

CHANGE DETECTION IN STREAMS OF SIGNALS WITH SPARSE REPRESENTATIONS

Cesare Alippi, Giacomo Boracchi*

Dipartimento di Elettronica, Informazione
e Bioingegneria, Politecnico di Milano

Brendt Wohlberg†

Theoretical Division
Los Alamos National Laboratory

ABSTRACT

We propose a novel approach to performing change-detection based on sparse representations and dictionary learning. We operate on observations that are finite support signals, which in stationary conditions lie within a union of low dimensional subspaces. We model changes as perturbations of these subspaces and provide an online and sequential monitoring solution to detect them. This approach allows extension of the change-detection framework to operate on streams of observations that are signals, rather than scalar or multivariate measurements, and is shown to be effective for both synthetic data and on bursts acquired by rockfall monitoring systems.

Index Terms— Change detection, sparse representation, dictionary learning, sequential monitoring.

1. INTRODUCTION

This paper addresses the problem of detecting, in an online and sequential manner, changes in an unknown data-generating process, i.e., departures of the process from its nominal (or original) stationary state. In particular, we address the specific case where each observation is a finite support signal, admitting a sparse representation on an unknown dictionary.

Most change-detection algorithms [1] operating on streaming data assume independent and identically distributed (i.i.d.) observations, and changes in stationarity are detected by applying sequential techniques directly to raw observations. Other techniques assume datastreams characterized by temporal dynamics [2], to be approximated by means of suitable predictive models. Change detection is carried out by either inspecting residuals (the discrepancy between real and predicted data) or model parameters estimated on a sliding window basis [1].

However, the above mechanisms are not viable when the process generates independent *signals*, each characterized by a deterministic structure that can not be described by a single model (e.g., in situations where signals arrive intermittently or are characterized by different modalities). An example of these processes is provided by the rockfall monitoring application. The enlargement of existing (micro)fractures within the rock face and their coalescing into larger dimensions generate bursts of microacoustic (microseismic) emissions whose signals can be retrieved with geophones or accelerometers [3]. Moreover, falling off stones hitting the neighbor of the sensors could also trigger burst recording, though these latter show a different structure. All these bursts can be seen as generated from

a unique burst-generating process in a stationary state, until the coalescing phenomenon evolves or a fault occurs in the sensing apparatus. Geologists claim that changes in stationarity are associated with macroscopic structural variation in the phenomenon, eventually leading to the rock collapse. Immediate detection of process changes allow civil engineers (within a structural risk monitoring application) and geophysicist (rockfall and rock toppling) to assess the risk and possibly build predictive mechanisms for it. Aside from this scenario, there are other environmental, industrial or structural health monitoring scenarios [4, 5] characterized by similar processes.

The solution we propose to detect changes in streams of signals is to model the stationary conditions as having some generating sparse representation [6] with respect to a fixed but unknown dictionary. An estimate of the dictionary is obtained via dictionary learning techniques [7], and a change in the process is detected when there is a change in the characteristics of the sparse representation estimated from the observed signals. The specific detection mechanism considered here is monitoring the residual error of the estimated sparse representation by means of a change-point method [8].

1.1. Observation Model and Problem Statement

We consider a process \mathcal{S} generating observed signals $\mathbf{s} \in \mathbb{R}^M$ modeled as

$$\mathbf{s} = D_0 \mathbf{x} + \boldsymbol{\nu} \quad (1)$$

for some $D_0 \in \mathbb{R}^{M \times N}$, $\mathbf{x} \in \mathbb{R}^N$, and $\boldsymbol{\nu} \in \mathbb{R}^M$. We assume that \mathbf{x} is *sparse* [6], i.e. $\|\mathbf{x}\|_0 = L \ll N$, where the ℓ^0 “norm” of \mathbf{x} is the number of non-zero components in \mathbf{x} or that \mathbf{s} is compressible [9] with respect to dictionary D_0 , i.e., that $|x_k|$, $k \in \{1, \dots, N\}$ have a power law decay with increasing k , x_k being the k^{th} sorted component of \mathbf{x} . In the latter case $D_0 \mathbf{x}$ does not have an exact sparse representation, but it can be approximated to high accuracy by such a representation. In (1), $\boldsymbol{\nu}$ represents a homoscedastic, zero-mean, noise term (i.e., $E[\nu_k] = 0$ and $E[\nu_k^2] = \sigma^2$, $k \in \{1, \dots, M\}$, where $E[\cdot]$ denotes the mathematical expectation). We do not make any assumption regarding \mathbf{x} other than that signals \mathbf{s} are either sparse or have suitable decay properties [10].

