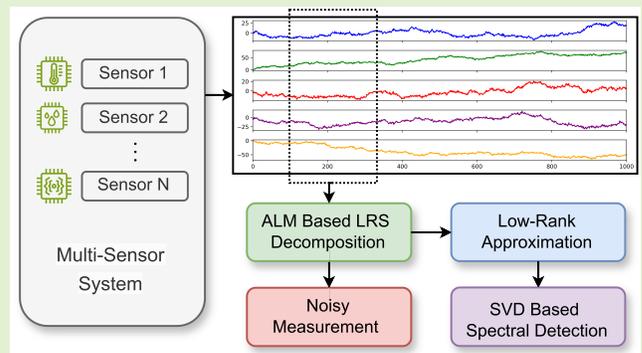


Multivariate Time Series Anomaly Detection via Low-Rank and Sparse Decomposition

Mohammed Ayalew Belay¹, Graduate Student Member, IEEE, Adil Rasheed¹,
and Pierluigi Salvo Rossi², Senior Member, IEEE

Abstract—In the era of data-driven decision-making, multisensor systems acquire complex, high-dimensional streams capturing temporal dynamics, and multivariate time series anomaly detection has become significantly relevant in several application domains. Conventional methods relying on supervised and semi-supervised learning require labeled data, which might not be available in various scenarios. Conversely, noise and outliers present in real-world sensor measurements negatively impact unsupervised methods. Furthermore, several methods rely on black-box architectures, which limit their use in safety-critical applications where interpretability and explainability are often necessary. To address these challenges, we propose a novel unsupervised multivariate time series anomaly detection method that exploits low-rank and sparse (LRS) decomposition combined with spectral detection. More specifically, we use augmented Lagrange multiplier (ALM)-based optimization with eigenvalue soft thresholding for decomposition. Data points are projected onto a low-dimensional subspace, capturing the underlying data structure and enabling robust anomaly detection in noisy multisensor environments. Finally, the effectiveness of the proposed approach is presented via performance comparison to several existing methods using publicly available datasets collecting real-world sensor measurements from testbeds of water treatment systems.

Index Terms—Anomaly detection, low-rank approximation, multisensor systems, multivariate time series, sparse decomposition, unsupervised learning.



I. INTRODUCTION

ANOMALY detection plays a pivotal role in various sectors, driven by the imperative to identify deviations from normal behavior. In industrial contexts, such as power plants and manufacturing facilities, early detection of anomalies can prevent equipment failures, enhance the safety of operations, and reduce downtime and maintenance costs [1], [2]. System

Received 12 August 2024; accepted 24 August 2024. Date of publication 13 September 2024; date of current version 31 October 2024. This work was supported in part by the Research Council of Norway through the Project DIGITAL TWIN within the PETROMAKS2 Framework under Project 318899. The associate editor coordinating the review of this article and approving it for publication was Prof. Qing Li. (Corresponding author: Mohammed Ayalew Belay.)

Mohammed Ayalew Belay is with the Department of Electronic Systems, Norwegian University of Science and Technology, 7034 Trondheim, Norway (e-mail: mohammed.a.belay@ntnu.no).

Adil Rasheed is with the Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7034 Trondheim, Norway (e-mail: adil.rasheed@ntnu.no).

Pierluigi Salvo Rossi is with the Department of Electronic Systems, Norwegian University of Science and Technology, 7034 Trondheim, Norway, and also with the Department of Gas Technology, SINTEF Energy Research, 7491 Trondheim, Norway (e-mail: salvorossi@iee.org).

Digital Object Identifier 10.1109/JSEN.2024.3452318

safety and reliability are especially crucial nowadays in the mobility sector, which is experiencing the growth of technologies for autonomous systems [3]. In cybersecurity, it is essential to recognize threats such as network intrusions, thus securing data integrity and mitigating risks [4], [5]. In healthcare, anomaly detection supports early diagnosis by identifying unusual patient conditions, leading to timely interventions with increased positive outcomes [6]. More generally, the continued expansion and complexity of data, especially those gathered in multisensor systems, continue to drive the need for advanced anomaly detection techniques, highlighting their significance in modern technological applications [2].

Recently, the adoption of deep neural networks (DNNs) in unsupervised anomaly detection (USAD) has seen remarkable success [7], [8], with a variety of architectures being explored, such as recurrent networks [9], [10], [11], [12], convolutional networks [13], [14], autoencoders (AEs) [14], [15], generative adversarial networks [16], [17], transformers [18], [19], and graph neural networks [20]. These methods have demonstrated significant advances in the management of complex multivariate time series data, offering substantial performance improvements over traditional techniques. However, the

success of these models comes with challenges: their black-box nature poses significant hurdles in critical applications where interpretability is crucial, as stakeholders must understand and trust the models' decisions; the substantial computational complexity required for training and inference poses practical limitations, making these models less viable for real-time or on-device applications. Moreover, deep learning models often struggle with generalization when faced with unseen data, which can severely impact their reliability in dynamic and unpredictable environments. This scenario underscores the ongoing need for simpler and more interpretable methods that can be easily implemented in operational settings, balancing performance with transparency and computational efficiency.

One such approach involves principal component analysis (PCA), which operates under the assumption that high-dimensional data can be embedded in a lower dimensional subspace where normal instances and anomalies are distinctly separable [21], [22], [23], [24]. PCA is appealing because of its robust mathematical foundation and inherent simplicity, which lead to computational efficiency and facilitate its confident use in high-stakes applications. However, the simplicity of PCA also introduces vulnerabilities, particularly in environments prone to data corruption, such as multisensor systems where noise and outliers are common [25]. Improving the robustness of PCA-based anomaly detection methods against data corruption and the ability to handle noise and outliers without compromising simplicity and interpretability is essential for ensuring reliable and accurate performance in operational settings, especially in critical applications.

To tackle the limitations inherent in traditional PCA-based anomaly detection, particularly its vulnerability to data corruption, we developed a novel approach grounded in low-rank and sparse (LRS) decomposition. This method is based on the assumption that high-dimensional data can typically be represented by low-dimensional latent variables and that anomalies tend to be sparse [26], [27]. Specifically, our contribution can be summarized as follows.

- 1) We introduce a novel USAD method for multivariate time series, using robust principal analysis to enhance detection accuracy in monitoring of multisensor systems.
- 2) Our approach leverages LRS decomposition techniques, using augmented Lagrange multiplier (ALM)-based optimization and eigenvalue soft thresholding, to effectively minimize the impact of noisy sensor measurements, potentially present during the training process, on the anomaly detection process.
- 3) We present several results for performance evaluation, comparing our method against other related anomaly detection techniques using publicly available multisensor datasets from testbeds of water treatment systems to demonstrate its efficacy and robustness.

This article is organized into five sections. Section II formulates the specific problem statement and describes the mathematical approach developed for robust anomaly detection. Section III describes the multisensor datasets used in this study, the evaluation metrics used for comparison, the

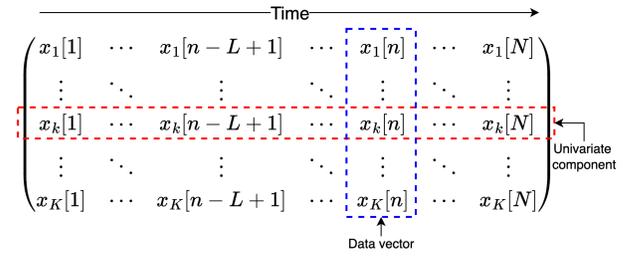


Fig. 1. Data structure of the multisensor systems.

baseline methods, and provides some relevant implementation details; the results are presented and discussed in Section IV. Finally, Section V summarizes the work and suggests potential directions for future research.

Notation—Vectors and matrices are denoted with bold lower-case and bold upper-case letters, respectively; rank and transpose operators are represented as $\text{rank}(\cdot)$ and $(\cdot)^T$, respectively; while $\|\cdot\|_j$, $\|\cdot\|_*$, $\|\cdot\|_F$, and $\langle \cdot, \cdot \rangle$ denote the ℓ_j -norm, the nuclear norm, the Frobenius norm, and the Frobenius inner product, respectively; and $\lceil \cdot \rceil$ is the ceiling operator.

