INSTITUTE OF MATHEMATICS

THE CZECH ACADEMY OF SCIENCES

Analysis of a low memory implementation of the Orthogonal Matching Pursuit greedy strategy

Laura Rebollo-Neira
Miroslav Rozložník
Pradip Sasmal

Preprint No. 78-2017

PRAHA 2017

Analysis of a low memory implementation of the Orthogonal Matching Pursuit greedy strategy

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

November 12, 2017

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 provides 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. Furthermore, the low memory requirement of the method suits it for massive parallelization, via Graphics Processing Unit, to tackle systems which can be broken into a large number of subsystems of much smaller dimension.

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

1 Introduction

The process by which a signal is transformed, in order to significantly reduce its dimensionality, is refereed to as sparse representation of the signal. For the class of signals known as compressible, such as images and audio signals, this process can be realized without much loss of information content. The degree of the achieved sparsity depends on the suitability of the transformation for representing the particular signal. Traditional methods implement the transformation using fast orthogonal transforms. However, much higher levels of sparsity are attained, in many cases, if the transformation is carried out using a large redundant set called a 'dictionary'. The gain comes at expenses of increment in complexity. However, advances

in computational facilities, including multiprocessors for personal computers, have encouraged the developments of techniques for signal representation using dictionaries. For the most part these techniques comprise strategies based on 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 [4–8, 10–19]. 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 though [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. Consequently, in addition to failing to minimize the norm of the approximation error at each step, it may select linearly dependent elements. As illustrated in [22], this feature significantly compromises sparsity in some cases.

A refinement to MP, which does yield an orthogonal projection at each iteration, is refereed to as Orthogonal Matching Pursuit (OMP) [5]. This technique is broadly used for signal processing and the analysis of its suitability in the context of some applications is a topic of resent research [23–26]. If implemented by direct methods, the OMP approach is very effective up to some dimensionality. When processing large signals, however, the storage requirement of the implementations by direct methods frequently exceed the memory capacity of a personal computer used for research purposes. Some techniques addressing this matter are known as Gradient Pursuits [16]. They are based on approximations of the conjugate gradient method. While these techniques produce a satisfactory approximation to the OMP criterion in many practical situations, they are not guaranteed to choose linearly independent atoms and may not be satisfactory in cases where the basic MP method performs badly. An alternative implementation of OMP, which requires much less memory than standard implementations is considered in [22]. The approach is termed Self Projected Matching Pursuit (SPMP), because it produces the orthogonal projection of the signal, at each iteration, by applying MP using a dictionary consisting only of the already selected elements. The SPMP implementation of OMP has been shown to be effective for sparse representation of astronomical images [22] and X-Ray medical images [27]. The convenient feature in that context is that SPMP fully exploits the separability of dictionaries. In the case of one dimensional signals such as melodic music, which are well approximated by trigonometric dictionaries, the SPMP approach allows to implement OMP without having to save the dictionary and performs the calculation via the Fast Fourier Transform [28]. In addition to its suitability for solving the least squares problem of large systems, or systems which otherwise would require large memory availability, the SPMP algorithm has been shown appropriate for implementations in Graphics Processing Unit (GPU) to tackle systems which can be broken into a large number of subsystems of smaller dimension [29]. Nevertheless, until now the method had not been analyzed. Hence, this work focusses on the convergence and numerical analysis of the SPMP approach.

The paper is organized as follows: Sec. 2 recalls the SPMP algorithm, proves the geometrical 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)

Before reviewing the general SPMP technique let's define some basic notation: $\mathbb{R}$ 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 called a *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

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

with the atom $\mathbf{d}_{\ell_{k+1}}$ 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

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

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 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 [46], i.e., $\lim_{k\to\infty} \|\mathbf{r}^k\|^2 = b$. We prove next that $b = 0$. Since

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

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

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

Thus, $\lim_{k\to\infty} |\left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle| = 0$, which using (4) implies $\lim_{k\to\infty} |\left\langle \mathbf{d}_n, \mathbf{r}^k \right\rangle| = 0$, $n = 1, \ldots, M$. 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. □

