The p-value as a New Similarity Function for Spectral Clustering in Sensor Networks

Mael BOMPAIS, Hamza AMEUR, Dominique PASTOR, Elsa DUPRAZ
IMT Atlantique, Lab-STICC, UBL, 29238 Brest, France

***Abstract*—In this paper, we consider spectral clustering over data collected by a network of sensors. In this context, the spatial data distribution is not necessarily uniform and can further be affected by sensor noise. This is why we propose a new similarity measure for spectral clustering in sensor networks. This similarity function is derived as the p-value of an hypothesis test that has to decide whether two sensor measurements belong to the same cluster. Unlike other existing similarity measures, the p-value takes into account both the local data densities and the fact that the noise variance can vary from sensor to sensor. Simulation results show that the p-value leads to a better spectral clustering performance than the standard Gaussian kernel when there is some noise in the collected data.**

I. INTRODUCTION

Learning over data collected by a network of sensors is now an important issue in a wide range of applications such as event detection, target localization, or object tracking, see [1] for a review. In this context, clustering is an unsupervised learning task that consists of separating the collected measurements into clusters: measurements assigned to the same cluster are similar while measurements assigned to two different clusters are dissimilar. As an alternative to standard clustering algorithms such as K-means or Expectation-Minimization (EM), Spectral clustering [2] permits to separate the data into clusters that cannot be described by convex boundaries (see Figure 1 for an example).

Spectral clustering consists of two steps. It first maps the set of collected measurements into a similarity graph by computing all the similarities between every pair of measurements. It then applies the spectral clustering algorithm that partitions the similarity graph into subgraphs in order to constitute the clusters. A key aspect in spectral clustering resides in the choice of the similarity function that is used to build the similarity graph.

The most standard similarity function is the Gaussian kernel [3], although other similarity functions have been proposed [4]–[7]. The Gaussian kernel expression depends on a parameter that was shown to be homogeneous to the local density of data. This parameter can be estimated for every measurement by calculating the average distance between the measurement and its M nearest neighbors [3]. The local densities are supposed to capture the distribution of the data in the space. But in sensor networks, the data distribution can further be affected by the sensor noise, and the noise variance may vary from one sensor to another. As a result, the standard Gaussian kernel is not sufficient to completely characterize the similarity between measurements. To the best of our knowledge, the other existing similarity measures cannot take into account at the same time the local data density, the noise variance, and the fact that the noise variance can vary from sensor to sensor.

This paper proposes a new similarity function for spectral clustering. The proposed similarity function is derived as the p-value of a hypothesis test called Random Distortion Testing (RDT) [8]. The p-value measures the plausibility that two measurement vectors belong to the same cluster. It permits to take into account both the local data densities and the noise variances, even when each sensor has a different variance. The local data densities can be estimated as in [3]. The noise variance of a given sensor is either known from the sensor physical characteristics, or can locally be estimated via signal processing techniques [9], [10]. Simulation results show that the p-value yields better clustering performance than the Gaussian kernel when there is some noise in the collected data.

This paper is organized as follows. Section II introduces the statistical signal model for the sensors measurements. Section III describes the standard spectral clustering approach. Section IV derives our new similarity function as the p-value of a RDT test. Section V shows the simulation results.

II. SIGNAL MODEL

This section describes the notations and assumptions that will be used throughout the paper for data collection in a network of sensors. Consider a network of N sensors, where sensor $i \in \{1, \cdots, N\}$ observes a random d-dimensional measurement vector $\boldsymbol{Y}_i$. Denote by $\mathcal{Y} = \{\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_N\}$ the set of N measurement vectors. For every $\boldsymbol{Y}_i \in \mathcal{Y}$, we assume a Gaussian observation model $\boldsymbol{Y}_i = \boldsymbol{\theta}_i + \mathbf{Z}_i$, where $\boldsymbol{\theta}_i \in \mathbb{R}^d$ is unknown but deterministic and $\mathbf{Z}_i \sim \mathcal{N}(0, \sigma_i^2 I_d)$ is a random vector that has independent and identically distributed (i.i.d.) components with mean 0 and variance σ_i^2. The deterministic vector $\boldsymbol{\theta}_i$ represents the unknown original data and the random vector $\mathbf{Z}_i$ describes the sensor noise. Note that here, the statistical model is not on the original data $\boldsymbol{\theta}_i$ but only on the measurement noise $\mathbf{Z}_i$ from the sensors. We consider a standard Gaussian model for the noise, although the proposed

