FOI-R--1297--SE
May 2004
ISSN 1650-1942
Scientific report

Per Hyberg

Array Pre-processing for the SESAM System:

DOA Mean Square Minimization

Command and Control Systems
SE-581 11 Linköping

SWEDISH DEFENCE RESEARCH AGENCY

Command and Control Systems
P.O. Box 1165
SE-581 11 Linköping

FOI-R--1297--SE

May 2004

ISSN 1650-1942

Scientific report

Per Hyberg

Array Pre-processing for the SESAM System: DOA Mean Square Minimization

Issuing organization	Report number, ISRN	Report type
FOI – Swedish Defence Research Agency Command and Control Systems P.O. Box 1165 SE-581 11 Linköping	FOI-R--1297--SE	Scientific report
	Research area code 6. Electronic Warfare and deceptive measures	
	Month year May 2004	**Project no.** E 7059
	Customers code 5. Commissioned Research	
	Sub area code 61 Electronic Warfare including Electromagnetic Weapons and Protection	

Author/s (editor/s)	Project manager
Per Hyberg	Rolf Gustavsson
	Approved by Martin Rantzer
	Sponsoring agency Swedish Armed Forces
	Scientifically and technically responsible Per Hyberg

Report title

Array Pre-processing for the SESAM System: DOA Mean Square Minimization

Abstract (not more than 200 words)

This report treats interpolation (mapping) of the output vector from an existing antenna array onto the output vector of another (virtual) array when the signal directions of arrival (DOAs) are known only to within a wide sector. In an earlier companion paper a first order condition for zero DOA bias under such mapping was derived, and also used to construct a design algorithm for the mapping matrix that minimized the DOA estimate bias. This bias minimizing theory is now extended to minimize not only bias, but DOA mean square error (MSE), i.e. bias squared plus variance.

We first derive an analytical expression for mapped DOA MSE. Thereafter we propose a design algorithm for the transformation matrix that generates mapping errors minimizing this MSE. Generally, DOA MSE is not reduced by minimizing the size of the mapping errors, but instead by rotating these errors and the associated noise subspace into optimal directions relative to a certain gradient of the DOA estimator cost function.

The analytical MSE expression and the design algorithm are supported by simulations that show not only conspicuous MSE improvements in relevant scenarios, but also a resulting pre-processing of better robustness for low SNRs as compared to the pure bias minimizing design described in the earlier paper.

Keywords

Array mapping, array pre-processing,array interpolation, DOA MSE minimization

Further bibliographic information	Language English
ISSN 1650-1942	**Pages** 25 p.
	Price acc. to pricelist

Utgivare Totalförsvarets Forskningsinstitut - FOI Ledningssystem Box 1165 581 11 Linköping	**Rapportnummer, ISRN** FOI-R--1297--SE	**Klassificering** Vetenskaplig rapport
	Forskningsområde 6. Telekrig och vilseledning	
	Månad, år Maj 2004	**Projektnummer** E 7059
	Verksamhetsgren 5. Uppdragsfinansierad verksamhet	
	Delområde 61 Telekrigföring med EM-vapen och skydd	
Författare/redaktör Per Hyberg	**Projektledare** Rolf Gustavsson	
	Godkänd av Martin Rantzer	
	Uppdragsgivare/kundbeteckning FM	
	Tekniskt och/eller vetenskapligt ansvarig Per Hyberg	

Rapportens titel (i översättning)

Array-signalbehandling i SESAM-systemet: DOA MSE minimering

Sammanfattning (högst 200 ord)

Denna rapport behandlar array-mappning (array-interpolation) för SESAM-systemet. Signalernas DOA är okänd varför mappningen (avbildningen) från den verkliga till den virtuella array:en måste ske över en relativt bred sektor, typiskt 30^0. I rapporten utvecklas en teori enligt vilken denna avbildning kan ske med minimerade fel. Såväl deterministiska som slumpmässiga fel behandlas och den storhet som minimeras är MSE (minsta kvadratfelet = bias^2+varians).

Först härleds ett analytiskt uttryck för MSE-felet vid array-mappning. Därefter presenteras en algoritm med vars hjälp en transformationsmatris (mappningsmatris) kan konstrueras som ger minimerade MSE-fel i bäringsestimaten. Principen för minimeringen är att vrida hela observationsrummet till en optimal orientering relativt gradienten för kostnadsfunktionen för den DOA-estimator som används. Principen är generell men i de redovisade simuleringarna har mappning skett från en cirkulär array med 8 element (SESAM-systemet) till en linjär array (ULA). Därefter har root-MUSIC –algoritmen använts.

Simuleringarna bekräftar såväl de härledda uttrycken för MSE-felen som fel-minimeringsförmågan hos de transformationsmatriser som beräknats med hjälp av den härledda algoritmen.

Nyckelord

Array-mappning, signalbehandling, pejling, felminimering

Övriga bibliografiska uppgifter	**Språk** Engelska
ISSN 1650-1942	**Antal sidor:** 25 s.
Distribution enligt missiv	**Pris: Enligt prislista**

CONTENTS

Array Pre-processing for the SESAM System: DOA Mean Square Error Minimization

Per Hyberg, per.hyberg@foi.se

I. Background

PRE-PROCESSING Mapping of the output data from an illuminated antenna array before this data is fed to a DOA estimator can be used to gain different advantages. One such advantage treated here, is the possibility to optimize the array geometry and DOA estimator individually. Using a circular array (for omni-directionality) together with root-MUSIC or other fast DOA estimators that presume linear arrays, is one such application, attractive for signal surveillance. This is also the underlying application behind the presented work.

A relevant problem in this context is under what circumstances the pre-processing retains accuracy in the DOA estimates, since both a deterministic (bias) and random (variance) error increase can be caused by the pre-processing as such.

The present paper addresses an application (wide band signal surveillance) where these error types are of comparable size and neither can be neglected. Conditions for minimizing the variance under mapping has been studied by many authors, [1], [2], and under the related application of dimension reduction by [3]. Bias minimizing conditions were given in a.o. [4], but here the *combined* error, the Mean Square Error MSE (defined as $bias^2 + variance$), is analyzed. Furthermore an analytical expression for this MSE is derived and also a special design algorithm for the pre-processing matrix that minimizes the MSE of the DOA estimates.

A. Interpolation

When an array with unknown errors is calibrated at a grid of calibration directions, interpolation techniques can be used to generate the response vectors from emitters between the calibrated directions. This is done by calculating a correction matrix that matches the real array to a mathematical model of the error-free array in the calibrated directions. After multiplication with this correction matrix the data is treated as error-free.

First proposed in 1988, [5], [6], [1], this technique can also be viewed as *mapping* between two similar array manifold sets, the *real* one and the *virtual* (error-free) one. Due to the array similarity that was presumed in these early studies, such a data transformation could be performed with very small added new errors, despite the directions of arrival, the DOAs, being only coarsely known.

B. Sector mapping

Signal surveillance often needs to combine omni-directionality and fast signal processing, the latter to catch sub-second burst transmitters. Hence there is a need to combine a circular array (for omni-directionality) with ULA based root estimators (for speed). One attractive solution to this problem is to map (interpolate) the circular array manifold onto that of an imaginary uniform linear manifold, usually with equally many elements, see Figure 1.

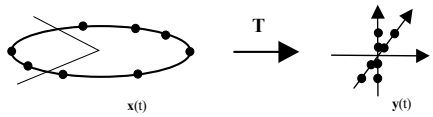

Fig. 1. Array mapping over sectors makes it possible to use fast rooting DOA estimators with arbitrary array shapes. Here the output vectors from a uniform circular (real) array UCA, are mapped (interpolated) onto the output vectors of a (virtual) ULA using the transformation matrix T. However, if the sector is wide, errors will occure, both increased bias and increased variance, unless special measures are taken. This is a typical situation where MSE minimization becomes important.

One appreciated feature of this mapping approach is its generality: For example we can equally well map upon two equal but displaced ULAs and use the ESPRIT or correspondingly any other suitable DOA estimator. The original real array can also be mapped on a displaced copy of itself [7].

Since in the signal surveillance application the DOAs initially are unknown, we have to process wide azimuthal sectors and design one mapping matrix for each sector. If the frequency band in addition is wide, such as in signal reconnaissance, we may also use one mapping matrix per frequency subband.

In this pre-processing several difficulties occur. First, for obvious geometrical reasons we cannot perfectly match a circular array to a linear one. Mapping errors will occur since the mapping matrix can at the best be a compromise over the mapped sector. The wider the sector the larger the errors.

The *deterministic part* of these errors are known, at least in the calibrated directions, a fact that can be used to correct them [4], [8], [9], [10] The *random part*, caused by the transformed noise, can result in added DOA estimate variance [2], [3] as well as bias, [11].

Conditions for minimizing variance and attaining the Cramér-Rao Bound (CRB) were derived by Andersson [3] but presume the DOAs to be known, at least to within a beamwidth. Hence when the emitter directions are completely unknown extra DOA error variance will be generated by the mapping. This motivates the development of techniques to reduce the extra variance as much as possible while at the same time keeping bias small.

Describing the mechanisms behind both these error types, the bias and the variance, (together expressed as Mean Square

Error, MSE), and finding ways to minimize them, is the aim of the present paper.

C. Dimension reduction

Processing speed can be enhanced if the DOA estimator is designed for fewer antenna elements. A large, possibly non-uniform, linear array (that provides good resolution) can then be mapped onto a ULA with fewer elements, to achieve this. To retain variance, this *dimension reduction* presumes that the DOAs are known, at least to within a beamwidth. The requirements on the dimension-reducing mapping matrix to retain variance was derived in [3], see equation (34).

We also note that dimension reduction helps to restore numerical robustness in the solution for the mapping matrix, since the number of columns in the mapping matrix is reduced. Dimension reduction may be necessary if the used array beamwidth is equal to, or larger than, the mapped sector, since in such cases linear dependence between the response vectors inside the sector may become a problem. This is further commented on in the Simulations section.

D. Dimension increase

It is in principal possible to map the real array onto a virtual array with a larger number of antenna elements, i.e. *dimension increase*. This is equivalent to using more base functions when matching the two array manifolds, a measure that should reduce the errors somewhat. However, since the number of required calculations increases as m_v^3,where m_v is the number of elements in the virtual array, this approach will not be considered in what follows.

II. Problem formulation

This report will address the general problem of DOA estimate accuracy under pre-processing. One such application is *array mapping (interpolation) over sectors*, a problem where the pre-processing as such can cause both *bias* and *added variance*. Our problem is to design a pre-processing matrix common to the sector, that minimizes both.

Although the underlying application is mapping from a circular array onto a uniform linear array, the argument and derivations will be of a general nature. The ULA and the associated ULA DOA estimator will be regarded as one entity, separated from the mapping operation. Hence all estimator cost functions, as well as derivatives and gradients thereof, will refer to the ULA and its ULA based DOA estimator, and not be explicitly parameterized by any pre-processing.

The abovementioned mapping errors will be regarded as errors in the (virtual) field that impinges on the ULA and parameterized as corresponding errors $\Delta\mathbf{e}_v$ in the eigenvectors $\mathbf{e}_v$ of the virtual signal subspace of the range space of the ULA output covariance matrix.

This view will simplify the analysis. The problem of best mapping matrix design can now be formulated as a best transformation on the errors $\Delta\mathbf{e}_v$. We can either minimize them, or rotate them into directions where they cause the least DOA errors, or both.

As a result the derived method of preventing the mappings errors from affecting the DOA measurements, is applicable to a much wider class of errors than those caused by imperfect mapping. In fact, the derived method is capable of making *any* signal subspace error almost invisible in the DOA estimates. At least as long as said errors are of limited size.

A. Error types

DOA estimates $\hat{\theta}$ often have errors $\Delta\theta = \hat{\theta} - \theta$ consisting of both deterministic and random parts: These errors manifest themselves as DOA *bias* and *variance*. Hence, a practical entity to describe the total DOA error is the MSE (Mean Square Error) and we thus have $MSE = bias^2 + variance$. For subspace based DOA estimators these DOA errors can be seen as caused by corresponding eigenvector errors $\Delta\mathbf{e}_v$ in the (virtual) signal subspace of the data covariance matrix $\mathbf{R}$.

Consider the virtual array from which we get an output data vector with certain signal subspace errors $\Delta\mathbf{e}_v$. With superscript denoting the *cause*, and subscript *effect* of these errors, we can write

$$\Delta\mathbf{e}_v = \Delta\mathbf{e}_{v,(bias)}^{(det)} + \Delta\mathbf{e}_{v,(bias)}^{(noise)} + \Delta\mathbf{e}_{v,(var)}^{(noise)} \qquad (1)$$

The first superscript (det) refers to various model errors, imperfect array mapping, or other non-fluctuating deterministic error mechanisms. This type of error is independent of signal to noise ratio SNR and the number of snapshots N. Array mapping errors are of this type, see Figure 1. Their correction was treated in [4], [8], [9] and [10].

Random errors $(noise)$ can cause both DOA bias and variance, the second and third error types in (1) respectively.

DOA variance, the third error type, has been studied extensively and conditions for the retainment of the Cramér-Rao Bound, CRB, without and with pre-processing described, [12], [13], [3] and [2]. Furthermore, DOA bias due to noise, the second error type, was analyzed for MUSIC in [11], where it was shown that this type of bias requires a third order Taylor expansion of the estimator cost function to be described properly. However, no procedure for *removal* of this bias was suggested.

The purpose of the present report is to study array pre-processing,especially array mapping, with the property of minimizing the *total MSE*, not just bias or variance. The argument will be of general nature although the underlying application is fast signal surveillance against emitters in unknown directions. Special emphasis is therefore put on the case where the DOAs initially are known only to within a wide sector and any pre-processing has to be a compromise over this sector.

For a bias free estimator in white noise, no pre-processing can yield a DOA variance below the CRB. It is the unnecessary extra variance that is caused by the pre-processing under the above compromise that we should try to minimize. If, in the Figure 1 scenario, a single mapping matrix $\mathbf{T}$ is to work over a 30^o sector (a typical width) and the bandwidth of the circular array is stressed, both bias and extra variance can grossly exceed the CRB. This is a typical situation where MSE minimization techniques are called for.

B. Notational remarks and assumptions

Throughout this paper $(\cdot)$ will denote scalar product, $(^*)$Hermitian transpose, $(^T)$ matrix transpose, and $(^-)$ complex conjugate. Measurement noise is considered Gaussian, zero mean and circularly symmetric. Furthermore, for ease of notation we will mainly analyze scenarios that contain a single emitter at the horizon, i.e. the DOA is fully characterized by the scalar entity θ (azimuth). In the bias analysis and the subsequent construction of the mapping matrix $\mathbf{T}$, the impinging signal will come from a known calibration emitter, whereas during the evaluation simulations it will come from an unknown emitter (presumed to be within the mapped azimuthal sector).

The argument is not limited to the single emitter case however but the error reduction effectiveness of the suggested mapping will be best for such cases. This is because the mapping error $\Delta\mathbf{e}_v = \mathbf{T}^*\mathbf{e}_s - \mathbf{e}_v$, where $\mathbf{e}_s$ and $\mathbf{e}_v$ are the eigenvectors associated with the real and virtual arrays respectively, is more difficult to rotate (through the selection of the mapping matrix) into simultaneous orthogonality with more than one signal eigenvector gradients.

Since from a signal reconnaissance point of view the impinging wave fields are of a non-predictable nature, and in addition only direction of arrival (DOA) is of interest, we use a Gaussian distributed (single) signal model $\mathbf{s}(t)$ with covariance matrix $\mathbf{S}$ (which in the presumed one emitter case reduces to a scalar representing the signal power), and the usual properties of being both temporally and spatially white. Furthermore, consider a general planar real array of m_r isotropic antenna elements. The array output at time t is then modeled as an $m_r \times 1$ vector $\mathbf{x}(t_i)$, where $i = 1, ..., N$:

$$\mathbf{x}(t) = \mathbf{A}(\theta)\mathbf{s}(t) + \mathbf{n}(t) \tag{2}$$

The columns of the $m_r \times 1$ matrix $\mathbf{A}(\theta)$ transform from the impinging signal $\mathbf{s}(t)$ to the array output $\mathbf{x}(t)$, and $\mathbf{n}(t)$ is the noise contribution from the m_r receiver channels. This noise is presumed to be an ergodic spatially white stochastic process with the second order moment $E\left\{\mathbf{n}(\mathbf{t})\mathbf{n}^*(t)\right\} = \sigma^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

To distinguish the real array from the virtual, we will denote the former $\mathbf{A}_r$ and the latter $\mathbf{A}_v$. The symbol $\hat{\theta}$ will be used in the sequel for estimates of the DOA azimuth θ.

