ROBUST ADAPTIVE BEAMFORMING BASED ON DOMAIN WEIGHTED PCA

Paul D Baxter John G McWhirter

Advanced Signal and Information Processing Group, QinetiQ Ltd
St Andrews Road, Malvern, Worcestershire, WR14 3PS, United Kingdom
phone: + (44) 01684 897458, fax: + (44) 01684 896502, email p.baxter@signal.qinetiq.com

ABSTRACT

A novel technique for robust adaptive beamforming (ABF) is proposed. The technique, referred to as Domain-Weighted PCA (DW-PCA), is founded on a basic paradigm shift from one of noise cancellation to one of signal separation. It uses the singular value decomposition (SVD) to perform second order blind signal separation after applying a simple transformation to the data. The transformation is designed to exploit prior knowledge in the form of one or more estimated steering vectors. The method is quite distinct from existing algorithms for robust ABF and can offer improved performance in many cases. The results of computer simulations, which demonstrate this point, are presented.

1. INTRODUCTION

Sensor array signal processing is widely used in many fields, including communications, radar, seismology and sonar. Fixed beamformers were developed to use simple prior information, usually the response of the array to the signal of interest (SOI). Unfortunately these are highly susceptible to interfering signals. This is not the case however for ABF techniques. If SOI-free training data are available, powerful ABF techniques have been proposed. However, in many applications SOI-free data are not obtainable. In these situations, the SOI is present in all the data; as a result ABF techniques become highly susceptible to errors in underlying assumptions made about the environment, the sources or the array. The performance of ABF techniques is shown to degrade significantly when the sample size is small or when the response of the array to the SOI is even slightly wrong. This means that robust ABF techniques are required.
The basic ABF technique is the look direction constrained least squares adaptive technique, as used in the Minimum Variance Distortionless Response (MVDR) beamformer [1] in which the sample matrix inversion (SMI) algorithm can be exploited. Techniques developed to be more robust include: adding diagonal loading to the sample covariance matrix [2]; penalty function approaches [3]; a worst-case optimisation approach [4] and many others, see [4] and references therein. Many of these techniques are robust against certain types of error, but not robust against others – a good summary is provided in [4].
This paper proposes a novel technique for robust adaptive beamforming using Principal Component Analysis (PCA). It is based upon two observations. Firstly, adaptive beamforming is a special case of signal separation; we are trying to separate one SOI from other interfering signals. Various blind signal separation (BSS) algorithms can do this with minimal underlying assumptions – but these do not emphasise the SOI over the other signals, nor in their original formulation can they utilise prior information. The initial stage of many of these BSS algorithms uses PCA. We observe that if the total power of the SOI across all the channels is significantly different from that of the interferers, then the PCA carries out most of the separation. Based on these observations, we propose a technique that uses information about the SOI and the sensor array to modify the total power of the SOI so it becomes distinct from that of the interferers; PCA is then applied to extract the SOI.
Our technique uses both data and prior information in the data analysis. The prior information is used in a 'soft' manner, which means the technique is more robust to errors in the information than the MVDR beamformer and compares well with other robust ABF techniques. However by using the prior information, the technique produces better results than purely data based techniques, such as standard PCA and BSS techniques, particularly when statistical estimates of quantities, e.g. the covariance matrix, are unreliable. We call our technique Domain-Weighted PCA because it applies weighting dependent upon the prior information followed by carrying out standard PCA.
The paper is organised as follows, section 2 describes the background and introduces the notation to be used. In section 3 we demonstrate that in a mixture of several signals the powerful signals are removed from the weaker signals by the PCA, but that the reverse is not true. The DW-PCA technique is introduced in section 4, and some results from using it are presented in section 5. Section 6 contains the conclusions.

2. BACKGROUND AND NOTATION

The classical instantaneous, stationary, and linear mixing model that is used throughout this paper is that the received data matrix $\mathbf{X}$ can be modelled as follows:

$$\underset{n\times T}{\mathbf{X}} = \underset{n\times m}{\mathbf{C}}\ \underset{m\times m}{\mathbf{P}}\ \underset{m\times T}{\mathbf{S}} + \underset{n\times T}{\mathbf{N}}$$

