Compressed sensing signal models - to infinity and beyond?

Thomas Blumensath and Michael Davies

IDCOM & Joint Research Institute for Signal and Image Processing
Edinburgh University, King's Buildings, Mayfield Road, Edinburgh, UK
thomas.blumensath@ed.ac.uk, mike.davies@ed.ac.uk

Abstract:

Compressed sensing is an emerging signal acquisition technique that enables signals to be sampled well below the Nyquist rate, given a finite dimensional signal with a sparse representation in some orthonormal basis. In fact, sparsity in an orthonormal basis is only one possible signal model that allows for sampling strategies below the Nyquist rate. We discuss some recent results for more general signal models based on unions of subspaces that allow us to consider more general *structured representations*. These include classical sparse signal models and finite rate of innovation systems as special cases.

We consider the dimensionality conditions for two aspects of the compressed sensing inverse problem: the existence of *one-to-one* maps to lower dimensional observation spaces and the smoothness of the inverse map.

On the surface Lipschitz smoothness of the inverse map appears to limit the applicability of compressed sensing to infinite dimensional signal models. We therefore discuss conditions where smooth inverse maps are possible even in infinite dimensions. Finally we conclude by mentioning some recent work [14] which develops the these ideas further allowing the theory to be extended beyond exact representations to *structured approximations*.

1. Introduction

Since Nyquist and Shannon we are used to sampling continuous signals at a rate that is twice the bandwidth of the signal. However recently, under the umbrella title of *compressed sensing*, researchers have begun to explore how and when signals can be recovered using much fewer samples, but relying on known signal structure. Importantly the papers by Candes, Romberg and Tao [4], [5], [6] and by Donoho [8] have shown that under certain conditions on the signal sparsity and the sampling operator (which are often satisfied by certain random matrices), finite dimensional signals can be stably reconstructed when the number of observations is of the order of the signal sparsity and only logarithmically dependent on the ambient space dimension. Furthermore the reconstruction can be performed using practical polynomial time algorithms.

Here we discuss a generalization of the sparse signal model that enables us to consider more structured signal types. We are interested in when the signals can be stably reconstructed (or in some cases approximated). We finish the paper by considering the implications of these results for ∞-dimensional signal models and extending from *structured representations* to *structured approximation*.

2. Signal models and problem statement

The problem can be formulated as follows. A continuous or discrete signal f from some separable Hilbert space is to be sampled. This is done by using M linear measurements $\{\langle f, \phi_n \rangle\}_n$, where $\langle \cdot, \cdot \rangle$ is the inner product and where $\{\phi_n\}$ is a set of vectors from the Hilbert space under consideration. Through the choice of an appropriate orthonormal basis, ψ we can replace f by the vector x such that $f = \sum_{i=1}^{N} \psi_i x_i$. Let $\Phi \in \mathbb{R}^{M \times N}$ be the sensing matrix with entries $\langle \psi_i, \phi_j \rangle$. The observation can then be written as

$$\mathbf{y} = \Phi \mathbf{x}. \quad (1)$$

In compressed sensing it is paramount to consider signals $\mathbf{x}$ that are highly structured and in the original papers, $\mathbf{x}$ was assumed to be an *exact k-sparse* vector, i.e. a vector with not more than k non-zero entries (we discuss a relaxation of this in section 6.). This naturally defines the signal model as a union of N-choose-k k-dimensional subspaces, $\mathcal{K}$.

A nice generalization of this model, introduced in [12], is to consider the signal $\mathbf{x}$ to be an element from a union of arbitrary subspaces $\mathcal{A}$, defined formally as

$$\mathcal{A} = \bigcup_{j}^{L} S_j, \; S_j = \{\mathbf{y} = \Omega_j \mathbf{a}, \Omega_j \in \mathbb{R}^{N \times k_j}, \mathbf{a} \in \mathbb{R}^{k_j}\}, \quad (2)$$

where the Ω_j are bases for linear subspaces. This general signal model incorporates many previously considered compressed sensing settings, including:

- The exact k-sparse signal model, $\mathcal{K}$
- Finite Rate of Innovation (FRI) [15] signal models, if we allow an uncountable number of subspaces (e.g. filtered streams of Dirac functions)
- signals that are k-sparse in a general, possibly redundant dictionary
- exact k-sparse signals whose non-zero elements form a tree

- multi-dimensional signals that are k-sparse with common support

Importantly this model allows us to incorporate additional structure which can in turn be advantageous by for example reducing signal complexity (as in the tree-constrained sparse model).

