Partial Discharge Pattern Recognition via Sparse Representation and ANN

Article *in* IEEE Transactions on Dielectrics and Electrical Insulation · April 2015

DOI: 10.1109/TDEI.2015.7076807

CITATIONS
63

READS
380

4 authors, including:

Mehrdad Majidi
University of Nevada, Reno
34 PUBLICATIONS **642** CITATIONS
SEE PROFILE

Sami Fadali
University of Nevada, Reno
286 PUBLICATIONS **3,737** CITATIONS
SEE PROFILE

Mohammad Oskuoee
11 PUBLICATIONS **171** CITATIONS
SEE PROFILE

Partial Discharge Pattern Recognition via Sparse Representation and ANN

Mehrdad Majidi, Mohammed Sami Fadali, Mehdi Etezadi-Amoli
Department of Electrical & Biomedical Engineering, University of Nevada, Reno (UNR), USA

and **Mohammad Oskuoee**
High Voltage Department, Niroo Research Institute (NRI), Tehran, Iran

ABSTRACT

In this study, seventeen samples were created for classifying internal, surface, and corona partial discharges (PDs) in a high voltage lab. Next, PDs were measured experimentally to provide a dictionary comprising the types. Due to the huge size of the recorded dataset, a new and straightforward preprocessing method based on signal norms was used to extract the appropriate features of various samples. The new sparse representation classifier (SRC) was computed using ℓ^1 and stable ℓ^1-norm minimization by means of Primal-Dual Interior Point (PDIP) and Basis Pursuit De-noise (BPDN) algorithms, respectively. The pattern recognition was also performed with an artificial neural network (ANN) and compared with the sparse method. It is shown that both methods have comparable performance if training process, tuning options, and other tasks for finding the best result from ANN are not taken into account. Even with this assumption, it is shown that SRC still performs better than ANN in some cases. In addition, the SRC technique presented in this paper converges to a fixed result, while the results after training the ANN vary with every run due to random initial weights.

Index Terms - **sparse representation, compressive sensing, partial discharges, pattern recognition, ℓ^1 and stable ℓ^1-norm minimization, ANN, signal norms.**

1 INTRODUCTION

PARTIAL discharge is one of the main causes of electrical insulation failures in the long term [1]. Because electrical insulation breakdowns mostly lead to electrical equipment outages especially in HV and EHV levels and cascading failures, on-line PD monitoring schemes have gained importance both in industry and academic studies in recent years [2]. Online monitoring can detect PDs and save the datasets for use in pattern recognition and feature extraction. This plays an essential role for evaluating the condition of on-line HV devices by identifying the types of defect that lead to the discharges [3]. Each type of PD source has a distinct behavior and pattern which can be recognized by some particular features [4]. Extracting these features is a preprocessing step for transforming the raw data to discriminatory features. A classical method which is commonly used is Phase Resolve Partial Discharge (PRPD) by which three main features of PDs are defined. They are maximum charge, average charge and number of PD pulses in each degree of a 360 degree cycle [5-16]. The main problem with this method is the lack of features and the need to provide most classifiers with more inputs. New methods, such as wavelets or fast Fourier transform techniques have been applied to PD signals to estimate features [17]. These methods not only increase the number of attributes, but also decrease the size of the required dataset to a reasonable scale. Popular preprocessing methods include Principle Component Analysis (PCA) [18], statistical computations [19], and the pulse shape parameter approach [20].

After the preprocessing stage, the extracted features must be trained to recognize patterns. There are many advanced tools for this step including ANN, support vector machines, decision function classifiers and fuzzy logic [4, 17].

In this paper, we extract the features of PD signals by a combination of PRPD and signal norms. Our main goal is to introduce the sparse representation classifier (SRC) for recognizing PD patterns. We compare the performance of SRC with a common method such as ANN.

To our knowledge, only [17] used SRC to classify the PD patterns using their pulse shapes. Pulse shape characterization is a PD separation technique which is limited to the experimental condition. We use a completely different experimental approach based on PD identification. PD identification is usually performed using PRPD recognition.

Manuscript received on 20 February 2014, in final form 11 July 2014, accepted 7 August 2014.

DOI 10.1109/TDEI.2014.004622

In this paper, we propose a new method that extracts nine features from raw measured data by observing PDs in each degree of each cycle, as opposed to the three only in PRPD that are extracted by observing PDs in defined windows through several cycles. The method is based on the 1-norm, 2-norm, and infinity-norm of the number of charges, peak charge, and mean charge in each degree of each cycle. The method also investigates several questions that were not addressed in [17]:

- In [17], ℓ^1-norm minimization was employed to calculate the sparse vector without comparing different minimization algorithms. Here, we utilize and compare several algorithms to minimize both the ℓ^1 and stable ℓ^1-norms.
- Reference [17] is primarily devoted to preprocessing using wavelets and the fast Fourier transform. SRC is used as one of several classifiers but details of its application are omitted. We present SRC in detail and evaluate its classification performance using a new preprocessing approach for features extraction via signal norms.
- Five object classes are considered in [17] and 94.9% separation percentage is obtained by SRC. We consider up to seventeen samples in four scenarios and achieve 99.71% identification among seven samples for the best scenario.
- In [17], no more than 44 data points are used for training and no more than 25 data test points are used for each class. In this study, we record a large dataset spanning 3000 cycles of measurement for each sample, then randomly select 200 cycles for training and 50 cycles for testing.

The remainder of the paper is organized as follows. In Section 2, we describe the measurement setup and the experimentally collected data. Section 3 describes classification and the SRC. Section 4 describes ℓ^1 and stable ℓ^1-norm minimization. Section 5 describes and analyzes the performance of SRC in four scenarios for PD pattern recognition. Section 6 compares the results and Section 7 presents conclusion.

2 MEASUREMENT SETUP AND ARTIFICIAL DEFECTS

In this study, PD signals of 17 samples were measured in a high voltage laboratory, where 15 samples are associated with internal PDs and the two other samples are for surface and corona PD measurements. The 15 samples consist of 1-5 air voids with dimensions 1, 1.5 and 2 mm. These classifications enable us to distinguish between different types of PD patterns such as corona, surface and internal PDs. In addition, we detect the number and size of the voids where internal PDs occur.

Collecting these samples is complicated by the need to ensure that there is no PD sources other than the one anticipated. Solid resin materials are used for surface and internal PDs and creating the voids for internal PDs is done by drilling on the surface of two circular solid resins and attaching them together. In practice, one hemisphere with a particular radius is created on each surface and the two hemispheres are accurately attached together to produce one sphere with that radius as the internal void. For internal PD measurements, the objects are sandwiched between two plane electrodes cylindrically and immersed in mineral oil to avert any other PDs in the test system. The radius and thickness of the planes are 50 mm and 10 mm, respectively. A Rogowski curve is made in the edge of the electrodes by metal bronze to create a uniform electrical field in the space between the two electrodes while decreasing the field in the edges. Corona is measured by needle-plane electrodes in air. The radius of the plane, needle tip, and resin objects for internal and surface PDs are 50 mm, 100 μm, 35 mm, and 50 mm, respectively.

