Preprint

File produced on
June 2, 2010

Compressed Sensing with Nonlinear Observations

Thomas Blumensath
School of Mathematics, University of Southampton, University Road, Southampton, SO17 1BJ, UK
thomas.blumensath@soton.ac.uk

Abstract

Compressed sensing is a recently developed signal acquisition technique. In contrast to traditional sampling methods, significantly fewer samples are required whenever the signals admit a sparse representation. Crucially, sampling methods can be constructed that allow the reconstruction of sparse signals from a small number of measurements using efficient algorithms.

We have recently generalised these ideas in two important ways. We have developed methods and theoretical results that allow much more general constraints to be imposed on the signal and we have also extended the approach to more general Hilbert spaces.

In this paper we introduce a further generalisation to compressed sensing and allow for non-linear sampling methods. This is achieved by using a recently introduced generalisation of the Restricted Isometry Property (or the bi-Lipschitz condition) traditionally imposed on the compressed sensing system. We show that, if this more general condition holds for the non-linear sampling system, then we can reconstruct signals from non-linear compressive measurements.

Key words and phrases : Compressed Sensing, Non-linear Sampling, Non-Convex Constraints, Inverse Problems

1 Introduction

Compressed sensing [6] deals with the acquisition of finite dimensional sparse signals. Let $\mathbf{x}$ be a sparse vector of length N and assume we sample $\mathbf{x}$ using M *linear* measurements. The M samples can then be collected into a vector $\mathbf{y}$ of length M and the sampling process can be described by a matrix $\mathbf{\Phi}$. If the observations are noisy, then the compressed sensing observation model is

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{e}$ is the noise vector. If $M < N$, then such a linear system is not uniquely invertible in general, unless we use additional assumptions on $\mathbf{x}$. Sparsity of $\mathbf{x}$ is such an assumption and compressed sensing theory tells us that, for certain $\mathbf{\Phi}$, we can recover $\mathbf{x}$ from $\mathbf{y}$ even if $M << N$, given that $\mathbf{x}$ has roughly $O(M)$ non-zero elements. However, in general, recovery of $\mathbf{x}$ is a combinatorial problem which is known to be NP-hard. Fortunately, under stricter conditions on $\mathbf{\Phi}$, a range of different polynomial time algorithms can be used to recover $\mathbf{x}$ whenever $\mathbf{x}$ has roughly $O(M/log(N))$ non-zero elements.

One of the conditions that guarantees that we can use efficient algorithms is the *Restricted Isomety Property.* A matrix $\mathbf{\Phi}$ satisfies the *Restricted Isomety Property* of order K [1] if

$$(1-\delta_K)\|\mathbf{x}\|^2 \leq \|\mathbf{\Phi x}\|^2 \leq (1+\delta_K)\|\mathbf{x}\|^2 \tag{2}$$

for all K-sparse $\mathbf{x}$. The *Restricted Isometry Constant* δ_K is defined as the smallest constant for which this property holds.

For example, [2] has shown that, for any $\mathbf{x}$, given an observation $\mathbf{y} = \mathbf{\Phi x} + \mathbf{e}$, where $\mathbf{\Phi}$ has the Restricted Isometry Property with $\delta_{2K} < \sqrt{2} - 1$, then the solution $\mathbf{x}^\star$ to the convex optimisation problem

$$\min_{\tilde{\mathbf{x}}} \|\tilde{\mathbf{x}}\|_1 \ : \ \|\mathbf{y} - \mathbf{\Phi}\tilde{\mathbf{x}}\|_2 \leq \|\mathbf{e}\|_2 \tag{3}$$

has an error bounded by

$$\|\mathbf{x}^\star - \mathbf{x}\| \leq cK^{-0.5}\|\mathbf{x} - \mathbf{x}_K\|_1 + c'\|\mathbf{e}\|, \tag{4}$$

where $\|\cdot\|_1$ is the vector 1 norm, $\mathbf{x}_K$ is the best K term approximation to $\mathbf{x}$ and where c and c' are two constants depending only on δ_{2K}.

