This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

Fault Classification for Transmission Lines Based on Group Sparse Representation

Shenxing Shi, *Senior Member, IEEE,* Beier Zhu, *Student Member, IEEE,* Sohrab Mirsaeidi, *Member, IEEE,* and Xinzhou Dong, *Fellow, IEEE*

***Abstract*—Fault classification is an important aspect of the protective relaying system for transmission lines. This paper proposes a new method based on group sparse representation for fault classification in transmission lines, in which half-cycle superimposed current signals are measured for the classification task. Compared with the conventional feature extraction methods, the proposed method in this paper alleviates the requirement to manually design feature. Signals are factorized over an over-complete basis, in which elements are the fault signals themselves. The algorithm of classification is based on the idea that the training samples of a particular fault type approximately form a linear basis for any test sample belonging to that class. Solved by $l^{2,1}$-minimization, the coefficient should be group sparse, and its non-zero entries correspond to particular group of correlated training samples. It is illustrated that the proposed classification method can be properly modified to deal with noise-containing signals. Moreover, dimension reduction is performed using random mapping technique. The results of several simulations carried out by PSCAD/EMTDC and field data in real system indicate that the proposed method is accurate and fast for fault classification, and has a high robustness to noise.**

***Index Terms*—fault classification, transmission lines, sparse representation, group sparse representation, compressed sensing, $l^{2,1}$-minimization.**

I. Introduction

FAULT classification plays an important role in the protection of a transmission line. Accurate and rapid fault classification can greatly facilitate fault location and fault isolation, thus reducing potential dangers in power system.

In recent years, various fault classification methods have been proposed in the technical literature. Feature extraction is the first step in fault classification which reveals the nature of fault signals. Some feature extraction methods such as Discrete Fourier Transform (DFT) [1]–[3] and Discrete Wavelet Transform (DWT) [4]–[6] are adopted to obtain the information of waveform signals in the frequency or time-frequency domains. Also, modal transforms such as Clarke transform and Karenbauer transform are applied to decouple a three-phase system to extract the modal values of signals.

After extracting features from waveform signals, the fault type can be classified by setting a logic flow [7]. Other novel methods mainly applied machine learning models such as Artificial Neural Network (ANN) [8] , Support Vector Machine (SVM) [9], fuzzy logic algorithm [2], decision trees [1], [3] and deep learning models [10].

However, application of these methods are accompanied by some challenges: (1) For fault classification methods based on logic flow, each node of the logic flow is basically an if-then condition with preset thresholds, which is obtained by analyzing characteristics of various faults in all possible situations. The setting value is related to a certain power system, which lacks generality. (2) Unlike classification methods based on logic flow, machine learning methods do not need to set the setting value. However, the combination of a certain feature extraction transform and a certain classification model is almost arbitrary, which takes much time to be evaluated by the researchers.

In this paper, an attempt is made to seek a method which can automatically extract features from fault signals without the need for design and prior knowledge. The traditional transform basis is independent from the fault signals. For example, the DFT results can be viewed as coefficient of signal in a finite-dimensional orthogonal basis, which is composed of complex sinusoids, but individual elements in the transform basis are not assumed to have any particular semantic meaning. In this paper, the discriminative nature of sparse representation is exploited to perform feature extraction and classification. Instead of using the data-independent dictionaries discussed before, the fault signal is presented in an over-complete dictionary (basis), in which base elements are the known fault signals themselves. $l^{2,1}$-minimization is adopted to solve the coefficient of signal in that basis. The coefficient is group sparse, which contains discriminative information of faults. The proposed method also alleviates the requirement to design a complex classifier or set thresholds.

This paper is organized as follows: Section II introduces the sparse representation technique; Section III describes the algorithm of fault classification based on group sparse representation; In Section IV, the robustness improvement and dimension reduction of the proposed method are discussed; Section V introduces the test network and reports the results; Finally, Section V concludes the paper.

II. Sparse Representation Technique

Sparse representation technique has achieved much success in the field of signal processing due to its appropriate properties in signal representation and reconstruction. The main idea of the technique is to use only a few elements from the

This work is supported in part by National Natural Science Foundation of China (Grant No.51120175001), in part by the National Key Research and Development Plan of China (Grant No. 2016YFB 0900600), and in part by the State 863 projects (Grant No. 2015AA050102)

The authors are with the State Key Lab of Power Systems, Dept. of Electrical Engineering, Tsinghua University, Beijing 100084, P.R. of China. (e-mail: shishenxing@tsinghua.edu.cn; zbe16@mails.tsinghua.edu.cn; m_sohrab@mail.tsinghua.edu.cn; xzdong@mail.tsinghua.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

dictionary to approximate the original signal. In recent years, sparse representation has widely been used in many fields such as state estimation, face recognition, digital recognition, objection detection and image classification [11]–[16].

However, construction of an effective dictionary is a key issue to sparse representation. Wright et al. [17] proposed a sparse representation method, in which the training samples are directly used to form a dictionary. If sufficient training samples are available from each class, it will be possible to represent the test samples as a linear combination of just those training samples from the same class. Meanwhile, training signals from other classes are not expected to contribute to the representation of the test sample. Therefore, this representation is supposed to be sparse, in which only a small fraction of the entire dictionary would be involved. It is also proposed in [17] to penalize the l^1-norm of the coefficients in the linear combination. However, the sparsity on the number of training samples does not consider the correlation among the training samples. In [18], the label information of training samples is considered; also, the sparsity of the class is emphasized rather than the number of training samples, which leads the group sparsity. With the recent development in the theory of sparse representation and compressed sensing [19]–[21], group sparse representation can be efficiently recovered by $l^{2,1}$-minimization [22].

It is noteworthy that the sparse representation based classification method does not require designing an extra classifier. Ideally, once the sparse solution is computed, all of the nonzero entries in the coefficient are associated with the training samples from a single class, and the test sample can be easily assigned to that class.

III. The Proposed Classification Algorithm

A. Sparse Representation of Fault signals

In order to classify faults, three-phase half-cycle superimposed current signals are first measured; and then, the signals are stacked in the order of phase-a, phase-b and phase-c to form the training and testing samples. Assume that the stacked signal is denoted as $\boldsymbol{v}$ and its length is m. The given n_i training samples from the i-th class were arranged as columns of a matrix $A_i = [\boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, ..., \boldsymbol{v}_{i,n_i}] \in \mathbb{R}^{m\times n_i}$. Given sufficient training samples of the i-th signal class, any test signal $\boldsymbol{y}$ from the same class will approximately lie in the linear span of the training samples associated with signal i.

$$\boldsymbol{y} = \alpha_{i,1}\boldsymbol{v}_{i,1} + \alpha_{i,2}\boldsymbol{v}_{i,2} + ... + \alpha_{i,n_i}\boldsymbol{v}_{i,n_i}, \tag{1}$$

for some scalars, $\alpha_{i,j}, j = 1, 2, ..., n_i$. Denote the training samples of all k classes as the matrix $A = [A_1, A_2, ..., A_k] \in \mathbb{R}^{m\times n}$, where $n = \sum_{i=1}^{k} n_i$. Then the linear representation of a testing signal $\boldsymbol{y}$ can be rewritten in terms of all training samples as:

$$\boldsymbol{y} = A\boldsymbol{x}_0, \tag{2}$$

where $\boldsymbol{x}_0 = [0, ..., 0, \alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,n_i}, 0, ..., 0]^T \in \mathbb{R}^n$ is a sparse vector, in which entries are zeros except those associated with the i-th class.