We can now model the array output covariance matrix $\mathbf{R}$ as

$$\mathbf{R} = \mathbf{A}(\theta)\mathbf{S}\mathbf{A}^*(\theta) + \sigma^2\mathbf{I} \tag{3}$$

The estimated signal- and noise subspaces $\hat{\mathbf{E}}_s$ and $\hat{\mathbf{E}}_n$ of the real array are formed by eigenvalue decomposition of an estimate of $\mathbf{R}$ obtained as

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}(t)\mathbf{x}^*(t) \tag{4}$$

where N is the number of snapshots.

Finally, this paper deals with the construction of mapping (interpolation-, pre-processing-) matrices. These matrices are calculated from a set of calibration response vectors $\left\{\theta^{(i)}\right\}$ that are assumed to be collected by moving a (single) calibration transmitter around the array at some suitable distance. Hence all the derivations in the sequel presume one emitter at a time.

The resulting mapping matrix is of course a linear operator that can be used in multi-emitter scenarios as well, as long as these are confined to the calibration sector. As illustrated in the simulations section, a somewhat reduced performance is expected in the multi emitter case however.

The additional problem of suppressing out of sector signals is only coarsely treated in this paper. The reader is referred to existing literature on this topic, i.e. [14], [15], [16].

III. The gradient of the cost function

For all cost function based DOA estimators bias corresponds to an off-set along the θ-axis in the extreme of the (scalar) cost function $V(\theta)$. We can equivalently study the off-set in the zero of the derivative with respect to θ, $\dot{V}(\theta)$, an approach that has some advantages and therefore will be used throughout this paper.

In the sequel the gradient $\nabla_{\mathbf{e}_v}\dot{V}(\theta)$ with respect to the signal eigenvector $\mathbf{e}_v$ of the virtual array (which is delivering data to the DOA estimator) will be important. Using Brandwood's conventions [17], the scalar product $2\,\mathrm{Re}\left\{\overline{\nabla_{\mathbf{e}_s}\dot{V}(\theta)}\cdot\Delta\mathbf{e}_v\right\}$ equals the change in the value of $\dot{V}(\theta)$ caused by the change $\Delta\mathbf{e}_v$ in $\mathbf{e}_v$. Hence, by dividing with the second derivative $\ddot{V}(\theta)$ we get the corresponding change in estimated DOA.

Below the gradient will be used to express and cast new light over both bias and variance in the DOA estimates.

We now observe that the above errors $\Delta\mathbf{e}_v$ in the virtual signal subspace eigenvectors can be divided into two classes, namely

1) Those with the scalar product

$$2\,\mathrm{Re}\left\{\overline{\nabla_{\mathbf{e}_s}\dot{V}(\theta)}\cdot\Delta\mathbf{e}_v\right\} = 0 \tag{5}$$

2) The remaining ones for which

$$2\,\mathrm{Re}\left\{\overline{\nabla_{\mathbf{e}_s}\dot{V}(\theta)}\cdot\Delta\mathbf{e}_v\right\} \neq 0 \tag{6}$$

Obviously errors that belong to the first class are harmless since they do not effect the value of $\dot{V}(\theta)$ and therefore do not cause bias or other DOA errors. We will come back to this in the next section.

The gradient $\nabla_{\mathbf{e}_v}\dot{V}(\theta)$ has another interesting feature connected with zero bias. This feature is formulated in the following theorem:

Theorem 1: With $\mathbf{a}(\theta)$ being the response vector, for an arbitrary array geometry in a single emitter scenario, if the estimator cost function consists of the quadratic form[1]

$$Q(\theta, \mathbf{e}_v) = \mathbf{a}^*(\theta)\mathbf{e}_v\mathbf{e}_v^*\mathbf{a}(\theta) \tag{7}$$

and is bias free at the correct DOA θ, then at this DOA

$$\overline{\nabla_{\mathbf{e}_v}\dot{V}(\theta_o)} \perp \mathbf{a}(\theta_o) \tag{8}$$

[1] This is the case for most subspace based estimators, especially WSF/MODE and MUSIC

Proof: The derivative w.r.t. θ of $V(\theta)$ is

$$\dot{V}(\theta) = 2\,\mathrm{Re}\left\{\mathbf{a}^*(\theta)\mathbf{e}_v\mathbf{e}_v^*\dot{\mathbf{a}}(\theta)\right\} \tag{9}$$

The gradient of $\dot{V}(\theta)$ w.r.t. the signal eigenvector $\mathbf{e}_v$ becomes (Brandwood, see (52) in the Appendix)

$$\mathbf{g} \triangleq \nabla_{\mathbf{e}_v}\dot{V}(\theta_o) = \overline{\dot{\mathbf{a}}}(\theta)\mathbf{e}_v^*\mathbf{a}(\theta) + \overline{\mathbf{a}}(\theta)\mathbf{e}_v^*\dot{\mathbf{a}}(\theta) \tag{10}$$

We now write $\mathbf{a}(\theta)^*\mathbf{a}(\theta) = q_1$ and furthermore $\mathbf{e}_v = q_2 \cdot \mathbf{a}(\theta)$, where q_1 and q_2 are a constants, the latter since the two vectors span the same subspace. Hence we get

$$\bar{\mathbf{g}}^* = \dot{\mathbf{a}}^*(\theta)q_1q_2 + \dot{\mathbf{a}}^T(\theta)\bar{\mathbf{e}}_v\mathbf{a}^*(\theta)$$

and, describing the array elements through their phase lags only, for a general array geometry,

$$\begin{aligned}\bar{\mathbf{g}}^*\mathbf{a}(\theta) &= (\dot{\mathbf{a}}^*(\theta)\mathbf{a}(\theta)q_1q_2 + \dot{\mathbf{a}}^T(\theta)\bar{\mathbf{a}}(\boldsymbol{\theta})q_1q_2) \\ &= 0\end{aligned} \tag{11}$$

■

We now conclude that

Lemma 2: Given an estimator cost function that contains the quadratic form $Q(\theta, \mathbf{e}_v)$, any error $\Delta\mathbf{e}_v$ parallel to $\mathbf{a}(\theta_o)$ and small enough to fulfill $\nabla_{\mathbf{e}_v} \approx \nabla_{\mathbf{e}_v+\Delta\mathbf{e}_v}$, does not cause other than higher order DOA bias.

Proof: The proof follows immediately from Theorem 1. ■

The above Lemma points at one possible strategy for obtaining zero bias: If the errors $\Delta\mathbf{e}_v$ are deterministic and controllable[2], a pre-processing that rotates them into orthogonality with $\overline{\nabla_{\mathbf{e}_v}\dot{V}(\theta_o)}$, will reduce or eliminate the bias.

In the next section, using a Taylor expansion, the result of Lemma 1 will be extended to encompass all cost functions for which the necessary derivatives exist, not only those containing the quadratic form $Q(\theta, \mathbf{e}_v)$.

IV. Reducing bias

Below the derivative of the DOA estimator cost function will be Taylor expanded around the estimated point $\dot{V}(\hat{\theta}, \hat{\mathbf{E}}_v)$. The purpose is to establish a relation between a certain (mapping) error $\Delta\mathbf{E}_v$ in $\mathbf{E}_v$ on the DOA estimate $\hat{\theta}$. Initially multi emitter formalism is used but later the analysis will be restricted to one emitter scenarios.

Third order Taylor expansion around the (single) correct DOA θ_o of the estimator cost function V is required to analyze the first two bias mechanisms in (1). Second order expansion of the derivative $\dot{V}$ with respect to θ is equivalent and yields

$$\dot{V}(\hat{\theta}, \hat{\mathbf{E}}_v) = \dot{V}(\theta_o, \mathbf{E}_v) + \ddot{V}(\theta_o, \mathbf{E}_v)\Delta\theta \tag{12a}$$

$$+2\,\mathrm{Re}\left\{\sum_{k=1}^{d}\nabla_{\mathbf{e}_v}\dot{V}(\theta_o, \mathbf{E}_v)^T\Delta\mathbf{e}_k\right\} \tag{12b}$$

$$+\frac{1}{2}\dddot{V}(\theta_o, \mathbf{E}_v)\Delta\theta^2 \tag{12c}$$

$$+2\,\mathrm{Re}\left\{\sum_{k=1}^{d}\nabla_{\mathbf{e}_v}\ddot{V}(\theta_o)^T\Delta\mathbf{e}_k\Delta\theta\right\} \tag{12d}$$

$$+\,\mathrm{Re}\left\{\sum_{k=1}^{d}\sum_{l=1}^{d}Tr\left(\mathbf{H}1_{kl}\Delta\mathbf{e}_l\Delta\mathbf{e}_k^*\right)\right\} \tag{12e}$$

$$+\,\mathrm{Re}\left\{\sum_{k=1}^{d}\sum_{l=1}^{d}Tr\left(\mathbf{H}2_{kl}\Delta\mathbf{e}_l\Delta\mathbf{e}_k^T\right)\right\} \tag{12f}$$

$$+o(N^{-1}) \tag{12g}$$

Here $\mathbf{E}_s$ contains the d signal eigenvectors[3]. The Hessian matrices $H1$ and $H2$ are given by

$$\mathbf{H}1_{kl}(i,j) = \left.\frac{\partial^2\dot{V}(\cdot,\cdot)}{\partial\bar{\mathbf{e}}_k(i)\partial\mathbf{e}_l(j)}\right|_{(\theta,\mathbf{E}_s)} \tag{13}$$

$$\mathbf{H}2_{kl}(i,j) = \left.\frac{\partial^2\dot{V}(\cdot,\cdot)}{\partial\mathbf{e}_k(i)\partial\mathbf{e}_l(j)}\right|_{(\theta,\mathbf{E}_s)} \tag{14}$$

We will now use the above expansion to analyze the effect of different types of errors.

A. Using the gradient to express bias

As shown in [4], under imperfect array mapping and other deterministic errors, the first error term in (1) can become large. This shows up in the Taylor expansion term (12b) which then will dominate over the higher order terms and we get the following simplified expression for the DOA bias b. (Note that if $\Delta\mathbf{e}_k$ is random we in stead get standard deviation)

$$\begin{aligned}\Delta\theta \triangleq b = &-\frac{2\,\mathrm{Re}\left\{\sum_{k=1}^{d}\nabla_{\mathbf{e}_v}\dot{V}(\theta_o)^T\Delta\mathbf{e}_k\right\}}{\ddot{V}(\theta_o, \mathbf{E}_s)} \\ &+o(N^{-1})\end{aligned} \tag{15}$$

Hence, for N not to small and reasonable SNR, the resulting DOA bias $\Delta\theta$ can be reduced if (15) is minimized by reshaping the (deterministic) errors $\Delta\mathbf{e}_k$ such that

$$\overline{\nabla_{\mathbf{e}_v}\dot{V}(\theta_o)} \perp \Delta\mathbf{e}_k, \quad \forall\, k \tag{16}$$

By Lemma 1 we know that if $\Delta\mathbf{e}_k$ can be rotated to become parallel with the response vector $\mathbf{a}(\theta_o)$ then θ_o will be estimated correctly without other than higher order bias.

While this orthogonality yields $\mathrm{Re}\left\{\overline{\nabla_{\mathbf{e}_v}\dot{V}(\theta_o)}\cdot\Delta\mathbf{e}_k\right\} = 0$, and $\mathrm{Im}\left\{\overline{\nabla_{\mathbf{e}_v}\dot{V}(\theta_o)}\cdot\Delta\mathbf{e}_k\right\} = 0$, and we need only the former to reduce 1^{st} order bias, it is easily verified by simulations

[2] Mapping between two known arrays of different known shapes is one example of a pre-processing which yields this type of errors

[3] Note that we earlier confined our application to $d = 1$ emitters

that the extra latter condition helps in reducing the higher order terms in (12), i.e. (12c)-(12f). These become visible when SNR is high and thus constitute the *residual bias* in such scenarios.

On the other hand minimizing both the Re- and the Im -terms ties up more degrees of freedom in the mapping matrix and may therefore compromise bias reduction. This measure is therefore only recommended in high SNR scenarios.

Both these effects are illustrated and commented on in the Simulations section.

When mapping from a matrix $\mathbf{A}_r$ of real response vectors $\mathbf{a}_r$ onto a matrix $\mathbf{A}_v$ of virtual response vectors $\mathbf{a}_v$ in a scenario with only one emitter, i.e. $d = 1$, the bias expression (15) can be simplified into a more useful form if we replace the eigenvector error $\Delta\mathbf{e}_v = \mathbf{T}^*\mathbf{e}_s - \mathbf{e}_v$ with the array mapping error $\Delta\mathbf{a} = \mathbf{T}^*\mathbf{a}_r - \mathbf{a}_v$ (presuming the response vectors have been normed to length 1). This is allowed since in the one emitter case these two vectors span the same subspace.

The bias expression (15) then becomes

$$b_T = -\frac{2\,\mathrm{Re}\left\{\bar{\mathbf{g}}_v^*\Delta\mathbf{a}\right\}}{\ddot{V}(\theta_o, \mathbf{e}_v)} \tag{17}$$

where we for convenience have denoted the gradient in (15) by $\mathbf{g}_v$ and used the notation b_T to indicate the dependence of the bias on the mapping matrix $\mathbf{T}$.

B. Reducing deterministically induced bias

To minimize bias one would naturally choose $\mathbf{T}$ as

$$\mathbf{T}_{opt} = \arg\min_{\mathbf{T}} |b_T|^2 \tag{18}$$

For the specific application of array mapping over a sector one would use a set $\{\theta_i\}_1^{N_{cal}}$ of adjacent calibration directions to comprise the sector and then choose $\mathbf{T}$ as

$$\mathbf{T}_{opt} = \arg\min_{\mathbf{T}} \sum_{i=1}^{N_{cal}} |b_{T,i}|^2 \tag{19}$$

where $b_{T,i}$ is the bias according to (17) in direction θ_i. However, since we also need keeping the eigenvector (mapping) errors $\Delta\mathbf{e}_{v,i}$ small we can extend the sum with the corresponding manifold matching terms into

$$\mathbf{T}_{opt} = \arg\min_{\mathbf{T}} \left\{(1-k)\sum_{i=1}^{N_{cal}} |\Delta\mathbf{a}(\theta_i)|^2 + k\sum_{i=1}^{N_{cal}} |b_{T,i}|^2\right\} \tag{20}$$

where $\Delta\mathbf{a}(\theta_i) = \mathbf{T}^*\mathbf{a}_r(\theta_i) - \mathbf{a}_v(\theta_i)$ and k is a weighting constant smaller than, but close to 1. Extending the sum as in (20) also ensures a sufficient number of equations to solve the corresponding least square problem for $\mathbf{T}$.

The weighting constant k is chosen smaller than, but close to 1 to ensure the orthogonality (16) in the real part. The value of k is a compromise between putting high emphasis on the orthogonality between the mapping error and the gradient on one hand, and on the other hand to keep the length of the mapping errors limited so that the linear terms in the Taylor expansion dominate. Low signal SNRs contribute to the mapping errors so we expect the optimal k to be chosen slightly lower in such cases. This aspect is illustrated in the Simulation section, see Figures 8 and 9.

1) Simplifications to the bias minimizing algorithm: When implementing the suggested design algorithm a few practical details are worth noting:

1) For a typical broadside mapping sector, comprised by the set $\{\theta_i\}$ of calibration directions, say up to $+/-15^o$ wide, the beamwidth (and hence the Hessian $\ddot{V}(\theta_i, \mathbf{e}_{v,i})$ of a $\lambda/2$ spaced ULA does not vary much. As verified by simulations we can therefore replace the Hessians $\ddot{V}(\theta_i, \mathbf{e}_{v,i})$ in (17) with their average value across the sector and then modify k accordingly.
2) The scalar products $\mathbf{g}_{v,i}^*\mathbf{a}_v(\theta_i)$ are all zero. This simplifies the (17) bias terms into
$$b_T = const. \times 2\,\mathrm{Re}\left\{\bar{\mathbf{g}}^*\mathbf{T}\mathbf{a}_r\right\} \tag{21}$$
3) As mentioned before, if we devote some of the degrees of freedom in $\mathbf{T}$ to reducing also the imaginary part of the scalar products $\mathbf{g}_{v,i}^*\mathbf{T}^*\mathbf{a}(\theta_i)$, i.e. we drop the Re operator in (21), then the rest bias can be noticeably reduced.This measure will deteriorate first order bias suppression slightly but in high SNR scenarios simulations show it to be a rewardable step. See the Simulations section.

