Detection of the Number of Signals in Noise with Banded Covariance Matrices

W. Chen,[*] J. P. Reilly,[†] K. M. Wong

Communications Research Lab,

McMaster University

1280 Main St. W.,

Hamilton, Ont.,

CANADA, L8S 4K1

telephone: + 1 905 525 9140 ext 22895

fax: + 1 905 521 2922

email: jim@reilly.crl.mcmaster.ca

Abstract

A new approach is presented to the array signal processing problem of detecting the number of incident signals in unknown *coloured* noise environments with banded covariance structure. The principle of canonical correlation analysis is applied to the outputs of two spatially separated arrays. The number of signals is determined by testing the significance of the corresponding sample canonical correlation coefficients. The new method is shown to work well in unknown coloured noise situations and does not require any subjective threshold setting. The medium/high-SNR error rate may be approximately specified at a certain prescribed level, and may be traded off against the detection performance characteristic at low SNR. Simulation results are included to illustrate the performance of the proposed canonical correlation technique (CCT). It is found that the method performs well in a wide variety of coloured background noise environments. It is also demonstrated that the method is robust in the case when the noise covariance is not truly banded.

[*]Now with COM DEV Ltd., Cambridge, Ont. Canada

[†]Person to whom correspondence should be addressed

1 Introduction

This paper uses canonical correlation analysis to solve the problem of detecting the number of sources incident onto an array of sensors, where the background noise covariance is coloured and unknown with a banded structure. Detection of the number of sources in array signal processing has been a widely studied problem for many years. This is not only because determination of the number of incident signals onto an array of sensors is important information in its own right, but also because modern direction of arrival estimation algorithms depend on knowledge of this quantity.

For the white noise situation, various elegant detection methods have been proposed. The original approach to this problem is based on hypothesis testing, which uses the eigenvalues of the sample covariance matrix of the observation vectors. The typical methods of this type can be found in the papers written by Bartlett [4], Lawley [13], and Anderson [2]. The common problem associated with this approach is that the threshold values used for hypothesis testing must be determined *subjectively*. To avoid this subjectivity, other methods for *known* noise have been developed. Wax and Kailath [20] applied the information theoretic criteria introduced by Akaike (AIC) [1] and by Schwartz and Rissanen (MDL) [18] [17] to the detection problem. Chen, Wong, and Reilly [7] introduced a predictive eigen-threshold (ET) method, whereas Wu and Fuhrmann [22] developed a high-performance parametric detection technique which is performed in conjunction with direction of arrival estimation. No subjective threshold is required for these methods. The difficulty with these methods for known (or equivalently white) noise is that their performance degrades very quickly as the degree of colour of the noise increases.

Le Cadre [14] has proposed a parametric method for detection for the *unknown* noise case, which involves finding the minimum of a specific information theoretic function to yield the estimated model order. Fuchs [10] has also proposed a method, which separates a linear sensor array into two non-overlapping segments. The assumption is made that the noises between the segments are uncorrelated. This assumption leads to a proposed function of the estimated noise from both segments, which is χ^2-distributed. A χ^2 test on an hypothesized model order may then be conducted.

The method [8][9] proposed in this paper has some ideas in common with that of Fuchs.

Specifically, each method proposes two distinct array segments, and it is assumed in each case the background noise is uncorrelated between the two segments. However, here, we propose two physically separated arrays (Fig. 1), instead of one contiguous array divided into two segments. Our proposed method uses the *canonical correlation coefficients* derived from the data received from the two arrays to test for the most likely number of incident signals. For this reason, it is referred to as the canonical correlation test (CCT) method. The CCT procedure offers several advantages. First, the medium/high SNR error rate may be specified at a certain prescribed level, which is traded-off against performance at low SNR. Also, the method is statistically rigorous, simple to apply, and offers relatively high performance.

The CCT method is based on the ideal assumption that the background noise is uncorrelated between the two arrays. In the practical setting, this condition is rarely exactly satisfied. However, the condition is approximately satisfied in many applications. For example, it is shown in [16] that the reverberation noise in a sonar environment decays relatively quickly in space. This means it is possible to find a separation δ shown in Fig. 1 between the two proposed arrays so that the noise correlations between them are small. We have demonstrated by simulation that the CCT method is robust when significant noise correlations do exist between the two arrays. Therefore, although ideally the CCT method depends on this uncorrelated assumption, in practice we find the method performs well when some degree of correlation exists.

The ultimate performance of any detection method in noise of unknown characteristic depends on the fact that the spatial correlation function of the noise is sufficiently narrow in the spatial dimension. Noise processes with spatially narrow autocorrelation functions have wide, or directionally diffuse wavenumber spectra, and thus are distinct from the wavenumber characteristics of a signal. It becomes more difficult for a coloured noise detection scheme to distinguish noise from a signal as the noise wavenumber spectrum becomes increasingly narrow. Therefore it is reasonable, as we have done, to restrict attention to those noise environments which are more directionally diffuse, with spatial correlation functions which decay relatively quickly. These are the conditions which more closely satisfy our assumption that the background noise is uncorrelated between the two arrays.

