A Sparse and Low Rank Penalized Signal Decomposition Model with Constraints: Anomaly Detection in PV Systems

Wei Yang*, Daniel Fregosi†, Michael Bolen† and Kamran Paynabar*
*Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
†Electric Power Research Institute (EPRI), Palo Alto, CA 94304, USA

***Abstract*—Recently, robust PCA has seen its wide application in various industries for its ability to perform the task of anomaly detection. The essence of robust PCA approach is to break down the signal into a low rank component and sparse component. In many applications, a simple breakdown of the signal without accounting for the signs of low rank components and sparse components would violate the physical constraints of the decomposed signal. In addition, often times, the signals in the real world collected for a long duration has smooth changes within a day and between days. As an example, the power signals collected in a photovoltaic (PV) system are cyclostationary, exhibiting these characteristics. Neglecting the smoothness of signals would result in miss detection of anomalous signals which are smooth within a day but non-smooth between days and vice versa. In this paper, we developed a signal decomposition approach for the purpose of anomaly detection based on the idea of low rank and sparse decomposition taking into consideration the signs of the decomposed low rank and sparse components and the within-day and between-day smooth changes in the original signals. The proposed unsupervised approach for fault detection eliminates the need for faulty samples required by other machine learning methods. It does not require the full I-V characteristics to work. Furthermore, there is no need for complex modelling of PV systems as in the case of power loss analysis. Using Monte Carlo simulations, we demonstrate the ability of our proposed approach for detecting anomalies of different duration and severity in PV systems.**

***Index Terms*—Anomaly detection; PV systems; signal decomposition; low rank and sparse decomposition; constrained optimization**

I. Introduction

Photovoltaic systems due to their ability of converting solar energy into electricity in a clean fashion, are playing a bigger role in the evolution of the energy sector [1]. According to the projections by U.S. Energy Information Administration [2], solar generation will account for 14% of the U.S. total electricity generation in 2035 and 20% in 2050. Despite its noticeable advantages over traditional energy resources such as fossil fuel, the faults in PV systems are hindering their wide applications across industries. Without responsive identification of the faults in PV systems, undetected faults do not only impact the power output but also accelerate system aging or result in even fire hazards in worst scenario [3].

This work is funded in part by the U.S. Department of Solar Energy Technologies Office, under Award Number DE-EE-0008976.

Generally, the operation of PV systems can be categorized into three states including ideal, normal and actual operating states. The ideal operating state represents the system performing at its rated efficiency, which is almost never the case in real world scenarios. Therefore, a normal operating state depicts the systems functioning as expected despite a slight deviation from the ideal performance due to unavoidable losses such as temperature loss and inverter conversion loss. The actual operating state is when faults occur in PV systems, which lead to additional drops in the system efficiency unexpectedly [4]. The smallest unit in a PV system is the solar cells which make up the PV module, equivalently, a solar panel. When the sun strikes the solar panel, the solar energy is then converted into electricity in the form of DC current which can be converted to AC current depending on the applications. Stacking the PV modules in either parallel or series configuration transforms the individual PV modules into an array form. The array power equals to the summation of power output from individual PV modules. In field conditions, the power output of the PV arrays might not be close to the predicted power. There are numerous factors contributing to this deviation. The faults in PV systems are therefore defined as any factor reducing the power output. Based on the duration of the faults, they can be classified into two categories including temporary faults (shadows, bird droppings, and dust or snow accumulation on the surface of the PV panel) and permanent faults (electrical disconnection, wiring losses and ageing) [5].

