# Sparse and Redundant Representation Modeling — What Next?

Michael Elad (*Fellow, IEEE*)

***Abstract*—Signal processing relies heavily on data models; these are mathematical constructions imposed on the data source that force a dimensionality reduction of some sort. The vast activity in signal processing during the past several decades is essentially driven by an evolution of these models and their use in practice. In that respect, the past decade has been certainly the era of *sparse and redundant representations*, a popular and highly effective data model. This very appealing model led to a long series of intriguing theoretical and numerical questions, and to many innovative ideas that harness this model to real engineering problems. The new entries recently added to the IEEE-SPL EDICS reflect the popularity of this model and its impact on signal processing research and practice.**

**Despite the huge success of this model so far, this field is still at its infancy, with many unanswered questions still remaining. This paper[1] offers a brief presentation of the story of sparse and redundant representation modeling and its impact, and outlines ten key future research directions in this field.**

## I. Introduction — Who Needs Models?

One could not imagine the vast progress made in signal and image processing in the past fifty years without the central contribution of data models. Consider the following example as a way of illustrating the need for a model: A signal of interest $\mathbf{x} \in \mathbb{R}^d$ is measured in the presence of additive noise, $\mathbf{v} \sim \mathbb{N}(0, \sigma^2\mathbf{I})$, producing $\mathbf{y} = \mathbf{x} + \mathbf{v}$. Given $\mathbf{y}$ we would like to recover $\mathbf{x}$, essentially seeking a decomposition of $\mathbf{y}$ into its two parts, $\mathbf{x}$ and $\mathbf{v}$. Despite the fact that we have a full statistical characterization of the noise, such a separation is impossible, as the noise model only implies a Gaussian distribution for $\mathbf{x}$, peaked at none other than $\mathbf{y}$ itself. To depart from this triviality, we must characterize $\mathbf{x}$ as well, so that the two parts can be told apart.

A model for the signal $\mathbf{x}$ is exactly this — a mathematical characterization of the signal. As an example for a possible model and its use, if we know that $\mathbf{x}$ resides in a subspace of dimension $r \ll d$ spanned by the columns of the matrix $\mathbf{Q} \in \mathbb{R}^{d\times r}$, this constitutes a model, and denoising (cleaning the noise) becomes possible. One could project $\mathbf{y}$ onto this subspace, applying the operation $\mathbf{Q}\mathbf{Q}^\dagger\mathbf{y}$, in order to find the closest signal to $\mathbf{y}$ that complies with the model. Put formally, this projection is obtained as the solution of the problem

$$\hat{\mathbf{x}} = \text{argmin}_{\mathbf{x}} \ \|\mathbf{x}-\mathbf{y}\|_2^2 \quad s.t. \quad \mathbf{x} = \mathbf{Q}\mathbf{Q}^\dagger\mathbf{x}, \qquad (1)$$

where $\mathbf{Q}\mathbf{Q}^\dagger\mathbf{x}$ represents our subspace constraint for the signal.[2] This will lead to effective denoising, with noise attenuation by a factor $r/d$ ($\ll 1$) on average.

The signal and image processing literature has seen numerous attempts to handle the above-described denoising problem. Explicitly or implicitly, each and every one of these many thousands of published methods relies on a specific model, proposing a way to characterize the signal and a method to exploit this for the recovery of $\mathbf{x}$. While the above model example is very simple, it sheds light on key properties of models in general. An effective model typically suggests a dimensionality reduction of some sort; the original $d$ samples in the signal $\mathbf{x}$ are believed to be redundant and a much shorter description (in our example, of length $r$) can be given, reflecting the true dimensionality of the signal. Another issue is the migration from the core model formulation to its deployment in the processing task. In the example we suggested a projection of $\mathbf{y}$, which is very natural. However, when the model becomes more expressive and complex, leaving room for more than one approximation, various possible "estimators" may be proposed; thus, in general there is no one-to-one correspondence between a model and the way to practice it, and this leaves much room for original and creative ideas.

While the above example discussed the denoising problem, models are necessary for almost every processing that $\mathbf{x}$ may need to undergo. Sampling of a signal relies on a prior assumption about its content, so as to guarantee no loss or limited loss of information. Compression of a signal is conceptually possible only because it has an inner structure that reduces its entropy,

M. Elad is with the Computer Science Department, Technion – Israel Institute of Technology, Haifa 32000, Israel (e-mail: elad@cs.technion.ac.il). 

[1]This is not a regular IEEE-SPL paper, but rather an invited contribution offering a vision for key advances in emerging fields.

