Exact Bayesian Test for a Common Rank-One Component in White Noise

Songsri Sirianunpiboon*, Stephen D. Howard*, Douglas Cochran†
*Defence Science and Technology Organisation
PO Box 1500, Edinburgh 5111, Australia
†School of Mathematical and Statistical Sciences
Arizona State University, Tempe AZ 85287–5706 USA

***Abstract*—Unknown and arbitrary rank-one signals are collected by two arrays of sensors. This paper constructs an *exact* invariant Bayesian detector for deciding if the two sets of data are generated by the same rank-one signal. This detector is compared to the generalized likelihood ratio test (GLRT) and shown in simulation to give significantly better performance.**

I. Introduction

Consider an unknown signal of known rank-1 received at a set of M spatially distributed sensors. The sensor channels are suitably sampled to obtain M complex measurement vectors, each of length N. These are organised into an $N \times M$ data matrix $\mathbf{X}$, each column $\boldsymbol{x}_j$ of which represents N samples of the noisy signal collected at the j^{th} sensor. The data $\mathbf{X}$ is modeled as

$$\mathbf{X} = \boldsymbol{s}\boldsymbol{a}^\dagger + \boldsymbol{\nu}. \tag{1}$$

where $\boldsymbol{s}$ is a signal vector in $\mathbb{C}^N$, $\boldsymbol{a}$ is vector in $\mathbb{C}^M$, whose elements a_j are the complex amplitude of the signal $\boldsymbol{s}$ from sensor j. Beyond these properties, $\boldsymbol{a}$ and $\boldsymbol{s}$ are unknown. The noise matrix $\boldsymbol{\nu}$ is Gaussian distributed with zero mean and covariance $2\sigma^2 I_{MN}$ where I_{MN} is the $MN \times MN$ identity matrix. In this work we assume that the noise variance σ^2 is known. As $\boldsymbol{a}$ and $\boldsymbol{s}$ are unknown the model represents a rank-one component in white noise.

This model has been considered by Besson et al. [1] in constructing an invariant generalized likelihood ratio detector a rank-one component in white noise. It has recently found application in passive radar detection where access to direct path signals is limited [2], [3], [4]. Here we consider a related problem. Unknown and arbitrary rank-one components are collected by two arrays of sensors. The problem is to decide if received signal across the two arrays are generated by a single rank-one component, or by two distinct rank-one components. This problem is related to matched subspace detection [5], but here, rather than matching known subspace to a noisy data record, the question is whether two noisy data records contain signals from a common subspace.

In this paper we derive an exact Bayesian detector for this problem using non-informative priors for the amplitudes $\boldsymbol{a}$ and the subspaces $\langle \boldsymbol{s} \rangle$. Somewhat surprisingly, the integral over all rank-one subspaces can explicitly computed. For comparison, the GLRT for this problem is also computed and the performance compared with the Bayesian test.

II. Problem Formulation

In this paper we consider the following problem. Two data collections are made according to the model (1).

$$\mathbf{X}^{(1)} = \boldsymbol{s}_1\boldsymbol{a}_1^\dagger + \boldsymbol{\nu}_1$$
$$\mathbf{X}^{(2)} = \boldsymbol{s}_2\boldsymbol{a}_2^\dagger + \boldsymbol{\nu}_2$$

where the number of channels for the first collection is m_1 and that for the second is m_2, that is, $\mathbf{X}^{(1)} \in \mathbb{C}^{N \times m_1}$ and $\mathbf{X}^{(2)} \in \mathbb{C}^{N \times m_2}$. Both $\boldsymbol{a}^{(1)}$ and $\boldsymbol{s}^{(1)}$, as well as $\boldsymbol{a}^{(2)}$ and $\boldsymbol{s}^{(2)}$ are unknown. The problem is then to decide if unit vectors $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ define the same one-dimensional subspace.

