Optimised Orthogonal Matching Pursuit Approach

Laura Rebollo-Neira and David Lowe
Aston University,
Birmingham B4 7ET,
United Kingdom

Abstract

An adaptive procedure for signal representation is proposed. The representation is built up through functions (atoms) selected from a redundant family (dictionary).

At each iteration the algorithm gives rise to an approximation of a given signal, which is guaranteed a) to be the orthogonal projection of a signal onto the subspace generated by the selected atoms, and b) to minimise the norm of the corresponding residual error.

The approach is termed Optimised Orthogonal Matching Pursuit because it improves upon the earlier proposed Matching Pursuit and Orthogonal Matching Pursuit approaches.

I. Introduction

Traditional methods for signal representation involve the use of orthogonal bases. Nevertheless, recent methodologies for this purpose have been developed outside the basis setting. These new methodologies operate by decomposing an arbitrary signal into a linear expansion of waveforms, called "atoms", that are selected from a large and, in general, redundant family of functions called a "dictionary"[1]. Most of the new dictionaries are redundant and have been constructed within the frame structure [3], [4], [5], [6], [7], [8], [9]. However, dictionaries of the same redundant nature also arise as the result of merging bases [10]. Among the existing methods for decomposing a signal in terms of dictionary atoms [1], [6], [7], [10], [11], [12], [13], [14] we focus here on the Matching Pursuit (MP) approach [1], since we wish to discuss a natural improvement to that methodology.

MP is a technique to compute adaptive signal representations. The general goal of this technique is to obtain a sparse signal representation by choosing, at each iteration, a dictionary atom that is best adapted to approximate part of the signal. Nonetheless, the MP algorithm in its original form [1] does not provide at each iteration the linear expansion of the selected atoms that approximate the signal at best. A later refinement which does provide such approximation has been termed Orthogonal Matching Pursuit (OMP) [15].

The OMP approach improves upon the MP in the following sense: from the selected atoms through the MP criterion, the OMP approach gives rise, at each iteration, to the set of coefficients yielding the linear expansion that minimises the distance to the signal. However, since it selects the atoms according to the MP prescription, the selection criterion is not optimal in the sense of minimising the residual of the new approximation.

We show here that a natural requirement within the MP philosophy results in a selection criterion amenable to be implemented by the Orthogonal Least Square (OLS) learning approach earlier introduced in the neural networks context [16]. Nevertheless, the mathematical derivation we present in

this letter yields an iterative procedure allowing us to simultaneously adapt the signal representation so as to provide, at each iteration, the coefficients of the linear expansion in terms of the selected atoms that represent the signal optimally. Such a procedure, that we term Optimised Orthogonal Matching Pursuit (OOMP), improves upon the OMP one by providing, at each step, the coefficients of the linear expansion that minimises the distance to the signal, and the dictionary atom which is optimal in the same sense.

The letter is organised as follows: In section II the MP procedure is briefly discussed. Section III gives the foundations of the proposed OOMP approach and sketches the algorithm for its implementation. The final remarks are presented in section IV.

II. The MP approach

Let $\mathcal{H}$ be a Hilbert space, Γ a set of indices, and $\mathcal{D} = \{\alpha_n \; ; \; n \in \Gamma\}$ a family of functions in $\mathcal{H}$ each of them normalised to unity. Since in practice one deals with a finite number of such functions we shall assume that Γ is a finite set of, say, N indices and re-label them as: $n = 1, \ldots, N$. The family $\mathcal{D}$ is referred to as a dictionary and we denote S as the linear span of the dictionary functions, which are called atoms. Thus, a dictionary is complete if for $\lim_{N\to\infty}$ the closure of S is equal to $\mathcal{H}$. As already mentioned, a dictionary is in general redundant, which implies that in the finite dimensional case the dimension of S is less than N and in the $\lim_{N\to\infty}$ a complete dictionary is actually overcomplete.

