Analysis of the Self Projected Matching Pursuit Algorithm

Laura Rebollo-Neira
Mathematics Department, Aston University
B4 7ET, Birmingham, United Kingdom

Miroslav Rozložník
Institute of Mathematics, Czech Academy of Sciences
Žitná 25, CZ – 115 67 Praha 1, Czech Republic

Pradip Sasmal
Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore
Karnataka, 560012, India

April 24, 2020

Abstract

The convergence and numerical analysis of a low memory implementation of the Orthogonal Matching Pursuit greedy strategy, which is termed Self Projected Matching Pursuit, is presented. This approach renders an iterative way of solving the least squares problem with much less storage requirement than direct linear algebra techniques. Hence, it is appropriate for solving large linear systems. The analysis highlights its suitability within the class of well posed problems.

Keywords: Sparse Representation; Greedy Pursuit Strategies; Orthogonal Matching Pursuit; Self Projected Matching Pursuit; Least Squares of Large Systems; Iterative Projections.

1 Introduction

Sparse representation refers to the process by which a signal is transformed in order to reduce its dimensionality. Traditional methods implement the transformation using fast orthogonal transforms. Higher levels of sparsity are attained, in many cases, if the transformation is carried out using a large redundant set called a dictionary. For the most part this method is implemented by minimization of the l_1-norm [1–3] and the so-called greedy strategies. The latter consist in adaptively constructing a signal representation as a linear superposition of elements taken from the dictionary. In this contribution we focus on the analysis of a low memory implementation of a particular method within this category.

Greedy strategies have been the subject of extensive research in the last two decades [4–16] and currently support diverse applications [17–19]. The simplest, yet very effective greedy

algorithm for the sparse representation of large signals, was introduced to the signal processing community in [4] with the name of Matching Pursuit (MP). It had previously appeared as a regression technique in statistics [20, 21], where the convergence property was established. While MP converges asymptotically to a signal in the linear span of the dictionary, or to its orthogonal projection if the signal is out of that space, the approach is not stepwise optimal because it does not yield an orthogonal projection at each step. A refinement to MP which fulfils this requirement is refereed to as Orthogonal Matching Pursuit (OMP) [5]. If implemented by direct methods the OMP approach is very effective up to some dimensionality. When processing large signals, however, the storage requirements frequently exceed the memory capacity of a standard computer. An alternative implementation of OMP, which requires much less memory than direct implementations is considered in [22]. The approach is termed Self Projected Matching Pursuit (SPMP). It produces the orthogonal projection of the signal, at each iteration, by applying MP using a sub-dictionary consisting only of the already selected elements. A convenient feature of SPMP when applied in 2D (SPMP2D) [22, 23] and 3D (SPMP3D) [24] is that it fully exploits the separability of dictionaries. Nevertheless, until now the method had not been analyzed. Thus, the main contributions of this paper are:

- The convergence analysis of the SPMP approach, which deals with those cases where the standard implementation of the OMP method is not feasible due to storage requirements.
- The error analysis of the SPMP approach.

Additionally, the SPMP approach is extended to consider its Hierarchized Block Wise (HBW) version for approximating a signal partition subjected to a global constraint on sparsity.

The paper is organized as follows: Sec. 2 recalls the SPMP algorithm, proves the power law for the convergence rate of the self projection step and develops its numerical analysis. In Sec. 3 the applicability of the method is extended by dedicating the algorithm to the approximation of non stationary signals by partitioning. The final conclusions are presented in Sec. 4.

2 Self Projected Matching Pursuit (SPMP)

Throughout the paper $\mathbb{R}$ and $\mathbb{N}$ represent the sets of real and natural numbers, respectively. Boldface fonts are used to indicate Euclidean vectors or matrices and standard mathematical fonts to indicate components, e.g., $\mathbf{d} \in \mathbb{R}^N$ is a vector of N-components $d(i) \in \mathbb{R}\,, i = 1, \ldots, N$ and $\mathbf{A} \in \mathbb{R}^{N_x \times N_y}$ a matrix of elements $A(i,j) \in \mathbb{R}\,,\ i = 1, \ldots, N_x,\ j = 1, \ldots, N_y$. The transpose of $\mathbf{A}$ is denoted as $\mathbf{A}^\top$. The operation $\langle \cdot, \cdot \rangle$ indicates the Euclidean inner product and $\| \cdot \|$ the induced norm, i.e. $\|\mathbf{d}\|^2 = \langle \mathbf{d}, \mathbf{d} \rangle$, with the usual inner product definition: For $\mathbf{g} \in \mathbb{R}^N$ and $\mathbf{f} \in \mathbb{R}^N$

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{i=1}^{N} f(i) g(i). \tag{1}$$

Let's consider a finite set $\mathcal{D}$ of M of normalized vectors $\mathcal{D} = \{\mathbf{d}_n \in \mathbb{R}^N ; \|\mathbf{d}_n\| = 1\}_{n=1}^M$ and let's define $\mathbb{S}_M = \text{span}(\mathcal{D})$, which could be $\mathbb{R}^N$. For $M > \dim(\mathbb{S}_M)$ the set $\mathcal{D}$ is a redundant dictionary and the elements are called *atoms*. Given a signal, as a vector $\mathbf{f} \in \mathbb{R}^N$, the k-term *atomic decomposition* for its approximation takes the form

$$\mathbf{f}^k = \sum_{j=1}^{k} c(j) \mathbf{d}_{\ell_j}. \tag{2}$$

The problem of how to select from $\mathcal{D}$ the smallest number of k atoms $\mathbf{d}_{\ell_j}$, $j = 1 \ldots, k$, such that $\|\mathbf{f}^k - \mathbf{f}\| < \rho$, for a given tolerance parameter ρ, is an NP-hard problem [6]. In practical applications one looks for 'tractable sparse' solutions. This is to say a representation involving a number of k-terms, with k acceptably small in relation to N. The simplest approach to tackle this problem is MP. It evolves by successive approximations as follows [4]: Setting $k = 0$ and starting with an initial approximation $\mathbf{f}^0 = 0$ and residual $\mathbf{r}^0 = \mathbf{f}$, the algorithm progresses by sub-decomposing the k-th order residual in the form

$$\mathbf{r}^k = \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \mathbf{d}_{\ell_{k+1}} + \mathbf{r}^{k+1}, \tag{3}$$

where $\mathbf{d}_{\ell_{k+1}}$ is the atom corresponding to the index selected as

$$\ell_{k+1} = \underset{n=1,\ldots,M}{\arg\max} \left| \left\langle \mathbf{d}_n, \mathbf{r}^k \right\rangle \right|. \tag{4}$$

This atom is used to update the approximation $\mathbf{f}^k$ as

$$\mathbf{f}^{k+1} = \mathbf{f}^k + \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \mathbf{d}_{\ell_{k+1}}. \tag{5}$$

From (3) it follows that $\|\mathbf{r}^{k+1}\| \leq \|\mathbf{r}^k\|$, since

$$\|\mathbf{r}^k\|^2 = \left| \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \right|^2 + \|\mathbf{r}^{k+1}\|^2. \tag{6}$$

Lemma 1. *In the limit $k \to \infty$, the sequence $\mathbf{f}^k$ given in (5) converges to $\mathbf{f}$, if $\mathbf{f} \in \mathbb{S}_M$, or to $\hat{\mathbf{P}}_{\mathbb{S}_M}\mathbf{f}$, the orthogonal projection of $\mathbf{f}$ onto $\mathbb{S}_M$, if $\mathbf{f} \notin \mathbb{S}_M$.*

This lemma is just a particular case of the well established and more general convergence results for MP [4, 8, 21]. However, for pedagogical reasons, due to its crucial importance for this work, we present here a particular proof holding *only* for finite dimension spaces which, for this reason, is very simple.