First some redundancy in the parameterization of the problem need to be addressed. The signal $\boldsymbol{s}$ is only defined up to multiplication by a complex number, as any such complex multiplier can be absorbed into the complex amplitudes $\boldsymbol{a}$. Thus, a rank-1 signal really corresponds to a line or one-dimensional subspace in $\mathbb{C}^N$ and the parameter $\boldsymbol{s}$ is a label for this subspace which is denoted by $\langle \boldsymbol{s} \rangle$. The set of lines in $\mathbb{C}^N$ is called the complex projective space and denoted by $\mathbb{C}P^{N-1}$. Some of the redundancy in $\boldsymbol{s}$ is removed by taking it to be unit vector which we do from this point on. A point in $\mathbb{C}P^{N-1}$ can be uniquely associated with rank 1 orthogonal projection P, that is, an $N \times N$ hermitian matrix satisfying the conditions $P^2 = P$ and $\mathrm{Tr}(P) = 1$. Thus, an orthogonal projection

$$P_{\boldsymbol{s}} = \boldsymbol{s}\boldsymbol{s}^\dagger$$

where † denotes hermitian conjugate, can be uniquely associated with each signal subspace.

Thus our detection problem compares

$$\begin{aligned} H_1 &: \langle \boldsymbol{s}^{(1)} \rangle = \langle \boldsymbol{s}^{(2)} \rangle \\ H_0 &: \langle \boldsymbol{s}^{(1)} \rangle \neq \langle \boldsymbol{s}^{(2)} \rangle. \end{aligned} \tag{2}$$

Under H_1, the two set of measurements $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ constitute a single set of measurements, which is denoted by $\mathbf{X} = \mathbf{X}^{(1)} \cup \mathbf{X}^{(2)}$ are derived from a common subspace $\boldsymbol{s}$, with the number of sensors $M = m_1 + m_2$ and the amplitude vector is $\boldsymbol{a} = \boldsymbol{a}^{(1)} \oplus \boldsymbol{a}^{(2)}$. Under the null hypothesis H_0, the two set of measurements are generated from two distinct rank-1 subspaces.

Under hypothesis H_1, the joint probability density function (pdf) of $\mathbf{X}$ conditioned on the parameters $\boldsymbol{s}$ and $\boldsymbol{a}$,

$$p(\mathbf{X}|\boldsymbol{a},\boldsymbol{s}) = (2\pi\sigma^2)^{-MN} \times \exp\left(-\frac{1}{2\sigma^2}\mathrm{Tr}\left((\mathbf{X}-\boldsymbol{s}\boldsymbol{a}^\dagger)^\dagger(\mathbf{X}-\boldsymbol{s}\boldsymbol{a}^\dagger)\right)\right) \tag{3}$$

Under H_0, the pdf of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ conditioned on the parameters $\Theta^{(i)} = \{\boldsymbol{s}^{(i)}, \boldsymbol{a}^{(i)}\}, i = 1, 2,$

$$p(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}|\Theta^{(1)}, \Theta^{(2)}) = \prod_{i=1}^{2}(2\pi\sigma^2)^{-Nm_i} \times \exp\left(-\frac{1}{2\sigma^2}\mathrm{Tr}\left((\mathbf{X}^{(i)}-\boldsymbol{s}^{(i)}\boldsymbol{a}^{(i)\dagger})^\dagger(\mathbf{X}^{(i)}-\boldsymbol{s}^{(i)}\boldsymbol{a}^{(i)\dagger})\right)\right) \tag{4}$$

III. Generalized Likelihood Ratio Test

In this section the generalized likelihood ratio test is derived for the purpose of comparison with the Bayesian detector. The GLRT statistic for the detection problem (2) is obtained by considering the ratio of maximal values of the joint likelihood functions of the unknown parameters under hypothesis H_1 and H_0 as follows:

$$\frac{\max p(\mathbf{X}|\boldsymbol{a},\boldsymbol{s})}{\max\prod_{i=1}^{2} p(\mathbf{X}^{(i)}|\boldsymbol{a}^{(i)},\boldsymbol{s}^{(i)})} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