Given a signal $f \in \mathcal{H}$ the aim is to represent it as a linear combination of atoms selected from $\mathcal{D}$. The MP approach proposes to make the selection by successive approximations of f. At each step the corresponding residue is sub-decomposed by projecting it on the dictionary atom that matches the residue at best. Let R_k be the kth order residue and l_k the index n for which the corresponding dictionary atom α_{l_k} yields a maximal value of $|\langle \alpha_n, R_k \rangle| \; ; \; n = 1, \ldots, N$, where $\langle ., . \rangle$ indicates the inner product in $\mathcal{H}$.

Starting with an initial approximation $f_1 = 0$ and residue $R_1 = f$ the algorithm evolves by sub-decomposing the kth order residue into

$$R_k = \langle \alpha_n, R_k \rangle \alpha_n + R_{k+1}, \tag{1}$$

which defines the residue at order $k+1$. Since R_{k+1} given in (1) is orthogonal to α_n

$$||R_k||^2 = |\langle \alpha_n, R_k \rangle|^2 + ||R_{k+1}||^2 \tag{2}$$

hence, the dictionary atom α_{l_k} yielding a maximal value of $|\langle R_k, \alpha_n \rangle|$ minimises $||R_{k+1}||^2$.

From Eq (1) it follows that at iteration k the MP algorithm results in an intermediate representation of the form:

$$f = f_k + R_{k+1} \tag{3}$$

with

$$f_k = \sum_{n=1}^{k} \langle \alpha_{l_n}, R_n \rangle \alpha_{l_n}. \tag{4}$$

Provided that the dictionary is a complete one, in the limit $k \to \infty$ the sequence f_k given by (4) has been shown to converge to f [1]. Nevertheless, if the algorithm is stopped at iteration k the function f_k recovers an approximation of f with an error equal to R_{k+1}. Since the family of atoms α_{l_n} ; $n = 1, \ldots, k$ is in general not orthogonal, the residual R_{k+1} will not be in general orthogonal to the subspace generated by the selected atoms. Let us denote V_k to such a subspace and let operator $\hat{P}_{V_k}$ be the orthogonal projector on V_k. Then $\hat{P}_{V_k}$ is self-adjoint (i.e. it satisfies the relation $\langle \hat{P}_{V_k} f, g \rangle = \langle f, \hat{P}_{V_k} g \rangle$, for all f and $g \in \mathcal{H}$) and idempotent (i.e. $\hat{P}_{V_k} \hat{P}_{V_k} = \hat{P}_{V_k}$). It is appropriate to recall at this point that the closest function to f that can be written as a linear expansion of the $k-$atoms α_{l_n} ; $n = 1, \ldots, k$ is $\hat{P}_{V_k} f$. Indeed, let g be an arbitrary function in V_k and let us write it as $g = g - \hat{P}_{V_k} f + \hat{P}_{V_k} f$. If we calculate the distance $||f - g||^2$, since $(f - \hat{P}_{V_k} f) \in V_k^{\perp}$ (the orthogonal complement of V_k in $\mathcal{H}$) we have: $||f - g||^2 = ||f - g - \hat{P}_{V_k} f + \hat{P}_{V_k} f||^2 = ||f - \hat{P}_{V_k} f||^2 + ||\hat{P}_{V_k} f - g||^2$, hence the distance is minimised if $g \equiv \hat{P}_{V_k} f$. Since, unless the family α_{l_n} ; $n = 1, \ldots, k$ is orthogonal, f_k given in (4) is not the orthogonal projection of f on V_k, (4) is not the closest approximation of f that we can obtained in V_k. A refinement to the MP method, which does yield an orthogonal projection approach at each step, has been termed OMP in [15]. The OMP approach improves MP convergence rate and therefore amounts to a better approximation of a signal after a finite number of iterations. However, at each iteration OMP keeps selecting the dictionary atom as prescribed by the MP approach, although such a selection is no longer an optimal one. In the next section we introduce a new method that we term Optimised Orthogonal Matching Pursuit (OOMP) which, in addition to providing the orthogonal projection approximation, at each step selects the dictionary atom that minimises the corresponding residual error.

