Minimum Mean Square Distance Estimation of Subspaces in presence of Gaussian sources with application to STAP detection

R. Ben abdallah 1, A. Breloy 2, M. N. El Korso 3, D. Lautru 4, H. Hafdallah Ouslimani 5
LEME, Universite Paris Ouest, 50, rue de Sevres, 92410 Ville dAvray, France
E-mail: 1rayen.benabdallah@u-paris10.fr, 2abreloy@u-paris10.fr, 3m.elkorso@u-paris10.fr, 4dlautru@u-paris10.fr, 5houslima@u-paris10.fr

Abstract. We consider the problem of signal subspace estimation from a given sample set. We rely on the Bayesian framework developed in [1] to obtain minimum mean square distance (MMSD) estimators, which minimize the expected distance between the true projection matrix $UU^T$ and its estimate $\hat{U}\hat{U}^T$. In this work, we extend the estimators of [1] to the context of linear model with Gaussian sources, with respectively Bingham and von Mises Fisher priors for the basis $U$. Numerical simulations are given in order to assess the performance of the proposed estimator. The interest of the considered approach is finally illustrated on a real data set, for a Space Time Adaptive Processing (STAP) application.

Index Term Subspace estimation, minimum mean square distance estimation, Bayesian setting, Bingham distribution, von Mises Fisher distribution, STAP detection.

1. Introduction

Subspace estimation is an ubiquitous issue in signal processing. It aims at estimating the low dimensional subspace where information (usually signals of interest) lies in. This problem arises in numerous applications, such as in PCA, DOA estimation [2], interference cancellation [11] or signal detection [10].

In the literature, a widely used model is the linear model where the data matrix is modeled as $Y = US + N$, in which $U$ spans the subspace of interest, $S$ is the signal of interest and $N$ is an additive noise. The addressed problem is to estimate $U$ from the observations $Y$. The most usual mean of estimating this subspace $U$ is the singular value decomposition (SVD) of the sample covariance matrix (SCM), i.e., taking the largest eigenvectors of the SCM. This provides usually an accurate estimator of range space $R(U)$ of $U$ for high signal to noise ratio (SNR) and/or for large number of sample support contexts. However, this method shows its limits outside these asymptotic regimes. A possible solution to overcome this issue is to incorporate some prior knowledge into the estimation process, as proposed in [1], [5]. For example, in a Bayesian setting, the minimum mean square error (MMSE) [5] which minimizes the average Euclidean distance between the true subspace $U$ and its estimate $\hat{U}$, and the minimum mean square distance (MMSD) [1] which aims to minimize the distance between the true projection matrix $UU^T$ and its estimate $\hat{U}\hat{U}^T$, were proposed. More specially, in [1] the MMSD is derived for both linear and covariance models with uniform prior distributions for $S$. In this paper, we extend the work of [1] for the linear model, for the case where, the signal model $S$ is assumed to follow a Gaussian law and for $U$ which has respectively a Bingham and a von Mises Fisher prior distributions as [1].

In this work, we extend the estimators of [1] to the context of linear model with Gaussian sources, with respectively Bingham and von Mises Fisher priors for the basis $U$. Numerical simulations are given in order to assess the performance of the proposed estimator. The interest of the considered approach is finally illustrated on a real data set, for a Space Time Adaptive Processing (STAP) application.

2. Theory background and model setup

Along this paper, we denote by $Y \in \mathbb{R}^{N \times K}$ the data matrix, $U \in \mathbb{R}^{N \times P}$ the unknown subspace of interest, $S \in \mathbb{R}^{P \times K}$ the signal of interest and $N \in \mathbb{R}^{N \times K}$ the additive noise, in which, $N$...
denotes the number of antennas, \( K \) the number of snapshots, \( P \) the dimension of subspace of interest (\( P \ll N \)) and \( p(Y \mid U) \) the conditional probability of \( Y \) given \( U \). \( \text{Tr} \{ \cdot \} \) stands for trace of a given matrix and \( E \{ \cdot \} \) is the expectation of a given random variable.