Under the hypothesis H_1, maximizing the joint pdf (3) with respect to $\boldsymbol{a}$, yields the maximum likelihood estimate $\hat{\boldsymbol{a}}$ as

$$\hat{\boldsymbol{a}} = \boldsymbol{s}^\dagger\mathbf{X} \tag{5}$$

Substituting $\hat{\boldsymbol{a}}$ into (3), and writing

$$\mathbf{Z} = \frac{\mathbf{X}\mathbf{X}^\dagger}{\mathrm{Tr}(\mathbf{X}^\dagger\mathbf{X})} \tag{6}$$

we obtain the marginalized likelihood function

$$\ell(\boldsymbol{s}, H_1|\mathbf{X}) = (2\pi\sigma^2)^{-MN}\exp\left(-q(1-\boldsymbol{s}^\dagger\mathbf{Z}\boldsymbol{s})\right) \tag{7}$$

where

$$q = \frac{1}{2\sigma^2}\mathrm{Tr}(\mathbf{X}^\dagger\mathbf{X}). \tag{8}$$

The likelihood function (7) is maximized when $\boldsymbol{s}^\dagger\mathbf{Z}\boldsymbol{s}$ is maximized. According to the Rayleigh principle ([6], p.176), the maximum likelihood (ML) estimate $\hat{\boldsymbol{s}}$ is the eigenvector of $\mathbf{Z}$ corresponding to its maximum eigenvalue. Substituting $\hat{\boldsymbol{s}}$ into (7) and let λ be the largest eigenvalue of $\mathbf{Z}$, we obtain

$$\ell(H_1|\mathbf{X}) = (2\pi\sigma^2)^{-MN}\exp(-q(1-\lambda)) \tag{9}$$

Similarly, under hypothesis H_0,

$$\ell(H_0|\mathbf{X}) = (2\pi\sigma^2)^{-NM}\prod_{i=1}^{2}\exp(-q^{(i)}(1-\lambda^{(i)})) \tag{10}$$

where $\lambda^{(i)}, i = 1, 2$ are the largest eigenvalues of $\mathbf{Z}^{(i)}$ and $\mathbf{Z}^{(i)}$ and $q^{(i)}$ are defined in (6) and (8) with corresponding $\mathbf{X}^{(i)}$. The GLR between H_1 and H_0 is

$$\gamma_G = \frac{\exp(-q(1-\lambda))}{\prod_{i=1}^{2}\exp(-q^{(i)}(1-\lambda^{(i)}))} \tag{11}$$

From (8), we have $q = q^{(1)} + q^{(2)}$, then the generalized log likelihood ratio test statistic is

$$\log\gamma_G = (q^{(1)} + q^{(2)})\lambda - q^{(1)}\lambda^{(1)} - q^{(2)}\lambda^{(2)}. \tag{12}$$

IV. Bayesian Likelihood Ratio

In this section we derive a Bayesian posterior probability for the hypotheses test (2). In the Bayesian approach, instead of maximising over the nuisance parameters as in the GLRT, each parameter is marginalized out of the likelihood functions in which it appears by integration with respect to a corresponding prior probability distribution.

The essential element in Bayesian framework is choosing an appropriate prior distribution to represent our prior knowledge of the unknown, or nuisance, parameters, $p(\boldsymbol{a}, \boldsymbol{s})$, then marginalizing with respect to this prior. The prior for the nuisance parameters $\boldsymbol{a}$ and $\boldsymbol{s}$ is taken to be as proper and as non-informative as possible corresponding to little or no prior knowledge of the parameters. We consider prior of $\boldsymbol{a}$ and $\boldsymbol{s}$ to be independent and introduce proper prior for $\boldsymbol{a}$ that is unitarily invariant and non-informative in an appropriate limiting sense [7],

$$p(\{\boldsymbol{a}\}|\sigma^2, \beta^2) = \frac{1}{(2\pi\sigma^2\beta^2)^M}\exp\left(-\frac{\boldsymbol{a}^\dagger\boldsymbol{a}}{2\sigma^2\beta^2}\right), \tag{13}$$

where the dimensionless parameter β^2 controls the a priori maximum signal to noise ratio. Then under the hypothesis H_1, the posterior distribution of $\boldsymbol{s}$, conditioned on $\mathbf{X}$ is

