Q = 50$	$Q = 150$	$Q = 10$	$K = 400$	$K = K_{\max}$
0.6471	0.6291	0.6234	0.6142	0.6016	0.5169

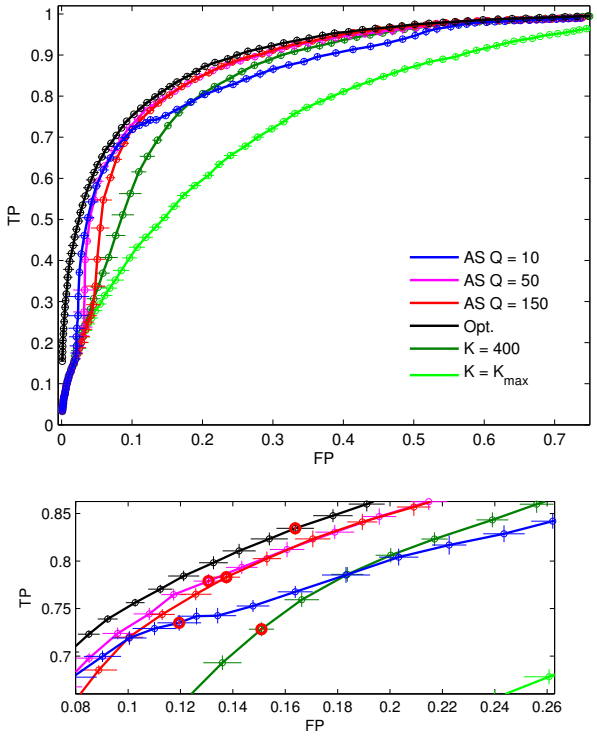

Fig. 3. *Top:* Receiver operating characteristics: For every adaptation scheme γ was varied from -0.5 up to 1.5 in steps of 0.025, and the resulting average relative number of false positive (FP) and true positive (TP) were plotted together with their standard deviations. The black curve corresponds to the theoretically best possible adaptation scheme, in which the filter was calculated based on the true steering vector, i.e. $\bar{\boldsymbol{p}} = \bar{\boldsymbol{q}}(t)$, $t = k \cdot T$. The green lines resulted, when a non-optimal number of detections was used for adaptation, i.e. $K = 400$ (dark green) rsp. $K = K_{\max}$ (light green). The color coding of the other lines is the same as in Fig. 1. The corresponding areas under the curves are listed in Table III. *Bottom:* Magnification of a part of the top figure. In the herein considered case when exactly half of the snapshots contain the signal, the optimal threshold which minimizes the total error is $\gamma = 0.5$. The corresponding points on the ROC curves are marked with a red circle.

IV. CONCLUSION

The performance with respect to P_D and P_{FA} of the MVDR in the presence of steering vector mismatches has been also studied in [17], [18]. Therein, it was assumed that the nominal steering vector is deterministic and that there is a large error in the noise covariance matrix estimation. However, in the case of continuous recordings containing time-varying sparse signals, the mentioned assumptions do not hold.

In [19] a probabilistic nominal steering was assumed, but only the average signal-to-interference-plus-noise ratio (SINR) was analyzed. The average SINR provides less information than the probability of detection and false alarm. P_D and P_{FA} can be used to directly account for the number of true positive and true negative detections, as well as to evaluate more complex performance measures as in [20].

In this study P_D and P_{FA} were derived in the case when both actual and nominal steering vector are random. It was shown that these probabilities are equivalent to quantities

studied by the econometrics and mathematical statistics communities. A numerical method for their efficient evaluation was proposed, implemented and compared to existing techniques based on saddlepoint approximations.

Furthermore, this analysis allowed to propose an optimal adaptation scheme for the nominal steering vector. In this sense, an algorithm for robust beamforming via target tracking was proposed. In our algorithm the shape of the steering vector as well as its temporal evolution can be arbitrary. This makes the algorithm also suitable for applications beyond radar and antenna systems; amongst others, it can be used for digital communication systems or in biomedical signal processing.