The T columns of $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{N}$ contain snapshots of the received data, original signals and sensor noise respectively, indexed by time, so for example $\mathbf{X} = [\mathbf{x}(1) \quad \mathbf{x}(2) \quad \cdots \quad \mathbf{x}(T)]$. Each snapshot is a complex vector, $\mathbf{x}(t) = [x_1(t) \quad x_2(t) \quad \cdots \quad x_n(t)]^T$ where $[\cdot]^T$ denotes the vector transpose. It is assumed that the vector quantities $\mathbf{n}(t)$, $\mathbf{s}(t)$ and $\mathbf{x}(t)$ are all zero mean, and that the rows of $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{N}$ are zero sum. $\mathbf{P}$ is a diagonal matrix containing the powers of the individual signals. $\mathbf{C}$ is constructed columnwise so that the p'th column of $\mathbf{C}$, i.e. $\mathbf{c}_p$, contains a unit norm vector corresponding to the pointing vector of the p'th signal.
The spatial covariance matrix of a zero mean wide sense stationary random vector quantity, e.g. $\mathbf{x}(t)$, is denoted by $\mathbf{R}$ with a subscript corresponding to the quantity, e.g. $\mathbf{R}_{xx}$. It is defined by:

$$\mathbf{R}_{xx} = E\{\mathbf{x}(t)\mathbf{x}(t)^H\}$$

where $(\cdot)^H$ denotes Hermitian transpose, and $E\{\cdot\}$ is the statistical expectation operator.
We would like to work with the covariance matrices, but we do not have access to them, and so must estimate them from the samples. The standard, consistent, estimators of these are denoted by a hat and are calculated from the matrices $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{N}$; e.g.

$$\hat{\mathbf{R}}_{xx} \stackrel{\Delta}{=} \frac{\mathbf{X}\mathbf{X}^H}{T} \approx \mathbf{R}_{xx}$$

A beamformer calculates a weight vector $\mathbf{w} \in C^{n \times 1}$ that can be applied to the data snapshots to produce the beamformer output:

$$z_{bf}(t) = \mathbf{w}^H \mathbf{x}(t)$$

Following the definition of signal-to-interference-plus-noise ratio (SNIR) as in [4], each data snapshot $\mathbf{x}(t)$ can be broken down into the sum of statistically independent components, $\mathbf{x}_s(t)$, $\mathbf{x}_i(t)$ and $\mathbf{n}(t)$, i.e. the SOI, interfering signals and noise. The covariance matrices of $\mathbf{x}_s(t)$ and of $\mathbf{x}_i(t) + \mathbf{n}(t)$ are defined by:

$$\mathbf{R}_{SOI} = E\left\{\mathbf{x}_s(t)\mathbf{x}_s(t)^H\right\}$$

$$\mathbf{R}_{i+n} = E\left\{(\mathbf{x}_i(t) + \mathbf{n}(t))(\mathbf{x}_i(t) + \mathbf{n}(t))^H\right\}$$

The SNIR is then calculated by the following formula:

$$SNIR = \frac{\mathbf{w}^H \mathbf{R}_{SOI} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}}$$

PCA is a linear data analysis tool that is entirely data based, i.e. it does not use any prior information. It is used in exploratory data analysis, extracting dominant signal components from linear mixtures, and as a pre-processing step within many established BSS algorithms. In this paper, we only use the data domain version carried out by an SVD. However the ideas presented can be transferred into the covariance domain with only small modifications to the technique required.

Symbolically, the SVD operates on a $n \times T$ data matrix **Y**. It produces the following decomposition of this data matrix into an $n \times n$ unitary matrix **U**, an $n \times n$ diagonal matrix $\mathbf{\Lambda}^{1/2}$, and an $n \times T$ modified data matrix, **V**:

$$\mathbf{Y} = \mathbf{U}^H \mathbf{\Lambda}^{1/2} \mathbf{V}$$

The values on the diagonal of $\mathbf{\Lambda}^{1/2}$ are known as the singular values. They are non-negative. It is a convention to order the rows of **V** so that the singular values are arranged in descending order. The matrix **V** has the property that $\mathbf{V}\mathbf{V}^H = \mathbf{I}_n$.

When using the SVD to carry out PCA, it is sufficient to calculate **U** and use this to calculate a transformed data matrix **Z** according to:

$$\mathbf{Z} = \mathbf{U}\mathbf{Y} = \mathbf{\Lambda}^{1/2}\mathbf{V}$$

Thus the estimated covariance matrix of $\mathbf{z}(t) = \mathbf{U}\mathbf{y}(t)$ is:

$$\hat{\mathbf{R}}_{zz} = \mathbf{Z}\mathbf{Z}^H / T = \mathbf{\Lambda}/T$$

Thus the estimated correlation between two different rows of **Z** is zero following the SVD; usually this is referred to as the data matrix **Z** being decorrelated. Under the constraint of using a unitary matrix, such as **U**, the SVD also ensures that the sum of the powers of the first p rows of **Z** is maximised, for any p.