The aim of compressed sensing is to select a linear sampling operator, Φ, such that there exists a unique inverse map $\Phi|_{\mathcal{K}}{}^{-1} : \Phi(\mathcal{K}) \mapsto \mathcal{K}$. Moreover, for stability, we generally desire $\Phi(\mathcal{K})$ to be a bi-Lipschitz embedding of $\mathcal{K}$. In standard compressed sensing this stability is captured by the restricted isometry property [1].

When considering the union of subspaces model we can similarly look for a Φ with a unique stable (Lipschitz) inverse map $\Phi|_{\mathcal{A}}{}^{-1} : \Phi(\mathcal{A}) \mapsto \mathcal{A}$. Below we will discuss both necessary and sufficient conditions for this.

3. Existence of a unique inverse map

In [12] it was shown that a necessary condition for a unique inverse map to exist is that $M \geq M_{\min} := \max_{i \neq j} k_i + k_j$. If this is not the case we can find a vector $\mathbf{x} \in S_i \oplus S_j, \mathbf{x} \neq 0$ such that $\Phi x = 0$. The authors further go on to show that when there are a *countable* number of finite dimensional subspaces then the set of such sampling operators, Φ giving a unique inverse is dense.

In [3] we presented a slight refinement of this result for the case where the number of subspaces is finite. In this case *almost every* sampling operator, Φ, $M \geq M_{\min}$ has a unique inverse on $\mathcal{A}$. Furthermore even when $\max_i k_i < M < M_{\min}$ for almost every Φ the set of points in $\mathcal{A}$ without a unique inverse has zero measure (with respect to the largest subspace).

All this suggests that we might be able to perform compressed sensing from only slightly more observations than the dimension of the signal model, i.e. $M > \dim(\mathcal{A})$. Unfortunately we have so far ignored the issue of stability which we will see presents additional complications.

4. Stability of the inverse map

We now consider when the inverse mapping for the union of subspaces model is stable. Here we are particularly interested in the Lipschitz property of this inverse map and we derive conditions for the existence of a bi-Lipschitz embedding from $\mathcal{A}$ into a subset of $\mathbb{R}^M$.

The Lipschitz property is an important aspect of the map which ensures stability of any reconstruction to perturbations of the observation and in effect specifies the robustness of compressed sensing against noise and quantization errors. Furthermore, in the k-sparse model, the bi-Lipschitz property has also played an important role in demonstrating the existence of efficient and robust reconstruction algorithms through the k-restricted isometry property (RIP) [4, 5, 6, 8].

A natural extension of the k-restricted isometry for the union of subspaces model is [12, 3]:

Definition: ($\mathcal{A}$-restricted isometry) For any matrix Φ and any subset $\mathcal{A} \subset \mathbb{R}^N$ we define the $\mathcal{A}$-restricted isometry constant $\delta_{\mathcal{A}}(\Phi)$ to be the smallest quantity such that

$$(1 - \delta_{\mathcal{A}}(\Phi)) \leq \frac{\|\Phi \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \leq (1 + \delta_{\mathcal{A}}(\Phi)), \qquad (3)$$

holds for all $\mathbf{x} \in \mathcal{A}$.