Similar results have been obtained for other algorithms, such as the CoSaMP and Subspace Pursuit algorithms [3, 4] and the Iterative Hard Thresholding algorithm [5].

The question now is, can we derive similar results for a more general setup. For example, what if our signal $\mathbf{x}$ is not an element of Euclidean space? What if $\mathbf{x}$ is not sparse, but has some other known structure? What if we want to measure the error $\mathbf{e}$ using some other norm than the standard euclidean norm? What if $\mathbf{\Phi}$ is non-linear?

In this paper, we show that all of these restrictions can be relaxed and that similar results to those derived in compressed sensing also hold much more generally. We here assume that $\mathbf{x}$ is an element of a general Hilbert space, that $\mathbf{x}$ lies in, or close to, a union of subspaces $\mathcal{A}$, that we measure the error $\mathbf{e}$ with a more general norm and that $\mathbf{\Phi}$ is a non-linear operator. Similar to the compressed sensing results from the literature, we assume that $\mathbf{\Phi}$ satisfies a certain condition for all elements of $\mathcal{A}$. This condition is a generalisation of the Restricted Isometry Property and is associated with two constants. If these two constants are not too dissimilar, then we show that a projected Landweber algorithm (a generalisation of the Iterative Hard Thresholding algorithm) can be used to recover $\mathbf{x}$ with near optimal accuracy.

2 Several Generalisations

2.1 Non-linear measurements

We here generalise the notions of compressed sensing. We start by relaxing the requirement that $\mathbf{x}$ and $\mathbf{y}$ are finite dimensional vectors in Euclidean space. Instead, we assume $\mathbf{x}$ to be an element from some Hilbert spaces H and assume that the measurements $\mathbf{y}$ lie in a Banach space B. In this setting, we need to use a more general description for the sampling system and assume that $\mathbf{\Phi}$ is a possibly *nonlinear* mapping between H and B. In this setting, the observations are then

$$\mathbf{y} = \mathbf{\Phi}(\mathbf{x}) + \mathbf{e}, \tag{5}$$

where $\mathbf{e} \in B$ is an unknown error term.

2.2 A union of subspaces signal model

Again, we are interested in sampling systems $\mathbf{\Phi}$ that are non-invertible or ill-conditioned. To cope with this, additional constraints need to be imposed on $\mathbf{x}$. Instead of restricting our discussion to sparse signals (however we might define these in general Hilbert spaces) we instead assume that $\mathbf{x}$ lies in or close to a known set $\mathcal{A}$, where $\mathcal{A} \subset H$ is a possibly non-convex subset of H.

In this paper, our results are derived for models in which $\mathcal{A}$ is the union of subspaces, that is, for arbitrary closed subspaces $\mathcal{A}_i \subset \mathcal{H}$, we have

$$\mathcal{A} = \bigcup \mathcal{A}_i. \tag{6}$$

This more general model includes many problems of interest. See for example [7] and the references therein. Whilst we here restrict the discussion to union of subspaces, following the approach in [8], more general non-convex sets $\mathcal{A}$ could be considered and we conjecture that our main result might also hold in this more general setting, though our proof currently relies on the subspace structure.

2.3 The Cost and Subgradient

Our next generalisation is in the way we measure the error in the observation space. In fact, we will allow distance to be measured in any norm, which we will write as $\|\cdot\|_B$ from now on. Of importance is then the error $\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x})\|_B$ and the way it changes when we vary $\mathbf{x}$. To measure this variation, we assume that for each $\mathbf{y}$ and $\mathbf{x}$, the squared norm $\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x})\|_B^2$ has a subgeradient with respect to $\mathbf{x}$, that is, there is an element $\nabla(\mathbf{x}) \in H$ that, for fixed $\mathbf{y}$ and arbitrary $\mathbf{x}_1$ and $\mathbf{x}_2$ satisfies the condition

$$Re\langle\nabla(\mathbf{x}_1), \mathbf{x}_2\rangle + \|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}_1)\|_B^2 \leq \|\mathbf{y} - \mathbf{\Phi}(x_1 + \mathbf{x}_2)\|_B^2. \tag{7}$$