We note that a unitary transformation of a vector quantity can be considered to be energy preserving in the following way. We define the total energy of a data vector $\mathbf{y}(t)$ as the sum of the diagonal terms of its covariance matrix, e.g.

$$p_{y1} = E\left\{y_1(t)y_1(t)^*\right\} = \left[\mathbf{R}_{yy}\right]_{1,1}$$

Thus the total energy of a data vector $\mathbf{y}(t)$ is:

$$\sum_{i=1}^{n} p_{yi} = \mathrm{trace}\left(E\left\{\mathbf{y}(t)\mathbf{y}(t)^H\right\}\right)$$

If the data is transformed according to $\mathbf{z}(t) = \mathbf{U}\mathbf{y}(t)$ then, using the linearity of the expectation operator, the new total energy is:

	Singular	Signals			
		-10dB 0 degrees	10dB 44 degrees	30dB 86 degrees	Noise
Output	100.1	0.0 %	0.2 %	99.8 %	0.0 %
	9.2	0.1 %	98.5 %	0.3 %	1.2 %
	1.4	44.6 %	0.0 %	0.0 %	55.4 %

Table 1: The proportion of the channel powers due to each of the signals and the noise

$$\sum_{i=1}^{n} p_{zi} = \mathrm{trace}\left(E\left\{\mathbf{z}(t)\mathbf{z}(t)^H\right\}\right) = \mathrm{trace}\left(E\left\{\mathbf{U}\mathbf{y}(t)\mathbf{y}(t)^H\mathbf{U}^H\right\}\right) = \sum_{i=1}^{n} p_{yi}$$

A similar result holds for the sample estimated total energy when a unitary transform is applied to a data matrix. This is a consequence of a unitary matrix performing only a rotation on a vector to which it is applied.

3. POWER BASED SEPARATION

It is understood that applying the SVD is generally not sufficient to separate out the original signals from each other. This is because the SVD finds a new data matrix **V** such that the rows of the data matrix have been decorrelated from each other and their powers have been normalised, i.e. $\mathbf{V}\mathbf{V}^H = \mathbf{I}_n$. Then transforming **V** to $\mathbf{V}'$ by applying a unitary matrix **Q** does not change this property:

$$\mathbf{V}' = \mathbf{Q}\mathbf{V} \Rightarrow \mathbf{V}'\mathbf{V}'^H = \mathbf{Q}\mathbf{V}\mathbf{V}^H\mathbf{Q}^H = \mathbf{Q}\mathbf{I}_n\mathbf{Q}^H = \mathbf{I}_n$$

BSS algorithms use higher order statistics (implicitly or explicitly) to find the correct unitary matrix to separate out the signals. It is, however, incorrect to assume that the unitary matrix needed to separate the signals after the SVD is uniformly distributed over the unitary group of matrices [5]. A simple experiment can show that if a signal has significantly greater power than the others, the SVD comes close to separating it from the other signals. In this case, we demonstrate that it is possible to achieve good signal separation using only second order statistics. This is as expected by the energy-compaction nature of the SVD.

We demonstrate this for a simulated ten sensor circular array, where the spacing between two adjacent elements is half the wavelength of the signals. Three Quaternary Phase Shift Keying (QPSK) signals of length 10,000 samples are simulated, impinging on the array from angles of arrival (AOAs) of $\{0°, 44°, 86°\}$ with SNRs of -10dB, 10dB and 30dB respectively. This means that the weakest signal has enough power so that it is just possible for the SVD to distinguish it from noise, and there is at least a 20dB power difference between any two signals.

The SVD was applied to the simulated data. Table 1 displays the behaviour found; it shows how much of the power of each channel is due to each of the signals, and how much is due to noise. The three dominant channels between them contained almost all the power of the three signals.

Table 1 shows that a good degree of signal separation has been achieved. The dominant channel is comprised of 99.8% of the +30dB signal, 0.2% of the +10dB signal and trivial amounts of the weakest signal and noise: Thus this channel has succeeded in extracting the most powerful signal. There is a small amount of signal leakage between signals that are only 20dB apart in power, e.g. the 0.2% just mentioned. Although this is a very small corruption of the +30dB signal, it actually comprises a significant amount (16%) of the total power of the +10dB signal. Thus significant proportions of weak signals can corrupt the strong signals, where their influence is small.