method can be extended to other observation and noise models.

III. Spectral Clustering

The objective of spectral clustering is to separate the set $\mathcal{Y}$ into K clusters. In spectral clustering [2], the value of K is usually assumed known. Although this assumption is not always easy to verify in practical situations, we do it here in order to focus on the effect of the similarity function on the clustering performance. In the following, we consider a similarity function $w : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, and we denote by $w_{i,j} = w(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ the similarity measure between each pair of vectors $(\boldsymbol{Y}_i, \boldsymbol{Y}_j) \in \mathcal{Y}^2$. The function w is positive and it is assumed symmetric in the sense that $w(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = w(\boldsymbol{Y}_j, \boldsymbol{Y}_i)$.

In this section, we describe standard spectral clustering from a generic similarity function w [2]. From the subsequent measures $w_{i,j}$, one can construct a similarity graph over the dataset $\mathcal{Y}$. Spectral clustering then determines the K clusters by partitioning the similarity graph into K subgraphs. In the following, we first describe the construction of the similarity graph from the dataset. We then present the spectral clustering algorithm that permits to partition the graph into K subgraphs.

A. Similarity Graph

The set of measurement vectors $\mathcal{Y}$ can be mapped into an undirected weighted graph $\mathcal{G} = (V, E)$ called the similarity graph. In $\mathcal{G}$, the set V is composed by N vertices, and vertex $v_i \in V$ represents the measurement vector $\boldsymbol{Y}_i$. There is an edge $e_{i,j} \in E$ $(i \neq j)$ between two vertices v_i and v_j as long as $w_{i,j} > 0$. Edge $e_{i,j} \in E$ is weighted by the similarity measure $w_{i,j}$. Also denote by $\mathcal{N}_i$ the index set of vertices that are connected to v_i.

From the similarity graph $\mathcal{G}$, the problem of clustering the dataset $\mathcal{Y}$ can be reformulated as a graph partitioning problem. The objective of graph partitioning is to separate $\mathcal{G}$ into K subgraphs such that edges within a subgraph have high weights whereas edges between two subgraphs have low weights. The Spectral Clustering algorithm provides a relaxed solution to the above graph partitioning problem. This algorithm relies on the so-called Laplacian matrix which we now define.

B. Laplacian Matrix

The Laplacian matrix L is defined as [2]

$$L = D - W. \tag{1}$$

In the above expression, the similarity matrix W is a square matrix of size $N \times N$. The matrix W contains the similarity measures $w_{i,j}$ at positions (i, j), $i \neq j$, and its diagonal terms $w_{i,i}$ are set to 0. Note that W is the matrix representation of the similarity graph $\mathcal{G}$. The degree matrix $D = \text{diag}(d_i)$ is a diagonal matrix of size $N \times N$. The diagonal component $d_i = \sum_{j \in \mathcal{N}_i} w_{i,j}$ contains the sums of the input weights for vertex v_i.

It can be shown that the Laplacian matrix L always admits 0 as eigenvalue. The multiplicity of the eigenvalue 0 is equal to the number of independent connected components in the graph. By independent connected component, we mean a subgraph $\mathcal{G}' \subseteq \mathcal{G}$, such that there is a path between any pair of vertices $\mathcal{G}'$ but there is no edge between any pair of vertices in $\mathcal{G}' \times (\mathcal{G} \setminus \mathcal{G}')$. The kernel E_0 of L is further generated by the indicator vectors of these independent connected components. The spectral clustering algorithm provides a partition of $\mathcal{G}$ into K disjoint subgraphs that are close to independent connected components in the sense that two different subgraphs have a limited amount of connections between each other.