$$\begin{aligned} p(\boldsymbol{s}|\mathbf{X}, H_1) &= (2\pi\sigma^2)^{-NM}(2\pi\sigma^2\beta^2)^{-M} \\ &\times\exp\left(-\frac{1}{2\sigma^2}\mathrm{Tr}(\mathbf{X}^\dagger\mathbf{X})\right)\exp\left(\frac{\alpha}{2\sigma^2}\boldsymbol{s}^\dagger\mathbf{X}\mathbf{X}^\dagger\boldsymbol{s}\right) \\ &\times\int\exp\left(-\frac{1}{2\alpha\sigma^2}\mathrm{Tr}\left((\boldsymbol{a}-\alpha\boldsymbol{s}^\dagger\mathbf{X})^\dagger(\boldsymbol{a}-\alpha\boldsymbol{s}^\dagger\mathbf{X})\right)\right)d\boldsymbol{a} \\ &= (2\pi\sigma^2)^{-NM}(2\pi\sigma^2\beta^2)^{-M}(2\pi\sigma^2\alpha)^M \\ &\times\exp\left(-\frac{1}{2\sigma^2}\mathrm{Tr}(\mathbf{X}^\dagger\mathbf{X})\right)\exp\left(\frac{\alpha}{2\sigma^2}\boldsymbol{s}^\dagger\mathbf{X}\mathbf{X}^\dagger\boldsymbol{s}\right) \end{aligned} \tag{14}$$

where $\alpha = \beta^2/(1+\beta^2)$. This can be simplified to

$$p(\boldsymbol{s}, \mathbf{X}|H_1) = \frac{\exp(-q)}{(2\pi\sigma^2)^{NM}(1+\beta^2)^M}\exp(\alpha q\boldsymbol{s}^\dagger\mathbf{Z}\boldsymbol{s})$$

where $\mathbf{Z}$ and q are defined in (6) and (8), respectively.

If we have no prior information about the subspace $\langle\boldsymbol{s}\rangle$ then the appropriate prior distribution is the uniform, or unitary invariant probability distribution on $\mathbb{C}P^{N-1}$ [8], [7], which we denote by $d\mu(\boldsymbol{s})$. This distribution is unique. With this prior, the marginalized likelihood for the data $\mathbf{X}$ given the hypothesis H_1 is

$$p(\mathbf{X}|H_1) = \frac{\exp(-q)}{(2\pi\sigma^2)^{NM}(1+\beta^2)^M}\int_{\mathbb{C}P^{N-1}}\exp(\alpha q\boldsymbol{s}^\dagger\mathbf{Z}\boldsymbol{s}))d\mu(\boldsymbol{s})$$

Similarly, under H_0, $\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}$ are marginalized with respect to the corresponding proper invariant prior as given in (13), then the posterior distribution of $\boldsymbol{s}^{(1)}$ and $\boldsymbol{s}^{(2)}$ conditioned on $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ is

$$\begin{aligned} &p(\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)}|\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \\ &= \frac{\exp(-q)}{(2\pi\sigma^2)^{NM}(1+\beta^2)^M}\prod_{i=1}^{2}\exp(\alpha q^{(i)}\boldsymbol{s}^{(i)\dagger}\mathbf{Z}^{(i)}\boldsymbol{s}^{(i)})) \end{aligned} \tag{15}$$

The marginalized likelihood for the data given H_0

$$p(\mathbf{X}^{(1)},\mathbf{X}^{(2)}|H_0) = \frac{\exp(-q)}{(2\pi\sigma^2)^{NM}(1+\beta^2)^M} \times \prod_{i=1}^{2} \int_{\mathbb{C}P^{N-1}} \exp(\alpha q^{(i)} \boldsymbol{s}^{(i)\dagger}\mathbf{Z}^{(i)}\boldsymbol{s}^{(i)})\, d\mu(\boldsymbol{s}^{(i)}) \tag{16}$$

The Bayesian likelihood ratio between H_1 and H_0, in the limit as $\beta \to \infty$ (completely non-informative prior for $\boldsymbol{a}$), is given by

$$\gamma_B = \frac{\int_{\mathbb{C}P^{N-1}} \exp(q\boldsymbol{s}^\dagger\mathbf{Z}\boldsymbol{s})\, d\mu(\boldsymbol{s})}{\prod_{i=1}^{2} \int_{\mathbb{C}P^{N-1}} \exp(q^{(i)}\boldsymbol{s}^{(i)\dagger}\mathbf{Z}^{(i)}\boldsymbol{s}^{(i)})\, d\mu(\boldsymbol{s}^{(i)})} \tag{17}$$