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 [31–33]. 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 produce an alternative iterative implementation of the OMP approach. Given a signal and a dictionary it proceeds as follows [22]: Set $\mathcal{L}_0 = \{\emptyset\}$, $\mathbf{f}^0 = 0$ and $\mathbf{r}^0 = \mathbf{f}$. Starting from $k = 1$, at each iteration k implement the steps below.

i) Apply the MP criterion for selecting one atom from $\mathcal{D}$, i.e., select ℓ_k such that

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

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

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

If $\left|\left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1} \right\rangle\right| < \epsilon$ stop. Otherwise continue with the next steps.

(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.

Continue with steps i) - ii) until, for a required tolerance error ρ, the condition $\|\mathbf{r}^k\| < \rho$ is reached.

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. Let's recall that such a 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

$$\mathbf{r}^{k,j} = \mathbf{r}^{k,j-1} - \left\langle \mathbf{d}_{l_j}, \mathbf{r}^{k,j-1} \right\rangle \mathbf{d}_{l_j}, \tag{11}$$

where

$$l_j = \underset{i=1,\ldots,k}{\arg\max} | \left\langle \mathbf{d}_{\ell_i}, \mathbf{r}^{k,j-1} \right\rangle |. \tag{12}$$

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 (11) 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

$$\|\hat{\mathbf{P}}_{\mathbb{S}_k} \mathbf{r}^{k,j-1}\|^2 = \|\hat{\mathbf{P}}_{\mathbb{S}_k} \mathbf{r}^{k,j-1}\|^2 - | \left\langle \mathbf{r}^{k,j-1}, \mathbf{d}_{l_j} \right\rangle |^2. \tag{13}$$

By definition of the index l_j (cf.(12)), 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

$$\|\hat{\mathbf{P}}_{\mathbb{S}_k} \mathbf{r}^{k,j}\|^2 \leq \left(1 - \frac{\lambda_k^k}{k}\right) \|\hat{\mathbf{P}}_{\mathbb{S}_k} \mathbf{r}^{k,j-1}\|^2, \tag{14}$$

and applying the inequality back j-times

$$\|\hat{\mathbf{P}}_{\mathbb{S}_k} \mathbf{r}^{k,j}\|^2 \leq \left(1 - \frac{\lambda_k^k}{k}\right)^j \|\hat{\mathbf{P}}_{\mathbb{S}_k} \mathbf{r}^{k,0}\|^2 \leq \left(1 - \frac{\lambda_k^k}{k}\right)^j \|\mathbf{r}^{k,0}\|^2. \tag{15}$$

The above bound indicates the geometric convergence 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 ([34], 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 geometric convergence of MP in terms of the dictionary's coherence [13] is derived in [14] 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 (15) 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.

The numerical accuracy of most used direct methods for calculating a projection is well studied [31–33, 35, 38–40] and also the subject of resent research in particular contexts [41–43]. Contrarily, the numerical analysis of the SPMP algorithm has not yet been addressed. Therefore, the next section discusses the accuracy of the self projection procedure, when implemented in finite precision arithmetic.

2.3 On the accuracy of self projections

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 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 ([32], p. 99)

$$|\mathrm{fl}(\langle \mathbf{f}_1, \mathbf{f}_2 \rangle) - \langle \mathbf{f}_1, \mathbf{f}_2 \rangle| \leq Nu\|\mathbf{f}_1\|\|\mathbf{f}_2\| + \mathrm{O}(u^2). \tag{16}$$

The computation of the saxpy operation $\alpha\mathbf{f}_1 + \mathbf{f}_2$, with α a number, is bounded as ([32], p. 100)

$$\|\mathrm{fl}(\alpha\mathbf{f}_1 + \mathbf{f}_2) - (\alpha\mathbf{f}_1 + \mathbf{f}_2)\| \leq u(2\|\alpha\mathbf{f}_1\| + \|\mathbf{f}_2\|) + \mathrm{O}(u^2). \tag{17}$$