The level of signal separation just demonstrated, achieved by using the SVD, is good. However it can only be achieved by the SVD when the signals do have significantly different powers. Also, BSS techniques can produce superior results, further reducing the cross leakage by normalising the output channels and finding an appropriate rotation matrix.

4. DOMAIN-WEIGHTED PCA

We introduce a robust adaptive beamformer that uses prior information to impose artificially different signal powers, where they did not exist before. This is followed by applying an SVD, which separates the signals much better than they would have been had the SVD been carried out on the raw data. This can be considered as applying some pre-emphasis to the PCA.

The proposed DW-PCA technique has three distinct stages as shown in Figure 1. In the following discussion of the stages, the case where the prior information from the domain is in the form of a pointing vector for the SOI, i.e. $\hat{\mathbf{c}}_1$, is considered.

Domain Transform: In a similar way to the Griffiths-Jim transformation, [6], domain knowledge is used to produce a transformation of the data into a primary channel and a set of auxiliary channels by means of a unitary transform. i.e.

$$\mathbf{X}_1 = \mathbf{U}\mathbf{X} = \begin{bmatrix} \mathbf{u}_1^{\mathrm{H}} \\ \overline{\mathbf{U}} \end{bmatrix} \mathbf{X}$$

The first row of $\mathbf{U}$, $\overset{\Delta}{=} \mathbf{u}_1^{\mathrm{H}}$, is chosen so that $\hat{\mathbf{c}}_1^{\mathrm{H}} \mathbf{u}_1 = \lambda$, where λ is near 1. The simplest choice is $\hat{\mathbf{c}}_1^{\mathrm{H}} = \mathbf{u}_1^{\mathrm{H}}$, but other choices are possible e.g. the projection of $\hat{\mathbf{c}}_1$ onto the null space of $\{\hat{\mathbf{c}}_2 \cdots \hat{\mathbf{c}}_p\}$. The remaining rows of $\mathbf{U}$ are chosen so that the rows of $\mathbf{U}$ form a basis. Hence $\mathbf{U}$ is unitary, and the transform is energy preserving. The aim of this stage of processing is to ensure, by means of an energy preserving transform, that the primary data channel is mainly composed of the SOI, rather than other signals or noise. In the DW-PCA, as opposed to the GSLC, [6], *it is not necessary to ensure that the SOI does not leak into the auxiliary channels*. The DW-PCA only requires that the SNIR of the primary channel is better than the SNIR of the original data channels.

Primary Channel Enhancement: Now that a channel has been produced to consist mainly of the SOI, this channel is enhanced by a factor μ. Symbolically, this is achieved through the application of a diagonal matrix $\mathbf{D}$:

$$\mathbf{Y} = \mathbf{D}\mathbf{X}_1$$

The top-left entry of $\mathbf{D}$ is μ, but the remaining diagonal entries are 1. The effect of varying μ will be considered in section 5, however taking μ=1 reduces $\mathbf{D}$ to the identity, and so is equivalent to doing nothing. A normal value of μ to take is 2. This stage is not energy preserving –it enhances the energy in the primary channel.

Power Based Separation: Following the enhancement of the primary channel, it remains to apply PCA to all of the channels. This applies an energy preserving transform in the form of a unitary matrix $\mathbf{Q}$:

$$\mathbf{Y}_1 = \mathbf{Q}\mathbf{Y} = \mathbf{Q}\mathbf{D}\mathbf{U}\mathbf{X}$$

Although the ordering of the rows in $\mathbf{Y}_1$, the output of the power-based separation, is ambiguous, it is trivial to either assign an ordering based on the powers of the outputs or to decide which output is derived the most from the primary channel. $\mathbf{Y}_1$ is the output of the DW-PCA technique. The output of interest for robust ABF may be the dominant powered signal, or the signal best aligned with the prior pointing vector. The beamforming vector, $\mathbf{w}$, can be found from the corresponding row of the matrix product **QDU.**

The computational cost of the DW-PCA technique is not much greater than that of the PCA. For an $n \times T$ data matrix, calculating and applying $\mathbf{U}$ and $\mathbf{D}$ takes $O(n^2T+n^3)$ operations. The PCA also takes $O(n^2T+ n^3)$ operations. Empirically, the cost of the PCA is observed to be the dominant factor.