C. Spectral Clustering algorithm

For a fixed value K, the spectral clustering algorithm [2] proceeds as follows. As a first step, it finds the first K eigenvectors $(\mathbf{u}_1, \cdots, \mathbf{u}_K)$ of the Laplacian matrix L. It then constructs a matrix U of size $N \times K$ that contains the eigenvectors $(\mathbf{u}_1, \cdots, \mathbf{u}_K)$ as columns. As a second step, it extracts the N rows $(\mathbf{r}_1, \cdots, \mathbf{r}_N)$ of U and applies the K-means algorithm over this set of rows in order to obtain K classes $\mathcal{A}_1, \cdots, \mathcal{A}_K$. These K classes provide a clustering of the set of data $\mathcal{Y}$.

This algorithm has shown good performance in various applications [4], [11]–[13]. The performance of the algorithm however depends on the choice of the similarity function w that computes the similarity measures $w_{i,j}$ between each pair of data $(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$. We now describe the standard Gaussian kernel [4] as well as the new similarity function we propose for spectral clustering in sensor networks.

IV. Similarity Functions

In this section, we first describe the standard Gaussian kernel. We then introduce a new similarity function that is defined as the p-value of a RDT test. Unlike the Gaussian kernel, the p-value takes both the local data densities and the sensor noise variances into account.

A. Gaussian Kernel

For any pair $(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ of measurements, the Gaussian kernel is defined as [4]

$$w_G(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = \exp\left(-\frac{\|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|^2}{\beta_i \beta_j}.\right) \tag{2}$$

For every measurement vector $\boldsymbol{Y}_i \in \mathcal{Y}$, the local parameter β_i represents the density of observations in the neighborhood of $\boldsymbol{Y}_i$. The parameter $\beta_i > 0$ can be estimated as the average Euclidian distance between $\boldsymbol{Y}_i$ and its M nearest neighbors [3].

Apart from the Gaussian kernel, several other similarity functions have been proposed in the literature, see [5] for an

overview. For instance, in [6], an extension of the Gaussian kernel was proposed to take both the local density and the measurement noise into account. This similarity function however assumes that all the sensors have the same noise variance, which is not always true in practical situations. This is why we now introduce our new similarity function which we derive as the p-value of a hypothesis test.

B. The Random Distortion Test

Our new similarity function is defined from a RDT hypothesis test [14]. In this section, we describe the general RDT test. In Section IV-C, we specify this test in the context of clustering, and derive the corresponding similarity function. Fix a parameter $\tau \in [0, \infty)$ and consider the following testing problem

$$\begin{cases} \text{Observation: } \boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2 I_d), \\ \text{Hypotheses: } \begin{cases} \mathcal{H}_0 : \|\boldsymbol{\theta}\| \leq \tau, \\ \mathcal{H}_1 : \|\boldsymbol{\theta}\| \geq \tau. \end{cases} \end{cases} \tag{3}$$

The null hypothesis $\mathcal{H}_0$ corresponds to the case where the norm of the unknown deterministic $\boldsymbol{\theta}$ is less than τ. The meaning of the parameter τ will be specified later in the paper, in connection with the clustering problem. Although $\boldsymbol{\theta}$ is unknown, problem (3) can be solved via the following hypothesis test applied to $\boldsymbol{Y}$.

