B. Project leader: Bogdan Dumitrescu

B1. Important scientific achievements of the project leader (maximum 2 pages)

From 1998 to 2010. Most results are gathered in the monograph *Positive trigonometric polynomials and signal processing applications*, Springer 2007 (revised 2017). The book has more than 300 GS citations and 16k downloads of the electronic versions, is cited by significant researchers and is characterized as "a goldmine of many elegant results" in [29]. Some relevant results:

- The parameterization of *positive trigonometric polynomials* through positive definite matrices (2001, *IEEE Transactions on Signal Processing*), allowing the use of semidefinite programming for solving optimization problems with such polynomials. The parameterization is exact and efficient, while earlier methods either were approximate or had high complexity. I extended the result to the multidimensional case, for bounded regions, in another *IEEE TSP* paper (2006).
- A *Bounded Real Lemma* for FIR systems in linear matrix inequality (LMI) form, including the multidimensional case. This unique type of result has been essentially used in, among others, the seminal highly-cited paper of super-resolution theory based on atomic norm [3]. The result appeared in 2005 in *IEEE Signal Processing Letters*, but is cited mostly through the book.
- The use of convex optimization techniques for the design of oversampled filter banks. This work has led to two journal papers and three patents together with Nokia Research Center, Tampere, Finland, and is effectively used in Nokia/Lumia and Huawei phones.

Since 2010 I worked on topics relevant to the current project. Contributions to the general topic of sparse representations:

- *Online* (adaptive) algorithms for sparse representations are given in [7,20] and [A1] (see section B2 in this document). Novelty lies in the use of Information Theoretic Criteria (ITC) for sparsity level estimation, a low storage version of orthogonal triangularization with permutations and a compact and implicit application of Householder reflectors.
- An algorithm for sparse *total least squares* (TLS) was presented and analyzed in [5].
- In [A6], where the *matching pursuit* algorithm was analyzed and reinterpreted from a statistics viewpoint and ITC had an important role, I contributed with an efficient online algorithm for computing the degrees of freedom associated with the representation.

On the topic of dictionary learning, to which this project belongs, I worked with my former PhD students Cristian Rusu and Paul Irofti. The main results are:

- Algorithms for finding the true (for synthetic data) or efficient (for real data) *dictionary size* were proposed in [25] and [A5]. In both of them the size is adapted during the learning process, using ITC in [A5].

- The design of *incoherent frames* [A4] can also be mentioned, although adaptation to data is not performed. I proposed a simple and computationally efficient algorithm, that is still the best for dictionaries with a large number of atoms.

- *Regularized K-SVD* [A3] is a very simple (and exact) version of the well known K-SVD [1] when regularization is added to the optimization objective, with benefits to the overall representation. Unlike its predecessors, the proposed algorithm has the same complexity as K-SVD.

The monograph *Dictionary Learning Algorithms and Applications*, published by Springer in 2018, contains these results and many others, giving an algorithmic view of the DL problem. It is the first book covering the whole field. It contains all significant approaches and forms of the DL problem, starting from the standard one and going to kernel DL. MATLAB implementations corresponding to all algorithms from the book are freely given. It has more than 8k downloads.

Main author. During my whole research career I worked in small groups and I enjoyed much freedom, often deciding the direction and the means of obtaining the results. As a result, I am main author of most of my publications, in particular of those cited in this project.

Citations. Although my work has not a very high number of citations (1580 in June 2021), it has been cited by many relevant and well known researchers. More importantly, many times the results were effectively used in their work. As a proof of the quality of my research as perceived by others, here is a list of the most significant names, in decreasing order of their own citation count on Google Scholar: Stephen Boyd (195000), Emmanuel Candès (138000), Lieven Vandenberghe (77000), Robert Heath (74000), Petre Stoica (67000), Michael Elad (67000), Stephen Wright (63000), Francis Bach (49000), Yinyu Ye (46000), Lihua Xie (40000), José Principe (39000), Venkataramanan Balakrishnan (38000), Erik G. Larsson (38000), Yonina Eldar (37000), Benjamin Recht (36000), Zhi-Quan Luo (35000), Babak Hassibi (35000), P.P. Vaidyanathan (33000), Truong Nguyen (27000), Nicholas J. Higham (25000), Rogelio Lozano (23000), Pablo Parrilo (22000), Pierre Apkarian (20000), Geert Leus (20000), Nikolaos Sidiropoulos (19000), Paul van Dooren (17000), Ivan Selesnick (16000). Note the range of the main research interests of these researchers: signal processing, optimization, machine learning, control, communications, numerical methods, applied mathematics, and sparse representations. Three of them were my coauthors (Lieven Vandenberghe, Petre Stoica, Ivan Selesnick).

B2. The visibility and the impact of the scientific contribution of the project leader

a. Number of citations: >600 WoS, 700 Scopus, 1580 GS

b. Hirsch index: 14 (WoS), 15 (Scopus), 20 (GS)

c. Brainmap link: U-1700-039W-5496

d. Profiles. Publons (ResearcherID): B-5839-2011, Google Scholar (GS): QEf1T4gAAAAJ, Scopus: 6603839944.

e. Representative publications.

Articles. Since 2013 I have published 21 papers in journals indexed in WoS. Excepting [A2], the papers given below have either dictionary learning as a topic or, more generally, deal with sparse representations.

Identification data	**A1**. A.Onose, B.Dumitrescu. *Adaptive Randomized Coordinate Descent for Sparse Systems: Lasso and Greedy Algorithms*, IEEE Trans. Signal Proc., vol.63, no.15, pp.4091-4101, Aug. 2015. Q1
Main author?	YES (corresponding author)
In project domain?	YES
Number of citations	2
DOI	10.1109/TSP.2015.2436369

Identification data	**A2**. B.Dumitrescu. *On the relation between the randomized extended Kaczmarz algorithm and coordinate descent*, BIT Numerical Mathematics, vol.55, no.4, pp.1005-1015, 2015. Q1
Main author?	YES
In project domain?	NO
Number of citations	4
DOI	10.1007/s10543-014-0526-9

Identification data	**A3**. B.Dumitrescu, P.Irofti. *Regularized K-SVD*, IEEE Signal Processing Letters, vol.24, no.3, pp.309-313, Mar. 2017. Q1
Main author?	YES
In project domain?	YES
Number of citations	13
DOI	10.1109/LSP.2017.2657605