The fault detection in PV systems is an essential task to increase system reliability, efficiency and safety. To detect the fault in a signal, it is natural to think of a detection strategy which decomposes the signal into a faulty component and non-faulty component. In our case, the faulty component is the influence of the faults which brings down the power output of the systems. The non-faulty component is the normal power output of the systems free from the impact of faults. However, directly monitoring the power signal of the system would treat the cloud influence over the system as faults which are not the actual faults of the PV systems although cloud influence reduces the power output the systems. To address this issue, the power signal used in this study is normalized by the irradiance which is measured by the irradiance sensors at the site. As the cloud influence reduces the irradiance and power output measurements simultaneously, this normalization uncovers the actual system efficiency by removing the effect of the cloud on the systems. Considering the periodic behavior of the signals across multiple days, the one-dimensional signal of the power to irradiance ratio collected at every minute is reformulated into a matrix form where the row represents the time within a day while the column represents the index of a day within a given period. Anomaly detection in the reformulated matrix can be realized through modelling the mean as a low-rank component while modelling the anomalies as the sparse matrix as the within-day and between-day correlations of the power to irradiance ratio signal implies low-rank structure of the background and the anomaly is assumed to be sparse.

Recently, low rank and sparse decomposition (LRSD) has demonstrated its successful performance in anomaly detection in various applications involving hyperspectral images [6], infrared thermal images [7] and face images [8]. Let $\boldsymbol{Y} \in \mathbb{R}_{m\times n}$ denotes the original data matrix, $\boldsymbol{L} \in \mathbb{R}_{m\times n}$ denotes the low-rank component and $\boldsymbol{S} \in \mathbb{R}_{m\times n}$ denotes the sparse component. The problem of LRSD can be mathematically described using the following convex optimization problem:

$$\min_{\boldsymbol{L},\boldsymbol{S}} \|\boldsymbol{L}_*\| + \lambda\|\boldsymbol{S}\|_1 \quad s.t. \quad \boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{S} \tag{1}$$

where $\|\boldsymbol{L}_*\| = \sum_r \sigma_r(\boldsymbol{L})$ denotes the nuclear norm of $\boldsymbol{L}$, $\sigma_r(\boldsymbol{L})$ $(r = 1, 2, \ldots, min(m, n))$ is the r_{th} singular value of $\boldsymbol{L}$, $\|\boldsymbol{S}\|_1$ denotes the L_1 norm of $\boldsymbol{S}$.The well-known implementation of low rank and sparse decomposition is robust PCA [9]. Extending the concept of LRSD to detect anomalies in a smooth background, Hao et al. [10] proposed using a smooth basis matrix to extract the coefficients corresponding to the smooth background while using a predetermined basis matrix to extract the coefficients corresponding to the anomaly. A roughness matrix was used to enforce the smoothness of the extracted coefficients of the smooth component. Their proposed method demonstrated superior performance over traditional methods designed for anomaly detection in images including Sobel edge detection [11], jump regression with local polynomial kernel regression [12], the Otsu global thresholding method [13], and the Nick local thresholding method [14]. However, although their method penalizes smoothness of the coefficients corresponding to the smooth component, their method does not permit the separate control of smoothness of within-day variations and between-day variations corresponding to the row wise and column wise changes in the reformulated matrix, respectively. In addition, the low rank based method for anomaly detection in the images and the aforementioned traditional image detection methods do not take into consideration the signs of the decomposed components which carry physical meanings. In our case, the mean signal corresponds to the system efficiency defined as the ratio of power output to irradiance, which should be non-negative. The anomaly corresponds to the negative impact of the faults on system efficiency, which reduces the ratio of power output to irradiance by reducing the expected power output of the system.

In this paper, we proposed a sparse and low rank penalized signal decomposition model with constraints to decompose the signals into faulty and non-faulty components by considering within-day and between-day smoothness of the reformulated signal matrix. In addition, the proposed model takes care of the signs of decomposed components in order to meet their physical constraints. The proposed methodology in this work can be easily adapted for anomaly detection in other cyclostationary signals whose statistical properties vary cyclically with time [15]. The periodic characteristics of the cyclostationary signal is preserved through controlling the smoothness of the changes between periods.