III. The OOMP approach

The above mentioned OMP approach provides a decomposition of the signal as given by:

$$f = \sum_{n=1}^{k} c_n^{(k)} \alpha_n + \tilde{R}_k \tag{5}$$

where the coefficients $c_n^{(k)}$ are computed so as to guarantee that $\sum_{n=1}^{k} c_n^{(k)} \alpha_n = \hat{P}_{V_k} f$. The superscript of $c_n^{(k)}$ indicates the dependence of these figures on the approximation step, which entails that at each iteration the coefficients are to be re-calculated. Although the OMP approach gives rise to the best approximation of f that one can obtain by means of the selected atoms, since the atoms are selected as prescribed by the MP criterion, the choice is not optimum. Indeed, maximisation of $|\langle \alpha_n, R_k, \rangle|$

minimises the residue R_{k+1} of the MP decomposition (3) but it does not guarantee that the residue of the OMP decomposition (5) is minimised. In order to achieve optimality in the latter sense we propose a method, that we term OOMP, which, at each step yields the coefficients $c_n^{(k)}$ giving rise to the orthogonal projection approximation of f, and selects the dictionary atom that minimises the corresponding residue.

Let α_{l_1} be a function taken out of the set $\alpha_n\,;\, n=1,\ldots,N$ and let us define $V_1=\alpha_{l_1}$ and $V_{k+1}=V_k\oplus\alpha_{l_{k+1}}$. Thus, by denoting W_{k+1} as the orthogonal complement of V_k in V_{k+1}, the orthogonal projector operator onto V_{k+1} can be written as: $\hat{P}_{V_{k+1}}=\hat{P}_{V_k}+\hat{P}_{W_{k+1}}$. Consequently, since by definition $\alpha_{l_{k+1}}\in V_{k+1}$, the orthogonal projection of $\alpha_{l_{k+1}}$ onto W_{k+1} is a function, ψ_{k+1} say, which satisfies

$$\psi_{k+1}=\hat{P}_{W_{k+1}}\alpha_{l_{k+1}}=\hat{P}_{V_{k+1}}\alpha_{l_{k+1}}-\hat{P}_{V_k}\alpha_{l_{k+1}}=\alpha_{l_{k+1}}-\hat{P}_{V_k}\alpha_{l_{k+1}}. \tag{6}$$

The subspace W_{k+1} is spanned by the single function ψ_{k+1} so that disregarding those functions ψ_{k+1} of zero norm we can define the normalised to unity functions

$$\tilde{\psi}_{k+1}=\frac{\psi_{k+1}}{||\psi_{k+1}||} \tag{7}$$

each of which provides the representation of the corresponding orthogonal projection operator onto W_{k+1} as given by: $\hat{P}_{W_{k+1}}f=\tilde{\psi}_{k+1}\langle\tilde{\psi}_{k+1},f\rangle$.

Our aim is to select, at each step, the function $\alpha_{l_{k+1}}$ minimising the residue $||f-\hat{P}_{V_{k+1}}f||^2$. The following proposition is in order.

Proposition 1: Let functions $\tilde{\psi}_{k+1}$ and ψ_{k+1} be given by (7) and (6), respectively, with $\psi_1=\tilde{\psi}_1=\alpha_{l_1}$ Let function $\beta_{l_1}^{(1)}$ be equal to α_{l_1} and functions $\beta_{l_n}^{(k+1)}$ be defined as:

$$\beta_{l_n}^{(k+1)} = \beta_{l_n}^{(k)}-\frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}\langle\alpha_{l_{k+1}},\beta_{l_n}^{(k)}\rangle \quad ; \quad n=1,\ldots,k \tag{8}$$

$$\beta_{l_{k+1}}^{(k+1)} = \frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}. \tag{9}$$