Besides, the magnitude of fault components is different under different fault conditions, each atom of dictionary A is normalized to have unit l^2-norm, i.e., for an atom $\boldsymbol{v}$, it is normalized to $\boldsymbol{v}/\|\boldsymbol{v}\|_2$.

In sparse representation based classification problems, m is typically smaller than n; Therefore, Equation (2) is unknown and its solution is not unique. However, if the solution is known to be sparse, studies in compressed sampling literature [19] show that the sparse solution is unique as well. A valid test sample $\boldsymbol{y}$ can be sufficiently represented using only the training samples from the same class. This representation is naturally sparse if k is large. Thus, Equation (2) is a linear inverse problem with a unique sparse solution. The coefficient $\boldsymbol{x_0}$ is solved by the following optimization problem:

$$(l^0): \ \hat{\boldsymbol{x}}_0 = \arg\min \|\boldsymbol{x}\|_0 \ s.t. \ A\boldsymbol{x} = \boldsymbol{y}, \tag{3}$$

where $||\boldsymbol{x}||_0$ denotes the l^0-norm, which counts the number of nonzero entries in $\boldsymbol{x}$. Although solving the l^0-minimization problem, which is NP-hard, is theoretically possible, it is impractical since it requires an exhaustive search of all possible combinations. An approximate solution is obtained by replacing the l^0-norm by l^1-norm, as follows:

$$(l^1): \ \hat{\boldsymbol{x}}_1 = \arg\min \|\boldsymbol{x}\|_1 \ s.t. \ A\boldsymbol{x} = \boldsymbol{y}, \tag{4}$$

The theory of compressed sensing proves that if certain conditions on the sparsity is satisfied, i.e., the solution is sparse enough, and the solution of Equation (3) is equivalent to the solution of Equation (4), which can be efficiently solved. To solve the l^1-minimization problem, as reviewed in [23], there are five representative fast approaches: Gradient Projection, Homotopy, Iterative Shrinkage-Thresholding, Proximal Gradient, and Augmented Lagrange Multiplier (ALM).

B. Criterion of Classification: Minimal Reconstruction Error

Ideally, after test signal $\boldsymbol{y}$ is recovered, the nonzero entries in the estimate $\hat{\boldsymbol{x}}_1$ will all be associated with columns of A from a single object class; The test signal $\boldsymbol{y}$ should be labeled to that class. However, some small non-zero entries associated with multiple classes may appear in the coefficient $\boldsymbol{x}$, which is caused by modeling error or noise. Thus, a simple classifier based on the reconstruction error was designed. For each class i, a function $\delta_i(\hat{\boldsymbol{x}}_1)$ retains the non-zero entries in $\boldsymbol{x}$ that associates with class i, and assigns the rest of coefficients to zero. An approximation of test signal $\boldsymbol{y}$ corresponding to i-th class can be represented as $A\delta_i(\hat{\boldsymbol{x}}_1)$. The reconstruction error for i-th class is the Euclidean distance between the test signal $\boldsymbol{y}$ and the approximation signal $A\delta_i(\hat{\boldsymbol{x}}_1)$. Then, the test signal can be classified by assigning it to the fault class that minimizes the reconstruction error:

$$\min r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\delta_i(\hat{\boldsymbol{x}}_1)\|_2 \tag{5}$$

The complete sparse representation based classification method is stated in Algorithm 1.

C. Group Sparse Representation of Fault Signals

In the previous subsection, the fault classification task was cast as the problem of finding the sparsest representation of a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

Algorithm 1 Sparse Representation Based Classification (SRC)

Input: matrix $A = [A_1, A_2, .., A_k] \in \mathbb{R}^{m\times n}$ for k classes, a test signal $\boldsymbol{y} \in \mathbb{R}^m$

1: Normalize the columns of A to have unit l^2-norm.
2: Solve the l^1-minimization problem:
$(l^1): \ \hat{\boldsymbol{x}}_1 = \arg\min \|\boldsymbol{x}\|_1 \ s.t. \ A\boldsymbol{x} = \boldsymbol{y}$,
3: Compute the residuals:
$r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\delta_i(\hat{\boldsymbol{x}}_1)\|_2$ for $i = 1, .., k$

Output: identity$(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$

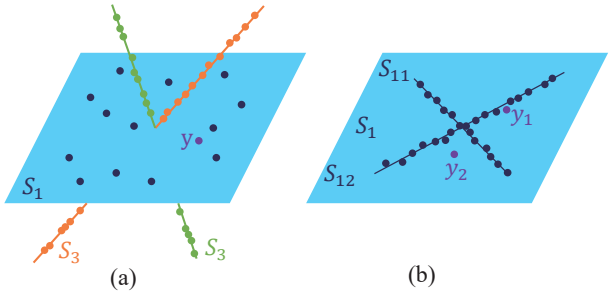


Fig. 1. (a) sparsest representation of a test sample is not necessarily from the correct class. (b) training data might be separated into several group, a test sample can be written as a linear combination of a few groups in each class.

test sample in the dictionary of all training samples. In [24], it is argued that the sparsest representation of a test example might not be the best criterion for the classification. In the classification problems, sometimes the training samples from certain classes are highly correlated, and the sparsest representation might select the samples from all of the correlated classes rather than one single class.

In order to validate this argument, an example is shown in Fig. 1 (a) including three classes which the data are sampled from three subspaces: S_1 being a 2-dimensional subspace, S_2 and S_3 being 1-dimensional subspaces. The data from the three classes are highly correlated. The test sample $\boldsymbol{y}$, which belongs to class 1, can be written as a linear combination of any two points from S_1 (class 1), while it can also be written as linear a combination of one data point from S_2 (class 2) and one from S_3 (class 3). Thus, based upon the criterion of sparsest representation, there is no difference between these two representations because they both have two non-zero entries, while the first one is the desired solution.

In contrast, instead of searching the sparsest solution, looking for a representation that imposes group sparsity on the entire class can obtain the desired solution for classification. In a general classification task, the dictionary of the training data has a group structure where a few groups of the dictionary correspond to the training data in each class. Therefore, a test sample can be represented as a linear combination of training samples from a few blocks of dictionary. For example, in Fig. 1 (a), if we consider the data from the same class to construct a group, then the test data $\boldsymbol{y}$ can be written as a linear combination of group 1 or as linear combination of group 2 and 3. The former has a group sparsity of 1 and the latter's group sparsity is 2. Based upon the criterion of group sparsest representation, the high correlated data can be classified correctly. In Fig. 1 (b), the test example $\boldsymbol{y}_1$ can be written as a linear combination of one group (S_{12}), while $\boldsymbol{y}_2$ can be written as a linear combination of two groups (S_{11} and S_{12}).

In order to consider the constraint of group sparse structure, the group representation based classification is proposed in [23], [24], in which the mixed norm is deployed as the sparsity measure as follows:

$$(l^{2,0}): \ \hat{\boldsymbol{x}}_{2,0} = \arg\min \|\boldsymbol{x}\|_{2,0} \ s.t. \ \|A\boldsymbol{x} - \boldsymbol{y}\|_2 = 0, \quad (6)$$

The mixed norm $\|\cdot\|_{2,0}$ is defined as:

$$\|\boldsymbol{\alpha}\|_{2,0} = \sum_{i=1}^{k} I(\|\boldsymbol{\alpha}_i\|_2 > 0) \quad (7)$$

where

$$\boldsymbol{\alpha} = [\underbrace{\alpha_{1,1}, .., \alpha_{1,n_i}}_{\boldsymbol{\alpha_1}}, ..., \underbrace{\alpha_{k,1}, .., \alpha_{k,n_k}}_{\boldsymbol{\alpha_k}}]^T$$

$\boldsymbol{\alpha}_i = [\alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,n_i}]$

associates the i-th group of the training samples and $I(\|\boldsymbol{\alpha}_i\|_2 > 0) = 1$ if $\|\boldsymbol{\alpha}_i\|_2 > 0$.