Then, denoting the computed quantities by $\bar{\mathbf{r}}^{k,j}$ and by $\bar{l}_j$ the indices selected with the computed quantities, using (17) we have

$$\bar{\mathbf{r}}^{k,j} = \bar{\mathbf{r}}^{k,j-1} - \mathrm{fl}(\langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle)\mathbf{d}_{\bar{l}_j} + \delta\bar{\mathbf{r}}^{k,j}, \tag{18}$$

where, since $\|\mathbf{d}_{\bar{l}_j}\| = 1$,

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

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 (16), we finally obtain

$$\|\delta\bar{\mathbf{r}}^{k,j}\| \;\leq\; u\left(3\|\bar{\mathbf{r}}^{k,j-1}\| + 2Nu\|\bar{\mathbf{r}}^{k,j-1}\|\right) + \mathrm{O}(u^2) = 3u\|\bar{\mathbf{r}}^{k,j-1}\| + \mathrm{O}(u^2). \tag{19}$$

Moreover, (18) can be rewritten as

$$\bar{\mathbf{r}}^{k,j} = \bar{\mathbf{r}}^{k,j-1} - \langle \bar{\mathbf{r}}^{k,j-1}, \mathbf{d}_{\bar{l}_j} \rangle \mathbf{d}_{\bar{l}_j} + \Delta\bar{\mathbf{r}}^{k,j}, \tag{20}$$

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 (19) and (16) we have the bound for the norm of $\Delta\bar{\mathbf{r}}^{k,j}$ in the form

$$\|\Delta\bar{\mathbf{r}}^{k,j}\| \leq Nu\|\bar{\mathbf{r}}^{k,j-1}\| + 3u\|\bar{\mathbf{r}}^{k,j-1}\| + \mathrm{O}(u^2) = u(N+3)\|\bar{\mathbf{r}}^{k,j-1}\| + \mathrm{O}(u^2). \tag{21}$$

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

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

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 $\bar{\mathbf{T}}_j = (\mathbf{I} - \mathbf{d}_{\bar{l}_j}\mathbf{d}_{\bar{l}_j}^\top)$, where $\mathbf{I} \in \mathbb{R}^{N\times N}$ is the identity matrix, equation (20) 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 $\Delta\bar{\mathbf{r}}_T^{k,j} = \sum_{i=1}^{j}\Delta\bar{\mathbf{r}}^{k,i}$. Since all the $\Delta\bar{\mathbf{r}}^{k,i}$, $i = 1,\ldots,j$ are bounded as in (21) it follows that $\Delta\bar{\mathbf{r}}_T^{k,j}$ is bounded as

$$\|\Delta\bar{\mathbf{r}}_T^{k,j}\| \;\leq\; \sum_{i=1}^{j}\|\Delta\bar{\mathbf{r}}^{k,i}\| \leq u(N+3)\sum_{i=1}^{j}(1 + (N+3)u)^i\|\mathbf{r}^{k,0}\| + \mathrm{O}(u^2). \tag{22}$$

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

$$\|\Delta\bar{\mathbf{r}}_T^{k,j}\| \lessapprox (N+3)ju\|\mathbf{r}^{k,0}\| + \mathrm{O}(u^2). \tag{23}$$

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 (23).

2.4 Numerical Example I

We illustrate here the suitability of the SPMP approach for tackling the following large dimension problem: The representation by non-orthogonal frequency components of a flute tone, consisting of $N = 61285$ points, depicted in Fig.1.

