Digit. Signal Process. 145 (2024) 104338

Contents lists available at ScienceDirect

Digital Signal Processing

journal homepage: www.elsevier.com/locate/dsp

Joint analysis and segmentation of time-varying data with outliers

Stefania Colonnese [a,*], Gaetano Scarano [a], Marcello Marra [a], Panos P. Markopoulos [b,1], Dimitris A. Pados [c]

[a] Sapienza University of Rome, Rome, Italy
[b] The University of Texas at San Antonio, San Antonio, TX, 78249, USA
[c] Florida Atlantic University, Boca Raton, FL 33431, USA

ARTICLE INFO

Keywords:
Data segmentation
L1-norm principal-component analysis
Time series
Outliers
Subspace clustering

ABSTRACT

Principal-Component Analysis (PCA) is a fundamental tool in data science and machine learning, used for compressing, analyzing, visualizing, and processing large datasets. At the same time, temporal segmentation is important for coherent component analysis of big data collections generated by time-varying distributions. However, both segmentation and PCA can be critically affected and misled by corrupted points that often exist in big data collections. To address these issues, we propose a novel and robust method for joint segmentation and principal-component analysis of time-varying data, based on L1-norm formulations. Our proposed method estimates robust L1-norm principal components (L1-PCs) over different temporal horizons and combines them to perform outlier detection, data segmentation, and subspace estimation. Numerical studies on real-world data, including videos and smartphone-sensed human body motion measurements, corroborate the merits of the proposed method in terms of segmentation, PCA, and outlier detection/removal.

1. Introduction

Principal-Component Analysis (PCA) is commonly used in data science and machine learning. By applying PCA to large datasets, we can identify the main subspace and use it for various tasks such as denoising, dimensionality reduction, compression, feature extraction, and visualization. However, in many applications, the data subspaces change over time [1], smoothly or abruptly, as for example in digital healthcare [2], wireless communications [3], and autonomous vehicles [4], among others. Subspace shifts can either be continual or more abrupt, in the form of successive subspace-coherent segments (i.e., data segments within which the data subspace remains largely invariant).

Regardless of the kind of temporal subspace shifting, any PCA on the entire time series data will fail to capture the temporal coherence or the principal-component evolution. This motivated us to explicitly address the joint data segmentation and analysis for the first type of temporal variation (continual) while providing in the process an approximate solution for the second type of variation (abrupt).

In essence, time-series segmentation can be viewed as a form of data clustering [24]. Indeed, there exist several methods in the literature that extend clustering to temporal segmentation and change detection. A main difference is that standard clustering methods (such as K-means) do not consider any temporal order of the data, which is a critical factor in time-series segmentation. For example, subspace clustering methods, such as Generalized PCA (GPCA), do segment (cluster) the dataset based on subspace coherence, but the segments may not be successive in time, as desired in the time-series analysis problem at hand [6–8].

Apart from clustering, subspace tracking methods have also been used for temporal segmentation. While tracking methods do take into account the temporal order, they are better suited for capturing continual subspace shifts rather than explicitly identifying data segments [30,31,51].

Clustering, tracking, and segmentation methods alike can be susceptible to corrupted and outlying points in the dataset. Cluster centers can be misled, tracked subspaces can be skewed, and non-existing data segments might be falsely identified. Indeed, real-world big data collections often contain outliers resulting, for example, from sensor mal-

* Corresponding author.
E-mail address: stefania.colonnese@uniroma1.it (S. Colonnese).

[1] P. P. Markopoulos's work was supported in part by the U.S. National Science Foundation under Award OAC-1808582 and in part by the Air Force Office of Scientific Research under Award FA9550-20-1-0039.

https://doi.org/10.1016/j.dsp.2023.104338

Available online 5 December 2023

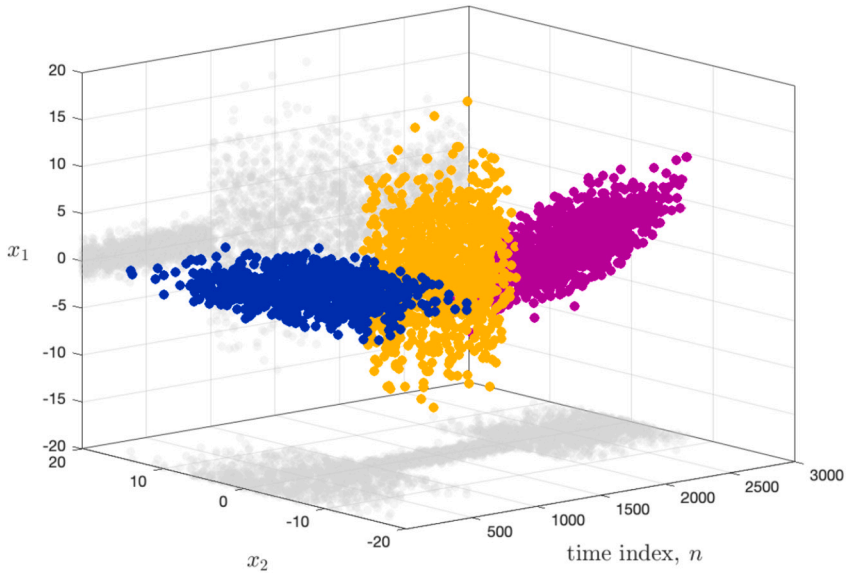


Fig. 1. Two-dimensional data $\{\mathbf{x}_n\}_{n=1}^{3000}$ from a time-varying Gaussian distribution, plotted versus time index n.

functions, errors in data storage/transcription, sensing interference, and even adversarial data poisoning attacks [25–28].

Therefore, there is a need for data analysis methods that are robust against outliers. In the area of PCA, two successful robust variants of standard PCA are L1-PCA [29,47] and Robust PCA (RPCA). Both are designed for batch processing of a complete data collection. Their respective Adaptive L1-PCA [30,31] and Robust Subspace Tracking [42] variants handle the problem of streaming data. However, none of these robust component analysis and subspace tracking methods have been specifically designed for joint data segmentation and subspace analysis of full rank data. In this work, we fill this gap in the literature, proposing a novel robust method for joint data segmentation and component analysis (R-JSCA): R-JSCA is built on the foundations of L1-PCA [29] and inherits its robustness against corrupted and outlying data. Specifically, R-JSCA

- employs the Hierarchical L1-PCA (H-L1-PCA) computation proposed in [32];
- uses the computed L1-PCs within a coarse-to-fine segmentation approach;
- the coarse stage discards outliers and detects the main temporal data segments;
- the fine stage acts on each data segment to detect the smaller temporal data changes.

As we demonstrate, R-JSCA is robust, computationally efficient, and can be easily extended to streaming-data processing.

The rest of the paper is organized as follows. In Sec. 2, we present our data model and problem statement. In Sec. 3, we review the technical background on the topic. In Sec. 4 we present the theoretical basis of the proposed R-JSCA method and in Sec. 5 we describe the R-JSCA algorithm. Sec. 6 presents numerical studies and comparisons on synthetic and real-world data. Concluding remarks are drawn in Sec. 7.

2. Data model and problem statement

We consider time-series data $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \subset \mathbb{R}^D$, partitioned into T successive subspace-coherent segments as follows:

$$\mathcal{X} = \{\mathbf{x}_1, \ldots, \underbrace{\mathbf{x}_{n(t)}, \ldots, \mathbf{x}_{n(t+1)-1}}_{t\text{-th segment}}, \ldots, \mathbf{x}_N\} \subset \mathbb{R}^D,$$

where we denote by $n(t)$ the index of the first data point in segment $t \in [1:T]$. We also denote by

$$\mathcal{N}(t) := [n(t) : n(t+1) - 1]$$

the index set of all data points in segment t.[2] Based on these definitions: $n(1) = 1$, $\bigcup_{t=1}^{T} N(t) = [1:N]$, $\mathcal{N}(t) \cap \mathcal{N}(\tau) = \emptyset$ if and only if (iff) $\tau > t$ and $n(\tau) > n(t)$ iff $\tau > t$. Then, for any $n \in \mathcal{N}(t)$–i.e. for each data point in the t-th segment–we define

$$\mathbf{x}_n = \mathbf{U}(t)\mathbf{r}_n + \mathbf{w}_n \in \mathbb{R}^D, \tag{1}$$

where $\mathbf{U}(t) \in \mathbb{V}(D, K(t)) := \{\mathbf{U} \in \mathbb{R}^{D \times K} : \mathbf{U}^\top \mathbf{U} = \mathbf{I}_{K(t)}\}$ is a basis for the subspace of segment t and $\mathbf{r}_n$ is a random subspace coordinate vector. We also define the signal subspace $\mathcal{S}(t) := \text{span}(\mathbf{U}(t))$, with dimensionality $K(t) \leq D$. $\mathbf{w}_n \in \mathbb{R}^D$ is a additive zero-mean noise vector, independent of $\mathbf{r}_n$, with covariance matrix $\mathbf{\Sigma}(t)$. Based on this model, the mean-squared distance of $\mathbf{x}_n \in \mathcal{N}(t)$ from $\mathcal{S}(t)$ is

$$\begin{aligned} &\mathbb{E}\{\|\mathbf{x}_n - \mathbf{U}(t)\mathbf{U}(t)^\top \mathbf{x}_n\|_2^2\} \\ &= \mathbb{E}\{\| \left(\mathbf{I}_D - \mathbf{U}(t)\mathbf{U}(t)^\top\right) \mathbf{w}_n\|_2^2\} \\ &\leq \mathbb{E}\{\|\mathbf{w}_n\|_2^2\} = \text{Tr}(\mathbf{\Sigma}(t)) \end{aligned} \tag{2}$$

Thus, as noise variance $\text{Tr}(\mathbf{\Sigma}(t))$ tends to 0, the noise is benign and $\mathbf{x}_n$ tends to be entirely in subspace $\mathcal{S}(t)$. However, in real-world applications, a small fraction of the data-points might also be corrupted by high-variance outliers of unknown statistics, so they highly deviate from the nominal subspace of their segment. For example, in Fig. 1 we show ($D = 2$)-dimensional zero-mean Gaussian data plotted over time, defining $T = 3$ successive subspace-coherent segments.

The problem of interest is described as follows. Given the time-series data $\mathcal{X}$, we wish to jointly detect segment thresholds $\{\hat{n}(t)\}_{t=2}^{\top}$ by estimating the first component of the segment bases $\{\mathbf{U}(t)\}_{t \in [1:T]}$. Importantly, we need a robust method that will allow us to accomplish these tasks even if the time-series is corrupted by sporadic outliers of unknown statistics, at unknown frequency and locations.

[2] For $\forall z_1, z_2 \in \mathbb{Z}$ such that $z_1 < z_2$, we denote $[a:b] = \{z_1, z_1 + 1, \ldots, z_2\}$.

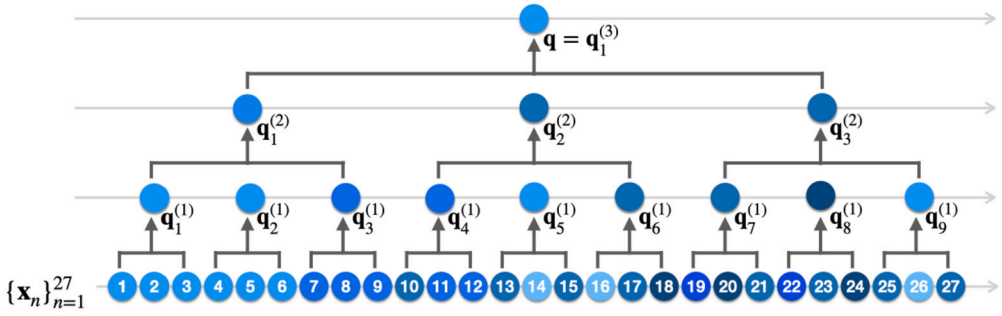


Fig. 2. Illustrative example of hierarchical L1-PCA [32], for $N = 27$, $L = 3$, and $W_1 = W_2 = W_3 = 3$.

3. Related work

3.1. Segmentation and component analysis approaches

Time-series segmentation and subspace analysis, as described in Sec. 2, are typically addressed using tracking-based or clustering-based analysis methods. Robust tracking methods provide a time-varying representation of the data subspace, which expresses a suitably selected data subset sliding in time [41,42,30,43]. On the other hand, clustering-based analysis can be modified to cluster time-series data and provide a basis for each cluster. Segment-based analysis can also be achieved by first segmenting the data [10] and then performing a component analysis on each segment.