In the sequel we consider a streaming scenario where signals (1) arrive over time and changes in the generating process \mathcal{S} are inspected online, in a possibly infinite stream of signals *drawn* from \mathcal{S} :

$$S = \{\mathbf{s}_i\}_{i=1, \dots,} \quad (2)$$

where \mathbf{s}_i denotes the i^{th} signal. More precisely, if we denote by I^* the change point, we consider changes affecting the signals as follows:

$$\mathbf{s}_i = \begin{cases} D_0 \mathbf{x}_i + \boldsymbol{\nu}_i & i < I^* \\ D_1 \mathbf{x}_i + \boldsymbol{\nu}_i & i \geq I^* \end{cases}, \quad (3)$$

where D_1 is a new dictionary having one or more columns that cannot be sparsely or compressibly represented on D_0 . Thus, we expect

*This research has been funded by the European Commissions 7th Framework Program, under grant Agreement INSFO-ICT-270428 (iSense).

†This research was supported by the U.S. Department of Energy through the LANL/LDRD Program.

to be able to detect a change in \mathcal{S} if we monitor a suitable measure of the degree to which the observed signals are sparse or compressible on D_0 . Note that D_1 could consist of D_0 together with some atoms that are not sparsely representable with respect to D_0 : in this case, the change model in (3) corresponds to the presence of additive components that are not sparsely representable on D_0 .

The change-detection problem can be formulated as computing, in an online manner, an estimate, \hat{I} , of the change point I^* . We do not address the problem of detecting at which sample of \mathbf{s}_{I^*} the change has occurred, since each signal is considered to be acquired at once. For change-detection purpose, we require a suitable training set containing I_0 signals generated by \mathcal{S} in stationary conditions and insert each training signals in a column of the training matrix $T_0 = \{\mathbf{s}_i\}_{i \in \{1, \dots, I_0\}}$.

1.2. Related Works

The use of sparse representations for detection problems has mainly been considered in the compressive sensing literature [11, 12, 13, 14, 15], where *signal detection* algorithms have been proposed. These works [12, 13, 14, 15] assume that observations provided in stationary conditions are just noise, and detect the presence of any structured signal from compressive measurements of the observations. In contrast, we assume that in stationary conditions observations live in a union of subspaces (3), and – by means of standard signal sampling – we address the *change detection* problem, i.e., the detection of whatever departure of \mathcal{S} from its initial stationary state that can be represented as (3). The algorithm in [11] instead aims at identifying a signal of interest in the superimposition of noise and interference signals. The algorithm requires two dictionaries, each yielding sparse representations of the interference and signal of interests, while we do not assume any specific information concerning the signal to be detected.

Furthermore, the above algorithms leverage *one shot* detection techniques, such as the Neyman-Pearson detector, and make decisions considering each observed signal alone, while we enforce a sequential change-detection technique. Sequential techniques allow the detection of small perturbations when these are persistent, while subtle changes become difficult to detect by means of one-shot tools.

It is also worth mentioning the anomaly detection literature (see [16] and references there in), which addresses the detection of anomalous patterns in datastreams. However, these works target off-line processing, and provide data analysis method rather than online detection techniques; furthermore, none of these works considers sparse representations to model processes.

2. PROPOSED APPROACH

We proceed as follows to detect changes in \mathcal{S} : first we characterize the stationary state of \mathcal{S} by learning a dictionary \hat{D}_0 that yields sparse representations of signals generated from \mathcal{S} . Then, we use \hat{D}_0 to compute a sparse representation (sparse coding) of each incoming signal \mathbf{s}_i , and detect a change in \mathcal{S} when there is a change in the degree to which an incoming signal can be sparsely represented w.r.t. \hat{D}_0 . To quantitatively assess the extent to which a signal is sparse w.r.t. \hat{D}_0 , we compute a sparsity-related feature, which is monitored by a change-detection algorithm during the operational period. There are many possible ways to apply this general model; in the remainder of this section we describe the specific solution we have adopted in the experiments of Section 3.