With the above simplifications, when the mapping is onto a ULA and SNR is high, (20) takes the simplified form

$$\mathbf{T}_{opt} = \arg\min_{\mathbf{T}} \left\{(1-k)\sum_{i=1}^{N_{cal}} |\Delta\mathbf{a}(\theta_i)|^2 + k\sum_{i=1}^{N_{cal}} \left|\bar{\mathbf{g}}_{v,i}^*\mathbf{T}\mathbf{a}_r(\theta_i)\right|^2\right\} \tag{22}$$

where the weighting constant k now includes the average of the Hessians. It is presumed that the exact value of k is chosen empirically.

2) Some comments: Note that the suggested algorithm (20) reduces DOA estimation bias not by minimizing the size of the virtual array mapping errors $\Delta\mathbf{e}_{v,i}$ but instead by giving said errors such a structure that they no longer affect the DOA estimates. This approach makes mapping between dissimilar arrays feasible over sectors much wider than would be possible for a corresponding (pure) array manifold match (where $k = 0$).

Furthermore, note that due to Theorem 1 we can usually replace $\Delta\mathbf{a}(\theta^{(i)})$ in (20) by $\mathbf{T}^*\mathbf{a}_r(\theta^{(c)})$. This simplification will be used also in the MSE analysis later on.

The action of the design algorithm (20) can be described as generating one transformation matrix $\mathbf{T}$ for each of the N_{cal} calibration directions, and then, in a least square sense, finding the best compromise across all the N_{cal} transformation matrices. If the uncorrected bias is large its reduction by adding the last sum term can be dramatic. In [4] a reduction factor exceeding 100 was demonstrated despite the sum terms not being individually weighted by the corresponding Hessians.

Finally note that (15) only uses a parabolic, i.e. symmetrical, approximation of the cost function V, hence in this description bias generated by zero mean noise will not be visible, [11].

C. An expression for minimized bias

Algorithm (20) reduces bias primarily by enforcing orthogonality between the gradient $\mathbf{g}_v$ and the error $\Delta\mathbf{e}_v$ in (15), an approach that will be shown to be quite effective. However, after having performed a mapping using a transformation matrix designed according to (20), it should again be noted that for the remaining bias, expression (15) is no longer valid.

This is evident from an inspection of the Taylor expansion (12) where the first order term (12b) through the enforced orthogonality now is small but the mapping error $\Delta\mathbf{e}_v$ still may be relatively large.

Assuming good orthogonality, and hence a small $\Delta\theta$ in (12) we can neglect the terms (12b) and (12c). Thus an higher order expression for minimized bias (the remaining bias *after* a mapping according to (20) has been performed) is

$$\Delta\theta_{rem} = \tag{23a}$$

$$-\frac{2\,\mathrm{Re}\left\{\sum_{k=1}^{d}\sum_{l=1}^{d} Tr\left(\mathbf{H}1_{kl}\Delta\mathbf{e}_l\Delta\mathbf{e}_k^*\right)\right\}}{\ddot{V}(\theta_o,\mathbf{E}_v)+2\,\mathrm{Re}\left\{\sum_{k=1}^{d}\nabla_{\mathbf{e}_s}\ddot{V}(\theta_o,\mathbf{E}_v)^T\Delta\mathbf{e}_k\right\}} \tag{23b}$$

$$-\frac{2\,\mathrm{Re}\left\{\sum_{k=1}^{d}\sum_{l=1}^{d} Tr\left(\mathbf{H}2_{kl}\Delta\mathbf{e}_l\Delta\mathbf{e}_k^T\right)\right\}}{\ddot{V}(\theta_o,\mathbf{E}_v)+2\,\mathrm{Re}\left\{\sum_{k=1}^{d}\nabla_{\mathbf{e}_s}\ddot{V}(\theta_o)^T\Delta\mathbf{e}_k\right\}} \tag{23c}$$

$$+o(N^{-1})$$

This entity will be plotted as reference in the Simulations section. It illustrates the limits of the effectiveness of the first order based bias reduction approach (20).

Since for a ULA the second Hessian $\mathbf{H}2 = 0$, it is evident from (23) that in order to minimize the remaining bias $\Delta\theta_{rem}$ the mapping should have the property to leave errors $\Delta\mathbf{e}_k$ that fall in the nullspace of the first Hessian $\mathbf{H}1$. As supported by simulations, due to the similar structure in the ULA case between $\mathbf{H}1$ and the gradient $\nabla_{\mathbf{e}_v}\dot{V}(\theta_o)$ (see Appendix I) this is advocated by requiring (16) to hold in both the Re and Im parts.

This explains the lower residual bias for high SNR cases illustrated in the Simulations section if the Re operator is omitted in algorithm (20).

D. Reducing noise induced bias

Assuming no deterministic errors, measurement noise will dominate and we have $\Delta\mathbf{e}_v = O(N^{-1/2})$. It was shown in [11] that for relatively few snapshots ($N = 20$ was used as an example) the bias caused by noise can dominate over STD caused by noise. Using the full Taylor expansion (12) and the result in Theorem 2 (see below), [11] derives the following expression for MUSIC noise induced bias

$$E\{\Delta\theta\} \approx$$

$$-\frac{1}{N}\frac{2\sum_{k=1}^{d}\frac{(m-d-1)\lambda_k\sigma_n^2}{(\lambda_k-\sigma_n^2)^2}\,\mathrm{Re}\left\{\dot{\mathbf{a}}(\theta_o)\mathbf{e}_k\mathbf{e}_k^*\mathbf{a}(\theta_o)\right\}}{\ddot{V}(\theta_o,\mathbf{E}_v)}$$

$$-\frac{\dddot{V}(\theta_o,\mathbf{E}_v)}{\ddot{V}(\theta_o,\mathbf{E}_v)}var(\Delta\theta) \tag{24}$$

In (24) m denotes the number of array elements, d the number of emitters, λ_k and σ_n^2 the signal and noise eigenvalues respectively of the data covariance matrix.

The last term in (24) can be interpreted as a measure of the degree of asymmetry in the cost function V at the true DOA:s, whereas the first term describes the interaction between individual DOA:s. This interaction is larger if at least one λ_k is close to σ_n^2 which is the case for adjacent and/or coherent emitters.

For one single emitter however, this first term vanishes and the resulting bias depends only on the variance and on the second and third order derivatives of the cost function, i.e. the above mentioned asymmetry.

To reduce noise induced bias in the single emitter case we can then use that $\ddot{V}$ and $\dddot{V}$ are analytically known. As an example, to avoid the well known end-fire bias for ULA MUSIC we can estimate $\sin\theta$ instead of θ which makes $\ddot{V}(\theta_o,\mathbf{E}_v) = 0, \forall\theta$ and the noise induced bias is avoided, [11].

A corresponding measure for the multi emitter case is in principle also possible, but eliminating the first term analytically seems tedious and in general requires knowledge of the true DOA. Pre-processing to yield $\dot{\mathbf{a}}^*(\theta)\perp\mathbf{e}_k\forall k$ (in the real part), or choosing an array geometry for which $\dot{\mathbf{a}}^*(\theta)\perp\mathbf{a}^*(\theta)$, may be options however.

Assuming a real array $\mathbf{A}_r$ that is mapped by $\mathbf{T}$ onto a virtual array $\mathbf{A}_v$ using a set $\left\{\theta^{(c)}\right\}$ of N_{cal} calibrated directions across a sector, we can give the mapped data approximately this property (with the aim of zeroing the first term in (24)) by taking

$$\begin{aligned}\mathbf{T}_{opt} &= \arg\min_{\mathbf{T}}(1-k)\left\|\Delta\mathbf{A}(\boldsymbol{\theta}^{(c)})\right\|_F^2 + \\ &\quad k\cdot\sum_{i=1}^{N_{cal}}\left|\mathrm{Re}\left\{\dot{\mathbf{a}}_v(\theta^{(i)})\mathbf{T}\mathbf{a}_r(\theta^{(i)})\right\}\right|^2\end{aligned} \tag{25}$$

where again $\Delta\mathbf{A}(\boldsymbol{\theta}^{(c)}) = \mathbf{T}^*\mathbf{A}_r(\boldsymbol{\theta}^{(c)}) - \mathbf{A}_v(\boldsymbol{\theta}^{(c)})$ is the manifold mapping error.

Because both terms in (24) scale with $\frac{1}{N}$, this design of $\mathbf{T}$ is expected to improve the DOA estimates only if these are based on relatively few snapshots, and the scenario is one for which noise induced bias is expected to dominate over noise induced STD. A typical such scenario would involve about 20 snapshots [11], and/or adjacent or coherent emitters near ULA end fire directions.

As in (20) the value of the weighting constant k has to be determined empirically. Likewise we cannot allow $k = 0$ for then the LS solution to both (20) and (25) becomes $\mathbf{T} = 0$.

V. Reducing variance

To formulate an expression for DOA error variance we need the following result also used by [12], p. 723.

Theorem 3: If measurement noise is both spatially white and circularly symmetric, then the noise induced signal eigenvector errors $\Delta\mathbf{e}_{s,i}^{(noise)}$ are asymptotically (for large N) jointly

Gaussian distributed with zero mean and

$$E\left\{\Delta\mathbf{e}_{s,i}^{(noise)}\Delta\mathbf{e}_{s,i}^{(noise)*}\right\} \triangleq cov(\Delta\mathbf{e}_{s,i}^{(noise)})$$
$$= \frac{\lambda_i}{N}\cdot\sum_{\substack{k=1\\k\neq i}}^{d}\frac{\lambda_k}{(\lambda_k-\lambda_i)^2}\mathbf{e}_{s,k}\mathbf{e}_{s,k}^* + \frac{\lambda_i}{N}\cdot\sum_{k=1}^{m-d}\frac{\sigma_k^2}{(\sigma_k^2-\lambda_i)^2}\mathbf{e}_{n,k}\mathbf{e}_{n,k}^* \quad (26)$$

Here λ_i and σ_k^2 are eigenvalues of the data covariance matrix, and $\mathbf{e}_{n,k}$ are the noise eigenvectors. In our application the number d of emitters d is limited to 1.

Proof: See [12] and the references therein. ■

The first sum in (26) describes the interaction between different signal eigenvectors $\mathbf{e}_{s,k}$, it therefore disappears if the scenario contains one emitter only. In this case $cov(\Delta\mathbf{e}_{s,i}^{(noise)})$ is determined solely by the noise eigenvectors $\mathbf{e}_{n,k}$ and the factor $\frac{\lambda_i}{N}\cdot\frac{\sigma_k^2}{(\sigma_k^2-\lambda_i)^2}$, a measure of SNR.

A. Using the gradient to express variance

If the errors $\Delta\mathbf{e}_v$ are limited and we presume one emitter only, by (15) we can write any DOA error, deterministic or random, as

$$\Delta\theta = -\frac{2\,\mathrm{Re}\left\{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v\right\}}{\ddot{V}(\theta_o,\mathbf{E}_v)} = -\frac{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v+\overline{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v}}{\ddot{V}(\theta_o,\mathbf{E}_v)} \quad (27)$$

where $\Delta\mathbf{e}_v$ is the error in the single signal eigenvector $\mathbf{e}_v$ contained in $\mathbf{E}_v$.

To study the total effect on $\dot{V}$ of $\Delta\mathbf{e}_{v,(var)}^{(noise)}$, i.e. *variance*, we need to examine the entity $2\,\mathrm{Re}\left\{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_{v,(var)}^{(noise)}\right\}$. For notational convenience we now write this expression as $2\,\mathrm{Re}\left\{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v\right\}$.

Since this term is random we use second order statistics and assess its magnitude by forming the expectation

$$E\left\{(\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v+\overline{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v})(\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v+\overline{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v})^*\right\} =$$
$$E\left\{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v\Delta\mathbf{e}_v^*\overline{\mathbf{g}_v}\right\}+ \quad (28a)$$
$$E\left\{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v\Delta\mathbf{e}_v^T\mathbf{g}_v\right\}+ \quad (28b)$$
$$E\left\{\overline{\overline{\mathbf{g}}^*\Delta\mathbf{e}_v}\Delta\mathbf{e}_v^*\overline{\mathbf{g}_v}\right\}+ \quad (28c)$$
$$E\left\{\overline{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v}\Delta\mathbf{e}_v^T\mathbf{g}\right\} \quad (28d)$$

By the Theorem 3 the entities (28a) and (28b) are both zero. We can therefore write the $\dot{V}$ variance as the expectation

$$E\left\{2\,\mathrm{Re}(\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v)2\,\mathrm{Re}(\Delta\mathbf{e}_v^*\overline{\mathbf{g}_v})\right\} =$$
$$E\left\{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v\mathbf{e}_v^*\overline{\mathbf{g}_v}\right\}+E\left\{\overline{\overline{\mathbf{g}_v}^*\Delta\mathbf{e}_v}\Delta\mathbf{e}_v^T\mathbf{g}_v\right\} =$$
$$2\,\mathrm{Re}\left\{\overline{\mathbf{g}_v}_v^*cov(\Delta\mathbf{e}_v)\overline{\mathbf{g}}_v\right\} \quad (29)$$

Theorem 2 now gives an expression for $cov(\Delta\mathbf{e}_v)$, hence for a non mapped one emitter case, using the gradient, we can write the (non-mapped) DOA error variance as

$$var(\Delta\theta) = \frac{2\frac{\lambda_1}{N}\cdot\sum_{k=1}^{m-d}\frac{\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}\overline{\mathbf{g}}_v^*\mathbf{e}_{n,k}\mathbf{e}_{n,k}^*\overline{\mathbf{g}}_v}{\ddot{V}^2(\theta_o,\mathbf{E}_s)} \quad (30)$$

where we have dropped the Re operator since $\overline{\mathbf{g}_v}^*\mathbf{e}_{n,k}\mathbf{e}_{n,k}^*\overline{\mathbf{g}}_v = |\overline{\mathbf{g}}^*\mathbf{e}_{n,k}|^2$ is real.

Note that we have also dropped the higher order terms corresponding to (12c)-(12f). For noise like errors this is well motivated since no realistic pre-processing can null out all noise eigenvectors, see Figure 2 and the associated argument, so the effect of the higher order terms is expected to be negligible..

We now observe that for the one emitter case the variance is determined by the $m-1$ scalar products $\overline{\mathbf{g}}_v^*\mathbf{e}_{n,k}$. This opens up a prospect for minimizing variance when pre-processing is involved: If the pre-processing rotates $\overline{\mathbf{g}}_v$ and $\mathbf{e}_{n,k}$ relative to one another variance could possibly be controlled. We will address this issue in the sequel.

B. Using the gradient to express mapped variance

Due to our partitioning of $\Delta\mathbf{e}_v$ into separate bias- and variance generating parts, putting the former to zero we can for the latter assume

$$E\left\{\Delta\mathbf{e}_{v,(var)}^{(noise)}\right\} = E\left\{\mathbf{T}\hat{\mathbf{e}}_{s,r}-\mathbf{e}_{s,v}\right\} = 0 \quad (31)$$

The virtual signal eigenvector error $\Delta\mathbf{e}_v = \mathbf{T}\hat{\mathbf{e}}_{s,r}-\mathbf{e}_{s,v}$ has all its noise power is in the first term $\mathbf{T}\hat{\mathbf{e}}_{s,r}$, i.e. associated with the real array. Hence, to derive an expression for DOA variance we can study the variance in $\Delta\mathbf{e}_{s,r,T} \triangleq \mathbf{T}(\hat{\mathbf{e}}_{s,r}-\mathbf{e}_{s,r})$ where $\mathbf{e}_{s,r}$ is the true non-mapped signal eigenvector associated with the real array. This variance is expressed as

$$var(\Delta\mathbf{e}_{s,r,T}) = E\left\{\mathbf{T}(\mathbf{T}\hat{\mathbf{e}}_{s,r}-\mathbf{e}_{s,r})(\mathbf{T}\hat{\mathbf{e}}_{s,r}-\mathbf{e}_{s,r})^*\mathbf{T}^*\right\} = \mathbf{T}cov(\Delta\mathbf{e}_{s,r})\mathbf{T}^* \quad (32)$$

The above expression has the desired property of being quadratic in $\mathbf{T}$ and therefore lends itself well to least square minimization.