Subspace clustering of time-varying data series [12,11,14] assumes that there is a finite number of distinct subspace segments in the data. Thus, clustering as in [13] is inherently challenged by slow/smooth subspace changes and is often addressed by considering substantially low rank [15] or application-specific [16–19] priors. Deep-learning approaches to change detection are either suited for low-dimensional time series data [20] or require large training datasets [21]. For segmentation and component analysis purposes, noisy streaming data with weak or no priors can be handled by robust component analysis algorithms. Besides, possible trends on data mean can also be addressed [22]. Segmentation can be also accomplished on weights profile obtained in subspace tracking, as in [23]. However, subspace tracking is a more complex operation addressing recursive computation of the complete subspace matrix and of its pseudo inverse, whereas not all the subspace components are necessary to detect a change in data statistics. Furthermore, change detection on the subspace weight profile needs thresholding operations which are not directly related to the data features (e.g., burstiness of outliers) making it difficult to fine-tune the segmentation to the actual data characteristics.

3.2. Robust component analysis and L1-PCA

Regarding robustness, several works have addressed the principal component analysis (PCA) of data affected by noise and outliers [33,35]. Outlier-resistant PCA has been widely studied in the literature [5,33]. Several methods aim to robustly and compactly represent the observed data by a constant vector [36–38], possibly leveraging a low-rank prior on the data subspace [34,39,40].

Outlier-resistant PCA methods, such as L1-norm based PCA (L1-PCA), bring robustness in data analysis. The L1-PC of a collection of data points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ is a vector $\mathbf{q}$ of normalized Euclidean norm that maximizes $\sum_{n=1}^{N} ||\mathbf{q}^\top \mathbf{x}_n||_1$. L1-PCA can be solved exactly as a combinatorial optimization problem [29], or approximately, e.g., by alternating optimization [44], fast Fourier transform [45], bit-flipping [46] (L1-BF), or hierarchical computation [32]. L1-PCA has also been extended to multi-way array (tensor) processing in [43,48–50]. Streaming versions of L1-PCA were proposed in [31] as well as in [30], where the authors associate a reliability to each sample by computing its scalar product with respect to (w.r.t.) the estimated principal component and leverage it to discard outliers.

3.3. Hierarchical L1-PCA

The proposed method presented below is based on robust hierarchical L1-PCA [32], summarized as follows. First, the $N_0 = N$ input data points $\{\mathbf{x}_n \in \mathbb{R}^D\}_{n=1}^{N_0}$ are divided into N_1 groups of size $W_1 = N_0/N_1$. Then, the L1-PC of each of these groups is calculated, e.g., by means of the bit-flipping algorithm [46]. The L1-PC of group $n \in [1 : N_1]$ is denoted by $\mathbf{q}_n^{(1)}$. The superscript (1) denotes that this is the first layer of L1-PC computations. Then, the N_1 layer-1 L1-PCs $\{\mathbf{q}_n^{(1)}\}_{n=1}^{N_1}$ are divided into N_2 layer-2 groups of size $W_2 = N_1/N_2$ and the L1-PC of each group is computed. The L1-PC of group $n \in [1 : N_2]$ in layer-2 is denoted by $\mathbf{q}_n^{(2)}$. Similarly, at layer $l = 3, 4, \ldots$, we divide the N_{l-1} L1-PCs from layer $l-1$, $\{\mathbf{q}_n^{(l-1)}\}_{n=1}^{N_{l-1}}$, into $W_l = N_{l-1}/N_l$ groups and we calculate the L1-PC of each group. We terminate at layer L such that $N_L = 1$. The resulting top-layer L1-PC $\mathbf{q} = \mathbf{q}_1^{(L)}$ constitutes an approximate L1-PC of the entire dataset, found in a hierarchical way. Moreover, importantly, hierarchical L1-PCA returns a group of hierarchical L1-PCs (H-L1-PCs):

$$\mathcal{Q}(\mathcal{X};\ \mathcal{W}) := \left\{ \{\mathbf{q}_n^{(l)}\}_{n \in [1:N_l]} \right\}_{l \in [1:L]}, \tag{3}$$

where $\mathcal{W} = \{W_l\}_{l \in [1:L]}$. The H-L1-PCs in $\mathcal{Q}(\mathcal{X};\ \mathcal{W})$ are used by the proposed method for joint segmentation and component analysis as described in the following sections. An illustrative example of hierarchical L1-PCA is offered in Fig. 2.

4. Proposed method

Herein, we propose to detect subspace changes assessing how well each data sample aligns with sets of neighbors of different widths. The underlying idea is that the alignment random fluctuations are expected to be smaller in absence than in presence of subspace changes. Therefore, in this section we firstly introduce the multi-scale conformity and then we analyze its statistical properties in a reference case. The next section will describe the segmentation algorithm based on the multi-scale conformity statistic.

4.1. Definition of multi-scale conformity

In this work, we utilize the H-L1-PCs presented above and build a method for multi-scale analysis of the input data. First, we notice that, for any layer $l > 1$ and index $n \in [1 : N_l]$, only a subset of the data points $\{x_m\}_{m \in [1:N]}$ have contributed to the formation of the H-L1-PC $\mathbf{q}_n^{(l)}$. We denote by $\omega_n^{(l)}$ the index set of those data points and find that $\omega_n^{(1)} = [(n-1)W_1 + 1 : nW_1]$ and, for every $l > 2$, $\omega_n^{(l)} = \bigcup_{j \in [(n-1)W_l + 1\ :\ nW_l]} \omega_j^{(l-1)}$. Expanding this recursive definition, we find that for any $l \geq 1$ and $n \in [1 : N_l]$,

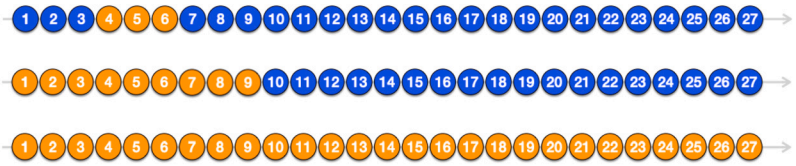


Fig. 3. We consider $N = 27$ points and calculate the conformity of $\mathbf{x}_6$ at $L = 3$ layers ($W_1 = W_2 = W_3 = 3$). Row-wise from top to bottom, the orange points correspond to the conformity cohorts $\{\mathbf{x}_n\}_{n\in\omega_6^{(1)}}$, $\{\mathbf{x}_n\}_{n\in\omega_6^{(2)}}$, and $\{\mathbf{x}_n\}_{n\in\omega_6^{(3)}}$, respectively.

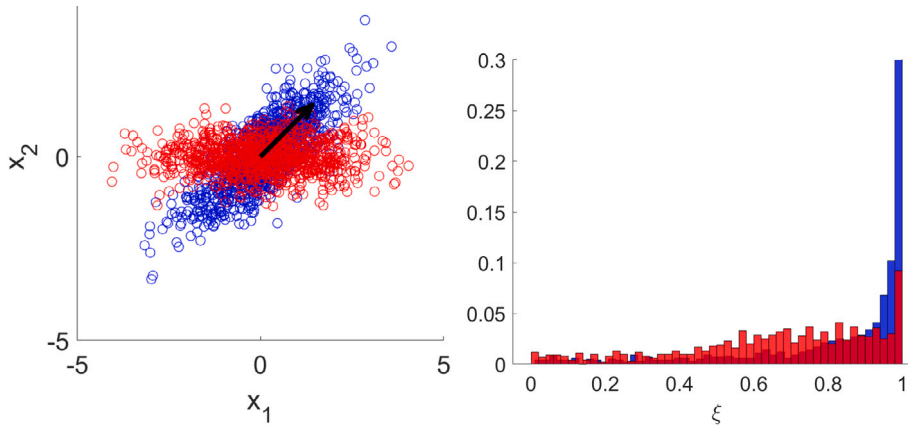


Fig. 4. (a) Points from $\mathcal{D}(\pi/4, 0.8)$ (blue) and $\mathcal{D}(0, 0.8)$ (red). (b) Histogram of conformity of blue points (blue) and red points (red), w.r.t. the eigenvector of the blue points.

$$\omega_n^{(l)} = \left[(n-1)L_l + 1 : nL_l\right], \tag{4}$$

where $L_l = \prod_{j=1}^{l} W_j$. For example, $\omega_2^{(2)} = [(W_2)W_1 + 1 : (W_2+1)W_1] \cup [(W_2+1)W_1 + 1 : (W_2+2)W_1] \cup \cdots \cup [(2W_2-1)W_1 + 1 : 2W_2W_1] = [W_1W_2 + 1 : 2W_1W_2]$. Moreover, we notice that the cardinality of $\omega_n^{(l)}$ is $|\omega_n^{(l)}| = L_l = \prod_{j=1}^{l} W_j$ (constant across n). That is, the number of data points that participate to the formation the H-L1-PCs at layer l increases as l increases. In other words, H-L1-PCs of higher layers summarize data points across a wider temporal span.

At the same time, each data point $\mathbf{x}_n$ contributes to the formation of a single H-L1-PC at each layer l. Let $\nu(n,l) \in [1 : N_l]$ denote the index of that H-L1-PC (it holds $\nu(n,l) \in \omega_n^{(l)}$). That is, at layer l, data point $\mathbf{x}_n$ is only represented by H-L1-PC $\mathbf{q}_{\nu(n,l)}^{(l)}$. Our method uses $\mathbf{q}_{\nu(n,l)}^{(l)}$ to assess the reliability or conformity of $\mathbf{x}_n$ at layer l. Specifically, for any $n \in [1 : N]$ and $l \in [1 : L]$, we define the layer-l conformity of $\mathbf{x}_n$ as

$$\xi_n^{(l)} = \frac{|\mathbf{x}_n^\top \mathbf{q}_{\nu(n,l)}^{(l)}|}{\|\mathbf{x}_n\|_2} \leq 1. \tag{5}$$

Based on the above, through $\mathbf{q}_{\nu(n,l)}^{(l)}$, $\xi_n^{(l)}$ captures the conformity of $\mathbf{x}_n$ to its layer-l cohort $\{\mathbf{x}_m\}_{m\in\omega_n^{(l)}}$. Importantly, since $|\omega_n^{(l)}|$ increases along l, $\xi_n^{(l)}$ at higher layers captures the conformity of $\mathbf{x}_n$ w.r.t. a broader cohort around it. For instance, $\xi_n^{(1)}$ captures the conformity of $\mathbf{x}_n$ w.r.t. just W_1 neighboring points, while $\xi_n^{(L)}$ captures the conformity of $\mathbf{x}_n$ w.r.t. the entire dataset. An example of this is offered in Fig. 3.

4.2. Note on conformity distribution

Depending on $\mathbf{x}_n$ and its layer-l cohort $\{\mathbf{x}_m\}_{m\in\omega_n^{(l)}}$, conformity $\xi_n^{(l)}$ is a random variable the distribution of which varies both along n and l –unless the noise is low enough and the data distribution/subspace remains stationary.

We study the stationary case with the following example. We consider zero-mean ($D = 2$)-dimensional normal distribution, with dominant eigenvector $\mathbf{q}$ at angle $\frac{\pi}{4}$ rad with the x-axis, unit marginal variances, and correlation coefficient equal to ρ. For simplicity in notation, we denote this distribution as $\mathcal{D}(\pi/4, \rho)$. For any data point $\mathbf{x}$ from this distribution, we denote by $\Xi = |\mathbf{x}^\top \mathbf{q}|/\|\mathbf{x}\|_2$ its conformity w.r.t. the dominant eigenvector $\mathbf{q}$ [53]. Then, we derive the probability density function (PDF) of Ξ. Denoting by ϕ the angle between $\mathbf{x}$ and the x-axis, we find $\Xi = |\cos(\Phi - \pi/4)|$ –that is, conformity captures the angular proximity of the data point to the dominant rank-1 subspace of the distribution, spanned by $\mathbf{q}$. Based on [54], the PDF of ϕ is

$$p_\Phi(\theta) = \frac{\sqrt{1-\rho^2}}{2\pi\,(1-\rho\sin(2\phi))} \quad \forall \phi \in [-\pi, \pi]. \tag{6}$$

Accordingly, through simple algebraic manipulations, it follows that the PDF of Ξ is

$$p_\Xi(\xi) = \frac{\sqrt{1-\rho^2}}{\pi\left(1+\rho-2\rho\,\xi^2\right)\sqrt{1-\xi^2}} \quad \forall \xi \in [0, 1]. \tag{7}$$

If instead of $\mathbf{q}$ we measure conformity w.r.t. a norm-1 vector $\mathbf{q}_0$ that is at angle Φ_0 with $\mathbf{q}$, the conformity PDF becomes

$$p_\Xi(\xi) = \frac{\sqrt{1-\rho^2}}{\pi\left(1+\rho s_0 - 2\rho\, s_0\xi^2 - 2\rho\, c_0\sqrt{1-\xi^2}\right)\sqrt{1-\xi^2}}, \tag{8}$$

where $c_0 = \cos(\pi/2 - 2\phi_0)$ and $s_0 = \sin(\pi/2 - 2\phi_0)$. Clearly, for $\phi_0 = \pi/4$, (8) boils down to (7). Next, we illustrate how the PDF of the conformity, when calculated with respect to the true eigenvector, can serve as an identifier of points drawn from the nominal distribution.