The modelling of the physical environment is described in Sect. 2 of this paper. The idea of canonical correlations and its application to the detection problem is addressed in Sect. 3. In Sect. 4, the main idea of using the canonical correlation coefficients for detection is discussed. Simulation results are given in Sect. 5, and conclusions drawn in Sect. 6.

2 Formulation of the problem

Let us consider k independent narrowband signals arriving from k distinct directions, onto two spatially separated arrays. These arrays are denoted by X and Y with p and q sensors respectively. Fig. 1 shows an example of the arrangement of the two arrays, which in this particular case, are both linear with uniformly separated sensors. Signals are incident on the two arrays with physical angles $\boldsymbol{\varphi}_x$ and $\boldsymbol{\varphi}_y$ respectively.[1] We denote the corresponding electrical direction of arrival (DOA) of the k signals relative to the normal of the two arrays by $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$ respectively, such that each has k elements representing the DOA's of the k incident signals with respect to the positions of Array X and Array Y. The outputs of these two arrays are denoted by vectors $\mathbf{x}(n)$, $\mathbf{y}(n)$ which can be expressed as

$$\mathbf{x}(n) = \mathbf{A}_x(\boldsymbol{\theta}_x)\boldsymbol{\alpha}_x(n) + \boldsymbol{\nu}_x(n) \tag{1}$$

$$\mathbf{y}(n) = \mathbf{A}_y(\boldsymbol{\theta}_y)\boldsymbol{\alpha}_y(n) + \boldsymbol{\nu}_y(n), \qquad n = 1, \ldots, N, \tag{2}$$

where N is the number of snapshots (observations), and $\mathbf{A}_x(\boldsymbol{\theta}_x)$ and $\mathbf{A}_y(\boldsymbol{\theta}_y)$ are $p \times k$ and $q \times k$ unambiguous directional matrices of the signals with respect to the geometries of the arrays X and Y respectively. The vectors $\boldsymbol{\alpha}_x(n)$ and $\boldsymbol{\alpha}_y(n)$ are $k \times 1$ signal vectors received by the two arrays and are modelled to be complex zero-mean jointly-distributed Gaussian vectors. If the travelling of the signals from Array X to Array Y involves no distortion or alteration, then $\boldsymbol{\alpha}_y$ is merely a delayed version of $\boldsymbol{\alpha}_x$. In this case, narrowband signals appearing on the two arrays may be modelled identically and the actual delay delay between the two arrays can be absorbed into the directional matrices.

The vectors $\boldsymbol{\nu}_x(n)$ and $\boldsymbol{\nu}_y(n)$ are p and q dimensional vectors respectively, representing noise components in the outputs of array X and array Y. These noise vectors are assumed uncorrelated with the signals. Ideally, these noises are assumed to be Gaussian, complex,

[1] In Fig. 1 for simplicity we have set $\varphi_x = \varphi_y = \varphi$, although in general this relation need not hold.

zero mean, with joint covariance matrix satisfying

$$\mathrm{E}\left\{\begin{bmatrix}\boldsymbol{\nu}_x \\ \boldsymbol{\nu}_y\end{bmatrix}\begin{bmatrix}\boldsymbol{\nu}_x^H & \boldsymbol{\nu}_y^H\end{bmatrix}\right\} = \begin{bmatrix}\boldsymbol{\Sigma}_{\nu_x} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_{\nu_y}\end{bmatrix} \tag{3}$$

where superscript H denotes Hermitian transpose, and $\mathbf{O}$ denotes a $p \times q$ null matrix, and $\boldsymbol{\Sigma}_{\nu_x}$ and $\boldsymbol{\Sigma}_{\nu_y}$ are unknown noise covariance matrices of the noise in the two arrays. Eq. (3) does not unduely restrict the method because, 1) as we see later from simulation results, the proposed method is robust against violations of this assumption, and 2) as discussed earlier, provided the background noise characteristic is not too directional, the spatial covariance function of the noise will decay at such a rate that there exists a separation δ (Fig.1) between the arrays so that (3) holds approximately.

To get an idea of the structure of the joint variance-covariance matrix of $\mathbf{x}(n)$ and $\mathbf{y}(n)$ under such assumptions, let us define a composite vector $\mathbf{z}(n)$ with components $\mathbf{x}(n)$ and $\mathbf{y}(n)$ as,

$$\mathbf{z}(n) = \begin{bmatrix}\mathbf{x}(n) \\ \mathbf{y}(n)\end{bmatrix} \tag{4}$$

and assume there exist k signals. The variance-covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{x}(n)$ and $\mathbf{y}(n)$ can be written as

$$\boldsymbol{\Sigma} = E[\mathbf{z}(n)\mathbf{z}(n)^H] = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\begin{matrix} p \\ q\end{matrix} \tag{5}$$
$$\qquad\qquad\qquad\qquad p \quad q$$

In Equation (5), we have partitioned $\boldsymbol{\Sigma}$ according to the dimensions of $\mathbf{x}(n)$ and $\mathbf{y}(n)$ respectively. Without loss of generality, in the following discussion, we assume that $p \leq q$.