2.4 Projections onto the Constraints

One approach to recover $\mathbf{x}$ from $\mathbf{y}$ would be to mirror compressed sensing ideas and to define a convex objective function which can then be optimised using standard tools. However, for our general setup, this is difficult. Instead, we use a generalisation of the Iterative Hard Thresholding algorithm. To define this, we need to replace the thresholding step with a more general map which can be understood as a form of projection. Let $P_{\mathcal{A}}^{\epsilon}$ be a map from H to $\mathcal{A}$ such that

$$\mathbf{x}_{\mathcal{A}} = P_{\mathcal{A}}^{\epsilon}(\mathbf{x}) \; : \; \mathbf{x}_{\mathcal{A}} \in \mathcal{A}, \; \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|^2 \leq \inf_{\hat{\mathbf{x}} \in \mathcal{A}} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + \epsilon. \tag{8}$$

Note that we require an arbitrary small $\epsilon > 0$ term in our general setup, as there might not exist an $\mathbf{x}_{opt}$, such that $\|\mathbf{x} - \mathbf{x}_{opt}\|^2 = \inf_{\hat{\mathbf{x}} \in \mathcal{A}} \|\mathbf{x} - \hat{\mathbf{x}}\|^2$. For sets $\mathcal{A}$ for which the existence of these optimal points are guaranteed (so called proximal sets), we can take $\epsilon = 0$. Note also that this map might not be defined uniquely, as for a given $\mathbf{x}$, there might be several elements $\mathbf{x}_{\mathcal{A}}$ that satisfy the condition in (8). However, all we require here is that the map $P_{\mathcal{A}}^{\epsilon}(\mathbf{x})$ returns a single element from the set of admissible $\mathbf{x}_{\mathcal{A}}$ (which is guaranteed to be non-empty [8]). How this selection is done is of no consequence for our arguments here.

2.5 The Projected Landweber Algorithm

We are now in a position to define an algorithmic strategy to recover an element $\mathbf{x}$ from its (noisy) measurements $\mathbf{y}$. We replace the Hard Thresholding step in the Iterative Hard Thresholding algorithm with the more general 'projection' $P_{\mathcal{A}}^{\epsilon}(\mathbf{x})$ and also include the appropriate subgradient term. The algorithm is then more appropriately called the Projected Landweber Algorithm [9], which we defined by the iteration

$$\mathbf{x}^{n+1} = P_{\mathcal{A}}^{\epsilon}(\mathbf{x}^n - (\mu/2)\nabla(\mathbf{x}^n)), \tag{9}$$

where $\mathbf{x}^0 = \mathbf{0}$ and μ is a step size parameter chosen to satisfy the condition in Theorem 1.

2.6 The Recovery Condition RSGP

Finally, in order to study the performance of the Projected Landweber Algorithm, we generalise the Restricted Isometry Property. This is done using the *Restricted Strong Convexity Property* (RSGP) recently introduced in [10]. The

Restricted Strong Convexity Constants α and β are the largest respectively smallest constants for which [10]

$$\alpha \leq \frac{\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_1)\|_B^2 - \|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_2)\|_B^2 - Re\langle \nabla(\mathbf{x}_2), (\mathbf{x}_1-\mathbf{x}_2)\rangle}{\|\mathbf{x}_1-\mathbf{x}_2\|^2} \leq \beta, \tag{10}$$

holds for all $\mathbf{x}_1, \mathbf{x}_2$ for which $\mathbf{x}_1+\mathbf{x}_2 \in \mathcal{A}+\mathcal{A}$, where the set $\mathcal{A}+\mathcal{A} = \{\mathbf{x} = \mathbf{x}_a + \mathbf{x}_b : \mathbf{x}_a, \mathbf{x}_b \in \mathcal{A}\}$. Importantly, we *don't* require that the $\mathbf{x}_1$ and $\mathbf{x}_2$ are in $\mathcal{A}$! Note that the (non-symmetric) Restricted Isometry property is recovered if $\mathcal{H}$ and $\mathcal{B}$ are Euclidean spaces, if $\mathbf{\Phi}$ is linear and if $\|\cdot\|_B$ is the Euclidean norm.