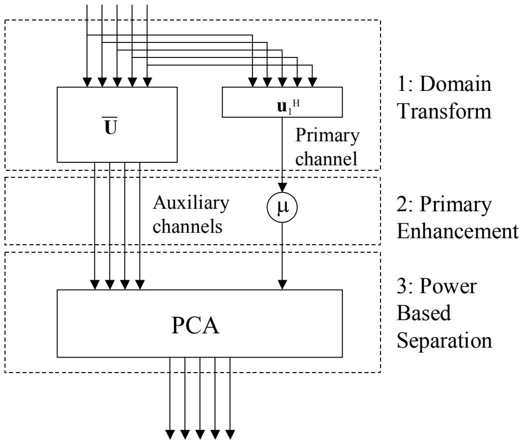

Figure 1: Diagrammatic representation of the DW-PCA technique

dB	Prior Error 0%			Prior Error 2%			Prior Error 8%		
Jammer	None	Weak	Strong	None	Weak	Strong	None	Weak	Strong
Fixed	4.0	3.6	-34	3.9	3.5	-35	3.7	3.3	-36
SMI	3.7	3.5	3.3	2.8	2.7	2.5	0.9	0.8	0.7
Diag	4.0	3.6	3.5	3.9	3.5	3.4	3.6	3.2	3.1
Wong	4.0	3.6	-58	3.9	3.5	-57	3.7	3.3	-58
μ=1.1	4.0	3.3	3.5	4.0	3.3	3.5	4.0	3.3	3.5
μ =2	4.0	3.6	3.5	3.9	3.5	3.5	3.8	3.3	3.3

Table 2: Output SNIRs achieved, in several different levels of jamming and prior error, by several techniques for extracting a single SOI. The best results for each scenario are shaded.

5. RESULTS

Three simulations were used to demonstrate the behaviour of the DW-PCA technique. The following algorithms were used for comparison: the fixed beamformer, which only uses the prior information; the SMI MVDR beamformer, which uses prior information and the data; the SMI MVDR beamformer with diagonal loading of 30, and the beamformer proposed by Wong et al. in [4] with parameters γ=30 and ε=16.

The first simulation was of a ten sensor circular array upon which the SOI, composed of 500 samples of a QPSK signal, impinged at –6dB SNR on each sensor. If a jammer was present it arrived from the sidelobes at a SNR of either –10dB (weak) or +44dB (strong). Three different levels of prior information were used, either $\hat{\mathbf{c}}_1$ was accurate ($= \mathbf{c}_1$) or it consisted of 98% or 92% of $\mathbf{c}_1$ with the remaining 2% or 8% being random error vectors chosen from a $N(0,\sigma^2 I)$ distribution. The output SNIRs, averaged over 100 Monte-Carlo runs, are shown in table 2.

Table 2 shows that the DW-PCA technique is more robust to errors in the prior information than diagonal loading techniques, while still being adaptive enough to cope with sidelobe jamming. In the case of strong jamming a value of μ near unity performed the best, but in the case of weak jamming μ=2 produced the best results.

To study the effects of different μ values a second simulation was run. This was of a 20 element linear array, as in [4], upon which the QPSK SOI impinged with varying SNR. The prior information was incorrect by 3°, i.e. $\mathbf{c}(\hat{\theta}_1) = \mathbf{c}_1(\theta_1 + 3°)$. A jamming signal was present at a level of +20 dB, impinging on the array from a random location in the sidelobes of the SOI. Figure 2 shows the performance curves, obtained by averaging over 100 Monte-Carlo runs, for the DW-PCA technique with various values of μ. The optimal performance was calculated analytically by using the pseudo-inverse of $\mathbf{CP}$ to separate the signals from each other.

Figure 2 shows that when the level of enhancement actually takes the signal power to a level similar to the jammer power, then the DW-PCA technique does not lead to signal separation. This is consistent with the observations made at the beginning of section 2. For large μ the performance tends to that of a fixed beamformer.

When the DW-PCA technique does not lead to signal separation, two of the singular values obtained by the SVD are similar. It is possible to use this to detect when an unsuitable value of μ has been chosen, and the technique could be carried out again with a different value. Further work needs to be done on quantifying the link between the differences between the singular values and the performance of the techniques.

The third simulation used the same setup as the second simulation, only both the SOI and the jammer powers were fixed at 0dB, while the error in the prior information was allowed to vary from 0° to 5°. The performance curves for the various comparison algorithms and for the DW-PCA technique with several different μ values are shown in Figure 3.