2.1. Dictionary Learning

Since the dictionary D_0 characterizing the stationary state of \mathcal{S} is unknown, it is necessary to learn it from training data sampled from the stationary state. The dictionary learning problem can be formalized as a joint optimization over both dictionary and coefficients of a sparse representation of the training data; the usual approach consisting of alternating between a sparse coding stage (minimization with respect to the coefficients) and a dictionary update stage (minimization with respect to the dictionary). The specific form of the problem we adopt here is

$$\hat{D}_0 = \arg \min_{D \in \mathbb{R}^{M \times N}, X \in \mathbb{R}^{N \times I_0}} \|DX - T_0\|_F \quad \text{such that } \|\mathbf{x}_k\|_0 \leq L_0 \quad \forall k \quad (4)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $D \in \mathbb{R}^{M \times N}$ is a matrix representing a dictionary, $X \in \mathbb{R}^{N \times I_0}$ is such that each column contains the coefficients of the corresponding signal in T_0 w.r.t. the dictionary D , \mathbf{x}_k is the k^{th} column of X , and L_0 determines the required sparsity of the solutions. Exact solution of this optimization problem is computationally intractable (the problem is NP-hard), but a greedy solution can be obtained via the K-SVD dictionary learning algorithm [17], which utilizes Orthogonal Matching Pursuit (OMP) [18] for the sparse coding stage of the algorithm.

The choice of L_0 depends on the underlying process generating \mathbf{x} . When \mathcal{S} is such that \mathbf{x} is sparse, a natural choice is $L_0 = L$, which is viable if L is known or can be determined from prior information regarding \mathcal{S} . Otherwise, L_0 should be estimated from the behavior of the plot of the reconstruction error $\|DX - T_0\|_F$ against L_0 in (4). More generally, if \mathbf{x} has power law decay, then it is not strictly sparse, and we must choose L_0 to balance the sparsity against the accuracy of approximation of the recovered representation. A larger L_0 represents a greater fraction of the signal energy by the sparse model, leaving a smaller unmodelled residual, but also reduces the residual in the changed state. Conversely, decreasing L_0 increases the residual in the changed state, but at the expense of a larger residual error in the stationary state.

This parameter can be considered to control the strength of the sparsity-based model; when L_0 is small the model is very restrictive, becoming weaker as L_0 increases. Once $L_0 \geq M$, the sparsity-based model is vacuous since any subset of M dictionary elements (assuming that they are linearly independent) spans the entire signal space. The best choice of L_0 will depend on the different decay rates of the elements of \mathbf{x} in the stationary and changed states, and should be set empirically.

2.2. Sparse Coding

Exploiting the model (1) requires the estimation of the sparse generating coefficients for a given signal with respect to dictionary \hat{D}_0 . When the degree to which the incoming signals can be sparsely represented changes, it is likely that \mathcal{S} has changed. The natural choice of sparse coding optimization corresponding to dictionary learning problem (4) is

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \|\hat{D}_0 \mathbf{x} - \mathbf{s}\|_2 \quad \text{such that } \|\mathbf{x}\|_0 \leq L_0, \quad (5)$$

with the minimization performed via OMP. The value of L_0 selected for the dictionary learning stage is also appropriate here.

We quantitatively assess how likely \mathbf{s}_i is to have been generated by \mathcal{S} by computing the error of its sparse coding w.r.t. \hat{D}_0 , i.e.,

$$e_i = \|\hat{D}_0 \hat{\mathbf{x}}_i - \mathbf{s}_i\|_2, \quad (6)$$

where \hat{x}_i are the coefficients computed in (5). We assume that learned dictionary \hat{D}_0 is a sufficiently good approximation of D_0 for (3) to hold with \hat{D}_0 substituted for D_0 . Thus, the problem of detecting a change in the stream of signals $\{s_i\}_{i>I_0}$ can be reformulated as monitoring the scalar sequence $\{e_i\}_{i>I_0}$ to detect changes in the distribution of the reconstruction error (6), and this latter problem can be conveniently treated by means of sequential change-detection techniques.