The covariance between $\mathbf{x}(n)$ and $\mathbf{y}(n)$ is represented by the upper right submatrix $\boldsymbol{\Sigma}_{12}$ of $\boldsymbol{\Sigma}$ which may be expressed as

$$\begin{aligned}\boldsymbol{\Sigma}_{12} &= E[\mathbf{x}(n)\mathbf{y}^H(n)] \\ &= \mathbf{A}_x(\boldsymbol{\theta}_x)\boldsymbol{\Sigma}_{\alpha xy}\mathbf{A}_y^H(\boldsymbol{\theta}_y),\end{aligned} \tag{6}$$

where $\boldsymbol{\Sigma}_{\alpha xy}$ is the $k \times k$ cross–covariance matrix of the signals. $\boldsymbol{\Sigma}_{\alpha xy}$ is full rank because we have assumed the signals are independent. Here, we have allowed for the general case that the travelling of the signals from Array X to Array Y may involve some distortion of the signals such that $\boldsymbol{\alpha}_y$ may no longer be a merely distorted version of $\boldsymbol{\alpha}_x$. The matrices $\mathbf{A}_x$, $\mathbf{A}_y$ are full column rank because the angles of arrival are all distinct; therefore, rank $(\boldsymbol{\Sigma}_{12}) = k$, the number of signals. Hence, if we let $\mathbf{S}_{12}$ be the finite sample estimate of $\boldsymbol{\Sigma}_{12}$, then the detection problem can be treated as equivalent to the determination of the rank of $\mathbf{S}_{12}$.

According to (3), the noise component of $\boldsymbol{\Sigma}_{12}$ is zero. Therefore, in the presence of signal and noise, the rank of $\mathbf{S}_{12}$ may be estimated by testing the number of singular values of $\mathbf{S}_{12}$ which are significantly different from zero.

3 Canonical Correlations

3.1 Canonical correlations and variables in the population

Canonical correlations are discussed in e.g., [3][11]. Consider the random vector $\mathbf{z}(t)$ having variance-covariance matrix $\boldsymbol{\Sigma}$ as defined in Eqs. (4) and (5) respectively.

Theorem 1 *For $\boldsymbol{\Sigma}$ being defined as in (5) and rank$(\boldsymbol{\Sigma}_{12}) = k$, define the matrix*

$$\tilde{\boldsymbol{\Sigma}}_{12} = \boldsymbol{\Sigma}_{11}^{-1/2}\,\boldsymbol{\Sigma}_{12}\,(\boldsymbol{\Sigma}_{22}^{-1/2})^{H} \tag{7}$$

on which a singular value decomposition (SVD) [12] can be performed such that

$$\boldsymbol{\Sigma}_{11}^{-1/2}\,\boldsymbol{\Sigma}_{12}\,(\boldsymbol{\Sigma}_{22}^{-1/2})^{H} = \mathbf{U}_x\mathbf{P}\mathbf{U}_y \tag{8}$$

where $\mathbf{U}_x$ and $\mathbf{U}_y$ are unitary matrices of dimension $p \times p$ and $q \times q$ respectively, and $\mathbf{P}$ is a $p \times q$ matrix given by

$$\mathbf{P} = \begin{bmatrix} \tilde{\mathbf{P}}_k & \mathbf{O} \end{bmatrix} \tag{9}$$

with

$$\tilde{\mathbf{P}}_k = diag[\rho_1, \rho_2, \ldots, \rho_k, 0, \ldots, 0]. \tag{10}$$

Then, there exists a linear transformation $\mathbf{L}$ on $\mathbf{x}$ and a linear transformation $\mathbf{M}$ on $\mathbf{y}$, respectively defined as

$$\mathbf{L} \;=\; \boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{U}_x \tag{11}$$

$$\mathbf{M} = \mathbf{\Sigma}_{22}^{-1/2}\mathbf{U}_y \tag{12}$$

such that

$$E\begin{bmatrix}\mathbf{w}_x\\ \mathbf{w}_y\end{bmatrix}\begin{bmatrix}\mathbf{w}_x^H & \mathbf{w}_y^H\end{bmatrix} = \begin{bmatrix}\mathbf{L}^H & \mathbf{O}\\ \mathbf{O} & \mathbf{M}^H\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{L} & \mathbf{O}\\ \mathbf{O} & \mathbf{M}\end{bmatrix} = \begin{bmatrix}\mathbf{I}_p & \mathbf{P}\\ \mathbf{P}^H & \mathbf{I}_q\end{bmatrix} \tag{13}$$

where $\mathbf{w}_x = \mathbf{Lx}$ *and* $\mathbf{w}_y = \mathbf{My}$.

