International Journal of Pure and Applied Mathematics
Volume 15 No. 1 2004, 25-34

# A MODIFIED MATCHING PURSUIT ALGORITHM APPLIED TO THE APPROXIMATION OF VECTO-CARDIOGRAM DATA

Catherine Bonnet[1 §], Jonathan R. Partington[2], Michel Sorine[3]
[1,3]INRIA Rocquencourt
Domaine de Voluceau
B.P. 105, 78153 Le Chesnay Cedex, FRANCE
[1]e-mail: Catherine.Bonnet@inria.fr
[3]e-mail: Michel.Sorine@inria.fr
[2]School of Mathematics
University of Leeds
Leeds LS2 9JT, UK
e-mail: J.R.Partington@leeds.ac.uk

**Abstract:** A modification of the matching pursuit algorithm is presented which allows the recovery of a vector-valued signal from partial data (a relaxed version of the algorithm is also introduced). The new algorithm is illustrated by its application to the analysis of a vecto-cardiogram signal.

**AMS Subject Classification:** 94A12, 92C50
**Key Words:** vectorial matching pursuit algorithm, vecto-cardiogram signal

## 1. Introduction

The Matching Pursuit algorithm (MP) presented in Mallat et al [13], Dudley Ward et al [8], Partington [15] and Davis et al [6] is an algorithm that has been used in such settings as signal and image processing when one has large amounts

Received: April 6, 2004

§Correspondence author

of information available from which one must select the most important features. This enables one to obtain a more easily-analysed approximation to the given data. Versions of this algorithm have also been used in medical contexts for the analysis of breathing rhythms in Akay et al [1] and brain patterns in Durka et al [9].

In this paper we present two refinements of the algorithm. One is a *relaxed* version which in some situations leads to faster convergence. The other refinement allows us to handle vector-valued data, and this is illustrated by an application to the analysis of vecto-cardiogram (VCG) data. Other techniques have been applied to the analysis of scalar-valued (ECG) data, and we mention the papers of Berger et al [2], Daskalov et al [4] and DeBoer et al [7] for general background information.

In Section 2 we present the mathematical basis of the modified MP, and in Section 3 we see how this can be applied to signal processing. Section 4 contains the numerical example, illustrating the technique.

## 2. The Matching Pursuit Algorithm

### 2.1. The Basic Algorithm and its Relaxed Version

The standard MP, as given in Mallat et al [13] and Davis et al [6] allows one to construct approximations to an element of a Hilbert space $H$ using vectors from a fixed set, with the notation of Partington [15, Section 5.2], a *dictionary* of normalized vectors, $\{g_\alpha \in H : \alpha \in A\}$. For computational purposes, $A$ is normally finite, but this is not theoretically necessary. For $f \in H$, we choose $\gamma$ with $0 < \gamma < 1$ and set $f_0 = f$. For each $n \geq 0$ we choose $\alpha_n \in A$ and $f_{n+1} \in H$ such that

$$\left\{ \begin{array}{ccc} |\langle f_n, g_{\alpha_n} \rangle| & \geq & \gamma \sup_{\alpha \in A} |\langle f_n, g_\alpha \rangle|, \\ f_n & = & \lambda \langle f_n, g_{\alpha_n} \rangle g_{\alpha_n} + f_{n+1} := h_n + f_{n+1}, \end{array} \right. \tag{1}$$

where $0 < \lambda < 2$. One can also make $\lambda$ to depend on $n$, providing that the sequence of values does not accumulate near 0 or 2, but we shall treat only the simpler case. Then for $\lambda = 1$ (basic algorithm), $h_0 + h_1 + \ldots$ converges in norm to the orthogonal projection of $f$ onto the closed linear span of the dictionary elements. Note that, if the dictionary is an orthonormal basis, then the expansion produced by the MP is a standard orthonormal expansion, in (approximately) decreasing order of coefficients.

We study the case $\lambda \neq 1$ (relaxed algorithm): although this modified algorithm is definitely suboptimal if the dictionary vectors are orthonormal, it leads to faster convergence in some other circumstances, for example the two-dimensional example $H = \mathbb{R}^2$, $f = (0,1)$, $A = \{1,2\}$, $g_1 = (1,0)$, $g_2 = (\cos\delta, \sin\delta)$, with $\delta > 0$ (small), given in Dudley Ward et al [8]. See also Pati et al [16].