If we define the set $\bar{\mathcal{A}} = \{\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 : \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}\}$ then $\delta_{\bar{\mathcal{A}}} < 1$ controls the Lipschitz constants of Φ and $\Phi|_{\mathcal{A}}{}^{-1}$ (in the standard compressed sensing this is directly equivalent to δ_{2m}). Specifically let us define:

$$\|\Phi(\mathbf{y}_1) - \Phi(\mathbf{y}_2)\|_2 \leq K_F \|\mathbf{y}_1 - \mathbf{y}_2\|_2 \qquad (4)$$

$$\|\Phi|_{\mathcal{A}}^{-1}(\mathbf{x}_1) - \Phi|_{\mathcal{A}}^{-1}(\mathbf{x}_2)\|_2 \leq K_I \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \qquad (5)$$

then a straight forward consequence of the $\bar{\mathcal{A}}$-RIP definition is that:

$$K_F \leq \sqrt{1 + \delta_{\bar{\mathcal{A}}}} \qquad (6)$$

$$K_I \leq \frac{1}{\sqrt{1 - \delta_{\bar{\mathcal{A}}}}} \qquad (7)$$

Note, as always with RIP, it is prudent to consider appropriate scaling of Φ to balance the upper and lower inequalities in (3).

The following results, proved in [3], give neccessary and sufficient conditions for Φ to be an $\mathcal{A}$-restricted isometry.

4.1 Sufficient conditions

Theorem 1 *For any $t > 0$, let*

$$M \geq \frac{2}{c\delta_{\mathcal{A}}} \left(\ln(2L) + k \ln\left(\frac{12}{\delta_{\mathcal{A}}}\right) + t \right), \qquad (8)$$

then there exist a matrix $\Phi \in \mathbb{R}^{M \times N}$ and a constant $c > 0$ such

$$(1 - \delta_{\mathcal{A}}(\Phi))\|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2 \leq (1 + \delta_{\mathcal{A}}(\Phi))\|\mathbf{x}\|_2^2 \qquad (9)$$

holds for all $\mathbf{x}$ from the union of L arbitrary k dimensional subspaces $\mathcal{A}$. What is more, if Φ is generated by randomly drawing i.i.d. entries from an appropriately scaled subgaussian distribution then this matrix satisfies equation (9) with probability at least

$$1 - e^{-t}. \qquad (10)$$

The proof follows the same lines as the construction of random matrices with k-RIP [1].

In contrast to the previous results on the existence of a unique inverse map this sufficient condition is logarithmic in the number of subspaces considered.

4.2 Necessary conditions

We next show that the logarithmic dependence on L is in fact necessary. This can be done by considering the distance between the optimally packed unit norm vectors in $\mathcal{A}$ as a function of the number of observations. To this end it is useful to define a measure of separation between vectors in the different subspaces:

Definition: ($\Delta(\mathcal{A})$ subspace separation) Let $\mathcal{A} = \bigcup_i S_i$ be the union of subspaces S_i and let $\mathcal{A}/S_i$ be the union of subspaces with the i^{th} subspace excluded. The subspace separation of $\mathcal{A}$ is defined as

$$\Delta(\mathcal{A}) = \inf_i \left[\sup_{\substack{\mathbf{x}_i \in S_i \\ \|\mathbf{x}_i\|_2=1}} \left[\inf_{\substack{\mathbf{x}_j \in \mathcal{A}/S_i \\ \|\mathbf{x}_j\|_2=1}} \|\mathbf{x}_i - \mathbf{x}_j\|_2 \right]\right] \quad (11)$$

We can now state the following necessary condition for the existence of an $\mathcal{A}$-restricted isometry in terms of $\Delta(\mathcal{A})$ and the observation dimension.

Theorem 2 *Let $\mathcal{A}$ be the union of L subspaces of dimension no more than k. In order for a linear map $\Phi : \mathcal{A} \mapsto \mathbb{R}^N$ to exist such that it has a Lipschitz constant K_F and such that its inverse map ${\Phi_{\mathcal{A}}}^{-1} : \Phi(\mathcal{A}) \mapsto \mathcal{A}$ has a Lipschitz constant K_I, it is necessary that*

$$M \geq \frac{\ln(L)}{\ln\left(\frac{4K_F K_I}{\Delta(\mathcal{A})}\right)}. \quad (12)$$

Therefore, for a fixed subspace separation, the *necessary* number of samples grows logarithmically with the number of subspaces.
This last fact suggests that extending the compressed sensing framwork to infinite dimensional signals may be problematic. For example, it implies that the $\log(N)$ dependence in the standard k-sparse signal model is necessary (from the easily derived bound $\Delta(\mathcal{A}) \geq \sqrt{2/k}$) and therefore such a framework does not directly map to infinite dimensional signal models.