The proof of this theorem can be found in [8][11][21]. □

The $\{\rho_i\}$ are called *canonical correlation coefficients,* and the columns of $\mathbf{L}^H$ and $\mathbf{M}^H$ are called *canonical vectors.*

3.2 Sample canonical correlations

In practice, canonical correlations must be estimated from sampled data. Let $\mathbf{z}_1, ..., \mathbf{z}_N$ be N observations from a $(p+q)-$ variate Gaussian distribution $N(\mathbf{0}, \mathbf{\Sigma})$. It is well known [3] that the maximum likelihood estimate of $\mathbf{\Sigma}$ is

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i{\mathbf{z}_i}^H = \begin{bmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{bmatrix} \triangleq \mathbf{S}\ . \tag{14}$$

The maximum likelihood estimates γ_i [3] of the ρ_i are the singular values of

$$\tilde{\mathbf{S}}_{12} \triangleq \mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\left(\mathbf{S}_{22}^{-1/2}\right)^H = \mathbf{D}_1\mathbf{\Gamma}\mathbf{D}_2^H \tag{15}$$

where the right-hand term is the singular value decomposition on $\tilde{\mathbf{S}}_{12}$, $\mathbf{D}_1$ and $\mathbf{D}_2$ are unitary matrices, and

$$\mathbf{\Gamma} = [\bar{\mathbf{\Gamma}}_p\ \ \mathbf{O}], \tag{16}$$

$$\bar{\mathbf{\Gamma}}_p = diag(\ \gamma_1,\ \gamma_2,\ ...\ \gamma_p\ ), \tag{17}$$

where $1 \geq \gamma_1 \geq \gamma_2 \geq ... \geq \gamma_p \geq 0$ [15] . The γ_i are the sample canonical correlation coefficients.

We now present some additional theorems which are useful for later analysis.

Theorem 2 *In a narrow-band system as the SNR* $\to \infty$*, for* $N \geq k$*, the largest* k *sample canonical correlation coefficients approach unity, whereas the smallest* $p-k$ *coefficients approach zero.*

Proof: The proof of this theorem can be found in [8]. The two groups of coefficients mentioned in Theorem 2 are referred to as the *signal* and *noise* coefficients, respectively.

Theorem 3 *The canonical correlations are* invariant *with respect to the transformation*

$$\tilde{\mathbf{x}} = \mathbf{C}\,\mathbf{x} \tag{18}$$

$$\tilde{\mathbf{y}} = \mathbf{B}\,\mathbf{y} \tag{19}$$

where $\mathbf{C}$ *and* $\mathbf{B}$ *are nonsingular matrices.*

Proof: See [8].

Theorem 4 *The* γ_i *are consistent estimates of the true coefficients* ρ_i*.*

Proof: The proof follows directly from the fact that $\mathbf{S} \to \mathbf{\Sigma}$ as $N \to \infty$ in (14).

4 Detection by Testing the Sample Canonical Correlation Coefficients

The Neyman-Pearson detection criterion [19] constrains the probability of false alarm[2] to be less than or equal to a certain chosen value and designs a test to minimize the probability of missing. This leads to a likelihood ratio (LR) test in which the threshold T is chosen so that the constraint on the probability of false alarm is satisfied. In the following, we use the

[2]Here, we have generalized the meaning of *miss* and *false alarm* compared to their usual usage in the radar context. In the latter, we are dealing with only one signal. However, in this analysis, we are dealing with multiple signals, so a miss and false alarm in our context means declaring fewer or greater signals respectively than what are actually present.

Neyman-Pearson principle to develop a test procedure for the determination of the number of signals in unknown correlated noise.

Assume the true canonical correlation coeficients corresponding to the outputs of array X and array Y are arranged in descending order of magnitude[11]:

$$1 \geq \rho_1 \geq \rho_2 \geq \ldots \geq \rho_p \geq 0. \tag{20}$$

A similar relation is applied to the sample coefficients γ_i. We consider the following set of hypotheses:

$$\mathrm{H}_s : \quad \rho_1 \neq 0,\ \rho_2 \neq 0,\ \ldots,\ \rho_s \neq 0,\ \rho_{s+1} = \rho_{s+2} = \ \ldots\ = \rho_p = 0 \tag{21}$$

for $s = 0, 1 \ldots, p-1$. The index s is the number of signals under test. The detection problem is thus equivalent to a multiple hypothesis test to determine which value of s is most likely.

Here, the multiple hypothesis test is decomposed into p elementary binary hypothesis tests. In each elementary binary hypothesis test, a primary hypothesis H_s is tested against its alternative hypothesis H_{s+}, where H_{s+} denotes the hypothesis there are more than s signals present. Then the test is iterated for $s = 0, 1, \ldots, p-1$ until a primary hypothesis is accepted. The estimated number of signals $\hat{k}$ is assigned the value s.

We first establish a generalized LR defined as

$$\Upsilon_s = \frac{\max_{\Omega_s} \Lambda(\mathbf{Z}|\Omega_s)}{\max_{\Omega_{s+}} \Lambda(\mathbf{Z}|\Omega_{s+})}, \tag{22}$$

where $\Lambda(\mathbf{Z}|\Omega_s)$ is the likelihood function of the observed data matrix $\mathbf{Z}$ when the hypothesis H_s is true, and Ω_s is the parameter space of the observations constrained by the hypothesis H_s.[3] The following theorem provides a relationship between Λ and the canonical correlation coefficients:

Theorem 5 *Given that the observed data is Gaussian, the maximization of* $\Lambda(\mathbf{Z}|\Omega_s)$ *is given by*

$$\max_{\Omega_s} \Lambda(\mathbf{Z}|\Omega_s) \propto \left[\prod_{i=1}^{s} (1 - \gamma_i^2)\right]^{-N}, \tag{23}$$

where N *is the number of snapshots.*

[3]If the received signals are Gaussian, a sufficient parameter set for Ω_s is the true covariance matrix $\mathbf{\Sigma}$ defined by (5), corresponding to s signals.