If $M < N$ the matrix $\mathbf{Z}$ in (6). has rank-M and its M non-zero eigenvalues satisfy $0 < \lambda_1 \leq \cdots \leq \lambda_M \leq 1$. Using the integral formula given in [9], we have

$$\int_{\mathbb{C}P^{N-1}} e^{q\boldsymbol{s}^\dagger\mathbf{Z}\boldsymbol{s}}\, d\mu(\boldsymbol{s}) = \sum_{\ell=1}^{M}(L^{-1})_{M\ell}\lambda_\ell^{M-1}\,{}_1F_1(1;N;q\lambda_\ell). \tag{18}$$

Computing the integrals in (17), we obtain the exact Bayesian likelihood ratio, for non-informative priors on the nuisance parameters

$$\gamma_B = \frac{\sum_{\ell=1}^{M}(L^{-1})_{M\ell}\lambda_\ell^{M-1}\,{}_1F_1(1;N;q\lambda_\ell)}{\prod_{i=1}^{2}\sum_{\ell=1}^{m_i}(L_i^{-1})_{m_i\ell}\lambda_\ell^{(i)m_i-1}\,{}_1F_1(1;N;q_i\lambda_\ell^{(i)})} \tag{19}$$

where L^{-1} is the lower triangular matrix with elements [10]

$$(L^{-1})_{ij} = \begin{cases} 1 & \text{if } i=j=1 \\ 0 & \text{if } i<j \\ \prod_{\substack{\ell=1\\ \ell\neq j}}^{i} 1/(\lambda_j-\lambda_\ell) & \text{otherwise} \end{cases} \tag{20}$$

and ${}_1F_1(a;b;x)$ is a confluent hypergeometric function defined in [11],

$${}_1F_1(a;b;x) = \sum_{n=0}^{\infty}\frac{(a)_n x^n}{(b)_n n!} \tag{21}$$

Having the exact Bayesian likelihood ratio for this hypothesis test means that we can use the optimal Bayesian technique for setting the decision threshold for the test rather than Neyman-Pearson technique of choosing the threshold based on false alarm probabilities. Let $C_{ij}, i,j = 0,1$ denote the loss function for our decision problem; i.e., C_{ij} is the loss incurred by deciding in favour of H_i when H_j is true and take $C_{00} = C_{11} = 0$. The Bayesian decision threshold is set to minimize the expected loss and is given by

$$\gamma_B \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p(H_1)C_{10}}{p(H_0)C_{01}} \tag{22}$$

where $p(H_1)$ and $p(H_0)$ are the prior probabilities of H_1 and H_0 and γ_B is given by (19).

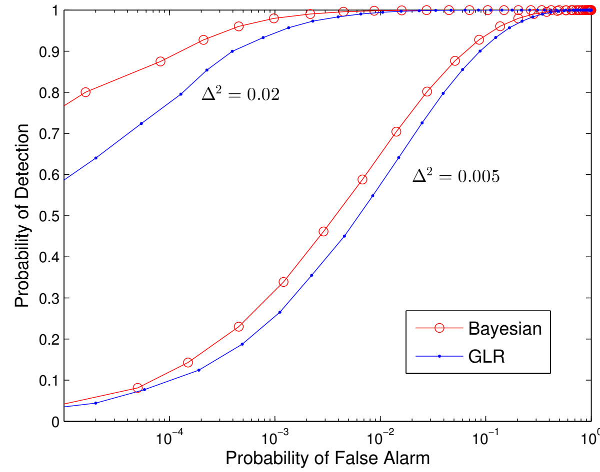

Fig. 1. ROC curves for signals with -5dB SNR and $N = 64$.