[2]If $\mathbf{x}$ is known to be spanned by the columns of $\mathbf{Q}$, it can be written as $\mathbf{x} = \mathbf{Q}\alpha$ for an arbitrary vector $\alpha \in \mathbb{R}^r$. Multiplying both sides by $\mathbf{Q}^\dagger$ we obtain $\mathbf{Q}^\dagger\mathbf{x} = \mathbf{Q}^\dagger\mathbf{Q}\alpha = \alpha$, since $\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}$. Substituting this relation back into $\mathbf{x} = \mathbf{Q}\alpha$ we get the constraint $\mathbf{x} = \mathbf{Q}\alpha = \mathbf{Q}\mathbf{Q}^\dagger\mathbf{x}$, as appears in Equation (1).

which is captured by a model that describes this signal with few parameters. Detection of anomalies in a signal or detection of target-content can be done only when we rely on models for the different contents. Similarly, separation of superimposed signals is done by using models for the distinct parts. Solving inverse problems such as a tomographic reconstruction from projections, recovery of missing samples, super-resolving a signal, inpainting (filling-in missing values), extrapolating a signal, and deblurring, all rely on a model for the signals in question in order to regularize the typically highly ill-posed inversion process when recovering the desired signal.

Models can take various forms, and through the years they have been gradually improving. What does this mean? It is important to understand that for most signal sources there is no notion of a “correct model”. This is reminiscent of unified theories in physics that cannot be proven correct, but can be demonstrated to align well with experiments. Models, as a set of mathematical properties that the data is believed to follow, or as a probability density function in the Bayesian point of view, are necessarily erroneous. The quest for better models is a search for a more flexible and accurate mathematical construction that reduces the overall model error. A careful study of the vast literature in signal and image processing from the past several decades reveals that there has been an evolution of models, constantly improving along time by reducing their modeling error, and therefore consistently leading to better performance in the applications they were brought to serve. In that respect, the more successful models are those that rely on signal examples to tune their characteristics, thus fitting better to the signals they describe. Figure 1 presents such an evolution of models along time in the image processing literature.

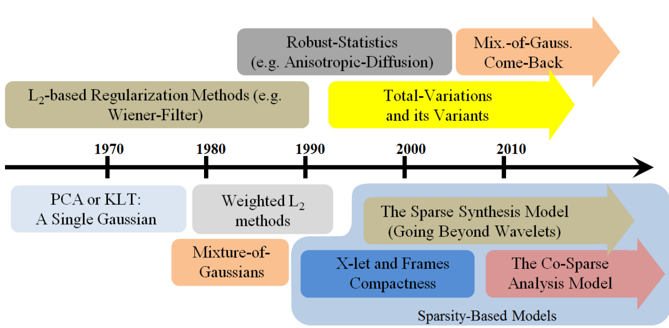


Figure 1. The evolution of models over the years in the field of image processing. This evolution shows a clear path from $L_2$-based methods to more challenging and sparsity-promoting models.

To summarize, and to answer the question posed in the title of this section, *we all need models*, as they are extensively used for many tasks. Their importance cannot be overstated — their impact on our abilities and ways to process data is central and irreplaceable. It is now time to discuss one of the most recent contributions to signal modeling — the model based on sparse and redundant representations, which will be referred to hereafter as *Sparseland*.

## II. Sparseland Modeling

In the past decade there has been tremendous progress in the construction and use of new signal models. One of the main achievements within this activity is the conception of sparse and redundant representation modeling [1], [2], [3], [4] and references therein.[3] The fundamental idea behind this model is a redundant transform of the signal $\mathbf{x} \in \mathbb{R}^d$ to a new representation $\boldsymbol{\alpha} \in \mathbb{R}^n$, where $n > d$ (thus leading to redundancy), such that the obtained representation is the simplest (i.e., sparsest) possible. This transform is semi-linear — the inverse transform from $\alpha$ to $\mathbf{x}$ is linear, given by $\mathbf{x} = \mathbf{D}\alpha$, where $\mathbf{D} \in \mathbb{R}^{d \times n}$ is a matrix commonly referred to as the dictionary. The forward transform is a highly non-linear one,

$$(P_0) \quad \hat{\alpha} = \text{argmin}_{\alpha} \ \|\alpha\|_0 \quad s.t. \quad \mathbf{x} = \mathbf{D}\alpha, \tag{2}$$

searching for the sparsest explanation for the signal $\mathbf{x}$. The $\ell^0$ cost function $\|\cdot\|_0$ counts the non-zero entries in this vector, and we expect a sparse outcome, $\|\alpha\|_0 = k \ll d$. This model can be interpreted as a chemistry of data sources: the columns of the dictionary are referred to as atoms (fundamental elements), and thus the dictionary serves as the “periodic table” in this chemistry. The signal is thought of as a molecule, which is believed to be a (linear) combination of only *few* atoms.

We mentioned before that models impose a dimensionality reduction – how does this happen here? A signal belonging to Sparseland is assumed to have $k \ll d$ non-zeros in its representation, and thus it must have one of the $\binom{n}{k}$ possibilities for a support (set of atoms used). Each such support defines a subspace of dimension $k$ (at most) in $\mathbb{R}^d$, and so the overall signal model is therefore a *union-of-subspaces* which the signal is believed to belong to. As such, this model is a natural extension of the naive single subspace model mentioned in the previous section, and the dimensionality reduction obtained here is somewhat more involved. It is evident from the above description that the dictionary $\mathbf{D}$ is

[3]Due to length constraints, it would be impossible to do justice and cite all the relevant literature.