2.3. Detecting Changes in Stationarity

For the final change-detection step we adopt a change-point method (CPM) [8] on the sequence of reconstruction errors (6). CPMs are statistical tests designed for i.i.d. sequences, which have been recently extended to operate online on data streams [20, 21] with bounded computational complexity and memory requirements. Since the distribution of (6) in stationary conditions is unknown, we have to adopt a nonparametric CPM, and in particular we choose one [20] based on the Lepage statistic \mathcal{L} [22], which is a nonparametric statistic used to detect changes in the location and scale of a scalar distribution¹. The advantages of CPMs over other nonparametric solutions (see [1, 23, 24] and reference therein), are that CPMs operate at a controlled average run length (ARL), namely, the expected number of samples before having a detection within an i.i.d. sequence. Furthermore, online CPMs do not need a proper training phase, and only few i.i.d. samples are required for configuration.

In what follows we denote by $\delta_i \in \{0, 1\}$, the output of the CPM based on the \mathcal{L} statistics applied on $\{e_j\}_{j \leq i}$, i.e. the sequence of the reconstruction errors until the i^{th} signal,

$$\delta_i = \text{CPM}_{\mathcal{L}}(\{e_j\}_{j \leq i}). \quad (7)$$

We apply the CPM at the arrival of each new signal, and we detect a change in \mathcal{S} as soon as $\text{CPM}_{\mathcal{L}}$ raises a detection, i.e.,

$$\hat{I} = \min_{\delta_i=1} (i). \quad (8)$$

3. EXPERIMENTS

We test the proposed approach on several sequences composed of 500 signals generated before the change and 500 after the change (such that $I^* = 501$). We report the final performance of $\text{CPM}_{\mathcal{L}}$ to indicate how these changes can be detected in sequential monitoring applications; these are:

- *FPR*, the false positive rate, i.e., the percentage of detections not corresponding to an actual change (here, when $\hat{I} \leq 500$).
- *DL*, the detection latency, i.e., the average delay of correct detections. Here, the delay for an individual correct detection is $\hat{I} - 501$.
- *FNR*, the false negative rate, i.e., percentage of missed detections. To keep *FPR* low, we configured the CPM *ARL* to 10000.

We also measure the change detectability comparing, by means of the Kolmogorov-Smirnov (KS) statistic, the empirical distributions of reconstruction errors (6) before and after the change. The KS statistic ranges in $[0, 1]$, and has to be considered as an *absolute* indicator of how good the reconstruction error is in indicating

¹By doing so, we implicitly assume the reconstruction errors to be independent in stationary conditions. This assumption is reasonable as far as \hat{D}_0 was successfully learned, and the bias term in the reconstruction error is negligible or independent of the signal. However, when a few subspaces of the signal space can be far better approximated than others, in practice, it is enough for the signals to arrive in a random order w.r.t. these subspaces.

changes of \mathcal{S} : the larger the more different the two distributions are. To illustrate the effectiveness of the proposed approach, we also assess $\text{CPM}_{\mathcal{L}}$ and the KS statistic on the ℓ^1 and ℓ^2 norms of the input signals, showing that, most often, the considered changes would not be perceivable in the signal domain. We test the proposed approach on both synthetic datasets and on a dataset of recorded bursts from the rockfall monitoring application.

3.1. Synthetically Generated Dataset

The synthetic dataset consists of signals of $M = 64$ samples, each of which has an L -sparse representation w.r.t. D_0 or D_1 , two dictionaries having some of their atoms in common. These L -sparse signals are synthesized from D_0 or D_1 by randomly defining the L nonzero coefficients, and then adding Gaussian white noise to achieve a specific SNR value. The dataset consists of three components: a training set T_0 containing 20000 signals generated from D_0 , which are used to learn \hat{D}_0 by means of K-SVD (4); a stationary condition testing set S_0 containing an additional 20000 signals generated from D_0 ; and a change condition testing set S_1 containing 20000 signals generated from D_1 . We set D_0 as a Daubechies-4 wavelet basis [25], augmented with the first 32 discrete Fourier basis elements, while D_1 contains the biorthogonal 3/7 wavelet basis [26] and the same Fourier basis elements. These experiments are performed in the ideal conditions where L_0 corresponds to the known signal sparsity L and the number of atoms of \hat{D}_0 coincides with that of D_0 .

We generated 10 datasets for each pair (L, SNR) with $L \in \{1, 2, 3, 4, 6, 8, 11, 16, 23, 32\}$ and $\text{SNR} \in \{5, 10, 20, 40\}$. In each dataset, the change-detection performance is assessed from 40 sequences of 1000 signals each ($I^* = 501$), which are prepared by suitably concatenating elements of S_0 and S_1 . The KS statistic is computed comparing all the reconstruction errors in S_0 and S_1 .