The $l^{2,0}$-minimization is an NP-hard problem. An approximation solution is given by:

$$(l^{2,1}): \ \hat{\boldsymbol{x}}_{2,1} = \arg\min \|\boldsymbol{x}\|_{2,1} \ s.t. \ \|A\boldsymbol{x} - \boldsymbol{y}\|_2 = 0, \quad (8)$$

where the mixed norm $\|\cdot\|_{2,1}$ is defined as:

$$\|\boldsymbol{\alpha}\|_{2,1} = \sum_{i=1}^{k} \|\boldsymbol{\alpha}_i\|_2, \quad (9)$$

There exist several algorithms for solving the $l^{2,1}$-minimization problem. In this paper, the Alternating Direction Method (ADM) algorithm [22] is adopted, which has high efficiency, and strong stability and robustness.

In the scenario of fault classification, the training samples are divided into $k = 10$ groups corresponding to their fault type, i.e., Single-Line-to-Ground (a-g, b-g, c-g), Line-to-Line-to-Ground (ab-g, bc-g, ac-g), Line-to-Line (ab, ac, bc) Three-Phase-to-Ground (abc-g) faults. Theoretically, because of symmetry, Three-Phase faults and Three-Phase-to-Ground faults are equivalent. Moreover, in the field of relay protection, the protection strategy of these two kinds of faults are the same. Therefore, we consider the two kinds of faults as the same faults.

Fig. 2 illustrates the proposed fault classification method based on group sparse representation. In Fig.2, three-phase half-cycle superimposed current signals are measured for the classification. The sampling frequency is set to 10 kHz. All signals are stacked in the order of phase-a (time range [1, 100]), phase-b (time range [101, 200]) and phase-c (time range [201, 300]). The proposed method represents a c-g fault signal in an over-complete dictionary, which is composed of 10 types of fault signals. Under the constraint of group sparsity, the group sparse coefficient was obtained. The test signal can be viewed as a sparse linear combination of all

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

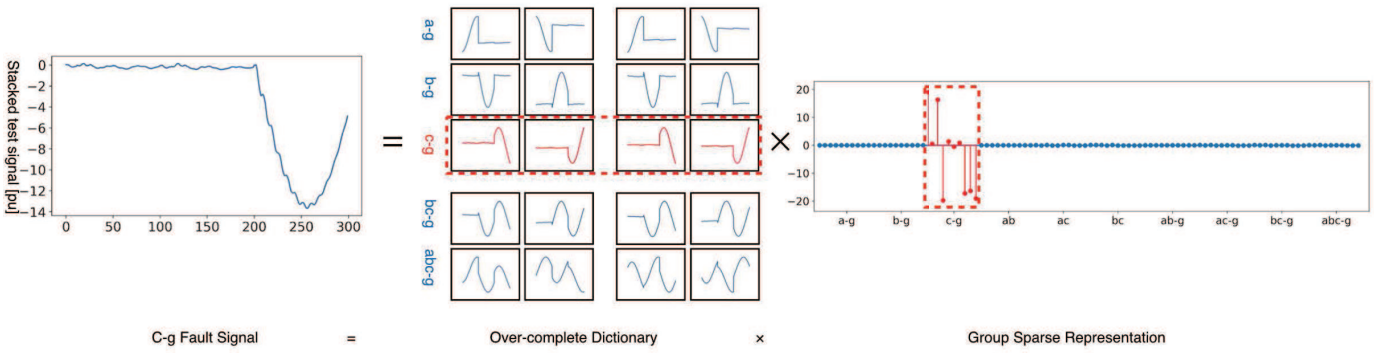


Fig. 2. Overview of our approach.

training samples, the red coefficients are associated with the correct fault class (c-g fault). The classification algorithm is summarized in Algorithm 2.

Algorithm 2 Group Sparse Representation Classification (GSRC)

Input: matrix $A = [A_1, A_2, .., A_k] \in \mathbb{R}^{m\times n}$ for k classes, a test signal $\boldsymbol{y} \in \mathbb{R}^m$

1: Normalize the columns of A to have unit l^2-norm.
2: Solve the $l^{2,1}$-minimization problem:
$(l^{2,1}):\ \hat{\boldsymbol{x}}_{2,1} = \arg\min \|\boldsymbol{x}\|_{2,1}\ s.t.\ \|A\boldsymbol{x} - \boldsymbol{y}\|_2 = 0$
3: Compute the residuals $r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\delta_i(\hat{\boldsymbol{x}}_{2,1})\|_2$ for $i = 1, .., k$

Output: identity$(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$

IV. Robustness Improvement and Dimensionality Reduction

A. Robustness to Noise

In practical fault classification scenarios, real signals in electrical power system are often contaminated with noise, and hence, modeling error is also considered in real signals. Equation (2) can be modified as

$$\boldsymbol{y} = A\boldsymbol{x}_0 + \boldsymbol{z}, \tag{10}$$

where $\boldsymbol{z} \in \mathbb{R}^m$ is a noise term with bounded energy $\|\boldsymbol{z}\|_2 \leq \varepsilon$. The sparse solution $\boldsymbol{x}_0$ can be approximately recovered by solving:

$$(l^{2,1}_{\varepsilon}):\ \hat{\boldsymbol{x}}_{2,1} = \arg\min \|\boldsymbol{x}\|_{2,1}\ s.t.\ \|A\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \varepsilon, \tag{11}$$

This simple method is illustrated in Fig. 3. In this paper, the error tolerance ε is set to 0.05.

The above-mentioned denoising technique is designed for small noise. Sometimes, the errors may have great magnitude which cannot be treated using Equation (11). Hence, the need to deal with high levels of noise is also taken into account in the proposed group sparse representation. Considering $\boldsymbol{y}$ as a test sample corrupted by an error $\boldsymbol{e}$ and $\boldsymbol{y_0}$ as the uncorrupted sample, i.e., $\boldsymbol{y} = \boldsymbol{y_0} + \boldsymbol{e}$, the linear model of Equation (2) can be modified as:

$$\boldsymbol{y} = \boldsymbol{y}_0 + \boldsymbol{e} = A\boldsymbol{x}_0 + \boldsymbol{e} \tag{12}$$

where $\boldsymbol{e} \in \mathbb{R}^m$ is a vector of errors. The non-zero entries of $\boldsymbol{e}$ correspond with the noise in $\boldsymbol{y}$. The uncorrupted signal $\boldsymbol{y_0}$ can be written as a linear combination of groups from the training sample dictionary. Also, the vector $\boldsymbol{e}$ is associated with a new single group, $\boldsymbol{e}$ can be written as a linear combination of the identity matrix basis E [17], [24]. Thus, the corrupted test signal, $\boldsymbol{y}$, can be written as a linear combination of the groups from a new dictionary composed of the training samples A and the basis E. Equation (12) can be rewritten as:

$$\boldsymbol{y} = [A, E]\begin{bmatrix}\boldsymbol{x}_0\\ \boldsymbol{e}\end{bmatrix} = B\boldsymbol{w}_0 \tag{13}$$