Note that the main result in the next section requires the *Restricted Strong Convexity Property* to hold for all vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, such that $\mathbf{x}_1 - \mathbf{x}_2 \in \mathcal{A}+\mathcal{A}+\mathcal{A}$, where the set $\mathcal{A}+\mathcal{A}+\mathcal{A} = \{\mathbf{y} = \mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3 : \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathcal{A}\}$. This is obviously somewhat more restrictive than the requirement that it holds for $\mathbf{x}_1 - \mathbf{x}_2 \in \mathcal{A}+\mathcal{A}$. We believe this restriction to be an artifact of the proof and we conjecture that a similar result can be derived for problems for which the *Restricted Strong Convexity Property* holds for $\mathbf{x}_1 - \mathbf{x}_2 \in \mathcal{A}+\mathcal{A}$.

3 Main Result

Our main result states that, if $\mathbf{\Phi}$ and $\|\cdot\|_{\mathcal{B}}$ satisfy the Restricted Strong Convexity Property, then the Projected Landweber Algorithm can recover any signal in H from measurements $\mathbf{y} = \mathbf{\Phi}(\mathbf{x}) + \mathbf{e}$ with an error proportional to the size of $\mathbf{e}$ and the size of the error $\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_B$, where $\mathbf{x}_{\mathcal{A}} = P_{\mathcal{A}}^{\epsilon}(\mathbf{x})$. That is, if $\mathbf{x}$ is close to the constraint set $\mathcal{A}$ and if the error $\mathbf{e}$ is small, then we can recover $\mathbf{x}$ with small error. In particular, we have the following theorem.

Theorem 1. *Let $\mathcal{A}$ be a union of subspaces. Given $\mathbf{y} = \mathbf{\Phi}(\mathbf{x}) + \mathbf{e}$ where $\mathbf{x}$ is arbitrary, assume $\mathbf{\Phi}$ and $\|\cdot\|_{\mathcal{B}}^2$ satisfy the* Restricted Strict Convexity Property

$$\alpha \leq \frac{\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_1)\|_B^2 - \|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_2)\|_B^2 - Re\langle \nabla(\mathbf{x}_2), (\mathbf{x}_1-\mathbf{x}_2)\rangle}{\|\mathbf{x}_1-\mathbf{x}_2\|^2} \leq \beta, \tag{11}$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{H}$ for which $\mathbf{x}_1+\mathbf{x}_2 \in \mathcal{A}+\mathcal{A}+\mathcal{A}$ with constants $\beta \leq \frac{1}{\mu} \leq \frac{4}{3}\alpha$, then, after

$$n^\star = 2\frac{\ln\left(\delta\frac{\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}}}{\|\mathbf{x}_{\mathcal{A}}\|}\right)}{\ln 4(1-\mu\alpha)}, \tag{12}$$

iterations, the Projected Landweber Algorithm calculates a solution $\mathbf{x}^{n^\star}$ satisfying

$$\|\mathbf{x}^{n^\star} - \mathbf{x}\| \leq (2\sqrt{\mu}+\delta)\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}} + \|\mathbf{x}-\mathbf{x}_{\mathcal{A}}\| + \sqrt{2\epsilon}. \tag{13}$$

In the traditional compressed sensing setting, this result is basically that derived in [5].

In summary, our main result states that, under similar conditions on the *Restricted Strong Convexity Property* to those imposed on the Restricted Isometry Property in compressed sensing, the Projected Landweber Algorithm can recover the elements from $\mathcal{H}$ with a similar bound to that achieved for the recovery of sparse finite dimensional vectors in compressed sensing. The only difference in our more general setting is that the error depends on two terms $\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_B^2$ and $\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}\|$, rather than the error $K^{-0.5}\|\mathbf{x} - \mathbf{x}_{2K}\|_1$ achieved in traditional compressed sensing. As discussed in [7], this is expected for our general setting. Whilst for traditional compressed sensing, $\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_B^2$ is of the same order as $K^{-0.5}\|\mathbf{x} - \mathbf{x}_{2K}\|_1$, better bounds for the general setting would require the imposition of additional properties on the sampling system similar to the *Restricted Amplification Property* introduced for model based compressed sensing [11].