Results averaged over all the datasets are reported in Fig. 1. In particular, Fig. 1 (a) shows that, when the signals are sparse on D_0 ($L \leq 16$), the change is promptly detected even at low SNR, however, the change is not detectable when the signals are not sparse enough, as the FNR in Fig. 1 (b) shows. The FPR, which depends on the ARL of $\text{CPM}_{\mathcal{L}}$, was always between 4% and 7%, without any specific trend w.r.t. L or SNR (not reported in Fig. 1). The KS statistic in Fig. 1 (c) is consistent with the change-detection performance: the maximum is not achieved at low values of L because 32 atoms of D_0 and D_1 coincide so that there is quite a large probability that signals after the change are generated only from atoms that are also in D_0 . Remarkably, the considered change would not be detectable by monitoring the ℓ^1 or ℓ^2 norms of the input signals: the KS statistic of ℓ^1 norms in Fig. 1 (d) achieves very low values and similarly the KS statistic of the ℓ^2 norms was always below 0.02 (without any specific trend w.r.t. L and SNR), and for this reason has not been reported in Fig. 1).

3.2. Rock Face Monitoring Bursts Dataset

Bursts have been recorded by a hybrid wireless-wired monitoring system [3], which has been designed at Politecnico di Milano and deployed in the Alps to monitor rock collapses. Bursts are recorded by triaxial MEMS accelerometers, sampled at 2 kHz from seven sensing nodes deployed on the rock face. As a preliminary preprocessing step we register each 3D burst by projecting it along its principal direction (computed via principal component analysis) and cropping a signal of $M = 64$ consecutive samples about the burst peak.

From two years of monitoring activity we prepared a dataset of 40000 signals, 20000 of which were used for dictionary learning

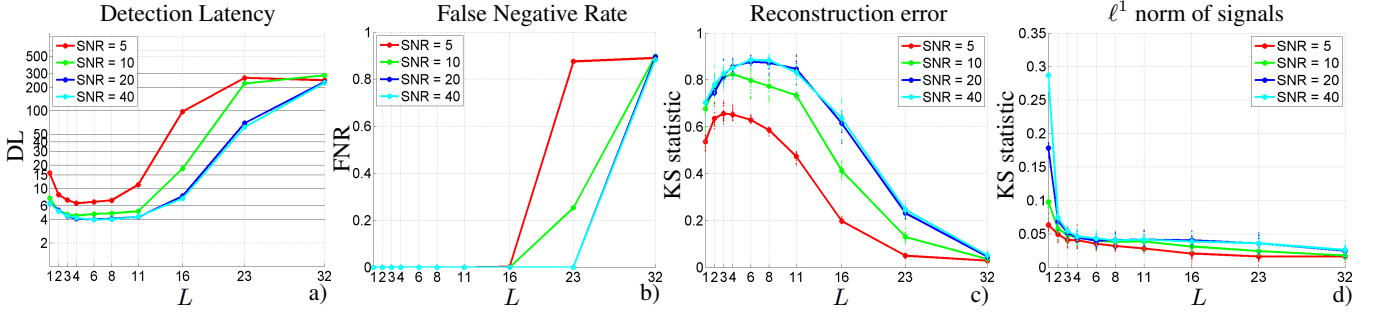


Fig. 1. Experiments on synthetically generated dataset: the detection latency of the proposed solution (a) together with the FNR (b), the values of the KS statistic for the reconstruction errors (c) and ℓ^1 norm of the input signals (d). The individual values of the KS statistic over the 10 datasets generated for each (L, SNR) pair are reported (small dots) together with their average (solid line). It emerges that reconstruction errors clearly indicate the occurred change as far as the signals are sparse ($L \leq 16$): very good change-detection performance can be achieved at low SNR when the signal is less than $1/4$ sparse. The KS statistic for the ℓ^2 norm of the input signals was below 0.02 (not reported in the plots); the FPR was always between 3% and 5%, without any relation to SNR and L (not reported in the plots).