Equation (13) is still underdetermined, since extended basis $B = [A, E] \in \mathbb{R}^{m\times(n+m)}$. As discussed in Subsection III-C, the $l^{2,1}$-minimization problem can be solved as:

$$(l^{2,1}_{\boldsymbol{e}}):\ \hat{\boldsymbol{w}}_{2,1} = \arg\min \|\boldsymbol{x}\|_{2,1} + \|\boldsymbol{e}\|_2\ s.t.\ B\boldsymbol{w} = \boldsymbol{y} \tag{14}$$

where we used the fact that the group of E has length m. Thus $\|\boldsymbol{e}\|_{2,1} = \|\boldsymbol{e}\|_2$. The classification criterion (5) should be modified as

$$\text{identity}(\boldsymbol{y}) = \arg\min_i \|\boldsymbol{y} - \boldsymbol{e} - A\delta_i(\hat{\boldsymbol{x}}_{2,1})\|_2 \tag{15}$$

To illustrate how Equation (14) deals with the noisy data, Fig. 4 shows the coefficient of a test signal contaminated with white Gaussian noise.

In this paper, an identity matrix has been selected as the error basis. To be more general, if one knows the prior knowledge of the noise, a well-designed error basis D might yield better performance. For example, if there are certain frequencies of harmonics in a power system, the error could be represented in a basis with sinusoid and cosine atoms. Then, matrix B is redefined by appending D instead of E.

B. Dimension Reduction by Random Mapping

Directly working with raw signals with high sampling frequency can be computationally expensive on microcontroller-unit-based devices. Therefore, it is necessary to reduce the data dimensionality.

The commonly used dimension reduction methods such as Principal Component Analysis (PCA) are tailored according to the nature of the data, which may be too costly. In the context of the fault classification task, a rapid dimension

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

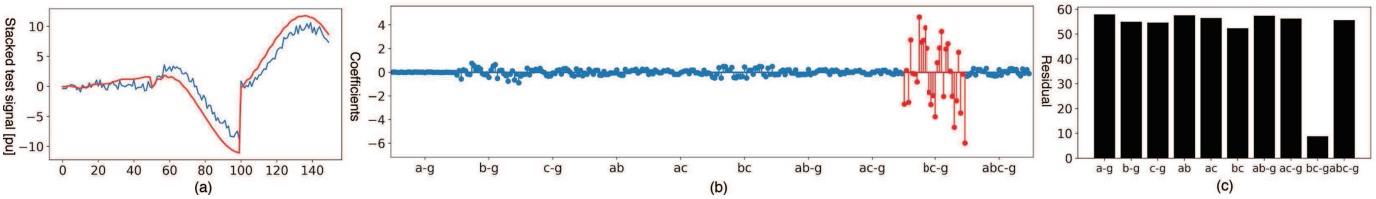


Fig. 3. Fault classification under small random Gaussian noise. (a) A bc-g fault signal is contaminated with noise (Blue). The sampling frequency is set to 5 kHz for demonstration. Phase-a of time range [1,50], phase-b of time range [51, 100] and phase-c of time range [101, 150] are stacked. The red signal is recovered by only using the bc-g fault signals. (b) The signal can be represented as a linear combination of bc-g fault signals (Red) and the contribution from the other group (Blue). (c) The reconstruction residuals of the test signal w.r.t the projection $\delta_i(\boldsymbol{x})$ are computed (Black). The bc-g fault has the lowest value.

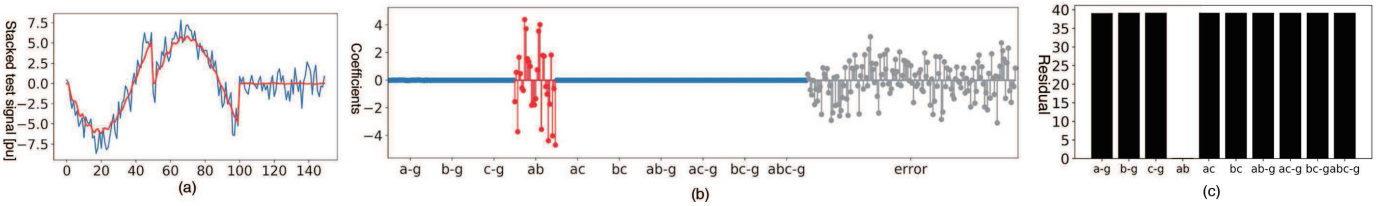


Fig. 4. Fault classification under strong random Gaussian noise. (a) An ab fault signal is contaminated with noise (Blue). The red signal is recovered by only using the ab fault signals, which filters out the noise. (b) The signal can be represented as a linear combination of ab fault signals (Red) with errors (Grey). The coefficients from the other fault type is zero. (c) The ab fault has the lowest reconstruction error.

reduction method is preferred. Random projection method is a linear projection, which can reduce the data dimensionality in a manner that preserves enough features of the original data. It has been previously studied as a general dimension reduction method for clustering problems in [25]–[27], as well as recognition problems in [17], [28].

Linear random projection from the raw signal space to the feature space can be represented as matrix $R \in \mathbb{R}^{d\times m}$ with $d \ll m$. Matrix R consists of random values drawn from an identical and independent Gaussian distribution with zero mean and unit covariance, and the Euclidean length of each column has been normalized to unity. Let $\widetilde{\boldsymbol{y}}$ and $\widetilde{A}$ denote mapped test signals and basis, respectively:

$$\widetilde{\boldsymbol{y}} = R\boldsymbol{y},\ \ \widetilde{A} = RA \tag{16}$$

Hence, the basis A and test signals $\boldsymbol{y}$ in the previous algorithm are replaced by $\widetilde{A}$ and $\widetilde{\boldsymbol{y}}$, respectively.

It has been theoretically shown in [29] that the group sparse representation classifier is robust to such dimension reduction method. In brief, the group sparse classifier can be categorized as compressive classifiers. The two conditions that guarantee the robustness of compressive classifiers are Restricted Isometry Property (RIP) and Generalized Restricted Isometry Property (GRIP).

RIP states that a small set of points in a high-dimensional space can be embedded into a space of much lower dimension in such a way that distances between the points are nearly preserved, i.e., when a sparse vector $\boldsymbol{u}$ is projected by a matrix Φ, then

$$(1-\delta)\|\boldsymbol{u}\|_2^2 \le \|\Phi\boldsymbol{u}\|_2^2 \le (1+\delta)\|\boldsymbol{u}\|_2^2 \tag{17}$$

Constant δ is an RIP constant whose value depends on the type of matrix Φ and nature of $\boldsymbol{u}$.

GRIP states as follows: for a matrix Φ which satisfies RIP for inputs $\boldsymbol{u}_i$, the inner product of two vectors is approximately maintained under the random projection, i.e., for two vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, which satisfy RIP with matrix Φ, the approximate form of GRIP is

$$\langle \Phi\boldsymbol{u}_1, \Phi\boldsymbol{u}_2\rangle \approx \langle \boldsymbol{u}_1, \boldsymbol{u}_2\rangle \tag{18}$$

where $\langle\cdot\rangle$ denotes the inner product.

Reference [29] proves that, in the $l^{2,1}$-minimization optimization problems, if a matrix satisfies the RIP and GRIP conditions, then the constraints are approximately preserved and the objective function remains the same after the random projection; and the random projection matrix satisfies the (RIP) and (GRIP) conditions with high probability. Thus, the solutions for before and after projection are approximately the same.