Identification data	**A4**. B.Dumitrescu. *Designing Incoherent Frames With Only Matrix-Vector Multiplications*, IEEE Signal Processing Letters, vol.24, no.9, pp.1265-1269, Sep. 2017. Q1
Main author?	YES
In project domain?	YES
Number of citations	6
DOI	10.1109/LSP.2017.2723044

Identification data	**A5**. B.Dumitrescu, C.D.Giurcaneanu. *Adaptive-Size Dictionary Learning Using Information Theoretic Criteria*, Algorithms, vol.12, no.9, 13 pages, art.no. 178, Sep. 2019
Main author?	YES
In project domain?	YES
Number of citations	2
DOI	10.3390/a12090178

Identification data	**A6**. F. Li, C.M. Triggs, B.Dumitrescu, C.D.Giurcaneanu. *The matching pursuit algorithm revisited: A variant for big data and new stopping rules*, Signal Processing, vol.155, pp.170-181, Feb. 2019. Q1
Main author?	NO
In project domain?	YES
Number of citations	1
DOI	10.1016/j.sigpro.2018.09.033

Identification data	**A7**. A.Băltoiu, B.Dumitrescu. *Sparse Bayesian Learning Algorithm for Separable Dictionaries*, Digital Signal Processing, vol.111, art.no.102990, Apr. 2021. Q2
Main author?	YES
In project domain?	YES
Number of citations	0
DOI	10.1016/j.dsp.2021.102990

Monographs. I have published two monographs since 2013. One is the revised edition of *Positive Trigonometric Polynomials and Signal Processing Applications* (Springer 2007, 2017), a book with many citations (more than 300 on Google Scholar). The second, given below, is directly related to the topic of this project and is an overview of dictionary learning; it gathered almost 50 citations on GS.

Identification data	**M1.** B. Dumitrescu, P. Irofti. *Dictionary Learning Algorithms and Applications*, Springer, 2018.
In project domain?	YES
No. of libraries (KVK)	176
University libraries	Technische Universität München, ETH-Bibliothek Zurich, Delft University of Technology, MIT Libraries, Princeton University, University of Illinois at Urbana Champaign, Purdue University, Texas A&M University, Virginia Tech, University of Auckland

Patents. The following patent belongs to a group of patents on audio signal processing, to which I contributed with design algorithms for oversampled filter banks. They were results of cooperation with Nokia Research Center, Tampere, Finland.

Identification data	**P1.** R.Niemistö, R.Bregovic, B.Dumitrescu, V.Myllylä. *Method and an apparatus for processing an audio signal using noise suppression or echo suppression*, US patent 9640187, 2017.
Issuing bureau	US Patent Office
In project domain?	NO

B3. The correspondence between the demonstrated experience of the project leader and the proposed theme (max. 1 page)

Since 2010, sparse representations and dictionary learning (DL) are my main research topic. About half of my papers published since 2010 are dedicated to this topic, see the full list here. Section B1 presents some results of my research in this area. The book [M1] (see section B2 in this documents), to which I am a coauthor, presents the state-of-the-art as it was a few years ago and also contains the contributions obtained by me and my PhD students.

Some of the papers, like [A3,A7], present theoretical results that allow a new or a faster solution to a DL problem. All my papers dedicated to DL contain algorithms as the end result, and implementations are provided or are easy to build. For example, the software for the book [M1] is here and some other programs are here. The same kind of approach will be used for AsyDiL, the current proposal.

Designing algorithms for the problems posed in AsyDiL requires also a good expertise in optimization. Most of my papers, for example [A1-A5, A7] from those listed in section B2, involve optimization in one form or another. In particular, Bayesian learning, which may have a role in this proposal, was used in [A7].

The applications foreseen for this project are in anomaly detection, especially for data organized in graph form and coming from bank transactions. I am the director of an ongoing project on this topic (PED, 2020–2022); the team comprises members from UPB, University of Bucharest, and an industrial partner (Tremend Software Consulting). We have good cooperation with Libra Bank and BRD and benefited from access to real data. Two members of the team are projected to work in AsyDiL. We have thus acquired good expertise in anomaly detection and testing new algorithms is a clearly manageable task.

Although I consider that my experience fits very well the proposed theme, I wouldn't say that the problems that will be tackled are entirely in my comfort zone. There are still many challenging aspects and some uncertainties on the final outcome. So, the project looks worth the effort.

B4. Curriculum Vitae (maximum 2 pages)

Full name: **Bogdan-Alexandru Dumitrescu**

Date/place of birth: May 7,1962, Bucharest, Romania

Publons (ResearcherID): B-5839-2011, Google Scholar: QEf1T4gAAAAJ, Scopus: 6603839944, UEF-ID: U-1700-039W-5496

EDUCATION

- 1993: Ph.D. degree at the University Politehnica of Bucharest (UPB). Thesis title: ”Parallel computation systems and algorithms”.
- 1987: M.Sc. degree at the Faculty of Automatic Control and Computers (ACC), UPB.

POSITIONS

- 2003 – present: professor at the Department of Automatic Control and Systems Engineering, ACC, UPB.
- 1998 – 2015 (part-time). Visiting researcher at Tampere University of Technology (TUT), Department of Signal Processing, Finland.
- 2010 – 2013: FiDiPro (Finnish Distinguished Professor) fellow, TUT.
- 1990 – 2003: various teaching positions at ACC, UPB. Visiting stages at Universite Polytechnique de Grenoble, France, for a total duration of about 15 months.
- 1987 – 1990: software engineer at FEPER Bucharest (Peripheral Equipments Factory).

ACTIVITY

Teaching. At UPB I currently teach Numerical Methods (2nd year), Scientific Writing and Computation for Complex Systems (master) and I taught several other courses like Parallel Algorithms, Signal Processing, Scientific Computation, Advanced Signal Processing, Convex Optimization, Control Systems. PhD advisor since 2007.

Research. Main current interest: numerical methods and optimization in signal processing, sparse representations and related problems like dictionary learning. Other interests: parallel algorithms, optimization with positive polynomials.