V. Performance

In this section we numerically test the performance of the Bayesian likelihood ratio test statistic (19) for deciding whether the two set of signal belong to the same rank-1 signal subspace. Here we will just consider the ROCs in a number of typical cases. The performance is compared to that of the GLRT statistic (12). The simulation was set up for a set of $M = 12$ and $m_1 = m_2 = 6$ channels. Two scenarios were considered, first, with a total signal SNR across the 6 channels of -5dB and $N = 64$ samples. In the second case, the total signal SNR was -8dB and $N = 128$samples. In the simulations first 6 channels randomly chosen signals for a fixed randomly chosen subspace $\langle \boldsymbol{s}^{(1)}\rangle$. The amplitudes $|\boldsymbol{a}_i^{(1)}|^2$ are chosen from a scaled Dirichlet distribution so that the total energy of this signals is fixed. The phases of the $\boldsymbol{a}_i^{(1)}$ are independent and uniformly distributed. The second 6 channels contain randomly chosen signals from either the same subspace as the first 6 channels or a randomly chosen chosen signal from a randomly chosen alternate subspace. The alternate subspaces are generated by adding zero mean IID white Gaussian noise with variance $\Delta^2\sigma^2$, to $\boldsymbol{s}_1$ and renomalizing. The dimensionless parameter Δ^2 is a measure of how different the alternate subspaces are from $\langle \boldsymbol{s}^{(1)}\rangle$, relative to the receiver noise. In each scenario, two values of Δ^2, 0.02 and 0.005 are considered. The ROC curves for both scenarios are shown in Figure 1 and Figure 2. It is seen that in all cases the exact Bayesian detector derived in this paper has significantly better performance than the GLRT detectors.

VI. Conclusion

An exact Bayesian likelihood ratio test statistic is derived for deciding if two multichannel data collections contain signals from a common subspace in the presence of white noise. The derivation of the detector involves an integral over complex projective space $\mathbb{C}P^{N-1}$ for which we have an exact result. Simulations show that the Bayesian test outperforms the GLRT by a significant margin, particularly at low probabilities of false alarm.

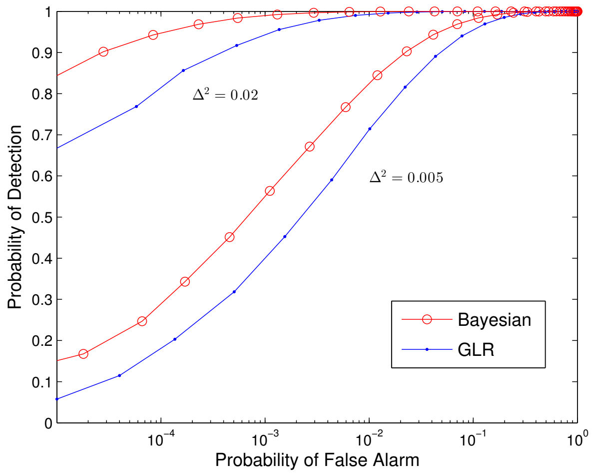

Fig. 2. ROC curves for signals with -8dB SNR and $N = 128$.

References

[1] O. Besson, S. Kraut, and L. Scharf, "Detection of an unknown rank-one component in white noise," *Signal Processing, IEEE Transactions on*, vol. 54, no. 7, pp. 2835–2839, July 2006.

[2] K. S. Bialkowski, I. V. L. Clarkson, and S. D. Howard, "Generalized canonical correlation for passive multistatic radar detection," in *Proceedings of the IEEE Statistical Signal Processing Workshop*, 2011, pp. 417–420.

[3] K. S. Bialkowski and I. L. Clarkson, "Passive radar signal processing in single frequency networks," in *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012, pp. 199–202.

[4] D. Hack, L. Patton, B. Himed, and M. Saville, "Centralized passive mimo radar detection without direct-path reference signals," *Signal Processing, IEEE Transactions on*, vol. 62, no. 11, pp. 3013–3023, June 2014.

[5] L. Scharf and B. Friedlander, "Matched subspace detectors," *Signal Processing, IEEE Transactions on*, vol. 42, no. 8, pp. 2146–2157, Aug 1994.

[6] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[7] S. Sirianunpiboon, S. D. Howard, and D. Cochran, "Maximum a posteriori estimation of signal rank," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2014.

[8] ——, "Multiple-channel detection of signals having known rank," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2013, pp. 6536–6540.

[9] S. D. Howard, S. Sirianunpiboon, and D. Cochran, "Integration over the Grassmannian manifolds," *in preparation*, 2015.

[10] L. R. Turner, "Inverse of the Vandermonde matrix with applications," Lewis Research Center, Cleveland, Ohio, Tech. Rep. NASA TN D-3547, 1966.

[11] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, graphs, and Mathematical Tables*. New york, Denver, 1965.