Proof. We notice, from (6), that $\|\mathbf{r}^k\|^2$ is a decreasing sequence which, since $\|\mathbf{r}^k\|^2 \geq 0$ for all k, is bounded. It is a classic result of analysis that a decreasing and bounded sequence converges to the infimum [28], i.e., $\lim_{k\to\infty} \|\mathbf{r}^k\|^2 = b$. We prove next that $b = 0$. Since

$$\|\mathbf{r}^{k+1}\|^2 = \|\mathbf{r}^k\|^2 - \left| \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \right|^2,$$

taking $\lim_{k\to\infty}$ of both sizes, we have:

$$b^2 = b^2 - \lim_{k\to\infty} \left| \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \right|^2.$$

Thus, $\lim_{k\to\infty} \left| \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \right| = 0$, which using (4) implies $\lim_{k\to\infty} \left| \left\langle \mathbf{d}_n, \mathbf{r}^k \right\rangle \right| = 0$, $n = 1, \ldots, M$. Consequently, either $\lim_{k\to\infty} \mathbf{r}^k = 0$ or, if the dictionary is incomplete, $\lim_{k\to\infty} \mathbf{r}^k$ is orthogonal to all the elements in $\mathcal{D}$. This result is readily obtainable here, because of the finite dimension framework. Indeed, in finite dimension the existence of a reciprocal (dual) dictionary $\tilde{\mathcal{D}} = \{\tilde{\mathbf{d}}_n \in \mathbb{R}^N\}_{n=1}^M$ spanning the same space as $\mathcal{D}$ is guaranteed [26, 27]. Hence, even if due to the redundancy of $\mathcal{D}$ the decomposition is not unique, all $\mathbf{g} \in \mathbb{S}_M = \text{span}(\mathcal{D}) = \text{span}(\tilde{\mathcal{D}})$ can be decomposed in the form

$$\mathbf{g} = \sum_{n=1}^{M} \mathbf{d}_n \left\langle \tilde{\mathbf{d}}_n, \mathbf{g} \right\rangle = \sum_{n=1}^{M} \tilde{\mathbf{d}}_n \left\langle \mathbf{d}_n, \mathbf{g} \right\rangle.$$

Furthermore, every vector in $\mathbb{R}^N$, and in particular $\mathbf{r}^k$, can be split as $\mathbf{r}^k = \hat{\mathbf{P}}_{\mathbb{S}_M}\mathbf{r}^k + \hat{\mathbf{P}}_{\mathbb{S}_M^\perp}\mathbf{r}^k$, where $\hat{\mathbf{P}}_{\mathbb{S}_M}\mathbf{r}^k$ is the orthogonal projection onto $\mathbb{S}_M$ and $\hat{\mathbf{P}}_{\mathbb{S}_M^\perp}\mathbf{r}^k$ is the orthogonal projection onto the subspace $\mathbb{S}_M^\perp$, which is the orthogonal complement of $\mathbb{S}_M$ in $\mathbb{R}^N$. From the relation

$$\hat{\mathbf{P}}_{\mathbb{S}_M}\mathbf{r}^k = \sum_{n=1}^{M} \mathbf{d}_n \left\langle \tilde{\mathbf{d}}_n, \mathbf{r}^k \right\rangle = \sum_{n=1}^{M} \tilde{\mathbf{d}}_n \left\langle \mathbf{d}_n, \mathbf{r}^k \right\rangle,$$

and because it involves a *finite* sum, we conclude that $\lim_{k\to\infty} |\left\langle \mathbf{d}_n, \mathbf{r}^k \right\rangle| = 0$, $n = 1, \ldots, M \implies \lim_{k\to\infty} \hat{\mathbf{P}}_{\mathbb{S}_M}\mathbf{r}^k = 0$. Then, either $\lim_{k\to\infty} \mathbf{r}^k = 0$ or $\lim_{k\to\infty} \mathbf{r}^k \in \mathbb{S}_M^\perp$. Consequently, since $\mathbf{f}^k = \mathbf{f} - \mathbf{r}^k \in \mathbb{S}_M$, it follows that $\lim_{k\to\infty} \mathbf{f}^k = \hat{\mathbf{P}}_{\mathbb{S}_M}\mathbf{f}$. □

2.1 Adding Self Projections

The obvious way of improving the MP algorithm is to calculate the coefficients in (2) so as to minimize the norm of the residual error $\|\mathbf{f} - \mathbf{f}^k\|$ for every value of k. In other words, to require that, at each iteration, the coefficients in (2) should fulfill the condition $\mathbf{f}^k = \hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{f}$, where $\mathbb{S}_k = \text{span}\{\mathbf{d}_{\ell_j}\}_{j=1}^k$. Hence the name, OMP, of the approach achieving this. When the dimension of the problem is such that memory requirement is not an issue, a number of convenient direct linear algebra methods for performing the projection $\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{f}$ are available [29–31]. However, it is the need of calculating orthogonal projections with much less storage demands than direct methods what originated the SPMP approach described below.

SPMP relays on Lemma 1 to realize the orthogonal projection step and produces an alternative iterative implementation of the OMP approach. Given a signal $\mathbf{f}$, a tolerance error ρ for the approximation, and a dictionary $\mathcal{D}$, the SPMP algorithm proceeds as follows [22]: Set $\mathcal{L}_0 = \{\emptyset\}$, $\mathbf{f}^0 = 0$ and $\mathbf{r}^0 = \mathbf{f}$. Starting from $k = 0$, at each iteration implement the steps below.

i) While $\|\mathbf{r}^k\| > \rho$ increment $k \leftarrow k+1$ and apply the MP criterion for selecting from $\mathcal{D}$ the atom $\mathbf{d}_{\ell_k}$ to be placed in the atomic decomposition i.e., select ℓ_k such that

$$\ell_k = \underset{n=1,\ldots,M}{\arg\max} |\left\langle \mathbf{d}_n, \mathbf{r}^{k-1} \right\rangle|. \tag{7}$$

Update the set $\mathcal{L}_k = \mathcal{L}_{k-1} \cup \{\ell_k\}$. Compute $c(k) = \left\langle \mathbf{d}_{\ell_k}, \mathbf{r}^{k-1} \right\rangle$, update the approximation of $\mathbf{f}$ as $\mathbf{f}^k = \mathbf{f}^{k-1} + c(k)\mathbf{d}_{\ell_k}$, and evaluate the new residual $\mathbf{r}^k = \mathbf{f} - \mathbf{f}^k$.

ii) Realize the orthogonal projection by subtracting from $\mathbf{r}^k$ the component in $\mathbb{S}_k = \text{span}\{\mathbf{d}_{\ell_i}\}_{i=1}^k$, via the MP algorithm, as follows. Let ϵ be a given tolerance for the projection error. Set $j = 1$, $\mathbf{r}^{k,0} = \mathbf{r}^k$ and at iteration j implement the steps below:

(a) Choose, out of the set $\mathcal{L}_k$, the index l_j such that

$$l_j = \underset{i=1,\ldots,k}{\arg\max} \left|\left\langle \mathbf{d}_{\ell_i}, \mathbf{r}^{k,j-1} \right\rangle\right|.$$

If $\left|\left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1} \right\rangle\right| < \epsilon$ set $\mathbf{r}^k \leftarrow \mathbf{r}^{k,j-1}$ and return to i). Otherwise continue with steps (b) and (c) as follows.