To the knowledge of the authors the analysis of the beamformer performance when both steering vectors as well as the estimate of the covariance matrix deviate from the true quantities has not been carried out yet. This problem is left for further research.

ACKNOWLEDGMENT

This research was supported by the Federal Ministry of Education and Research (BMBF) with the grants 01GQ0743 and 01GQ0410. The authors would like to thank Małgorzata M. Wójcik for help with proof-reading.

REFERENCES

[1] H. L. V. Trees, *Detection, Estimation, and Modulation Theory Part IV - Optimum Array Processing*. JOHN WILEY & SONS, 2002.

[2] L. Du, T. Yardibi, J. Li, and P. Stoica, "Review of user parameter-free robust adaptive beamforming algorithms," *Digital Signal Processing*, vol. 19, pp. 567–582, 2009.

[3] A. B. Gershman, "Robust adaptive beamforming: an overview of recent trends and advances in the field," in *Proc. 4th International Conference on Antenna Theory and Techniques*, vol. 1, 9–12 Sept. 2003, pp. 30–35.

[4] S. Kim and J. McNames, "Automatic spike detection based on adaptive template matching for extracellular neural recordings." *J Neurosci Methods*, vol. 165, pp. 165–174, 2007.

[5] P. H. Thakur, H. Lu, S. S. Hsiao, and K. O. Johnson, "Automated optimal detection and classification of neural action potentials in extra-cellular recordings." *J Neurosci Methods*, vol. 162, pp. 364–376, 2007.

[6] M. Natora, F. Franke, M. Munk, and K. Obermayer, "Blind source separation of sparse overcomplete mixtures and application to neural recordings," *Independent Component Analysis and Signal Separation*, vol. 5441, pp. 459–466, 2009.

[7] E. Branchaud, J. Burdick, and R. Andersen, "An algorithm for autonomous isolation of neurons in extracellular recordings," in *BioRob 2006*, 2006, pp. 939–945.

[8] S. Affes, S. Gazor, and Y. Grenier, "Robust adaptive beamforming via lms-like target tracking," in *ICASSP-94*, vol. 4, 1994, pp. 269–272.

[9] S. Gazor, S. Affes, and Y. Grenier, "Robust adaptive beamforming via target tracking," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1589–1593, June 1996.

[10] S. Timofeev, A. Bahai, and P. Varaiya, "Adaptive acoustic beamformer with source tracking capabilities," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2812–2820, 2008.

[11] F. Champagnat, Y. Goussard, and J. Idier, "Unsupervised deconvolution of sparse spike trains using stochastic approximation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 2988–2998, 1996.

[12] G. Forchini, "The distribution of a ratio of quadratic forms in noncentral normal variables," *Communications in statistics. Theory and methods*, vol. 34, p. 999 - 1008, 2005.

[13] P. Marsh, "Saddlepoint approximations for noncentral quadratic forms," *Economic Theory*, vol. 14, pp. 539 – 559, 1998.

[14] R. Butler and M. Paolella, "Uniform saddlepoint approximations for ratios of quadratic forms," *Bernoulli*, vol. 14, pp. 140 – 154, 2008.

[15] J. Imhof, "Computing the distribution of quadratic forms in normal variables," *Biometrika*, vol. 48, pp. 419 – 426, 1961.

[16] R. Butler and M. Paolella, "Uniform saddlepoint approximations for ratios of quadratic froms," Department of Statistical Science, Southern Methodist University, Tech. Rep., 2007.

[17] S. Z. Kalson, "An adaptive array detector with mismatched signal rejection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 195–207, Jan. 1992.

[18] R. S. Blum and K. F. McDonald, "Analysis of stap algorithms for cases with mismatched steering and clutter statistics," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 301–310, Feb. 2000.

[19] M. Wax and Y. Anu, "Performance analysis of the minimum variance beamformer in the presence of steering vector errors," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 938–947, 1996.

[20] M. Natora, F. Franke, and K. Obermayer, "Optimal convolutive filters for real-time detection and arrival time estimation of transient signals," *Proceedings of World Academy of Science, Engineering and Technology*, vol. 55, pp. 235–240, 2009.