**Theorem 1.** *In the relaxed MP given by the iteration (1), where* $0 < \lambda < 2$, *the sequence* $(f_n)$ *converges in norm to the orthogonal projection of* $f$ *onto the closed linear span of the dictionary.*

*Proof.* This is a fairly straightforward modification of the proof for the case $\lambda = 1$ given in Mallat et al [13]; we are using the same notation as in Partington [15]. First, it is easy to see that we may assume without loss of generality that the closed linear span of the dictionary is the whole of $H$, since vectors orthogonal to the dictionary play no part in the algorithm. Next, we have

$$f_{n+1} = f_n - \lambda \langle f_n, g_{\alpha_n} \rangle g_{\alpha_n}, \tag{2}$$

so that

$$\begin{aligned} \|f_{n+1}\|^2 &= \|f_n\|^2 - 2\lambda |\langle f_n, g_{\alpha_n}\rangle|^2 + \lambda^2 |\langle f_n, g_{\alpha_n}\rangle|^2 \\ &= \|f_n\|^2 + (\lambda^2 - 2\lambda)|\langle f_n, g_{\alpha_n}\rangle|^2 \le \|f_n\|^2, \end{aligned} \tag{3}$$

since $0 < \lambda < 2$. Let $m > n$ and write $f_m = f_n - \sum_{k=n}^{m-1} h_k$, so that

$$\begin{aligned} \|f_n - f_m\|^2 &= \langle f_n - f_m, f_n\rangle - \langle f_n - f_m, f_m\rangle \\ &= \langle f_n, f_n\rangle - \langle f_m, f_n - f_m + f_m\rangle - \langle f_n - f_m, f_m\rangle \\ &= \|f_n\|^2 - \|f_m\|^2 - 2\,\mathrm{Re}\langle f_m, f_n - f_m\rangle \\ &= \|f_n\|^2 - \|f_m\|^2 - 2\,\mathrm{Re}\langle f_m, \sum_{k=n}^{m-1} h_k\rangle. \end{aligned} \tag{4}$$

Now

$$\left|\langle \sum_{k=n}^{m-1} h_k, f_m\rangle\right| \le \sum_{k=n}^{m-1} \lambda |\langle f_k, g_{\alpha_k}\rangle|\, |\langle g_{\alpha_k}, f_m\rangle| \le \sum_{k=n}^{m-1} \|h_k\|\, \|h_m\|/(\lambda\gamma). \tag{5}$$

Also

$$\|h_k\|^2 = \lambda^2 |\langle f_k, g_{\alpha_k}\rangle|^2 = \frac{\lambda^2}{2\lambda - \lambda^2}(\|f_k\|^2 - \|f_{k+1}\|^2), \tag{6}$$

by (3), and hence $C^2 := \sum_{k=1}^{\infty} \|h_k\|^2 < \infty$. Hence, for each $\epsilon > 0$, it is the case that

$$\|h_m\| \le \epsilon/\sqrt{m} \tag{7}$$

for infinitely many $m$. Now, by the Cauchy–Schwarz inequality,

$$\sum_{k=n}^{m-1} \|h_k\| \le \sqrt{m-n} \left( \sum_{k=n}^{m-1} \|h_k\|^2 \right)^{1/2} \le C\sqrt{m}, \tag{8}$$

for all $m > n$. Finally

$$\|f_n - f_m\|^2 \le \|f_n\|^2 - \|f_m\|^2 + 2(C\sqrt{m})(\epsilon/\sqrt{m})/(\lambda\gamma), \tag{9}$$

for infinitely many $m$ and all $n < m$, by (4), (5), (8) and (7), which implies that $(f_n)$ is a Cauchy sequence. We now observe that $(f_n)$, being a bounded sequence, has a subsequence which tends weakly to a limit, necessarily zero, and hence the norm limit of $(f_n)$ is also zero. □

### 2.2. A Vectorial Version

The MP, with $0 < \lambda < 2$, can also be applied in the situation when the given data is vector-valued. We shall now give a rather abstract formulation of the vectorial algorithm: in Section 3.2 we shall describe how this works in a more concrete situation. A related but apparently less general approach can be found in Rao et al [17].