The measurement system in the high voltage laboratory included a 220 V/200 kV high-voltage transformer which is PD free, a 1 nF coupling capacitor installed in parallel with the HV side of the transformer, a PD detector in series with the test object, and an LDS-6 PD measurement system, with an LDIC monitor [21]. An autotransformer was connected to the LV side of the transformer to regulate the applied voltage on the HV side.

In accordance with the IEC 60270 Standard [22], the test voltage was increased smoothly to a level higher than the Partial Discharge Inception Voltage (PDIV) at which PD behavior becomes stable. When the applied voltage is gradually decreased from a level higher than PDIV, continuous PDs no longer occur below the Partial Discharge Extinction Voltage (PDEV). The measurement process was continued for about 3000 cycles and the PDs were recorded at 0.1 degree intervals. More than 10 million data points were saved for each sample including voltage, PD charges, and the phases in which PD occurred. The measurement voltages of, PDIV and PDEV for each sample are presented in Table 1.

3 DATA PREPROCESSING, FEATURE EXTRACTION AND SPARSE REPRESENTATION

Raw measured data can be converted into discriminatory features by means of signal norms [23]. Three main signal norms, 1, 2 and infinity, can provide suitable signal features. Schematic of data preprocessing method is demonstrated in Figure 1 where the extracted features for the raw measured data of a sample with 4 voids 1.5 mm in each step are shown. Due to the negative applied voltage in second half cycle, partial discharges are negative between 180 and 360 degrees. Therefore, absolute values of the partial discharges are used to eliminate the negative data in the process. Based on Figure 1, 250 cycles, each comprising 3600 measurements, were selected randomly from the 3000 cycles of measured data for each sample. Note that each cycle does not necessarily include 3600 PD events. But our measurement system is able to record the possible PD events in each 0.1 degree interval.

Table 1. Measurement conditions.

Sample	Corona	Surface	1 mm					1.5 mm					2 mm				
			1 void	2 voids	3 voids	4 voids	5 voids	1 void	2 voids	3 voids	4 voids	5 voids	1 void	2 voids	3 voids	4 voids	5 voids
RMS of Measurement Voltage (kV)	5	6.3	14	14.7	14.5	14	12	16	16	14	12	12	16	16	14	12	12
PDIV (kV)	3.5	5.1	11.8	11.6	11.3	11	9	11	10.5	10.3	8.5	8.5	10.5	10	10	8.3	8
PDEV (kV)	3	4.4	10.9	10.7	10.5	9	7.5	10	8.5	8.4	7.2	7	9.5	8	7.5	6.5	6.4

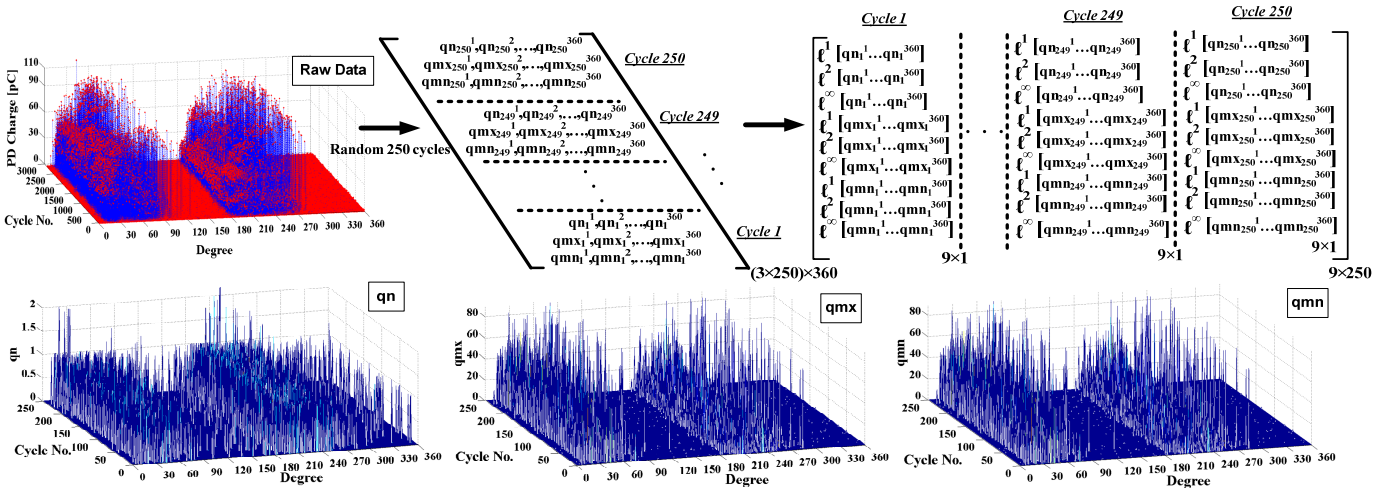


Figure 1. Schematic of preprocessing algorithm.

For the 10 measurements in each degree, we calculated: the maximum discharge (q_{mx}), the mean discharge (q_{mn}) and the rate of PD repetition (q_n). We used the data to form a 3×360 matrix corresponding to each cycle. In Figure 1, qn_i^j , qmx_i^j, and qmn_i^j denote the number of PD pulses, maximum discharge, and mean discharge given by the sum of all discharge magnitudes divided by number of PDs in i^{th} cycle and j^{th} degree, respectively, where i=1,…,250 and j=1,…,360. The figure also shows the qn, qmx, and qmn curves versus cycle numbers and degrees. Next, we calculated the 1-norm, 2-norm, and infinity-norm for each row of the matrix to obtain a 9×1 vector. In addition, norm operators were used to convert each row including 360 coefficients to one single value:

$$\|S\|_1 = \sum_{j=1}^{360} |S^j| \quad (1)$$

$$\|S\|_2 = \left[\sum_{j=1}^{360} (S^j)^2\right]^{1/2} \quad (2)$$

$$\|S\|_\infty = \max_j |S^j| \quad (3)$$

where $\|.\|_1$, $\|.\|_2$, and $\|.\|_\infty$ are ℓ^1, ℓ^2, and ℓ^∞-norms, respectively and S can be qn, qmx, and qmn signals in each Figure 1 cycle. Each of the three signals in each cycle is a 1×360 vector and three norms operators compress each signal to a 3×1 vector. Linking the three converted vectors together produces a 9×1 vector for all three signals.

The process was repeated for all 250 selected cycles to obtain a 9×250 matrix whose columns are the 9×1 vectors of each cycle. Thus, each sample can be represented with a 9×250 matrix to which we can apply pattern recognition methods. For instance, Figure 2 shows the PD patterns of the samples with 1-5 voids 1.5 mm and patterns of corona and surface PDs. In each cycle, the nine extracted features are normalized using their maximum value. Since the PD is a random event, it does not follow a particular pattern in each cycle. Therefore, the pattern recognition becomes difficult and time consuming for other methods such as ANN.