The above defined functions provide a representation of the orthogonal projection operator onto V_{k+1} as given by:

$$\hat{P}_{V_{k+1}}f=\sum_{n=1}^{k+1}\langle\beta_{l_n}^{(k+1)},f\rangle\alpha_{l_n}. \tag{10}$$

The proof is achieved by induction. For $k+1=1$ $\langle\beta_{l_1}^{(1)},f\rangle\alpha_{l_1}=\langle\alpha_{l_1},f\rangle\alpha_{l_1}$ is obviously equal to $\hat{P}_{V_1}f$ because the residue $R_1=f-\langle\alpha_{l_1},f\rangle\alpha_{l_1}$ is orthogonal to α_{l_1} as it readily follows, i.e. $\langle\alpha_{l_1},R_1\rangle=0$. Assuming that (10) holds for label k we shall show it holds for $k+1$. Indeed, using (9) and (8) we obtain:

$$\sum_{n=1}^{k+1}\langle\beta_{l_n}^{(k+1)},f\rangle\alpha_{l_n} = \sum_{n=1}^{k}\langle\beta_{l_n}^{(k+1)},f\rangle\alpha_{l_n}+\langle\frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||},f\rangle\alpha_{l_{k+1}}$$

$$= \sum_{n=1}^{k} \langle \beta_{l_n}^{(k)}, f \rangle \alpha_{l_n} - \sum_{n=1}^{k} \langle \beta_{l_n}^{(k)}, \alpha_{l_{k+1}} \rangle \langle \frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}, f \rangle \alpha_{l_n} + \langle \frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}, f \rangle \alpha_{l_{k+1}} \quad (11)$$

By hypothesis $\sum_{n=1}^{k} \langle \beta_{l_n}^{(k)}, f \rangle \alpha_{l_n} = \hat{P}_{V_k} f$ and $\sum_{n=1}^{k} \alpha_{l_n} \langle \beta_{l_n}^{(k)}, \alpha_{l_{k+1}} \rangle = \hat{P}_{V_k} \alpha_{l_{k+1}}$ so that from (11) we have:

$$\begin{aligned} \sum_{n=1}^{k+1} \langle \beta_{l_n}^{(k+1)}, f \rangle \alpha_{l_n} &= \hat{P}_{V_k} f - \hat{P}_{V_k} \alpha_{l_{k+1}} \langle \frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}, f \rangle + \langle \frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}, f \rangle \alpha_{l_{k+1}} \\ &= \hat{P}_{V_k} f + (\alpha_{l_{k+1}} - \hat{P}_{V_k} \alpha_{l_{k+1}}) \langle \frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}, f \rangle = \hat{P}_{V_k} f + \hat{P}_{W_{k+1}} f = \hat{P}_{V_{k+1}} f \square \quad (12) \end{aligned}$$

Corollary 1: The coefficients $c_{l_n}^{(k+1)}$ yielding the orthogonal projection of f onto V_{k+1} can be recursively obtained as:

$$c_{l_n}^{(k+1)} = c_{l_n}^{(k)} - \langle \beta_{l_n}^{(k)}, \alpha_{l_{k+1}} \rangle \langle \frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}, f \rangle \quad ; \quad n = 1, \ldots, k \quad (13)$$

$$c_{l_{k+1}}^{k+1} = \langle \frac{\tilde{\psi}_{k+1}}{||\psi_{k+1}||}, f \rangle, \quad (14)$$

with $c_1^1 = \langle \alpha_{l_1}, f \rangle$.

Proof: For coefficients $c_{l_n}^{(k+1)}$ to provide the orthogonal projection of f onto V_{k+1} they should satisfy

$$\hat{P}_{V_{k+1}} f = \sum_{n=1}^{k+1} c_{l_n}^{(k+1)} \alpha_{l_n} \quad (15)$$

Thus (13) and (14) readily follow by using (8) and (9) in (10) and identifying $c_{l_n}^{(k+1)}$ with $\langle \beta_{l_n}^{(k+1)}, f \rangle \square$

We are now ready to settle the OOMP selection criterion.

