Alternative methods to ordinary least squares : matching pursuit and orthogonal matching pursuit

February 18, 2016

Premia 18

1 Introduction

The following methods deal with two alternatives to Least Squares Method : the matching pursuit method, introduced in [1] and the orthogonal matching pursuit, introduced in [2].

2 Matching pursuit (MP)

The paper introduces an algorithm, called matching pursuit, that decomposes any signal into a linear expansion of waveforms that are selected from a redundant dictionary of functions. Theses waveforms are chosen in order to best match the signal structures.

Let H be an Hilbert space and define a dictionary as a family $D = (x_i)_{i\in\mathbb{N}}$ of vectors in H. Let $y \in H$. We want to compute a linear expansion of y over a set of vectors selected from D, in order to best match its inner structures. This is done by successive approximations of y with orthogonal projections on elements of D. Let $x_{\gamma_0} \in D$. The vector y can be decomposed into

$$y = \langle y, x_{\gamma_0}\rangle x_{\gamma_0} + Ry$$

where Ry is the residual vector after approximating y in the direction x_{γ_0}. Clearly x_{γ_0} is orthogonal to Ry, hence

$$\|y\|^2 = \|\langle y, x_{\gamma_0}\rangle\|^2 + \|Ry\|^2.$$

To minimize $\|Ry\|^2$, we must choose $x_{\gamma_0} \in D$ such that $\|\langle y, x_{\gamma_0}\rangle\|$ is maximum.

A matching pursuit is an iterative algorithm that subdecomposes the residue Ry by projecting it on a vector of D that matches Ry at best, as it as been done for y. This procedure is repeated as many times as necessary.

Description of the algorithm

Let $\hat{y}^{(k)}$ denote the approximation of y at step k and $r^{(k)}$ denote the k^{th} residue, i.e. $r^{(k)} := y - \hat{y}^{(k)}$.

1. Initialisation : $\hat{y}^{(0)} = 0$, $r^{(0)} = y$ and $k = 1$. At step k
2. we compute $(\frac{\|\langle r^{(k-1)}, x_i \rangle\|}{\|x_i\|})_{i=1,\cdots,n}$.
3. Among the dictionary directions, find the direction γ_k that best describes the residual vector and compute the associated coefficient

$$\gamma_k := \arg\max_i \frac{\|\langle r^{(k-1)}, x_i \rangle\|}{\|x_i\|},$$
$$a_k := \frac{\langle r^{(k-1)}, x_{\gamma_k} \rangle}{\|x_{\gamma_k}\|^2}.$$

4. Substract all the information along the best direction from the data and compute the residual vector

$$\hat{y}^{(k)} = \hat{y}^{(k-1)} + a_k x_{\gamma_k},$$
$$r^{(k)} = y - \hat{y}^{(k)} = r^{(k-1)} - a_k x_{\gamma_k}$$

5. Increment k and repeat steps 2,3 and 4 until $\|y - \hat{y}^{(k)}\| \leq \varepsilon$, ε is specified at the beginning of the procedure.

3 Orthogonal matching pursuit (OMP)

The matching pursuit method is very easy to implement and very fast. However, when the dictionary we consider does not contain orthogonal functions, the convergence is very slow. At the first step, we have $\langle r^{(1)}, x_{\gamma_1} \rangle = 0$. At the second step, we compute γ_2 such that $\langle r^{(2)}, x_{\gamma_2} \rangle = 0$, but we also have $\langle r^{(2)}, x_{\gamma_1} \rangle \neq 0$. Then it is possible to choose again x_{γ_1} at an other step in the algorithm. The following algorithm enables to circumvent this problem.
It is a modified version of the mathing pursuit algorithm. At each step of the algorithm, we find the direction that best describes the residual vector but, contrary to MP algorithm, we recompute the estimation $\hat{y}^{(k)}$. $\hat{y}^{(k)}$ is now defined as the projection of y on the space spanned by $\{x_{\gamma_1}, \cdots, x_{\gamma_k}\}$. At the k^{th} iteration, we have, for all $i \in \{1, \cdots, k\}$, $\langle r^{(k)}, x_{\gamma_i} \rangle = 0$.

Description of the algorithm
Let $\hat{y}^{(k)}$ still denote the approximation of y at step k and $r^{(k)}$ denote the k^{th} residue, i.e. $r^{(k)} := y - \hat{y}^{(k)}$. We also introduce the matrix $X_k := [x_{\gamma_1}| \cdots |x_{\gamma_k}]$ and $\hat{a}^{(k)}$ the vector composed by the coefficients of the linear regression. We recall that all the coefficients are recomputed at each iteration, so that $r^{(k)}$ does not belong to the space spanned by X_k.

1. Initialisation : $\hat{y}^{(0)} = 0$, $r^{(0)} = y$, $X_0 = [\,]$ and $k = 1$. At step k
2. we compute $(\frac{\|\langle r^{(k-1)}, x_i \rangle\|}{\|x_i\|})_{i=1,\cdots,n}$.
3. Among the dictionary directions, find the direction γ_k that best describes the residual vector and compute the associated coefficient

$$\gamma_k := \arg\max_i \frac{\|\langle r^{(k-1)}, x_i \rangle\|}{\|x_i\|},$$

4. Substract all the information along the best direction from the data and compute the residual vector

$$X_k = [X_{k-1}|x_{\gamma_k}],$$
$$\hat{a}^{(k)} = X_k^+ y \text{ then } \hat{y}^{(k)} = X_k \hat{a}^{(k)}$$
$$r^{(k)} = y - \hat{y}^{(k)}$$

5. Increment k and repeat steps 2,3 and 4 until $\|y - \hat{y}^{(k)}\| \leq \varepsilon$, ε is specified at the beginning of the procedure.

where $X_k^+ = (X_k^T X_k)^{-1} X_k^T$ is the pseudo-inverse of Moore-Penrose of X_k.

References

[1] Stéphane G. Mallat and Zhifeng Zhang Matching pursuits with time frequency dictionaries IEEE Transactions on Signal Processing, Vol 41, N° 12, December 1993

[2] Y.C. Pati, R. Rezahfar and P.S. Krishnaprasad Orthogonal matching pursuit : recursive function approximation with application to wavelet decomposition Processes of the 27th annual Asilomar conference on Signals systems and computers, November 1-3th 1993 1

1