(b) Use $\left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1} \right\rangle$ to update the coefficient $c(l_j)$, the approximation $\mathbf{f}^k$, and the residual, as

$$\begin{aligned} c(l_j) &\leftarrow c(l_j) + \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1} \right\rangle, \\ \mathbf{f}^k &\leftarrow \mathbf{f}^k + \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1} \right\rangle \mathbf{d}_{l_j}, \\ \mathbf{r}^{k,j} &= \mathbf{r}^{k,j-1} - \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1} \right\rangle \mathbf{d}_{l_j}. \end{aligned}$$

(c) Increment $j \leftarrow j+1$ and repeat steps (a) $\rightarrow$ (c) until the stopping criterion is met.

As proved in Lemma 1, by means of the self-projections implemented by steps (a) – (c), at each iteration k the SPMP algorithm asymptotically delivers an approximation $\mathbf{f}^k = \hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{f}$ with residual $\mathbf{r}^k = \mathbf{f} - \hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{f}$. The next Lemma stresses the fact that, as a consequence, the SPMP algorithm selects only linearly independent atoms.

Lemma 2. *If the atoms $\mathbf{d}_{\ell_i}$, $i = 1, \ldots, k$ are selected by criterion (7), and the residual $\mathbf{r}^k$ is refined by self projections at each iteration, the selected atoms constitutes a linearly independent set.*

Proof. For $k = 1$ the lemma is triviality true. Assuming that it is true for the first k atoms we prove that it is true for $k+1$ atoms.

Suppose, on the contrary, that $\left|\left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle\right| > 0$ and $\mathbf{d}_{\ell_{k+1}} = \sum_{i=1}^{k} a_i \mathbf{d}_{\ell_i}$, where a_i, $i = 1, \ldots, k$ are numbers such that $\sum_{i=1}^{k} |a_i|^2 > 0$. Since at the iteration k the SPMP algorithm asymptotically gives a residual that satisfies $\mathbf{r}^k = \mathbf{f} - \hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{f}$ we have:

$$\left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle = \left\langle \sum_{i=1}^{k} a_i \mathbf{d}_{\ell_i}, \mathbf{f} - \hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{f} \right\rangle = 0,$$

which contradicts the assumption that $\left|\left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle\right| > 0$. It is concluded then that $\mathbf{d}_{\ell_{k+1}}$ cannot be expressed as a linear combination of the previously selected atoms. □

2.2 Convergence rate of the self projection steps

We start by recalling some properties of symmetric matrices, which will be used for the analysis. Let the atoms $\mathbf{d}_{\ell_i}$, $i = 1, \ldots, k$ be the columns of the matrix $\mathbf{S}_k$. Since the atoms are linearly independent, the symmetric matrix $\mathbf{H}_k = \mathbf{S}_k\mathbf{S}_k^\top$ has k nonzero eigenvalues, which are also the k eigenvalues of the Gram matrix $\mathbf{G}_k = \mathbf{S}_k^\top\mathbf{S}_k$. In terms of the corresponding eigenvectors $\mathbf{H}_k$ can be expressed as

$$\mathbf{H}_k = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^\top, \tag{8}$$

where $\mathbf{\Lambda}_k$ is a diagonal matrix, containing in the diagonal its eigenvalues $\lambda_i^k > 0$, $i = 1, \ldots, k$ in descending order. Since all the atoms are normalized, it holds that

$$\text{Trace}(\mathbf{H}_k) = \sum_{i=1}^{k} \lambda_i^k = k.$$

This relation implies that $k\lambda_k^k \leq k \leq k\lambda_1^k$, which ensures that $\lambda_k^k \leq 1$. The columns of matrix $\mathbf{U}_k$ are the normalized eigenvectors of $\mathbf{H}_k$ corresponding to the eigenvalues $\lambda_i^k > 0$, $i = 1, \ldots, k$. Since $\mathbf{H}_k$ is symmetric these eigenvectors constitute an orthonormal basis for $\mathbb{S}_k = \text{Range}(\mathbf{S}_k)$. Accordingly, the orthogonal projector $\hat{\mathbf{P}}_{\mathbb{S}_k}$ admits a representation of the form:

$$\hat{\mathbf{P}}_{\mathbb{S}_k} = \mathbf{U}_k\mathbf{U}_k^\top. \tag{9}$$

Then, the following inequality arises from (8) and (9),

$$\|\mathbf{S}_k^\top\mathbf{g}\|^2 = \left\langle \mathbf{g}, \mathbf{S}_k\mathbf{S}_k^\top\mathbf{g} \right\rangle \geq \lambda_k^k \|\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{g}\|^2, \quad \forall\, \mathbf{g} \in \mathbb{R}^N. \tag{10}$$

This inequality will be used for the analysis of the convergence rate of the self-projection step.

Proposition 1. *At iteration j the component in $\mathbb{S}_k$ of the residual $\mathbf{r}^{k,j}$ is bounded as*

$$\|\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,j}\|^2 \leq \left(1-\frac{\lambda_k^k}{k}\right)^j \|\mathbf{r}^{k,0}\|^2. \tag{11}$$

Proof. Let's recall that the projection step operates by setting $\mathbf{r}^{k,0}=\mathbf{r}^k$ and at the j-th iteration decomposing the residual $\mathbf{r}^{k,j}$ as

$$\mathbf{r}^{k,j} = \mathbf{r}^{k,j-1} - \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1}\right\rangle \mathbf{d}_{l_j}, \tag{12}$$

where

$$l_j = \underset{i=1,\ldots,k}{\arg\max} |\left\langle \mathbf{d}_{\ell_i}, \mathbf{r}^{k,j-1}\right\rangle|. \tag{13}$$

Since $\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{d}_{l_j}=\mathbf{d}_{l_j}$, applying the operator $\hat{\mathbf{P}}_{\mathbb{S}_k}$ on both sides of (12) we have,

$$\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,j} = \hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,j-1} - \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1}\right\rangle \mathbf{d}_{l_j},$$

and consequently

$$\|\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,j}\|^2 = \|\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,j-1}\|^2 - |\left\langle \mathbf{r}^{k,j-1}, \mathbf{d}_{l_j}\right\rangle|^2. \tag{14}$$

By definition of the index l_j (cf.(13)), and using (10), we assert that

$$|\left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1}\right\rangle|^2 \geq \frac{1}{k}\sum_{i=1}^{k} |\left\langle \mathbf{d}_i, \mathbf{r}^{k,j-1}\right\rangle| = \frac{1}{k}\|\mathbf{S}_k^\top \mathbf{r}^{k,j-1}\|^2 \geq \frac{\lambda_k^k}{k}\|\hat{\mathbf{P}}_{\mathcal{S}_k}\mathbf{r}^{k,j-1}\|^2.$$

Then, we finally obtain

$$\|\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,j}\|^2 \leq \left(1-\frac{\lambda_k^k}{k}\right)\|\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,j-1}\|^2, \tag{15}$$

and applying the inequality back j-times

$$\|\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,j}\|^2 \leq \left(1-\frac{\lambda_k^k}{k}\right)^j\|\hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{r}^{k,0}\|^2 \leq \left(1-\frac{\lambda_k^k}{k}\right)^j \|\mathbf{r}^{k,0}\|^2. \tag{16}$$

□

The bound (16) gives a power form for the worst-case convergence rate to a residual vector having no component in $\mathbb{S}_k$. It also shows the dependence of the convergence rate on the smallest eigenvalue of the Gram matrix $\mathbf{G}_k$ of the selected atoms up to iteration k. According to the interlacing theorem ([32], p 189–190) it is true that $\lambda_{k+1}^{k+1} < \lambda_k^k$. Hence, in general one could expect the convergence rate of the self projection to slow down as the iterative selection of atoms progresses.