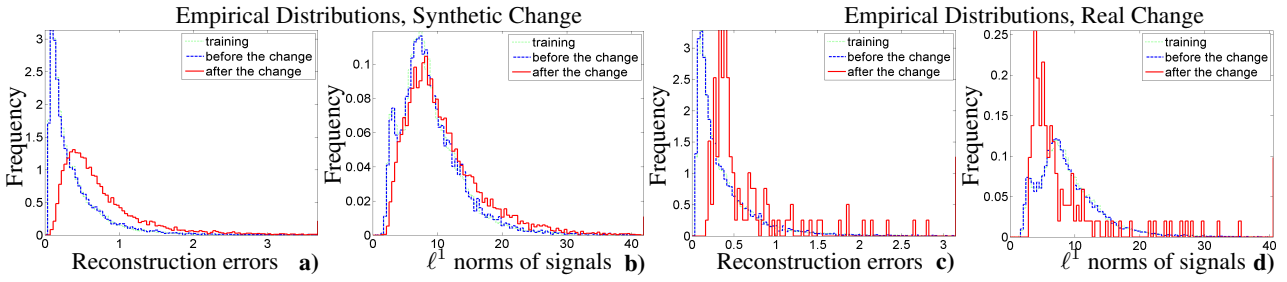


Fig. 2. Experiments on bursts dataset with synthetically induced change: (a), and (b), or simulating a real change by plugging 120 bursts recorded from a different deployment: (c) and (d). The distributions of the reconstruction errors in (a) (KS statistic value 0.42) and (c) (KS statistic value 0.45) clearly highlights the change, while it is more difficult to detect the change from the ℓ^1 norms of the input signals (b) (KS statistic value 0.12) and (d) (KS statistic value 0.23).

(T_0), 10000 were left untouched (S_0) and the remaining 10000 (S_1) were modified by summing an element of Daubechies-4 wavelet basis scaled by a random value chosen between 10% and 50% of burst amplitude. We experimentally set $\hat{L} = 4$ for both K-SVD and OMP, and overcompleteness of \hat{D}_0 to 2.5.

We generated 40 sequences of signals as in Section 3.1, and we achieved very satisfactory change-detection performance when monitoring the reconstruction error since $\text{FPR} = 0\%$, $\text{DL} = 25.52$ and $\text{FNR} = 0\%$ (the KS distance was 0.42). When applying the same CPM_L on the ℓ^1 norms of the input signals, i.e., $\|s_i\|_1$, the performance dropped to $\text{FPR} = 0\%$, $\text{MD} = 124.71$ and $\text{FNR} = 33.33\%$, (KS distance was 0.12), and similarly when testing CPM_L on the ℓ^2 norms of the input signals, i.e., $\|s_i\|_2$, where $\text{FPR} = 9\%$ $\text{FNR} = 91\%$ and no detection provided (KS distance was 0.03). Figure 2 illustrates the empirical distributions over T_0 (training), S_0 (before the change) and S_1 (after the change) for both the e_i (Fig. 2 a) and $\|s_i\|_1$ (Fig. 2 b). These results confirm the effectiveness of the proposed approach on real data, as it allows the detection of changes not be detectable by monitoring the ℓ^1 or ℓ^2 norms in the signal domain.

Finally, we run an additional experiment where S_1 consists of 120 signals acquired from a similar sensing unit deployed in a completely different location (i.e. generated from a different process), thus simulating a real change by concatenating signals from the two process (only one sequence was available for change detection). In this case the change was promptly detected, with a delay of 40 signals, and the empirical distributions of the restoration errors (Fig. 2

c) clearly highlights the change. This change was not detected by monitoring the ℓ^1 norms of the signals (see also the empirical distributions in Fig. 2 d).

4. CONCLUSIONS

We approach the problem of detecting changes in processes generating signals by learning a dictionary yielding sparse representation of signals generated in stationary conditions. Changes are then detected online by performing sparse coding of each input signal, and monitoring the reconstruction error with a sequential change-point method. The solution has been successfully tested on simulated and real changes in bursts acquired from a rockfall monitoring system.

Ongoing work includes the study of alternative sparse coding and dictionary learning methods (e.g. based on ℓ^1 regularization) and alternative choices of sparsity-related features for change detection. We will also investigate extensions of the proposed approach to include structured sparsity for multichannel monitoring system.