As in (30) we can now use Theorem 2 and express the mapped DOA variance as

$$var_T(\Delta\theta) = \frac{2\frac{\lambda_1}{N}\cdot\sum_{k=1}^{m-d}\frac{\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}\overline{\mathbf{g}}_v^*\mathbf{T}\mathbf{e}_{n,k}\mathbf{e}_{n,k}^*\mathbf{T}^*\overline{\mathbf{g}}_v}{\ddot{V}^2(\theta_o,\mathbf{E}_v)} \quad (33)$$

Note that (33) expresses true mapped variance only if the transformation matrix $\mathbf{T}$ is unitary. If this is not the case multiplication of the data with $\mathbf{T}$ can mean a net power gain and thus a scaling on the resulting DOA variance. This has to be taken into account when comparing different signal processing solutions.

If $\mathbf{T}$ is well conditioned we can in principle correct for the non-unitarity by the pre-multiplication $\mathbf{T}(\mathbf{T}\mathbf{T}^*)^{-\frac{1}{2}}$. The multiplications in (33) will then yield a weighted matrix product of

the structure $\mathbf{T}(\mathbf{T}\mathbf{T}^*)^{-\frac{1}{2}}\mathbf{e}_{n,k}\mathbf{e}_{n,k}^*\mathbf{T}(\mathbf{T}\mathbf{T}^*)^{-\frac{1}{2}*}$ which can be seen as a weighted projector onto the range space of $\mathbf{T}$. We thus see that the resulting mapped variance will depend on the eigenvectors of the real array output covariance matrix, the projection properties of $\mathbf{T}$, and on the derivatives comprising the gradient $\overline{\mathbf{g}}_v$.

With the above correction term $(\mathbf{T}\mathbf{T}^*)^{-\frac{1}{2}}$ the expression (33) for mapped MSE no longer is quadratic in $\mathbf{T}$. This complicates the minimization. However, since it is the *mapped* DOA MSE we want to minimize and the same $\mathbf{T}$ and gradient $\overline{\mathbf{g}}_v$ are used in the bias expression, we can still use (33) as a minimization criterion for the variance part of the MSE:

Without the pre-multiplication the above variance term (33) contains the form $\overline{\mathbf{g}}_v^*\mathbf{T}\mathbf{e}_{n,k}\mathbf{e}_{n,k}^*\mathbf{T}^*\overline{\mathbf{g}}_v$ where $\mathbf{T}$ can be seen as a rotator and re-scaler operating on the noise eigenvectors $\mathbf{e}_{n,k}$. The variance is then determined by the scalar product $\overline{\mathbf{g}}_v^*\mathbf{T}\mathbf{e}_{n,k}$. This can be given a geometrical interpretation and will be developed further in the next sections.

VI. General requirements on the pre-processing matrix

Minimization of noise induced DOA variance under mapping has been studied in great detail. Aside from improving the SNR by spatial filtering (which assumes some pre-knowledge of the searched DOA:s) the best we can do is to avoid impeding the estimator effectiveness and try to retain the Cramér-Rao Bound CRB under any pre-processing that we want to apply. We thus require that the unmapped CRBs should be retained, then DOA estimation variance will not be impeded by the pre-processing. Assuming spatially white noise and a unitary $\mathbf{T}$ it has been shown in [3] and [2] that if the pre-processing matrix $\mathbf{T}$ fulfills

$$\mathbf{T}^*\mathbf{T}\left[\begin{array}{cc}\mathbf{A}(\theta^{(i)}) & \mathbf{D}(\theta^{(i)})\end{array}\right]=\left[\begin{array}{cc}\mathbf{A}(\theta^{(i)}) & \mathbf{D}(\theta^{(i)})\end{array}\right]\ \forall i \tag{34}$$

then at $\theta^{(i)}$ the DOA estimator variance performance will be retained.

Here the matrix $\mathbf{D}(\theta^{(i)})$ contains the derivatives w.r.t. θ of the response vectors $\mathbf{a}$ of the manifold matrix $\mathbf{A}$ evaluated at $\theta^{(i)}$. $\mathbf{T}$ is also presumed to be unitary.

Obviously, if $\mathbf{T}$ is full rank we have $\mathbf{T}^*\mathbf{T}=\mathbf{I}$ and (34) is trivially fulfilled. However if $\mathbf{T}$ is of poor condition some dimensions in the space $\mathbf{T}^*\mathbf{T}$ will not be well spanned. This plus the wide sector compromise for $\mathbf{T}$ in the signal reconnaissance application are the main reasons behind variance increase in the sector mapping application. A corresponding situation is *dimension reduction* where $\mathbf{T}$ has fewer rows than columns.

In the formulation (34) $\mathbf{T}$ is assumed unitary. Hence $\mathbf{T}^*\mathbf{T}$ can be seen as a projector, and this projector must not project away the dimensions needed to attain the CRB, especially not those spanned by the derivatives. However in the sector mapping application, minimizing both manifold mapping error and bias (as in (20)) usually results in a non-unitary $\mathbf{T}$. This is easily seen from simulations.

Generalizing (34) for this case we write

$$\mathbf{\Pi}_T\left[\begin{array}{cc}\mathbf{A}(\theta^{(i)}) & \mathbf{D}(\theta^{(i)})\end{array}\right]=\left[\begin{array}{cc}\mathbf{A}(\theta^{(i)}) & \mathbf{D}(\theta^{(i)})\end{array}\right]\ \forall i \tag{35}$$

where $\mathbf{\Pi}_T=\mathbf{T}^*(\mathbf{T}\mathbf{T}^*)^{-1}\mathbf{T}\triangleq\mathbf{T}^\dagger\mathbf{T}$.
A slight reformulation of (34) now gives

$$\mathbf{T}^\dagger\left[\begin{array}{cc}\mathbf{T}\mathbf{A}(\theta^{(i)}) & \mathbf{T}\mathbf{D}(\theta^{(i)})\end{array}\right]=\left[\begin{array}{cc}\mathbf{A}(\theta^{(i)}) & \mathbf{D}(\theta^{(i)})\end{array}\right]\ \forall i \tag{36}$$

An interesting question now is wether the fulfillment of (35) in any way is impeded by the zero bias condition $\mathrm{Re}\{\bar{\mathbf{g}}^*\Delta\mathbf{e}_k\}=0$. We shall see that this can be the case.

The last term in the design algorithm (20) advocates the zero bias condition by pushing $\mathrm{Re}\left\{\bar{\mathbf{g}}^{(i)*}\Delta\mathbf{A}\right\}=\mathrm{Re}\left\{\bar{\mathbf{g}}^{(i)*}(\mathbf{T}\mathbf{A}_r(\theta^{(i)})-\mathbf{A}_v(\theta^{(i)}))\right\}$ towards zero. However from Theorem 1, for a wide class of DOA estimators we have that $\bar{\mathbf{g}}^*\mathbf{A}_v=0$ so the zero bias condition can be simplified into $\mathrm{Re}\left\{\bar{\mathbf{g}}^{(i)*}\mathbf{T}\mathbf{A}_r(\theta^{(i)})\right\}=0$.

Applying this to the left membrum in (36) we see that a $\mathbf{T}$ designed for minimum bias, as seen in the real part, will tend to rotate all column vectors $\mathbf{T}\mathbf{A}_r(\theta^{(i)})$ into orthogonality with $\bar{\mathbf{g}}^{(i)}$, i.e. the vector $\bar{\mathbf{g}}^{(i)}$ forms a null-space of $\mathbf{T}$ (at least in the real part). This will tend to remove a dimension from $\mathbf{\Pi}_T$ and therefore somewhat restricts the space spanned by the set of said column vectors. According to (35) this can make it more difficult to attain the CRB.

In the Simulations section this effect is illustrated in Figure 5 which shows MUSIC spectra for unmapped and mapped data, the latter using a matrix $\mathbf{T}$ obtained from algorithm (20) with, and without, the Re operator in the sum terms.

VII. Reducing mean square error MSE

In order to design a "best" pre-processing matrix $\mathbf{T}$ when neither deterministic nor noise induced errors can be neglected, we need a suitable criterion function. Preferably this criterion function should be scalar and also convex in $\mathbf{T}$ to yield a practicable optimization problem. The Mean Square of the total error, $MSE=bias^2+variance$, provides one such criterion function which will be used here.

Using the earlier results we can now form expressions for both unmapped and mapped MSE. The latter expression has the desired feature of being quadratic in $\mathbf{T}$.

A. An analytical expression for unmapped DOA MSE

Using (15) and (33), we get the following expression for the total Mean Square Error of the (unmapped) DOA estimates. It describes both bias and variance but presumes validity of the linear expansion (12a) and (12b) of the estimator cost function derivative $\dot{V}$.

$$MSE=bias^2+Var=\frac{4\,\mathrm{Re}\left\{\bar{g}^*\Delta\mathbf{e}_{s,(bias)}^{(det)}\right\}^2}{\ddot{V}^2(\theta_o,\mathbf{E}_s)} \tag{37a}$$

$$+\frac{2\sum_{k=1}^{m-d}\overline{\mathbf{g}}_v^*\mathbf{e}_{n,k}\frac{\lambda_1}{N}\cdot\frac{\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}\mathbf{e}_{n,k}^*\overline{\mathbf{g}}_v}{\ddot{V}^2(\theta_o,\mathbf{E}_s)}, \tag{37b}$$

where we have assumed the measurement noise to spatially white. Therefore all σ_k^2 are equal.

At this point we recall that both (37a) and (37b) were derived using the Taylor expansion (12) around the true DOA. Hence, in the mapped case this expansion becomes more and more approximative the larger the mapping errors.

Any pre-processing based on said (first order) Taylor expansion must meet the requirement that these errors are limited.

B. An analytical expression for mapped DOA MSE

We now assume that the DOA estimator receives pre-processed data and view the mapped data that is fed from the virtual array to the DOA estimator. Assuming one emitter $d = 1$, and aiming at a quadratic expression in $\mathbf{T}$, for the corresponding mapped MSE_T, we therefore write

$$MSE_T = bias_T^2 + Var_T \tag{38}$$

and by the earlier formulations (15) and (33) get the following expression for pre-processed MSE. Apart from assuming limited sized errors $\Delta\mathbf{e}_v = \mathbf{T}^*\mathbf{e}_s - \mathbf{e}_v$ it also assumes that the higher order terms (12c) - (12f) are negligible.

$$MSE_T = \frac{4\,\mathrm{Re}\left\{\bar{g}^*\Delta\mathbf{a}_v\right\}^2}{\ddot{V}^2(\theta_o, \mathbf{e}_v)} \tag{39a}$$

$$+\frac{2\sum_{k=1}^{m-d}\overline{\mathbf{g}}_v^*\mathbf{T}\hat{\mathbf{e}}_{n,k}\frac{\lambda_1}{N}\cdot\frac{\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}\hat{\mathbf{e}}_{n,k}^*\mathbf{T}^*\overline{\mathbf{g}}_v}{\ddot{V}^2(\theta_o, \mathbf{e}_v)} \tag{39b}$$

We observe from (39) that the MSE minimizing $\mathbf{T}$ will depend on SNR through the factor $\sigma_k^2/(\sigma_k^2 - \lambda_1)^2$. This means that we either have to estimate the prevailing SNR or, which is preferable, use a pre-calculated "design SNR", f. ex. 10 dB, according to which the mapping matrix $\mathbf{T}$ is constructed.

In a practical sector mapping application one would (per sector) pre-calculate a number of matrices $\mathbf{T}$ for different SNR values and then, once the proper sector has been determined, via an eigenvalue decomposition performed on the non-mapped data, estimate the SNR. Thereafter the best $\mathbf{T}$ would be used.

Finally notice that both (39a) and (39b) in the MSE_T expression (39) could in principle be extended with the higher order (23) terms to gain a somewhat better accuracy. However, any procedure to minimize MSE_T is complicated by such an extension since MSE_T would then no longer be quadratic in $\mathbf{T}$. This option will therefore not be further considered.

VIII. Minimizing mapped MSE: A geometrical interpretation

From (39) it is clear that, disregarding the proportionality constants involved and the higher order terms, mapped variance is essentially determined by the quadratic form $\sum_k \overline{\mathbf{g}}_v^*\mathbf{T}\hat{\mathbf{e}}_{\mathbf{e}_{n,k}}\hat{\mathbf{e}}_{n,k}^*\mathbf{T}^*\overline{\mathbf{g}}_v$. Formally, we can view $\mathbf{T}$ as a mapping on $\overline{\mathbf{g}}_v$, and in doing so we see that in order to minimize variance $\mathbf{T}$ should have the property to rotate $\overline{\mathbf{g}}_v$ into orthogonality with all (real array) noise eigenvectors. This means that $\overline{\mathbf{g}}_v^*\mathbf{T}$ as much as possible should lie in the corresponding signal subspace $\Re\{\hat{\mathbf{e}}_s\} \approx \Re\{\mathbf{a}_r\}$ .For $\mathbf{T}$ to map into $\Re\{\hat{\mathbf{e}}_s\}$ it thus needs the property

1) $\Re\{\mathbf{T}\} \supseteq \Re\{\hat{\mathbf{e}}_s\}$

We also need the property of $\mathbf{T}$ to be able to map *all components* of $\overline{\mathbf{g}}_v$ onto the signal subspace, i.e. $\Re\{\mathbf{a}_r\}$. No matter if the DOA estimator uses orthogonality between the array manifold and the noise subspace as in MUSIC, or tries to fit the signal subspace to the array manifold as in WSF[4] , for a single emitter the resulting cost function gets the same structure $\mathbf{a}^*(\theta)(\mathbf{I} - \hat{\mathbf{e}}_s\hat{\mathbf{e}}_s^*)$, see f. ex. [18].

The corresponding (unconjugated) gradient with respect to $\hat{\mathbf{e}}_s$becomes

$$g_v = \dot{\overline{\mathbf{a}}}(\theta)\hat{\mathbf{e}}_s^*\mathbf{a}(\theta) + \overline{\mathbf{a}}(\theta)\hat{\mathbf{e}}_s^*\dot{\mathbf{a}}(\theta) \tag{40}$$

so obviously we also need the property of $\mathbf{T}$ that

2 $\Re\{\mathbf{T}\} \supseteq \Re\{\dot{\mathbf{a}}_r\}$

As expected, these two properties of $\mathbf{T}$ are in full accordance with (34).

A. A geometrical interpretation

If the used estimator is bias free and our strategy is to avoid adding new bias, the search for the variance minimizing $\mathbf{T}$ can be visualized in Figure 2 as a rotation of the (mapped) noise subspace relative to the gradient $\bar{\mathbf{g}}_v$. If $\bar{\mathbf{g}}_v \perp \mathbf{a}$ (for minimum bias) and measurement noise covariance is a scaled identity matrix this rotation does not affect the variance. For coloured measurement noise however, when the noise eigenvalues are unequal, there exist optimal directions for the gradient inside the noise subspace.

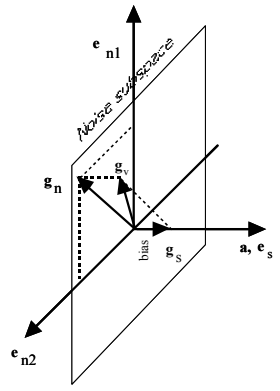

Fig. 2. The mapped signal- ($\mathbf{e}_s$, $\mathbf{a}$) and noise subspaces ($\mathbf{e}_{n1}$ and $\mathbf{e}_{n2}$) are rotated relative to the gradient $\overline{\mathbf{g}}_v$ by the mapping matrix $\mathbf{T}$ for minimum DOA MSE. Bias is represented by by the real part of the projection of $\overline{\mathbf{g}}_v$ onto the signal subspace ($\mathbf{a}$-axis). This figure illustrates the orthogonal nature of bias reduction in relation to variance reduction. No pre-processing exist by which both can be zeroed simultaneously.

Then, by Lemma 1 we observe that those samples of $\Delta\mathbf{e}_{s,(var)}^{(noise)}$ that happen to fall parallel to the response vector $\mathbf{a}(\theta_o)$ do not contribute to DOA estimation variance. However, those that have components inside the corresponding orthogonal space, i.e. the noise subspace, do.

To minimize MSE_T both the deterministic and random parts of the signal subspace eigenvector errors should be orthogonal

[4]Weighted Subspace Fitting

to the gradient $\overline{\mathbf{g}}_v$, i.e. parallel to $\mathbf{a}(\theta_o)$, at least in the real part. This would be the ideal situation, but one obvious problem is the amount of control we can exercise over the random part $\mathbf{a}(\theta_o)$.