Remark 1: The convergence of MP in terms of the dictionary's coherence [12] is derived in [13] for the case of quasi incoherent dictionaries. That condition is too stringent for signals of practical interest, which are far more compressible when using a highly coherent dictionary than when using an orthogonal or quasi orthogonal basis. Contrarily, the expression (16) gives a realistic appreciation with respect to the broad range of effective applicability of the SPMP approach. Regardless of the dictionary coherence, SPMP can be an effective low memory implementation of the OMP greedy strategy as long as the least squares problem, for the determination of the coefficients in the decomposition (2), is a well posed problem.

2.3 Numerical Example I

We illustrate here some features of the numerical convergence of the SPMP method in relation to the particular application to sparse signal decomposition.

The quality of the k-term approximation $\mathbf{f}^k$ of a signal $\mathbf{f}$ is assessed by the Signal to Noise Ratio (SNR), which is defined as

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{f}\|^2}{\|\mathbf{f} - \mathbf{f}^k\|^2}.$$

As an example we approximate, up to SNR = 35 dB, the $N = 1024$ samples of a music signal shown on Fig. 1. This SNR value produces a high quality approximation of the signal, indistinguishable from the original signal in the scale of Fig. 1.

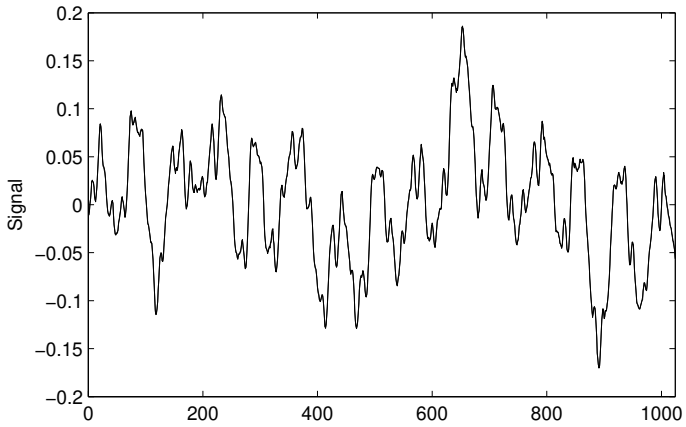

Figure 1: $N = 1024$ samples of a music signal and its approximation.

In the first instance we consider a uniformly random dictionary with redundancy four, which is certainly not an appropriate dictionary for representing music. Indeed, the SPMP method requires $k = 648$ atoms for approximating the 1024 samples up to SNR = 35 dB. The left graph in Fig. 2 shows the number of iterations spent in the orthogonal projection step vs the number of atoms involved in the corresponding step.

In order to obtain a sparse representation of the same signal we now change the random dictionary to the trigonometric one, $\mathcal{D}^{cs} = \mathcal{D}^c \cup \mathcal{D}^s$, with $\mathcal{D}^c$ and $\mathcal{D}^s$ as given below

$$\mathcal{D}^c = \{w^c(n)\, \cos(\frac{\pi(2i-1)(n-1)}{2M}),\, i = 1, \ldots, N\}_{n=1}^{M}. \tag{17}$$

and

$$\mathcal{D}^s = \{w^s(n)\, \sin(\frac{\pi(2i-1)n}{2M}),\, i = 1, \ldots, N\}_{n=1}^{M}, \tag{18}$$

where $w^c(n)$ and $w^s(n)$ are normalization factors. Taking $M = 2N$ the dictionary $\mathcal{D}^{cs}$ has the same redundancy as the previous one, but is suitable for representing music. The SPMP method uses now 137 atoms for approximating the signal in Fig. 1 up to SNR = 35 dB (the same number of atoms the OMP method needs). The right graph in Fig. 2 shows the iterations needed by the orthogonal projection step with dictionary $\mathcal{D}^{cs}$. It is clear that, up to the same numerical precision, the iterations to achieve the orthogonal projection depend on the dictionary.

Next we keep using the dictionary $\mathcal{D}^{cs}$ for tackling the following large dimension problem: The representation by non-orthogonal frequency components of the flute tone depicted in Fig.3,

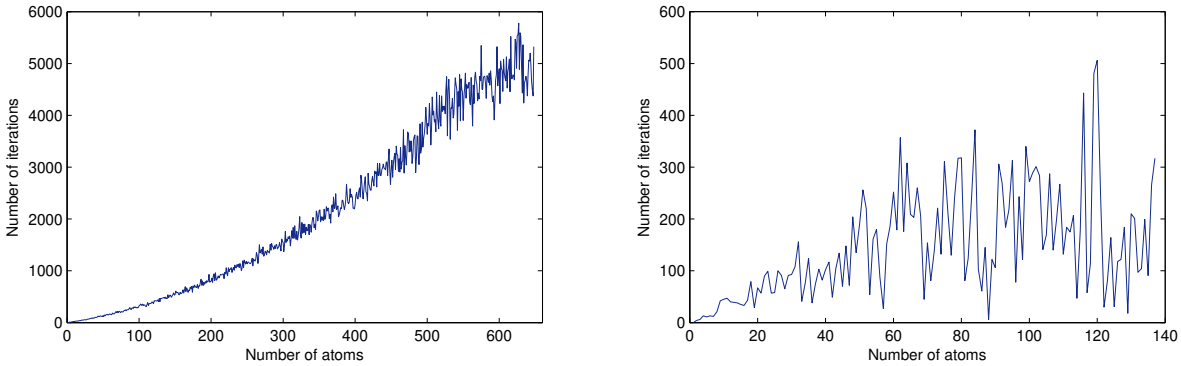

Figure 2: The left graph shows the number of iterations needed by SPMP for the approximation of the signal in Fig. 1 using a random dictionary. The right graph has the same description as the left graph but using the trigonometric dictionary $\mathcal{D}^{cs}$.

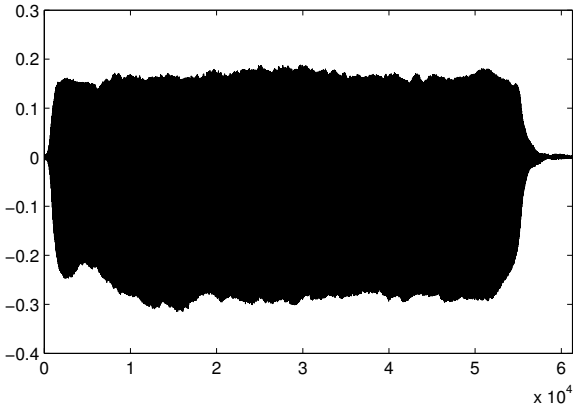

Figure 3: Baroque flute tone C#5. Sound clip Csharp5.baroque.wav available on `https://newt.phys.unsw.edu.au/music/flute/baroque/Csharp5.baroque.html`

which consists of $N = 61285$ samples. A particularity of dictionary $\mathcal{D}^{cs}$ is that, because by padding with zeros the inner products with its elements can be computed via the Fast Fourier Transform [25, 43], there is no need to store the dictionary as such (otherwise in this example it would be a matrix of dimension 61285×245140). The left graph of Fig. 4 shows the number of iterations vs the number of atoms in the signal approximation. The right graph is the histogram of the values in the left graph. The mean value of the number of iterations in the whole approximation is 44.