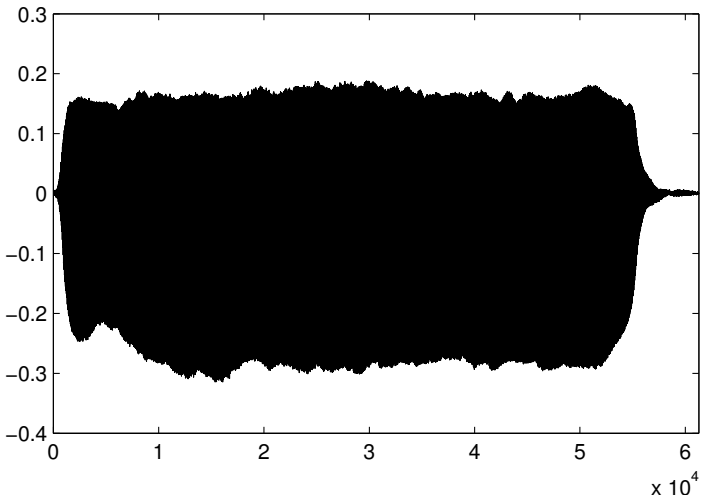

Figure 1: Baroque flute tone C#5. Sound clip Csharp5.baroque.wav available on `https://newt.phys.unsw.edu.au/music/flute/baroque/Csharp5.baroque.html`

As oppose to the melodic music clips (which will be considered in Sec.3.1) the flute tone is characterized by being stationary. This implies that its frequency components do not vary along the clip and this is approximated with fewer components without partitioning the signal. For the approximation we use the trigonometric dictionaries $\mathcal{D}^{cs} = \mathcal{D}^c \cup \mathcal{D}^s$, with $\mathcal{D}^c$ and $\mathcal{D}^s$ as given below

$$\mathcal{D}^c = \{\frac{1}{w^c(n)}\,\cos(\frac{\pi(2i-1)(n-1)}{2M}),\, i = 1,\ldots,N\}_{n=1}^{M}. \tag{24}$$

and

$$\mathcal{D}^s = \{\frac{1}{w^s(n)}\,\sin(\frac{\pi(2i-1)n}{2M}),\, i = 1,\ldots,N\}_{n=1}^{M}, \tag{25}$$

with

$$w^c(n) = \begin{cases} \sqrt{N} & \text{if} \quad n = 1, \\ \sqrt{\frac{N}{2} + \frac{\sin(\frac{\pi(n-1)}{M})\sin(\frac{2\pi(n-1)N}{M})}{2(1-\cos(\frac{2\pi(n-1)}{M}))}} & \text{if} \quad n \neq 1. \end{cases}$$

and

$$w^s(n) = \begin{cases} \sqrt{N} & \text{if} \quad n = 1, \\ \sqrt{\frac{N}{2} - \frac{\sin(\frac{\pi n}{M})\sin(\frac{2\pi nN}{M})}{2(1-\cos(\frac{2\pi n}{M}))}} & \text{if} \quad n \neq 1. \end{cases}$$

The trigonometric dictionary $\mathcal{D}^{cs}$ with redundancy four has been shown to produce high quality approximation of music, involving much less terms than what are needed when using an orthonormal trigonometric basis [28, 45].

A particularity of dictionary $\mathcal{D}^{cs}$ is that, because by padding with zeros the inner products with its elements can be computed via Fast Fourier Transform, the calculations are fast and there is no need to store the dictionary as such (otherwise in this example it would be a matrix of dimension 61285×245140).

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

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

In this numerical example the approximation error is set to yield a SNR of 35dB. This value produces a high quality approximation of the signal. It corresponds to a residual error with very small mean ($\text{O}(10^{-6})$) and very small variance ($\text{O}(10^{-6})$). For an approximation of this quality the SPMP algorithm uses $K = 373$ dictionary atoms while the basic MP uses $K = 450$ atoms.

As already mentioned, music signals which are not stationary are to be approximated by partitioning. The next section discusses an application of the SPMP methodology dedicated to that type of processing.

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 technique approximates each element of a signal partition independently of each other, but links the approximations by a global constraint on sparsity [44, 45]. The approach simply ranks the partition units for their sequential stepwise approximation.