In Fig. 4(a), the blue points are drawn from $\mathcal{D}(\pi/4, 0.8)$ and the red points from $\mathcal{D}(0, 0.8)$. In Fig. 4(b), we plot histograms of the conformity of the blue points (blue bars) and the red points (red bars) w.r.t. the dominant eigenvector of the blue points, $\mathbf{q}$. Conformity of the blue points w.r.t. $\mathbf{q}$ takes most values very close to 1. On the other hand, conformity of red points w.r.t. $\mathbf{q}$ takes most values around 0.707, which is the cosine of the angle between $\mathbf{q}$ and the eigenvector of the red points

(i.e., the x-axis). Both histograms are in accordance with the theoretical PDF formulas presented above.

In practical computation, the dominant direction is not known but estimated via the rank-1 L1 PCA component. The random fluctuation discussed earlier serve as a reliable representation for the random fluctuation of the points with respect to the estimated rank-1 L1 PCA component, which accurately and robustly captures the principal covariance matrix eigenvector.

5. R-JSCA via multi-scale conformity

R-JSCA is a method that calculates the multi-scale conformity of a time-series samples w.r.t. H-L1-PCs (calculated as discussed in Sec. 3.3 above). The conformity evaluation leads to a transition detection that happens in two stages: a *coarse* detection stage and a *fine* detection stage. The coarse stage checks the conformity of each sample with respect to its closest neighbors, and it performs a preliminary coarse data segmentation. The fine stage operates within each segment by evaluating the sample conformity on widening neighbor sets, thus refining the coarse segmentation.

5.1. Coarse stage

Let us consider a time-varying data distribution. As far as new samples arrive, the PCs over different time spans are computed, up to the higher available layer L. The R-JSCA coarse stage analyzes the samples conformity $\xi_n^{(l)}\big|_{l=L}$ with respect to L versus the time index n. If the conformity $\xi_n^{(l)}\big|_{l=L}$ is above a threshold $\theta_0 = \theta_{\max}$, the n-th sample is deemed stationary and noise-free. If the conformity $\xi_n^{(l)}\big|_{l=L}$ is below the threshold $\theta_0 = \theta_{\max}$, a set of up to $\tau_{\max}$ consecutive samples is considered and matched towards decreasing thresholds $\theta_0 = \theta_{\max}$, $\theta_i = \alpha\,\theta_{i-1}$, $i = 1, \cdots, \tau_{\max}$, $\alpha < 1$, identifying extending confidence intervals as depicted in Fig. 5. Given the Union of Confidence Intervals (UCI) $\mathrm{UCI}_n = \bigcup_{i=0}^{t_n} (\theta_{i+1}, \theta_i]$ of increasing probability $p_i^{(\mathrm{UCI})} = \mathcal{P}r\left\{\xi \in (\theta_{i+1}, 1]\right\}$, R-JSCA computes the UCI test written as follows:

$$\xi_n^{(L)} \in \mathrm{UCI}_n = \bigcup_{i=0}^{t_n} (\theta_{i+1}, \theta_i], \tag{9}$$

where $t_n \in [0 : \tau_{\max}]$ is a counter of the number of consecutive samples violating the condition in (9). If $t_n < \tau_{\max}$ consecutive values violate the UCI condition (9), they are discarded as outliers. This is exactly the outlier rejection strategy proposed in [30] (case of adaptive threshold). Differently from [30] though, in this work, a recycling buffer stores the $t_n < \tau_{\max}$ samples that do not fit in the UCI. When $\tau_{\max}$ consecutive samples are stored in the recycling buffer, a change in the data pdf is inferred, the $\tau_{\max}$ samples in the buffer form the initial set of the new data segment, the buffer is emptied, the counter t_n is set to zero and the online data analysis restarted.

Both the threshold $\theta_{\max}$ and the buffer size $\tau_{\max}$ can be tuned to the statistics of the data and outlier process. The threshold $\theta_{\max}$ determines the ability to detect isolated outliers. If a prior model of the observations is available, the threshold can be tuned to the desired trade-off between false alarm probability and detection probability; details on the adjustment can be found in Appendix A. If no prior is available, but a set of data is available for training purposes, the threshold can still be set on the frequency of occurrences of the multi-scale conformity.

Also the parameter $\tau_{\max}$ can be set based on the desired trade-off between probability of detection and probability of false alarms. Appendix A examines in detail the impact of $\tau_{\max}$ on the probability of detection and probability of false alarms in case of independent outliers. Following the same line of reasoning, in case of burst of outliers, $\tau_{\max}$ is set above the minimum outlier burst length to be detected.

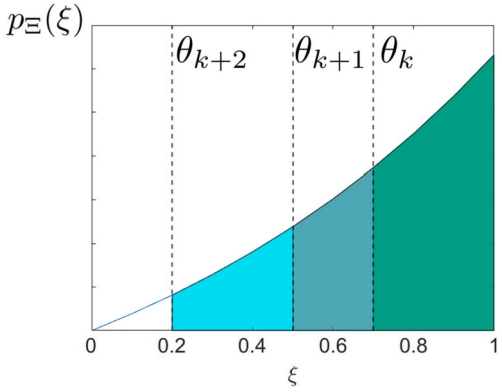


Fig. 5. Union of Confidence Interval criterion: PDF $p_\Xi(\xi)$ (black line), decreasing thresholds $\theta_k, \theta_{k+1}, \theta_{k+2}$ associated to increasing probabilities $p_k^{(UCI)}, p_{k+1}^{(UCI)}, p_{k+2}^{(UCI)}$ (shaded areas).

The points that fail the UCI conformity test (9) are stored in the recycling buffer: if they are less than $\tau_{\max}$ they are just discarded (green circle) as in [30] and [52]. Still, if the UCI test fails on $\tau_{\max}$ samples, this triggers the start of the new data segment. In summary, the coarse step checks for systematic failures of the UCI conformity test over consecutive samples to detect subspace changes in the observation time series. Differently from [30], the PC analysis window is not updated but restarted, and the following R-JSCA applies to detect further subspace changes. The performance of R-JSCA (coarse) is discussed in Appendix A.1, where a closed form analysis is carried out under simplifying assumptions.

5.2. Fine stage

The R-JSCA fine stage analyzes the samples conformity $\xi_n^{(l)}\big|_{l=L}$ with respect to the layer index l, corresponding to different temporal span. The conformity of the estimates achieved over different layers of the hierarchical L1-PCA architecture should consistently increase. Changes in the underlying pdf are spotted by inconsistencies of the estimates.

In case of stationary data, the conformity estimated at the $\overline{n}$-th time index, $\xi_{\overline{n}}^{(l)}$, $l = 0, \cdots, L-1$, is a constant mean r.v., with variance decreasing over increasingly larger data samples [56]. In presence of a subspace change, the conformity over larger data samples changes when the PC estimation interval spans across the transition, as illustrated in Fig. 6.

Herein, we detect a change by observing the conformity confidence interval [59,60]. Specifically, we adopt the Intersection of Confidence Intervals criterion [57,58] and we test the conformity $\xi_n^{(l)}$ over increasing values of l, corresponding to wider temporal intervals: if the data samples are stationary, the conformity pdfs the $\xi_n^{(l)}$ are expected to overlap, as in Fig. 6(a); on the contrary, if there is a subspace change, the conformity at higher layers (larger time span) is expected to lower and the overlap is lost, as in Fig. 6(b). Turning to math, the confidence intervals $\mathcal{I}_j(n)$ of $\xi_n^{(j)}$ over different layers $j = 1, \cdots, L$ are as follows

$$\mathcal{I}_n^{(j)} = \left(\xi_n^{(j)} - \Delta\xi_n^{(j)}, \xi_n^{(j)} + \Delta\xi_n^{(j)}\right], \tag{10}$$

where $\Delta\xi_n^{(j)}$ is a suitable function of the standard deviation of the sample conformity.[3] Then, the ICI is computed as

[3] Following the approach in [57], we set $\Delta\xi_n^{(j)} = \Gamma^{(j)} \cdot \sigma_n^{(j)}$, where $\sigma_n^{(j)}$ is the sample conformity standard deviation, estimated over a local window $\mathcal{W}$ and $\Gamma^{(j)}$ is equal to the $(1-\eta/2)^{th}$-percentile of the standard Gaussian distribution $\mathcal{N}(0, \eta/2)$ plus a correction factor $\frac{k}{W^l}$ which decreases as the layer index l increases.

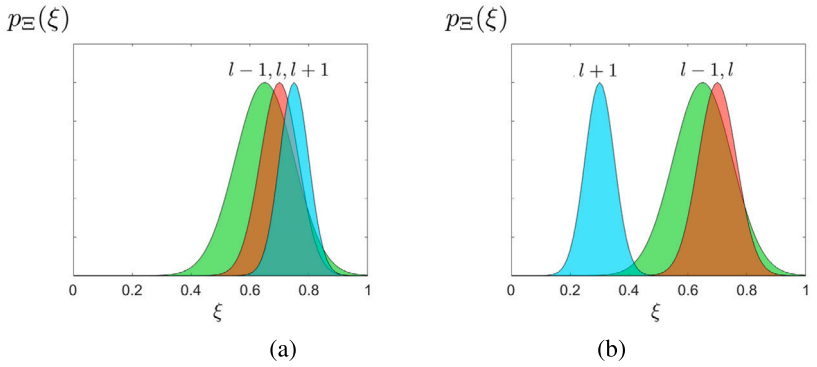


Fig. 6. Intersection of Confidence Interval criterion: PDFs of the j-th layer conformity $\xi_n^{(j)}$ for the n-th data point; (a) no subspace change (overlapping PDFs for increasing j), (b) subspace change (non overlapping PDFs for increasing j).

Algorithm 1 R-JSCA Coarse Stage algorithm pseudocode.

Input: Data matrix $\mathbf{X} \in \mathbb{R}^{D \times N_0}$, and currently estimated L1 PC of the L-th layer, $\mathbf{q}^{(L)}$.

$n = 0$, $s = 0$, $\mathbf{q}^{(L)} \leftarrow$ H-L1-PCA($\mathbf{X}_{:,0:N-1}$)
while $n < N - 1$
  $t = 0$, $\theta_0 = \theta_{\max} \in (0, 1)$, $RB = \emptyset$
  $n \leftarrow n + 1$
  while $\mathbf{x}_n^{\top}\mathbf{q}^{(L)}/|\mathbf{x}_n|^2 < \alpha^{t_n}\theta_0$ *and* $t \leq t_{max}$
    $RB \leftarrow RB \bigcup \{\mathbf{x}_n\}$
    $t \leftarrow t + 1$
    if $t = t_{max}$
      $T \leftarrow n$, $s \leftarrow s + 1$, $S \leftarrow s$
      $\mathbf{q}^{(L)} \leftarrow$ H-L1-PCA($\mathbf{X}_{:,T:N-1}$)
    end if
    $n \leftarrow n + 1$
  end while
  if $t \neq t_{max}$
    $\mathbf{X} \leftarrow \mathbf{X} \setminus RB$
  end if
end while

Return: Number of Coarse transitions S^C, and their locations n_s^C, $s = 0, \cdots, S^C - 1$, outlier-free data matrix $\mathbf{X}$.

$$\mathrm{ICI}_n^{(l)} = \bigcap_{j=1}^{l} \mathcal{I}_n^{(j)} . \tag{11}$$

The maximum layer index $l_n^{\max}$ for which the intervals overlap,

$$l_n^{\max} = \max_{l:\ \mathrm{ICI}_n^{(l)} \neq \emptyset} l \tag{12}$$

is monitored over n and, when $l_n^{\max}$ changes, a fine transition is detected, and the desired rank-1 PC (or higher rank PCs) are computed. Thereby, R-JSCA achieves a change detection resolution below the window size W, detecting changes at each and every sample.

5.3. Key takeaways

The multiscale conformity analysis carried out by the R-JSCA coarse and fine stages is summarized in Algorithms 1, 2. For clarity, also Figs. 7, 8 depict the R-JSCA coarse and fine stages, respectively: the original measurements come from different distributions (represented by different colors) and are affected by outliers (the green circle).

A few remarks are in order. Firstly, R-JSCA relies on the only assumption that the subspace change affects the first principal component of the data. Due to this mild assumption, R-JSCA tackles full-ranked time variant real data, which are rarely addressed in the literature.