3.1 Proof of the Main Result

Proof of Theorem 1. The proof requires the orthogonal projection onto a subspace Γ. The subspace Γ is defined as follows. Let Γ be the direct sum of no more than three subspaces of $\mathcal{A}$, such that $\mathbf{x}_{\mathcal{A}}, \mathbf{x}^n, \mathbf{x}^{n+1} \in \Gamma$. Let P_Γ be the orthogonal projection onto the subspace Γ. We write $\mathbf{a}_\Gamma^n = P_\Gamma \mathbf{a}^n$ and $P_\Gamma \nabla(\mathbf{x}^n) = \nabla_\Gamma(\mathbf{x}^n)$. Note that this ensures that $P_\Gamma \mathbf{x}^n = \mathbf{x}^n$, $P_\Gamma \mathbf{x}^{n+1} = \mathbf{x}^{n+1}$ and $P_\Gamma \mathbf{x}_{\mathcal{A}} = \mathbf{x}_{\mathcal{A}}$.

We note for later that with this notation

$$
\begin{aligned}
Re\langle \nabla_\Gamma(\mathbf{x}^n), (\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\rangle &= Re\langle P_\Gamma \nabla(\mathbf{x}^n), (\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\rangle \\
&= Re\langle \nabla(\mathbf{x}^n), P_\Gamma(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\rangle \\
&= Re\langle \nabla(\mathbf{x}^n), (\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\rangle \qquad (14)
\end{aligned}
$$

and

$$
\begin{aligned}
\|\nabla_\Gamma(\mathbf{x}^n)\|^2 = \langle \nabla_\Gamma(\mathbf{x}^n), \nabla_\Gamma(\mathbf{x}^n)\rangle &= \langle P_\Gamma \nabla(\mathbf{x}^n), P_\Gamma \nabla(\mathbf{x}^n)\rangle \\
&= \langle \nabla(\mathbf{x}^n), P_\Gamma^* P_\Gamma \nabla(\mathbf{x}^n)\rangle \\
&= \langle \nabla(\mathbf{x}^n), \nabla_\Gamma(\mathbf{x}^n)\rangle, \qquad (15)
\end{aligned}
$$

We also need the following lemma.

Lemma 2. *Under the assumptions of the theorem,*

$$
\|\frac{\mu}{2}\nabla_\Gamma(\mathbf{x}^n)\|^2 - \mu\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}^n)\|_{\mathcal{B}}^2 \leq 0. \qquad (16)
$$

Proof. Using the *Restricted Strict Convexity Property* we have

$$
\begin{aligned}
\|\frac{\mu}{2}\nabla_\Gamma(\mathbf{x}^n)\|^2 &= -\frac{\mu}{2} Re\langle \nabla(\mathbf{x}^n), -\frac{\mu}{2}\nabla_\Gamma(\mathbf{x}^n)\rangle \\
&\leq \frac{\mu}{2}\beta\|\frac{\mu}{2}\nabla_\Gamma(\mathbf{x}^n)\|^2 + \frac{\mu}{2}\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}^n)\|_{\mathcal{B}}^2 - \frac{\mu}{2}\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}^n - \frac{\mu}{2}\nabla_\Gamma(\mathbf{x}^n))\|_{\mathcal{B}}^2 \\
&\leq \frac{\mu}{2}\beta\|\frac{\mu}{2}\nabla_\Gamma(\mathbf{x}^n)\|^2 + \frac{\mu}{2}\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}^n)\|_{\mathcal{B}}^2. \qquad (17)
\end{aligned}
$$