2.4 On the accuracy of self projections

The numerical accuracy of most used direct methods for calculating a projection is well studied [29–31, 33–38] and also the subject of recent research in particular contexts [39–41]. Contrarily, the numerical analysis of the SPMP algorithm has not yet been addressed. Therefore, this section discusses the accuracy of the self projection procedure, when implemented in finite precision arithmetic.

Since the self projection steps (a) - (c) in Sec. 2.1 are based on recursive calculation of inner products, we base the numerical analysis of the method on two basic results. As usual

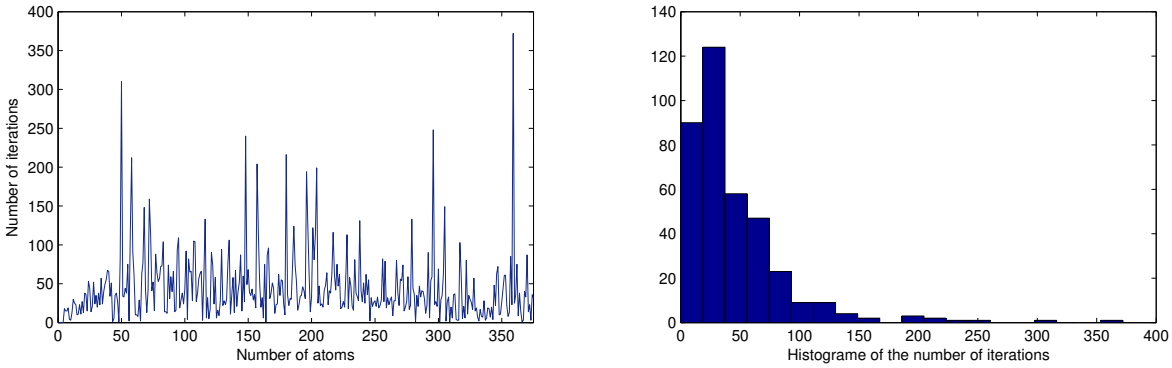

Figure 4: The left graph shows the number of iterations needed by SPMP for the approximation of the flute tone in Fig. 3. The right graph is the histogram of the values in the left graph.

the evaluation of an arithmetic operation is denoted as $\mathrm{fl}(\cdot)$ and the unit roundoff as u. Thus, for $\mathbf{f}_1 \in \mathbb{R}^N$ and $\mathbf{f}_2 \in \mathbb{R}^N$ the numerical error in the calculation of the inner product $\langle \mathbf{f}_1, \mathbf{f}_2 \rangle$ is bounded as ([30], p. 99)

$$|\mathrm{fl}(\langle \mathbf{f}_1, \mathbf{f}_2 \rangle) - \langle \mathbf{f}_1, \mathbf{f}_2 \rangle| \leq Nu\|\mathbf{f}_1\|\|\mathbf{f}_2\| + \mathrm{O}(u^2). \tag{19}$$

The computation of the saxpy operation $\alpha\mathbf{f}_1 + \mathbf{f}_2$, with α a number, is bounded as ([30], p. 100)

$$\|\mathrm{fl}(\alpha\mathbf{f}_1 + \mathbf{f}_2) - (\alpha\mathbf{f}_1 + \mathbf{f}_2)\| \leq u(2\|\alpha\mathbf{f}_1\| + \|\mathbf{f}_2\|) + \mathrm{O}(u^2). \tag{20}$$

Theorem 1. *An approximate bound for the error produced by implementing the projection step in finite precision arithmetics is give as*

$$\|\Delta\bar{\mathbf{r}}_T^{k,j}\| \lessapprox (N+3)ju\|\mathbf{r}^{k,0}\| + \mathrm{O}(u^2). \tag{21}$$

Proof. Denoting the computed quantities by $\bar{\mathbf{r}}^{k,j}$ and by $\bar{l}_j$ the indices selected with the computed quantities, using (20) we have

$$\bar{\mathbf{r}}^{k,j} = \bar{\mathbf{r}}^{k,j-1} - \mathrm{fl}(\langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle)\mathbf{d}_{\bar{l}_j} + \delta\bar{\mathbf{r}}^{k,j}, \tag{22}$$

with

$$\|\delta\bar{\mathbf{r}}^{k,j}\| \leq u\left(\|\bar{\mathbf{r}}^{k,j-1}\| + 2|\mathrm{fl}(\langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle)|\right) + \mathrm{O}(u^2).$$

Through straightforward manipulation we further have

$$\|\delta\bar{\mathbf{r}}^{k,j}\| \leq u\left(\|\bar{\mathbf{r}}^{k,j-1}\| + 2|\mathrm{fl}(\langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle) - \langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle| + 2|\langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle|\right) + \mathrm{O}(u^2)$$

so that, using (19), we finally obtain

$$\|\delta\bar{\mathbf{r}}^{k,j}\| \leq u\left(3\|\bar{\mathbf{r}}^{k,j-1}\| + 2Nu\|\bar{\mathbf{r}}^{k,j-1}\|\right) + \mathrm{O}(u^2) = 3u\|\bar{\mathbf{r}}^{k,j-1}\| + \mathrm{O}(u^2). \tag{23}$$

Moreover, (22) can be rewritten as

$$\bar{\mathbf{r}}^{k,j} = \bar{\mathbf{r}}^{k,j-1} - \langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle \mathbf{d}_{\bar{l}_j} + \Delta\bar{\mathbf{r}}^{k,j}, \tag{24}$$

where $\Delta\bar{\mathbf{r}}^{k,j} = -\mathrm{fl}(\langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle)\mathbf{d}_{\bar{l}_j} + \langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle \mathbf{d}_{\bar{l}_j} + \delta\bar{\mathbf{r}}^{k,j}$. Using now (23) and (19) we have the bound for the norm of $\Delta\bar{\mathbf{r}}^{k,j}$ in the form

$$\|\Delta\bar{\mathbf{r}}^{k,j}\| \leq Nu\|\bar{\mathbf{r}}^{k,j-1}\| + 3u\|\bar{\mathbf{r}}^{k,j-1}\| + \mathrm{O}(u^2) = u(N+3)\|\bar{\mathbf{r}}^{k,j-1}\| + \mathrm{O}(u^2). \tag{25}$$

Thus, due to rounding errors instead of the theoretical result $\|\mathbf{r}^{k,j}\| \leq \|\mathbf{r}^{k,j-1}\|$ we only have

$$\|\bar{\mathbf{r}}^{k,j}\| \leq (1+(N+3)u)\|\bar{\mathbf{r}}^{k,j-1}\| + \mathrm{O}(u^2) \leq (1+(N+3)u)^j\|\bar{\mathbf{r}}^{k,0}\| + \mathrm{O}(u^2).$$