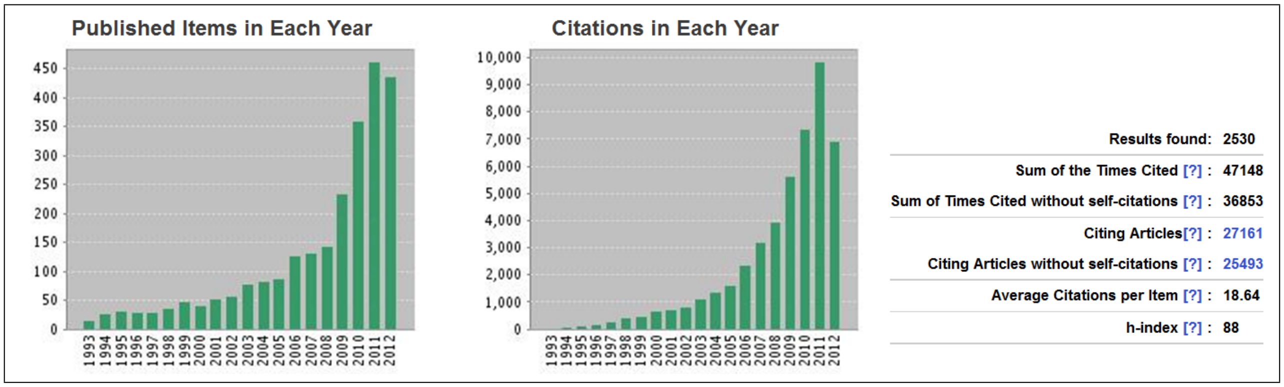


Figure 2. Paper and citation counts over the years for work related to the Sparseland model. These results were obtained via ISI Web-Of-Science searching the SCI-Expanded database with: Topic=(((parsim* or spars*) and (represent* or approx* or sol* or convex) and (pursuit or dictionary or transform)) or (compres* and (sens* or sampl*) and (spars* or parsim*))). This search was done on September $22^{nd}$, 2012. Since this search was done on September 2012, the paper and citation counts for 2012 are only partial, explaining the sudden drop in the graphs.

central to the characterization of the signal family which $\mathbf{x}$ belongs to.

In order to illustrate how this model should be used in practice, let us return to the denoising task as described above: We are given $\mathbf{y}$, a noisy version of the signal $\mathbf{x}$, and we assume that $\mathbf{x}$ belongs to Sparseland, i.e., it is known to have a $k$-sparse representation with respect to some dictionary $\mathbf{D}$. In order to denoise $\mathbf{y}$, we should project it onto the model, essentially searching for the signal $\hat{\mathbf{x}}$ that is the closest to $\mathbf{y}$, while also belonging to one of the $\binom{n}{k}$ subspaces that this model covers. Put formally, we should solve the problem

$$(P_0^{noise}) \qquad \hat{\alpha} = \text{argmin}_{\alpha} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 \; s.t. \; \|\alpha\|_0 = k, \quad (3)$$

which is a direct extension of Equation (2) for the noisy setup. The denoised signal would be $\mathbf{D}\hat{\alpha}$, the multiplication of the found representation $\hat{\alpha}$ by the dictionary. This description reveals the high complexity that characterizes the required denoising process (in fact, it has been proven to be NP-hard), as we need to sweep through all the possible subspaces. In order not to leave our readers worried at this stage, we should add that in recent years there has been enormous progress in devising highly efficient methods to approximate the solution of $(P_0^{noise})$, with provable guarantees of success.

The Sparseland model is a very effective and successful model. It has gained substantial popularity in the past decade, as it offers a unique mixture of theoretical depth, numerical challenges, and successful applications. Beyond the simplicity and elegance of this model, it is very appealing and popular because of the following key reasons:

**1. Universality:** As said above, a signal belonging to Sparseland must reside in a union-of-subspaces. This high-dimensional structure is very rich, very flexible, and yet, it emerges from a relatively concise set of parameters — the dictionary $\mathbf{D}$. Due to this universality, this model has found numerous applications in processing various imaging sources [5], [6], [7], [8], audio [9], video [10], [11], seismic data [12], financial data [13], and more.

**2. Theory:** This model leads to a wide set of interesting theoretical questions. Given a candidate solution for $\mathbf{D}\alpha = \mathbf{x}$, could we claim its global optimality for the problem defined as $(P_0)$ in Equation (2)? As this problem is NP-Hard, can we approximate it? If so, what guarantees can we pose for these algorithms? For a noisy signal, could we recover the correct support of its representation? What is the minimal possible error we should expect? and is it achievable by practical algorithms? Finally, if a signal is replaced by a small set of projections of it, can this be sufficient for recovering it? These and many other questions have formed the grounds for fascinating research activity that flourished in the past decade, leading to very constructive and elegant answers [14], [15], [16], [17], [18].

**3. Practicality:** There is a clear path from the above theory to practical algorithms and applications. A large family of pursuit algorithms (e.g. [19], [20]) has been proposed for approximating the solution of $(P_0)$ and its noisy variant, $(P_0^{noise})$. Various ideas have been explored for harnessing this model to applications, using wavelets (e.g., [21]) or learning the dictionary from examples [22], [23], [24], [25], leading often to state-of-the-art results.

