CLUSTERING COMPLEX SUBSPACES IN LARGE DIMENSIONS

Roberto Pereira[⋆] *Xavier Mestre*[⋆] *David Gregoratti*[†]

[⋆] Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Barcelona, Spain
[†] Software Radio Systems, Barcelona, Spain

ABSTRACT

A methodology to cluster multiple sets of Gaussian multivariate complex observations based on the alignment of their column spaces is presented. These subspaces are identified with points in the Grassmann manifold and compared according to a similarity measure drawn from a chosen manifold distance, which is proportional to the squared projection–Frobenius norm. In order to guarantee that distances between subspaces of different dimensions are comparable, we proposed to normalise the corresponding decision statistics with respect to their asymptotic mean and variance, assuming that (i) the dimensions of both the observation and the involved subspaces are large but comparable in magnitude and (ii) both subspaces are generated by the same statistical law. A procedure is derived to estimate these normalisation parameters, leading to a new statistic that can be built exclusively from the observations. The method is applied to a MIMO wireless channel clustering problem, where is shown to outperform conventional similarity measures in terms of classification performance.

***Index Terms*—** Complex Subspace Clustering, Grassmann Manifold, Random Matrix Theory, Central Limit Theorem.

1. INTRODUCTION

Multiple applications nowadays require the use of subspace clustering, a powerful mechanism that classifies sets of multivariate observations into clusters according to the alignment of their spanned subspaces. This technique is useful in both dynamic and non–dynamic systems, such as wireless communications [1], electroencephalography [2], transfer learning [3] or image segmentation [4] to name a few. Several methods have been developed to extract meaningful subspace–structured features that can typically be fed into conventional clustering algorithms (see [5] for a detailed review). A common approach is to represent these subspaces as points in a Grassmann manifold [6]. By doing so, one can take advantage of the geometrical properties of the Grassmannian in order to perform subspace alignment/comparison or design optimized kernel methods [5, 7, 8].

The main objective of the present paper is to propose a method to cluster collections of complex multivariate Gaussian observations according to their proximity in terms of the subspace they span. The motivation behind it can be found in the field of multi–antenna wireless communications, where these collections of observations correspond to different MIMO Rayleigh channels. The idea is to optimize the use of the available spatial degrees of freedom by identifying groups of users that are seen from similar angles (i.e. span a similar subspace) as a single spatial entity. It is typically easier to spatially multiplex different signals among well separated groups rather than attempt individual user multiplexing. Once these groups have been spatially multiplexed, one can process the signals within each group by either orthogonal (FDMA, TDMA) or non–orthogonal (NOMA, Rate Splitting) techniques [9,10]. It is therefore important that MIMO wireless channels are clustered according to their proximity in terms of the subspace they span.

In [11] we proposed a hierarchical subspace–based clustering mechanism that assumes observations to be complex multi–variate circularly symmetric Gaussian observations. The main idea is to build the hierarchical structure by using a similarity measure which exploits the inherent geometry embedded in the observed subspaces, i.e. by identifying groups of observations as points in the Grassmann manifold. In order to find a similarity measure which is relevant when the compared subspaces have different dimensions, we propose to normalise the original metric according to its large–dimensional asymptotic behavior (see Section 2.2 for details). The idea is to guarantee that, whenever the two compared subspaces are well aligned, the similarity measure will approximately behave as a standard Gaussian random variable regardless of dimensions of the involved subspaces. This provides a valid tool in order to compare the proximity of multiple subspaces by checking their pairwise similarity.

One of main open issues of the described approach is the fact that the normalisation of the similarity statistics requires perfect knowledge of the statistics of the underlying observations. Unfortunately, this information is typically unknown in real–world applications and needs to be estimated from the observations themselves. The one of the main objectives of this work is to propose an alternative solution and directly estimate the normalisation parameters from the given observations. In that sense, this work can be seen as an extension of [11] for the case where the statistics of the observations are no longer available.