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 $\hat{\text{J}}$ the concatenation operator, the signal $\mathbf{f} \in \mathbb{R}^N$ is 'assembled' from the blocks as $\mathbf{f} = \hat{\text{J}}_{q=1}^{Q} \mathbf{f}_q$. 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 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:

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

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

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

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 $k_{q^\star} > 1$ starting from $j = 1$, $\mathbf{r}_{q^\star}^{k,0} = \mathbf{r}_{q^\star}^{k}$ apply the self projection step of (a) $\rightarrow$ (c) 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 such that

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

If $\left|\left\langle \mathbf{d}_l, \mathbf{r}^{k_{q^\star},j-1} \right\rangle\right| < \epsilon$ stop. Otherwise proceed with the next step.

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

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

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

3) Check if, for the given numbers K the 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 $q^\star$

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

- Repeat steps 2) and 3).

Remark 2: Notice that the memory requirements of the SPMP approach and its HBW version are equivalent to that of the MP approach and its HBW version, respectively. This implies a significant saving in memory with respect of the standard implementations of OMP and the HBW version of it. Certainly, for the implementation of the orthogonal projection step, through Gram Schmidt orthogonalization for instance, the OMP approach would require to construct k_q vectors, each of dimension N_b. Thus, its HBW version would need to save these vectors for each of the Q blocks, only for the calculation of the projection. On the contrary, the HBW-SPMP implementation saves directly the coefficients of the approximation. Thereby the storage demands for the projection step are reduced from $O(N_b K)$ to $O(K)$.

3.1 Numerical Example II

We construct now the atomic decomposition of the Pop Piano and Classic Guitar clips shown in Fig. 2. Both clips consists of $N = 262144$ samples at 44100Hz each (5.94 secs length). The global sparsity of the signal approximation is measured by the Sparsity Ratio (SR) which is defined as $\mathrm{SR} = \dfrac{N}{K}$, 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.

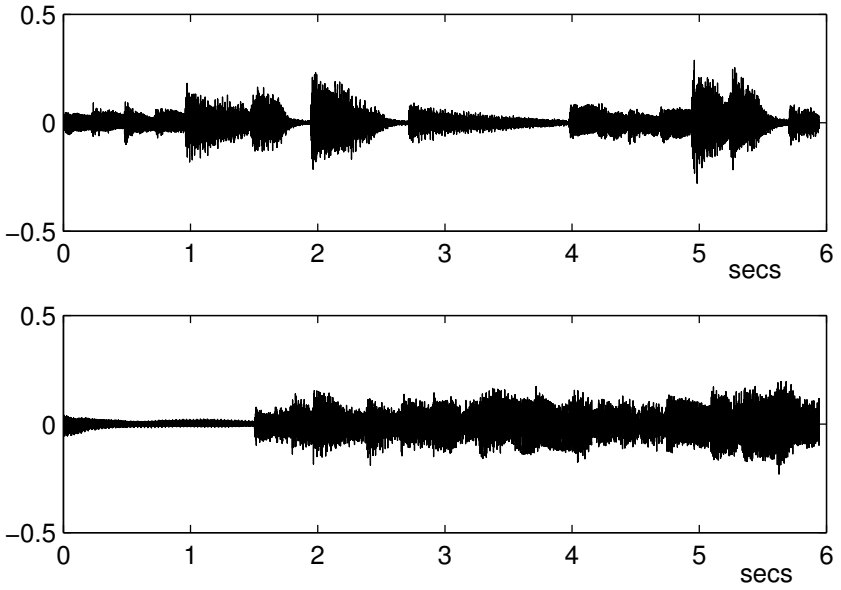

Figure 2: Pop Piano (top graph) and Classic Guitar music signals. Both clips consist of $N = 262144$ samples at 44100Hz each (5.94 secs length).

The convenience of applying the HBW version of a greedy approach for approximating by partitioning using the dictionary introduced in Sec. 2.4 was previously discussed in [45]. Here we are simply extending the applicability of the HBW-OMP approach by implementing it via SPMP. Certainly, the numerical study of this section would not be possible to realize otherwise in a standard computer.