Today, this field and its branches have become very popular topics, with many active researchers from leading universities devoting their efforts to decipher some of the above mentioned riddles and challenges. Figure 2 shows paper and citation counts over the years for work related to the Sparseland model. While this is a crude and somewhat inaccurate search, it is sufficient for demonstrating the massive (exponential) growth in the interest in this field. Figure 3 presents the spread of these publications in various universities and research institutes around the world.

| Field: Organizations | Record Count | % of 2530 |
|---|---|---|
| STANFORD UNIV | 86 | 3.399 % |
| TECHNION ISRAEL INST TECHNOL | 86 | 3.399 % |
| ECOLE POLYTECH FED LAUSANNE | 53 | 2.095 % |
| UNIV CALIF BERKELEY | 51 | 2.016 % |
| CALTECH | 47 | 1.858 % |
| UNIV ILLINOIS | 46 | 1.818 % |
| UNIV MINNESOTA | 46 | 1.818 % |
| UNIV BRITISH COLUMBIA | 45 | 1.779 % |
| MIT | 44 | 1.739 % |
| TSINGHUA UNIV | 41 | 1.621 % |

Figure 3. A list of the universities and research institutes where most of the publications on Sparseland originate from. This relies on the same ISI Web-of-Science search as in Figure 2.

The interest in this model is growing dramatically, which only magnifies the community's appetite for more breakthroughs in this field, and the belief that such achievements are within reach. Incidently, the work on compressed[4] sensing (CS) that emerged in 2005-2006 [26], [27] has also contributed in a large part to the popularity of this model, as it added another layer of theory and practice to sparse representations, in the context of signal-sampling. So strong was the impact of CS that many today identify the entire field of sparse and redundant representations with it.

In the recently updated EDICS for the IEEE-Signal Processing Letters, five new entries related to sparse and redundant representations have been added. These include:

- Theory of sparse representations and compressed-sensing (MLSAS-SPARSE),
- Sparse signal representations and recovery - algorithms and applications (DSP-SPARSE)
- Audio processing via sparse representations (AEA-SPARSE),
- Sparse representation in imaging (IMD-SPAR), and
- Compressed Sensing (SAM-SPARCS).

This reflects the popularity of this model and its impact on signal processing research and practice.

A few words about the history of this field: It is quite hard to single-out the origin of the ideas that stand behind the above model, but it is clear that the vast activity in the years 1980-2000 on transforms in general, and wavelets and frame theory in particular, substantially contributed to it, setting the stage for its conception. Central contributions were made by a handful of influential works in the mid-nineties [28], [29], [30], and these marked the birth of the field as an individual research area. Following these, a truly extensive work on Sparseland modeling took place mostly in the past decade. It was the daring paper in 2001 by Donoho and Huo [31], which ignited the burst of interest in this field, by establishing for the first time a theoretical connection between sparsity-seeking transforms and the $\ell_1$-norm measure.

During the past decade, the interest in sparsity and redundancy has grown dramatically. The scientists working on this topic come from various disciplines — mathematicians (applied and theorists), statisticians, engineers from various fields, geophysicists, physicists, neuropsychologists, computer science theoreticians, and others. Various conferences, workshops, and special sessions have been organized to gather scientists working on these topics, and various journals have allocated special issues for this and related topics. All of these testify to the great popularity this field has gained.

## III. So, What Next?

With the impressive achievements that sparse and redundant representation modeling has gathered in the past years, one might be tempted to believe that not much is left to be done. However, this could not be further from the truth — there are numerous unanswered questions and unexplored avenues for future research in the fields of signal modeling in general and sparsity-based models in particular. The success of sparsity and redundancy as core forces in signal modeling naturally leads to an appetite to further extend sparsity-based signal modeling techniques, and stretch their limits in various ways.

In trying to map the future work in this arena, we see three key directions that are central for advancing this field, and those are likely to draw the attention of researchers in coming years:

- Theoretical frontiers
- Model improvements, and
- Applications.

We shall now detail each of these directions, and mention ten key open problems that await future treatment.

At this point we would like to draw the reader's attention to Thomas Strohmer's paper published in this issue, whose focus is Compressive-Sensing (CS) [32]. In his article, Strohmer discusses the recent progress in CS, and points to key challenges and opportunities in this field. As CS and Sparseland are closely related, many of the topics covered in [32] are of great relevance to the discussion here, and the reader would benefit much from reading the two papers together.

[4] Sometimes the term "compressive" is used instead.

*And just before we start, a few words of caution: First, the open problems described hereafter are often so because they are simply very hard to handle. Second, the discussion to come is somewhat biased by the author's own perspective, and the ideas discussed should be taken as such.*

### A. Theoretical Frontiers

As said above, one of the main achievements of the vast work on Sparseland is the theoretical backbone that establishes the goodness of this model and the algorithms serving it. In the past decade, many papers explored the performance limits of pursuit algorithms, reconstruction bounds for various inverse problems serving sparsely represented signals, and fundamental properties of the core model. Nevertheless, the main questions on all these fronts remain essentially open, simply because the obtained results are often unrealistically restrictive.