This inequality gives rise to the recurrence for bounding the total error in the calculation of $\mathbf{r}^{k,j}$. In terms of the matrices $\bar{\mathbf{T}}_i = (\mathbf{I} - \mathbf{d}_{\bar{l}_i}\mathbf{d}_{\bar{l}_i}^\top)$, $i = 1, \ldots, j$, where $\mathbf{I} \in \mathbb{R}^{N\times N}$ is the identity matrix, equation (24) can be expressed in the form

$$\bar{\mathbf{r}}^{k,j} = \bar{\mathbf{T}}_j\bar{\mathbf{T}}_{j-1}\cdots\bar{\mathbf{T}}_1\mathbf{r}^{k,0} + \Delta\bar{\mathbf{r}}_T^{k,j},$$

where $\Delta\bar{\mathbf{r}}_T^{k,j} = \sum_{i=1}^{j}\bar{\mathbf{T}}_j\ldots\bar{\mathbf{T}}_{i+1}\Delta\bar{\mathbf{r}}^{k,i}$ (with the notation $\bar{\mathbf{T}}_j\bar{\mathbf{T}}_{j+1} = \mathbf{I}$). Since $\|\bar{\mathbf{T}}_i\| = 1$ for all $i = 1, \ldots, j$ and $\Delta\bar{\mathbf{r}}^{k,i}$ is bounded as in (25), it follows that $\Delta\bar{\mathbf{r}}_T^{k,j}$ is bounded as

$$\|\Delta\bar{\mathbf{r}}_T^{k,j}\| \;\leq\; \sum_{i=1}^{j}\|\Delta\bar{\mathbf{r}}^{k,i}\| \leq u(N+3)\sum_{i=1}^{j}(1+(N+3)u)^i\|\mathbf{r}^{k,0}\| + \mathrm{O}(u^2). \tag{26}$$

Restricting considerations to $Nu \ll 1$ we have the approximate bound

$$\|\Delta\bar{\mathbf{r}}_T^{k,j}\| \lessapprox (N+3)ju\|\mathbf{r}^{k,0}\| + \mathrm{O}(u^2). \tag{27}$$

□

Even if, as discussed in Sec. 2.2, in the limit $j \to \infty$ the convergence $\mathbf{r}^{k,j} \to \mathbf{f} - \hat{\mathbf{P}}_{\mathbb{S}_k}\mathbf{f}$ is theoretically guaranteed, the size of $\Delta\bar{\mathbf{r}}_T^{k,j}$ gives a limit for the maximum number of recursive operations. Beyond that limit the calculations in the self projection algorithm are dominated by rounding errors. However, in situations of practical interest the numerical convergence is fast enough for the algorithm to operate within the boundary of reliability established in (27).

3 Hierarchized Block Wise SPMP

The Hierarchized Block Wise (HBW) version of pursuit strategies is an implementation of those techniques dedicated to approximating by partitioning. The method approximates each element of a signal partition independently of each other, but links the approximations by a global constraint on sparsity [42, 43]. The strategy proceeds simply by ranking the partition units for their sequential stepwise approximation. This section discusses the HWB version of the OMP approach (HBW-OMP) [42, 43] but implemented via the SPMP method (HBW-SPMP).

Let's suppose that a given signal $\mathbf{f}$ is split into Q disjoint 'blocks' $\mathbf{f}_q$, $q = 1, \ldots, Q$, where each $\mathbf{f}_q$ is an element of $\mathbb{R}^{N_b}$, with $N_b = N/Q$. Denoting by $\hat{\mathrm{J}}$ the concatenation operator, the signal $\mathbf{f} \in \mathbb{R}^N$ is 'assembled' from the blocks as $\mathbf{f} = \hat{\mathrm{J}}_{q=1}^{Q}\mathbf{f}_q$. This operation implies that the first N_1 components of the vector $\mathbf{f}$ are given by the vector $\mathbf{f}_1$, the next N_2 components by the vector $\mathbf{f}_2$ and so on. The HBW version of SPMP for approximating the signal's partition using K atoms in total is implemented by the following steps.

1) For $q = 1, \ldots, Q$ set $\mathbf{r}_q^0 = \mathbf{f}_q$, $\mathbf{f}_q^0 = 0$, $\mathcal{L}_0^q = \{\emptyset\}$ and $k_q = 1$. Initialize the algorithm by selecting the 'potential' first atom for the atomic decomposition of every block q, according to the MP criterion:

$$\ell_{k_q}^q = \arg\max_{n=1,\ldots,M}\left|\left\langle \mathbf{d}_n, \mathbf{r}_q^{k_q-1}\right\rangle\right|, \; q = 1, \ldots, Q.$$

2) Select the block $q^\star$ such that

$$q^\star = \arg\max_{q=1,\ldots,Q} \left| \left\langle \mathbf{d}_{\ell_{k_q}^q}, \mathbf{r}_q^{k_q-1} \right\rangle \right|.$$

Update the set $\mathcal{L}_{k_{q^\star}}^{q^\star} = \mathcal{L}_{k_{q^\star}-1}^{q^\star} \cup \{\ell_{k_{q^\star}}^{q^\star}\}$ and the atomic decomposition of the block $q^\star$ by incorporating the atom $\mathbf{d}_{\ell_{k_{q^\star}}^{q^\star}}$ i.e., use $c^{q^\star}(k_{q^\star}) = \left\langle \mathbf{d}_{\ell_{k_{q^\star}}^{q^\star}}, \mathbf{r}_{q^\star}^{k_{q^\star}-1} \right\rangle$ to compute

$$\begin{aligned} \mathbf{f}_{q^\star}^{k_{q^\star}} &= \mathbf{f}_{q^\star}^{k_{q^\star}-1} + c^{q^\star}(k_{q^\star}) \mathbf{d}_{\ell_{k_{q^\star}}^{q^\star}}, \\ \mathbf{r}_{q^\star}^{k_{q^\star}} &= \mathbf{f}_{q^\star} - \mathbf{f}_{q^\star}^{k_{q^\star}}. \end{aligned}$$

If $k_{q^\star} > 1$ set $\mathbf{r}_{q^\star}^{k,0} = \mathbf{r}_{q^\star}^{k}$ and starting from $j = 1$ realize the projection as indicated below.

(a) Choose, out of the set $\mathcal{L}_{k_{q^\star}}^{q^\star} = \{\ell_i^{q^\star}\}_{i=1}^{k_{q^\star}}$, the index l_j such that

$$l_j = \arg\max_{i=1,\ldots,k_{q^\star}} \left| \left\langle \mathbf{d}_{\ell_i^{q^\star}}, \mathbf{r}^{k_{q^\star},j-1} \right\rangle \right|.$$

If $\left| \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k_{q^\star},j-1} \right\rangle \right| < \epsilon$ jump to 3). Otherwise proceed with steps b) and c).