2. PRELIMINARIES

2.1. Signal Model

Let us consider a collection of K complex independent multi–variate stochastic processes, all of the same dimension M. We assume that each multi–variate process is Gaussian distributed, with zero mean and covariance matrix $\mathbf{R}_k \in \mathbb{C}^{M\times M}$ (assumed to be full rank). Moreover, we assume that $N_k < M$ observations (independent from each other) are available from the kth multidimensional process and denote as $\mathbf{X}_k \in \mathbb{C}^{M\times N_k}$ the observation matrix that contains these N_k observations as columns, $k = 1, \ldots, K$. This means that the observation matrix of the kth process is decomposable as

$$\mathbf{X}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{G}_k \tag{1}$$

This work has been partially funded by the European Commission under the Windmill project (contract 813999) and the Spanish government under the Aristides project (RTI2018-099722-B-I00).

where the entries of the $M \times N_k$ matrix $\mathbf{G}_k$ are assumed to be independent and identically distributed (i.i.d.) complex circularly symmetric Gaussian random variables, with zero mean and unit variance. Here the covariance matrice is inspired by the wireless communications context and can be interpreted as the $M \times M$ receive covariance matrices of the multi–antenna channel. In that particular setting, this matrix is inherently dependent on the scattering structure of the scenario, in particular on the angles of arrival (AoA) of the multiple transmission paths.

We are primarily interested in establishing the similarity between the subspace spanned by the columns of pairs of observation matrices $\mathbf{X}_k, \mathbf{X}_j$, for $k \neq j$. By checking the structure of these observations in (1), one can readily anticipate that two observation matrices will span similar subspaces with high probability when they are generated with the same left covariance matrix $\mathbf{R}_k$. However, since this matrix is in practice unknown, we will instead resort to the sample covariance matrix, defined for the kth process as

$$\hat{\mathbf{R}}_k = \frac{1}{N_k} \mathbf{X}_k \mathbf{X}_k^{\mathrm{H}}. \tag{2}$$

We emphasise that the sample covariance matrix is now rank deficient, i.e., $\mathrm{rank}(\hat{\mathbf{R}}_{\mathrm{k}}) = N_k < M$. In the next section we will see that this is closely related to the dimension of the manifold which the kth observation is projected onto.

2.2. Subspace Comparison Using Grassmann Manifolds

It is well known [8, 12–14] that the distance between two (complex) N–dimensional subspaces can be geometrically characterised by their principal angles. Interestingly enough, these angles induce several metrics on the (complex) Grassmann manifold $\mathcal{G}(N, M)$, which provides a topological structure to the set of all N–dimensional subspaces in a (complex) M–dimensional space.

The Grassmannian $\mathcal{G}(N, M)$ can be identified with the space of symmetric orthogonal projection matrices of size $M \times M$ and rank N [8]. In other words, we can represent the column space of the kth observation matrix $\mathbf{X}_k$ as a point $\hat{\mathbf{P}}_k \in \mathcal{G}(N_k, M)$, where $\hat{\mathbf{P}}_k$ is the projection matrix

$$\hat{\mathbf{P}}_k = \mathbf{X}_k (\mathbf{X}_k^{\mathrm{H}} \mathbf{X}_k)^{-1} \mathbf{X}_k^{\mathrm{H}}. \tag{3}$$

Alternatively, both the column subspace of $\mathbf{X}_k$ and the projection matrix $\hat{\mathbf{P}}_k$ are fully determined by the first N_k left singular vectors of $\mathbf{X}_k$, which we denote here as $\hat{\mathbf{V}}_k$. Therefore, it can be seen [11, 13] that the cosine of the principal angles between the subspaces spanned by the columns of $\mathbf{X}_k$, $\mathbf{X}_j$ are equal to the singular values of the matrix $\hat{\mathbf{V}}_k^{\mathrm{H}} \hat{\mathbf{V}}_j$.