The calculation of the Sparse Representation (SR) vector is an active research topic in the compressive sensing literature. One approach used for face recognition [24] creates a dictionary of feature matrices extracted from the experimentally collected data. For our application, this is the data obtained in the preprocessing step for each sample. A new input signal is classified based on comparison to the feature vectors in the dictionary if SR vector coordinates are nonzero for a particular vector and zero or negligible for others. Thus, the non-zero elements in the vector determine the class to which input signal belongs. In other words, the input signal is characterized by a linear combination of extracted features. As most entries of the vector are zero or negligible, it is called a SR vector. When there are M classes in our dictionary and each class has $n_i, i = 1, \dots, M$ feature vectors consisting of k attributes; the feature matrix for each sample can be defined by

$$T_i = [f_{i1}, f_{i2}, \dots, f_{in_i}] \in R^{k \times n_i}, i = 1, \dots, M \quad (4)$$

where f_{ij} denotes the j^{th} feature vector for the i^{th} class.

Note that the number of the training (feature) vectors in each sample can vary. By concatenating the feature matrices of the M classes, we define the combined matrix

$$A = [T_1 \vdots T_2 \vdots T_3 \vdots \cdots \vdots T_M] \in R^{k \times p} \quad (5)$$

where $p=\sum_{i=1}^{M} n_i$

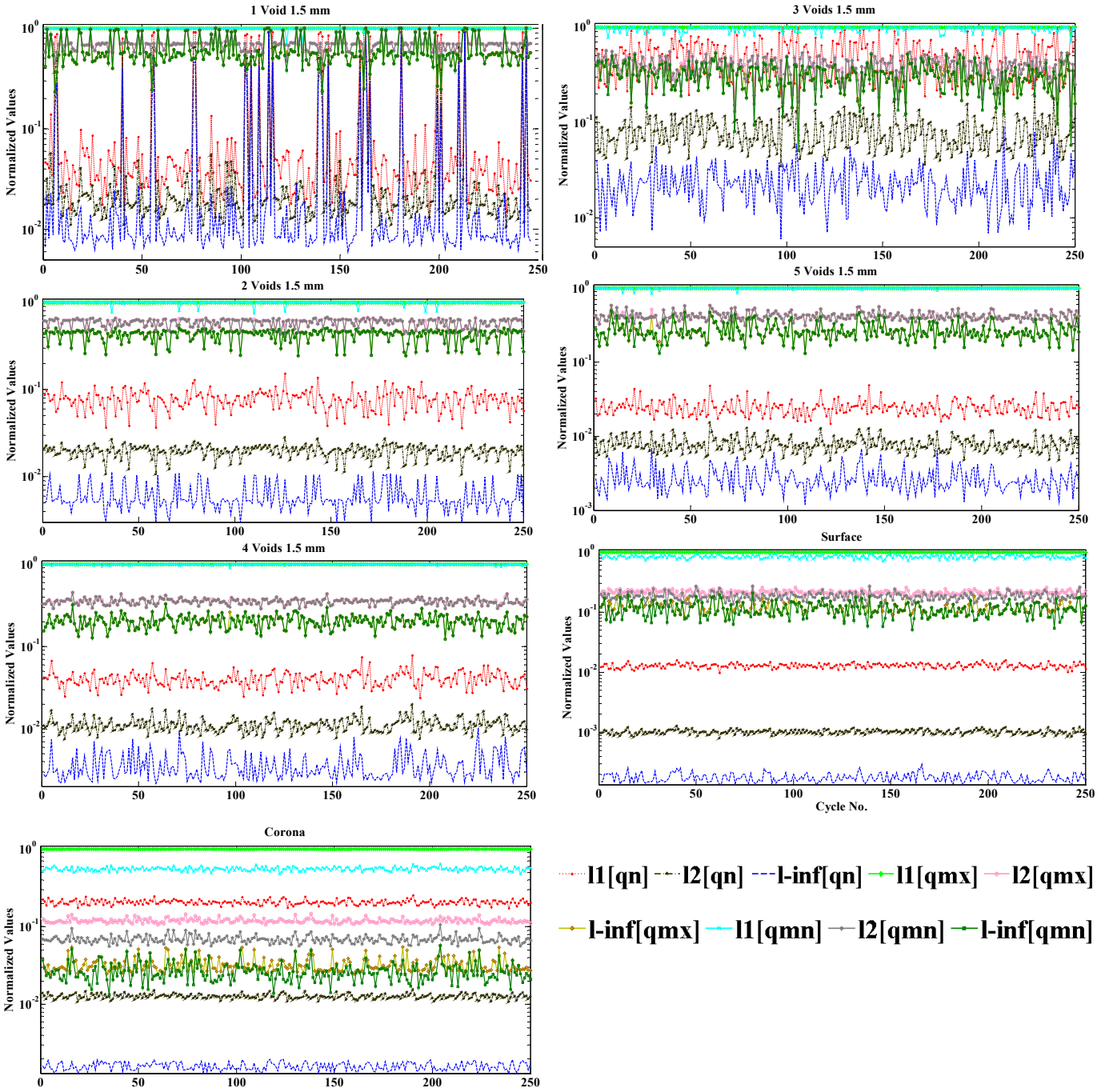


Figure 2. PD patterns of the samples with 1-5 voids 1.5 mm and pattern of corona and surface PDs.

If a new measurement is taken from a new sample, a feature vector $y \in R^{k\times 1}$ is produced. The vector can be written in terms of the matrix A as

$$y = Ax \tag{6}$$

If $k > p$, equation (6) will be over determined and a unique solution $x \in R^{p\times 1}$ is obtained. If $k < p$, (6) will be underdetermined and there are multiple solutions for x. The least-squares error solution is given by [25]

$$\bar{x} = (A^T A)^{-1} A^T y \tag{7}$$

The least-squares solution $\bar{x}$ may include several small but nonzero entries and does not provide a sparse representation that assigns the test sample y to a specific class.

Determining the sparsest solution requires finding the nonzero entries of x in equation (6) which is equivalent to solving the ℓ^0–norm minimization

$$\tilde{x}_0 = argmin \|x\|_0 \quad \text{subject to } y = Ax \tag{8}$$

This problem is known to be NP-hard [26], i.e. computationally intractable. However, it has been shown that,

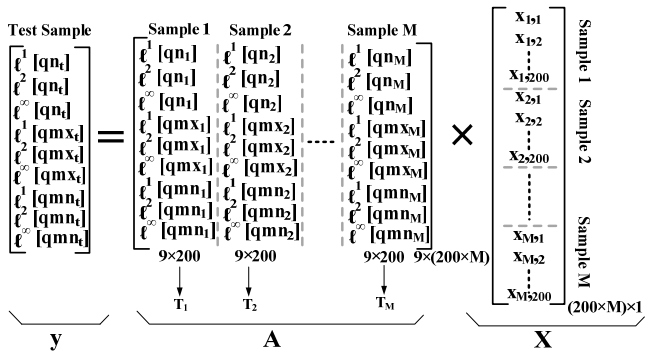


Figure 3. Equation system of SRC

with high probability and under mild conditions, the solution of equation (8) is the same as that of the ℓ^1-norm minimization provided that a sufficiently sparse solution for x exists [27].

$$\tilde{x}_1 = argmin \|x\|_1 \quad \text{subject to } y = Ax \tag{9}$$

In practice, our measurements usually include significant measurement noise which makes the data impractical to use equation (9). Instead, we rewrite equation (9) as the following stable ℓ^1-norm minimization subject to a ℓ^2-norm inequality constraint [24]

$$\tilde{x}_1 = argmin \|x\|_1 \quad \text{subject to } \|Ax - y\|_2 \leq \varepsilon \tag{10}$$

where the noise energy is restricted by ε. Although several algorithms are available to solve equations (9) and (10) [28-35], most are not suited to our problem. Some algorithms require an orthogonal basis set and others require knowledge of the sparsity of the vector x; neither of which is available for our problem. Hence, we use the PDIP algorithm to solve equation (9) and the BPDN algorithm to solve equation (10).