The proof of this theorem was first published in [23], revised and corrected in [21].

Further, because Ω_p is sufficient to parameterize all possible outcomes in H_{s+}, from Theorem 5 it follows that

$$\max_{\Omega_{s+}} \Lambda(\mathbf{Z}|\Omega_{s+}) \equiv \max_{\Omega_p} \Lambda(\mathbf{Z}|\Omega_p) \propto \left[\prod_{i=1}^{p}(1-\gamma_i^2)\right]^{-N} . \tag{24}$$

Substituting (24) and (23) in (22), the LR can be re-written as

$$\Upsilon_s = \frac{\max_{\Omega_s} \Lambda(\mathbf{Z}|\Omega_s)}{\max_{\Omega_{s+}} \Lambda(\mathbf{Z}|\Omega_{s+})} = \prod_{i=s+1}^{p} (1-\gamma_i^2)^N . \tag{25}$$

The distribution of Υ_s can now be determined through the use of Bartlett's Approximation, [5][6] which can be stated as follows:

Bartlett's Approximation 1 [4] *When the received data is Gaussian and the hypothesis*

$$\mathrm{H}_s : \quad \rho_1 \neq 0,\ \rho_2 \neq 0,\ ...,\ \rho_s \neq 0,\ \rho_{s+1} = \ ... \ = \rho_p = 0 \tag{26}$$

is true, the asymptotic distribution of the statistic

$$C(s) = -[2N - (p+q+1)] \sum_{i=s+1}^{p} \ln(1-\gamma_i^2) \tag{27}$$

is approximately χ^2 *with* $2(p-s)(q-s)$ *degrees of freedom.*

Using this result, we can propose a hypothesis test for determining the number of signals under the assumption that all signals are independent. For an assumed number of signals s, we propose to use the following test:

$$C(s) = -[2N - (p+q+1)] \sum_{i=s+1}^{p} \ln(1-\gamma_i^2) \ \underset{\mathrm{H}_s}{\overset{\mathrm{H}_{s+}}{\gtrless}} \ T_s \tag{28}$$

The distribution on the left side of (28) can be obtained directly from Bartlett's Approximation. The threshold T_s should be set so that the allowable probability of false alarm, c, is achieved. It is thus given by

$$c = \int_{T_s}^{\infty} \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} y^{\frac{m-2}{2}} e^{-\frac{y}{2}} dy\ , \qquad s = 0, 1, 2, ..., p-1 \tag{29}$$

[4] The formula given here has been modified for complex data.

where $m = 2(p-s)(q-s)$. A procedure for solving (29) is given in [8]. Therefore, for a specified false alarm rate c, a set of thresholds $\{T_s, s = 0, 1, \ldots, p-1\}$ can be calculated beforehand and a sequential hypothesis testing procedure can be constructed using successively increasing values of s, starting from $s = 0$ until the LR test is satisfied.

We have seen from Theorem 3 that the true canonical correlation coefficients are invariant with respect to linear transformations in $\mathbf{x}(t)$ or $\mathbf{y}(t)$. Since changes in the noise or signal parameters can be modelled by such transformations, then asymptotically, the thresholds generated according to (29) are constant over all signal and noise conditions, and also, the performance of this proposed detection procedure is invariant to changes in either the noise or signal parameters.

We denote the probability of error as P_e. We have $P_e = P_M + P_F$, where P_M is the probability of a miss, and P_F is the probability of a false alarm. We now show that P_F is the dominant error mechanism at medium/high SNR. First, according to (27), $C(s)$ is χ^2–distributed for $s = k$, independent of SNR. On the other hand, we see from Theorem 2, that for a specific N, the signal coefficients approach unity as the SNR becomes large. Hence, with fixed N and increasing SNR, the probability of $C(s)$ exceeding the threshold remains approximately constant for $s = k$ (false alarm[5]), whereas the probability of $C(s)$ being smaller than the threshold for $s \leq k$ (miss) becomes small. Hence, P_F dominantes at medium/high SNR. By considering only this range of SNR, we see the CCT method is *quantitatively controllable.* The specified false-alarm rate at medium/high SNR is achieved by trading-off against the miss rate at low SNR.