Publications. More than 50 journal articles, some in top 25% based on current AIS: IEEE Trans. on Signal Processing (9 articles), IEEE Signal Processing Letters (10), Signal Processing (10), IEEE

Trans. on Circuits and Systems I (1), Computers Graphics Forum (1), Optimization and Engineering (1), IEEE J. Selected Topics in Signal Processing (1). Other articles in IEEE Trans. on Circuits and Systems II (2), Digital Signal Processing (2), Linear Algebra and Its Applications (1), BIT Numerical Mathematics (1), Numerische Mathematik (1). I was principal author of most of these articles. About 90 conference papers. 3 international patents (two by US patent office). 5 books (3 in Romanian, 2 in English), including the monographs *Positive Trigonometric Polynomials and Signal Processing Applications*, Springer, 2007 (2nd edition 2017) and *Dictionary Learning Algorithms and Applications* (with P.Irofti), Springer, 2018. 3 book chapters, in books published by Springer, CRC Press and Diderot (in French). Complete list here.

Hirsch index: 14 WoS, 15 Scopus, 20 Google scholar

Citations: 600 WoS, 700 Scopus, 1580 Google scholar (June 2021)

Grants and contracts. ”Sparse representations and signal processing applications”, IDEI 2011-2016. ”Positivity in the analysis and synthesis of multidimensional systems”, IDEI 2007-2010. Both are PCE-type grants for which I was the director. PED (experimental demonstration) project ”Graphomaly – software package for anomaly detection in graphs modeling financial transactions”, PN-III-P2-2.1-PED-2019-3248, 2020–2022; I am the director and the parties involved are UPB, the University of Bucharest and Tremend Software Consulting SRL. Several industry contracts for filter bank optimization with Nokia Research Center (2004–2009), Microsoft Finland (2015), Huawei Finland (2020).

Other activities. Associate editor (2008-2012) and area editor (2010-2014) at IEEE Transactions on Signal Processing; the duration of these tasks is exceptionally long, since a regular AE term has a two-years length, with an optional third. Associate editor (2015-) at Mathematical Problems in Engineering (Hindawi). Associate editor (2018-) at Algorithms (MDPI). Reviewer for more than 20 journals (IEEE Trans. Signal Processing, IEEE Signal Processing Letters, IEEE Trans. Circuits and Systems, Signal Processing, Automatica, etc.). Technical Program Committee member at several editions of EUSIPCO since 2007 and at other signal processing conferences, like IEEE Statistical Signal Proc. Workshop (2021, 2018, 2016), ICUWB 2015, Globecom 2014, CISP-BMEI (several editions since 2012).

Visiting professor. The high point of the many years spent in Finland was the FiDiPro position, through a program in which top international researchers were invited to work about half of the time in a Finnish university. In the last few years I was three times invited at the University of Auckland, New Zealand (2017, 2018, 2019) for periods of 2-3 weeks. I gave a presentation on dictionary learning. In 2016, at Tampere University of Technology, I gave a two-weeks course on dictionary learning. Participations in thesis committees at CentraleSupélec and Lille.

Spoken languages: English, French, and a bit of Finnish.

C. Funding application (maximum 11 pages)

Title: **Asymmetric Dictionary Learning (AsyDiL)**

C1. Motivation of the proposed theme in the current scientific context. Originality and degree of innovation

Introduction. Sparse representation (SR) [34] is the technique with most impact in signal and image processing in the latest 10-15 years, changing the paradigm in many operations like denoising, inpainting, compression, coding, compressed sensing, and also for machine learning tasks like classification. The range of applications is simply stunning, due especially to SR versatility and to their natural property of giving a parsimonious model. Here are a few: face recognition, music analysis, MRI denoising, forgery detection in paintings, seismic signals denoising, landmine classification, etc.

Dictionary learning (DL) [6,23,31] enhances the power of sparse representations. Given a collection $\boldsymbol{Y} \in \mathbb{R}^{m \times N}$ of N training signals of length m, the standard DL problem consists of finding an overcomplete dictionary $\boldsymbol{D} \in \mathbb{R}^{m \times n}$ ($m < n$) minimizing the representation error $\|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F$, with the condition that the representation matrix $\boldsymbol{X}$ is sparse, typically with s elements on each column, with $s \ll n$. So, a column $\boldsymbol{y}$ of $\boldsymbol{Y}$ is a linear combination of s atoms (columns of $\boldsymbol{D}$). It was proved that learned dictionaries have clear advantages over fixed ones in most SR applications.

Our problem and its importance. In SR and DL, an atom is a vector. Representations are obtained by linear combinations of atoms and thus they reside in a union of low dimensional subspaces. This is a simple model, which is both intuitive and rich.

The core of our proposal is to essentially change the notion of atom, using an infinite set instead of a vector. We replace an atom with an *atom-set*. For the purpose of illustration, consider a single signal $\boldsymbol{y}$ that we want to approximate with a sparse representation $\boldsymbol{r}$ (we also call $\boldsymbol{r}$ the recovered signal). In standard SR we have to find the support $\mathcal{S}$ (the s atoms that are chosen for the representation) and to express

$$\boldsymbol{r} = \sum_{j \in \mathcal{S}} x_j \boldsymbol{d}_j = \boldsymbol{D}_{\mathcal{S}} \boldsymbol{x}_{\mathcal{S}}, \tag{1}$$

where $\boldsymbol{d}_j$ is an atom and the coefficients x_j are usually determined using an optimality measure like $\|\boldsymbol{y} - \boldsymbol{r}\|$. Once the support is chosen, which is an NP-hard problem, the coefficients can be computed by solving a least squares problem.

We propose a representation having the form $\boldsymbol{r} = f(\boldsymbol{D}, \boldsymbol{y})\boldsymbol{x}$, where f is a function that has no explicit expression. This is still a sparse linear representation, as the vector $\boldsymbol{x}$ is sparse like in the standard model. Instead of an atom $\boldsymbol{d}_j$, we have now an atom-set $\mathcal{C}(\boldsymbol{d}_j)$. As a simple example, imagine that instead of a single vector, we have a cone with given radius around a central vector. The representation is now

$$\boldsymbol{r} = \sum_{j \in \mathcal{S}} x_j \tilde{\boldsymbol{d}}_j = f(\boldsymbol{D}_{\mathcal{S}}, \boldsymbol{y})\boldsymbol{x}_{\mathcal{S}}, \quad \text{with } \tilde{\boldsymbol{d}}_j \in \mathcal{C}(\boldsymbol{d}_j). \tag{2}$$

The actually used atoms $\tilde{\boldsymbol{d}}_j$ and their coefficients are chosen to optimize the residual $\|\boldsymbol{y} - \boldsymbol{r}\|$, like in the standard case. This representation is still linear, once the atoms $\tilde{\boldsymbol{d}}_j$ have been chosen. However, there is essential nonlinearity in the representation since the atoms are chosen from a set and so are not fixed. The dictionary is no longer a matrix of size $m \times n$; for each signal, we have an $m \times s$ matrix that is specific to the signal.