After recovering $\tilde{x}_1$, nonzero terms related to several classes may appear in the entries which are not associated with feature vectors of one specific class to which the test sample belongs. The following residual function can be considered to minimize the ℓ^2-distance between the actual signal feature vector and the recovered

$$r_i = \|y - A\tilde{x}_1(i)\|_2 \ , i = 1, \dots, M \tag{11}$$

where $\tilde{x}_1(i) \in R^{p\times 1}$ is equal to $\tilde{x}_1$ vector but with all entries other than the n_i entries corresponding to the i^{th} class set equal to zero. The residual of (11) must be calculated for all M classes and the test sample y is assigned to the class with the minimum residual.

As discussed earlier in our case study, a 9×250 matrix (F_i) is extracted for each sample. We consider 200 feature vectors for training and 50 vectors for testing. Hence, a training matrix (T_i) is a 9×200 matrix, $k = 9$ and $n_i = 200$ in equation (4) for all samples. Furthermore, a 9×50 test matrix (W_i) is defined for each sample. The elements of y, x, and A are described in Figure 3. To evaluate the accuracy of the proposed algorithm, a misclassification matrix is calculated. When we have M classes in our study, we create a matrix $C \in R^{M\times M}$ whose

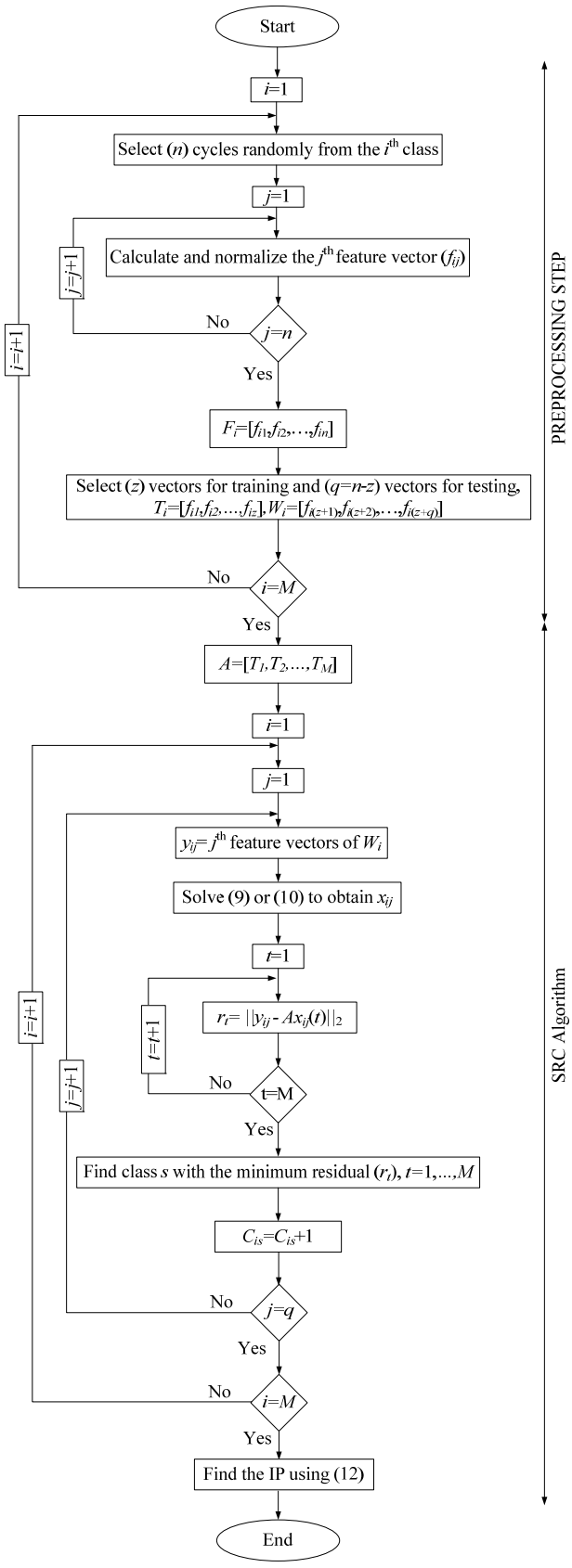


Figure 4. Preprocessing and SRC algorithm

diagonal entries are equal to the number of test samples (q) in the ideal condition where all test samples are identified correctly. When a test sample is selected from the i^{th} class, it

is associated with class s if that is the class that corresponds to the minimum residual in equation (11). This yields the array C_{is} for calculating the Identification Percentage (IP)

$$IP = (\frac{1}{M}) \sum_{i=1}^{M} \frac{C_{ii}}{q} \quad (12)$$

The preprocessing and pattern recognition algorithm is shown in Figure 4.

4 ℓ^1 -NORM MINIMIZATION

ℓ^1–norm minimization has applications in geophysics, data compression, image processing, sensor networks, and computer vision. Many optimization algorithms solve the ℓ^1-norm minimization problem such as: linear programming (LP), greedy algorithm, Gradient Projection (GP), Homotopy, Iterative Shrinkage-Thresholding (IST), Proximal Gradient (PG), Augmented Lagrange Multiplier (ALM), subspace pursuit, CoSaMP, and Bregman iterative algorithm, to name a few [28]. However, these methods provide different performance in terms of computational time and accuracy. We evaluated the performance of the following methods in PDs pattern recognition: CoSaMP [29], block-based recovery [30], Yall1 Basic [31], FPC_AC (fixed-point continuation and active set) [32-33], De-biasing [33], Primal-Dual interior point (PDIP) [34], and Basis Pursuit De-noise (BPDN) [35].