The main advantage of formulating the problem of subspace classification in the framework of Grassmannian learning is that we can consider conventional distance measures that define the topological structure of this manifold, see [12] for a detailed review. In particular, in this paper we consider the squared projection–Frobenius distance, which is defined as

$$d^2_{\mathrm{PF}}(\mathbf{X}_k, \mathbf{X}_j) = \sum_i \sin^2(\alpha_{k,j}(i)) = \min(N_j, N_k) - \mathrm{tr}(\hat{\mathbf{P}}_k \hat{\mathbf{P}}_j) \tag{4}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix and $\alpha_{k,j}(i)$ denotes the ith principal angle between the column spaces generated by $\mathbf{X}_k$ and $\mathbf{X}_j$. This particular choice of metric has the advantage that it can be computed without eigenvalue decompositions. Moreover, it can be shown [11] that this metric is particularly useful for the task at hand.

Finally, we can directly convert (4) into a similarity measure by considering

$$s_{k,j} = \frac{1}{M} \mathrm{tr}(\hat{\mathbf{P}}_k \hat{\mathbf{P}}_j) \tag{5}$$

when comparing two observations. Observe that this similarity measure takes values between 0 (when the observations span orthogonal subspaces) to $M^{-1} \min(N_k, N_j)$ (when they span parallel subspaces). Notice, however, that even in the case where $\mathbf{R}_k = \mathbf{R}_j$, the probability that two observations (k, j) span exactly the same subspace is zero due to the fact that those observations are drawn from absolutely continuous Gaussian probability distributions.

3. HIERARCHICAL CLUSTERING OF SUBSPACES

We have now introduced the main concepts necessary to formulate an agglomerative hierarchical clustering based on the similarity measure in (5). In this bottom–up approach, the goal is to combine different observations into larger collections based on how similar their corresponding subspaces are. To do so, we consider the collection of all pairs of clustered subspaces and compare them according to the similarity measure described above. The pair with highest similarity (or, equivalently, lower distance) is merged and forms a new cluster, represented by the concatenation of the different columns that constitute the corresponding observation matrices.

In order to study the problem from the statistical perspective, we formulate each merging step in the hierarchical clustering procedure as a binary hypothesis test. We declare that the two observation matrices belong to the same cluster (in terms of proximity of their corresponding column subspaces) if the observations are generated from the same left covariance matrix, that is

$$\begin{aligned} \mathrm{H}_0(k, j) &: \mathbf{R}_k = \mathbf{R}_j \\ \mathrm{H}_1(k, j) &: \mathbf{R}_k \neq \mathbf{R}_j. \end{aligned} \tag{6}$$

If the null hypothesis is accepted, then the new cluster is formed by the concatenation $\mathbf{X}_{[k,j]} = [\mathbf{X}_k, \mathbf{X}_j]$, which is equivalent to concatenating the left singular vectors associated to their respective column subspaces, assuming that $N_i + N_j \leq M$. Moreover, notice that the concatenated matrix $\mathbf{X}_{[k,j]}$ – of dimension $M \times (N_i + N_j)$ – can still be modeled according to (1), where $\mathbf{R}_k$ is the common left covariance matrix of the new cluster.

Now, one of the main problems with the use of the similarity measure in (5) is that its statistics are inherently dependent on the pair of observations that are being compared, especially with regard to the dimensionality of the associated subspaces. This problem is a direct consequence of the fact that the the distance measures are functions of the dimensionality of the underlying Grassmann manifold.

There are a few existing proposals for distances between subspaces of different dimensions [15–17], although usually the solution follows from embedding both Grassmanianns into a higher dimensional one. A more effective alternative consists in normalising the similarity measure $s_{i,j}$ with respect to its asymptotic null distribution (under $\mathrm{H}_0(i, j)$), by subtracting the corresponding mean and dividing by the corresponding standard deviation. This guarantees that, under the null hypothesis, all the statistics associated to pairs of observations will have the same standardized behavior, so that they can be considered to be comparable.