By (33), to minimize variance $\mathbf{T}$ should have the property that

$$\overline{\mathbf{g}}_v \in N\left\{\mathbf{T}\hat{\mathbf{E}}_n(\theta_o)\right\} \tag{41}$$

where N denotes the null-space. But by Lemma 1, for bias-free cost functions containing the quadratic form $Q(\theta, \mathbf{e}_v)$ we have for the (sufficient) minimum bias condition

$$\overline{\mathbf{g}}_v \perp \mathbf{T}\mathbf{a}_r(\theta_o) \tag{42}$$

or equivalently

$$\overline{\mathbf{g}} \in N\left\{\mathbf{T}\mathbf{a}_r(\theta_o)\right\} \approx \Re\left\{\mathbf{T}\hat{\mathbf{E}}_n(\theta_o)\right\} \tag{43}$$

where the last equality is exact if the mapping is error free and SNR is sufficiently good.

Obviously no pre-processing exist for which both (41) and (43) can be met simultaneously.

If we take into view that we only require

$$\mathrm{Re}\left\{\mathbf{g}_v \cdot \mathbf{T}\mathbf{a}_r(\theta_o) \cdot \mathbf{T}\mathbf{a}_r(\theta_o)\right\} = 0 \tag{44}$$

the zero bias condition can be relaxed into

$$\mathrm{Re}\left\{\overline{\mathbf{g}}_v\right\} \cdot \mathrm{Re}\left\{\mathbf{T}\mathbf{a}_r(\theta_o)\right\} = \mathrm{Im}\,\overline{\mathbf{g}}_v\,\mathrm{Im}\left\{\mathbf{T}\mathbf{a}_r(\theta_o)\right\} \tag{45}$$

At least if $\mathbf{T}$ is designed around the true DOA θ_o, a larger set of solutions should exist for (45) than for the combination of (41) and (43). When $\mathbf{T}$ is a sector compromise this may still be true but the set will be smaller.

The orthogonal character of (deterministic) bias reduction in relation to variance reduction is illustrated in Figure 2 for a 3D case. Using the notation in Figure 2 the real part of the projection $\overline{\mathbf{g}}_s$ of $\overline{\mathbf{g}}$ on the (mapped) signal subspace vector $\mathbf{a}$ determines the bias, and the projection $\overline{\mathbf{g}}_n$ of $\overline{\mathbf{g}}$ on the noise subspace the variance.

Bias will be small if $\overline{\mathbf{g}}$ lives entirely in the noise subspace but also if only (45) is met. If the noise is spatially white, all noise eigenvectors have a common scaling factor σ^2 so in this case there exists no pre-processing $\mathbf{T}$ on the data by which the DOA variance can be reduced while the gradient is confined to the noise subspace.

If the estimator has some (deterministic) bias, then $\overline{\mathbf{g}}^*\mathbf{e}_s \neq 0$ and some components of $\overline{\mathbf{g}}$ do fall outside the noise subspace. This leaves fewer components of $\overline{\mathbf{g}}$ inside the noise subspace, and therefore a somewhat reduced variance. Hence, as far as the relative direction between $\overline{\mathbf{g}}$ and the (pre-processed) noise subspace is concerned, we indeed can reduce variance, but, if the noise is spatially white, always at the cost of increased bias. And *vice versa*.

Still using the notation in Figure 2, an optimal transformation matrix would align $\mathrm{Im}\left\{\overline{\mathbf{g}}\right\}$ along the signal eigenvector axis (where it does not cause bias but moves maximum power out of the noise subspace), and $\mathrm{Re}\left\{\overline{\mathbf{g}}\right\}$ into orthogonality with said signal eigenvector axis, i.e. inside the noise subspace. Inside the noise subspace $\mathrm{Re}\left\{\overline{\mathbf{g}}\right\}$ would be aligned along the eigenvector with the smallest associated eigenvalue.

A "best" pre-processing matrix $\mathbf{T}$ should exploit this while at the same time be retained in as good a condition as possible. If condition becomes a problem we may apply dimension reduction, i.e. design a $\mathbf{T}$ with fewer rows than columns. All along the lines described in [3] to regain numerical robustness. Another remedy against poor conditioning of $\mathbf{T}$ is diagonal loading.

B. Spatially coloured noise

If some of the pre-amplifiers in the antenna array have higher noise figures than the others the noise covariance matrix will still be diagonal but with unequal diagonal elements σ_k^2. This will also be the case for the array output covariance matrix. Returning to the virtual array subspace notation we would then pick a $\mathbf{T}$ that aligns the eigenvector $\mathbf{e}_{n,k}$ that corresponds to the smallest factor $\frac{\lambda_1\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}$ along the gradient $\overline{\mathbf{g}}_v$. We can do this without generating bias if the gradient is retained inside the noise subspace.

The degree of variance reduction obtainable through this operation will depend on the ratio of the largest $\frac{\lambda_1\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}$ -value to the smallest. See Figures 12 & 13 in the Simulations section for an illustration.

If the external noise is spatially coloured, f. ex. by noise jammers, the sun etc., we have a similar situation. Then some gradient directions (inside the noise subspace) are better than the others in terms of MSE. An optimal pre-processing or mapping matrix should take this into account (spatial filtering). This problem has been treated in [14] where pre-whitening of the data using the inverse of an estimate of the (coloured) noise covariance matrix is used to restore the equal scaling of the different dimensions in the noise subspace.

IX. Mapping design algorithms for best MSE

Taking the MSE perspective into view, and assuming an N large, and $\Delta\mathbf{e}_v = \mathbf{T}^*\mathbf{e}_s - \mathbf{e}_v$ small enough for the Taylor expansion terms (12a) and (12b) to dominate, the used mapping matrix should possess the following properties as well as possible across all directions comprising the processed mapped sector

1) The direction of the gradient $\overline{\mathbf{g}}_v \triangleq \overline{\nabla_{\mathbf{e}_k} V(\theta, \mathbf{e}_v)}$ relative to $\Delta\mathbf{e}_v$, $\mathbf{e}_s$ and $\mathbf{e}_{n,i}$ should be such as to minimize bias and variance in optimal proportions.
2) The columns of the pre-processing matrix $\mathbf{T}$ should span all the dimensions required by (34) in order to avoid variance increase.
3) We must prevent the $\mathbf{T} = \mathbf{0}$ solution in any design algorithm
4) As in (20), the design algorithm should have a structure that facilitates a least square solution.

For large SNR property 1 and 3 are secured through the design algorithm (20), but generally the "best" $\mathbf{T}$ will depend also on the noise so this must be estimated and used in the design criterion.

A. A design algorithm for best mapped MSE

Since both terms in the expression (39) for mapped MSE are quadratic in $\mathbf{T}$ we know a global minimum exists. In designing $\mathbf{T}$, to obtain some numerical robustness we should preferably use more calibration directions $\theta^{(c)}$ than the number m_r of antenna elements and seek a best least square solution.

Assuming this and again replacing $\Delta\mathbf{e}_{v,i}$ with $\Delta\mathbf{a}(\theta^{(i)}) = \mathbf{T}^*\mathbf{a}_r(\theta^{(i)}) - \mathbf{a}_v(\theta^{(i)})$, we re-write (39) as

$$\ddot{V}^2(\theta_o, \mathbf{E}_s) \cdot MSE = 4\,\mathrm{Re}\left\{\bar{g}^*\Delta\mathbf{e}^{(det)}_{s,(bias)}\right\}^2 \quad (46a)$$

$$+2\sum_{k=1}^{m-d}(\overline{\mathbf{g}}_v^*\mathbf{T}\mathbf{e}_{n,k}\sqrt{\frac{\lambda_1}{N}\cdot\frac{\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}}\cdot$$

$$\sqrt{\frac{\lambda_1}{N}\cdot\frac{\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}}\mathbf{e}_{n,k}^*\mathbf{T}^*\overline{\mathbf{g}}_v) = \quad (46b)$$

$$\left|2\,\mathrm{Re}\left\{\bar{g}^*\Delta\mathbf{a}(\theta^{(i)})\right\}\right|^2 + \quad (46c)$$

$$2\sum_{k=1}^{m-d}\left|\overline{\mathbf{g}}^*\mathbf{T}^*\mathbf{e}_{n,k}\sqrt{\frac{\lambda_1}{N}\cdot\frac{\sigma_k^2}{(\sigma_k^2-\lambda_1)^2}}\right|^2 \quad (46d)$$

Here the terms (46a) and (46c) represents $bias^2$, (46b) and (46d) variance. Our problem now is to find a pre-processing $\mathbf{T}$ that minimizes their sum. In doing this we are aware of the facts that the variance cannot be reduced below the CRB, and that the accuracy of (46a) in describing the bias is limited by the size of the rest terms (12c) - (12f).

We now return to the application of array mapping over a sector comprised by the N_{cal} calibration directions $\theta^{(i)}$. As mentioned before this is a typical case where (i), the deterministic errors $\Delta\mathbf{e}^{(\mathrm{det})}_{v,(bias)}$ are controllable, and (ii), the variance, due to a $\mathbf{T}$ that due to the sector mapping spans the wrong subspace, easily gets much larger than the CRB.

In the calibration process we collect the N_{cal} response vectors $\mathbf{a}_r(\theta^{(i)})$, presumably at good SNR, and calculate the corresponding virtual vectors $\mathbf{a}_v(\theta^{(i)})$ analytically. For any given $\mathbf{T}$ the mapping errors $\Delta\mathbf{a}(\theta^{(i)}) = \mathbf{T}^*\mathbf{a}_r(\theta^{(i)}) - \mathbf{a}_v(\theta^{(i)})$ can now also be calculated.

To find a mapping matrix $\mathbf{T}$ that, in a DOA error minimizing sense, is a best least square compromise for all calibration response vectors within this sector, and at the same time preventing the $\mathbf{T} = \mathbf{0}$ solution, we thus propose the following design algorithm

$$\mathbf{T} = \arg\min_{\mathbf{T}}\left\{(1-k)\left\|\Delta\mathbf{A}(\theta^{(c)})\right\|_F^2 + \right. \quad (47a)$$

$$k\left[\sum_{i=1}^{N_{cal}}\frac{1}{\ddot{V}^2(\theta^{(i)})}\left|2\,\mathrm{Re}\left\{\overline{\mathbf{g}}_{v,i}^*\mathbf{T}^*\mathbf{a}_r(\theta^{(i)})\right\}\right|^2 + \right. \quad (47b)$$

$$\sum_{i=1}^{N_{cal}}\frac{1}{\ddot{V}^2(\theta^{(i)})}\sum_{j=1}^{m_r-1}\left|\overline{\mathbf{g}}_{v,i}^*\mathbf{T}^*\mathbf{e}_{n,i,j}(\theta^{(i)})\cdot\right.$$

$$\left.\left.\left.\sqrt{\frac{\lambda_{1,i}}{N}\cdot\frac{\sigma_{i,j}^2}{(\sigma_{i,j}^2-\lambda_{1,i})^2}}\right|^2\right]\right\} \quad (47c)$$

where F denotes the Frobenius norm and $\mathbf{g}_{v,i} \triangleq \mathbf{g}_v(\theta^{(i)})$ are the N_{cal} (calibration) gradient vectors.

The first term (47a) minimizes the manifold mapping error $\Delta\mathbf{A}(\theta^{(c)})$ and prevents the $\mathbf{T} = \mathbf{0}$ solution, the second term (47b) advocates minimum DOA bias, and the third (47c) minimum DOA variance.

In analogy with (20), we expect the weighting constant k to be close to 1 to enforce low MSE. We also expect the optimal k to be slightly smaller for lower SNR since then control over the mapping error becomes more important. Furthermore, the balance between correcting the deterministic or stochastic parts of the MSE is set by the root expression, i.e. essentially SNR. We see that if SNR or N is large, the variance part (47c) disappears and (47) reduces to (20).

In (47b) we have again used that $\mathbf{a}_r(\theta^{(i)})$ and $\mathbf{e}_s(\theta^{(i)})$ span the same subspace, and also, from Theorem 1, that $\overline{\mathbf{g}}_{v,i}^*\mathbf{a}_v(\theta^{(i)}) = 0, \forall i$.

Finally we observe that the space $\Re\left\{\Delta\mathbf{a}(\theta^{(i)})\right\} \approx \Re\left\{\Delta\mathbf{e}^{(det)}_{s,(bias)}(\theta^{(i)})\right\} \perp \mathbf{e}_{n,j}(\theta^{(i)}), \forall i$, so again we see the conflict previously illustrated by the pair of equations (41) and (43): No $\mathbf{T} \neq \mathbf{0}$ exist for which both the bias (39a) and the variance (39b) can be put to zero. However, as long as the linear Taylor expansion holds, a mapping matrix according to (47) is expected to minimize DOA MSE across the sector.

Since the variance part of the MSE cannot be reduced below the CRB the amount of improvement that can be expected by adding the last term (47c) depends on how close to the CRB the variance of the estimates already get without use of this last term.

Furthermore, to strong a linear dependence among the calibration response vectors will tend to cause numerical instability in the solution for $\mathbf{T}$ and thus extra variance. In such cases the resulting variance may substantially exceed the CRB and the last term nevertheless be effective.

B. Calibration sector width

To solve (47) for $\mathbf{T}$ we need at least m_r linearly independent column vectors in $\mathbf{A}_r(\theta^{(i)})$ where m_r is the number of antenna elements in the real array. This is decisive since in the solution for $\mathbf{T}$ we need to invert the matrix $\mathbf{A}_r(\theta^{(i)})$.

In addition, as mentioned before the zero bias condition will tend to further restrict $\Re\{\mathbf{\Pi}_T\}$ so any linear dependence among the calibration response vectors is important to control.

For the UCA to ULA mapping application this problem is evident:

The maximum number N_{rv} of (almost) independent response vectors inside a certain sector of width S_w can be approximated by $\frac{S_w}{B_w}$ where B_w is the array beamwidth. We should have $N_{rv} \geq m_r$. Strict inequality is to prefer since it leads to a numerically more robust least square solution for the best mapping matrix.

Taking a uniform $\lambda/2$ spaced circular array UCA as an example we roughly have $B_w = \lambda/(m_r\frac{\lambda}{2}\frac{1}{\pi}) = \frac{2\pi}{m_r}$. The above inequality then becomes $N_{rv} = S_w/B_w = S_w/(\frac{2\pi}{m_r}) \geq m_r$ which yields $S_w > 2\pi, \forall m_r$.

Since for a good circular to linear match this is impossibly wide we must increase the element spacing. In [4] it is shown

that up to 4λ is quite feasible for an UCA, and this would yield $B_w = \frac{\pi}{4m_r}$ and $S_w > \pi/4$ respectively, see Figure 3.

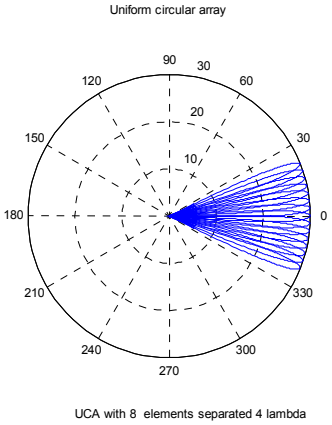

Fig. 3. With 4 λ of UCA element separation a 45^o sector can barely encompass m_r linearly independent response vectors, where m_r is the number of antenna elements.

For illustration, in figure 4 the condition number[5] of an 8 element UCA response vector matrix A_r is plotted against element spacing in wavelengths for two sector widths, 45^o and 30^o. The condition number increase towards smaller element separations is evident.

Allowing condition numbers up to 10^4 (above which further calibration response vectors contribute little to the solution for $\mathbf{T}$) figure 4 indicates a useful relative bandwidth of one octave for a uniform circular array. As a compromise between linear dependence (and a complicated calibration procedure) on one hand, and mapping errors on the other, 30^o will be used in the Simulations section.

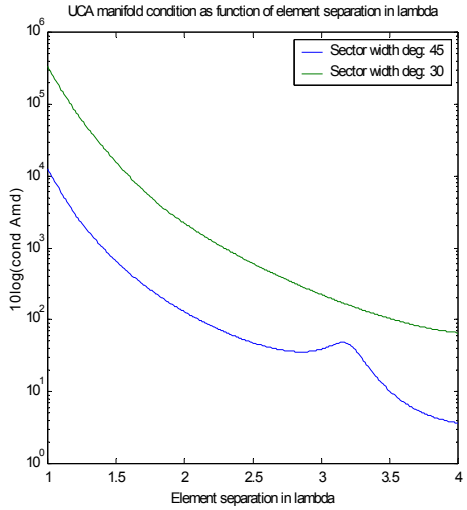

Fig. 4. Condition number of the response vector matrix $\mathbf{A}_r$ as function of element spacing, i.e. array size, for an 8 element uniform circular array, UCA. The bump is due to a geometrical resonance effect.