Our results show that PDIP and BPDN algorithms result in better outcomes for solving equations (9) and (10), respectively. Furthermore, the LP method is compared with PDIP and BPDN methods. Finally, a Feed Forward Back Propagation Neural Network (FFBP-NN) is trained and compared with the other three SRC methods. The ℓ^1-approximation of equation (9) is solved using LP as follows:

$$Minimize\ \tilde{c}^T \tilde{x} \quad subject\ to\ \tilde{A}\tilde{x} < \tilde{y} \quad (13)$$

where

$$\tilde{x} = \begin{bmatrix} x \\ y \end{bmatrix},\ \tilde{c} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ \tilde{A} = \begin{bmatrix} A & -I \\ -A & -I \end{bmatrix},\ \tilde{y} = \begin{bmatrix} y \\ -y \end{bmatrix} \quad (14)$$

5 SCENARIOS, IMPLEMENTION, AND RESULTS

We consider PDs pattern recognition for samples with one to five internal voids, corona, and surface discharge. The samples are considered in the following four scenarios:

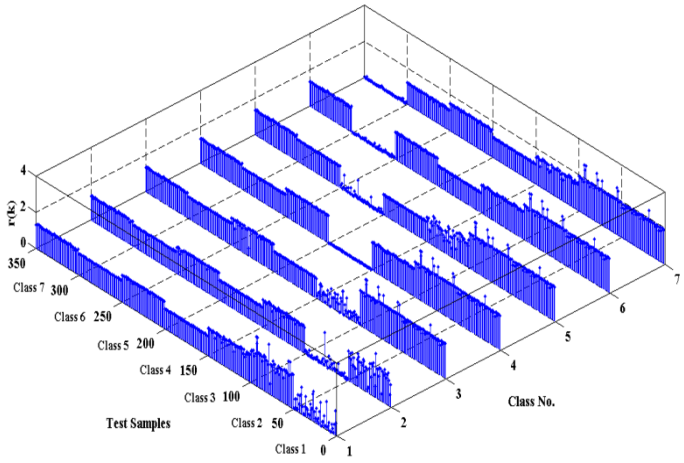


Figure 5. Residual values for scenario 3 by ℓ^1-norm minimization.

- **Scenario 1 (7 samples):** samples 1,2,…,5 include PDs of 1,2,…,5 internal voids 1 mm, sample 6 is corona, and sample 7 is surface discharge.
- **Scenario 2 (7 samples):** samples 1,2,…,5 include PDs of 1,2,…,5 internal voids 1.5 mm, sample 6 is corona, and sample 7 is surface discharge.
- **Scenario 3 (7 samples):** samples 1,2,…,5 include PDs of 1,2,…,5 internal voids 2 mm, sample 6 is corona, and sample 7 is surface discharge.
- **Scenario 4 (17 samples):** samples 1,2,…,15 include PDs of 1,2,…,5 internal voids 1 mm, 1,2,…,5 internal voids 1.5 mm, and 1,2,…,5 internal voids 2 mm, sample 16 is corona, and sample 17 is surface discharge.

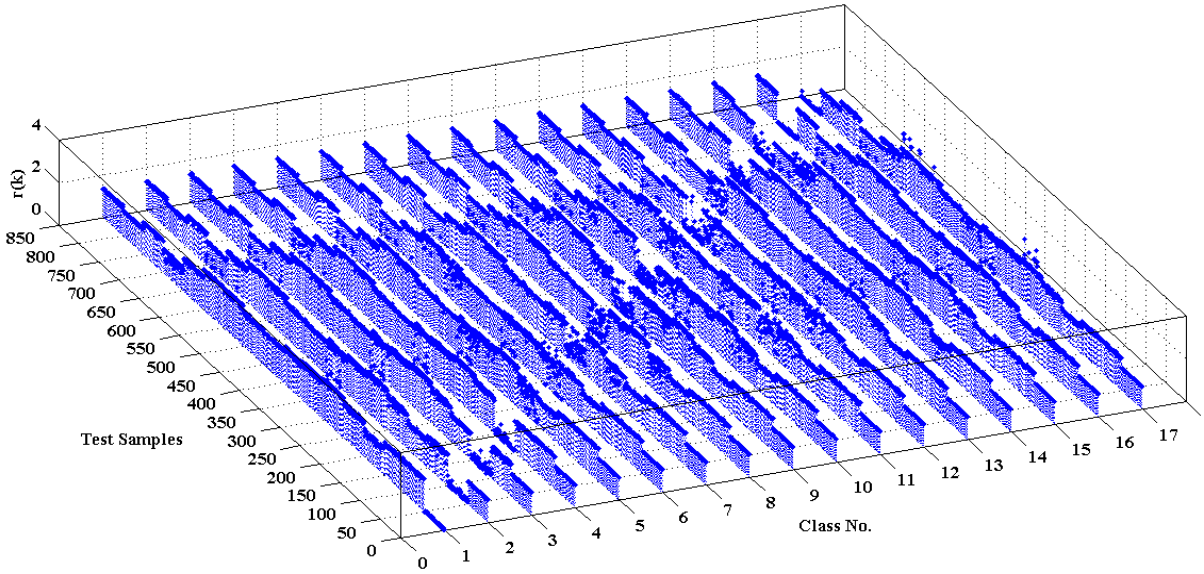


Figure 6. Residual values for scenario 4 by ℓ^1-norm minimization.

Table 2. Misclassification Matrices of scenarios 1-3 by solving the ℓ^1-norm problem.

Scenario 1	IP (%)		**99.7**						
	Output Class	1	50	0	0	0	0	0	0
		2	0	50	0	0	0	0	0
		3	0	1	49	0	0	0	0
		4	0	0	0	50	0	0	0
		5	0	0	0	0	50	0	0
		6	0	0	0	0	0	50	0
		7	0	0	0	0	0	0	50
Scenario 2	IP (%)		**92.9**						
	Output Class	1	48	2	0	0	0	0	0
		2	5	43	0	0	2	0	0
		3	0	5	44	0	1	0	0
		4	0	0	0	40	10	0	0
		5	0	0	0	0	50	0	0
		6	0	0	0	0	0	50	0
		7	0	0	0	0	0	0	50
Scenario 3	IP (%)		**94.0**						
	Output Class	1	41	9	0	0	0	0	0
		2	1	49	0	0	0	0	0
		3	0	0	40	0	10	0	0
		4	0	0	0	50	0	0	0
		5	0	0	1	0	49	0	0
		6	0	0	0	0	0	50	0
		7	0	0	0	0	0	0	50
			1	2	3	4	5	6	7
			Target Class						

Table 3. Misclassification Matrices of scenarios 1-3 by solving the stable ℓ^1-norm problem.

Scenario 1	IP (%)		**97.1**						
	Output Class	1	50	0	0	0	0	0	0
		2	0	43	4	0	3	0	0
		3	0	1	47	0	2	0	0
		4	0	0	0	50	0	0	0
		5	0	0	0	0	50	0	0
		6	0	0	0	0	0	50	0
		7	0	0	0	0	0	0	50
Scenario 2	IP (%)		**80.0**						
	Output Class	1	50	0	0	0	0	0	0
		2	27	18	0	5	0	0	0
		3	1	2	45	2	0	0	0
		4	0	0	0	50	0	0	0
		5	4	1	0	12	33	0	0
		6	0	0	0	0	0	50	0
		7	0	0	0	16	0	0	34
Scenario 3	IP (%)		**82.6**						
	Output Class	1	47	3	0	0	0	0	0
		2	28	22	0	0	0	0	0
		3	2	0	30	0	17	0	1
		4	0	0	0	50	0	0	0
		5	0	0	8	0	42	0	0
		6	0	0	0	0	0	50	0
		7	1	0	1	0	0	0	48
			1	2	3	4	5	6	7
			Target Class						

Table 4. Misclassification Matrix of scenarios 4 by solving ℓ^1-norm problem

IP (%)		**81.6**															
Output Class	1	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	49	0	0	0	0	0	0	0	0	1	0	0	0
	5	0	0	0	0	23	0	8	12	0	2	3	1	0	0	1	0	0
	6	0	0	0	0	0	16	0	0	0	1	19	14	0	0	0	0	0
	7	0	0	0	0	0	0	40	0	2	1	3	4	0	0	0	0	0
	8	0	0	0	0	9	0	1	37	0	0	2	1	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	42	8	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	49	1	0	0	0	0	0	0
	11	0	0	0	0	0	9	0	0	0	1	20	20	0	0	0	0	0
	12	0	0	0	0	0	3	0	0	0	0	1	46	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	2	6	0	0	25	0	15	0	2
	14	1	0	0	1	0	0	0	0	0	0	0	0	0	48	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	1	0	49	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		Target Class																

Table 5. Misclassification Matrices of all scenarios for three main PD types.