So, the standard DL model is considerably enriched by our model. Moreover, further nonlinearity can be added, since kernel techniques [19] or their approximations [9] are still applicable to our model. We can thus forecast that our model has the potential of clearly higher performance than standard DL.

An apparent drawback of our model is the lack of the reconstruction capability. Although there are not yet available algorithms, finding a good representation (2) for a given signal $\boldsymbol{y}$ is certainly possible. However, given only the dictionary (i.e., a description of the atom-sets) and the representation $\boldsymbol{x}$, it is impossible to build $\boldsymbol{r}$, since one cannot know what specific atoms $\tilde{\boldsymbol{d}}_j$ have been used if the target $\boldsymbol{y}$ is not known. Of course, storing all atoms $\tilde{\boldsymbol{d}}_j$ explicitly is impractical. So, applications in coding or compressed sensing are impossible. However, there are many other important applications where reconstruction is not needed, like denoising, inpainting, classification and especially (for this project) anomaly detection. Here, given signal $\boldsymbol{y}$, we build the approximation $\boldsymbol{r}$, which is used directly in denoising or inpainting; in anomaly detection, the error $\|\boldsymbol{y} - \boldsymbol{r}\|$ can be used as an anomaly score, although more involved methods can be imagined; in classification, both the error and the representation coefficients can be put at work. Due to the one-way construction mode, we name *Asymmetric Dictionary Learning* (AsyDiL) our model.

Difficulties. There are two main issues regarding difficulty: computational complexity and the sheer scientific complexity of solving the proposed problem.

Although the standard DL problem is NP-hard due to the sparsity constraint and has a large size, very good approximate solutions are given by algorithms like K-SVD [1] and its improved versions (among which our [A3,A5], see section B2 in this document). For moderate number of atoms ($n < 1000$) and number of signals ($N < 10000$) most methods give an optimized dictionary in under a minute on an ordinary computer. Although some questions appear to be still open, like finding the optimal size of the dictionary, the standard DL problem seems largely solved.

More challenging are variations of the problem where the dictionary is endowed with a structure (e.g., Kronecker product for multidimensional signals, block orthogonality, sparsity), has special properties (e.g., coherence bound, shift invariance) or the objective is altered to favor the application goal, like in classification based on DL. Even more difficult are problems where nonlinearity is introduced, like in kernel DL [19], where the problem size may increase considerably.

AsyDiL appears to be situated in the higher range of complexity, due the necessity of choosing optimal representations vectors from atom-sets, so the main challenge is to find algorithms for both representation and dictionary learning, including the case of nonlinear extensions, that can compete with their standard DL counterparts in terms of performance and complexity.

State-of-the-art and current limitations. There is no DL algorithm using atoms other than in vector form, although the number of atoms can be enlarged through, for example, implicit shifting. We thus review directions of DL research appealing to other means to achieve atom diversity. Sparse representations have a fundamental limitation, which is their linearity. Several extensions towards nonlinearity have been made. Kernel DL [19] adapts the usual kernel approach to DL, with the price of considerably increased complexity. Nyquist sampling [9] was used for (re)linearization and dimensional reduction; this method can be seen as an heuristic for building feature vectors from the training signals, mimicking the use of a kernel. Kernel optimization can be achieved by combining (fixed) kernels corresponding to several descriptors of the original signal [30] or directly [32] (deep kernel learning) by learning the kernel function. An alternative is the explicit construction of a feature function using neural networks [11]. It is noteworthy that most of these techniques are likely to be combined with atom-sets.

Going on to more nonlinear DL generalizations, the DL problem was moved from Euclidian space to a Riemannian manifold [10]. In [26] the signals are histograms in the probability simplex and linear combinations of atoms are replaced by the computation of the Wasserstein barycenter, with distances measured by optimal transport. A line of work that aims to connect DL with neural networks originates from multi-layer DL [16,24] where the dictionary is a product of (originally sparse) matrices; optimization algorithms are proposed in [27], [18], the latter with the advantage of learning also the classifier; the name ”deep DL” was used (somewhat prematurely) in [28] for a similar product structure, in which optimization is made in a greedy, clearly suboptimal way. The most general architecture builds on convolutional sparse coding [33], where the signal is not split into smaller m-length vectors like in standard DL, but treated into its integrality. The scheme from [21] inserts also thresholding between the layers of dictionaries, thus effectively introducing nonlinearity in a style that is similar to convolutional neural networks.

DL as a linear representation tool is close to fulfilling its potential. The existing extensions to nonlinearity show better performance in some cases, although not in general; some are intuitive but cumbersome (kernel DL), other share the lack of transparency and intuitiveness of neural networks. There is clearly space for new structures like AsyDiL, that seek nonlinearity in a way closer to the DL frameworks.

Anomaly detection (AD) is the main application of our DL algorithms especially for data in graph form obtained from banking transactions, with the purpose of discovering money laundering schemes and other frauds. There are many AD methods, so we mention mostly work related to graphs; two good sources of information are the reviews [2,14]. Here are some recent methods. Feature extraction, with the direct purpose to build an anomaly score, is proposed in [8]. DL only starts to be used in AD, some examples being [13,15]. Other fraud detection method are given in [4,17] the latter treating incremental learning (online algorithms). Many AD algorithms, very useful for comparisons, can be found in the PyOD library [35]. It is interesting that AD methods tend to behave better for some data sets and worse on others. There is no absolute champion. This is also true in the case of graph data and the best combination of feature extraction and AD algorithm is yet to be found. The scarcity of public bank transactions data is also a factor, but we have at least a source (see below).

Originality. While the general idea of dictionary learning is well established and AsyDiL does not deviate from its basic algebraic structure, the proposed atom-sets construction is clearly not explored. It can be seen as a new direction for extending the representation ability of a dictionary. Its asymmetry, in the sense that it is not possible to find which representative of an atom-set is used in a given representation, is actually not an impediment in many applications and opens the way to a novel DL interpretation. This approach is challenging both theoretically and algorithmically, and promises significant flexibility improvement.