Hence we see that if $\mathbf{T}$ retains the dimension, i.e. is quadratic and full rank, then sector mapping always has to be performed with some linear dependence among the used calibration response vectors, the more the narrower the sector. This can lead to increased variance and may enforce dimension reduction so that fewer than m_r calibration directions can be used. This is verified in figures 16 and 17 in the Simulations section.

If dimension reduction[6] is used to remove those dimensions in $\mathbf{A}_r$ that correspond to the smallest singular values, the condition numbers can be improved and lower frequencies used for a given array. This plus the fact that less inherent bias has to be countered if the frequency is reduced, so that the weighting constant k in (47) can be given a much smaller value, relieves the problem significantly.

Simulations show that we can reduce the operating frequency down below $\lambda/2$ element spacing if these measures are taken. Hence an operating bandwidth in excess of 3 octaves can be expected for a given $m_r = 8$ UCA.

Dimension reduction was treated extensively in [3] to which the interested reader is referred.

C. Spanning the space of the derivatives

While overlapping beams cause linear dependence and numerical problems, the CRB retainment condition (34) is automatically fulfilled by precisely said overlap. Using a dense grid $\theta^{(c)}$ of calibration directions we see that the corresponding set of response vectors $\mathbf{a}(\theta^{(c)})$ spans a space that also includes the derivatives $\mathbf{D}(\theta^{(c)})$ in (34).

This is seen by approximating the derivatives with the differentials.

$$\mathbf{d}(\theta^{(i)}) = \frac{\mathbf{a}(\theta^{(i+1)}) - \mathbf{a}(\theta^{(i)})}{(\theta^{(i+1)}) - (\theta^{(i)})} \tag{48}$$

The fulfillment of (34) generally requires the DOA to be known beforehand, at least to within a beamwidth. In signal surveillance this knowledge is not at hand and we have to rely on the LS compromise of sector mapping, at least in a first step. The resulting deviation from (34) will cause a variance increase in the mapped estimates. Reducing this increase as much as possible is one prime purpose of the proposed design algorithm (47).

D. The least square solution

The minimization problem (47) can be solved in many ways. One approach is to apply the vec operator to both terms inside each norm and solve for a vectorized version $\mathbf{t}$ of $\mathbf{T}$. In doing so we get a least square solution of the type

$$\mathbf{t} = \left[\begin{array}{l} vec(\mathrm{Re}\,\{\mathbf{T}\}) \\ vec(\mathrm{Im}\,\{\mathbf{T}\}) \end{array} = \mathbf{M}^{\dagger}\mathbf{m} \right] \tag{49}$$

with obvious definitions for $\mathbf{M}$ and $\mathbf{m}$. Solving it involves calculating the pseudo-inverse of the large matrix $\mathbf{M}$ of size $[2N_{cal} \times m_r + N_{cal} + 2N_{cal} \times (m_r - 1)] \times [2m_v m_r]$. If, as a typical example, $m_v = m_r = 8$ and $N_{cal} = 15$, the size of the matrix $\mathbf{M}$ is 465×128.

Omitting the Re operator in (47b) reduces the size of $\mathbf{M}$ to $[N_{cal} \times m_r + N_{cal} + 2N_{cal} \times (m_r - 1)] \times [m_v m_r]$ or $345 \times$

[5] The largest singular value divided by the smallest

[6] $\mathbf{T}$ has dimension $n \times m$ with $n < m$

64 in the example, but on the other hand enforces complex computations. However, this measure has the advantage of minimizing the modulus of $\overline{\mathbf{g}}_v^*(\theta^{(i)})\mathbf{T}\mathbf{A}_r(\theta^{(i)})$ and not just the real part, a feature that shows up in the simulations as lower bias during high SNR conditions.

This observation is explains as a reduced magnitude also of the residual terms (12c) - (12f) in the Taylor expansion, but has the accompanying disadvantage that some of the available degrees of freedom in $\mathbf{T}$ now are used to minimize also $\mathrm{Im}\left\{\overline{\mathbf{g}}^*(\theta^{(i)})\mathbf{T}\mathbf{a}_r(\theta^{(i)})\right\}$ which as such does not improve MSE. Conversely, as already pointed out, this increases the DOA error variance somewhat, see Figure 5 in the Simulations section for an illustration.

It should also be noted that solving (47) and inverting the rather large matrix $\mathbf{M}$ is part of the calibration process. During real time operation of a typical DF system all mapping matrices $\mathbf{T}$ would be calculated in advance and will therefore not slow down the real time generation of DOA estimates.

Finally note that in the above derivation of the MSE expression (39) no restrictions on the cost function V were made other than the existence of the necessary derivatives. Hence we conclude the proposed design algorithm (47) to be applicable to all DOA estimators that are based on cost functions of this class.

X. Dimension reduction and the span of T

A natural way to enforce some of the abovementioned CRB -property (34) into the transformation matrix $\mathbf{T}$ is to extend each term in the (20) sum with a new term that penalizes lack of said property (34), i.e. lack of ability in $\mathbf{T}$ to span the required subspace. This measure is motivated when dimension reduction is involved and the transformation matrix $\mathbf{T}$ therefore is much below full rank.

The design algorithm (20) then extends into

$$\mathbf{T}_{opt} = \arg\min_{\mathbf{T},\mathbf{M}} \left\{ (1-k) \left\| \Delta\mathbf{A}(\theta^{(c)}) \right\|_F^2 + \right. \quad (50a)$$

$$k(1-\lambda) \cdot \sum_{i=1}^{N_{cal}} \left| \mathrm{Re}\left\{ \overline{\mathbf{g}_v}^{(i)*} \Delta\mathbf{A}(\theta^{(c)}) \right\} \right|^2 + \quad (50b)$$

$$\left. \lambda \left\| \mathbf{T} - \mathbf{M}\mathbf{V} \right\|_F^2 \right\} \quad (50c)$$

where we see that if the weighting constant $\lambda = 0$ then (50) reduces to the earlier design algorithm (20).

The new penalty term involves a matching matrix $\mathbf{M}$, and the matrix $\mathbf{V}$ whose columns form an orthogonal basis for the space spanned by the right membrum in (34) at the true DOA θ_o. This basis can be obtained by performing a singular value decomposition on the matrix in said right membrum, [19].

To enforce some of the property (34), we take $\lambda > 0$, where the optimum value will depend on the amount of deterministic bias as well as on SNR and N, and above all, on the amount of space $\mathbb{C}^m$ outside $\Re\{\mathbf{V}\}$.

The design algorithm (50) is most effective in dimension reduction applications where $\Re\{\mathbf{V}\} \subset \mathbb{C}^m$. In this case an initially (for $\lambda = 0$) high or full rank matrix $\mathbf{T}$ can be "pruned" to span but the necessary dimensions.

For dimension retaining pre-processing, f. ex. array mapping over sectors, $\Re\{\mathbf{V}\} \approx \mathbb{C}^m$ and there will always exist an $\mathbf{M}$ that puts the last term (50c) to zero, regardless of $\mathbf{T}$. This renders (50) less useful in such cases and corresponds to trivial fulfillment of (34).

For (50) to work, the true DOA θ_o should in principle be known and if this is not the case a successive narrowing of the width of the mapped sector is necessary. If this operation is included and combined with dimension reduction, bias can be minimized and the CRB approached. See also [4], pp. 107-109 for verifying simulations.

XI. Simulations

In the simulations to follow the MSE of the DOAs obtained from a mapped circular array will be studied. The mapping is from a (real) 8 element planar uniform circular array UCA, spaced at 4λ to highlight bias, onto an 8 element (virtual) $\lambda/2$ ULA. Both arrays are confined to the same plane with the phase centers symmetrically placed. This is a typical case where the deterministic errors can get relatively large, are known (through a calibration process), and are *controllable* (by the design of $\mathbf{T}$).

Using a set of $N_{cal} = 15$ calibrated directions uniformly spread across a 30^o wide broadside azimuthal sector, a mapping (transformation) matrix $\mathbf{T}$ is constructed according to the proposed algorithm (47). This corresponds to a typical signal surveillance application where omni-directionality (the UCA) and bandwidth (the 4λ spacing), can be combined with estimator processing speed (ULA root MUSIC or root WSF). It is however also a scenario where DOA bias is large and can dominate over STD if no countermeasures are taken.

In the construction of the mapping matrix it is assumed that the calibration process is carried out against one (moving) calibration emitter at good SNR. The two sets of response vectors, $\mathbf{A}_r(\theta^{(c)})$ and $\mathbf{A}_v(\theta^{(c)})$ needed in (47), were hence calculated without errors, as were the noise subspace eigenvectors in (47c).

Furthermore, to avoid extra bias due to coloring of the measurement noise by (the usually non-unitary) $\mathbf{T}$, the signal eigenvectors were obtained from an eigenvalue decomposition of the (real) array output data covariance matrix. Thereafter said signal eigenvectors were mapped by $\mathbf{T}$, i.e. $\mathbf{e}_v = \mathbf{T}\mathbf{e}_s$, and the orthogonal noise subspace needed for root MUSIC obtained from a subsequent singular value decomposition on this mapped signal subspace.

Said eigenvalue decomposition of the real array output also yielded estimates of the noise eigenvectors $\mathbf{e}_{n,i}$ as well as the eigenvalues λ_1 and σ_k^2 needed in (47c).

Measurement noise is spatially white and Gaussian circularly symmetric. $N = 100$ snapshots and 400 Monte Carlo runs per azimuth are used to calculate the DOA means, biases and variances.

The N_{cal} Hessians $\ddot{V}(\theta^{(i)})$ vary only a few percent around the value 100 so to facilitate comparisons between the orthogonality enforcing algorithm (20) and the MSE minimizing (47), a (common) scaling factor of 100 was used in k for the latter algorithm.

In the simulations 4 different mapping matrices are used:

1) Pure LS manifold match, i.e. $k = 0$ in (22) and (47) respectively,
2) Pure bias reduction at high SNR, i.e. $k = 0.99$ in (22),
3) MSE minimization according to (47) at high SNR with $k = 0.99$.
4) MSE minimization according to (47) at moderate SNR with $k = 0.95$.

The lobewidth of the UCA array is about 5^o so with $30/15 = 2^o$ between the calibration directions there is a substantial amount of linear dependence in the calibration response vector matrix $\mathbf{A}_r(\theta^{(c)})$. The condition number, i.e. the ratio of the largest singular value to the smallest, of the resulting $\mathbf{A}_r(\theta^{(c)})$ is of order 10^4.

On the other hand, due to this proximity between the response vectors the (needed) space spanned by the derivatives D in (34) are to some extent spanned by the columns in $\mathbf{A}_r(\theta^{(c)})$.

The main obstacle in this latter respect[7] is the compromise across the 30^o wide sector, which hence prevents the simulated variances from attaining the CRB. Further simulations with smaller sector widths illustrate this effect clearly.

A. Bias reduction design

For later comparison with the MSE minimizing design (47), we start by illustrating the performance of the pure bias reduction design (22). First for a high SNR value, 60 dB, and then for a low, 10 dB. The result for 60 dB, see Figure 5, shows a conspicuous bias reduction factor, exceeding 100 in some azimuths, when the gradient orthogonality criterion is enforced in stead of the pure manifold match design.

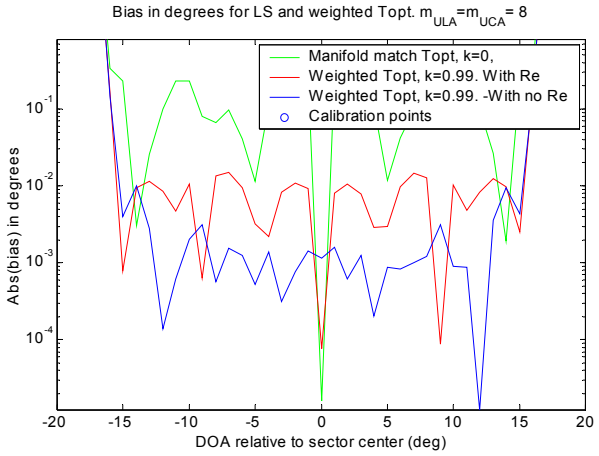

Fig. 5. Absolute value of the bias (deg) as function of azimuth for a single emitter that moves across a 30^o mapped sector. The manifild match design of $\mathbf{T}$ (green curve) generates the largest bias, and the $\mathbf{T}$ that uses all available degrees of freedom to counter bias, the lowest. Hence, the design algorithm (20) without the Re operator the lowest (blue curve).

The difference between the red and blue bias curves in figure 5 illustrates the effect of the higher order terms in the Taylor expansion (12), terms that to some extent are zeroed by enforcing the gradient orthogonality also in the imaginary part. See Appendix 1 for reference.

[7] Remember that the DOA should in principle be known for the requirement (34) to be met.

Also recall that the bias minimizing design criterions (20) and (22) were derived taking only deterministic effects into account. As expected it therefore works best at high $SNRs$. The rapidly deteriorated performance when SNR is reduced is apparent from a comparison with Figure 6 where SNR is reduced to 10 dB. Here we see a clear tendency of the algorithm to assign the emitters a DOA in the center of the mapped sector when noise is increased. Of the three illustrated designs, gradient orthogonality without the real part operator, still is the best.

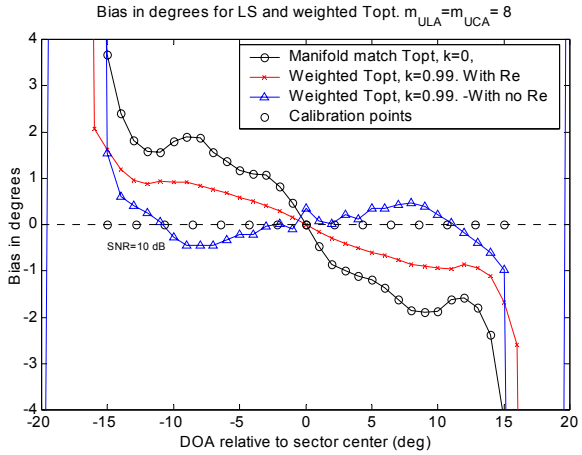

Fig. 6. At low SNR, 10 dB, the bias minimizing criterion (20) works less well. The reason is that it was derived using the pure deterministic errors $\Delta\mathbf{e}_{s,(bias)}^{(det)}$, in our case the mapping errors. The effect of the noise is to produce a net bias towards the center of the mapped sector.

As will be seen in the sequel however, the extension of the design criterion (20) to include also the variance terms appreciably improves performance for low SNRs.

1) Bias reduction and resolution: Figure 7 illustrates the effect on resolution of the various mapping matrix designs. For reference the non-mapped MUSIC UCA spectrum (black line) is also shown.

When $\mathbf{T}$ is forced to transform the signal eigenvector errors $\Delta\mathbf{e}_v$ into orthogonality with the gradient $\overline{\mathbf{g}}_v$ in *both* the real and imaginary part (top curve, i.e. the Re operator in (47b) is omitted), more degrees of freedom in $\mathbf{T}$ are used up to counter DOA bias than in the other two designs. As mentioned before, this leaves fewer degrees of freedom to fulfill the CRB criterion (34) and is consequently expected to result in higher DOA variance and loss of resolution.

Figure 7 clearly shows a loss in resolution for all mapped cases, and also, as lost sharpness in the two peaks, a corresponding increase in variance. The largest loss occurs for the best bias minimizing design, i.e. for (47b) without the real part operator.

As expected, with increasing amount of bias control we get more and more DOA variance, and correspondingly loose more and more resolution. *Bias control is paid for by increased variance and lost resolution*.