Theorem 1: The dictionary atom $\alpha_{l_{k+1}}$ that at iteration $k+1$ minimises the norm of the residue $\tilde{R}_{k+1}$ is the one yielding a maximal value of the functionals e_n ; $n = 1, \ldots, N$ given by

$$e_n = \frac{b_n}{d_n} = \frac{|\langle \alpha_n, \tilde{R}_k \rangle|^2}{1 - \langle \alpha_n, \hat{P}_{V_k} \alpha_n \rangle} \quad ; \quad b_n > 0. \quad (16)$$

Proof: As discussed above, at iteration $k+1$ the OOMP residue should verify: $\tilde{R}_{k+1} = f - \hat{P}_{V_{k+1}} f$ so that $||\tilde{R}_{k+1}||^2 = ||f||^2 - \langle \hat{P}_{V_{k+1}} f, f \rangle$. Thereby, in order to minimise $||\tilde{R}_{k+1}||^2$ we should maximise $\langle \hat{P}_{V_{k+1}} f, f \rangle$. From Preposition 1 (cf. eq. (12)) we have:

$$\langle \hat{P}_{V_{k+1}} f, f \rangle = \langle \hat{P}_{V_k} f, f \rangle + \frac{1}{||\psi_{k+1}||^2} |\langle \alpha_{l_{k+1}}, f \rangle - \langle \hat{P}_{V_k} \alpha_{l_{k+1}}, f \rangle|^2 \quad (17)$$

so that, since $\langle \hat{P}_{V_k} f, f \rangle$ is fixed in the previous iteration, (17) is maximised if

$$\frac{|\langle \alpha_{l_{k+1}}, f \rangle - \langle \hat{P}_{V_k} \alpha_{l_{k+1}}, f \rangle|^2}{||\psi_{k+1}||^2} = \frac{|\langle \alpha_{l_{k+1}}, f - \hat{P}_{V_k} f \rangle|^2}{1 - \langle \alpha_{l_{k+1}}, \hat{P}_{V_k} \alpha_{l_{k+1}} \rangle} = \frac{|\langle \alpha_{l_{k+1}}, \tilde{R}_k \rangle|^2}{1 - \langle \alpha_{l_{k+1}}, \hat{P}_{V_k} \alpha_{k+1} \rangle} \quad (18)$$

is maximised $\square$

From the last equation it is clearly seen that the atoms that at iteration $k+1$ have a linear dependence with the previously selected atoms yield a value of b_n (cf. eq. (16)) equal to zero, because they

fulfill the condition $\hat{P}_{V_k}\alpha_n = \alpha_n$. Hence, through the restriction $b_n > 0$, all the selected atoms are guaranteed to be linearly independent.

Notice that the additional computational task the OOMP condition introduces, with respect to the original MP one, is the evaluation of d_n, the denominator of (16). Corollary 2 below shows that this condition admits a fast implementation through the so called OLS technique, earlier introduced in the context of neural network structure optimisation [16].

Corollary 2 The OOMP selection criterion given in (16) is amenable to be implementing by maximising the expression:

$$\frac{|\langle \phi_n, f\rangle|^2}{||\phi_n||^2} \qquad ; \quad ||\phi_n|| \neq 0 \tag{19}$$

where for each n the corresponding function ϕ_n is obtained by orthogonalising α_n through the Gram Schmidt technique, with respect to the previously selected atoms, i.e., $\phi_n = \alpha_n - \sum_{j=1}^{k} \tilde{\phi}_j \langle \tilde{\phi}_j, \alpha_n\rangle$, where $\tilde{\phi}_j = \phi_j/||\phi_j||$.