II. Results

To examine the proposed method, we performed Monte Carlo simulation. We considered nine different faulty scenarios where the power percentage drop has three levels including 30%, 60% and 90% and the duration of the faults has three levels at 3, 5 and 7 hours. 100 normal data sets were generated by randomly drawing 4 weeks out of the 20 simulated normal weeks for 100 times. These 100 normal data sets were used to determine the mean and variance of the training false rate. The training false alarm rate is calculated as the percentage of falsely flagging normal days as faulty in a normal data set. For each of the faulty scenarios, 100 test data sets were generated to calculate the mean and variance of the miss detection rate. The miss detection rate is calculated as the percentage of falsely classifying faulty days as normal in a test data set. Each test data set has a duration of 20 weeks. In each week, there is one day injected with the fault. An example of fault detection for a week in which faults with 30% power loss with a duration of 3 hours is presented in Figure 1. The mean and variance of the false alarm rate are 0.03 and 0.001 using a control limit of 0.124. The mean and variance of the miss detection rate are summarized in Table I. The results indicate

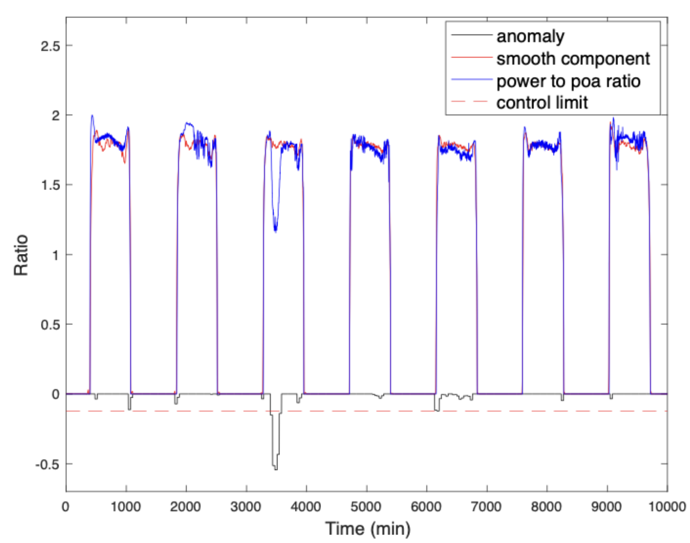

Fig. 1. An example of fault detection on the power to irradiance ratio signal with the presence of a fault leading to 30% power drop over 3 hours.

TABLE I
MEAN AND VARIANCE OF MISS DETECTION RATE

Duration of Faults (hours)	Percentage of Power Drop		
	30%	50%	70%
3	0.0180 $\pm$ 0.00160	0.0015 $\pm$ 0.00007	0.0005 $\pm$ 0.00003
5	0.0056 $\pm$ 0.00035	0.0005 $\pm$ 0.00003	0 $\pm$ 0
7	0.0030 $\pm$ 0.00014	0 $\pm$ 0	0 $\pm$ 0

that our proposed method can detect faults with a high level of accuracy even when the fault duration is relatively short and the percentage of power loss is relatively small. As the fault severity increases, the method can detect it with higher accuracy and smaller variance.

III. Discussion

The developed fault detection method taps into the power and irradiance signals collected at the site. This method does not require the full I-V characteristics to work, which is the disadvantage of the method of analyzing I-V characteristics [16], [17]. Another area of research for the fault detection in PV systems explores the use of power loss analysis technique [18], [19]. For the power loss analysis, the performance of the fault detection method is highly sensitive to the accuracy of the simulated model employed. Also, the development and calibration of the PV models for different sites are time consuming. Furthermore, the simulated model may not well represent the PV systems under complex environmental conditions. Different from the power analysis method, the proposed method does not involve complex modelling of PV systems. As recent advances in machine learning techniques, researchers also utilize machine learning tools to tackle fault detection challenges in PV systems [20], [21]. However, the machine learning based approaches require simulated or real faulty samples. On the one hand, collecting the faulty samples from the real sites is a costly and unsafe practice in some cases. On the other hand, it is impossible to include all the faulty scenarios in the training dataset. Therefore, the ability of the trained machine learning model to detect unseen faulty samples remains questionable. As an unsupervised method, this approach we developed does not require faulty samples to train the model. Our work provides a simple-to-implement yet robust tool for the detection of faults in PV systems.