A test $\mathfrak{T}$ is any measurable map from $\mathbb{R}^d$ to $\{0, 1\}$. Given $\boldsymbol{y} \in \mathbb{R}^d$, the value $\mathfrak{T}(\boldsymbol{y})$ returned by $\mathfrak{T}$ is the index of the hypothesis considered to be true and we say that $\mathfrak{T}$ accepts $\mathcal{H}_0$ (resp. $\mathcal{H}_1$) at $\boldsymbol{y}$ if $\mathfrak{T}(\boldsymbol{y}) = 0$ (resp. $\mathfrak{T}(\boldsymbol{y}) = 1$). Given $\alpha \in (0, 1)$, let $\lambda_\alpha(\tau)$ be the unique real value λ such that $Q_{d/2}(\tau, \lambda) = \alpha$, where $Q_{d/2}$ is the Generalized Marcum Function [15]. According to [14], the test defined for any $\boldsymbol{y} \in \mathbb{R}^d$ as

$$\mathfrak{T}_{\sigma\lambda_\alpha(\tau/\sigma)}(\boldsymbol{y}) = \begin{cases} 0 & \text{if} \quad \|\boldsymbol{y}\| \leqslant \sigma\lambda_\alpha(\tau/\sigma) \\ 1 & \text{if} \quad \|\boldsymbol{y}\| > \sigma\lambda_\alpha(\tau/\sigma). \end{cases} \tag{4}$$

guarantees a false alarm probability α. This test is called the RDT and it is shown to be optimal with respect to several criteria, see Appendix A and [14] for more details.

The p-value of a hypothesis test "*gives an idea of how strongly the data contradict the hypothesis*" [16, Sec. 3.3] and can be seen as a measure of the plausibility of the null hypothesis $\mathcal{H}_0$ given the observation.

Proposition 1. *The p-value of the test $\mathfrak{T}_{\sigma\lambda_\alpha(\tau/\sigma)}$ is given for any $\boldsymbol{y} \in \mathbb{R}^d$ by*

$$w(\boldsymbol{y}) = Q_{d/2}\left(\frac{\tau}{\sigma}, \frac{\|\boldsymbol{y}\|}{\sigma}\right). \tag{5}$$

Proof: The proof is given in Appendix A. ∎

We now restate the RDT test in the particular context of clustering, and derive our new similarity function as the p-value of this particular test.

C. The p-value as a Similarity Function

For a parameter $\tau_{i,j} \in [0, \infty)$, consider the following testing problem

$$\begin{cases} \text{Observation: } (\boldsymbol{Y}_i - \boldsymbol{Y}_j) \sim \mathcal{N}(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, (\sigma_i^2 + \sigma_j^2) I_d), \\ \text{Hypotheses: } \begin{cases} \mathcal{H}_0 : \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\| \leq \tau_{i,j}, \\ \mathcal{H}_1 : \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\| \geq \tau_{i,j}. \end{cases} \end{cases} \tag{6}$$

The RDT test for this problem is given by $\mathfrak{T}_{\sigma_{i,j}\lambda_\alpha(\tau_{i,j}/\sigma_{i,j})}$ defined from (4) and applied to $(\boldsymbol{Y}_i - \boldsymbol{Y}_j)$. For any pair $(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$, we define our new similarity function w_P as the p-value of this test. By setting $\sigma_{i,j} = \sqrt{\sigma_i^2 + \sigma_j^2}$ and according to Proposition 1,

$$w_P(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = Q_{d/2}\left(\frac{\tau_{i,j}}{\sigma_{i,j}}, \frac{\|\boldsymbol{Y}_i - \boldsymbol{Y}_j\|}{\sigma_{i,j}}\right). \tag{7}$$