Relation with previous work. I am the director of an ongoing project called ”Graphomaly – software package for anomaly detection in graphs modeling financial transactions” (August 2020 – March 2022). Among other methods, DL is adapted to anomaly detection, using small variations of its standard form. None of the techniques proposed in AsyDiL was tried. However, the experience in anomaly detection in graphs gained in Graphomaly will be very helpful in developing and testing the new algorithms using the new DL framework proposed in AsyDiL.

C2. Objectives, methodology and work plan

C2.1. *Objectives*

O1. Computation of asymmetric representations

Typical DL algorithms have iterative form and each iteration contains two operations: i) sparse representation, where the representations $\boldsymbol{X}$ are computed with the current dictionary $\boldsymbol{D}$, and ii) dictionary update, where the dictionary is optimized for the current representations. Due to its important significance and possible independent use, sparse representation is examined here separately.

O1.1. Sparse representations with uniform atom-sets. We assume that the atoms in the atom-set are equally likely and so we use the atom that is best for minimizing the error. At least 2-, 1-, and ∞-norm cones will be examined as atom-sets shapes.

O1.2. Sparse representations with probabilistic atom-sets. An atom-set can be interpreted as a probability density function. The sparse representation is made with vectors having each a certain probability to belong to the atom-set. The optimization objective is the representation error subject to a minimum norm of the associated probabilities. Note that this kind of representation is not similar to Gaussian mixtures or to fuzzy sets.

O1.3. Kernel versions of the above representations will be developed for adding further nonlinearity and modeling abilities.

O2. Algorithms for asymmetric dictionary learning

As sparse representation algorithms are now available, DL algorithms need only atom and possibly representation optimization for a fixed support; this is the dictionary update operation. Assuming that the structure of an atom-set is fixed, the issue is to optimize its parameters: position (the ”central” vector), size (e.g., cone radius) and possibly shape (e.g., ellipsoid axes and orientation, if we generalize a cone). Here, in most cases, the actually-used atoms from the atom-sets must be available or at least some partial information on them. This is not necessarily a problem, since this information is stored only internally in the DL process. Note that the objectives listed below apply to the representation frameworks of both objectives **O1.1** and **O1.2**, with the relevant specific.

O2.1. Dictionary update based on heuristics. Atom-sets can be improved with heuristics based on actually-used atoms, like averaging or using ideas from clustering. The possible advantage of such methods is that they can be incorporated in the representation algorithm and thus eliminate the need of any side information. They also can be very fast, at the level of standard DL algorithms. The main target is a combination with **O1.1** to get a full DL algorithm as first result of AsyDiL.

O2.2. Dictionary update using optimization. For developing true optimization methods, the challenge is to reach acceptable complexity, as the problem has large size. Candidate optimization methods are given in Section C2.2. The main goal is to get DL algorithms in combination with **O1.2**.

O2.3. Kernel AsyDiL. In direct connection with **O1.3**, kernel DL algorithms will be developed. Besides the extension of the algorithms from **O2.1** and **O2.2** to kernel form, we will try to simplify the computational burden by working with low-dimensional kernels, based on a selection of vectors.

O2.4. Optimization of atom-set parameters. Optimizing only the central atom of an atom-set may not be enough if the other parameters describing the atom-set are fixed. An atom-set can be described by a number of parameters: size (radius), shape (probability distribution), spatial aspect ("covariance" matrix, for a quadratic). In search of the best trade-off between complexity and quality, we will develop optimization methods for variate numbers and types of parameters.

O2.5. MATLAB and Python libraries will be developed, including all methods designed in this project. They will be published in at least two stages. Appropriate software will be published in conjunction with submitted articles, in the spirit of reproducibility.

O3. Applications to anomaly detection

Although we plan to illustrate our new methods with other applications, like denoising, the main focus will be on anomaly detection.

O3.1. Algorithm design and hyper-parameter tuning. The application of asymmetric DL to anomaly detection requires a basic DL structure to produce anomaly scores based on representation errors or atom-sets employment. The learning process may be different than in the general case, for example by using only a random sample of the signals at each iteration and using for atom updates only signals with small errors (to encourage good representations of normal signals and poor representations of anomalies). Also, 1-norm (or even lower) of the error may better than 2-norm, such that outliers have less weight in the objective function. So, we have to optimize not only some parameters of the atom-sets, but also hyper-parameters of the whole learning-for-anomaly-detection process.

O3.2. Anomaly detection for public data benchmarks. The success of an anomaly detection method is given by its success in practice. There are several popular databases, like ODDS and some insight is necessary to determine why a method is successful for some data and may fail for others. This insight is gained through experiments, for which we provide ample time.

O3.3. Anomaly detection for bank transactions data and other graph data. Graph data will be tackled by the classical procedure of feature extraction, but also by modeling sub-graphs topologies and working directly on them. The purpose is to obtain good detection ability and low-complexity algorithms.

C2.2. ***Methodology***

The first step is to define clearly the atom-set structures and a hierarchy of parameters associated to each structure, from the simplest to the most complex form. Optimization problems will be naturally associated with each representation and DL structure. As the optimization problems are nonconvex, NP-hard due to sparsity constraint, and have large size, the optimization techniques that will be used are simple: block coordinate descent (like in K-SVD and many other DL algorithms), (accelerated) gradient descent, proximal methods [22], and possibly ADMM. DL experience says that, to avoid early stops in local minima, exact optimization in each DL iteration is neither necessary nor desirable.