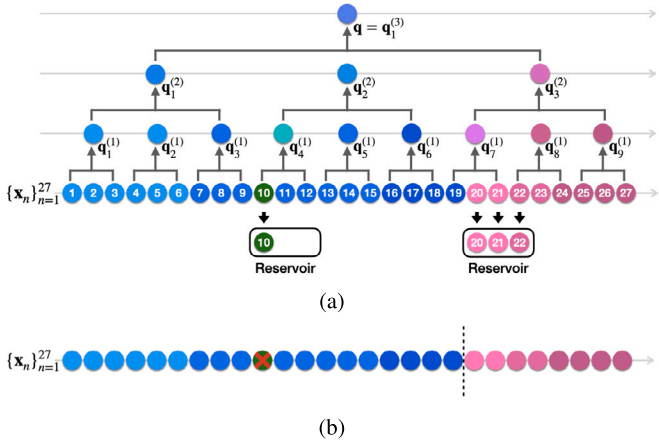


Fig. 7. R-JSCA (coarse): (a) Data points conformity test versus the time index n; unreliable points are identified and stored in reservoir; (b) Coarse stage result: outlier rejection and coarse segmentation.

Secondly, R-JSCA applies not only to batch data, but also to streaming data; the maximum layer L shall be selected so as to match the desired observation time span. Additionally, R-JSCA can manage both sporadic and consecutive outliers [55], by suitably setting the parameter τ_{max} of the Coarse stage performing outlier excision. Finally, herein we developed the analysis with reference to rank-1 L1-norm representation of a data subset, since it, as carried out in the foundation work, [46], captures its most important feature (basis vector) in the robust maximum-magnitude projection sense. We expect then, that monitoring the evolution of the rank-1 component alone across data subsets suffices to guide segmentation protocols. In arguably rare application cases where subtle changes over time do not show up in the first L1-norm principal component of the data, joint higher rank L1-norm decomposition may be pursued [29], [46] (or recursively projection of the data on to the subspace orthogonal to the first L1-norm principal component and disjoint computation of the next, etc. [36]).

In summary, the strength of R-JSCA in detecting subspace shifts, is related to a few key factors: i) the robustness of the adopted L1-PC, ii) the two stage computational architecture, which allows to distinctly address data errors versus slight subspace shifts, and iii) the tunability of the parameters to the desired false alarm versus detection trade-off. In terms of feasibility, the R-JSCA hierarchical computational structure assures that its computational complexity is reduced; for more details, the interested reader can refer to Appendix B). As for future work, R-JSCA relies on intermediate L1-PCs, and it can be extended to PCs computed with different metrics, provided that a layered architecture is adopted. Furthermore, R-JSCA relies on rank-1 PC analysis, but it can be generalized by i) introducing a conformity vector for the general rank-d PC

Algorithm 2 R-JSCA Fine Stage algorithm pseudocode.

Input: Outlier-free data matrix segments as identified by the fine stage $\mathbf{X}_{:,T^C:n^C_{s+1}}$.

$n = n^C_s$, $u = 0$, $S^F = 0$
compute $\mathbf{q}^{(l)}_{v(n,l)}$, $n = n^C_s \cdots n^C_{s+1}$, $l = 1, \cdots, L_{\max}$
while $n < n^C_{s+1}$
  $l = 1$, exit $= 0$
  while $(l < L_{\max})$ or (exit == 1)
    compute $\xi^{(l)}_n \leftarrow \mathbf{x}^\top_n \mathbf{q}^{(l)}_{v(n,l)} / |\mathbf{x}_n|^2$
    compute $\Delta\xi^{(l)}_n \leftarrow \text{confint}(n, l)$
    if $\left(\xi^{(l)}_n - \Delta\xi^{(l)}_n > \max_{j=1,\cdots,l-1} \left\{\xi^{(j)}_n + \Delta\xi^{(j)}_n\right\}\right)$
    or $\left(\xi^{(l)}_n + \Delta\xi^{(l)}_n < \min_{j=1,\cdots,l-1} \left\{\xi^{(j)}_n - \Delta\xi^{(j)}_n\right\}\right)$
      $n^F_u \leftarrow n$, $l_u \leftarrow l$ $u \leftarrow u + 1$
      exit $= 1$
    end if
  end while
  $n \leftarrow n + 1$
end while

Function: $\Delta\xi \leftarrow \text{confint}(n, l)$

$\hat{\sigma}^{(l)}_n = \left[Av \left\{ \left(\xi^{(l)}_n - Av\left\{\xi^{(l)}_n\right\}\right)^2 \right\} \right]^{1/2}$
$\Delta\xi^{(l)}_n = \hat{\sigma}^{(l)}_n \cdot \left(F_{1-\alpha/2} + k/W^l\right)$
Return $\Delta\xi$

Return: Number of Fine transitions S^F, their locations $n^F_u, u = 0, \cdots, S^F - 1$ and the associated selected layer $l_u, , u = 0, \cdots, S^F - 1$; the symbol Av denotes the average on $n \in \mathcal{W}$.

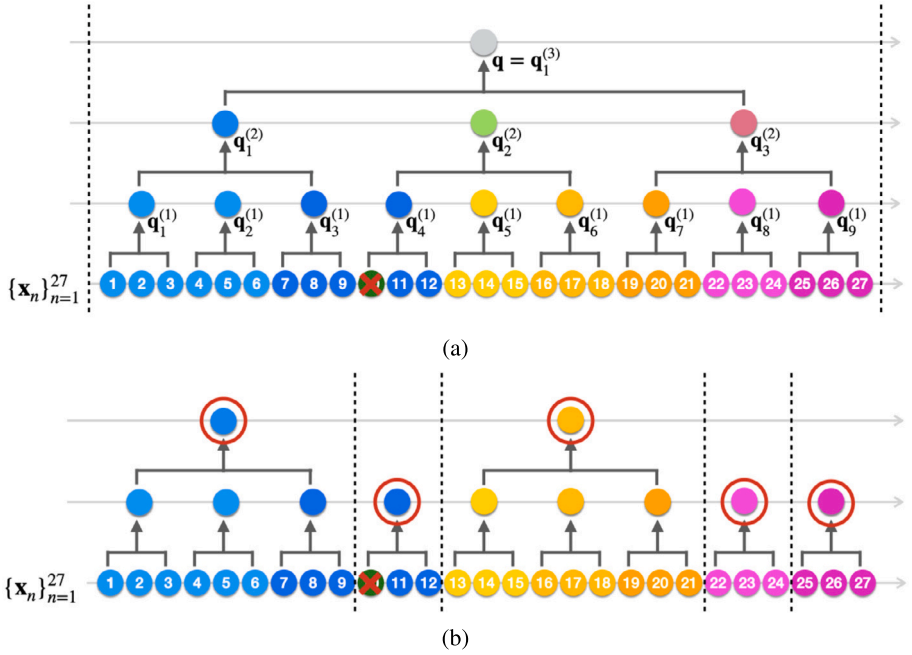


Fig. 8. R-JSCA (fine): (a) Outlier-free segmented data conformity test versus the layer index l; (b) Fine stage result: refined data segmentation and relevant rank-1 L1 PCs for each data segment.

case, and ii) analyzing how the distribution of the conformity vector changes corresponding to a subspace change. Finally, multiscale conformity can be exploited as a feature for a deep learning system. These topics are left for future studies.

6. Experiments

We assess the performance of R-JSCA on synthetic datasets as well as on two real datasets, referring to human activity related acceleration/speed signals acquired by a smartphone and to an outdoor, panning camera, highly dynamic video.

6.1. Synthetic datasets: off-on-off and rotating Gaussian

To assess the performance of R-JSCA, we first consider two $D = 2$, zero-mean, normally distributed data series, $N = 3740$ long, and differently changing in time. The covariance of the first data set changes at samples $n_1 = 935$ and $n_2 = 2806$, such that the data PDF assumes the

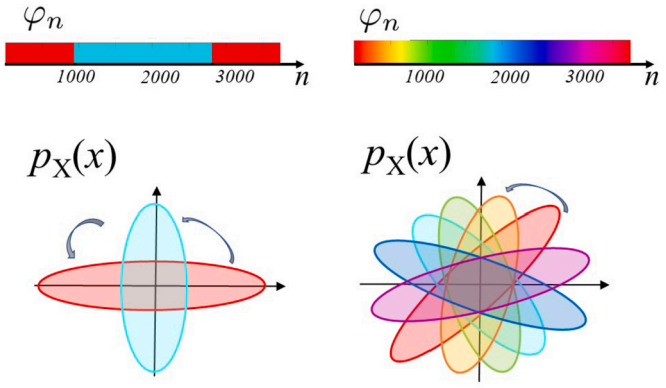


Fig. 9. Time varying distribution of the data: (left) off-on-off case and (right) linear phase shift case.

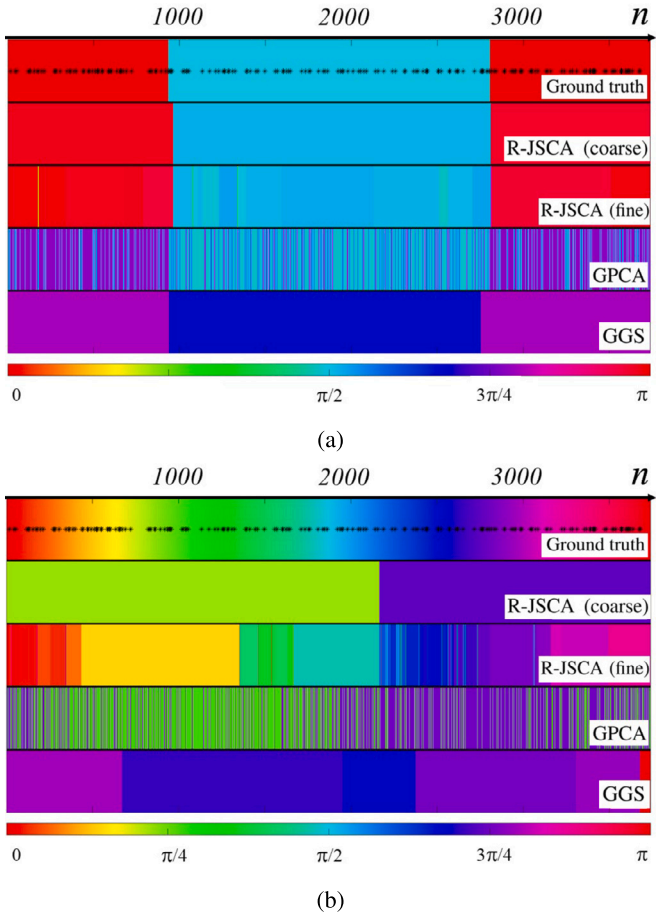


Fig. 10. Ground truth orientation φ_n and its estimates represented by pseudo-colored horizontal bars for the cases of (a) abrupt and (b) gradual transition; the time indexes of the 5% outliers are highlighted by black dots. R-JSCA (coarse) rejects the outliers and either grabs the abrupt transition or approximate the gradual one (a); R-JSCA (fine) step-wise approximates the gradual transition.

two principal axis directions: $\varphi_n = \frac{\pi}{2}$, for $n_1 < n < n_2$ and $\varphi_n = 0$ otherwise. Fig. 9(a) depicts the ground truth orientation by a pseudo-colored horizontal bar and sketches the concentration ellipses of the data distributions. The second dataset gradually rotates in time with $\varphi_n = \frac{n\pi}{N}$. Fig. 9(b) depicts φ_n and the associated rotating concentration ellipses. For both cases, 5% of data are randomly substituted with outliers, normally distributed around $(10, -10)$ with marginal variances $(1, 1)$.

The cascade of the R-JSCA coarse and fine segmentation stages is applied to both the datasets. Fig. 10 depicts the ground truth orientation φ_n, with outliers highlighted by black dots for the two cases. The estimated orientations at the output of the R-JSCA coarse and fine stages are also represented as horizontal colored bars. We compare the results with: GPCA [9], and GGS followed by segment-based PCA of [10].[4]

Table 1
Mean Square Error (dB) values for three synthetic datasets with off-on-off (first column), linear (second column), and constant (third column) trends (5% of outliers). R-JSCA (coarse) ranks first on off-on-off and constant datasets followed by R-JSCA (fine); R-JSCA (fine) ranks first on the linear trend, followed by R-JSCA (coarse).