5. 2 routes to infinity

Most of the results in compressed sensing assume that the ambient signal space, N, is finite dimensional. This also implies in the case of the k-sparse signal model ($k < \infty$) that the number of subspaces, L, in the signal model is also finite. In fact we would ideally like to understand when we can perform compressed sensing when either or both the quantities, N and L, are infinite. Specifically when might a stable unique inverse for $\Phi|_{\mathcal{A}}$ exist based upon a finite number of observations.
For example the Finite Rate of Innovation (FRI) sampling framework introduced by Vetterli *et al.* [15] provides sampling strategies for signals composed of the weighted sum of a finite stream of diracs. In this case both N and L are uncountably infinite while $M > 2k$ is sufficient to reconstruct the signal.
Below we consider two possible routes to infinity and comment on their stability. Note other routes to infinity also exist, such as when we let k, M and $N \to \infty$ while keeping k/M and M/N finite [9], or in the blind multi-band signal model [10, 13], where the sampling rate, M/N, is finite but where $M, N \to \infty$.

5.1 k, L finite and N infinite

We begin with the easy case that the reader might consider to be a bit of a cheat.

Consider a signal model $\mathcal{A} \subset \mathcal{H}$, where $\mathcal{H}$ is an infinite dimensional separable Hilbert space (i.e. $N = \infty$). Assume that both k and L are finite. In this case the union of subspace model $\mathcal{A}$ automatically lives within a finite dimensional subspace, $U \subset \mathcal{H}$ defined as:

$$U := \bigoplus_{i=1}^{L} S_i \quad (13)$$

Note that $\dim(U) \leq kL < \infty$. We can therefore first project onto the finite dimensional subspace U and then apply the above theory to guarantee both the existence and stability of inverse mappings in this setting.
Two signal models that naturally fit into this framework are: the block-based sparsity model [11], which is related to the multiple measurement vectors problem and has been used recently in a blind multi-band signal acquisition scheme [13]; and the tree-based sparsity model where the usual k-sparse model is constrained to form a rooted subtree where $L \leq \frac{(2e)^k}{k+1}$ independent of N [3] and naturally occurs in multi-resolution modelling. This model has also been recently extended to include tree-compressible signals [14]: see section 6..

5.2 k finite, L and N infinite

From Theorem 2 the only way in which the number of subspaces can be infinite (or even un-countable) while permitting a stable inverse mapping, $\Phi|_{\mathcal{A}}^{-1}$, with M finite is if the subspace separation, $\Delta(\mathcal{A}) = 0$. In such a case the union of subspace model may often form a nonlinear signal manifold. Note also that when we have an uncountable union of k-dimensional subspaces the dimension of the signal model may well be greater than k.
As an example let us consider the case of a simple Finite Rate of Innovation process [15]. Such models can be described as an uncountable union of subspaces and the key existence results from [12] immediately apply. However this tells us nothing about stability. For simplicity we will limit ourselves to a basic form of periodic FRI signal on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ which can be written as:

$$x(t) = G(\tau, \mathbf{a})(t) := \sum_{i=0}^{k-1} a_i \psi(t - \tau_i) \quad (14)$$

where ψ are also periodic on $\mathbb{T}$, $\tau = \{\tau_1, \ldots, \tau_k\}$ and $\mathbf{a} = \{a_1, \ldots, a_k\} \in \mathbb{R}^k$.
In [15] the possibility of a periodic Dirac stream is considered, i.e. $\psi(t) = \delta(t)$, $t \in [0, 1]$. Here we avoid the Dirac stream by restricting to the case where $\psi(t) \in \mathbf{L^2}(\mathbb{T})$ and directly consider the signal model defined by the parametric mapping:

$$G : U \times \mathbb{R}^k \mapsto \mathbf{L^2}(\mathbb{T}) \quad (15)$$

where $U = \{\tau \in \mathbb{R}^k : \tau_i < \tau_j, \forall i < j\}$. Individual subspaces can be identified with a given τ. Furthermore the continuity of the shift operator implies that for any $\psi(t) \in \mathbf{L^2}(\mathbb{T})$, the associated union of subspace model, $\mathcal{A}$ has $\Delta(\mathcal{A}) = 0$. Equivalently we can only find a finite number of subspaces, S'_j, whose union, $\mathcal{A}' := \bigcup_j^{L'} S'_j \subset$