2.1. Minimum mean square distance (MMSD)

The MMSD estimator minimizes the distance between the range spaces \( R(\hat{U}) \) and \( R(U) \), i.e., \( \| \hat{U}^T - UU^T \|_F^2 \) which is a natural metric in Grassman manifold \( G_{N,P} \) (the set of \( P \)-dimensional subspace in \( \mathbb{R}^N \)) as proven in [3], [4]. This estimator is therefore expressed as:

\[
\hat{U}_\text{mmsd} = \arg \min_U \| \hat{U}^T - UU^T \|_F^2 = \arg \max_U E(\text{Tr}(U^TUU^T\hat{U})).
\]

As shown in [1], the MMSD estimator is obtained in general from the \( P \) largest eigenvectors of the matrix \( M(p(U \mid Y)) = \int UU^T p(U \mid Y) dU \) [1] and [5] which is expressed below as:

\[
\hat{U}_\text{mmsd} = \mathcal{P}_P \left\{ \int UU^T p(U \mid Y) dU \right\} = \mathcal{P}_P \left\{ M(p(U \mid Y)) \right\}
\]  

where, \( \mathcal{P}_P \{ \cdot \} \) is an operator that stands for the \( P \) strongest eigenvectors of a given matrix. The generic expression of the MMSD depends on \( p(U \mid Y) \) which has to be specified based on both the data model and the prior distribution of \( S \) and \( U \).

2.2. Model setup

In this paper, we assume that the data is the sum of a signal of interest and a zero white Gaussian noise. Furthermore, we assume that the columns of the additive noise \( N \) are i.i.d. and each follows \( N(0, \sigma^2 I) \) distribution and the columns of the signal of interest \( S \) are i.i.d. and each follows \( N(0, \sigma^2 \mathcal{U}) \) distribution. By contrast, we note that in [1], \( S \) is assumed to follow a uniform law. The linear Gaussian model is given by:

\[
Y = US + N
\]  

(2)

Since, we adopt a Bayesian framework in this paper, a prior distribution must be assigned to \( U \). With regarding to Grassman and Stiefel manifold, the most known distributions are Bingham and von Mises Fisher (vMF) [7], [8] which are specified below.

2.2.1. General Bingham von Mises Fisher distribution (BMF)

The BMF distribution has as parameter matrices \( A \in \mathbb{R}^{N \times N}, B \in \mathbb{R}^{P \times P} \) and \( C \in \mathbb{R}^{N \times P} \) [6]. In this paper, it is sufficient to assume that \( B \) is a diagonal matrix, \( A \) is a symmetric matrix, \( C \) is an orthogonal matrix. The matrix \( U \) is said to have \( U \sim \text{BMF}(A, B, C) \) distribution, if its probability density function is given as below: \( p_{\text{BMF}}(U) \propto \text{etr}(C^T U + BU^T AU) \) in which, \( \text{etr} \{ \cdot \} \) stands for the exponential of the trace of a given matrix and \( \propto \) indicates ”proportional to”. The Bingham and vMF are special cases of this distribution.

2.2.2. Bingham distribution

The Bingham distribution [7] is a particular case of the BMF distribution where, \( C \) is a null matrix. The matrix \( U \) is said to have a Bingham distribution with parameters \( A \) and \( B \), i.e., \( U \sim B(A, B) \), if its probability density function reads as \( p_B(U) \propto \text{etr}(BU^T AU) \) in which, \( A \) gathers the prior knowledge about the subspace of interest. Here, we focus on the case where \( A \) is proportional to a projection matrix spanned by the basis \( \hat{U} \). Thus, we denote \( B = I \) and \( A = \kappa \hat{U} \hat{U}^T : \)

\[
p_B(U) \propto \text{etr}(\kappa U^T \hat{U} \hat{U}^T U)
\]  

(3)

where \( \hat{U} \) is therefore the center of the distribution and \( \kappa \) is the concentration parameter of the distribution.