Since obtaining the exact expression for the first and second order moments of $s_{i,j}$ appears to be extremely difficult, one can instead resort to a large dimensional approximation and consider the situation where all the involved dimensions $(M, N_1, \ldots, N_K)$ are large but comparable in magnitude. It can be seen that these are very good approximations even for relatively low dimensions.

In [11], we show that $s_{i,j}$ in (5) asymptotically fluctuates as a Gaussian random variable and provide expressions for the asymptotics of the first ($\eta_{k,j}$) and second order ($\sigma^2_{k,j}$) moments, which obviously depend on the covariance matrices $\mathbf{R}_k, \mathbf{R}_j$. We therefore propose to use the normalised similarity measurement

$$\tilde{s}_{k,j} = \frac{s_{k,j} - \eta_{k,j}}{\sigma_{k,j}} \tag{7}$$

where $\eta_{k,j}$ and $\sigma_{k,j}$ are particularized to the null hypothesis in (6), namely $\mathbf{R}_k = \mathbf{R}_j$. The main advantage of the proposed similarity measure is the fact that all distances are comparable under the null hypothesis (i.e. when the subspaces are generated with the same left covariance matrix) and this property holds asymptotically regardless of the inherent subspace dimensions. This means that the normalised similarity measure in (7) allows us to effectively compare the degree of alignment of different subspaces, in the sense that all similarity measures between observations in the null hypothesis (i.e. sharing the same covariance) will asymptotically behave according to the standard Gaussian law.

Observe that the construction of the normalised statistic in (7) requires a perfect knowledge of the common left covariance matrices of the signals that are being compared. In low dimensional scenarios, this is typically solved by computing the required parameters with the sample covariance matrix in (2) instead of the true one. However, in large dimensional scenarios this is far from optimal, mainly because the sample covariance matrix can hardly be regarded as a consistent estimate of the true one. In this paper, we propose to solve this issue by directly attempting to estimate both the asymptotic mean and variance $\eta_{k,j}, \sigma^2_{k,j}$ under the null hypothesis, directly from the observations. More specifically, we will provide estimators for these two quantities that are consistent even in the situation where the observation dimension increases with the sample size of each class under comparison.

4. ESTIMATOR OF THE DISTRIBUTION UNDER H_0

In this section, we present estimators for $\eta_{k,j}$ and $\sigma^2_{k,j}$ that can be built from the sample covariance matrices $\hat{\mathbf{R}}_k, \hat{\mathbf{R}}_j$ and are consistent even when the observation dimension increases without bound. In order to simplify the exposition, we begin by presenting the expression of these estimators and then later formulate the consistency result.

The proposed estimator of $\eta_{k,j}$ is denoted as $\hat{\eta}_{k,j}$ and can be expressed as

$$\hat{\eta}_{k,j} = \begin{cases} \frac{N_k}{2M}\left(1 - \frac{\kappa_k^2(1)}{\kappa_k(2)}\right) + \frac{N_j}{2M}\left(1 - \frac{\kappa_j^2(1)}{\kappa_j(2)}\right) & N_k = N_j \\ \frac{1}{2M}\left(\frac{N_k\hat{\mu}_k(k) - N_j\hat{\mu}_j(k)}{\hat{\mu}_k(k) - \hat{\mu}_j(k)} + \frac{N_k\hat{\mu}_k(j) - N_j\hat{\mu}_j(j)}{\hat{\mu}_k(j) - \hat{\mu}_j(j)}\right) & N_k \neq N_j \end{cases}$$

where we have defined, for $l \in \{k, j\}$,

$$\kappa_l(m) = \frac{1}{N_l}\mathrm{tr}\left[\left(\hat{\mathbf{R}}_l^{\#}\right)^m\right],\ m \in \mathbb{N}$$

and where $(\cdot)^{\#}$ denotes the Moore–Penrose pseudo–inverse. Furthermore, we have $\hat{\mu}_l(l) = \frac{-1}{\kappa_l(1)}$, whereas for $N_k \neq N_j$ we take

$$\hat{\mu}_j(k) = \gamma\left(1 - \frac{N_j}{N_k}\right)$$

where γ is the smallest solution to

$$\frac{1}{N_j}\mathrm{tr}\left[\hat{\mathbf{R}}_k\left(\hat{\mathbf{R}}_k - \gamma\mathbf{I}_M\right)^{-1}\right] = 1.$$