II. METHODOLOGY

A. Problem Statement

We consider a multivariate time series with K components generated by a multisensor system where $x_k[n] \in \mathbb{R}$ denotes the value (e.g., the sensor measurement) of the k th univariate component at discrete time n . The *data vector* $\mathbf{x}[n] = (x_1[n], x_2[n], \dots, x_K[n])^T \in \mathbb{R}^K$ collects all the components at discrete time n , and *data matrix* $\mathbf{X} = (\mathbf{x}[1], \mathbf{x}[2], \dots, \mathbf{x}[N]) \in \mathbb{R}^{K \times N}$ collects the data vectors related to N consecutive discrete times. The overall data structure is shown in Fig. 1.

The objective of the proposed framework is to construct, from a training data matrix ($\mathbf{X}_{\text{train}} \in \mathbb{R}^{K \times N}$), a model $\mathcal{G}(\cdot)$ that characterizes the behavior of the multisensor system under normal conditions and is capable of detecting deviations from normal behavior. More specifically, we assume that no guarantee is provided in relation to the absence of anomalies in the training data, so the challenge is to characterize the normal behavior of the multisensor system from unlabeled training data which is potentially corrupted (e.g., in the presence of noise and outliers). For model evaluation, we consider a test data matrix ($\mathbf{X}_{\text{test}} \in \mathbb{R}^{K \times M}$ with $M \ll N$) which includes measurements from both normal and anomalous conditions. In addition, a *label vector* $\mathbf{y} = (y_1, y_2, \dots, y_M)^T \in \{0, 1\}^M$ paired with the test data matrix represents ground-truth information, with $y_m = 1$ (resp. $y_m = 0$) denoting the presence (resp. absence) of an anomaly at discrete time m .

Measurements (during both training and testing) are pre-processed according to Z -score normalization, with mean and standard deviation (for each univariate component) obtained from the training data matrix. In the following, the data matrices refer to the normalized versions.

B. LRS Decomposition

We propose a spectral unsupervised multivariate anomaly detection method based on LRS decomposition. This approach

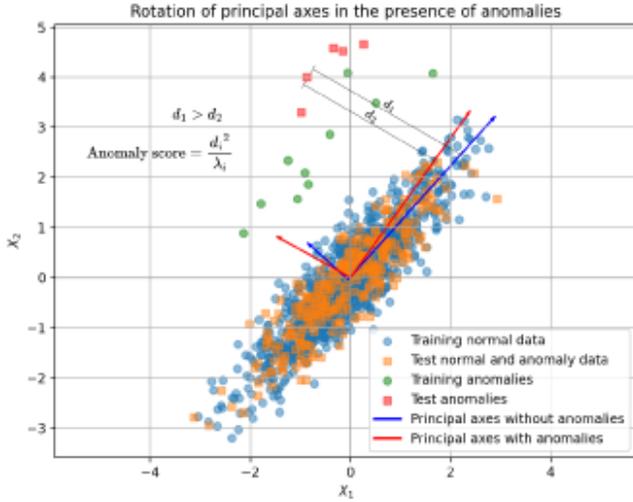


Fig. 2. LRS-based anomaly detection approach: rotation of the principal axis in the presence of anomalies and anomaly scoring using the principal axis of the low-rank training matrix.

is based on the idea that the noisy training data generated by measurements in multisensor systems can be separated into two components: 1) a structured, low-rank component representing normal behavior, and 2) a sparse component capturing anomalies and/or noise. More specifically, a data matrix $\mathbf{X} \in \mathbb{R}^{K \times N}$ can be expressed via LRS decomposition as

$$\mathbf{X} = \mathbf{L} + \mathbf{S} \quad (1)$$

where $\mathbf{L} \in \mathbb{R}^{K \times N}$ is a low-rank matrix representing the normal sensor readings and $\mathbf{S} \in \mathbb{R}^{K \times N}$ is a sparse matrix representing sensor anomalies and/or noise.

PCA can be applied to the data matrix to find the principal directions in which the data vary the most. However, the basic assumption of our framework is that in the presence of anomalies, applying PCA to the low-rank matrix (\mathbf{L}) is more robust than applying it directly to the data matrix (\mathbf{X}). Fig. 2 provides a visual representation of this concept, where the rotation of the principal axes in the presence of anomalies is depicted.

C. LRS Nonconvex Objective and Convex Relaxation

The nonconvex formulation of the LRS seeks to directly minimize the rank of the low-rank matrix (\mathbf{L}) and maximize the sparsity (i.e., minimize the number of nonzero elements) of the sparse matrix (\mathbf{S}). Hence, the optimization problem can be expressed as [28], [29], [30]

$$\min_{\mathbf{L}, \mathbf{S}} \text{rank}(\mathbf{L}) + \|\mathbf{S}\|_0, \quad \text{s.t. } \mathbf{L} + \mathbf{S} = \mathbf{X} \quad (2)$$

where minimizing the rank encourages the low-rank matrix (\mathbf{L}) to have fewer nonzero singular values and minimizing the ℓ_0 -norm encourages the sparse matrix (\mathbf{S}) to have fewer nonzero entries. Both the rank and the number of nonzero entries are nonconvex functions, so both the terms lead to a nonconvex optimization problem, which is NP-hard [31].

A surrogate convex relaxation approach can be used to make the optimization problem tractable [32]. A convex relaxation

is used where the rank function is replaced by the nuclear norm and the ℓ_0 -norm is replaced by the ℓ_1 -norm, leading to the following convex optimization problem [33]:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1, \quad \text{s.t. } \mathbf{L} + \mathbf{S} = \mathbf{X} \quad (3)$$

where $\lambda > 0$ is a regularization hyperparameter trading off between the LRS components. The nuclear norm minimization indirectly impacts the rank by promoting smaller singular values to be zero, thus achieving a low-rank matrix without directly minimizing the nonconvex rank function. The ℓ_1 -norm minimization encourages sparsity by penalizing nonzero entries, similar to the ℓ_0 -norm, but in a convex manner [34]. The optimal value for λ is influenced by the magnitude of outliers in the data and can be selected via cross-validation. These relaxations lead to the problem formulation in (3) that is convex and can be efficiently solved using optimization techniques such as ALM. Alternative techniques can be based on Moreau envelope alternating direction method of multipliers (ADMM) [35]. The advantage of using ADMM would be in terms of faster convergence and enhanced stability; however, it exhibits higher computational complexity and parameter sensitivity. The ALM approach transforms the constrained LRS optimization problem into an unconstrained using Lagrange multipliers alongside an augmented penalty term [36]. The augmented Lagrangian for the LRS optimization problem is formulated as

$$\begin{aligned} \mathcal{L}(\mathbf{L}, \mathbf{S}, \mathbf{Y}, \rho) = & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 + \langle \mathbf{Y}, \mathbf{X} - \mathbf{L} - \mathbf{S} \rangle \\ & + \frac{\rho}{2} \|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F^2 \end{aligned} \quad (4)$$

where $\mathbf{Y} \in \mathbb{R}^{K \times N}$ is the matrix of Lagrange multipliers, and $\rho > 0$ is a penalty parameter.