**1. Better Bounds:** A common flaw that shadows many of the existing results in this field is the reliance on worst-case measures of the dictionary properties, such as the mutual-coherence [14] or the restricted isometry property (RIP) [18]. Both these measures (and there are others suffering from the same weakness) are defined with respect to the worst-possible constellation of atoms and their influence on the pursuit success. As such, even if consequent analysis adopts a probabilistic point of view, the eventual results tend to be overly restrictive, leading to wide gaps between theoretical predictions and actual performance (which tends to be much better). Naturally, we should explore ways to replace these worst-case measures by average-case ones, or others. While such work has already started to appear (see [33], [34]), far more work should be invested on this topic in order to lead to simple yet effective theoretical predictions of the actual pursuit performance.

**2. Targeting General Dictionaries:** The above leads us naturally to the next issue, of providing theoretical results on the performance of pursuit algorithms for general content dictionaries. In recent years much progress has been made in making statements on the pursuit performance for random dictionaries. A major breakthrough along these lines has been obtained via the approximate message-passing algorithms and their machinery [35], [36]. However, none of these can be extended easily to more general content dictionaries, which means that these results seem to be applicable only for the compressed-sensing problem. In various image processing tasks, we have no control over the dictionary content, and often the dictionary is obtained through training. In these cases we would like to know that the CS results remain applicable, suggesting that a good (near-oracle) MSE recovery is possible of a signal that is known to be compressible (having a sparse representation, or nearly so).

**3. Theory of Dictionary Learning:** Dictionary learning is a prominent part of Sparseland, as it enables the extraction of an underlying sparsifying dictionary from a given set of signal examples. This process has been treated mostly empirically [22], [23], [25], [30], offering algorithms to learn the dictionary. Theoretically speaking, so far we have little justification for using these algorithms — we do not know whether the learning problem is stable (i.e., that noisy signals can be used in principle to learn a good quality dictionary), we do not have solid results that guarantee the success of the learning algorithms, nor do we have a thorough study of these algorithms' flaws. Here too there are very few recent, partial, but very daring attempts [37], [38], [39]. But more than anything, these testify to the complexity of these questions, and the dire need for broader and more constructive theoretical results that will establish the "safety" of using dictionary learning.

**4. Unified Theory of Simplicity Measures:** As a final topic in this sub-section, we take a few steps back and adopt a much wider view of Sparseland: The $\ell^0$ sparsity has been chosen as the driving force for this model because it serves as an ultimate measure of simplicity. In recent years, low-rankness of a matrix has been also popularized as an alternative simplicity measure [40]. How are the two connected? Recent results show a unified view of the two, including a common theoretical study of recovery problems [41]. Are there other measures of simplicity we should be thinking of? Is there a unified view of such measures that would include sparsity, low-rankness, and others (such as low-entropy), as special cases? This may lead to a new theory that considers an abstract notion of simplicity in general inverse problems.

### B. Model Improvements

Behind the deep and elegant theory developed for Sparseland, stand the desire to find better ways for modeling of data. As such, Sparseland has been shown to be quite successful and better than its predecessors. However, if the goal is indeed modeling, we must ask ourselves whether Sparseland is the ultimate model. Are there flaws in this model that call for modifications? Could we envision ways to improve it? We mentioned above the existence of an evolution of models, improving over time, and there is no reason to believe that this evolution ends here. Thus, Sparseland is nothing more than one stop in a long list of generations of models to come. And so, the natural question is — where do we go from here in order to improve this model?

**5. Introducing Structure:** The chemistry analogy to Sparseland is enlightening, as it offers natural extensions to this model. For example, molecules in nature are built of a few kinds of atoms, just like in the Sparseland model. However, in nature, the possible atom constellations are not equally likely; some atoms are more common while others are rare; some tend to appear together often, while others never co-appear, etc.. This suggests that the sparsity pattern of the representation should not be thought of as an i.i.d. binary vector, but rather assumed to have a structure. Imposing dependencies between the representation entries adds further constraints on the signal, reduces the effective number of permitted subspaces, and thus leads to deeper dimensionality reduction.

There are various ways to impose structure on the representation vector $\alpha$ and its support — see for example [42], [43], [44], [45]. Mixture of Gaussians, which recently got a revived interest for image processing purposes may be considered as a special case of such a structure [46], [47]. The existing work along these lines is only the beginning, however, as we struggle to adapt the model to actual data sources. Incorporating such a representation structure into the model should affect the whole processing chain, from design of novel pursuit algorithms, through adequately modifying dictionary learning methods, all the way to estimation of the parameters defining the structure-model. These should be accompanied by extensions of the existing theoretical guarantees, as discussed in the previous subsection.

**6. Structured Dictionaries:** Imposing structure should be considered not only in the context of the representation vector, but also in the learned dictionaries that bring the adaptation of the model to the data. Current dictionary learning methods operate on low-dimensional and typically fixed-size signals, and build a dictionary which is an unconstrained and unstructured matrix. When aiming to model high-dimensional data, existing methods cannot cope with the induced complexity and memory requirements. A structured dictionary is the answer — an arbitrarily large matrix that would serve the signal it aims to sparsify, and which will be defined via a reduced and manageable set of parameters.