Synthetic Dataset	Off-on-off	Linear	Constant
GPCA	-1.44 dB	-1.67 dB	-1.56 dB
GGS	-2.14 dB	-0.89 dB	1.96 dB
L2-PCA	-1.05 dB	-0.42 dB	1.96 dB
L1-PCA	-1.05 dB	-0.42 dB	-3.28 dB
H-L1-PCA (layer 5)	-1.05 dB	-0.42 dB	-12.12 dB
R-JSCA (coarse)	**-8.12 dB**	-2.95 dB	**-28.48 dB**
R-JSCA (fine)	-8.05 dB	**-5.95 dB**	-12.21 dB

In Fig. 10(a), R-JSCA (coarse) follows very well the ground truth trend, correctly detecting two subspace changes (with a delay of 28 and 5 samples). R-JSCA (fine) accurately estimates the data subspace orientation, with few small, spurious oscillations. Overall, R-JSCA outperforms competitors in jointly detecting the abrupt changes and adapting the L1-PCA estimation window to the time varying data. In Fig. 10(b), R-JSCA (coarse) detects one subspace change (at sample 2166), approximating the linear trend by two constants. Then, R-JSCA (fine) detects various smaller transitions, approximating the linear trend with several constant steps, and generating the segment-based PCs and achieving the best segmentation performance. R-JSCA performs consistently over different percentages of outliers, as shown in Fig. 11, illustrating, for different datasets (column a), the cases of absence of outliers (column b) and presence of 10% of outliers (column c).

In Table 1, we report the Mean Square Error (MSE) between the ground truth φ_n and the estimated orientation $\hat{\varphi}_n$, averaged over the time index n, for the two datasets, and for a reference constant dataset. R-JSCA (coarse) and R-JSCA (fine) are accurate estimators, outperforming the benchmarks GPCA and GGS as well as L2, L1 and H-L1-PCA, which approximate the orientation with a constant. On the first dataset, R-JSCA (coarse) and R-JSCA (fine) achieve MSE $= -8.12$ dB and MSE $= -8.05$ dB, respectively; on the second dataset R-JSCA ranks best (MSE $=$ -5.95 dB), followed by R-JSCA (coarse) (MSE $=$ -2.95 dB).

In summary, the proposed R-JSCA is accurate both on smoothly and abruptly changing data. We now show that R-JSCA captures meaningful transitions on two real datasets, referring to action recognition and video analysis, respectively.

6.2. *Real dataset: human activity smartphone accelerometer data*

Firstly, we challenge R-JSCA on the human activities datasets in [62,63], collecting time-variant acceleration/speed signals as acquired by a smartphone during different activities and postural transitions. R-JSCA applies to the spatial acceleration time series, resulting by moving average of the sensed samples. The observation matrix is built by 200 samples of the standing posture dataset, 23 samples of the stand-to-sit transition dataset and the 200 samples of sitting posture dataset. Fig. 12 plots one component of linear acceleration versus time, and it highlights in yellow the ground truth transition area (from sample 201 to sample 223). R-JSCA (coarse) correctly detects one transition at sample 206, no further transitions are detected by R-JSCA (fine). Thereby, R-JSCA grabs the statistical change associated with the postural change.

[4] The parameter setting is as follows: R-JSCA (coarse): $w_n = 20$, $\theta_{min} = 0.2$, $\theta_{max} = 0.9$, $\rho = 0.5$; R-JSCA (fine): $k = 4.10$, $\alpha = 0.05$; GGS: $\lambda = 0.1$, $K = 2$.

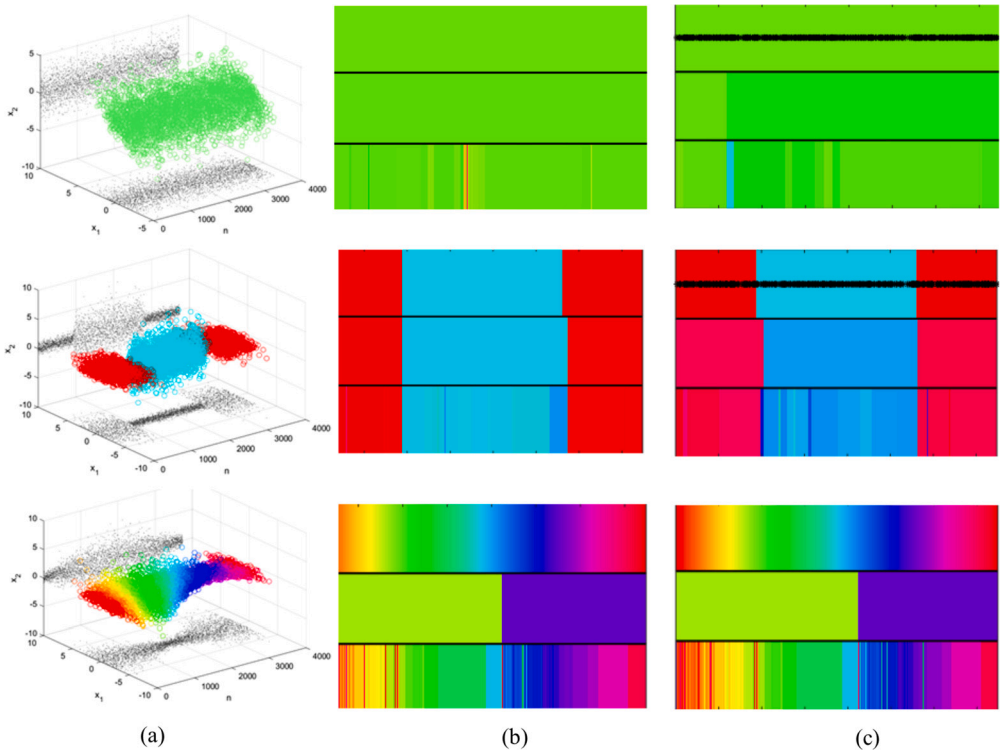


Fig. 11. R-JSCA of different time-varying synthetic data: (a) original dataset, (b) true, R-JSCA (coarse) and R-JSCA (fine) subspace directions in absence of outliers, and (c) true, R-JSCA (coarse) and R-JSCA (fine) subspace directions in presence of 10% of outliers.

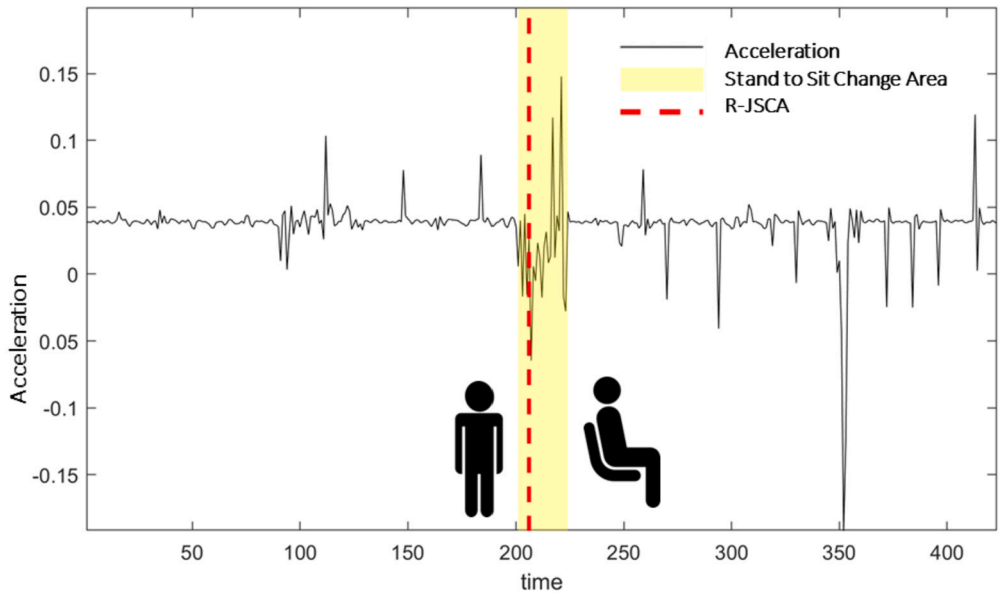


Fig. 12. R-JSCA application to the smartphone accelerometer dataset for human action recognition [62][63]; stand-to-sit analysis (b) of the first component of the acceleration vector versus time (black line). The ground truth interval of the stand-to-sit postural change is indicated by the yellow bar, and the transition detected by R-JSCA coarse stage is indicated by the dashed red line; no further transition was detected by the R-JSCA fine stage.

6.3. Real dataset: outdoor video with camera panning

Secondly, R-JSCA applies to the outdoor video sequence *Jumps* in Fig. 13, formerly introduced in [64] for the purpose of video key-frames extraction. Fig. 14 plots the subjective relevance of the sequence frames (black line) annotated by the authors in [64] based on users' frame scoring. The video camera rapidly pans and zooms throughout the video, posing challenges to the segmentation algorithms, since no low-rank assumptions stand. R-JSCA processes the grayscale, downsampled (from 480x270 a 160x90), vectorized video frames. After R-JSCA (coarse) discarded 78 outliers out of 950 frames, R-JSCA (fine) detected 26 subspace changes and computed the PCs of each segment. Fig. 14 depicts the time segments detected by R-JSCA as colored bars and the discarded outliers as narrow white stripes. Given the lack of ground truth segmentation, we compare R-JSCA segmentation with the subjective relevance measured in [64]. After an initial transitory (due to camera instability), the subspace changes identified by R-JSCA (fine) nicely match the peaks of the subjective score. R-JSCA catches sub-

Fig. 13. Selected thumbnails referring to different intervals of the video sequence Jumps: the scene undergoes major variations since the camera zooms and pans during the outdoor acquisition.

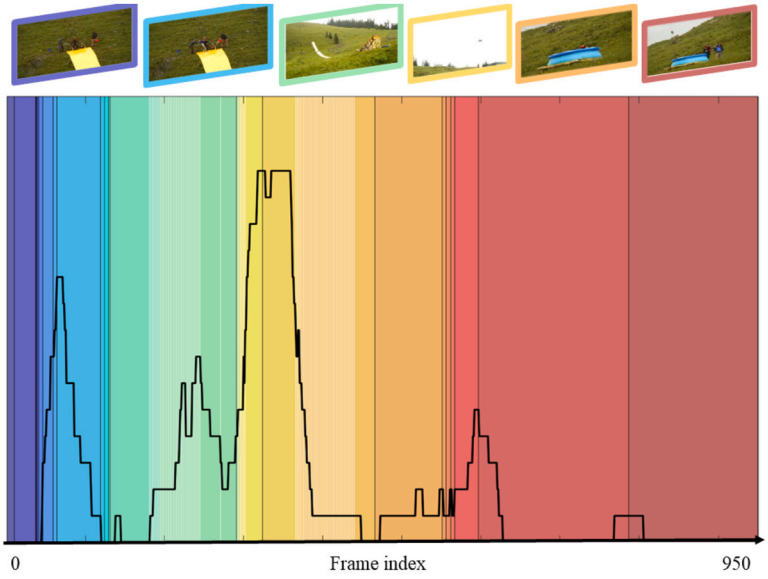


Fig. 14. R-JSCA results on the sequence Jumps: the segments detected by R-JSCA appear as colored bars while the outliers are represented by white stripes. The black line shows the fluctuations of the manually annotated frame visual relevance versus the frame index. R-JSCA detects the subspace changes on high innovation frames, corresponding to the peaks of the visual relevance, whereas larger segments span consecutive low-relevant, low-innovation frames.

space changes in correspondence to the high innovation frames that also trigger a high subjective interest. Instead, detected intervals span a larger number of frames when the subjective interest is low, i.e. on low innovation video segments. Thereby, R-JSCA provides a meaningful segmentation and analysis of the video data.

In summary, R-JSCA provides an accurate, meaningful segmentation and analysis of time-series, also rejecting outliers frequently occurring in real measurements. The two-tiers architecture, where R-JSCA (fine) detects small variations only after R-JSCA (coarse) has rejected outliers, avoids the curse of detecting gradual signal fluctuations amongst bigger outliers.

We remark that the R-JSCA coarse and fine stages are designed to work in synergy. The coarse stage rejects the outliers and it detects the larger changes; the fine stage detects smaller changes that would be challenging to identify without first removing outliers. The fine stage can be skipped when prior knowledge of the application suggests that a coarse segmentation is adequate.

Thanks to its mild assumptions, R-JSCA tackles data of different subspace structure kinds, including full-rank data, as shown above for a video shot by a moving camera. The analysis of identifiability conditions and performance guarantee on the basis of more restrictive hypotheses is left for further study.

7. Conclusion

This paper proposes a method for Joint Segmentation and Component Analysis (R-JSCA) based on multiscale data conformity analysis. By means of a two-tiers architecture, applicable both to batch and online data, R-JSCA provides a robust and accurate segmentation and analysis of time-series, managing both sporadic and consecutive outliers, without priors on the subspace structure of the observed data.