The ALM algorithm updates the matrices \mathbf{L} , \mathbf{S} , and \mathbf{Y} iteratively to minimize the augmented Lagrangian $\mathcal{L}(\mathbf{L}, \mathbf{S}, \mathbf{Y}, \rho)$. More specifically, the low-rank matrix (\mathbf{L}) and the sparse matrix (\mathbf{S}) are alternatively updated, and the algorithm uses the soft-thresholding method for the singular values of the low-rank matrix (\mathbf{L}) and the entries of the sparse matrix (\mathbf{S}), enforcing the constraint based on the data matrix ($\mathbf{X} = \mathbf{L} + \mathbf{S}$) while adjusting the penalty parameter (ρ) incrementally [37]. The soft-thresholding method is based on the shrinking operator defined as

$$S(a; \alpha) = \begin{cases} a - \alpha, & a > \alpha \\ a + \alpha, & a < -\alpha \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The update of the low-rank matrix (\mathbf{L}) is performed assuming the sparse matrix (\mathbf{S}) and the matrix of Lagrange multipliers (\mathbf{Y}) fixed in the following minimization problem:

$$\min_{\mathbf{L}} \left(\|\mathbf{L}\|_* + \frac{\rho}{2} \left\| \mathbf{X} - \mathbf{L} - \mathbf{S} + \frac{1}{\rho} \mathbf{Y} \right\|_F^2 \right) \quad (6)$$

where soft-thresholding is applied to the singular values of the following matrix $\mathbf{P} = \mathbf{X} - \mathbf{S} + \rho^{-1} \mathbf{Y}$. Denoting $\mathbf{P} = \mathbf{U}_P \mathbf{\Sigma}_P \mathbf{V}_P^T$ the corresponding singular value decomposition (SVD), where \mathbf{U}_P and \mathbf{V}_P are the orthogonal matrices and $\mathbf{\Sigma}_P$ is a diagonal matrix with the singular values $(\sigma_{P,1}, \dots, \sigma_{P,\min(K,N)})$ of \mathbf{P} on

Algorithm 1 LRS via ALM Optimization

Input: Training matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, λ
Output: Approximate \mathbf{L} , \mathbf{S}

```

1:  $\mathbf{Y}_0 = 0$ ,  $\rho > 1$ ,  $z = 0$ 
2: while not converged do
3:    $(\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}) \leftarrow \text{svd}(\mathbf{X} - \mathbf{S}_z + \rho_z^{-1} \mathbf{Y}_z)$ 
4:    $\mathbf{L}_{z+1} \leftarrow \mathbf{U} \mathcal{S}(\mathbf{\Sigma}; \rho_z^{-1}) \mathbf{V}^T$ 
5:    $\mathbf{S}_{z+1} \leftarrow \mathcal{S}(\mathbf{X} - \mathbf{L}_{z+1} + \rho_z^{-1} \mathbf{Y}_z; \lambda \rho_z^{-1})$ 
6:    $\mathbf{Y}_{z+1} \leftarrow \mathbf{Y}_z + \rho_z (\mathbf{X} - \mathbf{L}_{z+1} - \mathbf{S}_{z+1})$ 
7:    $\rho_{z+1} \leftarrow \rho_z$ 
8:    $z \leftarrow z + 1$ 
9: end while
    
```

the main diagonal. The soft-thresholding method applies the shrinking operator to the singular values ($\sigma_{p,i}$)

$$\tilde{\sigma}_{p,i} = \mathcal{S}(\sigma_{p,i}; \rho^{-1}), \quad i = 1, \dots, \min(K, N) \quad (7)$$

and updates the low-rank matrix as $\mathbf{L} = \mathbf{U}_P \tilde{\mathbf{\Sigma}}_P \mathbf{V}_P^T$, where $\tilde{\mathbf{\Sigma}}_P$ is the diagonal matrix with the updated singular values ($\tilde{\sigma}_{p,i}$) on the main diagonal.

The update of the sparse matrix (\mathbf{S}) is performed assuming the low-rank matrix (\mathbf{L}) and the matrix of Lagrange multipliers (\mathbf{Y}) fixed in the following minimization problem:

$$\min_{\mathbf{S}} \left(\lambda \|\mathbf{S}\|_1 + \frac{\rho}{2} \left\| \mathbf{X} - \mathbf{L} - \mathbf{S} + \frac{1}{\rho} \mathbf{Y} \right\|_F^2 \right) \quad (8)$$

where soft-thresholding via the shrinking operator is directly applied to the entries ($R_{i,j}$) of the following matrix $\mathbf{R} = \mathbf{X} - \mathbf{L} + \rho^{-1} \mathbf{Y}$, i.e.,

$$\tilde{R}_{i,j} = \mathcal{S}(R_{i,j}; \lambda \rho^{-1}), \quad i = 1, \dots, K, \quad j = 1, \dots, N. \quad (9)$$

Finally, the matrix of Lagrange multipliers (\mathbf{Y}) is updated to enforce the constraint $\mathbf{X} = \mathbf{L} + \mathbf{S}$ as iterations proceed, via the following updating rule:

$$\mathbf{Y}_{z+1} = \mathbf{Y}_z + \rho_z (\mathbf{X} - \mathbf{L}_{z+1} - \mathbf{S}_{z+1}) \quad (10)$$

where the indices z and $z + 1$ refer to the current and next iterations, respectively.

The ALM-based LRS decomposition is summarized in the Algorithm 1.

D. LRS-Based Anomaly Detection

In the proposed spectral anomaly detection approach, a normal instance is expected to have low projection scores along the principal components of the low-rank matrix, as they align with the direction of maximum variance defined by the normal data. Anomalies, on the other hand, will have higher projection scores due to their deviation from the normal structure (e.g., d_1 and d_2 in Fig. 2). We use SVD to perform PCA on the low-rank training matrix as $\mathbf{L} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ where $\mathbf{U} \in \mathbb{R}^{K \times K}$ and $\mathbf{V} \in \mathbb{R}^{N \times N}$ are the orthogonal matrices, and $\mathbf{\Sigma} \in \mathbb{R}^{K \times N}$ is a nonnegative diagonal matrix containing the singular values arranged in descending order (i.e., $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min(K,N)}$).

For anomaly detection purpose, we transform each test data vector ($\mathbf{x}[m]$) from the test data matrix (\mathbf{X}_{test}) using the

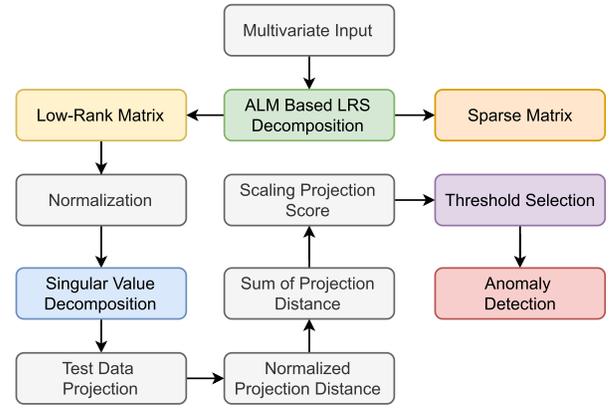


Fig. 3. LRS-based anomaly detection process.

TABLE I
SUMMARY OF DATASETS

Attributes	Datasets	
	SWaT	WADI
Entities	1	1
No. of channels	51	123
Average Train size	495,000	1,209,601
Average Test size	449,919	172,801
Anomaly rate	12.140%	5.99%

eigenvectors from the SVD of the low-rank training data as

$$\hat{\mathbf{x}}[m] = \mathbf{U} \mathbf{x}[m], \quad m = 1, \dots, M \quad (11)$$

where $\hat{\mathbf{x}}[n]$ is the projected test data. The anomaly score is calculated using the distance of the projected test data points along the principal axes of the low-rank training data. For each projected test vector ($\hat{\mathbf{x}}[m]$), the projection distance is given by

$$d_m = \sum_{i=1}^q \frac{\hat{x}_i^2[m]}{\sigma_i^2}, \quad m = 1, \dots, M \quad (12)$$

where q is the number of principal components used. The projection distance is then scaled using min–max scaling to determine the corresponding anomaly score (s_m) as follows:

$$s_m = \frac{d_m - d_{\min}}{d_{\max} - d_{\min}} \quad (13)$$

where d_{\min} and d_{\max} are computed from the score distribution. Finally, anomaly labeling (\hat{y}_m) is performed based on a threshold-based rule applied to the anomaly score (s_m), i.e.,

$$\hat{y}_m = \begin{cases} 1, & s_m > \tau \\ 0, & s_m \leq \tau \end{cases} \quad (14)$$

where the threshold (τ) is selected from the training score distribution according to the three-sigma rule.

A visualization of the steps involved is presented in Fig. 3. The overall steps for the proposed method are summarized in Algorithm 2.