V. Simulation Results

A. Test Networks

To verify the effectiveness of the proposed approach, two test networks are modeled using PSCAD/EMTDC software. The first test network is a single-circuit transmission line with sources at both ends. The second test network is a back-to-back Modular Multilevel Converter (MMC) HVDC System. The system frequency and voltage of both test networks are 50 Hz and 220 kV.

1) Test Network I: A three-phase double-ended power system is simulated, as shown in Fig. 5. The model uses the same parameters in [30].

The length of the transmission line is 200 km. The positive and zero sequence impedance of transmission line are $Z_1 = 5.38 + j84.55\ \Omega$ and $Z_0 = 64.82 + j209.53\ \Omega$, respectively. The impedance of both sources is $Z_s = 9.19 + j74.76\ \Omega$. The proposed classifier in this paper only uses the three phase current signals, which are collected by the relay installed at source 1. The sampling frequency is 10 kHz. The three phases of transmission line are untransposed.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

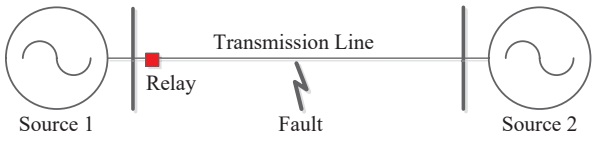


Fig. 5. Test Network I: Three-phase Double-ended System

A dataset (named dataset I) of voltage and current signal is generated by tuning the system parameters, as listed in Table I, where the fault distance is the distance between fault point and the relay, and the pre-fault power angle is the phase difference between source 1 and source 2 when faults happen.

TABLE I
SYSTEM PARAMETERS FOR THE TEST NETWORK I

System Parameter	Values or Type
Fault Distance (km)	$50, 70, 90, 110, 130, 150$
Fault Resistance (Ω)	$0.01, 5, 15, 30, 50$
Fault Inception Angle (degrees)	$0, 60, 120, 180, 240, 300$
Pre-fault Power Angle (degrees)	$10, 20, 30$
Fault Type	a-g, b-g, c-g, ab, ac, bc ab-g, ac-g, bc-g, abc-g

By considering the combination of the parameters, $6 \times 5 \times 6 \times 3 \times 10 = 5400$ samples are generated. Each sampling signal contained 3 cycles, and the fault is incepted at the end of the first cycle. In this paper, half-cycle current fault component signals are analyzed for the classification. Since the sampling frequency is 10 kHz, the length of half-cycle signals for each phase is 100; and then, the superimposed currents are obtained by subtracting the non-faulty signals associated with the same initial phase angle. It is required in the proposed algorithm that the three-phase superimposed current signals are concatenated to form new signals in the order of phase-a, phase-b, and phase-c. Thus, the length of the stacked signals m is $100 \times 3 = 300$ in p.u..

2) Test Network II: An MMC-based Back-to-Back HVDC system is studied as shown in Fig. 6.

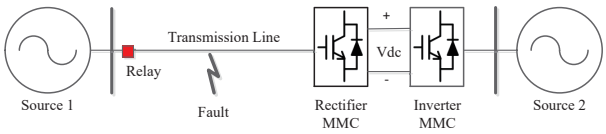


Fig. 6. Test Network II: MMC-based Back-to-Back HVDC System

The length of the transmission line is 100 km. The back-to-back MMC is installed in the vicinity of source 2. The positive sequence impedance and zero sequence impedance of the transmission lines are $Z_1 = 1.43 + j20.31\ \Omega$ and $Z_0 = 14.52 + j52.92\ \Omega$, respectively. Inverter MMC adopts $V_{\rm dc}$-control with $V_{\rm dc}^{\star} = 500$ kV and $V_{\rm ac}$-control with $V_{\rm ac}^{\star} = 215$ kV. The three-phase voltage and current signals are collected by the relay employed at transmission near source1 at the sampling frequency of 10 kHz.

We only consider the faults that occur in AC transmission lines. The rectifier MMC adopts P-control with $P^{\star}$ listed in Table II. With the rest of the parameters, a total of $9 \times 5 \times 10 \times 4 \times 3 = 5400$ signals are collected in the dataset II. Similarly, we chop half-cycle signals after the fault incident, and obtain the superimposed currents.

TABLE II
SYSTEM PARAMETERS FOR THE TEST NETWORK II

System Parameter	Values
Fault Distance (km)	$10, 20, 30, 40, 50, 60, 70, 80, 90$
Fault Resistance (Ω)	$0.01, 5, 15, 30, 50$
Fault Inception Angle (degrees)	$0, 30, 60, 90$
$P^{\star}$ (MW)	$400, 500, 600$
Fault Type	a-g, b-g, c-g, ab, ac, bc ab-g, ac-g, bc-g, abc-g

B. Selection of the Training Set Size

Sufficient training samples is the key to classify the faults. The accuracy of SRC and GSRC are computed under different training set size n. The first half-cycle signals after the fault event are taken to form the samples. For simplicity of the analysis, all classes are assumed to have the same number of training samples. Half of the fault data are randomly assigned to the training set (i.e., 270 signals per class) and the other half for testing. In each test, n training signals are randomly selected from the training set to form the dictionary A. Each classification accuracy is repeated for five times to compute the average accuracy. The results on dataset I and dataset II are shown in Table III.

As expected, the classification accuracy increases rapidly along with the increment of the training set size. When the training signals of each class are sufficient enough, the GSRC achieves higher accuracy than SRC, which is in line with the discussion in Section III-C.

For GSRC, no mistake is observed on dataset I when the number of training samples is larger than 80 (8 samples per class). The result also indicates that our method can classify the faults in the first half-cycle. Considering the computational cost and the reliability, 150 signals (15 samples per class) are selected as training samples for the following experiments on dataset I.

The GSRC does not remain a 100 % result on dataset II. The control of MMC makes the fault characteristics non-linear and more complex compared with traditional AC transmission systems. The proposed algorithm supposes a given signal can be linearly reconstructed by a subset of the training samples. The performance of GRSC would be jeopardized by non-linear features. For the following experiments on dataset II, 500 signals (50 samples per class) are selected to form the dictionary.

C. The Effect of Signal Type and Sampling Frequency

In the previous experiment, only current signals were measured for the classification. However, in cases where superimposed current signals are not available, the voltage signals

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

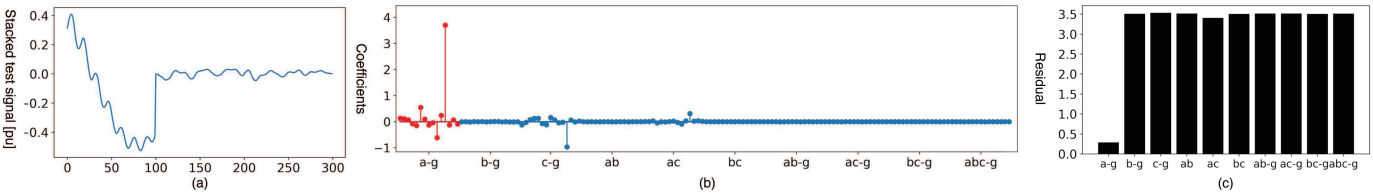


Fig. 7. Fault classification using superimposed voltage. (a) An a-g fault signal is contaminated with noise. (b) The signal can be represented as a linear combination of a-g fault signals (Red) and the contribution from the other group (Blue). (c) The reconstruction residuals of the test signal w.r.t the projection $\delta_i(\hat{\boldsymbol{x}}_{2,1})$ are computed. The a-g fault has the lowest value.