IP (%)		**100**			**99.9**		
Target Class	Internal	250	0	0	748	0	0
	Corona	0	50	0	0	50	0
	Surface	0	0	50	2	0	50
		Internal	Corona	Surface	Internal	Corona	Surface
		Scenarios 1-3			Scenario 4		
		Output Class					

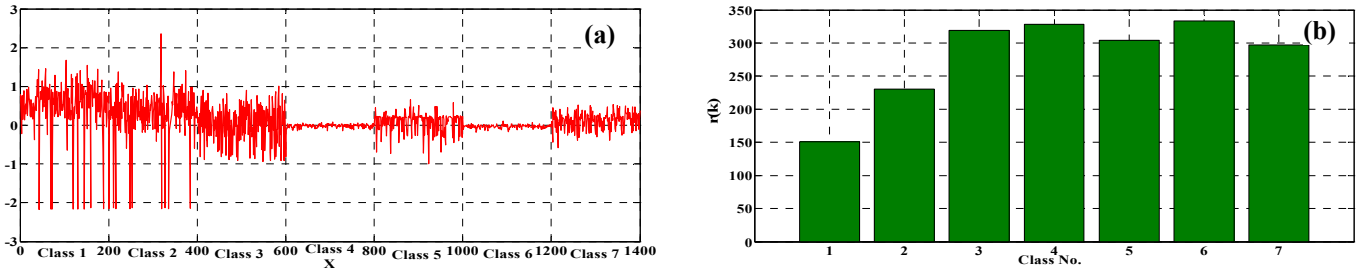


Figure 7. (a) Sparse Representation vector, (b) residual values when the ℓ^1-norm problem is solved by LP

SRC can be implemented using different ℓ^1–norm minimization algorithms. The results of different algorithms are presented in this paper for all scenarios. For example, the residual values extracted by ℓ^1-norm minimization in scenarios 3 and 4 are shown in Figures 5 and 6, respectively.

Scenarios 3 and 4 include seven and seventeen samples, respectively, and each sample contains 50 test samples. Each test sample is a 9×1 feature vector extracted from the raw measured data in one cycle by the preprocessing method. In scenario 3, the residual values corresponding to the correct

class are smaller for most test samples; hence, pattern recognition is satisfactory. In scenario 4, the minimum residual values of most test samples of some classes especially classes 5, 6, 11, and 13, correspond to the wrong class and the samples are misclassified. The misclassification matrices and IPs for all four scenarios calculated by ℓ^1-norm minimization are presented in Tables 2 and 4. In scenarios 1-3, PDs pattern recognition is excellent particularly in scenario 1. Only one misclassification occurs for case 3 where the test sample 3 is incorrectly assigned to class 2. Table 5 shows the misclassifications matrices in all scenarios for the most important PDs types; namely, internal, corona and surface PDs. As shown, all 50 test samples from corona and surface classes are recognized correctly in all scenarios. In scenarios 1-3, all 250 test samples from internal PDs of 5 test objects are identified without any misclassification. In scenario 4, only 2 of 750 test samples from internal PDs of 15 test objects are diagnosed wrongly as the surface PDs. Therefore, 100 and 99.9% IPs show that our method successfully identifies the three main internal, surface, and corona PDs.

In scenario 4 the large number of classes increases the number of misclassifications. This is also true when solving the stable ℓ^1- norm minimization as seen from Table 3. It is interesting to mention that this method produces an unacceptable 50.6% separation in scenario 4. The PDIP method adopted to solve the ℓ^1- norm minimization performs appropriately as long as the data are noise free, while the BPDN method is able to solve the stable ℓ^1- norm minimization when the test samples are affected by measurement noises. Since our measurements were performed in a low noise-level condition, the performance of PDIP is better than that of BPDN. Based on the results, the performance of SRC deteriorates when the sparse representation vector is calculated by minimizing the stable ℓ^1-norm in equation (10). Because the number of nonzero terms in the SR vector increases, the SRC is unable to allocate the test samples to their respective classes. Equations (9) and (10) were solved by a PDIP and a BPDN algorithm, respectively. Similarly, the algorithm of Figure 4 was simulated again using equations (13) and (14) to solve the ℓ^1-problem by LP. With LP, most entries of sparse vectors become nonzero and the residual function must be calculated to recognize the correct class. However, PDIP yields a sparse vector with many zero factors such that residual function calculation is unnecessary in most cases and nonzero factors are aggregated in the sections relevant to one class. For example, Figures 7 and 8 show the sparse representation vector and residual values extracted by LP and PDIP in scenario 3 for one test sample related to one 2 mm void (Class 1). Both algorithms are used to solve the ℓ^1-norm problem given by (9). Based on Figure 7, when LP solves the ℓ^1-norm problem, the nonzero elements in the sparse representation vector increase misclassification and worsen recognition performance. The IPs achieved by LP are presented in Table 6. Table 7 shows a comparison between IPs reported in tables 2-6. The performance of LP is weaker than that of PDIP in all four scenarios, especially in scenario 4. On the other hand, LP performs like BPDN in scenario 1 and even better in scenario 4, and much weaker than BPDN in scenarios 2 and 3.

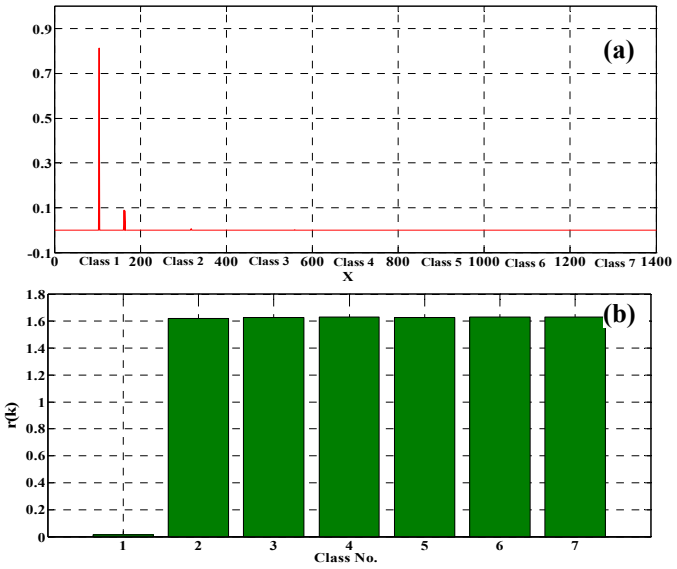


Figure 8. (a) Sparse Representation vector, (b) residual values when the ℓ^1-norm problem is solved by PDIP.