Figure 3 clearly demonstrates the difficulty the SMI ABF has with errors in the prior information. This can be mitigated by using diagonal loading, and further mitigated by the use of Wong's algorithm; but neither of these are as robust as using a fixed beamformer (although they will show better jammer rejection if the jammer is more powerful). The plain SVD, imposing PCA, is not very good in this situation as the jammer and signal powers are the same. However by increasing the μ value, the performance of the algorithms improves markedly, up to μ=2. Beyond that, a further increase to μ=5 leads to the performance tending to that of a fixed beamformer – this corresponds to an overemphasis on the prior information. It should be noted that this over emphasis does not lead to performance as bad as that of the SMI ABF.

6. CONCLUSIONS

The Domain Weighted PCA technique offers a new, and very simple, technique for data analysis that allows prior information about signals to be included in a soft manner. Thus we can avoid both the over-reliance on and the rejection of the prior information. The specific version of the DW-PCA introduced in this paper uses one simple form of prior information. The technique can be easily extended to incorporate more information, either for improved extraction of a single signal or to enable the robust determination of a signal+noise subspace.

We have demonstrated that the DW-PCA technique can operate as a robust adaptive beamformer, with performance which depends on μ, but which can exceed that of other ABF techniques. The correct selection of μ can be important to obtain the best performance from the technique. This can be aided by looking at the singular values obtained from the SVD. In offering a method of using both the data and 'softly' incorporating the prior information, the DW-PCA technique can offer performance that is startlingly distinct from other data analysis techniques.

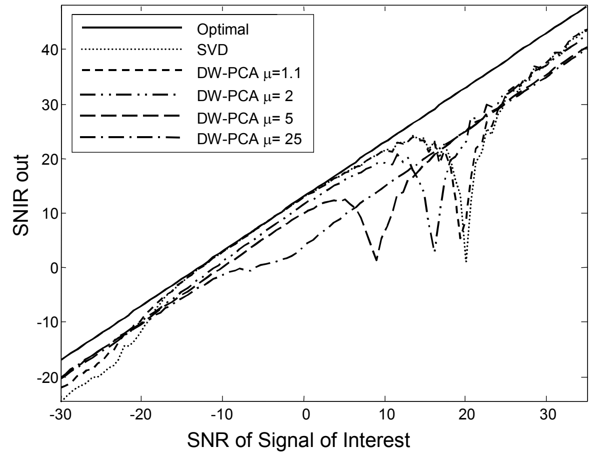

Figure 2: SNIR varying with SNR of input signal for a variety of μ levels in the DW-PCA

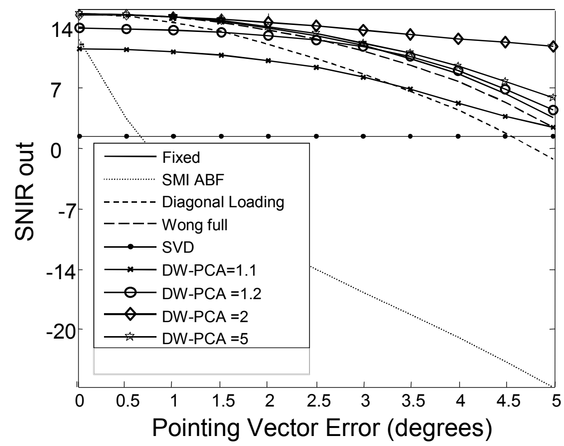

Figure 3: SNIR varying with error in pointing vector for a variety of signal extraction techniques

7. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory* 2'nd Edition: Prentice Hall, 1991

[2] B. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays", *IEEE Transactions on Aerospace and Electronic Systems*, vol.24, Issue 4, pp.397-401, July 1988

[3] D. Hughes and J. McWhirter, "Penalty Function Method for Sidelobe Control in Least Squares Adaptive Beamforming", *Proc. SPIE Conference on Advanced algorithms and Architectures for Signal Processing*, San Diego, CA , July 1995, pp170-181, 1995

[4] S. Shahbazpanahi, A. Gershman, Z. Luo and K.Wong, "Robust Adaptive Beamforming for General-Rank Signal Models*", IEEE Transactions on Signal Processing*, vol.51 Issue 9, pp.2257 - 2269, September 2003

[5] A. Joshi, "Elements of Group Theory for Physicists" 2'nd Edition, Wiley Eastern, 1976

[6] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming", IEEE Transactions on Antennas and Propagation, vol.30, pp. 27 – 34, Jan 1982

This research was sponsored by the MOD Corporate Research Programme.