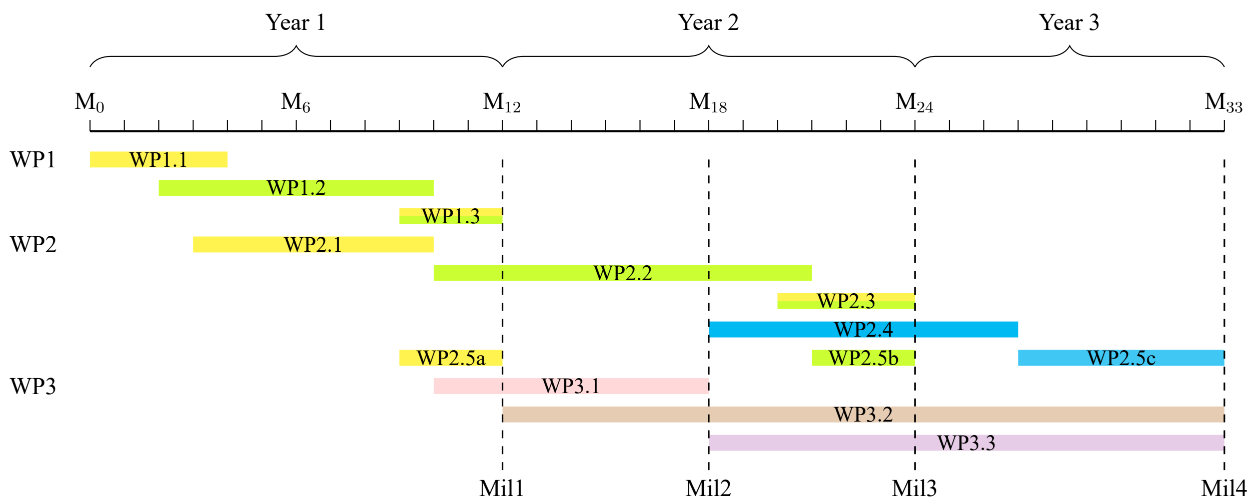
However, careful implementation is necessary and it is important to have efficient basic functions in MATLAB (mex files) and Python as early as the first tests show that a method has potential. The possibility of testing large size problems depends crucially on this issue.

Tests will be conducted on noisy artificial data, generated with atoms randomly selected from atom sets. Besides evaluating the representation error, we will also tackle standard problems like denoising. Once a method passes these tests, we will go to anomaly detection, on data that are publicly available (ODDS, again) or come from our bank sources.

We will also try to derive theoretical properties of the problems and their solving algorithms, regarding the properties of the sparsest solution, complexity, convergence, and guarantees of success.

C2.3. *Work plan*

The figure below shows the time organization of our project. Work packages (WP) correspond to objectives and have the same numbers.



WPs represented with the same color follow logically one another. WP2.5 is split into three parts, corresponding to the sections of the library that correspond to distinct DL structures and optimization problems. Anomaly detection tests will probably start earlier, when sufficient advance in DL algorithms is made; the start time from the diagram is the latest acceptable.

An activity that cannot be located very precisely in time is WP4–dissemination. We will write technical reports during each WP, most likely towards their end, and also articles for submission. Besides scientific WPs, we should add a permanent WP5 for administrative tasks.

The four milestones are meant for counting the minimum acceptable achievements: submitted articles, published software, etc. The milestones are as described in next section. Mil1: **D1–D5**. Mil2: **D6–D7**. Mil3: **D8–D9**. Mil4: **D10–D12**.

C2.4. *Deliverables*

- **D1** Project website, periodically updated. [M2]
- **D2** Technical report for SR algorithms with uniform atom-sets. [M6]
- **D3** Journal article on DL with uniform atom-sets. [M12]
- **D4** MATLAB software for DL with uniform atom-sets (on website). [M12]
- **D5** Conference paper on SR algorithms with probabilistic atom-sets. [M12]
- **D6** Conference paper on DL algorithms with probabilistic atom-sets. [M18]
- **D7** Technical report on anomaly detection using DL. [M18]
- **D8** Journal article and software for DL algorithms with probabilistic atom-sets. [M24]
- **D9** Conference paper on anomaly detection. [M24]
- **D10** Journal article on optimization of atom-sets structures. [M33]
- **D11** Journal article on anomaly detection in graphs. [M33]
- **D12** Complete software package in MATLAB and Python, published on website and github. [M33]

For all articles, the deadline is for submission, not publication. Table 1 shows the correspondence between WPs and the most important associated deliverables.

C3. Project feasibility: available resources, research team structure and preliminary results

Available resources. The team members work in the Department of Automatic Control and Systems Engineering, whose infrastructure is presented at erris.gov.ro/ACSE—UPB (see also acse.pub.ro). The research work will be carried out in room ED206, which is equipped with a graphic station with an 8-core processor, and other equipment like printer, scanner, copying machine. Each team member has a work or personal laptop. Due to its algorithmic nature, this project needs only computers. We need a computing system with a high performance processor like Intel Core i9, with 16 cores, and at least 64G RAM (128G desirable); it will help running more extensive tests with large data sets and thus speeding up our experiments. We also plan to buy 2 usual desktop computers, especially for the junior team members and 3 laptops, to replace some of the (rather old) existing ones.

Team. Besides the project leader (BD), the team will have other four members.

1. Andra Băltoiu (AB), postdoc. She will defend her PhD thesis (supervised by BD) before the end of this year and has good SR [A7] and DL skills [12].

2. Denis Ilie-Ablachim (DI) is a PhD student since 2019, supervised by BD. He is familiar with DL and kernel techniques and is an active member of project Graphomaly.

3. Cristian Zica (CZ) started his PhD with BD in 2020. He has dealt with anomaly detection and graph data.

Work package	Person-month					Deliv.
	BD	AB	DI	CZ	M	
WP1.1 SR for uniform atom-sets	1	1	0	1	0	**D2**
WP1.2 SR with probabilistic atom-sets	1	2	1	1	0	**D5**
WP1.3 Kernel SR	1	0	2	0	0	**D5**
WP2.1 DL based on heuristics	1	2	0	2	0	**D3**
WP2.2 DL using optimization	1	2	2	2	0	**D6**
WP2.3 Kernel DL	1	0	2	0	0	**D8**
WP2.4 Optimization of atom-set parameters	1	2	0	2	0	**D10**
WP2.5 MATLAB and Python libraries	1	1.5	3	2	1.9	**D4**, **D8**, **D12**
WP3.1 Anomaly detection algorithms	1	1	1	1	2	**D7**, **D9**
WP3.2 Tests on public data	0.5	1	1	1	3	**D9**
WP3.3 Tests on bank data (graph form)	0.5	1	2	2	3	**D11**
WP4 Dissemination	2	2	2	2	0	sec.C5
WP5 Management & administrative tasks	1.2	1	0.5	0.5	0	**D1**
Total	**13.2**	**16.5**	**16.5**	**16.5**	**9.9**	

Table 1: Team work distribution.

4. Master student (M). This position will be filled by 2-3 persons, for distinct periods of 4-6 months, for work during the research stages imposed by their master program. They will mainly help with implementation and especially with testing, including comparisons with other methods.

The work/time allocation plan is shown in Table 1.