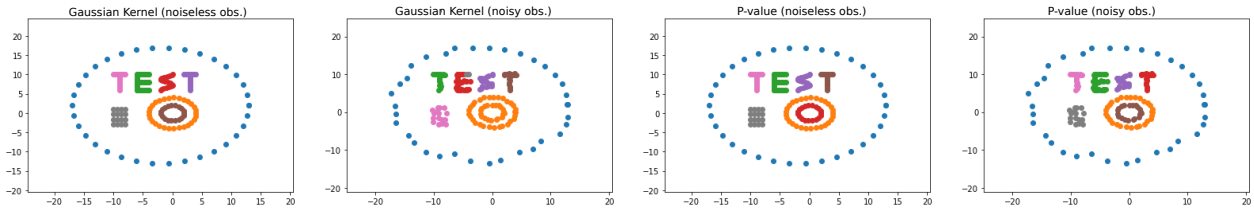
In the above expressions, $\tau_{i,j}$ is a design parameter that we choose as follows. We set $\tau_{i,j} = min(\tau_i, \tau_j)$ and the parameter τ_i (resp. τ_j) is estimated exactly as β_i (resp. β_j) in (2), that is as the average Euclidian distance between $\boldsymbol{Y}_i$ (resp. $\boldsymbol{Y}_j$) and its M nearest neighbors. The parameter τ_i is thus an estimate of the radius of the local neighborhood of $\boldsymbol{\theta}_i$. In more details, every $\boldsymbol{\theta}_j$ within the ball of radius τ_i and centered in $\boldsymbol{\theta}_i$ is considered to be in the same cluster as $\boldsymbol{\theta}_i$. We set $\tau_{i,j}$ as the minimum value between τ_i and τ_j in order to: 1) have a symmetric condition between $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$, 2) consider the most stringent radius between $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$, 3) penalize a high difference of local neighborhood between $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$.

In order to show the interest of our new similarity function, we now present simulation results that evaluate spectral clustering applied with either the Gaussian kernel or the p-value.

V. Simulation Results

In order to evaluate the performance of the spectral clustering algorithm with the two similarity functions described in Section IV, we consider the set of 211 data represented in Figure 1. In this Figure, we observe that each cluster has a different data density. Here, we consider two setups for the collected sensor measurements. In the first setup, we assume that the sensors directly collect the data $\boldsymbol{\theta}_i$ without any observation noise. In the second setup, each sensor collects a noisy version $\mathbf{y}_i$ of $\boldsymbol{\theta}_i$. In this setup, we assume that 160 sensors have variance $\sigma^2 = 0.1$, 16 sensors have variance $\sigma^2 = 0.2$, and 35 sensors have variance $\sigma^2 = 0.7$. The noise variance is chosen randomly for each sensor according to the above distribution.

In order to apply spectral clustering, we assume that the value of σ_i^2 is perfectly known for each sensor, and we estimate the values of τ_i with $M = 2$. We also set $K = 8$. Figures 1 (a) (b) show the results of Spectral Clustering with the Gaussian kernel over the two considered setups, and Figures 1 (c) (d) show the results with the p-value. In the noiseless setup, the two similarity functions correctly recover all the clusters. On the opposite, in the noisy setup, we observe



(a) (b) (c) (d)

Fig. 1. Spectral clustering over the set of 211 data applied over noiseless and noisy observation for the Gaussian kernel and the p-value

that the Gaussian kernel sometimes fails at retrieving the correct clusters. For example, the "E" was separated into two clusters when using the Gaussian kernel. Over this dataset, the p-value can hence lead to a better clustering performance when there is some noise in the collected data.

VI. Conclusion

In this paper, we proposed a new similarity measure for spectral clustering in sensor networks. Our new similarity measure was derived as the p-value of a RDT hypothesis test. The p-value takes into account both the local data density and the fact that each sensor has a different noise variance. In this case, the proposed similarity function shows better clustering performance than the standard Gaussian kernel.

In this paper, the p-value is obtained from a Gaussian model for the sensor noise, but the proposed method can be extended to other observation models (Gaussian with non i.i.d. components, Laplacian, etc.).

Appendix A
P-value of test $\mathfrak{T}_{\sigma\lambda_\alpha(\tau/\sigma)}$

In this section, we recall some basic properties of the test $\mathfrak{T}_{\sigma\lambda_\alpha(\tau/\sigma)}$ before deriving its p-value and thus establishing Proposition 1. Consider the problem of testing the null hypothesis $\mathcal{H}_0 : \|\boldsymbol{\theta}\| \leqslant \tau$ against the alternative hypothesis $\mathcal{H}_1 : \|\boldsymbol{\theta}\| > \tau$, when the observation is $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2\mathbf{I}_d)$, $\boldsymbol{\theta} \in \mathbb{R}^d$ is unknown and $\tau \in [0, \infty)$ is given.