The sparsity results of the clips in Fig. 2 are shown in Fig. 3, 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. 3 for the two clips in Fig. 2 the gain in sparsity achieved by implementing the SPMP approach in the HBW manner is significant. **Note:** The MATLAB functions implementing the methods, as well as the signals used in the numerical examples, have been made available on [47].

4 Conclusions

The geometric convergence 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

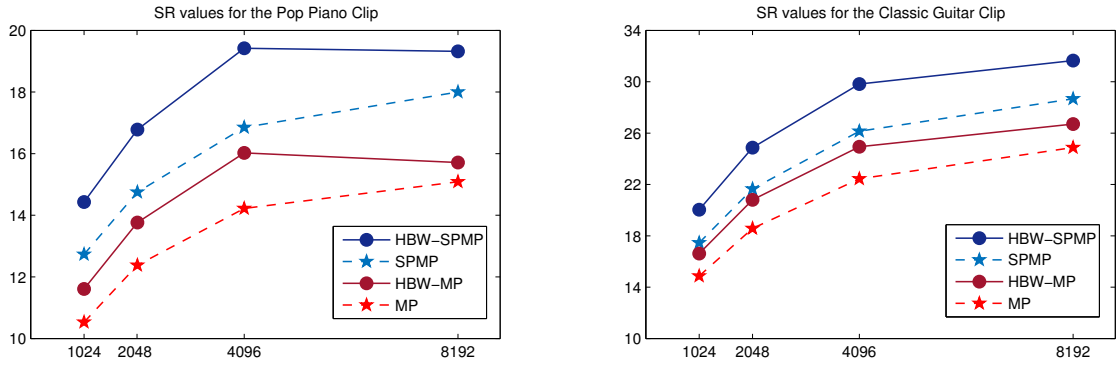

Figure 3: SR vs partition unity size $N_b = 1024, 2048, 4096$ and 8192 samples for the music clips of Fig. 2. The graph on left corresponds to the Pop Piano and the other to the Classic Guitar.

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. (15)) 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 geometric 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, 27–29] 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.

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] J. Wang, S. Kwon, B. Shim, "Generalized Orthogonal Matching Pursuit",*IEEE Transactions on Signal Processing*, **60**, 6202 – 6216 (2012).

[18] N. B. Karahanoglu, H. Erdogan, "Improving A∗OMP: Theoretical and empirical analyses with a novel dynamic cost model", *Signal Processing*, **118**, 62–74 (2016).

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

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

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

[22] J. Zhao, R. Song, J. Zhao, W-P. Zhu, "New conditions for uniformly recovering sparse signals via orthogonal matching pursuit", *Signal Processing*, **106**, 106–113 (2015).

[23] Bo Li, Yi Shen, Zhenghua Wu, Jia Li, "Sufficient conditions for generalized Orthogonal Matching Pursuit in noisy case", *Signal Processing*, **108**, 111 – 123 (2015).

[24] Bo Li, Yi Shen, Sreeraman Rajan, T. Kirubarajan, "Theoretical results for sparse signal recovery with noises using generalized OMP algorithm", *Signal Processing*, **117**, 270 – 278 (2015).

[25] G. Bai, R. Tao, J. Zhao, X. Bai, "Parameter-searched OMP method for eliminating basis mismatch in space-time spectrum estimation", *Signal Processing*, **138**, 11–15 (2017).

[26] 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`.

[27] 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);

[28] Daniel Whitehouse, "Sparse Image Approximation in 3-Dimensions with GPU Utilisation", BSc Computer Science and Mathematics Dissertation (2017) available on `https://dannyxd11.github.io/SIA_3D/docs/`.

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

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

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

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

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

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

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

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

[37] 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).

[38] 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).

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

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

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

[42] 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

[43] 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).

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

[45] `http://www.nonlinear-approx.info/examples/node04.html`