TABLE III
CLASSIFICATION ACCURACIES WITH DIFFERENT TRAINING SET SIZE

Method	Number of Samples per Class								
	5	**7**	**9**	**11**	**13**	**15**	**20**	**25**	**50**
Dataset I									
SRC	90.51	94.23	97.38	98.49	98.77	98.90	99.14	99.62	99.74
GSRC	95.87	98.38	100	100	100	100	100	100	100
Dataset II									
SRC	84.99	87.47	95.50	95.36	97.10	97.07	97.96	98.55	99.52
GSRC	87.39	90.42	95.15	95.86	96.73	97.35	98.37	99.00	99.98

TABLE IV
CLASSIFICATION FOR SUPERIMPOSED CURRENT AND VOLTAGE UNDER DIFFERENT SAMPLING FREQUENCIES

Sam. Fre. (kHz)		**0.5**	**1**	**2**	**2.5**	**5**	**10**
Dataset I	Sup. Curr.	100	100	100	100	100	100
	Sup. Volt.	92.63	93.64	99.76	99.28	99.86	99.88
Dataset II	Sup. Curr.	99.85	99.93	99.91	99.96	99.88	99.98
	Sup. Volt.	94.51	95.20	96.58	97.99	98.05	98.33

may be used instead. Further, the previous experiments were conducted when the sampling frequency was 10 kHz, but in practice, under the limits of devices, the sampling frequency may be lower than 10 kHz. Therefore, the experiments have been conducted under different signal type and sampling frequencies to examine the performance of the proposed method. Six sampling frequencies are selected namely, 0.5 kHz, 1 kHz, 2 kHz, 2.5 kHz, 5 kHz and 10 kHz. This process is repeated for five times and the average classification accuracies on dataset I and dataset II are reported in Table. IV.

The results indicate that the classification accuracy for scheme I (superimposed current signals) is basically unchanged when varying the sampling frequency. For scheme II (superimposed voltage signals), the classification accuracy is enhanced by an increase of sampling frequency. As predicted, since the superimposed current signals contain more low-frequency information about the specific fault type, low sampling rate is sufficient for the proposed method to represent the fault signals. Since the voltage signals contain more fault-induced transients, higher sampling rate will be more suitable to reveal the fault type. Further, the performance of scheme I is consistently better than scheme II. One explanation for this is that the transients negatively affect the signal representation and, to some extent, affect the recognition accuracy. According to the results, the proposed method of scheme I achieves high classification accuracies at low sampling frequency, which is practical for the real power system.

D. The Effect of Random Mapping

To examine how the random mapping projection affects the classification accuracy, the original signals are mapped to different dimensions. The first half-cycle signals after the fault event are taken to form the samples. This experiment is repeated for five times to compute the average accuracy. The results on dataset I and dataset II are presented in Fig. 8.

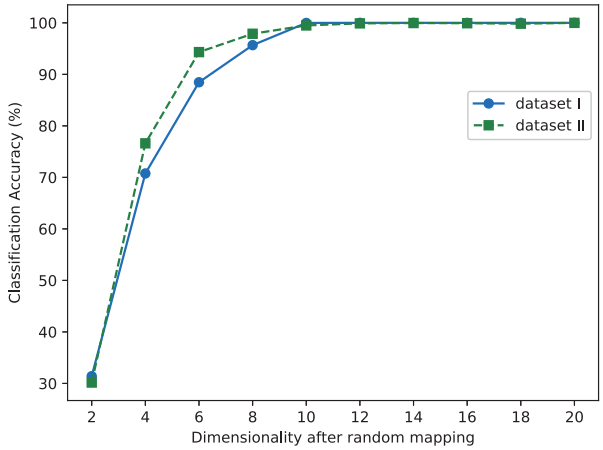


Fig. 8. The classification accuracies under random projection

The classification accuracy rises with increasing the feature dimension. The classification accuracy is actually related to the RIP and GRIP inequality constants. The results show that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

$d = 10$ random linear projections can retain the features of original signals, and suffice for group sparse recovery.

E. Performance on Remote Faults

To investigate the accuracy on remote faults, 160 faults at 190 km away from the relaying point are simulated in the test network I and classified. The tunable system parameters are listed in Table V. Results show that the accuracy of the method remains at 100 %.

TABLE V
SYSTEM PARAMETERS FOR REMOTE FAULTS

System Parameter	Values
Fault Resistance (Ω)	$0.05, 10, 25, 50$
Fault Inception Angle (degrees)	$0, 45, 90, 135$
Fault Type	a-g, b-g, c-g, ab, ac, bc ab-g, ac-g, bc-g, abc-g

F. Robustness to Noise

To ensure the method is noise-tolerant, the classification accuracy is tested on dataset I under six different Signal-to-Noise Rate (SNR) values of 10 db, 12 db, 14 db, 16 db, 18 db and 20 db. As discussed in Section IV-A, there exist two approaches to deal with noise, i.e., increasing ε (scheme I) and extending the basis with E (scheme II). Table VI compares the performance between these two approaches. The values used for ε are 0.05, 0.2 and 0.5. It should be noted that the random projection has not been implemented in this experiment.

TABLE VI
CLASSIFICATION ACCURACY ON DATASET I UNDER DIFFERENT SNRS FOR SCHEME I (TUNING ε) AND SCHEME II (EXPANDED BASIS)

SNR(db)	10	12	14	16	18	20
$\varepsilon = 0.05$	42.62	53.60	61.51	64.96	72.52	77.78
$\varepsilon = 0.2$	51.99	62.52	66.34	71.33	74.90	79.6
$\varepsilon = 0.5$	62.54	81.77	83.31	94.04	97.85	100
extended basis	99.09	99.78	99.94	99.98	100	100

TABLE VII
CLASSIFICATION ACCURACY ON DATASET I FOR DIFFERENT FAULT TYPE UNDER DIFFERENT SNRS FOR SCHEME II

SNR(db)	10	12	14	16	18	20
SLG	100	100	100	100	100	100
LLG	96.99	99.28	99.82	99.96	100	100
LL	100	100	100	100	100	100
TPG	99.96	100	100	100	100	100

The results also indicate that scheme II yields high accuracies, which are above 99 % under different noise amplitudes. When SNR is above 18 db, no incorrect classification is observed. Scheme I can effectively improve the robustness against noise. The classification accuracy increases with the increase in noise energy bound ε, given that the SNR is kept fixed. The proposed method is a task of reconstructing a given signal by selecting a relatively small subset of basis from a large basis pool. The uncorrupted training signals from a single class are not enough to span a given signal corrupted by strong noise. To accurately recover the noise, training signals from other classes are contributed, which jeopardizes the sparsity of the representation. The increment in noise energy bound alleviates the requirement of reconstructing the noise which keeps the sparsity to a certain degree. However, this method is designed to deal with small noise. If the ε is too large, the accuracy of the model will be affected.

In addition, the classification accuracies of Single-Line-to-Ground (SLG), Line-to-Line-to-Ground (LLG), Line-to-Line (LL) and Three-Phase-to-Ground (TPG) faults under the scheme II are listed in Table VII. Table VII suggests LLG fault classification to be the least robust.