Thus

$$(2-\mu\beta)\|\frac{\mu}{2}\nabla_\Gamma(\mathbf{x}^n)\|^2 \quad \leq \quad \mu\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}^n)\|_{\mathcal{B}}^2, \tag{18}$$

which is the desired result as $\mu\beta \leq 1$ by assumption. □

To proof the theorem, we start by bounding the distance between the current estimate $\mathbf{x}^{n+1}$ and the optimal estimate $\mathbf{x}_{\mathcal{A}}$. Let $\mathbf{a}_\Gamma^n = \mathbf{x}_\Gamma^n - \mu/2\nabla_\Gamma(\mathbf{x}^n)$. Because $\mathbf{x}^{n+1}$ is, up to ϵ the closest element in $\mathcal{A}$ to $\mathbf{a}_\Gamma^n$, we have

$$\begin{aligned}
\|\mathbf{x}^{n+1}-\mathbf{x}_{\mathcal{A}}\|^2 &\leq \left(\|\mathbf{x}^{n+1}-\mathbf{a}_\Gamma^n\| + \|\mathbf{a}_\Gamma^n-\mathbf{x}_{\mathcal{A}}\|\right)^2 \\
&\leq 4\|(\mathbf{a}_\Gamma^n-\mathbf{x}_{\mathcal{A}})\|^2 + 2\epsilon \\
&= 4\|\mathbf{x}^n-(\mu/2)\nabla_\Gamma(\mathbf{x}^n)-\mathbf{x}_{\mathcal{A}}\|^2 + 2\epsilon \\
&= 4\|(\mu/2)\nabla_\Gamma(\mathbf{x}^n)+(\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n)\|^2 + 2\epsilon \\
&= \mu^2\|\nabla_\Gamma(\mathbf{x}^n)\|^2 + 4\|\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n\|^2 + 4\mu Re\langle\nabla_\Gamma(\mathbf{x}^n),(\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n)\rangle + 2\epsilon \\
&= \mu^2\|\nabla_\Gamma(\mathbf{x}^n)\|^2 + 4\|\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n\|^2 + 4\mu Re\langle\nabla(\mathbf{x}^n),(\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n)\rangle + 2\epsilon \\
&\leq 4\|\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n\|^2 + \mu^2\|\nabla_\Gamma(\mathbf{x}^n)\|^2 \\
&\quad +4\mu[-\alpha\|\mathbf{x}^n-\mathbf{x}_{\mathcal{A}}\|^2 + \|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}}^2 - \|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}^n)\|_{\mathcal{B}}^2] + 2\epsilon \\
&= 4(1-\mu\alpha)\|\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n\|^2 + 4\mu\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}}^2 + 2\epsilon \\
&\quad +4[\|(\mu/2)\nabla_\Gamma(\mathbf{x}^n)\|^2 - \mu\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}^n)\|_{\mathcal{B}}^2] \\
&\leq 4(1-\mu\alpha)\|\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n\|^2 + 4\mu\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}}^2 + 2\epsilon.
\end{aligned} \tag{19}$$

Here, the second to last inequality is the RSCP and the last inequality is due to lemma 2.

We have thus shown that

$$\|\mathbf{x}^{n+1}-\mathbf{x}_{\mathcal{A}}\|^2 \leq 4(1-\mu\alpha)\|\mathbf{x}_{\mathcal{A}}-\mathbf{x}^n\|^2 + 4\mu\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}}^2 + 2\epsilon. \tag{20}$$

Thus, with $c = 4(1-\mu\alpha)$

$$\|\mathbf{x}^k-\mathbf{x}_{\mathcal{A}}\|^2 \leq c^k\|\mathbf{x}_{\mathcal{A}}\|^2 + 4\mu\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}}^2 + 2\epsilon, \tag{21}$$

so that, if $\frac{1}{\mu} < \frac{4}{3}\alpha$ we have $c = 4(1-\mu\alpha) < 1$, so that c^k decreases with k. Taking the square root on both sides and noting that for positive a and b, $\sqrt{a^2+b^2} \leq a+b$,

$$\|\mathbf{x}^k-\mathbf{x}_{\mathcal{A}}\| \leq c^{k/2}\|\mathbf{x}_{\mathcal{A}}\| + 2\sqrt{\mu}\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}} + \sqrt{2\epsilon}, \tag{22}$$