The steps involved in the execution of the CCT method are outlined below:

1. Use the sample outputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ of array X and $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ of array Y to form the sample product matrices $\mathbf{S}_{11}$, $\mathbf{S}_{22}$, and $\mathbf{S}_{12}$ by

$$\mathbf{S}_{11} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^H, \qquad \mathbf{S}_{22} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^H, \qquad \mathbf{S}_{12} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \mathbf{y}_i^H . \tag{30}$$

2. Calculate the singular values $\gamma_1, \gamma_2, \ldots, \gamma_p$, of the transformed matrix

[5]Here, we implicitly assume the probability of false alarming one signal dominates the probability of false alarming one signal or more. This assumption is verified through simulations.

$$\tilde{\mathbf{S}}_{12} = \mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}(\mathbf{S}_{22}^{-1/2})^{H} \ . \tag{31}$$

3. For a specified false alarm rate P_F a set of threshold values $\{T_s\}$ can be pre-calculated according to χ^2 distributions with $2(p-s)(q-s), \quad s = 0, \ldots, p-1$, degrees of freedom, where s is the assumed number of signals under test.

4. Hypothesis testing: Denote the hypothesis that there are s signals by H_s. The testing starts from $s = 0$. For each s, the criterion

$$C(s) = -[2N - (p+q+1)] \sum_{i=s+1}^{p} \ln(1-\gamma_i^2) \tag{32}$$

is compared with the threshold value T_s . If the criterion is less than the threshold, we accept H_s, stop the testing, and assign the value of $\hat{k} = s$. Otherwise we reject the hypothesis, increase s by one, and continue the testing until a H_s is accepted.

5 Simulation Results

The performance of the CCT method was evaluated by several simulations. The noise was generated from a spatially-varying moving average (MA) process of order $v = 3$. The corresponding autocorrelation function is zero for distance lags greater than or equal to vd. Thus, for $\delta \geq 3d$ the method is tested under the ideal case where (3) is satisfied. (Later, we discuss the case where (3) is violated).

The MA coefficients used for the simulation are $[1.00, 0.90\jmath, -0.81]$. The corresponding spatial spectrum is shown in Fig. 2. The remaining parameters used for the simulation are specified in Table 1. The arrays are linear, uniformly–spaced, and oriented along the same line.

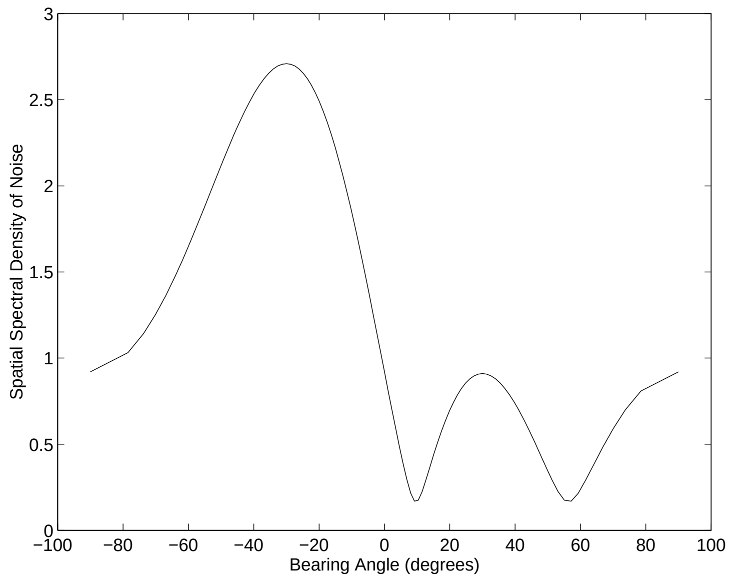

Figure 2: Spatial spectrum of the noise used in simulations.

Table 1 Parameters of the Simulation

Parameter		Value	Units
p	No. of sensors in array X	8	
q	No. of sensors in array Y	8	
d	Separation of sensors (uniformly spaced)	$\lambda/2$	metres
k	No. of signals	2	
N	No. of snapshots	100	
θ	Angles of arrival	$\pm 3.58^*$	degrees
δ	distance between the two arrays	$3d$	metres
N_t	No. of trials per SNR point	20,000	
SNR	Signal to noise ratio	-10 to 10	dB

* This separation corresponds to 0.5 standard beamwidths separation of an 8-element array.

The signal vector $\boldsymbol{\alpha}(n)$ at the nth snapshot is zero mean complex Gaussian such that

$$E\left[\boldsymbol{\alpha}(n)\boldsymbol{\alpha}^H(m)\right] = \delta_{mn}\sigma_s^2\mathbf{I} \tag{33}$$

where δ_{mn} is the Kronecker delta. The quantity σ_s^2 is defined for these simulations as $\sigma_s^2 = \frac{1}{k}$. Spatially filtered MA noise vector samples with coefficients given above, uncorrelated in time, were generated according to the method of [14] and added to the signal. The SNR is defined for the purposes of these simulations as

$$SNR = 10\log\left(\frac{\sigma_s^2}{\sigma_n^2}\right). \tag{34}$$

where σ_n^2 is the power of the noise before filtering.

In Table 2, the average error rates P_e of the CCT method obtained by simulation are compared with the corresponding theoretic values of false alarm rate, c in (29). These results were averaged over a wide range of coloured noise models, medium/high SNR values, and different signal locations. In these simulations, 195,000 trials are averaged over an SNR range of 6 to 15 dB. Over this range, P_e is almost equivalent to P_F. From these results, we see that the simulation error rates at medium/high SNR are in close agreement with the specified values of c.