2.2.3. vMF distribution

The vMF distribution [8] also is a special case of BMF for which \( A \) and \( B \) are nulls. \( U \) is said to have a vMF distribution with parameters \( C \), i.e., \( U \sim vMF(C) \), if its probability density
function is given by: \( p_{\text{vMF}}(U) \propto \text{etr}(C^T U) \). As previously, we focus on vMF distribution where \( C \) is proportional to a unitary matrix \( \hat{U} \). We denote \( C = \kappa \hat{U} \), thus,

\[
p_{\text{vMF}}(U) \propto \text{etr}(\kappa U^T U)
\]

(4)

**Remark 1:** Note that the Bingham distribution \( p_B \) characterizes the distribution of \( R(U) \) range space of \( U \). The vMF distribution \( p_{\text{vMF}} \) characterizes the distribution of the orthogonal base \( U \). Hence, as defined in (3) and (4), \( p_B \) is invariant by rotation while, \( p_{\text{vMF}} \) is not.

3. **Proposed MMSD estimators in presence of Gaussian sources**

Here, we develop the MMSD estimator from 2.1 for the model specified in 2.2 using the prior distributions of \( U \) given in 2.2. Let us derive the conditional probability of \( Y \) given \( U \):

\[
p(Y \mid U) = \int p(Y \mid S, U)p(S)dS \propto \int \text{etr} \left\{ -\frac{1}{2\sigma^2} (Y - US)^T (Y - US) \right\} \text{etr} \left\{ -\frac{1}{2\sigma^2} S^T S \right\} dS
\]

\[
\propto \text{etr} \left\{ \alpha_Y Y^T UU^T Y - \frac{1}{2\sigma^2} Y^T Y \right\}
\]

(5)

where \( \alpha_Y = \frac{\sigma^2}{2\pi \langle \sigma^2 + \mu_Y^2 \rangle} \). In the next subsection, we develop the MMSD estimators, for Bingham and vMF prior distributions on \( U \).

3.1. **Bingham prior**

We start by deriving the posterior probability of \( U \), which is given by: \( p_{B-G}(U \mid Y) \propto p(Y \mid U)p_B(U) \propto \text{etr} \left\{ U^T (\alpha_G YY^T + \kappa UU^T) U \right\} \), consequently, the posterior probability \( p_{B-G}(U \mid Y) \propto B(\alpha_G YY^T + \kappa UU^T, I) \). Then, we move to the derivation of the MMSD estimator using (1), we obtain:

\[
\hat{U}_{\text{mmsd-B-G}} = \mathcal{P}_P \left\{ \int U^T p_{B-G}(U \mid Y) dU \right\} = \mathcal{P}_P \left\{ \int U^T \text{etr} \left\{ U^T (\alpha_G YY^T + \kappa UU^T) U \right\} dU \right\}.
\]

At this point, we use proposition 1 of [1] which states that \( \int U^T \text{etr} \left\{ U^T (\alpha_G YY^T + \kappa UU^T) U \right\} dU \) and \( \alpha_G YY^T + \kappa UU^T \) share same eigenvectors and its correspondent eigenvalues have the same order. Consequently, we obtain:

\[
\hat{U}_{\text{mmsd-B-G}} = \mathcal{P}_P \left\{ \alpha_G YY^T + \kappa UU^T \right\}
\]

(6)

3.2. **vMF prior**

We start with deriving the posterior probability of \( U \) regarding the vMF prior: \( p_{\text{vMF-G}}(U \mid Y) \propto p(Y \mid U)p_{\text{vMF}}(U) \propto \text{etr} \left\{ \kappa U^T U + \alpha_G U^T YY^T U \right\} \). In this case, \( p_{\text{vMF-G}} \propto \text{BMF}(YY^T, \alpha_G I, \kappa U) \). Then, we use (1) to derive the estimator:

\[
\hat{U}_{\text{mmsd-vMF-G}} = \mathcal{P}_P \left\{ \int U^T \mathcal{P}_P \left\{ \text{BMF}(YY^T, \alpha_G I, \kappa U) \mid U \right\} dU \right\} = \mathcal{P}_P \left\{ \int U^T \text{etr} \left\{ \kappa U^T U + \alpha_G U^T YY^T U \right\} dU \right\}.
\]