Algorithm 2 LRS-Based Anomaly Detection

Input: X_{train} , X_{test} , significance level α
Output: Anomaly labels for X_{test}

```

1:  $L + S \leftarrow X_{\text{train}}$ 
2: for each time series  $j$  in  $L$  do
3:    $\mu_j \leftarrow \frac{1}{n} \sum_{i=1}^n L_j[i]$ 
4:    $\sigma_j \leftarrow \left( \frac{1}{n} \sum_{i=1}^n (L_j[i] - \mu_j)^2 \right)^{\frac{1}{2}}$ 
5:    $\tilde{L}_j[i] \leftarrow \frac{L_j[i] - \mu_j}{\sigma_j}$ 
6: end for
7:  $U \Sigma V^T \leftarrow \text{svd}(\tilde{L})$ 
8:  $\tilde{X}_{j,\text{test}}[i] \leftarrow \frac{X_{j,\text{test}}[i] - \mu_j}{\sigma_j}$ 
9:  $\hat{X}_{\text{test}} \leftarrow U \tilde{X}_{\text{test}}$ 
10: for each test vector  $\hat{x}_{\text{test}}$  in  $\hat{X}_{\text{test}}$  do
11:    $d_m \leftarrow \sum_{i=1}^q \frac{\hat{x}_{\text{test}}^2[m]}{\sigma_i^2}$ ,  $m = 1, \dots, M$ 
12: end for
13:  $s_m \leftarrow \frac{d_m - d_{\min}}{d_{\max} - d_{\min}}$ 
14:  $\tau \leftarrow \chi_q^2(\alpha)$ 
15: for each score  $s_m$  in scores do
16:   if  $s_m > \tau$  then
17:      $\hat{y}_m \leftarrow 1$ 
18:   else
19:      $\hat{y}_m \leftarrow 0$ 
20:   end if
21: end for

```

III. EXPERIMENTAL SETUP

A. Datasets

We conducted our experiments using two publicly available multivariate datasets from testbeds of multisensor systems: Secure Water Treatment (SWaT) [38], [39] and Water Distribution (WADI) [40]. More specifically, the SWaT dataset is collected from a testbed that mimics the physical process and control system of a real-world water treatment system. It contains various network traffic, sensor, and actuator measurements and spans 11 days of continuous operation (seven days of normal operation and four days under both normal and attack scenarios). The WADI dataset is collected from a testbed that expands upon the SWaT system, forming a comprehensive and realistic network for water treatment, storage, and distribution. It spans 16 days of operation (14 days of normal operation and two days under attack scenarios).

Table I presents a summary of the attributes and statistics of both the datasets and Fig. 4 depicts measurements from a selection of sensors and actuators.

B. Evaluation Metrics

In our performance evaluations, we consider binary classification metrics suitable for scenarios where anomalous samples are much less frequent than normal ones, namely, scenarios with imbalanced data. These metrics depend on the number of correctly detected anomalies [true positives (TP)], the number of erroneously detected anomalies (false positives (FP) or false alarms), the number of correctly identified normal samples [true negatives (TN)], and the number of erroneously identified normal samples [false negatives (FN)]. By defining the true

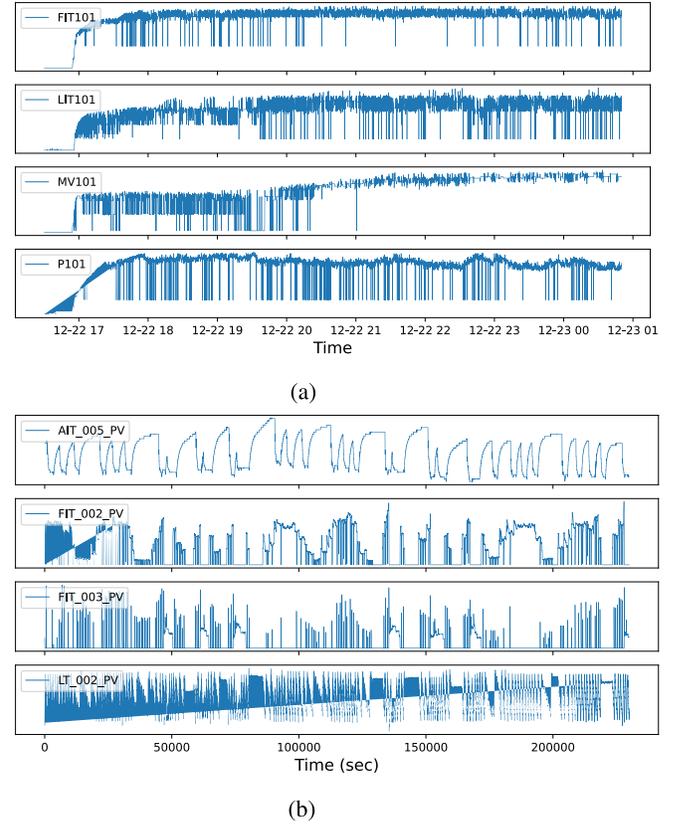


Fig. 4. Time series from selected sensors and actuators. (a) SWaT. (b) WADI.

positive rate (TPR) and the false positive rate (FPR) as

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (15)$$

the precision (P), recall (R), and F_1 -score (F_1) are given as

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad R = \text{TPR}, \quad F_1 = \frac{2 \cdot P \cdot R}{P + R} \quad (16)$$

while the receiver operating characteristic (ROC) is the curve representing TPR versus FPR and is often used to evaluate models at different threshold values. It is worth noting that the F_1 score and the area under precision–recall curve (AUPRC) are key performance metrics in anomaly detection given the common data-imbalance problem. In addition, we consider the area under the ROC curve (AUC).

C. Baseline Methods

To evaluate the performance of the proposed method, we performed a comprehensive comparison with the following state-of-the-art USAD algorithms.

- 1) Isolation forest (IF) [41], an anomaly detection approach based on decision trees and random forests.
- 2) Gaussian mixture model (GMM) [42], anomaly detection based on GMM and optimized using expectation–maximization algorithm.
- 3) Empirical cumulative distribution functions (ECOD) [43], a parameter-free and interpretable approach based on empirical cdf functions.

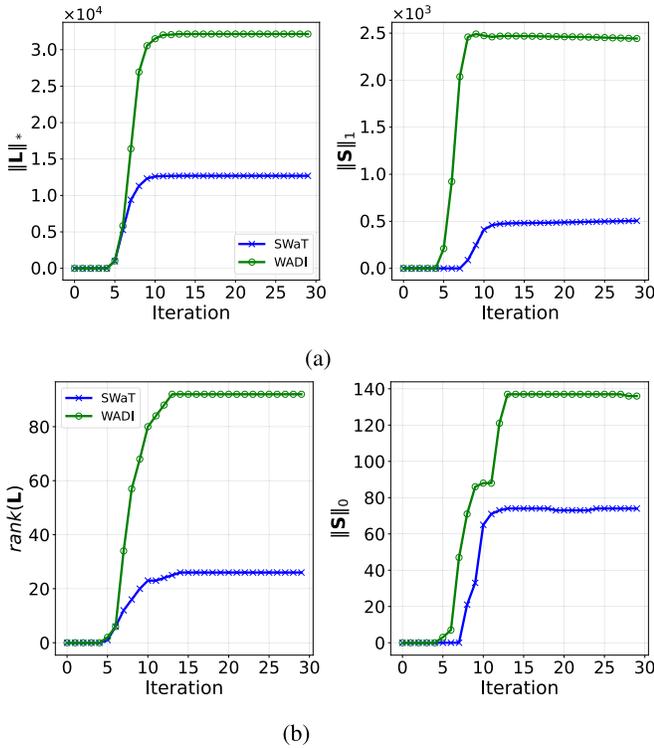


Fig. 5. Convergence of the LRS decomposition. (a) Convex formulation. (b) Nonconvex formulation.

- 4) Copula-based outlier detection (COPOD) [44], a parameter-free method based on empirical copula models.
- 5) Multilayer perceptron AEs [10], [45], a neural network architecture designed for reconstruction of latent representation.
- 6) Variational AEs (VAEs) [46], another class of AEs that represent latent space as a distribution.
- 7) Deep support vector data description (DeepSVDD) [47], a neural network designed for a one-class classification using a hypersphere.
- 8) Adversarially learned anomaly detection (ALAD) [48], a generative adversarial-network-based approach based on adversarially learned features.
- 9) USAD [49], a method via adversarially trained AEs.
- 10) Deep autoencoding GMM (DAGMM) [50], anomaly detection approach that combines a compression network and an estimation network.