Table 2 Comparison of False-Alarm Rates

Theoretical: c	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Simulation: P_e	0.93×10^{-1}	0.87×10^{-2}	0.83×10^{-3}	0.87×10^{-4}

The error performance of the CCT method vs. SNR is illustrated by the simulation results given in Fig. 3, corresponding respectively to a value of c of 10^{-1} and 10^{-2}. The performance of commonly–used white–noise methods (MDL and AIC [20]) are also plotted for comparison with CCT's performance. It is seen that the white-noise methods are not robust in coloured noise, whereas for CCT, we have good performance, as Fig. 3 verifies. In [8], simulation results for other values of c, down to 10^{-5} are presented. It is shown the CCT method also behaves well at this level of false–alarm rate. The robustness of the CCT method to variations in the noise characteristics can be demonstrated as follows. A direct consequence of (3) and Theorem 4 for a specified value of p and finite SNR is that the $\gamma_i \rightarrow 0$ as $N \rightarrow \infty, i = k+1, \ldots, p$; also, the $\gamma_i, i = 1, \ldots k$ are finite, regardless of the noise characteristic.

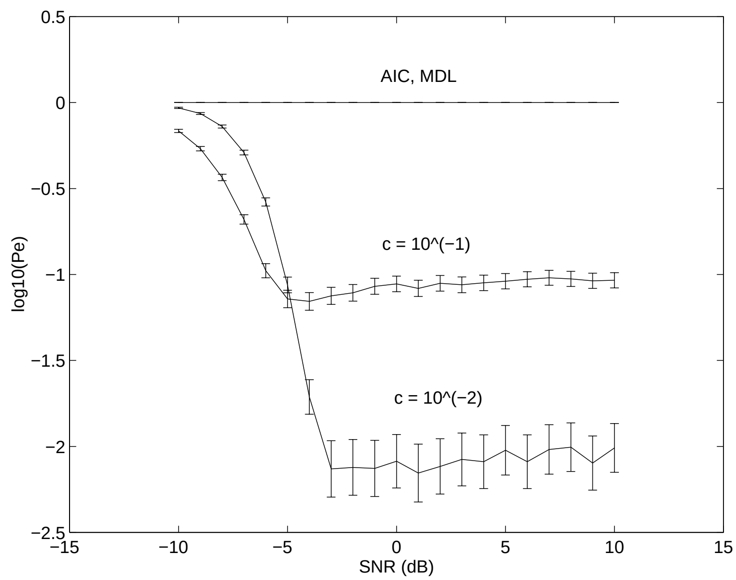

Figure 3: Probability of error for the CCT method, for theoretical false-alarm rates $c = 10^{-1}, 10^{-2}$. Error bars indicate approximate 95 % confidence intervals for the simulation results.

It follows that $P_M \to 0$ and $P_F \to c$; hence, $P_e \to c$. Since c is arbitrary, the performance of the CCT method can in principle be made arbitrarily good for large enough N. In contrast, the white-noise methods demonstrate a significant error floor in coloured noise, regardless of N. In practice, this analysis means that if the performance is not good enough for a certain situation, it can be made better for the CCT case by increasing N. On the other hand, for the white noise methods, there can exist coloured noise environments for which there is no value of N resulting in acceptable performance. However, it must be pointed out [8], that in white–noise environments, the white–noise methods show a modest performance advantage over CCT for the same total number of sensors. This is the penalty paid for flexibility with respect to the noise characteristics. Further comments and more detailed simulation results are available in [8].

Up to now, an MA model was used to represent the noise. This results in the ideal situation where the noise correlation between the two arrays can be made zero. What about the case where some degree of noise correlation exists between the two arrays, as happens for example when the background noise is spatially autoregressive (AR)? An experiment was performed where a spatially-varying AR noise model was used, with the parameters of the AR process adjusted so that the largest noise correlation between the two arrays was 0.35.

The degradation in performance in this case, relative to when the correlation between the two arrays is zero, is 2dB. This figure does not change much with changes in the AR noise model. We therefore see that the CCT method tends to be robust under violations of assumption (3).

A comparison of floating-point operations required for the AIC or MDL method vs. the CCT method reveals that the CCT technique requires about half the number of operations relative to AIC or MDL, for the parameter set chosen, for the same number of total array sensors in each case. This result is due in part to the fact that the matrices on which the singular value or eigen decompositions are performed (i.e., the operation which dominates the computation) are smaller for the CCT case.

6 Conclusions

A new approach to detecting number of signals in unknown noise environment has been presented. The approach is based on applying canonical correlation analysis to the outputs of two spatially separated arrays, where nominally, the noise between the arrays is uncorrelated. The new method, called the canonical correlation technique (CCT), has been shown to perform very well in a wide variety of coloured noise situations.