(b) Use $\left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k_{q^\star},j-1} \right\rangle$ to update the coefficient $c^{q^\star}(l)$, the approximation $\mathbf{f}_{q^\star}^{k_{q^\star}}$, and the residual as

$$\begin{aligned} c^{q^\star}(l_j) &\leftarrow c^{q^\star}(l_j) + \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k_{q^\star},j-1} \right\rangle, \\ \mathbf{f}^{k_{q^\star}} &\leftarrow \mathbf{f}^{k_{q^\star}} + \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k_{q^\star},j-1} \right\rangle \mathbf{d}_{l_j}, \\ \mathbf{r}^{k_{q^\star},j} &= \mathbf{r}^{k_{q^\star},j-1} - \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k_{q^\star},j-1} \right\rangle \mathbf{d}_{l_j}. \end{aligned}$$

(c) Increment $j \leftarrow j + 1$ and repeat steps (a) $\rightarrow$ (c) until the stopping criterion is met.

3) Check if for the given number K the stopping condition $\sum_{q=1}^{Q} k_q = K$ has been met. Otherwise:

- Increase $k_{q^\star} \leftarrow k_{q^\star} + 1$.
- Select a new potential atom for the atomic decomposition of block $q^\star$

$$\ell_{k_{q^\star}}^{q^\star} = \arg\max_{n=1,\ldots,M} \left| \left\langle \mathbf{d}_n, \mathbf{r}_{q^\star}^{k_{q^\star}-1} \right\rangle \right|.$$

- Repeat 2) and 3).

3.1 Numerical Example II

We construct here the atomic decomposition of the Pop Piano and Classic Guitar clips shown in Fig. 5. Both clips consists of $N = 262144$ samples at 44100Hz each (5.94 secs length). For the approximation we use the trigonometric dictionary $\mathcal{D}^{cs}$ introduced in Sec. 2.3.

The global sparsity of the signal approximation is measured by the Sparsity Ratio (SR) which is defined as $\text{SR} = \dfrac{N}{K}$, where K is the total number of coefficients in the signal representation. Hence, the larger the value of SR is the smaller the number of frequency components needed for the approximation.

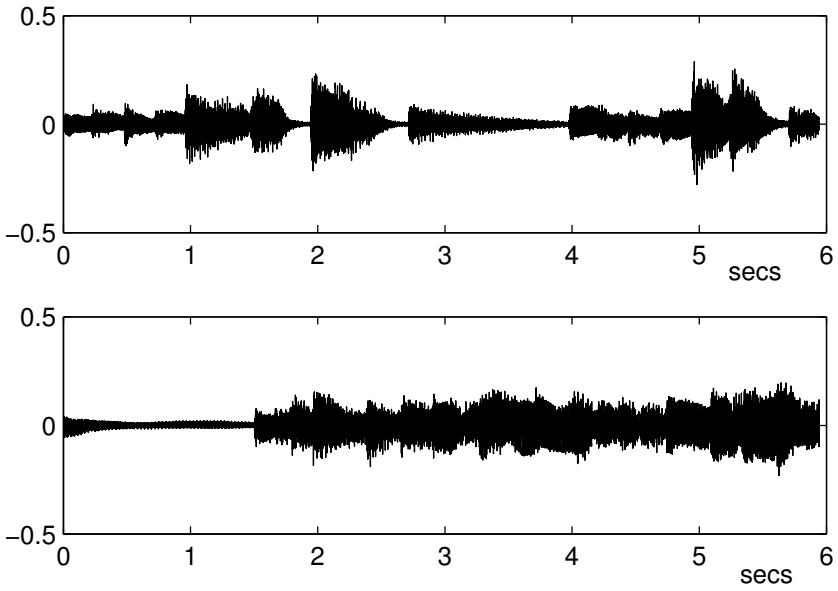

Figure 5: Pop Piano (top graph) and Classic Guitar music signals. Both clips consist of $N = 262144$ samples at 44100Hz each (5.94 secs length).

The sparsity results of the clips in Fig. 5 are shown in Fig. 6, for the MP, HBW-MP, SPMP, HBW-SPMP approaches and partitions of unit size N_b equal to 1024, 2048, 4096, and 8192 samples. For larger values of N_b the sparsity does not improve significantly. The quality of the approximation is fixed to yield a SNR of 35dB. As observed in Fig. 6 for the two clips in Fig. 5 the gain in sparsity achieved by implementing the SPMP approach in the HBW manner is significant.

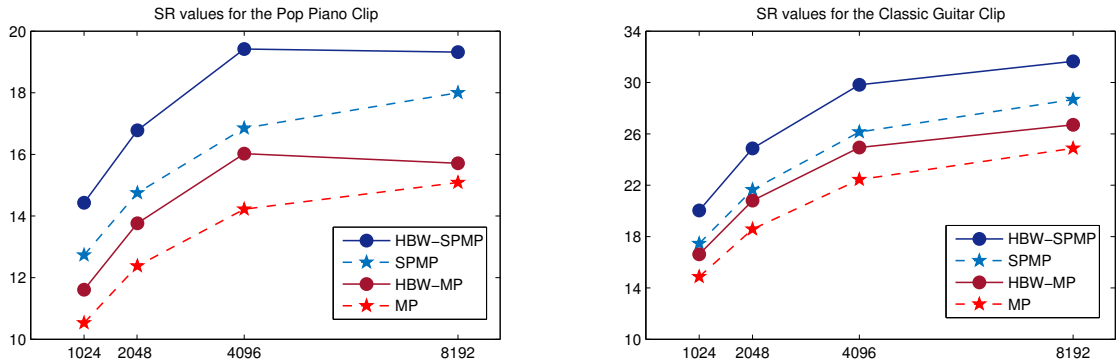

Figure 6: SR vs partition unity size $N_b = 1024, 2048, 4096$ and 8192 samples for the music clips of Fig. 5. The graph on left corresponds to the Pop Piano and the other to the Classic Guitar.

Note: The MATLAB function HBW-SPMP dedicated to reproducing the above example with the trigonometric dictionary $\mathcal{D}^{cs}$, via the FFT, has been have been made available on [44]. The MATLAB and C++ codes for implementing SPMP with general dictionaries, as well as the corresponding SPMP2D versions for separable dictionaries are available on [45]. The MATLAB and C++ codes for SPMP3D can be found on [46].

4 Conclusions

The convergence rate of the SPMP algorithm, which implements the OMP greedy strategy by means of the MP one, was derived. The orthogonal projection step, intrinsic to the OMP method, is realized within the SPMP framework by subtraction from the residual error its approximation using the MP algorithm with a dictionary consisting only of the already selected atoms, up to the particular step. Thus, the memory requirements are kept within the same scale as for MP. The bound for the self projection convergence rate (c.f. (16)) clearly highlights the broad range of cases for which the OMP greedy strategy can be implemented through the SPMP method. The cases for which the convergence could become very slow fall within the class of ill posed problems.

The analysis of the accuracy of the projection step, when implemented in finite precision arithmetics, produced a meaningful upper bound relating the number of iterations with the dimension of system and the unit roundoff. This worst-case behavior bound confirms that the SPMP method is suitable to be applied to solve well posed problems for which the convergence is fast. Otherwise, as the number of iterations increases the accuracy of the approach would be dominated by roundoff errors. Nevertheless, a number of applications to real world signals [22–25] have already confirmed that the approach is of assistance for practical implementations of the OMP greedy strategy in situations where, due to memory requirements, direct linear algebra techniques cannot be applied.

The HBW extension of a pursuit strategy for approximating a signal partition was considered in relation to the SPMP implementation for reduction in memory requirements. The suitability of the technique was highlighted by numerical tests which, due to memory limitations, could not have been realized in a standard computer by other implementations of OMP.