What structure should we impose? The natural answer that comes to mind is an imitation of existing analytical transforms, such as the curvelets [48], contourlets [49], and shearlets [50], incorporating near-shift-invariance, multi-directional and multi-scale relations between the atoms. This is the natural way to go, and yet it has not been done so far to the point where we could dispose of these more classical and analytically-oriented transforms, replacing them with a process we may refer to in the future as *trainlets*. Few exceptions to this statement exist (e.g., [11], [51]), but their construction is far from being satisfactory and competitive.

**7. The Analysis Co-Sparse Model:** Sparseland, as described above is typically referred to as a synthesis model, because the relation $\mathbf{x} = \mathbf{D}\alpha$ suggests a way to synthesize $\mathbf{x}$ by drawing a random representation with $k$ non-zero entries and multiplying it by $\mathbf{D}$. Most of the work done so far on sparse and redundant representations has been done in the context of this synthesis model. However, this is not the only way to obtain a union-of-subspaces construction. More specifically, there is a different viewpoint to sparse representations which has been left aside almost untouched – *the co-sparse analysis model* [52]. This model represents a signal $\mathbf{x}$ by multiplying it by an *analysis dictionary*, $\mathbf{\Omega}$, and requiring that the outcome, $\alpha = \mathbf{\Omega}\mathbf{x}$, is sparse.

We shall not expand on the properties of the co-sparse analysis model here, but we will mention that it is markedly different from the synthesis counterpart approach, it is more expressive as it leads to a much richer union-of-subspaces, and it promotes strong linear dependencies between its atoms for the model to perform well, something that stands as a complete opposite to the synthesis approach. In recent years there is a growing interest in this alternative, and a growing belief that it encompasses a potential for better signal modeling. The study of the co-sparse analysis model creates a series of new possibilities and opportunities for signal modeling.

**8. Model Errors:** Still in the context of model improvement, let us be bold and ask — how do we assess the quality of a model? A systematic way would be to evaluate the modeling errors — we mentioned earlier that every model induces some error. There are in fact *two* kinds of modeling errors, and they affect our applications very differently: The first corresponds to true signals that the model considers as unlikely (mis-detections), and the second refers to undesired signals that are endorsed by the model (false-alarms). These two errors, the balance between them, their formation, extent, and impact on applications, have never been systematically studied. Understanding these errors better may help in improving future models. In combating these errors, perhaps we should seek modeling techniques that are able to combine several simpler models, such as (i) a mixture of Sparseland models, (ii) using models that define their contribution using a negative force that carves out undesired regions from $\mathbb{R}^d$, or (iii) a hierarchical fusion of sparsity-based models.

### C. Applications

Returning to the basic goals of Sparseland, we need data models that can lead us to success in recovery, compression, sampling, detection, separation, and more. Various applications have already been addressed this way, and many others are likely to follow. Wireless communication, radar and sonar signal processing, speech and music processing, medical-imaging, computer-vision applications, machine-learning, data-mining, array-processing, and more, are a few of the many fields that will find use for sparsity-promoting models. In this subsection we mention two signal "customers" that are somewhat unconventional, thus calling for more unique treatment and original thinking.

**9. Sparsity in Computer Graphics:** 3D objects are commonly represented as polygon meshes, built as a collection of (many thousands of) vertices, edges, and 2D faces that define the eventual shape's surface. This is a signal defined on an unstructured grid, and as such, many of the available regular signal processing tools become irrelevant for it. How can we bring Sparseland to such signals? A naive attempt to re-sample the mesh and bring it to a uniformly-sampled 3D volume is of course possible, but this approach is highly sensitive to noise, it magnifies the data size substantially, and it introduces unavoidable sampling errors. As such, this approach loses much of the elegance that direct mesh-based processing methods have. Is there an invertible "transform" or "embedding" that could take a given mesh to a new domain, where Sparseland can be employed naturally? A positive answer to this question would lead to novel and highly effective tools in computer graphics.

**10. Processing Non-Conventional Signals:** The problem posed above is caused by the non-uniformity of the sampling in producing the data. This situation is also encountered in other cases, such as graph-based data and point clouds. For example, in recent years there is a growing interest in data available over social networks, organized as a graph with vertices and edges. Each vertex is characterized by a high-dimensional feature-vector, and the edges encode interrelations between vertices. The question arising is the same as above: How can we incorporate Sparseland into these cases? Sparsifying transforms for such data can be envisioned, and work has already started along these lines [53], [54], [55]. However, these are far from a full deployment of sparsity and redundancy to these needs.

## IV. Summary

In this paper we discussed the central role of data models in signal processing, and we introduced the story of Sparseland, which has come to be a leading model in recent years. This paper also offered a series of ten research directions that could be taken in order to push the frontiers of knowledge in this field. What we have not said (perhaps because it seems obvious) is that Sparseland is not the remedy to all the existing illnesses, and it does not belong everywhere. Scientists and engineers reading this article and broadly influenced by the buzz around Sparseland should be very careful to judge whether their problems call for the use of sparsity and redundancy. That said, the core idea of modeling signals based on their sparse representation seems natural, fundamental, and thus universally correct. For many applications this is a fantastic and winning tool, and the work and knowledge accumulated so far depict a hopeful future for this field.