We focused on computationally efficient methods that use both statistical and DNNs.¹

D. Implementation Details and Tools

We used the TensorFlow deep learning framework for training and evaluation and the Scikit-learn library for data preprocessing. All the models are trained in the Google Colaboratory Pro environment using NVIDIA T4 Tensor Core GPU processors. We conducted a comprehensive evaluation of the proposed approach using baseline implementation in

¹An implementation in the PyOD and TODS Python library is used for baselines.

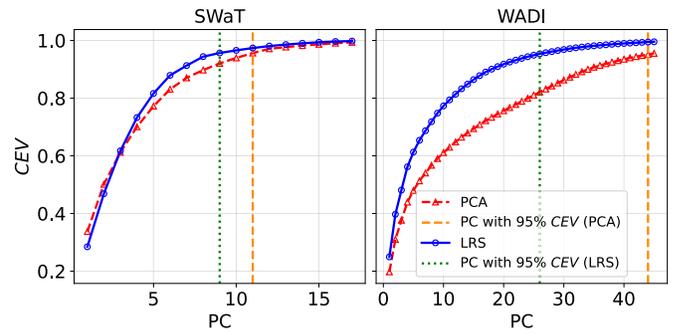


Fig. 6. Cumulative explained variance.

the PyOD Python library [51]. For the IF baseline, we used a 100-tree base estimator with a single feature to perform splits at each node. For GMM, we use a two-component mixture model with a full covariance matrix. The EM algorithm was used to train the model with 100 iterations. The AE consists of a three-layer encoder and a three-layer decoder. The number of neurons per hidden layer consists of [64, 32, 16, 32, 64], respectively. Therefore, the input vector is represented in a 16-D latent space before reconstruction. An ReLU activation function is used for hidden layers, and Sigmoid activation is used for the output layer. We train the model using mean square error loss using the Adam optimizer for 100 epochs (with 64 samples per gradient update). A dropout rate of 0.2 and $l-2$ regularization regularization strength of 0.1 used to avoid overfitting. In addition, we apply standardization to the data for faster convergence. For VAE, a similar architecture to AE is used. However, the VAE is trained using the sum of mean square error (reconstruction loss) and Kullback–Leibler divergence (KL loss). For DeepSVDD, a similar activation function, batch size, dropout rate, regularization, and standardization are applied. The hidden layer consists of neurons with 64, 32, and a neuron number equal to the dimension of the output size.

For the proposed method, we used a penalty on sparse errors $\lambda = 0.1$ and update criterion parameter $\rho = 10^{-5}$. The algorithm runs for 100 maximum number of iterations with a tolerance of 10^{-3} . Various amount of outliers were injected into the training data matrix (X_{train}). The total number of anomalies is denoted by $N_a = \lceil N \cdot r \rceil$, where r is the predefined anomaly rate, and the discrete time indices for the anomalies are uniformly generated. The amplitude values for the anomalies are generated independently for each sensor and drawn from a uniform distribution (ranging up to three times the maximum value of that sensor) and injected into the training data matrix (X_{train}) by replacing the sensor measurements

IV. NUMERICAL RESULTS AND DISCUSSION

Several experiments have been performed to assess the validity of the proposed algorithm and the related impact on robust anomaly detection for multisensor systems.

First, we explored the convergence of the LRS decomposition for both (surrogate) convex and nonconvex formulations via the SWaT and WADI datasets. Fig. 5(a) shows the

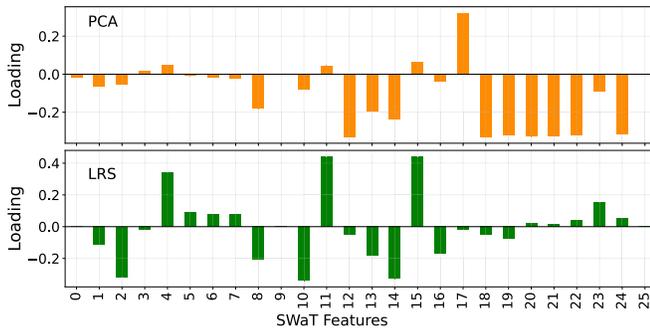


Fig. 7. SWaT PC-1 loadings (28.4% for PCA and 40.6% for LRS).

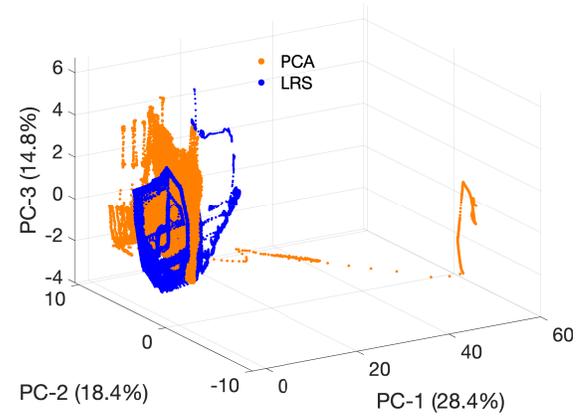
convergence over iterations of the nuclear norm of the low-rank matrix ($\|L\|_*$) and the ℓ_1 -norm of the sparse matrix ($\|S\|_1$) for both the SWAT and WADI datasets. Fig. 5(b) shows the convergence of LRS decomposition by considering the nonconvex parameters $\text{rank}(L)$ and $\|S\|_0$ for both the SWAT and WADI datasets.

The interaction among the measurements from multiple sensors and the related impact of noise and outliers was considered via the cumulative explained variance (CEV). More specifically, the CEV for k components (ζ_k) is defined as

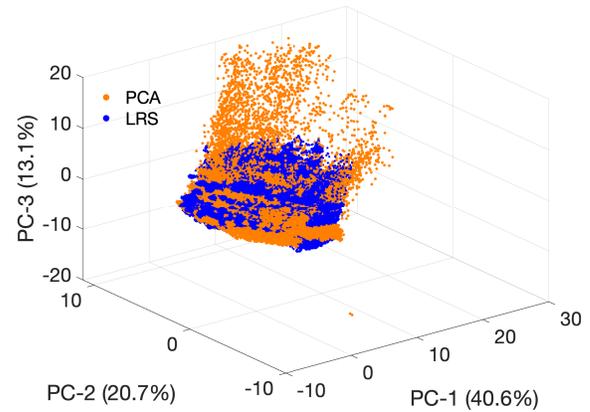
$$\zeta_k = \frac{\sum_{i=1}^k \sigma_i^2}{\sum_{i=1}^N \sigma_i^2}. \quad (17)$$

Fig. 6 shows the CEV versus the number of principal components for both the standard PCA and LRS methods. In particular, the LRS approach attains a CEV level equal to 95% for the SWAT (resp. WADI) dataset with 9 (resp. 26) components, with significantly fewer principal components compared with standard PCA, which requires 11 (resp. 44) components. The efficient variance accumulation with fewer components highlights the robustness of the LRS approach in dealing with noise and outliers in the training data, translating into a reduction of false positives during operation. Furthermore, the reduction in the required number of components underscores the potential of LRS to improve computational efficiency.

Fig. 7 presents the loadings for the first principal component (PC-1) derived from the standard PCA and LRS methods for the SWAT dataset. The loadings provide the relative contribution of the original features to the most informative principal component. Apparently, the loading spans positive and negative values in both the approaches, indicating a mix of direct and inverse relationships between the original features and the principal component. It is worth emphasizing that fewer original features contribute to the principal component in the case of LRS. The wide range of loadings also points to the vulnerability of the standard PCA to outliers. Differently, the loadings of the LRS approach present a distinct pattern with a higher concentration of loadings among fewer features. This behavior suggests a more targeted representation of the data structure, indicating that LRS prioritizes and accentuates features that are crucial to the dynamics of the underlying system. This selective emphasis is further demonstrated in the low-dimensional visualization depicted in Fig. 8.



(a)



(b)

Fig. 8. Low-dimensional visualization of the scores. (a) SWAT. (b) WADI.

A. Performance Evaluation

Fig. 9 shows the ROC curves for both the standard PCA and LRS methods applied on both the SWAT and WADI datasets with different numbers of PCs. LRS clearly outperforms the standard PCA approach in the case of the WADI dataset, while the improvement is reduced with the SWAT dataset.