The estimator of the asymptotic covariance $\sigma^2_{k,j}$ is denoted as $\hat{\sigma}^2_{k,j}$ and can be described as

$$\hat{\sigma}^2_{k,j} = \zeta^2_k(j) + \zeta^2_j(k) + \frac{\left(\mathrm{tr}\left[\hat{\mathbf{R}}_k^{\#}\hat{\mathbf{R}}_j^{\#}\right]\right)^2}{\mathrm{tr}\left[\left(\hat{\mathbf{R}}_k^{\#}\right)^2\right]\mathrm{tr}\left[\left(\hat{\mathbf{R}}_j^{\#}\right)^2\right]}. \tag{8}$$

The quantity $\zeta^2_j(k)$ can in turn be expressed as

$$\begin{aligned}\zeta^2_j(k) = &\left[\frac{\kappa_j(4)\kappa_j^2(1)}{\kappa_j^3(2)} - 2\left(\frac{\kappa_j(1)\kappa_j(3)}{\kappa_j^2(2)}\right)^2\right]\times \\ &\left[\frac{1}{\kappa_j(1)}\rho_j(\mathbf{B}_k, 1) - \left(\frac{1}{\kappa_j(1)}\rho_j(\mathbf{A}_k, 1)\right)^2\right] \\ &+ \frac{\kappa_j(1)\kappa_j(3)}{\kappa_j^2(2)}\left(\frac{\rho_j(\mathbf{B}_k, 1)}{\kappa_j(1)} + \frac{2\rho_j(\mathbf{A}_k, 2)}{\kappa_j(2)} - \frac{\rho_j(\mathbf{B}_k, 3)}{\kappa_j(3)}\right) \\ &- 2\frac{\kappa_j(3)\rho_j(\mathbf{A}_k, 1)}{\kappa_j^2(2)}\left(\frac{2\rho_j(\mathbf{A}_k, 2)}{\kappa_2(2)} - \frac{\rho_j(\mathbf{A}_k, 3)}{\kappa_j(3)}\right) \\ &+ \left(\frac{\rho_j(\mathbf{A}_k, 2)}{\kappa_j(2)}\right)^2 - \frac{\rho_j(\mathbf{B}_k, 2)}{\kappa_j(2)}\end{aligned} \tag{9}$$

where we have introduced the matrices $\mathbf{A}_k = -\hat{\mathbf{P}}_k^{\perp}$ and $\mathbf{B}_k = \hat{\mathbf{P}}_k^{\perp} - \frac{\kappa_k(1)}{\kappa_k(2)}\hat{\mathbf{R}}_k^{\#}$, together with $\hat{\mathbf{P}}_k^{\perp} = \mathbf{I}_M - \hat{\mathbf{P}}_k$ and

$$\rho_j(\mathbf{A}, n) = \frac{1}{N_j}\mathrm{tr}\left[\mathbf{A}\left(\hat{\mathbf{R}}_j^{\#}\right)^n\right]$$

valid for any squared matrix $\mathbf{A}$. Obviously, $\zeta^2_k(j)$ is defined in the same way, but swapping the two indexes k, j. Having introduced these two estimators, we are now in the position to formulate the main result of the paper.