## Acknowledgements

The author would like to thank Ron Rubinstein, Amir Adler, Tomer Peleg, Javier Turek, and Thomas Strohmer, for their help in proof-reading this paper and helping in bringing it to its final state.

## References

[1] 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*, 51(1):34–81, February 2009.

[2] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[3] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, 3rd edition, 2008.

[4] J.-L. Starck, F. Murtagh, and M.-J. Fadili, *Sparse Image and Signal Processing - Wavelets, Curvelets, Morphological Diversity* Cambridge University Press, 2010.

[5] M. Elad, M.A.T. Figueiredo, and Y. Ma, On the Role of Sparse and Redundant Representations in Image Processing, *Proc. of the IEEE*, 98(6):972–982, April 2010.

[6] M. Lustig, D.L. Donoho, J. Santos, and J. Pauly, Compressed Sensing MRI, *IEEE Signal Processing Magazine*, 27:72–82, March 2008.

[7] S. Samadi, M. Cetin, and M.A. Masnadi-Shirazi, Sparse Representation-Based Synthetic Aperture Radar Imaging, *IET Radar, Sonar and Navigation*, 5(2):182–193, February 2011.

[8] A. Rosenthal, D. Razansky, and V. Ntziachristos, Quantitative Optoacoustic Signal Extraction Using Sparse Signal Representations, *IEEE Trans. on Medical Imaging*, 28(12):1997–2006, Dec. 2009.

[9] M.D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M.E. Davies, Sparse Representations in Audio and Music: From Coding to Source Separation, *Proc. of the IEEE*, 98(6):995–1005, June 2010.

[10] M. Protter and M. Elad, Image Sequence Denoising Via Sparse and Redundant Representations, *IEEE Trans. on Image Processing*, 18(1):27–36, January 2009.

[11] J. Mairal, G. Sapiro, and M. Elad, Learning Multiscale Sparse Representations for Image and Video Restoration, *SIAM Multiscale Modeling and Simulation*, 7(1):214–241, April 2008.

[12] F.J. Herrmann and G. Hennenfent, Non-parametric Seismic Data recovery with Curvelet Frames, *Int. Geophysical Journal*, 173(1):233–248, April 2008.

[13] A. d'Aspremont, L. El Ghaoui, M.I. Jordan, and G.R.G. Lanckriet, A Direct formulation for Sparse PCA using Semidefinite Programming, *SIAM Review*, 49(3):434–448, September 2007.

[14] D.L. Donoho and M. Elad, Optimally Sparse Representation in General (nonorthogonal) Dictionaries via L1 Minimization, *Proc. National Academy of Sciences (PNAS)*, 100:2197–2202, March 2003.

[15] R. Gribonval and M. Nielsen, Sparse Representations in Unions of Bases, IEEE Trans. Information Theory, 49(12):3320–3325, December 2003.

[16] J.A. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Information Theory*, 50:2231–2242, 2004.

[17] J.A. Tropp, Just relax: Convex Programming Methods for Subset Selection and Sparse Approximation, *IEEE Trans. on Information Theory*, 51:1030-1051, 2006.

[18] E. Candès and T. Tao, The Dantzig Selector: Statistical Estimation When p is Much Larger than n, *Annals of Statistics*, 35(6):2313–2351, 2007.

[19] W. Dai and O. Milenkovic, Subspace Pursuit for Compressive Sensing Signal Reconstruction, *IEEE Trans. on Information Theory*, 55(5):2230–2249, may 2009.

[20] J.A. Tropp, J.S. Wright, Computational Methods for Sparse Solution of Linear Inverse Problems, *Proc. of the IEEE*, 98(6):948–958, June 2010.

[21] C. Chesneau, M.J. Fadili, J.-L. Starck, Stein Block Thresholding for Wavelet-Based Image Deconvolution, *Electronic Journal of Statistics*, 4:415–435, 2010.

[22] K. Engan, S.O. Aase, and J.H. Husoy, Multi-Frame Compression: Theory and Design, *Signal Processing*, 80:2121–2140, 2000.

[23] M. Aharon, M. Elad, and A.M. Bruckstein. K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation, *IEEE Trans. on Signal Processing*, 54:4311–4322, 2006.

[24] M. Elad and M. Aharon, Image Denoising Via Sparse and Redundant Representations over Learned Dictionaries, *IEEE Trans. on Image Processing*, 15:3736–3745, 2006.

[25] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, Online Learning for Matrix Factorization and Sparse Coding, *Journal of Machine Learning Research*, 11:19–60, March 2010.

[26] D.L. Donoho, Compressed-Sensing, *IEEE Trans. on Information Theory*, 52(4):1289–1306, April 2006.

[27] E.J. Candès and T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. on Information Theory*, 52(12):5406–5425, December 2006.

[28] S. Mallat and Z. Zhang, Matching Pursuits with Time-Frequency Dictionaries, *IEEE Trans. on Signal Processing*, 41:3397–3415, 1993.