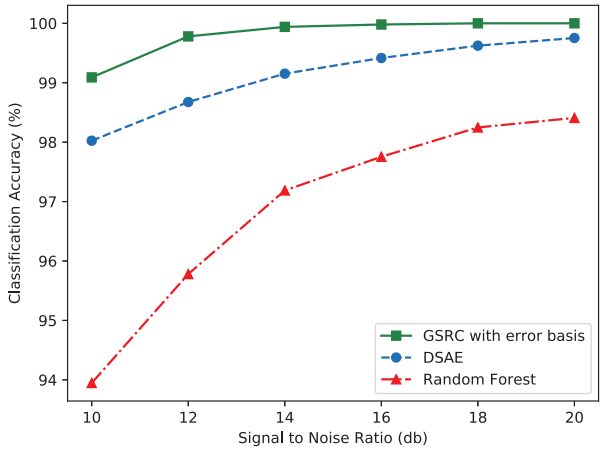


Fig. 9. Comparison with DASE and random forest under various SNR values

Comparative studies with sparse auto-encoder based method [30] and random forest based method [3] in terms of robustness to noise are conducted on dataset I.

Kunjin et al. [30] proposed the denoising sparse auto-encoder (DSAE) to deal with white Gaussian noise (WGN). The method requires half-cycle three-phase current and voltage signals from one side of the transmission line. Jamehbozorg et al. [3] used the random forest algorithm to classify faults in single-circuit transmission lines. The method requires a voltage of one phase and the currents of the other two phases at one side of the transmission line. Then, half-cycle discrete Fourier transform is applied to calculate the first ten odd harmonics. As recommended, 110 decision trees are used for the decision-making process. 70 % of the dataset instances are randomly selected for training and the rest for testing. During the training process, white Gaussian noise with different SNRs were added to the training signals. The comparison with DSAE and the random forest is presented in Fig. 9. It is demonstrated that our proposed method outperforms the other methods.

Moreover, we also considered the robustness to two kinds of representative measurement errors describe in [30]. The first type (Type I error) is "consecutive zero error" —a fraction of a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2866487, IEEE Transactions on Smart Grid

signal is zero. The second type (Type II error) is "consecutive high value"—a fraction of a signal rises to a high value (either positive of negative). For each test signal, one error randomly added to one of its phases and type I and type II errors account for 50 % of the errors. For type II error, the value is set two times of the rated amplitude of the signal. In this experiment, the sampling frequency of the fault signals is set to 20 kHz. The proportion of error in a single phase is 0.5 % to 3 %. [30] used sparse auto-encoder (SAE) with 10 % dropout to cope with the measurement errors. We enlarge the training set with corrupted data to train the random forest.

The results are shown in Table VIII, from which we clearly find the advantage of GSRC. When the high value is set to two times of the rated amplitude, GSRC with error basis achieves 100 % classification accuracy. Even amplifying the noise to 20 times of the rated amplitude, the performance is still satisfactory.

It should also be noted that the SAE and the random forest-based methods require adding noise to the training signals to enhance the robustness. However, our proposed method is only trained with uncorrupted signals, because the error basis is sufficient to deal with the WGN or measurement error.

TABLE VIII
COMPARISON WITH SPARSE AUTO-ENCODER WITH DROPOUT UNDER VARIOUS PROPORTION OF ERROR

Method	**High error value** (p.u.)	**Proportion of error (%)**					
		0.5	**1**	**1.5**	**2**	**2.5**	**3**
SAE	2	97.73	95.43	93.73	92.44	91.13	90.17
Random Forest	2	96.92	96.26	94.69	94.14	92.82	91.98
GSRC	2	**100**	**100**	**100**	**100**	**100**	**100**
	10	100	99.99	99.96	99.90	99.83	99.74
	20	99.96	99.47	98.78	97.78	97.12	96.48

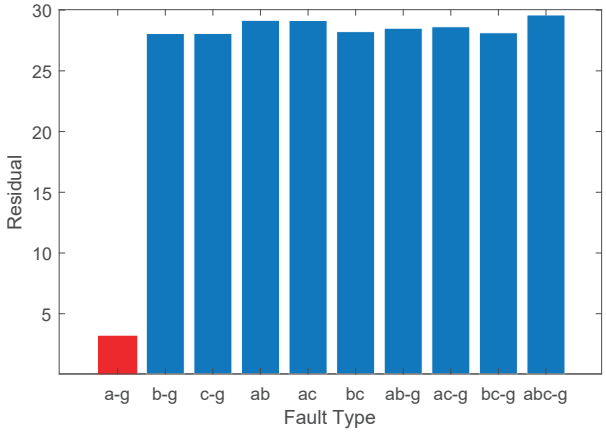


Fig. 10. The residuals of an a-g fault signal of Fuxing side

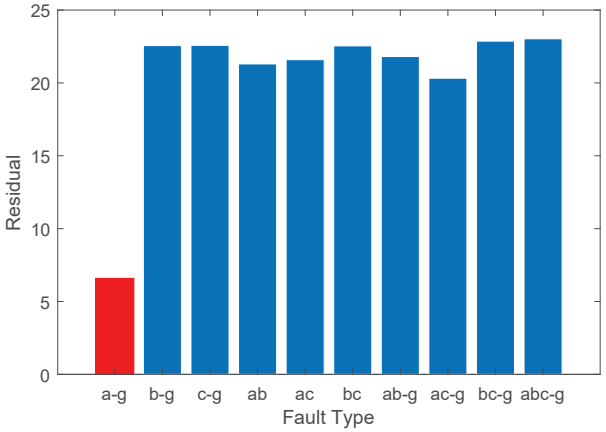


Fig. 11. The residuals of an a-g fault signal of Zhanggong side

G. Verification on Field Data

Field data recorded from the 220 kV Zhanggong-Fuxing transmission lines in China are tested. An a-g fault occurred 23.30 km from Zhanggong side and 45.23 km from Fuxing side. Three phase current signals are recorded at both sides. The sampling frequency is 1.2 kHz at Fuxing side and 4 kHz at Zhanggong side. The first half-cycle signals after the fault event are cropped from the field data to calculate the superimposed component. 500 signals (50 per class) are randomly selected from the dataset I and downsampled to form the dictionary. The two data are classified correctly. The reconstruction errors for each fault type is plotted in Fig 10 and Fig 11.

As shown in the figure, the residual of the correct class is much smaller than other classes. The ratio of the two smallest residuals is about 1:8.9 in Fig 10 and 1:3.1 in Fig 11.

VI. CONCLUSION AND FUTURE WORK

In this paper, an innovative method was presented for fault classification of single-circuit transmission lines. The GSRC method only requires three-phase superimposed current signals from one side of the transmission line, and the decision making is performed in a half-cycle. The main conclusions and suggestions according to the simulation results are summarized as follows:

1. When the signals of each class are sufficient to linearly represent the test signals, the group sparse representation based method achieves high classification accuracies by using the $l^{2,1}$-minimization. The GSRC method has better performance than SRC one.

2. Tests with varying sampling frequency show that superimposed current signals achieve favorable performance even under low sampling rates.

3. The random mapping matrix is independent of training data and can be easily generated. With random projection, the data dimension can be effectively reduced with less computational cost.

4. By adding error basis or increasing the ε, noise can be handled robustly within the same classification framework. The method which increases the ε is designed for treating with small noise, while the method using extended basis can deal with strong noise.

Although results have shown that directly using fault signals as dictionary basis guarantees favorable performance, a well-learned dictionary may lead to higher performance with fewer base element [31]. Moreover, learning a dictionary could mitigate the expensive computation cost if a dictionary has

a large number of training samples. Since the error basis is an identity matrix, a well-designed basis might yield better performance in addressing noise.