Proof: From the left-hand-side of (18) we have that (16) can be expressed as $\frac{|\langle \alpha_n, f\rangle - \langle \hat{P}_{V_k}\alpha_n, f\rangle|^2}{||\psi_n||^2} = \frac{|\langle \psi_n, f\rangle|^2}{||\psi_n||^2}$

On the other hand, if we define $\hat{P}_{V_0}$ as the zero operator, by successive applications of the relation $\hat{P}_{V_k} = \hat{P}_{V_{k-1}} + \hat{P}_{W_k}$ one arrives at $\hat{P}_{V_k} = \sum_{j=1}^{k} \hat{P}_{W_j}$. By using this in (6) it follows that $\psi_n = \alpha_n - \sum_{j=1}^{k} \tilde{\psi}_j \langle \tilde{\psi}_j, \alpha_n\rangle \equiv \phi_n \Box$

Algorithm Sketch

We sketch here an algorithm for selecting functions by implementing condition (19), and successively adapting the corresponding coefficients yielding the orthogonal projection of the signal onto the selected subspace.

From $\tilde{R}_0 = f$ and the inner products $\langle \alpha_n, f\rangle\,;\ n = 1, \ldots, N$ the OOMP algorithm evolves as follows:

i) Initially set:

$k = 1\,;\ \gamma_n = \alpha_n\,;\ d_n = 1\ (n = 1, \ldots, N)$ and l_1 equal to the index n for which $|\langle \alpha_n, f\rangle|$ takes a maximal value as n ranges from 1 to N.

Assign $\psi_1 = \alpha_{l_1} = \beta_1, c_1 = \langle \alpha_{l_1}, f\rangle$ and $||\tilde{R}_1||^2 = ||\tilde{R}_0||^2 - |c_1|^2$

ii) for $n = 1, \ldots, N$ compute:

$\gamma_n = \gamma_n - \psi_k \langle \psi_k, \alpha_n\rangle$

$b_n = \langle \gamma_n, f\rangle$

$d_n = d_n - |\langle \psi_k, \alpha_n\rangle|^2$ (or $d_n = ||\gamma_n||^2$)

if $|b_n| < \epsilon$ (tolerance parameter)$\,;\ e_n = 0$, otherwise $e_n = |b_n|^2/d_n$

iii) Increase k to $k+1$ and set l_k equal to the index n for which e_n takes a maximal value as n ranges from 1 to N. Assign $||\tilde{R}_k||^2 = ||\tilde{R}_{k-1}||^2 - e_{l_k}$. Assign $\psi_k = \gamma_{l_k}/\sqrt{d_{l_k}}$ and $\beta_k = \gamma_{l_k}/d_{l_k}$
Compute $c_k = \langle \beta_k, f \rangle$

iv) For $n = 1 \ldots, k-1$ compute:
$\beta_n = \beta_n - \beta_k \langle \alpha_{l_k}, \beta_n \rangle$
$c_n = c_n - \overline{\langle \alpha_{l_k}, \beta_n \rangle} c_k$ (where $\overline{\langle \alpha_{l_k}, \beta_n \rangle}$ indicates the complex conjugate of $\langle \alpha_{l_k}, \beta_n \rangle$)

v) Repeat steps *ii)*, and *iii)* and *iv)*.

As in the previous MP algorithms, the OOMP algorithm is to be stopped when some criterion is achieved. In the context of signal representation a popular stopping criterion is given by

$$||\tilde{R}_k||^2 \leq \delta \tag{20}$$

where δ is the desired precision.

It should be stressed that in the case of finite dictionaries, or in the case of incomplete dictionaries when $N \to \infty$, if such dictionaries are inappropriate for representing the signal at hand within the desired precision, all linearly independent atoms will be selected before the stopping criterion is satisfied. If that were the situation, after a number of iterations we would obtain $b_n = 0\,;\, n = 1, \ldots, N$. Then the algorithm should be stopped given the warning "convergence failure".