IV. Conclusion

In this work, we developed an unsupervised fault detection algorithm using the power to irradiance ratio signal. Our proposed method was tested under various fault scenarios. Detection accuracy remains high even as fault severity decreases. Assuming smoothness both within and between periods, the developed fault detection algorithm can also be adapted to detect faults in other cyclostationary signals. As opposed to the methods in the literature, our approach does not require obtaining full I-V characteristics, simulating PV systems, or generating faulty samples.

References

[1] S. K. Sansaniwal, V. Sharma, and J. Mathur, "Energy and exergy analyses of various typical solar energy applications: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1576–1601, 2018.

[2] U.S. Energy Information Administration, "Annual energy outlook 2021," Feb 2021.

[3] R. Hariharan, M. Chakkarapani, G. Saravana Ilango, and C. Nagamani, "A method to detect photovoltaic array faults and partial shading in pv systems," *IEEE Journal of Photovoltaics*, vol. 6, no. 5, pp. 1278–1285, 2016.

[4] S. Firth, K. Lomas, and S. Rees, "A simple model of pv system performance and its use in fault detection," *Solar Energy*, vol. 84, no. 4, pp. 624–635, 2010, international Conference CISBAT 2007.

[5] S. R. Madeti and S. Singh, "A comprehensive study on different types of faults and detection techniques for solar photovoltaic system," *Solar Energy*, vol. 158, pp. 161–185, 2017.

[6] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection," *IEEE Transactions on Cybernetics*, vol. 51, no. 9, pp. 4363–4372, 2021.

[7] Q. Wang, K. Paynabar, and M. Pacella, "Online automatic anomaly detection for photovoltaic systems using thermography imaging and low rank matrix decomposition," *Journal of Quality Technology*, vol. 0, no. 0, pp. 1–14, 2021.

[8] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, "Robust principal component analysis with complex noise," in *Proceedings of the 31st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research. Bejing, China: PMLR, 2014, pp. 55–63.

[9] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011. [Online]. Available: https://doi.org/10.1145/1970392.1970395

[10] H. Yan, K. Paynabar, and J. Shi, "Anomaly detection in images with smooth background via smooth-sparse decomposition," *Technometrics*, vol. 59, no. 1, pp. 102–114, 2017.

[11] Sobel and G. Feldman, "A 3x3 isotropic gradient operator for image processing," *Pattern Classification and Scene Analysis*, pp. 271–272, 1973.

[12] P. Qiu and J. Sun, "Local smoothing image segmentation for spotted microarray images," *Journal of the American Statistical Association*, vol. 102, no. 480, pp. 1129–1144, 2007.

[13] Otsu and Nobuyuki, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[14] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Computational Statistics and Data Analysis*, vol. 54, no. 4, pp. 1167–1178, 2010.

[15] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Processing*, vol. 86, no. 4, pp. 639–697, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168405002409

[16] A. Eskandari, J. Milimonfared, and M. Aghaei, "Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on i-v characteristics," *Solar Energy*, vol. 211, pp. 354–365, 2020.

[17] M. T. Boyd, S. A. Klein, D. T. Reindl, and B. P. Dougherty, "Evaluation and Validation of Equivalent Circuit Photovoltaic Solar Cell Performance Models," *Journal of Solar Energy Engineering*, vol. 133, no. 2, p. 021005, 2011.

[18] A. Chouder and S. Silvestre, "Automatic supervision and fault detection of pv systems based on power losses analysis," *Energy Conversion and Management*, vol. 51, no. 10, pp. 1929–1937, 2010.

[19] W. Chine, A. Mellit, A. M. Pavan, and S. Kalogirou, "Fault detection method for grid-connected photovoltaic plants," *Renewable Energy*, vol. 66, pp. 99–110, 2014.

[20] Y. Zhao, L. Yang, B. Lehman, J.-F. de Palma, J. Mosesian, and R. Lyons, "Decision tree-based fault detection and classification in solar photovoltaic arrays," in *2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2012, pp. 93–99.

[21] S. A. Zaki, H. Zhu, M. A. Fakih, A. R. Sayed, and J. Yao, "Deep-learning–based method for faults classification of pv system," *IET Renewable Power Generation*, vol. 15, no. 1, pp. 193–205, 2021.