Preliminary results. Besides the whole theoretical and algorithmic machinery for existing DL problems [M1], we have only some derivations and an algorithmic sketch made by BD, showing that the SR problem with cone atom-sets can be solved in reasonable time.

C4. Risks and alternative approaches

Like in all research activities, the main risk is not failing, but finding something that has already been discovered or that is not relevant. The idea of atom-sets is new, but it remains to prove that it is computationally efficient and better than standard DL in some contexts. Optimization tools are crucial and we will keep open options for other methods than those specified, for example for stochastic gradient and online methods; they can be useful for large data sets. Parallel implementations will be pursued in the early stages of development if the methods are too slow.

Delays and unexpected difficulties will be compensated by in-depth exploration of successful advances. On the other side, to minimize the risk of failure by reaching too far too quickly, we will also explore smooth transitions from standard DL, which is a particular case of AysDil where an atom-set is a single vector, by carefully increasing the ”size” of the atom-sets; such tests help understanding how some representation properties evolve.

A significant risk is that of work force volatility, especially as IT firms offer tempting salaries in Bucharest. DI has a three year assistant position. He and AB have manifested the desire to join our department on a permanent position after PhD defense. CZ seems also decently reliable. As precaution measures, we intend to identify master students that can work in this project and guide them to enroll as PhD students when they get the master degree; one such candidate already exists, working on anomaly detection in graphs. Also, we will create a larger pool of candidates via our collaborators Paul Irofti and Andrei Pătraşcu at the University of Bucharest. To keep team spirit strong, the monthly team meetings on technical issues will be followed by informal discussions and occasional fun activities. The team leader will have at least weekly individual discussions with each team member.

Finally, we will try to attract extra funding to this topic by helping the younger members of the team to apply for postdoc research grants in related topics and especially by seeking co-funding from our bank collaborators and other banks.

C5. Impact and dissemination

Estimated scientific results. If the projected results are obtained, the theoretical and practical advance can lead to publication in top journals. We aim to publish at least four articles in good journals (deliverables **D3**, **D8**, **D10**, **D11**); our first targets are top signal processing journal like IEEE TSP, Signal Proc. or IEEE SPL, and machine learning journals like JMLR. A good venue for the library is the JMLR track for open source software in machine learning. We also plan to have at least six papers at indexed conferences, of which at least three (deliverables **D5**, **D6**, **D9**) at the best signal processing (e.g., ICASSP, EUSIPCO) and machine learning (ICML, ICLR) conferences.

Impact. On the host institution and team members: strengthening our research group and growing the next generation of professors. Hopefully, AB will be able to get a permanent position, DI and CZ (at least one) to defend their PhD thesis and at least a master student to start PhD. Ideally, each team member will be the main author of at least a journal article.

On the economic environment: we expect to continue existing collaboration with Libra bank and BRD to develop new tools for money laundering and fraud detection.

Dissemination. Besides journal articles and participation to conferences, we intend to create ”light” versions of our results, attempting publications on blogs dedicated to signal processing and machine learning or even on popular sites like youtube. The proposed DL library will be made available on the project website in several steps and will contain convincing examples of use. All papers will have the associated code on our website. Locally, the team members will give presentations in the department, and also at common seminaries with our research associates and whenever visiting foreign universities.

C6. Requested budget

The whole budget for 33 months is 1068030 lei (213606 EUR) as detailed below.

Pre-calculation estimate (in lei, per calendar year).

Budget chapter	Year I (lei)	Year II (lei)	Year III (lei)	Total budget (lei)
Personnel expenses	268800	268800	201600	739200
Logistics expenses	50000	30000	10000	90000
Travel expenses	30000	42000	36000	108000
Indirect expenses	45570	48120	37140	130830
Total	394370	388920	284740	**1068030**

Pre-calculation estimate (in Euros, at the project level). 1 EUR = 5 lei.

Budget chapter	Total budget (euros)
Personnel expenses	147840
Logistics expenses	18000
Travel expenses	21600
Indirect expenses	26166
Total	**213606**

Personnel costs. The work load and full-time salaries are given in the next table. The work load over the duration of the project is uniform, although there will be flexibility, especially for M to work more intensely in shorter bursts. The last column shows the full-time salary, hence the actual monthly salary is workload/100*full salary. For BD, this is the main research project in the next years; work will be covered in part by the usual research duties as professor. For AB, DI and CZ, the workload is 50% in order to allow time for possible didactic activities and some research work not related to this project.

Person	Position type	Workload (%)	Man months	Full-time salary (lei/month)
BD	experienced researcher	40	13.2	14000
X	postdoc	50	16.5	12000
DI	PhD student	50	16.5	9000
CZ	PhD student	50	16.5	9000
M	master student	30	9.9	6000

Logistics expenses. As explained in Section C3, we need a high-end computing system (cost: up to 40000 lei), 2 usual desktop computers (cost: 10000 lei) and 3 laptops (cost: 15000 lei). The rest of the money will go to consumables, auxiliary devices (e.g., memory), IEEE member fees (for getting lower registration fees at IEEE conferences) and a reserve for unforeseen situations.

Travel expenses. Participating at conferences and short research stages abroad is extremely important for the development of young researchers and for maintaining the visibility of the older ones. We plan at least 6 conferences abroad, with an average budget of 12000 lei/conf. Depending on the outcome of our research, the remaining 36000 lei will be allocated to conferences in Romania and short stages/visits in foreign universities.

Indirect costs are 15% of the direct expenses (equipment cost removed), as required.

C7. Bibliography

[1] Aharon, M., M. Elad, and A. Bruckstein.“K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation”.In: *IEEE Trans. Signal Proc.* 54.11 (Nov. 2006), pp. 4311–4322.

[2] Akoglu, L., H. Tong, and D. Koutra.“Graph based anomaly detection and description: a survey”.In: *Data mining and knowledge discovery* 29.3 (2015), pp. 626–688.

[3] Candes, E.J. and C. Fernandez-Granda.“Towards a Mathematical Theory of Super-resolution”.In: *Comm. Pure Appl. Math.* 67.6 (June 2014), pp. 906–956.

[4] Carcillo, F., Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi.“Combining unsupervised and supervised learning in credit card fraud detection”.In: *Information sciences* 557 (May 2021), pp. 317–331.

[5] Dumitrescu, B.“Sparse Total Least Squares: Analysis and Greedy Algorithms”.In: *Lin. Alg. Appl.* 438.6 (Mar. 2013), pp. 2661–2674.