$\mathcal{A}$ has $\Delta(\mathcal{A}') \geq \epsilon > 0$). Theorem 2 can then be used to lower bound the Lipschitz constants of any embedding in terms of the number of subspaces, L' of any such $\mathcal{A}'$.
We have seen that Theorem 2 does not preclude a stable embedding for such systems. However there is clearly more work needed to determine when such models can have finite dimensional stable embeddings. One possible avenue of research would be to examine the recently derived sufficient conditions for stable embedding of general smooth manifolds [7, 2].

6. ...and beyond?

In reality all the union of subspace models we have considered are an idealization. In practise we can expect to, at most, be able to *approximate* a signal by one from a union of subspaces model. In traditional compressed sensing this is the difference between finding a sparse representaion of an exact k-sparse signal and finding a good sparse approximation of a compressible signal (i.e. one that is well approximated by a k-sparse signal).
Recent work at Rice university [14] has shown that for the special case of restricted k-sparse models (such as the tree-restricted sparsity) the exact union of subspace model can be extended to approximate union of subspace models that are subsets of compressible signal models.
In order to go beyond exact representations further conditions are introduced. Notably:

1. *Nested Approximation Property (NAP)* - this specifies sets of models, $\mathcal{M}_K$, that are naturally nested.

2. *Restricted Amplification Property (RAmP)* - this imposes additional regularity on the sensing matrix Φ when acting on the difference between the $\mathcal{M}_K$ subspaces and the $\mathcal{M}_{K-1}$ subspaces (in the k-sparse case it is interesting to note that the RAmP condition is automatically satisfied by the k-RIP condition).

There are therefore a number of interesting open questions. For example, are such additional conditions typically necessary to go beyond exact subspace representations? Furthermore can these additional tools be applied successfully to arbitrary union of subspace models (i.e. ones that are not subsets of the standard k-sparse model)?

7. Acknowledgements

This research was supported by EPSRC grants D000246/1 and D002184/1. MED acknowledges support of his position from the Scottish Funding Council and their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership.

References:

[1] R. Baraniuk, M. Davenport, R. De Vore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[2] R Baraniuk and M Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 2007.

[3] T. Blumensath and M. E. Davies. Sampling theorems for signals from the union of linear subspaces. *Awaiting Publication, IEEE Transactions on Information Theory*, 2008.

[4] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, Feb 2006.

[5] Emmanuel Candès and Justin Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Comput. Math*, 6(2):227 – 254, 2006.

[6] Emmanuel Candès and Terence Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. on Information Theory*, 52(12):5406 – 5425, 2006.

[7] K. L. Clarkson. Tighter bounds for random projections of manifolds. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 39–48, 2008.

[8] D. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52(4):1289–1306, 2006.

[9] D. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *Journal of the AMS*, 2009.

[10] Y. C. Eldar. Compressed sensing of analog signals. *submitted to IEEE Trans. on Signal Processing*, 2008.

[11] Y. C. Eldar and M. Mishali. Robust recovery of signals from a union of subspaces. *Submitted to IEEE Trans. Inf Theory*, arXiv.org 0807.4581, 2008.

[12] Y. Lu and M. Do. A theory for sampling signals from a union of subspaces. *IEEE transactions on signal processing*, 56(6):2334–2345, 2008.

[13] M. Mishali and Y. C. Eldar. Blind multiband signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. Signal Proc.*, 57(3):993–1009, 2009.

[14] M.F. Duarte R. G. Baraniuk, V. Cevher and C. Hegde. Model based compressed sensing. *Submitted to IEEE Transactions on Information Theory*, 2008.

[15] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, 2002.