R-JSCA outperforms state-of-the-art competitors on synthetic and real datasets, where it can detect abrupt as well as smooth transitions. The results show that, thanks to its mild assumptions, R-JSCA tackles data of different subspace structure kinds, including full-rank data, e.g., a video shot by a moving camera. Finally, we observe that the proposed method holds promise to develop into a data-driven approach, where the pattern assumed by the multi-scale sample conformity can serve as features for segment change detection. Further investigation is worthwhile on the adaptation of the R-JSCA multiscale conformity analysis to deep learning systems.

CRediT authorship contribution statement

Stefania Colonnese: Conceptualization, Writing – original draft. **Gaetano Scarano:** Methodology, Writing – review & editing. **Marcello Marra:** Software, Validation, Writing – review & editing. **Panos P. Markopoulos:** Formal analysis, Funding acquisition, Methodology, Writing – original draft. **Dimitris A. Pados:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

P. P. Markopoulos reports financial support was provided by Air Force Office of Scientific Research. P. P. Markopoulos reports financial support was provided by U.S. National Science Foundation.

Data availability

We used either synthetic data explained in the text or public datasets.

Appendix A. R-JSCA performance analysis

Herein we address the closed form performance analysis of the R-JSCA approach for the case of bidimensional Gaussian random variable, under few simplifying assumptions. The herein derived performance, although ideal, provides a theoretically grounded insight on the R-JSCA computational architecture.

A.1. R-JSCA coarse stage: multiscale conformity analysis over time index

Herein, we consider the change detection [61] as a binary classification problem and we derive the closed form expression of the probability of detection and the probability of false alarm for the case of bidimensional, i.i.d. normal data samples. Let $\mathcal{H}_0$, $\mathcal{H}_1$ denote the hypothesis of absence and presence of subspace changes, respectively. R-JSCA detects a change if and only if τ_{max} consecutive samples fail the UCI test. Let us also assume $q^L \approx q$, being q the dominant covariance matrix eigenvector. With these positions, the detection probability P_D depends on the subspace change, as measured by the angular shift ϕ_0, and equals to:

$$P_D(\phi_0) = \prod_{i=1}^{\tau_{max}} \int_0^{\alpha^i \theta_{max}} p_{\Xi|\mathcal{H}_0}(\xi_i|\mathcal{H}_1) d\xi_i$$
$$= \prod_{i=1}^{\tau_{max}} \int_0^{\alpha^i \theta_{max}} \sqrt{1-\rho^2} \left[\pi \left(1 + \rho \sin(\pi/2 - 2\phi_0) - 2\rho \, \sin(\pi/2 - 2\phi_0)\xi^2 \right.\right.$$
$$\left.\left. -2\rho \, \cos(\pi/2 - 2\phi_0)\xi\sqrt{1-\xi^2}\right)\sqrt{1-\xi^2}\right]^{-1} d\xi_i. \tag{A.1}$$

Conversely, the probability of false alarm equals to

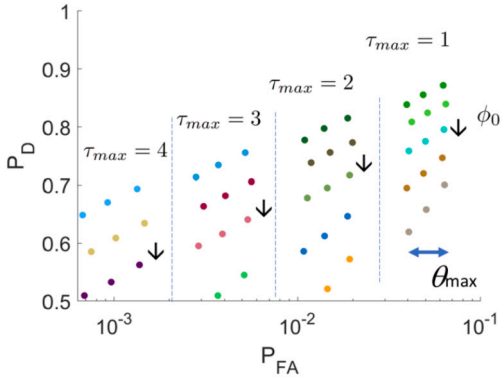


Fig. A.15. R-JSCA (coarse): probability of detection versus probability of false alarm for zero mean two-dimensional normal distribution with unitary marginal variances and correlation coefficient $\rho = 0.95$ ($\theta_{max} = 0.98 \pm 0.005$, $\tau_{max} = 1, 2, 3, 4$), for different values of ϕ_0, namely $\phi_0 \in \{0.25\pi, 0.225\pi, 0.2\pi, 0.175\pi, 0.15\pi\}$.

$$P_{FA} = \prod_{i=1}^{\tau_{max}} \int_0^{\alpha^i \theta_{max}} p_{\Xi|\mathcal{H}_0}(\xi_i|\mathcal{H}_0) d\xi_i$$
$$= \prod_{i=1}^{\tau_{max}} \int_0^{\alpha^i \theta_{max}} \frac{\sqrt{1-\rho^2}}{\pi\left(1 + \rho - 2\rho\,\xi_i^2\right)\sqrt{1-\xi_i^2}} d\xi_i. \tag{A.2}$$

In both the above equations we calculate the joint probability that a series of consecutive conformity samples fall below certain thresholds; the probability of this joint event is computed by a product, where the factors represent the probability that each sample ξ_i is under the corresponding threshold $\alpha^i \theta_{max}$. These factors vary in presence or in absence of a change (i.e. conditional to $\mathcal{H}_1$ or $\mathcal{H}_0$). The first case (i.e. all the samples under the thresholds in presence of a change) leads to a correct detection, while the second case (i.e. all the samples under the thresholds in absence of a change) results into a false alarm. Hence, the equations represent the probabilities of detection and false alarm in the coarse stage.

For the sake of concreteness, Fig. A.15 plots the probability of false alarm and the probability of detection computed for the case of zero mean two-dimensional normal distribution with unitary marginal variances and correlation coefficient $\rho = 0.95$, for different values of PC change ϕ_0. For increasing size τ_{max} of the recycling buffer, the false alarm rate drastically decreases, while the probability of detection remains consistent, also in presence of small fluctuations of θ_{max}. When the PC change ϕ_0 decreases from its maximum value $\pi/4$ the detection performance smoothly decreases.

A.2. R-JSCA fine stage: multi-scale conformity analysis over layer index

Let us outline the performance analysis of R-JSCA (fine) for the reference of two layers. R-JSCA (fine) detects a change when the difference between $\xi_n^{(l)}$ and $\xi_n^{(l+1)}$ exceeds a threshold $\Delta\xi_n^{(l)} + \Delta\xi_n^{(l+1)}$. i.e. R-JSCA fine detects change by analyzing the conformity $\xi_n^{(l)}$ over l. Herein, we consider two layers l and $l+1$, and the conformities

$$\xi_n^{(j)} = \frac{\mathbf{x}_n^\top \hat{\mathbf{q}}^{(j)}}{\|\mathbf{x}_n\|_2}, \; j = l, l+1, \tag{A.3}$$

where $\hat{\mathbf{q}}^{(j)}$ is the estimated PC at layer $j = l, l+1$. The false alarm and detection probability for the fine detection are written as:

$$P_D = \iint_{DR} p_{\Xi^{(l)},\Xi^{(l+1)}|\mathcal{H}_1}(\xi^{(l)}, \xi^{(l+1)}|\mathcal{H}_1) d\xi^{(l)} d\xi^{(l+1)} \tag{A.4}$$

and

$$P_{FA} = \iint_{\mathcal{DR}} p_{\Xi^{(l)}|\mathcal{H}_0}(\xi^{(l)}, \xi^{(l+1)}|\mathcal{H}_0) d\xi^{(l)} d\xi^{(l+1)}, \tag{A.5}$$

where $\mathcal{DR}$ is the decision region of the ICI criterion, defined as

$$\mathcal{DR} = \left\{ \xi^{(l)}, \xi^{(l+1)} \text{s.t.} \left(\xi^{(l)} - \Delta\xi^{(l)}\right) > \left(\xi^{(l+1)} - \Delta\xi^{(l+1)}\right) || \left(\xi^{(l)} + \Delta\xi^{(l)}\right) < \left(\xi^{(l+1)} - \Delta\xi^{(l+1)}\right) \right\}. \tag{A.6}$$

Let us assume $p(\xi^{(l)}, \xi^{(l+1)}|\mathcal{H}_i) \approx p(\xi^{(l)}|\mathcal{H}_1) \cdot p(\xi^{(l+1)}|\mathcal{H}_1)$, and let us denote by $\epsilon^{(j)}, j = l, l+1$ the phase of the estimated component $\mathbf{q}^{(j)}, j = l, l+1$. We obtain

$$p_{\Xi^{(j)}|\mathcal{H}_i}(\xi^{(j)}|\mathcal{H}_i) = \int p_{\Xi^{(j)}|E^{(j)},\mathcal{H}_i}(\xi^{(j)}|\epsilon^{(j)}, \mathcal{H}_i) \cdot p_{E^{(j)}|\mathcal{H}_i}(\epsilon^{(j)}|\mathcal{H}_i) d\epsilon, \quad i = 0, 1;\ j = l, l+1. \tag{A.7}$$

Approximating $p_{E^{(j)}|\mathcal{H}_i}(\epsilon^{(j)}|\mathcal{H}_i)$ with an uniform distribution in $[-\delta_l, \delta_l]$ under the hypothesis $\mathcal{H}_0$, and in $[m_e - \delta_l, m_e + \delta_l]$ under the hypothesis $\mathcal{H}_1$, we obtain

$$p_{\Xi^{(l)},\Xi^{(l+1)}|\mathcal{H}_0}(\xi^{(l)}, \xi^{(l+1)}|\mathcal{H}_0) = \left[\frac{1}{2\delta_l} \int_{-\delta_l}^{\delta_l} p_{\Xi^{(l)}|E^{(l)}|\mathcal{H}_0}(\xi^{(l)}|\epsilon^{(l)}|\mathcal{H}_0) d\epsilon^{(l)}\right] \cdot \left[\frac{1}{2\delta_{l+1}} \int_{-\delta_{l+1}}^{\delta_{l+1}} p_{\Xi^{(l+1)}|E^{(l+1)}|\mathcal{H}_0}(\xi^{(l+1)}|\epsilon^{(l+1)}|\mathcal{H}_0) d\epsilon^{(l+1)}\right] \tag{A.8}$$

and

$$p_{\Xi^{(l)},\Xi^{(l+1)}|\mathcal{H}_1}(\xi^{(l)}, \xi^{(l+1)}|\mathcal{H}_1) = \left[\frac{1}{2\delta_l} \int_{-\delta_l}^{\delta_l} p_{\Xi^{(l)}|E^{(l)},\mathcal{H}_1}(\xi^{(l)}|\epsilon^{(l)}, \mathcal{H}_1) d\epsilon^{(l)}\right] \cdot \left[\frac{1}{2\delta_{l+1}} \int_{m_e-\delta_{l+1}}^{m_e+\delta_{l+1}} p_{\Xi^{(l+1)}|E^{(l+1)},\mathcal{H}_1}(\xi^{(l+1)}|\epsilon^{(l+1)}, \mathcal{H}_1) d\epsilon^{(l+1)}\right], \tag{A.9}$$

where

$$p_{\Xi^{(j)}|E^{(j)},\mathcal{H}_0}(\xi^{(j)}|\epsilon^{(j)}, \mathcal{H}_0) = \frac{\sqrt{1-\rho^2}}{\pi\left(1 + \rho - 2\rho\,\xi_i^2\right)\sqrt{1-\xi_i^2}} \tag{A.10}$$

and

$$p_{\Xi^{(j)}|E^{(j)},\mathcal{H}_1}(\xi^{(j)}|\epsilon^{(j)}, \mathcal{H}_1) = \sqrt{1-\rho^2}\left[\pi\left(1 + \rho\sin(\pi/2 - 2 * \epsilon^{(j)}) \;\; -2\rho\,\sin(\pi/2 - 2 * \epsilon^{(j)})\xi^2 - 2\rho\,\cos(\pi/2 - 2 * \epsilon^{(j)})\xi\sqrt{1-\xi^2}\right)\sqrt{1-\xi^2}\right]^{-1}. \tag{A.11}$$

Fig. A.16 shows an example of the joint PDF[5] of two conformity measures $\xi_n^{(l)}$, $\xi_n^{(l+1)}$ under the hypothesis of absence and presence of a subspace change. The decision region where R-JSCA (fine) detects a change is highlighted. The probabilities of false alarm p_{FA} and of detection p_D are obtained by integration of the conditional joint PDFs under the hypothesis of absence and presence of a subspace change, respectively.

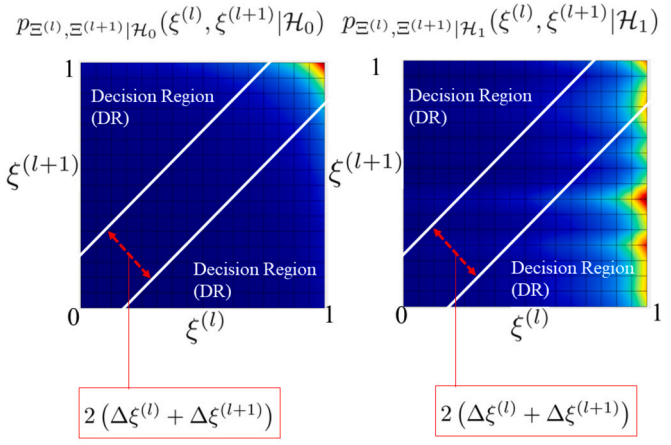


Fig. A.16. R-JSCA (fine): change detection decision region, for zero mean two-dimensional normal distribution with unitary marginal variances and correlation coefficient $\rho = 0.8$, $\phi_0 = \pi/4$; the thresholds $\Delta\xi^{(l)}, \Delta\xi^{(l+1)}$ determine the distance of the line boundaries $2\left(\Delta\xi^{(l)} + \Delta\xi^{(l+1)}\right)$.