For this problem, the power function $\beta_{\mathfrak{T}}$ of a given test $\mathfrak{T} : \mathbb{R}^d \to \{0, 1\}$ assigns to each $\boldsymbol{\theta} \in \mathbb{R}^d$ the probability value [16]:

$$\beta_{\mathfrak{T}}(\boldsymbol{\theta}) = \mathbb{P}\big[\mathfrak{T}(\boldsymbol{Y}) = 1\big], \tag{8}$$

and the size of this same test is

$$\alpha_{\mathfrak{T}} = \sup_{\boldsymbol{\theta}\in\mathbb{R}^d:\|\boldsymbol{\theta}\|\leqslant\tau} \beta_{\mathfrak{T}}(\boldsymbol{\theta}). \tag{9}$$

Given some level $\alpha \in (0, 1)$, there is no Uniformly Most Powerful (UMP) test for the binary hypothesis testing problem specified by $\mathcal{H}_0$ and $\mathcal{H}_1$ [16]. It is thus natural to seek a test optimal within a restricted class of tests relevant for the problem. In this respect, we note that the testing of $\mathcal{H}_0$ against $\mathcal{H}_1$ is invariant [16, Chapter 6] under the action of the group $\boldsymbol{O}_d$ of all orthogonal matrices [14, Sec. 3]. According to [14, Theorem 2], the test $\mathfrak{T}_{\sigma\lambda_\alpha(\tau/\sigma)}$ defined by (4) is UMP with size α among all $\boldsymbol{O}_d$-invariant tests and among all tests with constant power function on every sphere centered at the origin of $\mathbb{R}^d$ with radius $\rho \geqslant 0$. This result generalizes [17, Proposition III, p. 450] on the testing of the mean of a Gaussian, which corresponds to the specific case $\tau = 0$.

Given $\alpha \in (0, 1)$, the critical or rejection region S_α of $\mathfrak{T}_{\sigma\lambda_\alpha(\tau/\sigma)}$ is

$$\begin{aligned} S_\alpha &= \{\boldsymbol{y} \in \mathbb{R}^d : \mathfrak{T}_{\sigma\lambda_\alpha(\tau/\sigma)}(\boldsymbol{y}) = 1\} \\ &= \{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{y}\| > \sigma\lambda_\alpha(\tau/\sigma)\} \end{aligned}$$

The critical regions S_α are nested in the sense that $\alpha < \alpha' \Rightarrow S_\alpha \subset S_{\alpha'}$. This is a consequence of the following lemma.

Lemma 1. *Given $\tau \in [0, \infty)$, the map $\alpha \in (0, 1] \mapsto \lambda_\alpha(\tau) \in [0, \infty)$ is strictly decreasing.*

Proof: Given $\rho \in [0, \infty)$ and $\alpha, \alpha' \in (0, 1]$, we have $Q_{d/2}(\rho, \lambda_\alpha(\rho)) = \alpha$ and $Q_{d/2}(\rho, \lambda_{\alpha'}(\rho)) = \alpha'$. If $\alpha < \alpha'$ then $Q_{d/2}(\rho, \lambda_\alpha(\rho)) < Q_{d/2}(\rho, \lambda_{\alpha'}(\rho))$, since $Q_{d/2}(\tau/\sigma, \cdot)$ is strictly decreasing [15]. We thus have $\lambda_\alpha(\rho) > \lambda_{\alpha'}(\rho)$. ■