Table 6. IPs in all scenarios by solving the ℓ^1-norm problem with LP.

Scenario No.	1	2	3	4
IPs	97.1	76.7	79.4	57.4

6 COMPARISON OF RESULTS

A FFBP-NN Multilayer Preceptor (MLP) was used to identify PD classes in all 4 scenarios. The network has three layers; namely, the input, output and hidden layers. Neurons in the input layer feed the input signals to the neurons of the hidden layer. The output of each neuron is a linear, hyperbolic tangent sigmoid, logistic sigmoid, or a Gaussian radial basis function of a weighted sum of its inputs. The neurons of the output layer are similar to those of the hidden layer. The MLP is usually trained using the BP algorithm. BP iteratively changes the weights of the connections between input and hidden neurons with a specified learning rate [36]. The training matrices (T_i) and test matrices (W_i), which are extracted in the preprocessing step in Figure 4, are the input of the ANN. There are many options for training, learning and transfer function selection for FFBP networks. Extensive simulations showed that a combination of Levenberg-Marquardt training, gradient descent weight and bias learning, mean-square normalized error performance, and logistic sigmoid transfer function leads to better results.

During training, the initial weights of the ANN were selected randomly and the results changed in each simulation run of every scenario. Therefore, we repeated the simulations to obtain the best performance and summarized the results in Table 7. The IPs achieved by ℓ^1-norm minimization with PDIP (method 1) are higher than other methods in scenarios 1 and 3 with 99.7% and 94.0%, respectively. For scenarios 2 and 4, the recognition performance of method 1 is better than all but method 4 (ANN). The IP differences between the two methods in scenarios 2 and 4 are less than 3%. However, the

performance of method 1 in scenario 1 is 9% better than method 4.

Table 7. IPs in all scenarios with all investigated methods.

No	Method	IP (%)			
		Scenario 1	Scenario 2	Scenario 3	Scenario 4
1	ℓ^1-norm minimization by PDIP method	99.7	92.9	94.0	81.6
2	ℓ^1-norm minimization by LP method	97.1	76.7	79.4	57.4
3	Stable ℓ^1-norm minimization by BPDN	97.1	80.0	82.6	50.6
4	ANN	90.8	95.9	93.6	82.6

7 CONCLUSIONS

This paper presents a new SRC classifier for PD pattern recognition and applies it to 17 PD data sets. SRC simulation requires ℓ^1-norm minimization. Examining several compressive sensing methods showed that the PDIP and BPDN algorithms provide better performance with the extracted features and are more computationally efficient. The stable ℓ^1-norm minimization does not perform as well as ℓ^1-norm minimization due to the low noise level in our data. Also, the performance of ℓ^1-approximation via LP is similar to BPDN and worse than PDIP. A FFBP neural network was also trained to recognize PD patterns. By minimizing the ℓ^1-norm with PDIP, SRC gives similar IP results to FFBP, and much better accuracy than FFBP in one case. When the number of classes is increased in scenario 4, recognition accuracy drops for both ANN and SRC. Although misclassification of SRC with ℓ^1-norm minimization by PDIP in scenarios 2 and 4 is slightly lower than ANN, SRC has other valuable advantages. It is simpler to implement than ANN and does not require extensive training to tune multiple parameters such as selection of the training, learning, error functions, and number of neurons, for finding the best results. In addition, SRC always converges to the same results as long as the test and training matrices which are selected from the original 3000 samples remain the same, whereas the results of the ANN depends on the initial random weights used for training. Note that, we found the best result for the ANN by simulating each case several times with different combinations of tuning options. However, the SRC technique described in this paper finds the best outcome with no treed for the selection of various training options. The ANN can complement the SRC if the additional effort for its training to obtain the best result is not an issue. It is also possible to use the compact representation obtained with SRC to train a neural network or support vector machine. Such a combination will be the subject of future research.

8 REFERENCES

[1] R. Liao, Y. Fernandess, K. Tavernier, G. A. Taylor, and M. R. Irving "Recognition of Partial Discharge Patterns", IEEE Power and Energy Society General Meeting, pp.1-8, 2012.

[2] J. A. Ardila-Rey, J. M. Martinez-Tarifa, G. Robles and M. V. Rojas-Moreno, "Partial Discharge and Noise Separation by Means of Spectral-power Clustering Techniques" IEEE Trans. Dielectr. Electr. Insul., Vol. 20, No. 4, pp. 1436-1443, 2013.

[3] H. Ma, J.C. Chan, T.K. Saha, and C.Ekanayake," Pattern Recognition Techniques and Their Applications for Automatic Classification of Artificial Partial Discharge Sources", IEEE Trans. Dielectr. Electr. Insul., Vol.20, No.2, pp. 468-478, 2013.

[4] N. C. Sahoo, M. M. A. Salama, and R. Bartnikas, "Trends in Partial Discharge Pattern Classification: A Survey", IEEE Trans. Dielectr. Electr. Insul., Vol.12, No.2, pp. 248-264, 2005.

[5] R. Bartnikas," Partial Discharges: Their Mechanism, Detection and Measurement", IEEE Trans. Dielectr. Electr. Insul., Vol. 9, No.5, pp.763-808, 2002.

[6] E. Gulski and A. Krivda, "Neural network as a tool for recognition of partial discharges", IEEE Trans. Electr. Insul., Vol. 28, No.6, pp.984-1001, 1993.

[7] A. Krivda, "Automated Recognition of Partial Discharges", IEEE Trans. Dielectr. Electr. Insul., Vol. 2, No.5, pp.796-821, 1995.

[8] E.Gulski, "Digital Analysis of Partial Discharge", IEEE Trans. Dielectr. Electr. Insul., Vol. 2, pp.822-837, 1995.

[9] L. Satish and W.S. Zaengl,"Can fractal features be Used for recognizing 3-D partial Discharge Patterns", IEEE Trans. Dielectr. Electr. Insul., Vol. 2, No.3, pp. 352-359, 1995.

[10] A. Krivda and D. Birtwhistle, "Recognition of Multiple Partial Discharge Patterns", 11th Int'l. Sympos. High Voltage Eng. (Conf. Publ. No. 467), Vol. 5, pp. 17-20, 1999.

[11] R. Bartnikas, ''Note on Multichannel Corona Pulse-height Analysis'', IEEE Trans. Electr. Insul., Vol. 8, pp. 2-5, 1973.

[12] A. Kelen, ''The Functional Testing of HV Generator Stator Insulation'', Proc. CIGRE, Paris, France, Paper 15-03, 1976.

[13] I. C. Bapt, Bui Ai, and C. Mayoux, "Coronal Frequency Analysis in Artificial Cavities in Epoxy Resins", IEEE Conf. Electr. Insul. Dielectr. Phenomena, Washington, D.C., USA, pp. 282-288, 1974.

[14] H.-G. Kranz, ''Diagnosis of Partial Discharge Signals Using Neural Networks and Minimum Distance Classification'', IEEE Trans. Electr. Insul., Vol. 28, pp. 1016-1024, 1993.