Theorem 4.1. Consider two independent random matrices $\mathbf{X}_k, \mathbf{X}_j$ following the statistical structure in (1). Assume that M, N_k, N_j increase to infinity at the same rate, so that $N_k/M \to c_k$, $N_j/M \to c_j$ for some $c_k, c_j \in (0, 1)$. Assume that $\mathbf{R}_k = \mathbf{R}_j$ has bounded spectral norm for all M. Then, we have $|\eta_{k,j} - \hat{\eta}_{k,j}| \to 0$ and $|\sigma^2_{k,j} - \hat{\sigma}^2_{k,j}| \to 0$ with probability one.

Proof. See [18]. □

Theorem 4.1 offers a simple mechanism to build an asymptotic equivalent of (5) that uses only the available observations and is still able to behave asymptotically as the original one under the null hypothesis, namely

$$\hat{s}_{k,j} = \frac{s_{k,j} - \hat{\eta}_{k,j}}{\hat{\sigma}_{k,j}}. \tag{10}$$

We will see in the next section that this new statistic provides a very practical way of comparing subspaces of different dimensions, always keeping the same reference behavior under the null hypothesis.

5. NUMERICAL VALIDATION

In order to validate the results presented above, we compare the clustering performance of the proposed similarity measure $\hat{s}_{k,j}$ as defined in (10), where the normalising coefficients are estimated according to Theorem 4.1. The performance of this proposed similarity measure is denoted as "EST" in the figures. For comparison purposes, we also analyze the performance of the artificial statistic $s_{k,j}$ in (5), which cannot be built unless the left covariance matrices of the observations are known. This performance will be denoted as "ASYM" in what follows.

We also compare the performance against four other more conventional metrics, the conventional Projection–Frobenius similarity (denoted here by "PF"), the Fubini–Study based similarity [12] ("FS"), together with the conventional maximal ("C–S") and average ("C–AVG") similarity between one–dimensional subspaces belonging to the observed subspaces. More specifically, the PF (resp. FS) similarity measure between the jth and kth subspaces is given by the arithmetic (resp. geometric) average of the non–zero eigenvalues of the matrix $\hat{\mathbf{P}}_k\hat{\mathbf{P}}_j\hat{\mathbf{P}}_k$. On the other hand, the C–S (resp. C–AVG) similarity is defined as the maximum (resp. average) scalar product between column vectors of $\mathbf{X}_j$ and $\mathbf{X}_k$ after norm–one normalisation. For all these distances, we considered an agglomerative hierarchical clustering algorithm that at each step merges the pair of observed matrices with highest similarity measure.

Rather than characterizing the behavior of the whole hierarchical clustering method, we focus here on individual merging steps in the clustering process. More specifically, we evaluate a simplified scenario with three observation matrices, $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ and assume that $\mathbf{X}_1, \mathbf{X}_2$ are generated with the same left covariance matrix ($\mathbf{R}_1 = \mathbf{R}_2$), and should therefore be merged into a single cluster. The left covariance matrix of the third observation is different from the other two ($\mathbf{R}_3 \neq \mathbf{R}_1$). Notice that the hierarchical clustering of K observations boils down to $\frac{K}{(K-3)!3!}$ triplet comparisons. In this sense, analysing the merging of pairs based on triplets is closely related to the behaviour of the hierarchical clustering in large scenarios, i.e., $K > 3$. Moreover, by comparing observations of different sizes, we are essentially simulating different levels of the hierarchical clustering. Finally, for each of the six metrics described above, the algorithm first computes the three similarity measures between pairs of matrices, and chooses to merge the pair that has the highest similarity. We define the probability of success (POS) as the probability of making the right merge and evaluate it by considering a collection of 10^4 realizations of these observations.