The method is developed by showing that a generalized likelihood ratio for testing hypothesis H_s against H_{s^*} may be expressed as a function $C(s)$ of the canonical correlation coefficients $\gamma_s \ldots \gamma_p$. $C(s)$ is a χ^2 random variable. A set of thresholds is thus generated according to the number of degrees of freedom corresponding to each hypothesis. These thresholds are calculated so that the probability of them being exceeded corresponds to a specified false alarm error rate. A sequential hypothesis test is performed; the first instance where $C(s)$ exceeds the threshold corresponds to the most likely number of signals. The medium/high SNR performance is maintained at this desired false-alarm error rate, by trading-off against performance in the low-SNR range.

The need for a method which is robust in the presence of coloured noise has been made apparent, since the performance of techniques developed assuming white noise has been shown to be unsatisfactory in coloured background noise. We have shown the performance of CCT for varying degrees of noise colour remains relatively constant at a satisfactory level of performance. Also, the method is robust when significant noise correlations exist between the two arrays in contradiction to (3).

Thus, it may be concluded that the CCT method presented in this paper gives an attractive, simple, statistically rigorous approach for detecting number of signals in unknown noise environments.

7 Acknowledgements

The authors are grateful for useful discussions and input from Dr. Qiang Wu, now of Celwave Inc., Corvallis, Oregon, USA. Financial support from the following organizations is gratefully acknowledged: 1) Telecommunications Research Institute of Ontario, 2) the Natural Sciences and Engineering Research Council of Canada, and 3) COM DEV Ltd. of Cambridge, Ont. Canada.

References

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automatic Control*, vol.19, No.6, pp.716-723, 1974.

[2] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. of Math. Stat.*, vol. 34, pp.122-148, 1963.

[3] T. W. Anderson, *An introduction to Multivariate Statistical Analysis*, second edition, John Wiley & Sons, Inc, 1984.

[4] M. S. Bartlett, "A note on the multiplying factors for various χ^2 approximations," *J. Roy. Stat. Soc.,* ser.B, vol.16, pp.296-298, 1954.

[5] M. S. Bartlett, "Further aspects of the theory of multiple regression," *Proc. Camb. Phil. Soc.*, vol.34, p.33, 1938.

[6] M. S. Bartlett, "A note on tests of significance in multivariate analysis," *Proc. Camb. Phil. Soc.*, vol.35, p.180, 1939.

[7] W. G. Chen, K. M. Wong, and J. P. Reilly, "Detection of number of signals: a predicted eigen-threshold approach," *IEEE Trans. Signal Processing*, Vol. 39, No. 5, pp.1088-1098, May, 1991.

[8] W.G. Chen, "Detection of the number of signals in array signal processing", Ph.D. thesis, Dept. of Electrical and Computer Engineering, McMaster University, Hamilton, Ont., Canada, June, 1991.

[9] W. Chen, J.P. Reilly, K.M. Wong, "Application of canonical correlation analysis in detection in the presence of spatially correlated noise", SPIE Int'l Symposium on Optical Applied Science and Engineering- Advanced Signal Processing Algorithms, Architectures, and Implementations, II, 21-26 July, 1991, San Diego, CA.

[10] J.J. Fuchs, "Estimation of the number of signals in the presence of unknown correlated sensor noise", *IEEE Trans. Signal Processing*, Vol. 40, No. 5, pp. 1053-1061, 1992.

[11] A. M. Kshirsagar, *Multivariate Analysis*, Marcel Dekker, Inc., New York, 1972.

[12] P. Lancaster and M. Tismenetsky, "The Theory of Matrices", 2nd. Ed., Academic Press, 1985.

[13] D. N. Lawley, "Tests of significance of the latent roots of the covariance and correlation matrices," *Biometrika*, vol.43, pp.128-136, 1956.

[14] J. P. Le Cadre, "Parametric methods for spatial signal processing in the presence of unknown colored noise fields," *IEEE Trans. ASSP*, vol.37, No.7, pp.965-983, 1989.

[15] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc., New York, 1982.

[16] Y. Omichi, "An experimental study of covariance functions of reverberation from a lake surface", *ARL Tech. Report*, ARL-TR-75-25, ARL, University of Texas at Austin, May, 1975.

[17] J. Rissanen, "Modeling by shortest data description,"*Automatica*, vol.14, pp.465-471, 1978.

[18] G. Schwartz, "Estimating the dimension of a model," *Ann. Stat.*, vol.6, pp.461-464, 1978.

[19] H.L. van Trees, "Detection, estimation, and modulation theory", Pt. I, John Wiley and Sons, 1968.

[20] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. ASSP*, vol.33, No.2, pp.387-392, 1985.

[21] K.M. Wong, Q. Wu, and P. Stoica, "Generalized correlation decomposition applied to array processing in unknown noise environments", in "Advances in Spectrum Analysis and Array Processing", Vol. III, S. Haykin, Ed., Prentice Hall, 1995.

[22] Q. Wu and D.R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding", *IEEE Trans. Signal Processing*, Vol. 39, No. 8, pp.1848-1857, Aug. 1991.

[23] Q.T. Zhang and K.M. Wong, "Information theoretic criteria for the determination of the number of signals in spatially correlated noise", *IEEE Trans on SP*, Vol. 41, No. 4, pp. 1652-1663, April, 1993.