Unfortunately, there is no closed form of the the MMSD in this context. However, an efficient Gibbs sampling can be generated to draw the BMF distribution [6]. In order to compute the so-called induced arithmetic mean (IAM) [9] of the unitary matrix \( U^{(n)} \sim \text{BMF}(YY^T, \alpha_G I, \kappa U) \) and then, the MMSD is given by:

\[
\hat{U}_{\text{mmsd-vMF-G}} = \mathcal{P}_P \left\{ \frac{1}{N_r} \sum_{n=N_b+1}^{N_a+N_r} U^{(n)} U^{(n)\top} \right\}
\]

(7)

where \( N_b \) stands for the burn-in samples which is the number of thrown samples from the Markov chain, \( N_r \) is the number of samples used to approximate the estimator.

**Discussion:** We recall the expressions of the derived estimators in linear model context in [1], where \( S \) has a prior uniform distribution. For Bingham prior, the estimator is given by:

\[
\hat{U}_{\text{mmsd-B-U}} = \mathcal{P}_P \left\{ \alpha_U YY^T + \kappa UU^T \right\}.
\]

For the vMF prior, the MMSD estimator is obtained as (7) where \( U^{(n)} \sim \text{BMF}(YY^T, \alpha_U I, \kappa U) \) and \( \alpha_U = \frac{\sigma^2}{2\pi} \).

The main difference between our MMSD estimators (6), (7) and those presented previously [1] lies in the scalar \( \alpha_G \) from (6). Notice that for low SNR, \( \alpha_G \ll \alpha_U \) which means that our estimators are more shifted towards the prior \( \hat{U} \) than the ones given above. While, for high SNR, \( \alpha_G \approx \alpha_U \) which means that the estimators coincide, as shown in fig. 1.
4. Simulation
In this section, we illustrate the performance of the proposed estimator through evaluating the average fraction of energy (AFE). The AFE of a given estimator $\hat{U}$ is expressed as: $\text{AFE}(\hat{U}) = E\{\text{Tr}\{U^T\hat{U}^T\hat{U}\}\}$ which is considered as a criteria of performance, since, it evaluates the closeness of an estimator towards the random generated matrix $U$. We set $K=5$, $N=20$, $P=5$, $\text{SNR}=0\,\text{dB}$, $\kappa=20$ and $U = \begin{bmatrix} I_p \\ 0 \end{bmatrix}$. We compare in this part the proposed MMSD estimators denoted $\hat{U}_{\text{mmsd-B-G}}$ and $\hat{U}_{\text{mmsd-vMF-G}}$ with those presented in [1] denoted respectively $\hat{U}_{\text{mmsd-B-U}}$ and $\hat{U}_{\text{mmsd-vMF-U}}$ and with the SVD estimator of the SCM, denoted $\hat{U}_{\text{SCM}}$ which takes no prior knowledge of subspace, i.e., $U$ is deterministic. For computing the IAM, we fix $N_{bi} = 10$ and $N_r = 200$. Then, we illustrate figures below which examine the impact of SNR and $K$ on the performance of estimators and for Gaussian prior for $S$.

According to fig.2 and fig.3, the SVD estimator shows his poor performance outside the asymptotic regimes. In regard to fig.2 and fig.3, we notice that for low SNR and/or small number of snapshots, the proposed estimators (6), (7), prove a better performance compared respectively to those presented in discussion above. In asymptotic regimes, the proposed estimators have almost same performance as the ones given in discussion above.