B. MSE minimization with SNR 60 dB

To get a first insight into the operation of the proposed design algorithm (47) we start by assuming large SNR, 60 dB, so that

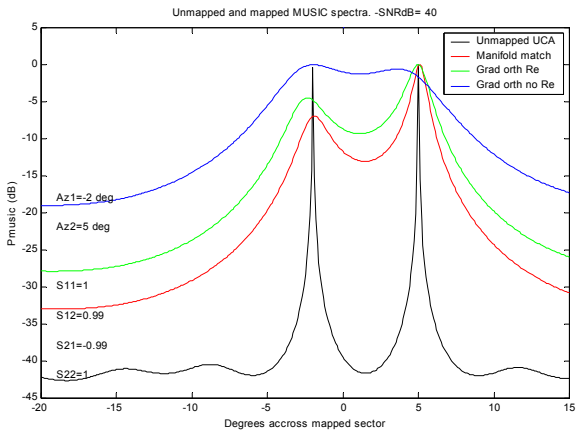

Fig. 7. MUSIC spectra for 2 emitters and various mapping matrix designs. The black line shows unmapped UCA MUSIC for reference. As seen the design that yields the lowest bias, i.e. without the Re operator, also shows the poorest resolution, (blue curve).

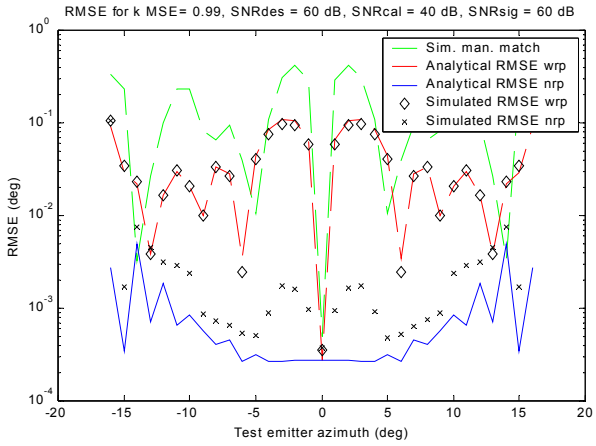

Fig. 8. RMSE for 8 element UCA to ULA mapping when both the design- and signal SNRs are large, 60 dB. The used algorithm (47) now essentially minimises bias, the difference from (20) and figure 6 being the inverse Hessian weghting in (47). Note the good agreement between the analytical (red and blue) and the simulated curves. The design with no Re operator (nrp) is best.

the MSE is dominated entirely by bias. The design algorithm (47) now selectively enforces the orthogonality $\overline{\mathbf{g}}_v \perp \mathbf{T}\mathbf{a}_r$ in the real part, but somewhat differently from algorithm (22) in that each term in the sum now is weighted by the Hessian inverse $\ddot{V}^{-2}(\theta_o^{(i)})$.

As previously mentioned, the resolution of an 8 element $\lambda/2$ ULA does not vary noticeably within $\pm 15^o$ from broadside, so the above weighting with a constant k only has marginal effect on the bias. Hence, at this SNR algorithms (22) and (47) produce mapping matrices with approximately the same bias suppression properties.

For the two alternatives with, (wrp) and with no (nrp), real part operator in (47), Figure 8 shows DOA RMSE, i.e. $\sqrt{MSE}$, when both design SNR and signal SNR are high, 60 dB. This means that in the corresponding design criterion for $\mathbf{T}$ the variance term (47c) is almost zero. The only difference now between the two design criterions (22) and (47) is the inverse Hessian weighting factor $\ddot{V}^{-2}(\theta_o^{(i)})$, so the influence of this factor, though small, can be studied by comparing figures 5 and 8.

Figure 8 also shows the analytically calculated $RMSE$ as non-marked coloured curves (red and blue for the alternatives *with* and *without* the Re operator respectively). These analytical values were calculated using the higher order expression (23) for the bias part, which explains the good agreement with the simulated values.

C. MSE minimization with SNR 6 dB

The 4λ element spacing of the (real) UCA generates slight, but due to the curved array geometry, not full ambiguities. These cause no problem at the high SNR 60 dB but do at the more realistic 6 dB level. To test the proposed design algorithm for $\mathbf{T}$ at SNR 6 dB we therefore reduce the UCA element spacing from 4λ down to 2λ.

Furthermore, the array mapping error $\Delta\mathbf{a}$ is still the same under an SNR decrease down to 6 dB but the estimated signal eigenvectors $\hat{\mathbf{e}}_s^{(i)}$ are no longer unaffected by the noise. The eigenvector errors $\Delta\mathbf{e}_v^{(i)}$ thereby become larger and need arises to keep the size of these vectors under control in order to reduce the higher order terms in the Taylor expansion (12). We therefore expect the optimal weighting factor k in (47) to be slightly smaller than before.

Hence in the simulations we use the value 0.95 instead of 0.99.

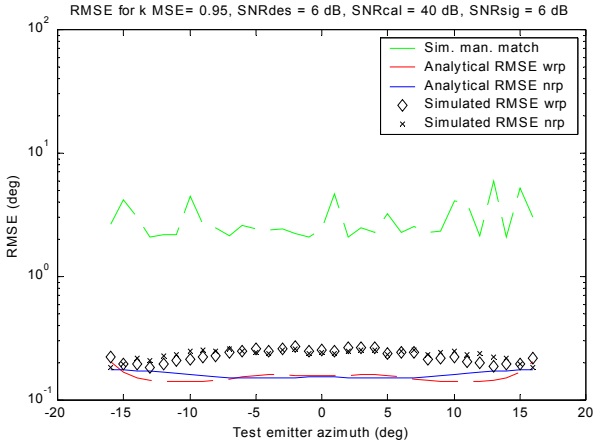

Fig. 9. DOA RMSE for 2λ UCA element spacing and SNR 6 dB. The proposed design of the mapping matrix $\mathbf{T}$ clearly outperforms the manifold match approach.

Figure 9 shows the $DOA\ RMSE$ when (47) is used to calculate the mapping matrix, with 6 dB as design- and signal SNR, and the weighting factor $k = 0.95$.

As seen, at SNR 6 dB, the design algorithm (47) clearly outperforms the manifold match approach (top green curve) as well as the (22) design. The reason is that the latter was derived presuming only deterministic errors, which at SNR 40 dB and up, may be a good approximation, but at 6 dB is not.

For the 6 dB SNR case the standard deviations STD of the DOA estimates are displayed in Figure 10. The STD values

follow reasonably well the corresponding CRB curves. The flat coinciding CRBs are expected because of the near omni-directionality of the non-mapped UCA and the full rank $\mathbf{T}$.

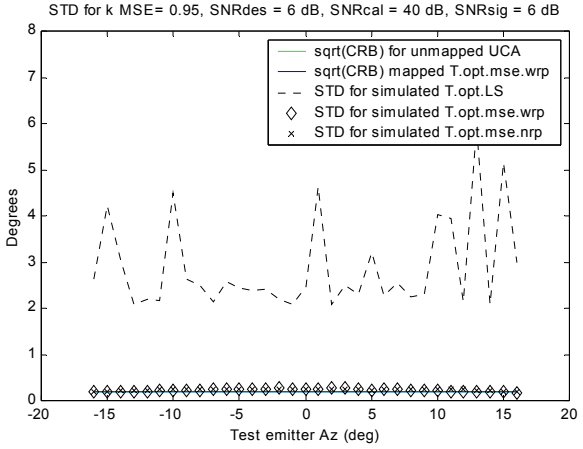

Fig. 10. The standard deviation part of the MSE in figure 9. Both the STD- and bias part are clearly reduced by the proposed design of $\mathbf{T}$ relative to the pure manifold match design (top dashed curve). Relative to the bias minimization design (22) a conspicuous robustness against measurement noise has been achieved.

As a comparison the STD values for the $k = 0$ case are also shown (top dashed curve), i.e. when $\mathbf{T}$ is designed for best array manifold match only. In the simulated case both the two mappings, with- and without the Re operator in (47c), produce mapped response vectors with derivatives that fulfill the CRB criterion (34) much better than does the manifold match designed $\mathbf{T}$.

When it comes to implementation in a practical signal surveillance system such as SESAM, the new transformation design (47) seems to be far better suited than both the manifold match- and the bias minimization (22) designs.

D. Spatially coloured noise

We now assume an unintentional change in the noise figure of one of the pre-amplifiers in the array. This case is not uncommon in a real direction finding system, it can be caused by increased attenuation in one of the antenna elements or the associated cabling, followed by an automatic gain increase in the corresponding receiver channel.

In this case the eigenvalues σ_k^2 of the noise subspace of the data covariance matrix will no longer be equal. Inside the now unsymmetrical noise subspace there exist directions for the gradient $\overline{\mathbf{g}}_v$ that are better in terms of variance than other directions.

In a practical DF system the unequal input noise factors are most easily detected and assessed when no external emitters illuminate the array. An $m_r \times m_r$ noise covariance matrix $\mathbf{Q}_{noise}$ can then be measured, stored, and, as a good approximation, be used to construct the variance terms (47c).

The simulation results displayed in Figures 11 and 12 involve an array with a 10 dB noise factor increase in one of the 8 receiver channels. Original signal SNR is 10 dB. All other parameters are the same as those in Figures 9 and 10.

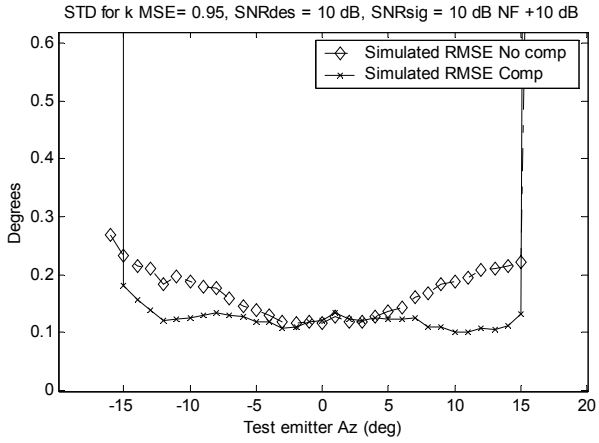

Fig. 11. Standard deviation for the DOA estimates when knowledge about a 10 dB noise factor increase in one of the receiver channels is used to construct the mapping matrix $\mathbf{T}$, (x-marked curve). For reference the diamond marked curve shows the same case but with no use of this knowledge.

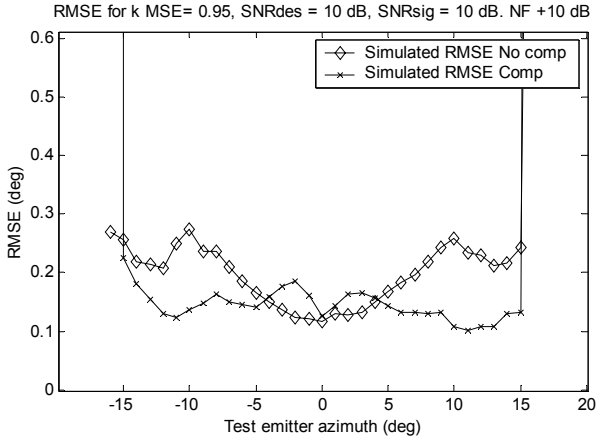

Fig. 12. RMSE for the DOA estimates when knowledge about a 10 dB noise factor increase in one of the receiver channels is used to construct the mapping matrix $\mathbf{T}$. For reference the diamond marked curve shows the same case but with no use of this knowledge.

The resulting overall noise level increase is visible when comparing Figures 11 and 10. In the latter figure, for the x- and diamond marked curves knowledge of the faulty receiver channel was not used when constructing the mapping matrix.

An important advantage of the suggested design algorithm (47) is that it allows us to exploit this knowledge if it is at hand and $\mathbf{Q}_{noise}$ is available. To do this, in the simulations the approximate noise eigenvectors $\mathbf{e}_{n,i}$, as well as the corresponding approximate eigenvalues $\lambda_{1.i}$ and $\sigma_{i,j}^2$ needed in (47c), were obtained from an eigenvalue decompositions of

$$\hat{\mathbf{R}}_i = \mathbf{a}(\theta^{(i)})\mathbf{a}^*(\theta^{(i)}) + \mathbf{Q}_{noise},\ i = 1, ..., N_{cal} \qquad (51)$$

The interpretation of the bias reduction visible when comparing Figures 11 and 12, is that the new $\mathbf{T}$ that was obtained through (47) and (51) rotated the mapped signal subspace further away from the gradient. According to Figure 2 this is expected to generate more DOA estimate variance, but this was

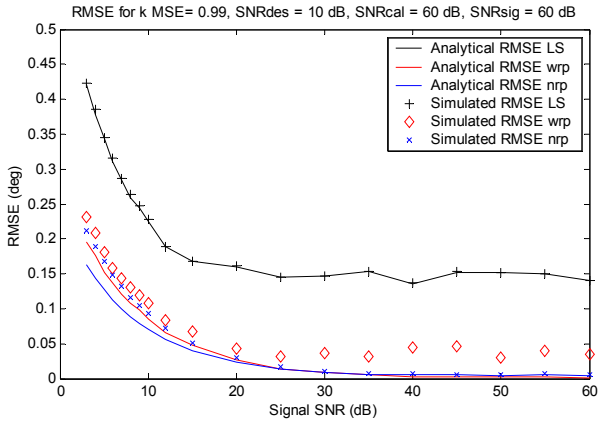

Fig. 13. RMSE as function of SNR when the sector mapping is designed according to the proposed algorithm (47).

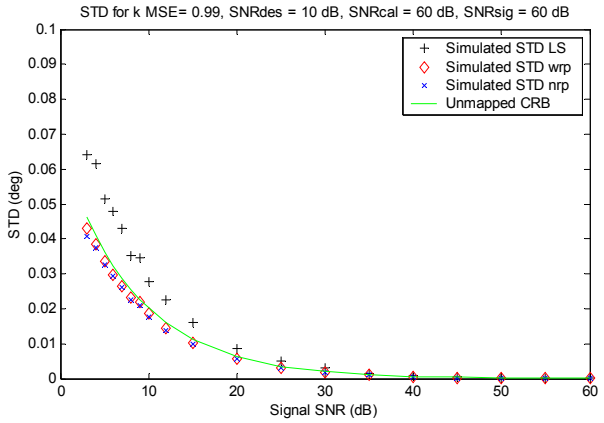

Fig. 14. DOA estimate standard deviation as function of SNR when the proposed algorithm (47) is used to construct the mapping.

avoided by rotating the (mapped) noise subspace around the signal subspace axis into an orientation that put the noise eigenvector with the smallest scaling factor along the gradient.

This illustrates a prominent feature of the proposed design (47): It can minimize the influence of various data errors, deterministic, stochastic as well as measurement setup errors, on the DOA estimates by rotating the mapped signal- and noise subspaces around the gradient into optimal positions.

Note: Another way to deal with coloured background noise is to pre-whiten, i.e. to pre-multiply the real array output data with $\mathbf{Q}_{noise}^{-1/2}$, see [14].

E. Dependence on SNR

For completeness the performance of the proposed design algorithm (47) is also verified as function of signal SNR, Figures 13, and 14. In doing so we retain all other simulation parameters, i.e. the "design" SNR is still 10 dB.

Furthermore, in order use signal SNR as a parameter, for each SNR the RMSE and STD values at a 1^o grid from -15^o to $+15^o$ were calculated. Thereafter a mean was formed across the sector for the particular SNR. The $bias^2$ and var values were first averaged. Thereafter $RMSE = bias^2 + var$ and $STD = \sqrt{var}$ were calculated and plotted. Since the grid goes from -15^o to $+15^o$ some (minor) end effects are included. Restricting the averaging to the center portion of the sector would yield smaller errors.

Clearly, for increasing signal SNR the MSE will be more and dominated by bias. This is also analytically visible in the second term (47c) which goes to zero when SNR goes to infinity. We therefore anticipate the results from the designs (20) and (47) to converge toward high SNRs.

As seen from Figures 13, and 14 these anticipations are confirmed by the simulations. The differences are most evident for low SNRs.

Figure 13 clearly illustrates the bias reduction capability of both the (20), (triangle- and star marked curves), and (47) algorithms (diamond- and x-marked curves). The (47) algorithm is best, especially at low SNRs.

Figure 14 shows the corresponding STD curves. By comparing with Figure 13 it is seen that the superior RMSE performance of the (47) algorithm mainly is due to lower STD. This is explained by the corresponding mapping matrices being designed to minimize RMSE at 10 dB SNR, and hence, at this and lower SNRs, yielding a more favorable orientation of the gradient $\bar{\mathbf{g}}_v$ relative to the prevailing noise- and signal subspaces.