Inspired by the channel clustering application in wireless communications, we simulate the three observation matrices as MIMO wireless channels, where the left covariance matrices are generated by averaging the contribution of random directions of arrival impinging on a uniform linear array with half a wavelength inter–element separation. We will denote by θ_j, $j = 1, 2, 3$ the mean angle of arrival associated to each channel (they are Gaussian distributed with angular spread of ten degrees), so that obviously we have $\theta_1 = \theta_2 \neq \theta_3$. The separation between θ_1 and θ_3 can be understood as a distance between the subspaces that are spanned by the different channels, hence, the distribution from the observations. For instance, for small $\Delta\theta = |\theta_1 - \theta_3|$, we have that two groups are close together, thus it becomes harder to distinguish between one another. In contrast, for $\Delta\theta$ large, both groups are generated by non-related subspaces.

Figure 1 illustrates the POS obtained in four distinct scenarios for different choices of the dimensions (M, N_1, N_2, N_3), where in all cases these quantities are taken to increase proportionally in the x–axis. Notice that in all four scenarios the proposed normalised metric generally outperforms the other similarity measures, especially in situations where the compared subspaces have very different dimensions. (We recall here again that the "ASYM" is merely shown for comparative purposes.) Furthermore, observe that the proposed metric outperforms the rest even in scenarios with a relatively low number of dimensions, in spite of the fact that the statistic is designed to perform well in large dimensional settings. Figure 2 illustrates the relationship between the mean angular distance in the covariance generation $|\theta_1 - \theta_3|$ and the POS for each of the considered similarity measures. Observe that, once again, both normalised metrics outperform all the other metrics, especially in the region where the two left covariance matrices are close, which corresponds to the region where $\Delta\theta$ is small.

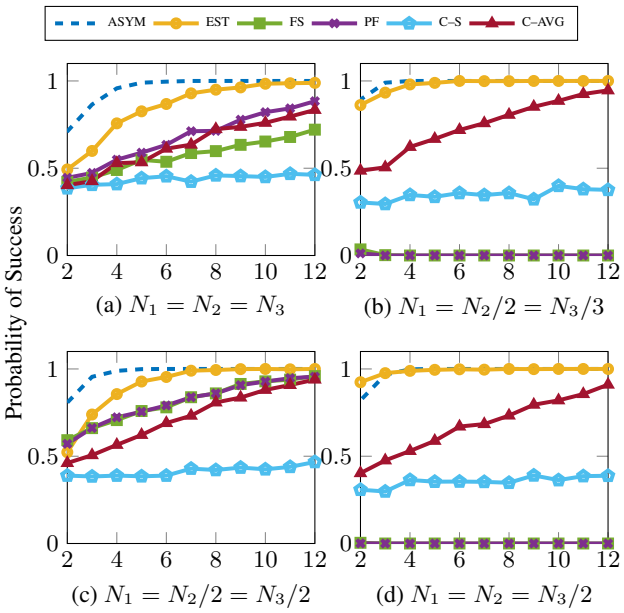

Fig. 1: Probability of success related to different metrics in four different scenarios (a)-(d) with respect to the growth of N_1 (x–axis).

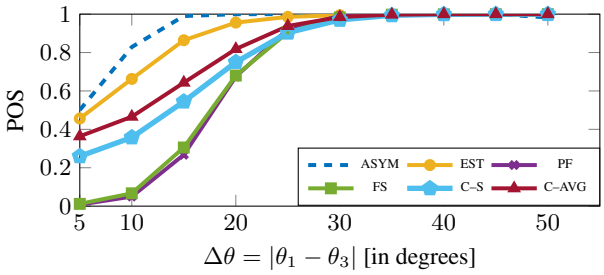

Fig. 2: Probability of success for different $\Delta\theta$. The dimensions of each observation can be described by $N_3/3 = N_2/2 = N_1 = 4$.

6. CONCLUSION

A novel similarity metric has been proposed that allows clustering complex Gaussian data according to the proximity of their column spaces. The main idea is to normalise the projection–Frobenius measure in the Grassmann manifold so that it behaves as a standard Gaussian random variable when the compared observation sets have the same left covariance matrix. The main achievement of the paper has been the proposal of consistent estimators of these normalisation parameters that are obtained exclusively from the observations. Numerical simulations confirm the validity of the presented results by illustrating a consistent and fast performance improvement of the normalised schemes when compared to other metrics. We finally stress that the approach described here is also applicable to other subspace clustering applications that are based on pairwise similarity comparisons.