Acknowledgements

We are grateful to anonymous Reviewers for the careful reading of the paper and their helpful comments and constructive remarks.

References

[1] S. S. Chen, D. L. Donoho, and M. A Saunders, "Atomic Decomposition by Basis Pursuit", *SIAM Journal on Scientific Computing*, **20**, 33–61 (1998).

[2] D. L. Donoho and J. Tanner "Sparse nonnegative solution of underdetermined linear equations by linear programming", *Proceedings of the National Academy of Sciences of the United States of America,* **102,** 9446–9451 (2005).

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

[4] S. G. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries", *IEEE Trans. Signal Process.,* **41**, 3397–3415 (1993).

[5] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers* ,**1**, 40–44 (1993).

[6] B. K. Natarajan, "Sparse Approximate Solutions to Linear Systems", *SIAM Journal on Computing*, **24**, 227–234 (1995).

[7] R.A. DeVore, and V.N. Temlyakov,"Some remarks on greedy algorithms", *Advances in Computational Mathematics*, **5**, 173–187(1996).

[8] V. N. Temlyakov, "Greedy algorithms and m-term approximation with regard to redundant dictionaries, *J. Approx. Theory*, vol. 98, pp.117145(1999).

[9] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach", *IEEE Signal Process. Letters*, **9**, 137–140 (2002).

[10] M. Andrle, L. Rebollo-Neira, and E. Sagianos, "Backward-optimized orthogonal matching pursuit approach", *IEEE Signal Proc. Let.*,**11**,705–708 (2004).

[11] M. Andrle and L. Rebollo-Neira, "A swapping-based refinement of orthogonal matching pursuit strategies", *Signal Processing*, **86**, 480–495 (2006).

[12] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation", *IEEE Transactions on Information Theory*, **50**, 2231–2242 (2004).

[13] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuits in quasi-incoherent dictionaries", *IEEE Transactions on Information Theory*, 255–261 (2006).

[14] D. L. Donoho , Y. Tsaig , I. Drori , and J. Starck, "Stagewise Orthogonal Matching Pursuit", *IEEE Transactions on Information Theory*, **58**, 1094–1121 (2006).

[15] T. Blumensath, M. E. Davies, "Gradient Pursuits", *IEEE Transactions on Signal Processing*, **56**, 2370 – 2382 (2008).

[16] D. Needell and J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples", *Applied and Computational Harmonic Analysis*, **26**, 301–321 (2009).

[17] M. Luo, F. Sun, H. Liu, Z Lid, "A novel TS fuzzy systems identification with block structured sparse representation", *Journal of the Franklin Institute* **351**, 3508–3523 (2014).

[18] C. You C., D. Robinson, R. Vidal, "Scalable sparse subspace clustering by orthogonal matching pursuit", Proceedings of Conference on Computer Vision and Pattern Recognition, 3918–3927 (2016).

[19] J. You, Y. Liu, J. Chen, F. Ding, "Iterative identification for multiple-input systems with time-delays based on greedy pursuit and auxiliary model", *Journal of the Franklin Institute*, **356,** 5819–5833 (2019).

[20] J. H. Friedman and W. Stuetzle, "Projection Pursuit Regression", *Journal of the American Statistical Association*, **76**, 817– 823 (1981).

[21] L. K. Jones, "On a conjecture of Huber concerning the convergence of Projection Pursuit Regression", *Ann. Statist.* **15**, 880–882 (1987).

[22] L. Rebollo-Neira and J. Bowley, "Sparse representation of astronomical images", *Journal of The Optical Society of America A,* **30**, 758–768 (2013).

[23] L. Rebollo-Neira, "Effective sparse representation of X-Ray medical image", *International Journal for Numerical Methods in Biomedical Engineering,* (2017) e2886. `https://doi.org/10.1002/cnm.2886`.

[24] L. Rebollo-Neira and D. Whitehouse, "Sparse representation of 3D images for piecewise dimensionality reduction with high quality reconstruction", *Array*, **1**, doi = 10.1016/j.array.2019.100001 (2019).

[25] L. Rebollo-Neira and G. Aggarwal "A dedicated greedy pursuit algorithm for sparse spectral representation of music sound", *Journal of the Acoustic Society of America*, **140**, 2933 (2016).

[26] L. Rebollo-Neira, "On non-orthogonal signal representation", in New Topics in Mathematical Physics Research, Nova Science Publisher (2006).

[27] P. G. Casazza and G. Kutyniok and F. Philipp, "Introduction to Finite Frame Theory", in Finite Frames: Theory and Applications, 1–54, Springer, 2012.

[28] R. G. Bartle and D, R. Sherbert, Introduction to Real Analysis, John Wiley & Sons, (1999). 4th Edition, 2012.

[29] A. Björck, "Numerical methods for least square problems," SIAM, 1996.

[30] G. H. Golub and C. F. Van Loan, Matrix Computations, The John Hopkins University Press, Baltimore, 4th Edition, 2012.

[31] N. J. Higham, "Accuracy and stability of numerical algoriths," SIAM, 2002.

[32] R. A. Horn and C. R. Johnson, Topics in Matrix Analysis, Cambridge University Press, (1991).

[33] A. Björck, "Solving linear least squares problems by Gram-Schmidt orthogonalization", *BIT Numerical Mathematics*, **7**, 1–21 (1967).

[34] A. Björck, "Numerics of Gram-Schmidt orthogonalization," *Linear Algebra and its Applications*, **197**, 297–316 (1994).

[35] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Oxford: Clarendon Press 1965.

[36] L. Giraud, J. Langou, M. Rozložník, and J. van den Eshof, "Rounding error analysis of the classical Gram-Schmidt orthogonalization process," *Numerische Mathematik*, **101**, 87–100 (2005).

[37] L. Giraud, J. Langou and M. Rozložník, "The loss of orthogonality in the Gram-Schmidt orthogonalization process," *Computers & Mathematics with Applications*, **50**, 1069–1075 (2005).

[38] G. W. Stewart, "Error Analysis of the Quasi-Gram-Schmidt Algorithm" *SIAM Journal on Matrix Analysis and Applications*, **27**, 493–506 (2005).

[39] E. S. Van Vleck, "On the Error in the Product QR Decomposition", *SIAM Journal on Matrix Analysis and Applications.*, **31**, 1775–1791 (2010).

[40] G. W. Stewart, "On the Numerical Analysis of Oblique Projectors", *SIAM Journal on Matrix Analysis and Applications*, **32**, 309–348 (2011).

[41] S. J. Leon, A. Björck and W. Gander, "Gram-Schmidt orthogonalization: 100 years and more" Authors", *Numerical Linear Algebra*, *Numer. Linear Algebra Appl.*, **20**, 492–532 (2013) doi:10.1002/nla.1839

[42] L. Rebollo-Neira, R. Matiol, and S. Bibi, "Hierarchized block wise image approximation by greedy pursuit strategies," *IEEE Signal Process. Letters*, **20**, 1175–1178 (2013).

[43] L. Rebollo-Neira, "Cooperative greedy pursuit strategies for sparse signal representation by partitioning", *Signal Processing,* **125**, 365–375 (2016).

[44] `http://www.nonlinear-approx.info/examples/node04.html` (Last access April 2020).

[45] `http://www.nonlinear-approx.info/examples/node1.html` (Last access April 2020).

[46] `http://www.nonlinear-approx.info/examples/node09.html` (Last access April 2020).