Appendix B. Computational complexity

The computational complexity of R-JSCA depends i) on the computational complexity $C_{N,w}^{(HL_1PCA)}$ of hierarchical L1 PCA evaluation, ii) on the complexity $C_{N,w}^{(coarse)}$ of the coarse stage, and iii) on the complexity $C_{N,w}^{(fine)}$ of the fine stage. The coarse and fine stages require a number of elementary operations which is straightforwardly related to N, w, namely $C_{N,w}^{(coarse)} \approx N$ and $C_{N,w}^{(fine)} \approx N \cdot L \approx N \cdot \log(N)/\log(w)$, respectively.

The computational complexity of R-JSCA firstly depends on the computational complexity $C_{N,w}^{(HL_1PCA)}$ of H-L1-PCA. Let us now develop on H-L1-PCA, computed according to the architecture in Fig. 2. For simplicity sake, we describe here a recursive L layers computation architecture using a fixed size w for the L1-PCA analysis window; the number of measurements satisfies $N = w^L$. The architecture can be generalized to variable size windows by means of a flexible data parsing strategy.

With these positions, X is partitioned in $N_1 = N/w$ submatrices $S_1(n) \in \mathbb{R}^{D \times w}$, with $n \in [1 : N_1]$: $X = [S_1(1), \ldots, S_1(N_1)]$. Then the L1-PC is computed inside every submatrix. At the 2^{nd}-layer, the matrix of the estimated L1-PC components $Q_{HL,1}$ is partitioned into $N_2 = N_1/w$ submatrices $S_2(n) \in \mathbb{R}^{D \times w}$: $Q_{HL,1} = [S_2(1), \ldots, S_2(n), \ldots, S_2(N_2)]$ with $n \in [1 : N_2]$. Then, same as above, L1-PCA is applied to each submatrix.

After having defined $Q_{HL_1,0} = X$ and $N_0 = N$, the process can be generalized to repeat iteratively $\forall l \in [1 : L]$, $\forall n \in [1 : N_l]$:

$$N_l = \frac{N_{l-1}}{w} \tag{B.1}$$

$$Q_{HL_1,l-1} = [S_l(1), \ldots, \underbrace{S_l(n)}_{w \;\; columns}, \ldots, S_1(N_l)] \tag{B.2}$$

$$\mathbf{q}_{HL_1,l}(n) = \underset{\substack{\mathbf{q} \in \mathbb{R}^{D \times 1} \\ ||\mathbf{q}|| = 1}}{\arg\max} ||S_l(n) \cdot \mathbf{q}||_1 \tag{B.3}$$

$$Q_{HL_1,l} = [\mathbf{q}_{HL_1,l}(1) \;\; \cdots \;\; \mathbf{q}_{HL_1,l}(N_l)] \in \mathbb{R}^{D \times N_l}. \tag{B.4}$$

Finally, the computational cost of classical PCA is given by covariance matrix computation, which is $O\left(N^2 \cdot D\right)$ and its eigenvalue decomposition, that is $O\left(N^3\right)$. So, the complexity of PCA is $O\left(N^2 \cdot D + N^3\right)$. The computational cost of H-L1-PCA is given by the computational cost of L1-PCA on each and every layer so we have:

$$C_{N,w} = \frac{N}{w} \cdot w^3 + \cdots + \frac{N}{w^L} \cdot w^3 = w^3 \cdot \sum_{k=1}^{L} \frac{N}{w^k}.$$

We can rewrite it also as

$$C_{N,w} = w^3 \cdot N \cdot \sum_{k=1}^{L} \frac{1}{w^k} = N \cdot w^3 \cdot \left(\sum_{k=0}^{L} \frac{1}{w^k} - 1\right).$$

[5] Herein, we assume $\xi_n^{(l)}$ and $\xi_n^{(l+1)}$ to be statistically independent.

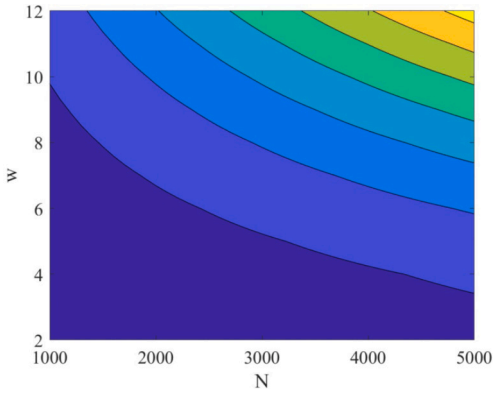


Fig. B.17. Computational complexity $C_{N,w}^{(R-JSCA)} = C_{N,w}^{(HL_1PCA)} + C_{N,w}^{(coarse)} + C_{N,w}^{(fine)}$ versus the number of samples N and the window length w.

Given that $\sum_{k=0}^{L} x^K = \frac{1-x^{L+1}}{1-x}$ we can write

$$C_{N,w} = N \cdot w^3 \cdot \left(\frac{1-x^{L+1}}{1-x} - 1 \right) \Big|_{x=\frac{1}{w}} = N \cdot w^3 \cdot \frac{w^L - 1}{w-1} \cdot \left(\frac{1}{w^L} \right)$$
$$\approx N \cdot w^3 \cdot w^{L-1} \cdot \left(\frac{1}{w^L} \right) = N \cdot w^2. \tag{B.5}$$

So the computational cost of H-L1-PCA is $C_{N,w}^{(HL_1PCA)} \approx N \cdot w^2$ versus the N^3 cost of conventional PCA.

The analysis above clarifies the role of the parameters N, L, w in the computation of H-L1-PCA and hence in R-JSCA. The number of layers L is a parameter depending on the overall number of samples N addressed by the analysis and by the window w adopted in the computation of the hierarchical PCs. The number of samples N is closely related to the problem under concern, e.g., whether there are big data involved, which is the expected stationarity interval, and so on. The window w determines the accuracy of the first layer estimates and the computational complexity of the multi-scale conformity computation. Coarsely speaking, a small value of w reduces the computational complexity of H-L1-PCA evaluation ($C_{N,w}^{(HL_1PCA)} \approx N \cdot w^2$), it does not affect the complexity of the coarse stage ($C_{N,w}^{(coarse)} \approx N$), and it slightly increases the complexity of the fine stage ($C_{N,w}^{(fine)} \approx N \cdot \log(N)/\log(w)$). The overall computational complexity is sketched in Fig. B.17; we recognize that, for any N, w should be kept as low as possible, compatibly with the accuracy of the individual L1 PCs computed on w data samples. In practice, in all our experiments we kept $w \approx 5$.

References

[1] V. Kapp, M.C. May, G. Lanza, T. Wuest, Pattern recognition in multivariate time series: towards an automated event detection method for smart manufacturing systems, J. Manuf. Mater. Process. 4 (3) (2020) 88.

[2] V. Raj, A. Renjini, M.S. Swapna, S. Sreejyothi, S. Sankararaman, Nonlinear time series and principal component analyses: potential diagnostic tools for COVID-19 auscultation, Chaos Solitons Fractals 140 (2020) 110246.

[3] I.G.A. Poornima, B. Paramasivan, Anomaly detection in wireless sensor network using machine learning algorithm, Comput. Commun. 151 (2020) 331–337.

[4] A. Miglani, N. Kumar, Deep learning models for traffic flow prediction in autonomous vehicles: a review, solutions, and challenges, Veh. Commun. 20 (2019) 100184.

[5] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, R. Otazo, On the applications of robust PCA in image and video processing, Proc. IEEE 106 (8) (2018) 1427–1457.

[6] L. Parsons, E. Haque, H. Liu, Subspace clustering for high dimensional data: a review, ACM SIGKDD Explor. Newsl. 6 (1) (2004) 90–105.

[7] R. Agrawal, J. Gehrke, D. Gunopulos, P. Raghavan, Automatic subspace clustering of high dimensional data for data mining applications, in: Proc. ACM SIGMOD, 1998, pp. 94–105.

[8] R. Vidal, Subspace clustering, IEEE Signal Process. Mag. 28 (2) (2011) 52–68.

[9] R. Vidal, Yi Ma, S. Sastry, Generalized principal component analysis (GPCA), IEEE Trans. Pattern Anal. Mach. Intell. 27 (12) (2015) 1945–1959.

[10] D. Hallac, P. Nystrup, S. Boyd, Greedy Gaussian segmentation of multivariate time series, IEEE Trans. Pattern Anal. Mach. Intell. 13 (Sep. 2019) 727–751.

[11] S. Matsushima, M. Brbic, Selective sampling-based scalable sparse subspace clustering, Advances in Neural Information Processing Systems, vol. 32.

[12] Y. Chen, L. Chun-Guang, Y. Chong, Stochastic sparse subspace clustering, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4155–4164.

[13] S. Tierney, J. Gao, Y. Guo, Subspace clustering for sequential data, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1019–1026.

[14] J. Duan, L. Guo, Variable-length subsequence clustering in time series, IEEE Trans. Knowl. Data Eng. 34 (2) (2020) 983–995.

[15] Y. Guo, J. Gao, F. Li, S. Tierney, M. Yin, Low rank sequential subspace clustering, in: 2015 International Joint Conference on Neural Networks (IJCNN), IEEE, July 2015, pp. 1–8.

[16] L. Wang, Z. Ding, Y. Fu, Learning transferable subspace for human motion segmentation, Proc. AAAI Conf. Artif. Intell. 32 (1) (April 2018).

[17] L. Wang, Z. Ding, Y. Fu, Low-rank transfer human motion segmentation, IEEE Trans. Image Process. 28 (2) (2018) 1023–1034.

[18] M. Dimiccoli, L. Garrido, G. Rodriguez-Corominas, H. Wendt, Graph constrained data representation learning for human motion segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1460–1469.

[19] G. Xia, H. Sun, L. Feng, G. Zhang, Y. Liu, Human motion segmentation via robust kernel sparse subspace clustering, IEEE Trans. Image Process. 27 (1) (2017) 135–150.

[20] T. De Ryck, M. De Vos, A. Bertrand, Change point detection in time series data using autoencoders with a time-invariant representation, IEEE Trans. Signal Process. 69 (2021) 3513–3524.

[21] S. Zhang, C. You, R. Vidal, C.G. Li, Learning a self-expressive network for subspace clustering, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12393–12403.

[22] Jian-Xun Mi, Ya-Nan Zhang, Zhihui Lai, Weisheng Li, Lifang Zhou, Fujin Zhong, Principal component analysis based on nuclear norm minimization, Neural Netw. 118 (2019).

[23] A.M. Rekavandi, A.K. Seghouane, K. Abed-Meraim, TRPAST: a tunable and robust projection approximation subspace tracking method, IEEE Trans. Signal Process. (2023).

[24] P. Ji, T. Zhang, H. Li, M. Salzmann, I. Reid, Deep subspace clustering networks, Adv. Neural Inf. Process. Syst. 30 (2017).

[25] A. Breloy, S. Kumar, Y. Sun, D.P. Palomar, Majorization-minimization on the Stiefel manifold with application to robust sparse PCA, IEEE Trans. Signal Process. 69 (2021) 1507–1520.

[26] V. Menon, S. Kalyani, Structured and unstructured outlier identification for robust PCA: a fast parameter free algorithm, IEEE Trans. Signal Process. 67 (9) (2019) 2439–2452.

[27] M. Rahmani, P. Li, Closed-form, provable, and robust PCA via leverage statistics and innovation search, IEEE Trans. Signal Process. (2021) 3132–3144.

[28] J. Liu, B.D. Rao, Robust PCA via ℓ_0-ℓ_1 regularization, IEEE Trans. Signal Process. 67 (2) (2018) 535–549.

[29] P. Markopoulos, G. Karystinos, D. Pados, Optimal algorithms for L1-subspace signal processing, IEEE Trans. Signal Process. 62 (19) (2014) 5046–5058.

[30] P. Markopoulos, M. Dhanaraj, A. Savakis, Adaptive L1-norm principal-component analysis with online outlier rejection, IEEE J. Sel. Top. Signal Process. 12 (6) (2018) 1131–1142.

[31] M. Dhanaraj, P.P. Markopoulos, Stochastic principal component analysis via mean absolute projection maximization, in: Proc. 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, Canada, Oct. 2019.