F. Dependence on sector width

A narrowing mapped sector reduces the effect of $\mathbf{T}$ being a sector compromise. According to the results in [3] a sector width smaller that the beamwidth of the array is needed in order to design a $\mathbf{T}$ that fully preserves variance. However, on the other hand a narrow sector generated more calibration lobe overlap and thus more linear dependence in the matrix $\mathbf{A}\left(\theta^{(c)}\right)\mathbf{A}^*\left(\theta^{(c)}\right)$ that has to be inverted in order to solve for $\mathbf{T}$.

The performance of the mapping as function of sector width is therefore also of interest.

Figure 15 shows standard deviation of the mapped DOA estimates as function of mapped sector width. It is seen that the CRB is not approached until the sector width approaches one beamwidth (about 5^0). The same thing is observed in the RMSE plot, Figure 16.

As expected, both bias and variance get small when the sector narrows. The former because of geometrical reasons and the latter because the required condition (34) to attain the CRB of the real array, now is met. A comparison between the RMSE- and STD -values reveal that noise effects now form the main constituent of the total errors and not bias.

G. Lower frequency limit for the SESAM system

The SESAM surveillance and direction finding system uses a circular antenna array of 10 m radius and 8 elements. The nominal frequency range is 3-30 MHz corresponding to element separations of 0.08λ to 0.8λ.

Using the above schemes for bias- and MSE minimization it was show through simulations, Hyberg [1], that even for a

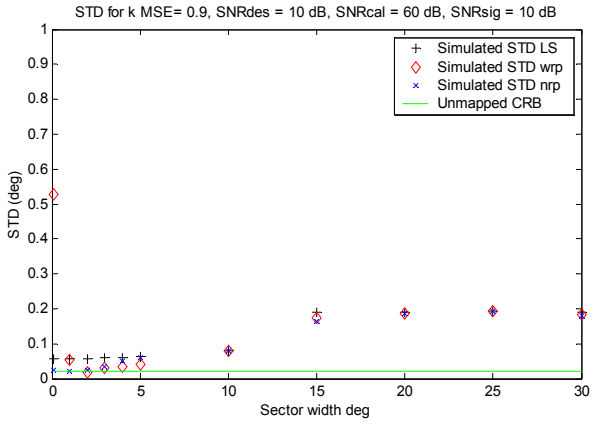

Fig. 15. Standard deviation STD as function of mapped sector width .

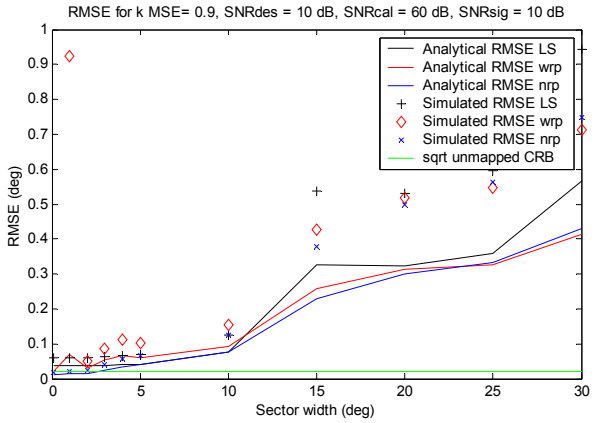

Fig. 16. RMSE of the DOA estimates as function of sector width.

non-randomized circular array, element spacings up to 4λ can be used for high SNRs. This corresponds to 150 MHz, a conspicuous extension in the upper frequency range, or put another way, improvement in DOA accuracy.

If we, for a given array, in stead go down in frequency, both bias problems and mapping errors, as well as increase in variance due to mapping errors, are reduced. This is because (i) bias tends to scale with frequency and (ii), the increased lobewidth that makes the CRB attainment requirement (34) easier to fulfill. However, as shown earlier the increased lobewidth causes more linear dependence among the calibration response vectors from which the mapping matrix is constructed, and subsequent numerical problems. The lower frequency limit for a given array therefore depends on how well we can avoid and cope with, this increased linear dependence.

To handle lower frequencies, in analogy with the previous subsection the measures to take are the following

1) Dimension reduction, $m_v < m_r$
2) Corresponding reduction in N_{cal}
3) Reduction of sector width

The first measure is the most important one. The second and last measures may not be needed, but simulations show they do improve the results.

Numerous simulations have indicated that 0.1λ element separation still yields usable results, provided a dimension reduction down to 2 elements in the virtual array.

It should also be noted that despite the high input impedance active antenna elements of the SESAM array, at 0.1λ element separation, *coupling effects* are expected to become non-negligible. The array models (2) - (4) used in the simulations may therefore have to be modified to produce relevant results in this case. Coming field tests with the SESAM system will clarify this problem.

Finally, note that at this low frequency the (parallel) Adcock algorithm of the SESAM system is expected to produce almost equally good DOA estimates.

XII. Conclusions

In this paper we have analyzed both deterministic and random DOA estimation errors that arise from various types of array pre-processing, especially array mapping. Via a Taylor expansion of the derivative of the DOA estimator cost function, its *gradient with respect to the signal eigenvectors* was identified as an entity by which both deterministic (bias) and random (variance) DOA errors could be quantified, analyzed and minimized.

Using this gradient, expressions for DOA bias, DOA error variance and DOA MSE under pre-processing were derived. Furthermore design algorithms for the preprocessing matrix were formulated by which the resulting DOA bias and MSE could be minimized. A slightly modified version was also given with the property of further reducing the higher order terms and thus the rest bias.

DOA bias was reduced not by minimizing the mapping errors, but in stead by rotating these errors so that they get orthogonal to said gradient and this way no longer affect the value of the cost function.

DOA variance was minimized by rotating the mapped noise subspace into an optimal orientation relative to the same gradient. The usefulness of this was illustrated by handling a situation with excess noise in one receiver channel.

DOA MSE was minimized by rotating the mapped signal- and noise subspaces into optimal directions relative to the gradient.

It was shown that while bias due to imperfect mapping can be dramatically reduced in high SNR scenarios, variance reduction also depends on the degree of asymmetry in the noise subspace. Through a geometrical interpretation it was also shown that there exist no pre-processing by which both bias and variance generally can be minimized simultaneously.

Finally, in relation to the earlier described pure bias minimizing design, Hyberg [1], by adding the variance terms to the design criterion for the mapping matrix, a much larger robustness in low and moderate SNR scenarios was achieved.

The expressions and design algorithms were verified by simulations. So were their consistency with earlier results concerning the attainment of the CRB as well as the bandwidth requirements of the SESAM 3-30 MHz signal surveillance system.

APPENDIX

For the MUSIC estimator, which is used in the Simulations section, we will here derive the needed gradients and Hessians. The MUSIC cost function is a real scalar that depends on the scalar θ and also on the complex eigenvectors comprising the signal subspace, i.e. the column vectors of the matrix $\mathbf{E}_s$.

In deriving the gradient and the Hessian we use the conventions by Brandwood [17] :

Let $J(\mathbf{e})$ be a function of the complex vector $\mathbf{e}$ and its conjugate. Further, let e_k be the kth element of $\mathbf{e}$ and let x_k and y_k be the real and imaginary parts of e_k, respectively. Then the kth component of the gradient vector is defined as

$$[\nabla_e J(\mathbf{e})]_k = \frac{1}{2}(\frac{\partial J(\mathbf{e})}{\partial x_k} - j\frac{\partial J(\mathbf{e})}{\partial y_k}) \tag{52}$$

With this convention, the differential of $J(\mathbf{e})$ will be $dJ(\mathbf{e}) = 2\,\mathrm{Re}\{[\overline{\nabla_{\mathbf{e}} J(\mathbf{e})}]^* d\mathbf{e}\}$. Specifically[8] for $J(\mathbf{e}) = \mathbf{e}^*\mathbf{R}\mathbf{e}$ where $\mathbf{R}$ is any Hermitian matrix independent of $\mathbf{e}$, we get $\nabla_e J(\mathbf{e}) = \mathbf{R}\mathbf{e}$ and $dJ(\mathbf{e}) = 2\,\mathrm{Re}\{\mathbf{e}^*\mathbf{R}\; d\mathbf{e}\}$.

For MUSIC the commonly used cost function is

$$V(\theta, \mathbf{E}_s) = \frac{\mathbf{a}^*(\theta)\mathbf{a}(\theta)}{\mathbf{a}^*(\theta)(\mathbf{I} - \mathbf{E}_s\mathbf{E}_s^*)\mathbf{a}(\theta)} \tag{53}$$

which is a scalar function of the scalar θ and the complex column vectors in $\mathbf{E}_s$.

For analysis of bias and variance we have used the derivative w.r.t. θ of the denominator:

$$\dot{V}(\theta, \mathbf{E}_s) = 2\,\mathrm{Re}\left\{\mathbf{a}^*(\theta)(\mathbf{I} - \mathbf{E}_s\mathbf{E}_s^*)\dot{\mathbf{a}}(\theta)\right\} \tag{54}$$

The gradient $\nabla_{\mathbf{e}_i}\dot{V}(\theta, \mathbf{E}_s)$ of $\dot{V}(\theta, \mathbf{E}_s)$ with respect to the i^{th} eigenvector $\mathbf{e}_i$ in the signal subspace $\mathbf{E}_s$ becomes

$$\nabla_{\mathbf{e}_i}\dot{V}(\theta, \mathbf{E}_s) \triangleq \overline{\mathbf{g}}^{(i)} = -\overline{\dot{\mathbf{a}}}(\theta)\mathbf{e}_i^*\mathbf{a}(\theta) - \overline{\mathbf{a}}(\theta)\mathbf{e}_i^*\dot{\mathbf{a}}(\theta) \tag{55}$$

We also need the second derivative w.r.t. θ

$$\ddot{V}(\theta, \mathbf{E}_s) = 2\dot{\mathbf{a}}^*(\theta)(\mathbf{I} - \mathbf{E}_s\mathbf{E}_s^*)\dot{\mathbf{a}}(\theta) \tag{56}$$

According to the above convention the gradient of (56) becomes

$$\nabla_{\mathbf{e}_i}\ddot{V}(\theta, \mathbf{E}_s) = -\overline{\ddot{\mathbf{a}}}(\theta)\mathbf{e}_i^*\mathbf{a}(\theta) - \overline{\mathbf{a}}(\theta)\mathbf{e}_i^*\ddot{\mathbf{a}}(\theta) - 2\overline{\dot{\mathbf{a}}}(\theta)\mathbf{e}_i^*\dot{\mathbf{a}}(\theta) \tag{57}$$

Finally we need the higher order $m_v \times m_v$ Hessian matrices which become, respectively

$$\mathbf{H}1 = -(\mathbf{a}(\theta)\dot{\mathbf{a}}^*(\theta) - \dot{\mathbf{a}}(\theta)\overline{\mathbf{a}}(\theta)\dot{\mathbf{a}}(\theta)) \tag{58}$$

$$\mathbf{H}2 = \mathbf{0} \tag{59}$$

The entities (56), (57), (58) and (59) are needed to quantify the higher order term in the Taylor expansion (12) and thus the bias rest terms.

1) Bias minimization: Assuming one emitter only and taking only the first order terms of the Taylor expansion (12) into account, we expect zero bias if

$$\mathrm{Re}\left\{\overline{\nabla_{\mathbf{e}_l}\dot{V}(\theta, \mathbf{E}_s)}^*\,\Delta\mathbf{e}_{s,l}\right\} = 0 \tag{60}$$

However, after a pre-processing that at least approximately yields (60)=0, and thus $\Delta\theta \approx 0$, but still $\Delta\mathbf{e}_{s,l} > 0$,the remaining bias is determined by the higher order terms (12c) - (12f) in the Taylor expansion (12). Since we assume $\Delta\theta \approx 0$ we can omit (12c) so the remaining bias will depend on the terms

$$\begin{aligned} &2\,\mathrm{Re}\left\{\overline{\nabla_{\mathbf{e}_l}\dot{V}(\theta, \mathbf{E}_s)}^*\,\Delta\mathbf{e}_{s,l}\Delta\theta\right\} \\ &+\mathrm{Re}\left\{Tr(\mathbf{H}1\Delta\mathbf{e}_{s,l}\Delta\mathbf{e}_{s,l}^*)\right\} \\ &+0 + o(N^{-1}) \end{aligned} \tag{61}$$

Simulations indicate that imposing the tightened requirement on the mapping errors that

$$\overline{\nabla_{\mathbf{e}_l}\dot{V}(\theta, \mathbf{E}_s)}^*\,\Delta\mathbf{e}_{s,l} = 0 \tag{62}$$

in *both* the real and imaginary parts, helps in reducing the sum of the terms in (61). This is made possible by the evident commonality in the structure of (55), (57) and (58), but is complicated to show analytically in a general case.

The property (62) is imposed on the errors $\Delta\mathbf{e}_{s,l}$ simply by dropping the Re -operator in the suggested design criterion (47) for $\mathbf{T}$. As already pointed out this measure has the additional advantage of reducing the size of the equation system that has to be solved to find $\mathbf{T}$ from (47), but on the other hand necessitates complex calculations.

REFERENCES

[1] A. J. Weiss and B. Friedlander, "Performance analysis of spatial smoothing with interpolated arrays," *IEEE Transactions on Signal Processing*, vol. 41, pp. 1881–1892, May 1993.

[2] A. J. Weiss and B. Friedlander, "Preprocessing for direction finding with minimal variance degradation," *IEEE Transactions on Signal Processing*, vol. 42, pp. 1478–1485, June 1994.

[3] S. Andersson, *On Dimension Reduction in Sensor Array Processing*. PhD thesis, Linköping University, Linköping, Sweden, Dec. 1992.

[4] P. Hyberg, "Circular to linear array mapping and bias reduction," tech. rep., Royal Institute of Technology KTH, Stockholm, Sweden, Dept. Signals, Sensors & Systems, Sep. 2001. Licentiate Thesis.

[5] T. P. Bronez, "Sector interpolation of non-uniform arrays for efficient high resolution bearing estimation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, (New York, NY), pp. 2885–2888, Apr. 1988.

[6] B. Friedlander, "The root-MUSIC algorithm for direction finding with interpolated arrays," *Signal Processing*, vol. 30, pp. 15–29, January 1993.

[7] M. Gavish and A. Weiss, "Direction finding using ESPRIT with interpolated arrays," *IEEE Transactions on Signal Processing*, vol. 39, pp. 1473–1478, June 1991.

[8] P. Hyberg, M. Jansson, and B. Ottersten, "Array mapping: Optimal transformation matrix design," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, pp. 2905–2908, May 2002.

[9] P. Hyberg, "Array mapping: Reduced bias transformation matrix design," in *Proceedings of RadioVetenskap och Kommunikation 02*, (KTH, Stockholm, Sweden), 2002.

[10] P. Hyberg, M. Jansson, and B. Ottersten, "Array Interpolation and Bias Reduction," *IEEE Transactions on Signal Processing*, to appear.

[11] X.-L. Xu and K. M. Buckley, "Bias analysis of the MUSIC location estimator," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2559–2569, Oct. 1992.

[12] P. Stoica and A. Nehorai, “MUSIC, Maximum Likelihood, and Cramér-Rao Bound,” *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 37, pp. 720–741, May 1989.

[13] P. Stoica and A. Nehorai, “MUSIC, Maximum Likelihood, and Cramér-Rao Bound: Further Results and Comparisons,” *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, pp. 2140–2150, Dec 1990.

[14] J. Eriksson, *On Detection and Estimation of Multiple Sources in Radar Array Processing*. PhD thesis, Chalmers University of Technology CTH, Dept. Signals and Systems, Chalmers University of Technology CTH Gothenburg, Sweden, Feb. 2002.

[15] A. G. M. Pesavento and Z.-Q. Luo, “Robust array interpolation using second order cone programming,” *IEEE Signal Processing Letters*, vol. 9, pp. 8–11, Jan. 2002.

[16] C. P. Mathews and M. D. Zoltowski, “Beamspace esprit for multiple source arrival angle estimation employing tapered windows,” in *ICASSP 2002 Proceedings*, May 2002.

[17] D. H. Brandwood, “A complex gradient operator and its application in adaptive array theory,” *Proc. Inst. Electr. Eng.*, vol. 130, pp. 11–16, Feb. 1983. pts. F and H.

[18] Ottersten and M. Viberg, “Sensor array signal processing,” tech. rep., Royal Institute of Technology, KTH, Stockholm, Sweden, 1994. compendium.

[19] J. Eriksson and M. Viberg, “Data reduction in spatially colored noise using a virtual uniform linear array,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, (Istanbul, Turkey), pp. 3073–3076, 2000.