5. REFERENCES

- [1] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
- [2] C.-S. J. Chu, M. Stinchcombe, and H. White, "Monitoring

- structural change,” *Econometrica*, vol. 64, no. 5, pp. 1045–65, September 1996.
- [3] C. Alippi, R. Camplani, C. Galperti, A. Marullo, and M. Roveri, “A high-frequency sampling monitoring system for environmental and structural applications,” *ACM Trans. Sen. Netw.*, vol. 9, no. 4, pp. 41:1–41:32, Jul. 2013. doi:10.1145/2489253.2489258
 - [4] C. R. Farrar and K. Worden, “An introduction to structural health monitoring,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 303–315, 2007. doi:10.1098/rsta.2006.1928
 - [5] M. Bocca, L. Eriksson, A. Mahmood, R. Jntti, and J. Kulaa, “A synchronized wireless sensor network for experimental modal analysis in structural health monitoring,” *Comp.-Aided Civil and Infrastruct. Engineering*, no. 7, pp. 483–499, 2011. doi:10.1111/j.1467-8667.2011.00718.x
 - [6] A. M. Bruckstein, D. L. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009. doi:10.1137/060657704
 - [7] B. Mailhé and M. D. Plumbley, “Dictionary learning with large step gradient descent for sparse representations,” in *Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Computer Science, F. J. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, Eds. Springer Berlin Heidelberg, 2012, vol. 7191, pp. 231–238. doi:10.1007/978-3-642-28551-6_29
 - [8] D. M. Hawkins, P. Qiu, and C. W. Kang, “The changepoint model for statistical process control,” *Journal of Quality Technology*, vol. Vol. 35, No. 4, pp. 355–366, 2003.
 - [9] E. J. Candès, “Compressive sampling,” in *Proceedings of the International Congress of Mathematicians*, August 2006, pp. 1433–1452. doi:10.4171/022-3/69
 - [10] V. Cevher, “Learning with compressible priors,” in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, Eds., 2009, pp. 261–269.
 - [11] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, “Sparse Signal Detection from Incoherent Projections,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, 2006. doi:10.1109/ICASSP.2006.1660651
 - [12] J. Haupt and R. Nowak, “Compressive sampling for signal detection,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, 2007, pp. III–1509–III–1512. doi:10.1109/ICASSP.2007.367135
 - [13] Z. Wang, G. R. Arce, and B. M. Sadler, “Subspace compressive detection for sparse signals,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008, pp. 3873–3876. doi:10.1109/ICASSP.2008.4518499
 - [14] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, “Signal processing with compressive measurements,” *Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, Apr. 2010. doi:10.1109/JSTSP.2009.2039178
 - [15] B. S. M. R. Rao, S. Chatterjee, and B. E. Ottersten, “Detection of sparse random signals using compressive measurements,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2012, pp. 3257–3260. doi:10.1109/ICASSP.2012.6288610
 - [16] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009. doi:10.1145/1541880.1541882
 - [17] M. Aharon, M. Elad, and A. M. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006. doi:10.1109/tsp.2006.881199
 - [18] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, November 1993, pp. 40–44. doi:10.1109/acssc.1993.342465
 - [19] B. Wohlberg, “Noise sensitivity of sparse signal representations: Reconstruction error bounds for the inverse problem,” *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3053–3060, Dec. 2003. doi:10.1109/TSP.2003.819006
 - [20] G. J. Ross, D. K. Tasoulis, and N. M. Adams, “Nonparametric monitoring of data streams for changes in location and scale,” *Technometrics*, vol. 53, no. 4, pp. 379–389, 2011. doi:10.1198/TECH.2011.10069
 - [21] G. Ross and N. M. Adams, “Two nonparametric control charts for detecting arbitrary distribution changes,” *Journal of Quality Technology*, vol. Vol 44, No. 22, pp. 102–116, 2012.
 - [22] Y. Lepage, “A combination of Wilcoxon’s and Ansari-Bradley’s statistics,” *Biometrika*, vol. Vol. 58, No. 1, pp. 213–217, April 1974.
 - [23] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blak, and H. Kim, “Detection of intrusions in information systems by sequential change-point methods,” *Statistical Methodology*, vol. 3, no. 3, pp. 252 – 293, 2006. doi:10.1016/j.stamet.2005.05.003
 - [24] C. Alippi, G. Boracchi, and M. Roveri, “A Just-In-Time adaptive classification system based on the Intersection of Confidence Intervals rule,” *Neural Networks*, vol. 24, no. 8, pp. 791 – 800, 2011. doi:10.1016/j.neunet.2011.05.012
 - [25] I. Daubechies, *Ten lectures on wavelets*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992.
 - [26] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Academic Press, 2008.