Fig. 10 makes a similar comparison in terms of F_1 score, AUC, and AUPRC. Again, the benefit of the LRS approach over the standard PCA is consistent and significant in the case of the WADI dataset, while less significant with the SWAT dataset. One possible explanation for the behavior with the SWAT dataset is the relative high number of binary features in the dataset which might reduce the effectiveness of low-rank description with respect to the case of real-valued features.

B. Sensitivity Analysis

We explored the impact on the performance of the proposed algorithm of the anomaly rate (r) in the training data. Outliers were introduced into the training datasets at various anomaly rates ($r \in \{0.01, 0.05, 0.1, 0.2\}$) and the corresponding F_1 -score, AUC, and AUPRC computed and shown in Figs. 11 and 12 for the SWAT and WADI datasets, respectively. The performance degradation becomes more relevant with the

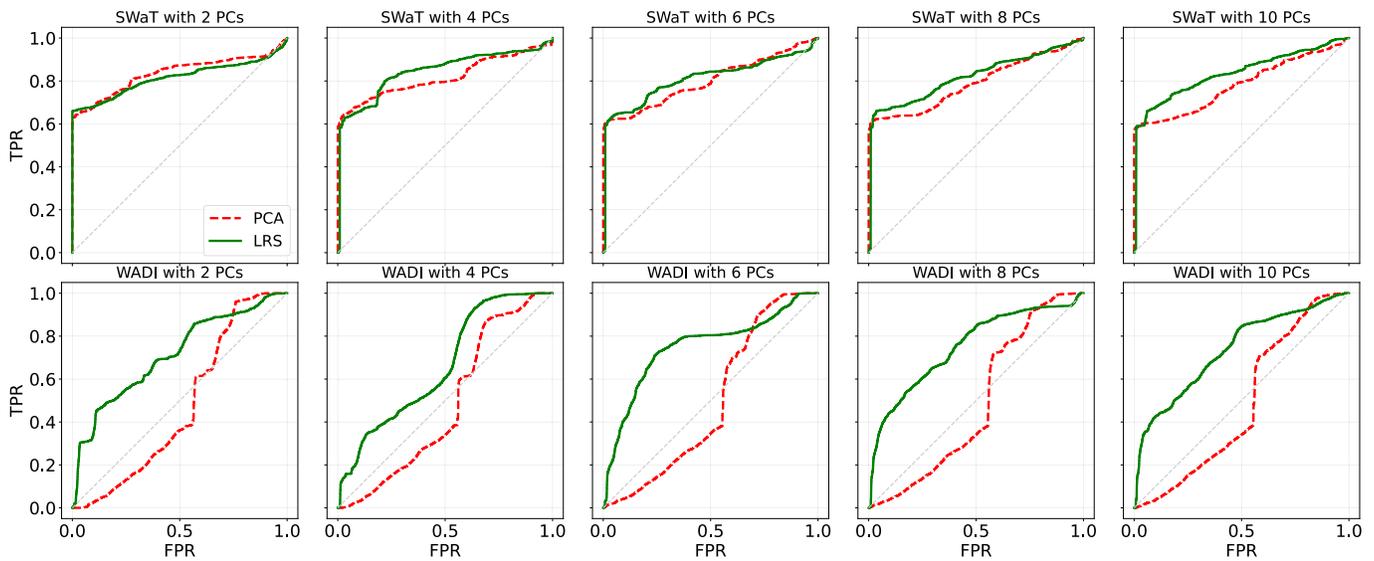


Fig. 9. LRS performance for different PCs using AUC metrics.

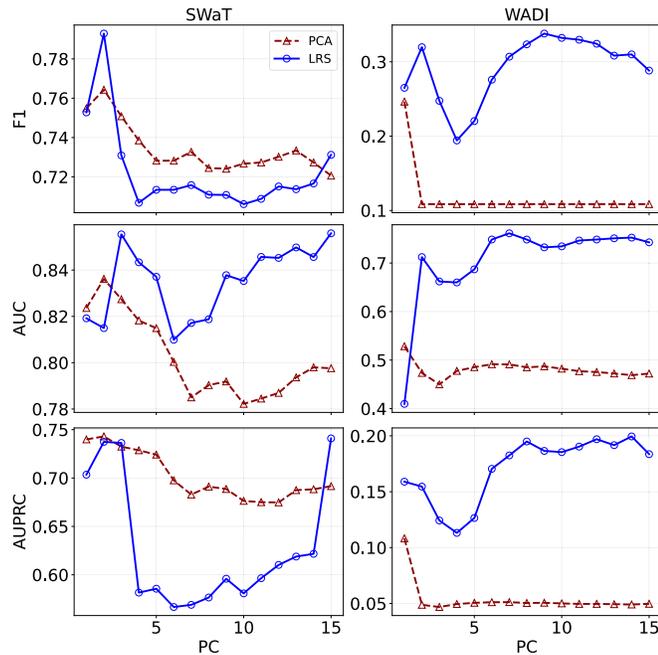


Fig. 10. Comparison of different approaches on two datasets.

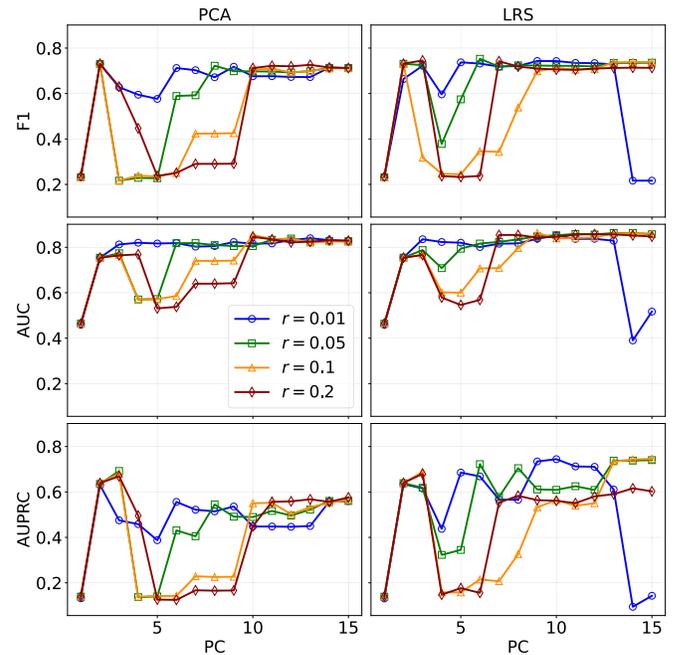


Fig. 11. Performance for noisy measurements on the SWaT dataset.

anomaly rate, which confirms that noisy and/or corrupted measurements in the training dataset pose critical challenges to anomaly detection systems. Notably, the proposed LRS-based method demonstrates substantial robustness with respect to those issues, especially in the presence of large anomaly rate.

C. Performance Comparison With Baselines

Table II presents a comprehensive performance comparison with the selected baseline anomaly detection methods. Here we refer to the LRS method with 15 principal components applied to the SWaT dataset. ECOD and COPOD show competitive performance, with COPOD slightly outperforming ECOD in

terms of F_1 -score and PRC. IF shows reasonable precision but low recall and PRC. GMM and AE show higher recall but low precision, leading to the lowest F_1 -score among the baseline methods. DeepSVDD shows a balance between precision and recall, achieving the highest F_1 -score. The proposed method outperformed the majority of the baseline methods in all the metrics, particularly in recall, F_1 -score, AUC, and PRC scores.