[15] B. A. Lloyd, S. R. Campbell and G. C. Stone, ''Continuous online PD Monitoring of Generator Stator Windings", IEEE Trans. Energy Conversion, Vol. 14, pp. 1131-1138, 1999.

[16] R. E. James and B. T. Phung, ''Development of Computer-based Measurements and their Application to PD Pattern Analysis", IEEE Trans. Dielectr. Electr. Insul., Vol. 2, pp. 838-856, 1995.

[17] R. Ambikairajah, B. T. Phung, J. Ravishankar and T.Blackburn, "Spectral features for the classification of partial discharge signals from selected insulation defect models", IET Sci., Measurement and Techn., Vol. 7, No. 2, pp. 104–111, 2013.

[18] K.X. Lai, B.T. Phung ,T.R. Bluckburn, "Partial Discharge analysis using PCA and SOM", IEEE Lausanne Power Tech, pp.2133-2138, 2007.

[19] R. Candela, G. Mirelli and R. Schifani, ''PD Recognition by Means of Statistical and Fractal Parameters and a Neural Network'', IEEE Trans. Dielectr. Electr. Insul., Vol. 7, pp. 87-94, 2000.

[20] A. A. Mazroua, R. Bartnikas and M. M. A. Salama, ''Neural Network System Using the Multi-layer Perceptron Technique for the Recognition of PD Pulse Shapes due to Cavities and Electrical Trees'', IEEE Trans. Power Delivery, Vol. 10, pp. 92-96, 1995.

[21] http://www.doble-lemke.eu/en/Products/Partial_Discharge.html

[22] IEC 60270, High-voltage test techniques, partial discharge measurement, third edition, 2000.

[23] M.Oskuoee, A.R.Yazdizadeh, H.R.Mahdiani, "A new feature extraction and pattern recognition of partial discharge in solid Material by Neural network", 8th Int'l. Conf. Natural Computation (ICNC), pp.183-187, 2012.

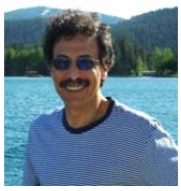
[24] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry and Y.Ma, "Robust face recognition via sparse representation", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.31, No. 2, pp. 210–227, 2009.

[25] R.G. Brown and P.Y.C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, Fourth Edition, John Wiley & Sons, Inc. 2012.

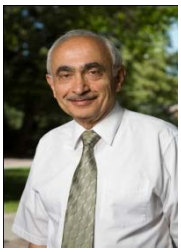
[26] E. Amaldi and V. Kann, "On the Approximability of Minimizing Nonzero Variables or Unsatisfied Relations in Linear Systems," Theoretical Computer Science, Vol. 209, pp. 237-260, 1998.

[27] E. Candes and T. Tao, "Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?" IEEE Trans. Information Theory, Vol. 52, No. 12, pp. 5406-5425, 2006.
[28] A. Y. Yang, A. Ganesh, Z. Zhou, S. S. Sastry, and Y. Ma, "A Review of Fast l^1-Minimization Algorithms for Robust Face Recognition," arXiv:1007.3753, 2010. http://ecovision.mit.edu/~sai/12S990/Yang10-SIAM.pdf.
[29] D. Needell, and J. A. Tropp "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples" Applied and Computational Harmonic Analysis, Vol. 26, No.3, pp.301–321, 2009.
[30] R.G. Baraniuk, V. Cevher, and M.F. Duarte, C.Hegde "Model-Based Compressive Sensing" IEEE Trans. Information Theory, Vol.56, No. 4, pp.1982- 2001, 2010.
[31] YALL1 basic solver code: Y. Zhang, J. Yang, and W. Yin. YALL1: Your Algorithms for L1, online at yall1.blogs.rice.edu, 2011.
[32] Z. Wen, W. Yin, D. Goldfarb, and Y. Zhang "A Fast Algorithm for Sparse Reconstruction Based on Shrinkage, Subspace Optimization, and Continuation" SIAM J. Sci. Comput., Vol.32, No.4, pp.1832–1857, 2010.
[33] E. T. Hale, W. Yin, and Y. Zhang" A Fixed-Point Continuation Method for ℓ_1-Regularized Minimization with Applications to Compressed Sensing" CAAM Technical Report TR07-07, Department of Computational and Applied Mathematics Rice University, Houston, Texas, 77005, U.S.A, 2007.
[34] E. Candes and J. Romberg, "1-Magic: Recovery of Sparse Signals via Convex Programming," http://www.acm.caltech. edu/l1magic/, 2005.
[35] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions", SIAM J. Scientific Computing, Vol.31, No.2, pp.890-912, 2008.
[36] A. A. Mazroua, M. M. A. Salama and R. Bartnikas, ''PD Pattern Recognition with Neural Networks Using the Multilayer Perceptron Technique", IEEE Trans. Electr. Insul., Vol. 28, pp. 1082-1089, 1993.

Mehrdad Majidi received the B.Sc. and M.Sc. degrees in power electrical engineering from Power and Water University of Technology (PWUT), Tehran, Iran with honors in 2009 and 2011, respectively. He is currently a Graduate Research Assistant in the Dept. of Electrical and Biomedical Engineering at University of Nevada, Reno (UNR), and is pursuing the Ph.D. degree. His research interest is pattern recognition, high voltage engineering, smart grids, power system automation, fault location, and applications of signal processing in power system.



Mohammed Sami Fadali (SM'91) earned a B.S. degree in electrical engineering from Cairo University in 1974, an M.S. degree from the Control Systems Center, UMIST, England, in 1977 and a Ph.D. degree from the University of Wyoming in 1980. He was an Assistant Professor of Electrical Engineering at the University of King Abdul Aziz in Jeddah , Saudi Arabia 1981-1983. From 1983-85, he was a Post Doctoral Fellow at Colorado State University. In 1985, he joined the Electrical Engineering Dept. at the University of Nevada, Reno, where he is currently Professor of Electrical Engineering. In 1994 he was a visiting professor at Oakland University and GM Research and Development Labs. He spent the summer of 2000 as a Senior Engineer at TRW, San Bernardino. His research interests are in the areas of fuzzy logic stability and control, state estimation and fault detection.



Mehdi Etezadi-Amoli (LSM'11) received the Ph.D. degree from New Mexico State University, Las Cruces, NM, USA, in 1974. From 1975 to 1979, he worked as an Assistant Professor of electrical engineering at New Mexico State and the University of New Mexico, Las Cruces, NM, USA. From 1979 to 1983, he worked as a Senior Protection Engineer at Arizona Public Service Company, Phoenix, AZ, USA. In 1983, he joined the faculty of the Electrical Engineering Department at the University of Nevada, Reno, NV, USA. His current interest is in large-scale systems, power system distribution and protection, distributed generation, and renewable energy. Dr. Etezadi is a registered Professional Engineer in Nevada.

Mohammad Oskuoee received the B.Sc. degree in power electrical engineering from Iran University of Science and Technology (IUST) in 1991 and the M.Sc. degree in control engineering from Shahid Beheshti University (SBU) in 2011. He was chair of the high voltage engineering department in Niroo Research Institute (NRI) from 1993-2011 and is currently the director of control research center in NRI. His research interests are in the field of high voltage engineering and transmission substations.