[29] S.S. Chen, D.L. Donoho, and M.A. Saunders, Atomic Decomposition by Basis Pursuit, *SIAM J. Scientific Computing*, 20:33–61, 1998.

[30] B.A. Olshausen and D.J. Field, Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images, *Nature*, 381(6583):607–609, June 1996.

[31] D.L. Donoho and X.-M. Huo, Uncertainty Principles and Ideal Atomic Decomposition, *IEEE trans. on Information Theory*, 47(7):2845–2862, 2001.

[32] T. Strohmer, Measure What Should be Measured: Progress and Challenges in Compressive Sensing, *IEEE Signal Processing Letters*, 2012.

[33] E.J. Candes and Y. Plan, A Probabilistic and RIPless Theory of Compressed Sensing, *IEEE Trans. on Signal Processing*, 57(11):7235–7254, 2011.

[34] W.U. Bajwa, A.R. Calderbank, and D.G. Mixon, Two are Better than One: Fundamental Parameters of Frame Coherence, *Applied Computational Harmonic Analysis (ACHA)*, 33:58Ű-78, 2012.

[35] D.L. Donoho, A. Maleki, and A. Montanari, Message-Passing Algorithms for Compressed Sensing, *Proc. National Academy of Sciences (PNAS)*, 106(45):18914–18919, 2009.

[36] D.L. Donoho, A. Maleki, and A. Montanari, The Noise-Sensitivity Phase Transition in Compressed Sensing, *IEEE Trans. on Signal Processing*, 57(10):6920–6941, 2011.

[37] R. Gribonval and K. Schnass, Dictionary Identification - Sparse Matrix-Factorisation via $\ell_1$-Minimisation, *IEEE Trans. on Information Theory*, 56(7):3523–3539, 2010.

[38] Q. Geng and J. Wright, On the Local Correctness of L1 Minimization for Dictionary Learning, to appear in *IEEE Trans. on Information Theory*.

[39] D.A. Spielman, H. Wang, and J. Wright, Exact Recovery of Sparsely-Used Dictionaries, 25th Annual Conference on Learning Theory, JMLR, vol 23, 2012.

[40] E.J. Candes and Y. Plan, Matrix Completion with Noise, *Proc. of the IEEE*, 98(6):925-936, 2009.

[41] E.J. Candes and B. Recht, Simple Bounds for Recovering Low-complexity Models, to appear in *Mathematical Programming*, 2012.

[42] L. He and L. Carin, Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing, *IEEE Trans. on Signal Processing*, 57(9):3488–3497, 2009.

[43] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hedge, Model-Based Compressive Sensing, *IEEE Trans. Information Theory*, 56(4):1982-Ű2001, 2010.

[44] Y.C. Eldar, P. Kuppinger, and H. Bolcskei, Block-Sparse Signals: Uncertainty Relations and Efficient Recovery, *IEEE Trans. on Signal Processing*, 58(6):3042–3054, 2010.

[45] T. Peleg, Y.C. Eldar, and M. Elad, Exploiting Statistical Dependencies in Sparse Representations for Signal Recovery, *IEEE Trans. on Signal Processing*, 60(5):2286–2303, 2012.

[46] G.Yu, G. Sapiro, and S. Mallat, Solving Inverse Problems with Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity, *IEEE Trans. on Image Processing*, 21(5):2481–2499, 2012.

[47] D. Zoran and Y. Weiss, From Learning Models of Natural Image Patches to Whole Image Restoration, International Conference on Computer Vision (ICCV), 2011.

[48] J.-L. Starck, E.J. Candes, and D.L. Donoho, The Curvelet Transform for Image Denoising, *IEEE Trans. on Image Processing*, 11(6)670–684, 2002.

[49] M. Do and M. Vetterli, The Contourlet Transform: an Efficient Directional Multiresolution Image Representation, *IEEE Trans. Image on Processing*, 14(12):2091-Ű2106, 2005.

[50] K. Guo and D. Labate, Optimally Sparse Multidimensional Representation using Shearlets, *SIAM Journal on Mathematical Analysis*, 39(1):298–318, 2007.

[51] B. Ophir, M. Lustig, and M. Elad, Multi-Scale Dictionary Learning using Wavelets, *IEEE Selected Topics in Signal Processing*, 5(5):1014–1024, 2011.

[52] S. Nam, M.E. Davies, M. Elad, and R. Gribonval, The Cosparse Analysis Model and Algorithms, to appear in *Applied and Computational Harmonic Analysis*.

[53] R.R. Coifman and M. Maggioni, Diffusion Wavelets, *Applied Computational Harmonic Analysis*, 21(1)53-Ű94, 2006.

[54] D.K. Hammond, P. Vandergheynst, and R. Gribonval, Wavelets on Graphs Via Spectral Graph Theory, *Applied Computational Harmonic Analysis*, 30(2):129–150, 2010.

[55] I. Ram, M. Elad, and I. Cohen, Redundant Wavelets on Graphs and High Dimensional Data Clouds, *IEEE Signal Processing Letters*, 19(5):291–294, 2012.