[6] Dumitrescu, B. and P. Irofti.*Dictionary Learning Algorithms and Applications*.Springer. 2018.

[7] Dumitrescu, B., A. Onose, P. Helin, and I. Tăbuş.“Greedy Sparse RLS”.In: *IEEE Trans. Signal Proc.* 60.5 (May 2012), pp. 2194–2207.

[8] Elliott, A., M. Cucuringu, M.M. Luaces, P. Reidy, and G. Reinert.“Anomaly detection in networks with application to financial transaction networks”.In: *Arxiv:1901.00402* (2019).

[9] Golts, A. and M. Elad.“Linearized Kernel Dictionary Learning”.In: *IEEE J. Sel. Topics Signal Proc.* 10.4 (June 2016), pp. 726–739.

[10] Ho, J., Y. Xie, and B. Vemuri.“On a nonlinear generalization of sparse coding and dictionary learning”.In: *International Conference on Machine Learning*.2013,pp. 1480–1488.

[11] Hu, J. and Y.-P. Tan.“Nonlinear dictionary learning with application to image classification”.In: *Pattern Recognition* 75 (2018), pp. 282–291.

[12] Irofti, P. and A. Băltoiu.“Malware Identification with Dictionary Learning”.In: *2019 27th European Signal Processing Conference (EUSIPCO)*.A Coruna, Spain, 2019,pp. 1–5.

[13] Irofti, P. and A. Băltoiu.“Unsupervised dictionary learning for anomaly detection”.In: *Arxiv preprint arxiv:2003.00293* (2020).

[14] Irofti, P., A. Pătraşcu, and A. Băltoiu.“Fraud detection in networks”.In: *Enabling AI Applications in Data Science*.Springer. 2021,pp. 517–536.

[15] Irofti, Paul, Florin Stoican, and Vicenç Puig.“Fault handling in large water networks with online dictionary learning”.In: *Journal of process control* 94 (2020), pp. 46–57.

[16] Le Magoarou, L. and R. Gribonval.“Chasing Butterflies: in Search of Efficient Dictionaries”.In: *Int. Conf. Acoustics Speech Signal Proc (ICASSP)*.Brisbane, Australia, Apr. 2015,pp. 3287–3291.

[17] Lebichot, B., G.M. Paldino, W. Siblini, L. He-Guelton, F. Oblé, and G. Bontempi.“Incremental learning strategies for credit cards fraud detection”.In: *International journal of data science and analytics* (2021), pp. 1–10.

[18] Mahdizadehaghdam, S., A. Panahi, H. Krim, and L. Dai.“Deep dictionary learning: A parametric network approach”.In: *IEEE Trans. Image Proc.* 28.10 (2019), pp. 4790–4802.

[19] Nguyen, V.H., V.M. Patel, N.M. Nasrabadi, and R. Chellappa.“Design of Non-Linear Kernel Dictionaries for Object Recognition”.In: *IEEE Trans. Image Proc.* 22.12 (Dec. 2013), pp. 5123–5135.

[20] Onose, A. and B. Dumitrescu.“Adaptive matching pursuit using coordinate descent and double residual minimization”.In: *Signal Proc.* 93.11 (2013), pp. 3143–3150.

[21] Papyan, V., Y. Romano, and M. Elad.“Convolutional neural networks analyzed via convolutional sparse coding”.In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 2887–2938.

[22] Parikh, N. and S. Boyd.“Proximal algorithms”.In: *Foundations and Trends in Optimization* 1.3 (2014), pp. 127–239.

[23] Rubinstein, R., A.M. Bruckstein, and M. Elad.“Dictionaries for Sparse Representations Modeling”.In: *Proc. IEEE* 98.6 (June 2010), pp. 1045–1057.

[24] Rubinstein, R., M. Zibulevsky, and M. Elad.“Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation”.In: *IEEE Trans. Signal Proc.* 58.3 (Mar. 2010), pp. 1553–1564.

[25] Rusu, C. and B. Dumitrescu.“Stagewise K-SVD to Design Efficient Dictionaries for Sparse Representations”.In: *Ieee signal proc. letters* 19.10 (Oct. 2012), pp. 631–634.

[26] Schmitz, M.A., M. Heitz, N. Bonneel, F. Ngole, D. Coeurjolly, M. Cuturi, G. Peyré, and J.-L. Starck.“Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning”.In: *SIAM Journal on Imaging Sciences* 11.1 (2018), pp. 643–678.

[27] Song, J., X. Xie, G. Shi, and W. Dong.“Multi-layer discriminative dictionary learning with locality constraint for image classification”.In: *Pattern Recognition* 91 (2019), pp. 135–146.

[28] Tariyal, S., A. Majumdar, R. Singh, and M. Vatsa."Deep dictionary learning".In: *IEEE Access* 4 (2016), pp. 10096–10109.

[29] Teke, O. and P.P. Vaidyanathan."On the role of the bounded lemma in the SDP formulation of atomic norm problems".In: *IEEE Signal Proc. Letters* 24.7 (July 2017), pp. 972–976.

[30] Thiagarajan, J.J., K.N. Ramamurthy, and A. Spanias."Multiple kernel sparse representations for supervised and unsupervised learning".In: *IEEE Trans. Image Proc.* 23.7 (2014), pp. 2905–2915.

[31] Tosic, I. and P. Frossard."Dictionary Learning".In: *IEEE Signal Proc. Mag.* 28.2 (Mar. 2011), pp. 27–38.

[32] Wilson, A.G., Z. Hu, R. Salakhutdinov, and E.P. Xing."Deep kernel learning".In: *19th Int. Conf. Artificial Intelligence and Statistics, AISTATS 2016*.2016,pp. 370–378.

[33] Wohlberg, B."Efficient Algorithms for Convolutional Sparse Representations".In: *IEEE Trans. Image Proc.* 15.1 (Jan. 2016), pp. 301–315.

[34] Zhang, Z., Y. Xu, J. Yang, X. Li, and D. Zhang."A Survey of Sparse Representation: Algorithms and Applications".In: *IEEE Access* 3 (2015), pp. 490–530.

[35] Zhao, Y., Z. Nasrullah, and Z. Li."PyOD: A Python Toolbox for Scalable Outlier Detection".In: *Journal of Machine Learning Research* 20.96 (2019), pp. 1–7.URL: http://jmlr.org/papers/v20/19-011.html.