Let us assume that the given stopping criterion is reached at iteration K. At such stage the above algorithm has selected K indexes $l_k\,;\, k = 1, \ldots, K$ and has computed the corresponding coefficients c_{l_k}. Then we are now in a position to represent the signal f as a linear combination of the corresponding K dictionary atoms $\alpha_{l_n}\,;\, n = 1, \ldots, K$, i.e.

$$f = \sum_{n=1}^{K} c_{l_n} \alpha_{l_n} + \tilde{R}_K. \tag{21}$$

IV. Conclusions:

A method for representing signals by selecting functions (atoms) from a redundant family (dictionary) has been advanced. The approach, that we termed OOMP, improves upon the previous MP and OMP approaches in that at each step it provides: a) the orthogonal projection of the signal onto the subspace generated by the selected atoms and b) the dictionary atoms minimising the norm of the corresponding residual error. It has been shown that the proposed selection criterion is amenable to be implemented by the OLS technique earlier introduced in the neural network context. Nevertheless, the derivation we give here amounts to a procedure for adapting, at each iteration, the coefficients of the linear expansion approximating the signal.

The above remarks, along with the fact that the technique allows for the use of atoms of different nature for representing different properties of a given signal, leads us to expect the proposed method will be of assistance in a broad range of signal processing applications.

References

[1] S. Mallat and Z. Zhang, "Matching Pursuit in time-frequency dictionary". IEEE Transactions on Signal Processing, Vol 41, pp 3397-3415 (1993)

[2] R. J. Duffin, A. C. Shaffer, "A Class of Nonharmonic Fourier Series", *Trans. Amer. Math. Soc.*, Vol 72, pp 341-366, 1952.

[3] R. M. Young, *An Introduction to Nonharmonic Fourier Series* (Academic Press, New York, 1980).

[4] I. Daubechies, A. Grossmann, Y. Meyer, "Painless nonorthogonal expansions", Journal of Mathematical Physics, Vol 27, pp. 1271-1283, (1986).

[5] I. Daubechies, "The Wavelets Transform, Time Frequency Localisation and Signal Analysis", IEEE Transactions on Information Theory Vol 36, pp 961-1005 (1990)

[6] I. Daubechies, *Ten Lectures on Wavelets* (CBMS-NSF, SIAM, Philadelphia, 1992).

[7] G. Kaiser, *A Friendly Guide to Wavelets* (Birkhäuser, Berlin, 1994).

[8] L. Rebollo-Neira, "Frames of cross- ambiguity functions", IEEE Signal Processing Letters, Vol. 7, 10, pp 293-296 (2000).

[9] L. Rebollo-Neira,"Frames in two dimensions arising from wavelet transforms", Proceedings of the Royal Society, series A, Vol 457, 2013, pp 2079-2091 (2001)

[10] A. P. Berg, W. B. Mikhael, "An efficient structure and algorithm for image representation using non-orthogonal basis images", Vol. 44, 10, pp 818-828 (1997).

[11] L. Rebollo-Neira, J. Fernandez-Rubio, A. Plastino, "Frames a Maximum Entropy Statistical Estimate of the Inverse Problem", Journal of Mathematical Physics, Vol 38, 9, pp 4863-4871 (1997).

[12] R. R. Coifman, M. V. Wickerhauser. 'Entropy-based algorithms for best-basis selection", IEEE Transactions on Information Theory , Vol 38, pp 713-718 (1992).

[13] R. R. Coifman, Y Meyer, M. V. Wickerhauser, "Wavelets analysis and signal processing" pp 153-178, and "Size properties of wavelet packets" pp 453-470, in *Wavelets and its applications*, M. B. Ruskai et al eds, Jones and Bartlett, Boston (1992).

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

[15] Y. C. Pati, R Rezaiifar, P. S. Krishnaprasad. "Orthogonal matching pursuits: recursive function approximation with applications to wavelet decomposition. "Proceedings of the 27th Asilomar Conference in Signals, Systems and Computers (1993)

[16] S. Chen, C. F. N. Cowan, and P. M Grant, "Orthogonal Least Square Algorithm for Radial Basis Function Networks, IEEE transaction on Neural Networks, Vol 2, 2, pp 302-309 (1991)