ACKNOWLEDGMENT

The authors would like to thank Professor Huaping Liu from Tsinghua University for his valuable and constructive advice on the theory of sparse representation.

REFERENCES

[1] A. Jamehbozorg and S. M. Shahrtash, "A decision tree-based method for fault classification in double-circuit transmission lines," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2184–2189, Oct 2010.

[2] B. Das and J. V. Reddy, "Fuzzy-logic-based fault classification scheme for digital distance protection," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 609–616, April 2005.

[3] A. Jamehbozorg and S. M. Shahrtash, "A decision-tree-based method for fault classification in single-circuit transmission lines," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2190–2196, Oct 2010.

[4] K. H. Kashyap and U. J. Shenoy, "Classification of power system faults using wavelet transforms and probabilistic neural networks," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, vol. 3, May 2003, pp. III–423–III–426 vol.3.

[5] U. B. Parikh, B. Das, and R. P. Maheshwari, "Combined wavelet-svm technique for fault zone detection in a series compensated transmission line," *IEEE Transactions on Power Delivery*, vol. 23, no. 4, pp. 1789–1794, Oct 2008.

[6] D. Chanda, N. K. Kishore, and A. K. Sinha, "A wavelet multiresolution-based analysis for location of the point of strike of a lightning overvoltage on a transmission line," *IEEE Transactions on Power Delivery*, vol. 19, no. 4, pp. 1727–1733, Oct 2004.

[7] P. Dutta, A. Esmaeilian, and M. Kezunovic, "Transmission-line fault analysis using synchronized sampling," *IEEE Transactions on Power Delivery*, vol. 29, no. 2, pp. 942–950, April 2014.

[8] T. Dalstein and B. Kulicke, "Neural network approach to fault classification for high speed protective relaying," *IEEE Transactions on Power Delivery*, vol. 10, no. 2, pp. 1002–1011, Apr 1995.

[9] P. K. Dash, S. R. Samantaray, and G. Panda, "Fault classification and section identification of an advanced series-compensated transmission line using support vector machine," *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 67–73, Jan 2007.

[10] W. Dong, Q. Gong, W. Lai, W. Bo, L. Dong, Q. Hui, and L. Gang, "Research on internal and external fault diagnosis and fault-selection of transmission line based on convolutional neural network," *Proceedings of the Csee*, 2016.

[11] S. M. S. Alam, B. Natarajan, and A. Pahwa, "Distribution grid state estimation from compressed measurements," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1631–1642, July 2014.

[12] Y. Fu and T. S. Huang, "Image classification using correlation tensor analysis," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 226–234, Feb 2008.

[13] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1794–1801.

[14] S. Li and Y. Fu, "Low-rank coding with b-matching constraint for semi-supervised classification," in *International Joint Conference on Artificial Intelligence*, 2013, pp. 1472–1478.

[15] J. Tang, R. Hong, S. Yan, T. S. Chua, G. J. Qi, and R. Jain, "Image annotation by k nn-sparse graph-based label propagation over noisily tagged web images," *Acm Transactions on Intelligent Systems and Technology*, vol. 2, no. 2, p. 14, 2011.

[16] K. Blekas and A. Likas, "Sparse regression mixture modeling with the multi-kernel relevance vector machine," *Knowledge and Information Systems*, vol. 39, no. 2, pp. 241–264, 2014.

[17] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb 2009.

[18] A. Majumdar and R. K. Ward, "Classification via group sparsity promoting regularization," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 861–864.

[19] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[20] T. T. Emmanuel J. Candes, Justin K. Romberg, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2005.

[21] E. J. Candes, "Compressive sampling," *Proceedings of the international congress of mathematicians*, vol. 3, pp. 1433–1452, 08 2006.

[22] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," in *Wavelets and Sparsity XV*, vol. 8858. International Society for Optics and Photonics, 2013, p. 88580R.

[23] A. Y. Yang, Z. Zhou, A. G. Balasubramanian, S. S. Sastry, and Y. Ma, "Fast l_1-minimization algorithms for robust face recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3234–3246, Aug 2013.

[24] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *CVPR 2011*, June 2011, pp. 1873–1879.

[25] S. Kaski, "Dimensionality reduction by random mapping: fast similarity computation for clustering," in *1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227)*, vol. 1, May 1998, pp. 413–418 vol.1.

[26] D. Achlioptas, "Achlioptas, d.: Database-friendly random projections: Johnson-lindenstrauss with binary coins. j. comput. syst. sci. 66, 671-687," *Journal of Computer and System Sciences*, vol. 66, no. 4, pp. 671–687, 2003.

[27] G. Li and Y. Gu, "Restricted isometry property of gaussian random projection for finite set of subspaces," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1705–1720, April 2018.

[28] D. Valsesia, G. Coluccia, T. Bianchi, and E. Magli, "Compressed fingerprint matching and camera identification via random projections," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1472–1485, July 2015.

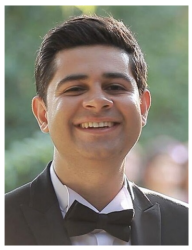
[29] A. Majumdar and R. K. Ward, "Robust classifiers for data reduced via random projections," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 5, pp. 1359–1371, Oct 2010.

[30] K. Chen, J. Hu, and J. He, "Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1748–1758, May 2018.

[31] Y. Sun, Q. Liu, J. Tang, and D. Tao, "Learning discriminative dictionary for group sparse representation," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3816–3828, Sept 2014.

Shenxing Shi (M'09 - SM'18) received the B.Sc. degree in electrical engineering from Northeast Electric Power University, Jilin, China, in 1999, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2006. Currently, he is an Associate Professor with the Department of Electrical Engineering, Tsinghua University, Beijing, China. His research interests include traveling-wave-based protective relaying and fault analysis in power systems.

Beier Zhu (S'18) received the B.Sc. degree in electrical engineering from Tsinghua University, Beijing, China, in 2016, where he is currently pursuing the M.Sc. degree with the Department of Electrical Engineering. His research interests included power system protection, sparse signal representation and pattern analysis in power systems.



Sohrab Mirsaeidi (M'17) received the Ph.D. degree in Electrical Engineering from Universiti Teknologi Malaysia (UTM), Johor, Malaysia in 2016. Currently, he is a Postdoctoral Fellow with the Department of Electrical Engineering, Tsinghua University, Beijing, China. Dr. Mirsaeidi has published more than 40 scientific papers and books in the field of control and protection of power systems. He has also been involved in several key research projects funded by the Chinese government. His main research interests include control and protection of large hybrid AC/DC grids, distributed generation, and microgrids.

Xinzhou Dong (M'99 - SM'01 - F'16) received the B.Sc., M.Sc., and Ph.D. degrees in Electrical Engineering from Xi'an Jiaotong University, China, in 1983, 1991, and 1996, respectively. He furthered his postdoctoral research at the Electrical Engineering Station of Tianjin University, Tianjin, China, from 1997 to 1998. Since 1999, he has been employed by Tsinghua University, Beijing, China. Currently, he is a Professor with Dept. of Electrical Engineering, Tsinghua University and Director of the International Union Research Center of Beijing on Green Energy and Power Safety. His research interests include protective relaying, fault location, and the application of wavelet transforms in power systems. Prof. Dong is a Fellow of IEEE and IET. He is an author or co-author of more than 200 journal papers.