D. Computational Complexity of LRS

The computational complexity of LRS decomposition via ALM optimization is primarily dominated by the SVD step. Each iteration of the algorithm involves performing SVD

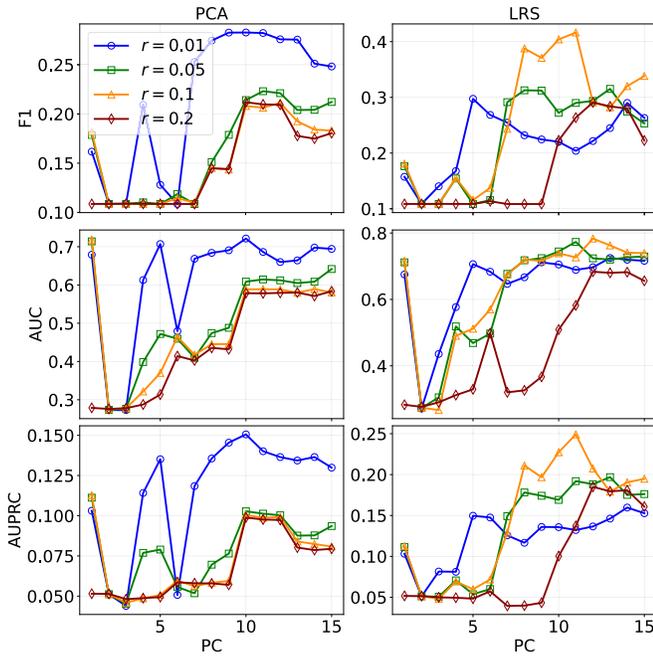


Fig. 12. Performance for noisy measurements on the WADI dataset.

TABLE II
PERFORMANCE RESULTS ON THE SWAT DATASET

Methods	P	R	F1	AUC	AUPRC
ECOD [43]	0.9764	0.5969	0.7409	0.8615	0.7571
COPOD [44]	0.9321	0.6240	0.7475	0.8600	0.7585
IF [41]	0.8510	0.5773	0.6880	0.8359	0.5033
GMM [42]	0.4107	0.6908	0.5152	0.7562	0.3757
AE [45]	0.9932	0.6234	0.7660	0.8188	0.7267
VAE [46]	0.9932	0.6234	0.7660	0.8187	0.7267
DeepSVDD [47]	0.9735	0.7060	0.8185	0.8499	0.7555
ALAD [48]	0.1221	0.9986	0.2176	0.2122	0.0884
USAD [49]	0.2018	0.8416	0.3256	0.8045	0.7030
DAGMM [50]	0.2031	0.8157	0.3253	0.8017	0.6917
LRS	0.9446	0.7137	0.8131	0.8799	0.7861

on a matrix of size $K \times N$, which has a complexity of $O(\min(KN^2, K^2N))$ based on divide and conquer algorithm. Assuming $K \leq N$, this simplifies to $O(KN^2)$. Other operations within each iteration, such as soft-thresholding and matrix updates, have lower complexities and do not significantly impact the overall computation time. Therefore, the total computational complexity of the LRS decomposition is $O(T \cdot KN^2)$, where T is the number of iterations required for convergence. This analysis highlights that the SVD step is the critical factor influencing the algorithm's efficiency. We fix $T = 10$ and analyze the time complexity. Table III presents the training and inference times' baselines and LRS. While the training time is higher compared with simpler models such as IF, it is significantly lower than deep learning models such as DeepSVDD and ALAD. The inference time of LRS is also low, making it a viable option for practical applications requiring a balance between training efficiency and inference performance.

TABLE III
COMPARISON OF TRAINING AND INFERENCE TIMES
ON THE SWAT DATASET

Model	Average Train time (sec)	Average Test time (sec)
ECOD	4.2746	7.8792
COPOD	4.1670	8.4628
IF	0.3591	0.1560
GMM	4.9242	0.4124
AE	8.6163	0.6625
VAE	84.4780	0.4630
DeepSVDD	100.3312	0.1415
ALAD	405.3361	1.7606
USAD	81.0371	4.4837
DAGMM	69.3916	3.2364
LRS	30.1502	1.1357

V. CONCLUSION AND FUTURE WORK

In this study, we introduced a novel, robust USAD method based on LRS decomposition for multivariate time series data generated by multisensor systems. The proposed approach leverages mathematical principles to provide a robust and interpretable solution to detect anomalies in complex, high-dimensional data streams from sensors that may be affected by noise and outliers. The main contributions are: enhanced dimensionality reduction through robust PCA and related improved computational efficiency, increased resilience to data noise during the training phase, and related increased reliability of anomaly detection during operation. Finally, comparative evaluations with the existing baseline methods on public datasets demonstrated the effectiveness and practical utility of the proposed approach. Despite LRS providing a robust anomaly detection approach, several research directions remain for future exploration, such as adapting the LRS framework for real-time operations of complex multisensor systems, integrating it with deep learning, and investigating alternative decomposition and optimization techniques. In addition, future work could explore the potential of graph signal processing and graph neural networks for multivariate time series anomaly detection, leveraging their capabilities in capturing complex dependencies and structures in data. It is important to note that the proposed anomaly detection method is currently most effective for point anomalies. Future research could extend the framework to address subsequence anomalies and contextual anomalies. Furthermore, while our method provides a robust foundation, we recognize the limitations of the convex relaxation approach in LRS decomposition, specifically its tendency to underestimate sparsity and rank. Future work should focus on integrating advanced nonconvex optimization techniques, such as iteratively reweighted methods and special norms, to overcome these limitations and improve the accuracy and robustness of anomaly detection. In addition, the use of the Moreau envelope ADMM algorithm could enhance convergence speed and stability of the decomposition process.