Therefore, given any $\boldsymbol{x} \in \mathbb{R}^d$, we can define the p-value [16, Sec. 3.3, p. 63] of the family $\{\mathfrak{T}_{\sigma\lambda_\alpha(\tau/\sigma)}\}_{\alpha\in(0,1)}$ of tests as the quantity

$$\begin{aligned} \widehat{p}(\boldsymbol{y}) &= \inf\{\alpha \in (0, 1) : \boldsymbol{x} \in S_\alpha\} \\ &= \inf\{\alpha \in (0, 1) : \|\boldsymbol{y}\| > \sigma\lambda_\alpha(\tau/\sigma)\}. \end{aligned} \tag{10}$$

We now prove that

$$\widehat{p} = Q_{d/2}(\tau/\sigma, \|\boldsymbol{y}\|/\sigma). \tag{11}$$

To this end, set $\alpha^* = Q_{d/2}(\tau/\sigma, \|\boldsymbol{y}\|/\sigma)$. By definition of $\lambda_{\alpha^*}(\tau/\sigma)$, we have $\alpha^* = Q_{d/2}(\tau/\sigma, \lambda_{\alpha^*}(\tau/\sigma))$. The bijectivity of $Q_{d/2}(\tau/\sigma, \cdot)$ implies that $\|\boldsymbol{y}\| = \sigma\lambda_{\alpha^*}(\tau/\sigma)$. It then follows from (10) that $\widehat{p}(\boldsymbol{y}) = \inf\{\alpha \in (0, 1) : \lambda_{\alpha^*}(\tau/\sigma) > \lambda_\alpha(\tau/\sigma)\}$. We then derive (11) from Lemma 1.

References

[1] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1996–2018, 2014.

[2] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[3] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Advances in neural information processing systems*, 2005, pp. 1601–1608.

[4] M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, "A survey of kernel and spectral methods for clustering," *Pattern recognition*, vol. 41, no. 1, pp. 176–190, 2008.

[5] H. Jia, S. Ding, X. Xu, and R. Nie, "The latest research progress on spectral clustering," *Neural Computing and Applications*, vol. 24, no. 7-8, pp. 1477–1486, 2014.

[6] X. Zhang, J. Li, and H. Yu, "Local density adaptive similarity measurement for spectral clustering," *Pattern Recognition Letters*, vol. 32, no. 2, pp. 352–358, 2011.

[7] X. Ye and T. Sakurai, "Similarity measure based on adaptive neighbors for spectral clustering," in *Proceedings of the 9th International Conference on Machine Learning and Computing*. ACM, 2017, pp. 405–409.

[8] D. Pastor and Q.-T. Nguyen, "Random distortion testing and optimality of thresholding tests," *IEEE Transactions on Signal processing*, vol. 61, no. 16, pp. 4161–4171, 2013.

[9] P. Huber and E. Ronchetti, *Robust Statistics, second edition*. John Wiley and Sons, 2009.

[10] D. Pastor and F. Socheleau, "Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, 2012.

[11] A. Kumar and H. Daumé, "A co-training approach for multi-view spectral clustering," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 393–400.

[12] X. Cai, F. Nie, H. Huang, and F. Kamangar, "Heterogeneous image feature integration via multi-modal spectral clustering," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1977–1984.

[13] B. Tian and W. Li, "Community detection method based on mixed-norm sparse subspace clustering," *Neurocomputing*, vol. 275, pp. 2150–2161, 2018.

[14] D. Pastor and Q.-T. Nguyen, "Testing the mahalanobis distance between a random signal with unknown distribution and a known deterministic model in additive and independent standard gaussian noise: the random distortion testing problem," Tlcom Bretagne, Research Report, RR-2012-04-SC, 2012.

[15] Y. Sun, A. Baricz, and S. Zhou, "On the Monotonicity, Log-Concavity, and Tight Bounds of the Generalized Marcum and Nuttall Q-Functions," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1166 – 1186, Mar. 2010.

[16] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses, 3rd edition*. Springer, 2005.

[17] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical Society*, vol. 54, no. 3, pp. 426 – 482, Nov. 1943.