The theorem then follows using the triangle inequality

$$\begin{aligned}
\|\mathbf{x}^{n+1}-\mathbf{x}\| &\leq \|\mathbf{x}^{n+1}-\mathbf{x}_{\mathcal{A}}\| + \|\mathbf{x}-\mathbf{x}_{\mathcal{A}}\| \\
&\leq c^{k/2}\|\mathbf{x}_{\mathcal{A}}\| + 2\sqrt{\mu}\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}} + \sqrt{2\epsilon} + \|\mathbf{x}-\mathbf{x}_{\mathcal{A}}\|.
\end{aligned} \tag{23}$$

The iteration count is found by setting

$$c^{k/2}\|\mathbf{x}_{\mathcal{A}}\| \leq \delta\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}}. \tag{24}$$

so that after

$$k = 2\frac{\ln\left(\delta\frac{\|\mathbf{y}-\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}}}{\|\mathbf{x}_{\mathcal{A}}\|}\right)}{\ln c}, \tag{25}$$

iterations

$$\|\mathbf{x}^k - \mathbf{x}\| \leq (2\sqrt{\mu} + \delta)\|\mathbf{y} - \mathbf{\Phi}(\mathbf{x}_{\mathcal{A}})\|_{\mathcal{B}} + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\| + \sqrt{2\epsilon}. \tag{26}$$

□

4 Conclusion

Compressed sensing ideas can be developed in much more general settings than considered traditionally. We have shown previously [7, 8] that sparsity is not the only structure that allows signals to be recovered and that the finite dimensional setting can be replaced with a much more general Hilbert space framework. In this paper we have made a further important generalisation and have introduced the concept of non-linear measurements into compressed sensing theory. Whenever the measurements and constraints satisfy a generalised version of the Restricted Isometry Property, then an Projective Landweber Algorithm can be used to recover signals that lie on or close to the constraint set.

We have here developed the main result based on ideas from [5] and [7], however, it appears reasonable to assume that a similar result also holds for more general sets $\mathcal{A}$ and for a RSCP with respect to $\mathcal{A}+\mathcal{A}$ rather than $\mathcal{A}+\mathcal{A}+\mathcal{A}$. We are currently working towards such a more general result based on the approaches in [8] and [12].

Acknowledgment

The author acknowledges support of his position from the School of Mathematics at the University of Southampton.

References

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

[2] E. Candès, "The restricted isometry property and its implications for compressed sensing.," *Compte Rendus de l'Academie des Sciences*, Paris, Serie I, 346 589–592. 2008.

[3] D. Needell and J. Tropp, "COSAMP: Iterative signal reovery from incomplete and inacurate samples.," *Appl. Comp. Harmonic Anal*, vol. 26, pp. 301–321, 2008.

[4] W. Dai and O. Milenkovic, "Subspace pursuit for compressed sensing: Closing the gap between performance and complexity," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[5] T. Blumensath and M.E. Davies "Iterative Hard Thresholding for Compressed Sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265-274, 2009

[6] E. Candès and M. B. Wakin "An introduction to compressive sampling," *IEEE Signal Processig Magazine*, 25(2), pp. 21–30, 2008

[7] T. Blumensath, "Sampling and reconstructing signals form a union of subspaces," *submitted*, 2010

[8] T. Blumensath, "Non-convexly constrained inverse problems," *submitted*, 2010

[9] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, 2000.

[10] T. Zhang, "Sparse Recovery with Orthogonal Matching Pursuit under RIP," *arXiv*:1005.2249v1

[11] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, Vol 56, no. 4, pp. 1982–2001, 2010

[12] S. Fucart, "Sparse Recovery Algorithms: Sufficient Conditions in terms of the Restricted Isometry Constants," *preprint*, 2010