7. REFERENCES

[1] Romain Couillet and Abla Kammoun, "Random matrix improved subspace clustering," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 90–94.

[2] Cong Ye, Konstantinos Slavakis, Pratik V Patil, Johan Nakuci, Sarah F Muldoon, and John Medaglia, "Network clustering via kernel-arma modeling and the grassmannian: The brain-network case," *Signal Processing*, vol. 179, pp. 107834, 2021.

[3] Tao Zhou, Huazhu Fu, Chen Gong, Jianbing Shen, Ling Shao, and Fatih Porikli, "Multi-Mutual Consistency Induced Transfer Subspace Learning for Human Motion Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[4] Leena Mathur and Maja J. Matarić, "Unsupervised Audio-Visual Subspace Alignment for High-Stakes Deception Detection," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 2255–2259.

[5] Jiayao Zhang, Guangxu Zhu, Robert W Heath Jr, and Kaibin Huang, "Grassmannian learning: Embedding geometry awareness in shallow and deep learning," *arXiv preprint arXiv:1808.02229*, 2018.

[6] Jeroen Boets, Katrien De Cock, Marcelo Espinoza, and Bart De Moor, "Clustering time series, subspace identification and cepstral distances," *Communications in Information & Systems*, vol. 5, no. 1, pp. 69–96, 2005.

[7] Yung-Chow Wong, "Differential geometry of Grassmann manifolds," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 57, no. 3, pp. 589, 1967.

[8] Lek-Heng Lim, Ken Sze-Wai Wong, and Ke Ye, "The Grassmannian of affine subspaces," *Foundations of Computational Mathematics*, vol. 21, pp. 537–574, 2021.

[9] Junyoung Nam, Ansuman Adhikary, Jae-Young Ahn, and Giuseppe Caire, "Joint Spatial Division and Multiplexing: Opportunistic Beamforming, User Grouping and Simplified Downlink Scheduling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 876–890, 2014.

[10] Mingbo Dai, Bruno Clerckx, David Gesbert, and Giuseppe Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4611–4624, 2016.

[11] Roberto Pereira, Xavier Mestre, and David Gregoratti, "Subspace Based Hierarchical Channel Clustering in Massive MIMO," in *2021 IEEE Globecom Workshops*, 2021, pp. 1–6.

[12] Alan Edelman, Tomás A. Arias, and Steven T. Smith, "The Geometry of Algorithms with Orthogonality Constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, Apr. 1999.

[13] David J. Love and Robert W. Heath, "Grassmannian beamforming on correlated MIMO channels," in *IEEE Global Telecommunications Conference, 2004. GLOBECOM '04.*, 2004, vol. 1, pp. 106–110 Vol.1.

[14] Uwe Helmke, Knut Hüper, and Jochen Trumpf, "Newton's method on Grassmann manifolds," *arXiv preprint arXiv:0709.2205*, 2007.

[15] Christopher A Beattie, Mark Embree, and Danny C Sorensen, "Convergence of polynomial restart Krylov methods for eigenvalue computations," *SIAM review*, vol. 47, no. 3, pp. 492–515, 2005.

[16] Xichen Sun, Liwei Wang, and Jufu Feng, "Further Results on the Subspace Distance," *Pattern recognition*, vol. 40, no. 1, pp. 328–329, 2007.

[17] Ke Ye and Lek-Heng Lim, "Schubert Varieties and Distances between Subspaces of Different Dimensions," *SIAM Journal on Matrix Analysis and Applications*, vol. 37, no. 3, pp. 1176–1197, 2016.

[18] Roberto Pereira, Xavier Mestre and David Gregoratti, "Clustering Subspaces in Large Dimensional Observations," *In preparation*.