5. Application to STAP low rank detection
In this section, we illustrate the interest of MMSD estimators for space-time adaptive processing (STAP) on a real data set, provided by the French agency DGA/MI [16]. STAP [12] is applied to airborne radar in order to detect moving targets. Typically, the radar receiver consists in an
array of $Q$ antenna elements processing $M$ pulses in a coherent processing interval ($N_c = MQ$). In the classical framework [15], we assume that the $K + 1$ samples are drawn as:

$$
\begin{aligned}
& z_0 = d + c_0 + n_0 \\
& z_k = c_k + n_k, \forall k \in \{1...K\}
\end{aligned}
$$

where the tested cell $z_0$ potentially contains a target $d = \alpha_0 p(\theta, \nu_t)$ to be detected, $\alpha_0$ is the amplitude and $p$ is the normalized steering vector, $\nu_t$ is the target speed and $\theta$ is the DOA of the target ($\alpha_0$, $\nu_t$ and $\theta$ are unknown). For each sample, the additive noise is the sum of the ground clutter $c$ and white Gaussian noise $n$. The secondary data $\{z_k\}$ are assumed to be i.i.d., signal and outlier free. Thanks to the Brennan rule [13], the ground clutter is known to be contained in a low dimensional subspace $\Pi_c = U_c U_c^H$ of (known) rank $R$. Hence, if we concatenate the secondary data, we obtain in matrix form:

$$
Z = \Pi_c C + N
$$

It is worth mentioning that, even if the fluctuation of the clutter $c$ may not be white Gaussian in practice, results reveal that the considered framework remains robust to such assumptions (c.f. Figure 4). We investigate the performance of the low rank adaptive normalized matched filter (ANMF) detector [12]:

$$
\hat{\Lambda}(\hat{\Pi}_c) = \frac{|d^H \hat{\Pi}_c^H z_0|^2}{|d^H \hat{\Pi}_c^H d||z_0^H \hat{\Pi}_c^H z_0|}
$$

where $\hat{\Pi}_c = \hat{U}_c \hat{U}_c^T$ is a clutter subspace projector (CSP) estimated from the secondary data $\{z_k\}$. The performance of a given detector depends on the accuracy of the CSP estimator from which it is built. This is why we compare the performance of the following detectors: $\hat{\Lambda}_{SCM}$, where the CSP is estimated from the SVD of the SCM, $\hat{\Lambda}_{SFPE}$, where the CSP is estimated from the SVD of the shrinkage fixed point estimator with minimal regularization parameter $\gamma$ [14]. $\hat{\Lambda}_{mmsd-B-G}$, where the CSP is estimated using our proposed schemes thanks to (6). Since the parameters $\alpha_G$ and $\kappa$ are here unknown, we will illustrate the performance of the estimator defined as $\hat{U}_{mmsd-B-G} = \mathcal{P}_R [Z Z^H + \beta \hat{U} \hat{U}^H]$ for different values of the ratio parameter $\beta = \kappa/\alpha_G$. The prior $\hat{U}$ is computed from the SVD of the STAP covariance matrix model in [12]. $\Lambda_{prior}$, the (non adaptive) detector using the prior $\hat{U}$ from the SVD of the STAP covariance matrix model in [12].

Figure 4 displays the output of the different detectors (i.e. the value of the ANMF for a grid of target parameters) for a configuration where the tested cell $z_0$ contains 10 targets and the number of secondary data is limited but still larger than the evaluated clutter rank ($2R < K = 100 < N_c/2$). We observe that, in this context, the detector $\Lambda_{SFPE}$ allows for target detection with a relatively small false alarm rate. The detector $\hat{\Lambda}_{mmsd-B-G}$ represents a trade off between $\Lambda_{SCM}$ (i.e. $\beta = 0$) and $\Lambda_{prior}$ ($\beta \to \infty$). As observed, $\Lambda_{SCM}$ does not allow for target detection since the target responses seems also canceled. This is probably due to the spreading of targets in the secondary data and the poor robustness of the SCM. The fully deterministic detector $\Lambda_{prior}$ does not cancel the ground clutter as we can only observe false alarms on the ridge and no distinguishable targets. However, the detector $\hat{\Lambda}_{mmsd-B-G}$ provides a trade off that allows for interference rejection and reliable target detection. This detector even outperforms $\Lambda_{SFPE}$ for appropriate values of the parameter $\beta$, which illustrates the interest of introducing some prior information in an adaptive subspace estimation process.