REFERENCES

- [1] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

- [2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, Jul. 2009, doi: [10.1145/1541880.1541882](https://doi.org/10.1145/1541880.1541882).
- [3] F. van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1264–1276, Mar. 2020.
- [4] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Comput. Secur.*, vol. 28, nos. 1–2, pp. 18–28, 2009.
- [5] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 303–336, 1st Quart., 2014.
- [6] A. Ukil, S. Bandyopadhyay, C. Puri, and A. Pal, "IoT healthcare analytics: The importance of anomaly detection," in *Proc. IEEE 30th Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Jun. 2016, pp. 994–997.
- [7] M. A. Belay, A. Rasheed, and P. S. Rossi, "MTAD: Multiobjective transformer network for unsupervised multisensor anomaly detection," *IEEE Sensors J.*, vol. 24, no. 12, pp. 20254–20265, Jun. 2024.
- [8] M. A. Belay, S. S. Blakseth, A. Rasheed, and P. S. Rossi, "Unsupervised anomaly detection for IoT-based multivariate time series: Existing solutions, performance analysis and future directions," *Sensors*, vol. 23, no. 5, p. 2844, Mar. 2023.
- [9] Y. Su, R. Liu, Y. Zhao, W. Sun, C. Niu, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 2828–2837.
- [10] Y. Guo, W. Liao, Q. Wang, L. Yu, T. Ji, and P. Li, "Multidimensional time series anomaly detection: A GRU-based Gaussian mixture variational autoencoder approach," in *Proc. Asian Conf. Mach. Learn.*, 2018, pp. 97–112.
- [11] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder-decoder for multi-sensor anomaly detection," 2016, [arXiv:1607.00148](https://arxiv.org/abs/1607.00148).
- [12] P. Malhotra, L. Vig, G. M. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proc. Eur. Symp. Artif. Neural Netw.*, 2015.
- [13] H. Ren et al., "Time-series anomaly detection service at Microsoft," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 3009–3017, doi: [10.1145/3292500.3330680](https://doi.org/10.1145/3292500.3330680).
- [14] C. Zhang et al., "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 1409–1416. [Online]. Available: www.aaai.org
- [15] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 665–674.
- [16] Y. Zhang, Y. Chen, J. Wang, and Z. Pan, "Unsupervised deep anomaly detection for multi-sensor time-series signals," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 2118–2132, Feb. 2021.
- [17] H. Zhao et al., "Multivariate time-series anomaly detection via graph attention network," in *Proc. Ind. Conf. Data Mining*, Nov. 2020, pp. 841–850.
- [18] Y. He and J. Zhao, "Temporal convolutional networks for anomaly detection in time series," *J. Phys., Conf. Ser.*, vol. 1213, no. 4, Jun. 2019, Art. no. 042050, doi: [10.1088/1742-6596/1213/4/042050](https://doi.org/10.1088/1742-6596/1213/4/042050).
- [19] S. Tuli, G. Casale, and N. R. Jennings, "TranAD: Deep transformer networks for anomaly detection in multivariate time series data," *Proc. VLDB Endow.*, vol. 15, no. 6, pp. 1201–1214, Feb. 2022.
- [20] X. Ma et al., "A comprehensive survey on graph anomaly detection with deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12012–12038, 2023, doi: [10.1109/TKDE.2021.3118815](https://doi.org/10.1109/TKDE.2021.3118815).
- [21] S. Schmidl, P. Wenig, and T. Papenbrock, "Anomaly detection in time series: A comprehensive evaluation," *Proc. VLDB Endow.*, vol. 15, no. 9, pp. 1779–1797, 2022.
- [22] I. T. Jolliffe, *Principal Component Analysis*. Cham, Switzerland: Springer, 1986, doi: [10.1007/978-1-4757-1904-8](https://doi.org/10.1007/978-1-4757-1904-8).
- [23] M.-L. Shyu, S. Chen, K. Sarinapakorn, and L. Chang, "A novel anomaly detection scheme based on principal component classifier," in *Proc. IEEE Found. New Directions Data Mining Workshop*, Nov. 2003, pp. 172–179.
- [24] Y.-J. Lee, Y.-R. Yeh, and Y. F. Wang, "Anomaly detection via online oversampling principal component analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 7, pp. 1460–1470, Jul. 2013.
- [25] D. Menges and A. Rasheed, "Computationally and memory-efficient robust predictive analytics using big data," 2024, [arXiv:2403.19721](https://arxiv.org/abs/2403.19721).
- [26] N. Vaswani and P. Narayanamurthy, "Static and dynamic robust PCA and matrix completion: A review," *Proc. IEEE*, vol. 106, no. 8, pp. 1359–1379, Aug. 2018.
- [27] T. Qiu, W. Huang, Y. Liao, C. Ding, and J. Wang, "Sparse optimization model based on sparse matrix and singular value vector for fault diagnosis of rolling bearings," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–9, 2024.
- [28] P. V. Giampouras, A. A. Rontogiannis, and K. D. Koutroumbas, "Alternating iteratively reweighted least squares minimization for low-rank matrix factorization," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 490–503, Jan. 2019.
- [29] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009, pp. 2080–2088. [Online]. Available: <http://perception.csl.illinois.edu/matrix-rank/home.html>
- [30] M. Rahmani and G. K. Atia, "High dimensional low rank plus sparse matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2004–2019, Apr. 2017.
- [31] R. Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5810–5819, Nov. 2012.
- [32] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.
- [33] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, Jun. 2012, doi: [10.1145/2184319.2184343](https://doi.org/10.1145/2184319.2184343).
- [34] M. Huang, S. Ma, and L. Lai, "Robust low-rank matrix completion via an alternating manifold proximal gradient continuation method," *IEEE Trans. Signal Process.*, vol. 69, pp. 2639–2652, 2021.
- [35] T. Qiu, W. Huang, Z. Zhang, J. Wang, and Z. Zhu, "A new approach for sparse optimization with Moreau envelope to extract bearing fault feature," *Mech. Syst. Signal Process.*, vol. 216, Jul. 2024, Art. no. 111493.
- [36] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York, NY, USA: Academic, 2014.
- [37] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, [arXiv:1009.5055](https://arxiv.org/abs/1009.5055).
- [38] J. Goh, S. Adepun, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," in *Proc. Int. Conf. Critical Inf. Infrastructures Secur.*, 2017, pp. 88–99.
- [39] A. P. Mathur and N. O. Tippenhauer, "SWaT: A water treatment testbed for research and training on ICS security," in *Proc. Int. Workshop Cyber-Phys. Syst. Smart Water Netw. (CySWater)*, Apr. 2016, pp. 31–36.
- [40] C. M. Ahmed, V. R. Palleti, and A. P. Mathur, "WAD: A water distribution testbed for research in the design of secure cyber physical systems," in *Proc. Int. Workshop CySWater*, Apr. 2017, pp. 25–28.
- [41] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discovery Data*, vol. 6, no. 1, pp. 1–39, Mar. 2012.
- [42] C. C. Aggarwal, *Probabilistic and Statistical Models for Outlier Detection*. Cham, Switzerland: Springer, 2017, pp. 35–64.
- [43] Z. Li, Y. Zhao, X. Hu, N. Botta, C. Ionescu, and G. Chen, "ECOD: Unsupervised outlier detection using empirical cumulative distribution functions," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12181–12193, Dec. 2022.
- [44] Z. Li, Y. Zhao, N. Botta, C. Ionescu, and X. Hu, "COPOD: Copula-based outlier detection," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 1118–1123.
- [45] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proc. MLSDA 2nd Workshop Mach. Learn. Sensory Data Anal.*, Dec. 2014, pp. 4–11.
- [46] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Found. Trends Mach. Learn.*, vol. 12, no. 4, pp. 307–392, 2019.
- [47] L. Ruff et al., "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, vol. 80, J. Dy and A. Krause, Eds., 2018, pp. 4393–4402.
- [48] H. Zenati, M. Romain, C. Foo, B. Lecouat, and V. Chandrasekhar, "Adversarially learned anomaly detection," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 727–736.
- [49] A. Julien, M. Pietro, G. Frédéric, M. Sébastien, and A. Z. Maria, "USAD: Unsupervised anomaly detection on multivariate time series," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, vol. 20, Aug. 2020, pp. 3395–3404, doi: [10.1145/3394486.3403392](https://doi.org/10.1145/3394486.3403392).

- [50] B. Zong et al., "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proc. Int. Conf. Learn. Represent.*, 2018. [Online]. Available: <https://openreview.net/forum?id=BJJLHbb0>
- [51] Y. Zhao, Z. Nasrullah, and Z. Li, "PyOD: A python toolbox for scalable outlier detection," *J. Mach. Learn. Res.*, vol. 20, no. 96, pp. 1–7, 2019. [Online]. Available: <http://jmlr.org/papers/v20/19-011.html>



Mohammed Ayalew Belay (Graduate Student Member, IEEE) received the B.Sc. degree in physics from Addis Ababa University, Addis Ababa, Ethiopia, in 2011, the M.Sc. degree in physics from Bahir Dar University, Bahir Dar, Ethiopia, in 2013, and the M.Sc. degree in computer science from Hawassa University, Awasa, Ethiopia, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electronic Systems, Norwegian University of Science and Technology (NTNU),

Trondheim, Norway.

His research interests include deep learning, unsupervised learning, anomaly detection, time series analysis, and physics-informed neural networks.



Adil Rasheed is a Professor at the Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway. There, he works to advance the development of novel hybrid methods that combine big data, physics-driven modeling, and data-driven modeling in the context of real-time automation and control. In addition, he also holds a part-time Senior Scientist position at the Department of Mathematics and Cybernetics, SINTEF Digital, Trondheim, where he previously served as the Leader of the Computational Sciences and Engineering Group from 2012 to 2018. His contributions in these roles have been the development and advancement of both the hybrid analysis and modeling and big data cybernetics concepts. Over the course of his career, he has been the driving force behind numerous projects focused on different aspects of digital twin technology, ranging from autonomous ships to wind energy, aquaculture, drones, business processes, and indoor farming. He is currently leading the digital twin and asset management-related work at the FME Northwind Center.



Pierluigi Salvo Rossi (Senior Member, IEEE) was born in Naples, Italy, in 1977. He received the Dr.Eng. (summa cum laude) degree in telecommunications engineering and the Ph.D. degree in computer engineering from the University of Naples "Federico II," Naples, in 2002 and 2005, respectively.

He is currently a Full Professor and the Deputy Head with the Department of Electronic Systems, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. He is also a (part-time) Senior Research Scientist with the Department of Gas Technology, SINTEF Energy Research, Trondheim. Previously, he worked with the University of Naples "Federico II"; the Second University of Naples, Naples; NTNU; and Kongsberg Digital AS, Asker, Norway. He held visiting appointments with Drexel University, Philadelphia, PA, USA; Lund University, Lund, Sweden; NTNU; and Uppsala University, Uppsala, Sweden. His research interests fall within the areas of communication theory, data fusion, machine learning, and signal processing.

Prof. Salvo Rossi was awarded as an Exemplary Senior Editor of IEEE COMMUNICATIONS LETTERS in 2018. He is (or was) in the Editorial Board of IEEE SENSORS JOURNAL, IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS, IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, IEEE COMMUNICATIONS LETTERS, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.