[32] G. Orrú, T. Cattai, S. Colonnese, G. Scarano, F. De Vico Fallani, P. Markopoulos, D. Pados, Deep L1-PCA of time-variant data with application to brain connectivity measurements, in: 27th European Signal Processing Conference (EUSIPCO), A Coruña, 2019.

[33] I. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. R. Soc. A, Math. Phys. Eng. Sci. 374 (2065) (2016) 20150202.

[34] Gilad Lerman, Tyler Maunu, An overview of robust subspace recovery, Proc. IEEE 106 (8) (2018) 1380–1410.

[35] N. Vaswani, et al., Robust subspace learning: robust PCA, robust subspace tracking and robust subspace recovery, IEEE Signal Process. Mag. 35 (4) (July 2018) 32–55.

[36] N. Kwak, Principal component analysis based on l1-norm maximization, IEEE Trans. Pattern Anal. Mach. Intell. 30 (9) (2008) 1672–1680.

[37] N. Kwak, Principal component analysis by lp-norm maximization, IEEE Trans. Cybern. 44 (5) (2014) 594–609.

[38] J. Brooks, J. Dula, Estimating L1-norm best-fit lines for data, in: Optimization Online, August 2019.

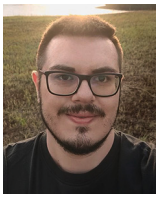
[39] H. Xu, C. Caramanis, S. Sanghavi, Robust PCA via outlier pursuit, Adv. Neural Inf. Process. Syst. 23 (2010).

[40] M. Rahmani, G.K. Atia, Coherence pursuit: fast, simple, and robust principal component analysis, IEEE Trans. Signal Process. 65 (23) (2017) 6260–6275.

[41] L.T. Thanh, N.V. Dung, N.L. Trung, K. Abed-Meraim, A brief survey on robust subspace tracking algorithms in signal processing, J. Electron. Commun. 11 (1) (January 2021) 16–25.
[42] P. Narayanamurthy, N. Vaswani, Fast robust subspace tracking via PCA in sparse data-dependent noise, IEEE J. Sel. Areas Inf. Theory 1 (3) (2020) 723–744.
[43] D.G. Chachlakis, M. Dhanaraj, A. Prater-Bennette, P. Markopoulos, Dynamic L1-norm Tucker tensor decomposition, IEEE J. Sel. Top. Signal Process. (2021), https://doi.org/10.1109/JSTSP.2021.3058846.
[44] F. Nie, H. Huang, C. Ding, D. Luo, H. Wang, Robust principal component analysis with non-greedy L1-norm maximization, in: Proc. Int. Joint Conf. Artif. Intell., Barcelona, Spain, Jul. 2011, pp. 1433–1438.
[45] S. Colonnese, P.P. Markopoulos, G. Scarano, D.A. Pados, FFT calculation of the L1-norm principal component of a data matrix, Signal Process. 189 (108286) (August 2021), Elsevier.
[46] P.P. Markopoulos, S. Kundu, S. Chamadia, D.A. Pados, Efficient L1-norm principal-component analysis via bit flipping, IEEE Trans. Signal Process. 65 (16) (2017) 4252–4264.
[47] H. Kamrani, A. Zoghadr Asli, P.P. Markopoulos, M. Langberg, D.A. Pados, G.N. Karystinos, Reduced-rank L1-norm principal-component analysis with performance guarantees, IEEE Trans. Signal Process. 69 (November 2020) 240–255.
[48] D.G. Chachlakis, A. Prater-Bennette, P. Markopoulos, L1-norm Tucker tensor decomposition, IEEE Access 7 (2019) 178454–178465.
[49] M. Mozaffari, P.P. Markopoulos, A. Prater-Bennette, Improved L1-Tucker via L1-fitting, in: Proc. European Signal Processing Conference (EUSIPCO 2021), Dublin, Ireland, August 2021.
[50] K. Tountas, D.G. Chachlakis, P. Markopoulos, D.A. Pados, Iteratively re-weighted L1-PCA of tensor data, in: 2019 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2019, pp. 1658–1661.
[51] L.T. Thanh, N.V. Dung, N.L. Trung, K. Abed-Meraim, Robust subspace tracking with missing data and outliers: novel algorithm with convergence guarantee, IEEE Trans. Signal Process. 69 (2021) 2070–2085, https://doi.org/10.1109/TSP.2021.3066795.
[52] D.G. Chachlakis, M. Dhanaraj, A. Prater-Bennette, P.P. Markopoulos, Dynamic L1-norm Tucker tensor decomposition, IEEE J. Sel. Top. Signal Process. (Feb. 2021).
[53] M. Dhanaraj, P.P. Markopoulos, On the asymptotic L1-PC of elliptical distributions, IEEE Signal Process. Lett. 29 (2022) 2343–2347, https://doi.org/10.1109/LSP.2022.3205274.
[54] P. Dharmawansa, N. Rajatheva, C. Tellambura, Envelope and phase distribution of two correlated Gaussian variables, IEEE Trans. Commun. 57 (4) (2009) 915–921.
[55] N. Vaswani, P. Narayanamurthy, Static and dynamic robust PCA and matrix completion: a review, Proc. IEEE 106 (8) (Aug. 2018) 1359–1379.
[56] H.V. Poor, An Introduction to Signal Detection and Estimation, 2nd ed., Springer-Verlag, Berlin, Heidelberg, 1994.
[57] V. Katkovnik, K. Egiazarian, J. Astola, A spatially adaptive nonparametric regression image deblurring, IEEE Trans. Image Process. (2005) 1469–1478.
[58] V. Katkovnik, K. Egiazarian, J. Astola, Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule, J. Math. Imaging Vis. 16 (05) (2002) 223–235.
[59] A. Foi, V. Katkovnik, K. Egiazarian, Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images, IEEE Trans. Image Process. 16 (5) (May 2007) 1395–1411, https://doi.org/10.1109/TIP.2007.891788.
[60] A. Foi, Anisotropic nonparametric image processing: theory, algorithms and applications, Ph.D. Thesis, Politecnico di Milano, April 2005, available at https://webpages.tuni.fi/foi/papers/Foi-Anisotropic_non_parametric_image_processing-2005.pdf.
[61] C. Truong, L. Oudre, N. Vayatis, Selective review of offline change point detection methods, Signal Process. 167 (2020).
[62] J.L. Reyes-Ortiz, L. Oneto, A. Samá, X. Parra, D. Anguita, Transition-aware human activity recognition using smartphones, Neurocomputing 171 (2016) 754–767.
[63] J.L. Reyes-Ortiz, L. Oneto, A. Samá, X. Parra, D. Anguita, UC Irvine machine learning repository, UC Irvine, available at http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions, January 2016.
[64] M. Gygli, H. Grabner, H. Riemenschneider, L. Van Gool, Creating summaries from user videos, in: European Conference on Computer Vision, 2014, pp. 505–520.

Stefania Colonnese is Associate Professor at the Department of Information Engineering, Electronics and Telecommunications (DIET) of Sapienza University of Rome, Italy. She received her Ph.D. in Electronics Engineering from the University of Roma Tre. She has participated in the MPEG-4 standardization activity within the ISO MPEG-4 Core Experiment on Automatic Video Segmentation. She is co-author of more than a hundred journal and conference papers, two book chapters and several ISO MPEG-4 Contributing Documents. Her research interests range from statistical signal processing, image deconvolution and restoration, biomedical signal processing, to video encoding, processing and networking. She served in several conferences as General, Technical Program, Publicity Chair, and Technical Program Committee member. She is an IEEE Senior Member and IEEE Women in Engineering Member.

Gaetano Scarano was born in Campobasso, Italy. He received the "Laurea" degree in Electronic Engineering (summa cum laude) from Università di Roma "La Sapienza," Rome, Italy, in 1982. In 1982, he joined the Istituto di Acustica, Consiglio Nazionale delle Ricerche, Roma, Italy, as Associate Researcher. Since 1988, he has been teaching digital signal processing at the University of Perugia, Perugia, Italy, where in 1991 he became Associate Professor of Signal Theory. In 1992, he joined the Dipartimento di Scienza e Tecnica dell'Informazione e della Comunicazione, now Dipartimento di Ingegneria dell'Informazione, Elettronica e Telecomunicazioni, Università di Roma "La Sapienza," first as Associate Professor of Image Processing, then as Professor of Signal Theory. His research interests lie in the area of signal and image processing, communications, estimation and detection theory, and include channel equalization and estimation, image restoration, and texture synthesis and classification. Prof. Scarano served as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS.



Marcello Marra was born in Taranto, Italy. He received from Sapienza University of Rome, Italy the B.Sc. degree in Clinical Engineering in 2017 and the M.Sc. degree in Biomedical Engineering in 2020. His research interests include principal component analysis, sensor data processing, biomedical image processing.

Dr. Panagiotis (Panos P.) Markopoulos, Ph.D., is an Associate Professor and Margie and Bill Klesse Endowed Professor with the Departments of Electrical and Computer Engineering and Computer Science at The University of Texas at San Antonio (UTSA). He is also a core faculty member of the UTSA School of Data Science and MATRIX: The UTSA AI Consortium for Human Well-Being. Prior to joining UTSA, Dr. Markopoulos was a tenured Associate Professor with the Rochester Institute of Technology (RIT). In the Summers of 2018, 2020, and 2021, he was a Visiting Research Faculty at the U.S. Air Force Research Laboratory (AFRL), Information Directorate, in Rome NY.

His expertise is in the areas of machine learning, data analysis, and adaptive signal processing. His research mission is to advance efficient, explainable, and trustworthy artificial intelligence. Together with students and collaborators, Dr. Markopoulos has co-authored more than 70 journal and conference articles and 3 book chapters. Since 2016, his research has received external funding awards in the order of $2M (both as PI and Co-PI) from sponsors including the US National Science Foundation (NSF), the US National Geo-Spatial Intelligence Agency, the US Air Force Office of Scientific Research (AFOSR), and the Air Force Research Laboratory (AFRL).

In October 2019, Dr. Markopoulos received the Young Investigator Program (YIP) Award, from the AFOSR. In 2021, Dr. Markopoulos was elevated to the grade of IEEE Senior Member.

Dimitris A. Pados (Senior Member, IEEE) received the Diploma degree in computer science and engineering (five-year program) from the University of Patras, Greece, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA. From 1997 to 2017, he was with the Department of Electrical Engineering, The State University of New York at Buffalo, as an Assistant Professor, an Associate Professor, a Professor, and a Clifford C. Furnas Chair Professor of electrical engineering. He also worked as the Associate Chair and was appointed as the Chair of the Department of Electrical Engineering. He was a University Faculty Senator (elected) four times and worked at the Faculty Senate Executive Committee for two terms. In 2017, he joined Florida Atlantic University, Boca Raton, FL, USA, as the Schmidt Eminent Scholar Professor of engineering and computer science and a fellow of the Institute for Sensing and Embedded Network Systems Engineering (I-SENSE). He is the Founding Director of the FAU Center for Connected Autonomy and Artificial Intelligence (https://ca-ai.fau.edu/). He has served as a Principal Investigator on federal grants (NSF and DoD) of about $17M and has been author/coauthor of 230 journal and conference proceedings articles in predominantly IEEE venues. Notable technical contributions from his team include small-sample-support adaptive filtering (auxiliary-vector filters), optimal total-squared-correlation multiple-access code sets (Karystinos-Pados bounds and designs), optimal spread-spectrum data hiding, L1-norm principal-component analysis (optimal algorithms for exact L1-norm PCA), and robust localization in extreme environments (L1-norm feature extraction from complex-valued data). Dr. Pados is a member of the IEEE Communications, IEEE Signal Processing, IEEE Information Theory, and IEEE Computational Intelligence Societies. Articles that he has coauthored with his students received the 2001 IEEE International Conference on Telecommunications Best Paper Award, the 2003 IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS Outstanding Paper Award, the 2010 IEEE International Communications Conference (ICC) Best Paper Award in signal processing for communications, the 2013 International Symposium on Wireless Communication Systems Best Paper Award in physical layer communications and signal processing, the Best of IEEE GLOBECOM 2014-Top 50 Papers Distinction, the Best Paper in the 2016 IEEE International Conference on Multimedia Big Data, and the Paper Distinctions at the International Workshop on Antenna Technology (iWAT)

2019 and IEEE/MTS Oceans 2020. He was a recipient of the 2009 SUNY-Wide Chancellor's Award for Excellence in Teaching and the 2011 University at Buffalo Exceptional Scholar-Sustained Achievement Award. He was presented with the 2021 Florida Atlantic Research and Development Authority Distinguished Researcher Award, Boca Raton, FL, USA. He has served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.