As prospects of this work, it will be interesting to develop adaptive methods for the selection of the parameter $\beta = \kappa/\alpha_G$. A second point to be investigated is the development of MMSD estimators that account for others signal models (e.g. covariance models, colored and non-Gaussian sources).
Figure 4. Output of the ANMF for a grid of target parameters $(\theta, v)$ where $\theta \in [-3, 3]$ and $v \in [-6, 6]$ for a given detector. The represented detectors are $\hat{A}_{SCM}, \hat{A}_{SPPE}, \hat{A}_{mmrd-B-G}$ for several values of $\beta$, $\hat{A}_{prior}$, the number of secondary data is $K=100$. The number of sensors is $Q = 4$ and the number of coherent pulses is $M = 64$, the size of the data is then $N_c = QM = 256$. The center frequency and the bandwidth are respectively equal to $f_0 = 10$GHz and the bandwidth $B = 5$MHz. The radar celerity is $V = 100m/s$. The inter-element spacing is $d = 0.3m$ and the pulse repetition frequency is $f_r = 1kHz$. The clutter rank, computed from Brennan Rule [13], is $R = 45$ and the CNR is evaluated around 20dB.

References
[1] O. Bessoon, N. Dobigeon and J. Tourneret, Minimum mean square distance estimation of a subspace, IEEE Trans., Signal Process., vol. 59, No. 12, pp 5709-5720, December 2011.
[2] R. O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans., Antennas Propag., vol. 34, no 3 , pp 276-280, March 1986.
[3] A. Edelman, T. Arias, and S. Smith, The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl., vol. 20, no. 2, pp. 303-353, 1998.
[4] G. Golub and C. V. Loan, Matrix Computations, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[5] A. Srivastava, A Bayesian approach to geometric subspace estimation, IEEE Trans. Sig. Process., vol. 48, no. 5, pp. 1390-1400, May 2000.
[6] P. D. Hoff, Simulation of the matrix Bingham-von Mises-Fisher distribution with applications to multivariate and relational data, J. Computat. Graph. Statist., vol. 18, no. 2, pp. 438-456, June 2009.
[7] K. V. Mardia and P. E. Jupp, Directional Statistics. New York: Wiley, pp. 167-183, 1999.
[8] Y. Chikuse, Statistics on Special Manifolds. New York: Springer-Verlag, pp. 27-52, 2003.
[9] A. Sarlette and R. Sepulchre, Consensus optimization on manifolds, SIAM J. Contr. Optimiz., vol. 48, no. 1, pp. 56-76, 2009.
[10] M. Haardt, M. Pesavento, F. Roemer and M. N. El Korso, Subspace Methods and Exploitation of Special Array Structures, Electro. Refer. in Sig. Process.: Array and Statis. Sig. Process. (M. Viberg, ed.), vol. 3, pp. 651-717, 2014.
[11] J. Liu, M. S. Sibat and V. Tuzulkov, Interference Cancellation and DOA Estimation by Generalized Receiver Applying LMS and MUSIC Algorithms, Progress In Electromagnetics Research Symposium Proceed., pp 187-190, March 2013.
[12] J. Ward, Space-time adaptive processing for airborne radar, Tech. Rep., Lincoln Lab., MIT, Lexington, Mass., USA, December 1994.
[13] L. E. Brennan and F. M. Staudaher, Subclutter visibility demonstration, Tech. Rep., RL-TR-92-21, Adapt. Sensors Incorporated, March 1992.
[14] F. Pascal, Y. Chitour and Y. Quek, Generalized robust shrinkage estimator and its application to STAP detection problem, Sig. Process., arXiv:1311.6567 , September 2014.
[15] G. Ginolhac, P. Forster, J. P. Ovarlez and F. Pascal, Détecteurs STAP Rang Réduit, Traitement du Signal, vol. 28, no. 1-2, pp 143-170, April 2011.
[16] J. P. Ovarlez, F. Le Chevalier and S. Bidon, les données de club STAP, Introduction au STAP, Traitement de signal, vol. 28, 2011.