Suppose that we are given a collection of bounded linear operators $\{T_\alpha : H \to K_\alpha : \alpha \in A\}$, where each $K_\alpha$ is a Hilbert space. The vectorial algorithm uses the dictionary $\{T_\alpha^* y : y \in K_\alpha, \|T_\alpha^* y\| = 1, \alpha \in A\}$ (here $T_\alpha^*$ denotes the adjoint of $T_\alpha$). Hence we select at each stage an $\alpha_n \in A$ and $y_n \in K_{\alpha_n}$ such that $\|T_{\alpha_n}^* y_n\| = 1$ and the quantity $\langle f_n, T_{\alpha_n}^* y_n \rangle = \langle T_{\alpha_n} f_n, y_n \rangle$ is large. We then define

$$f_{n+1} = f_n - \lambda \langle T_{\alpha_n} f_n, y_n \rangle T_{\alpha_n}^* y_n, \tag{10}$$

starting the iteration with $f_0 = f$.

The computations become simpler when $\|T_\alpha^* y\|$ is a constant multiple of $\|y\|$, independently of $\alpha$ and the choice of $y \in K_\alpha$. In that case, we select $\alpha_n \in A$ such that $\|T_{\alpha_n} f_n\|$ is large, and take $y_n = T_{\alpha_n} f_n / \|T_{\alpha_n}^* T_{\alpha_n} f_n\|$. Now

$$\begin{aligned} f_{n+1} &= f_n - \lambda \langle f_n, T_{\alpha_n}^* y_n \rangle T_{\alpha_n}^* y_n \\ &= f_n - \lambda \|T_{\alpha_n} f_n\|^2 T_{\alpha_n}^* T_{\alpha_n} f_n / \|T_{\alpha_n}^* T_{\alpha_n} f_n\|^2. \end{aligned}$$

We again obtain norm convergence onto the projection of $f$ onto the closed linear span of the dictionary vectors.

## 3. Application to Paley–Wiener Signals

In the approximation of VCG data we may wish to work with a general vectorial version of the Paley–Wiener space, which can be written as $PW(b, \mathbb{R}^N)$, for some $N \geq 1$: in our case the value $N = 3$ is appropriate. This space consists of vector-valued functions $f : \mathbb{R} \to \mathbb{R}^N$ given by $f(t) = \dfrac{1}{2\pi} \displaystyle\int_{-b}^{b} \hat{f}(x) e^{ixt}\, dx$, where $b > 0$ is the *bandwidth*, and $\hat{f} \in L_2([-b, b], \mathbb{C}^N)$, satisfying $\hat{f}(x) = \overline{\hat{f}(-x)}$ in order that $f$ should be real-valued. $\hat{f}$ is the Fourier transform of $f$.

The inner product is given by $\langle f, g \rangle = \int_{\mathbb{R}} f(t) \cdot g(t)\, dt$, where $(u, v) \mapsto u \cdot v$ is the usual inner product on $\mathbb{R}^N$.

The space $PW(b, \mathbb{R}^3)$ is a closed subspace of $L_2(\mathbb{R}, \mathbb{C}^n)$, and hence a Hilbert space $H$.

The application of MP in this case seems new.

### 3.1. The Scalar Case ($N = 1$)

A key role is played here by the sinc function, defined by

$$\operatorname{sinc} t = \begin{cases} \dfrac{\sin t}{t} & \text{if } \ t \neq 0, \\ 1 & \text{if } \ \text{t}= 0. \end{cases} \tag{11}$$

In particular the Whittaker–Kotel'nikov–Shannon sampling series or cardinal series (see e.g. Higgins [11] or Partington [15]) allows us to recover any $f \in PW(b)$ from regularly-spaced samples, by the formula

$$f(t) = \sum_{n=-\infty}^{\infty} f(n\pi/c) \operatorname{sinc}(ct - n\pi), \tag{12}$$

converging uniformly and in $L_2$ norm, for any $c \geq b$. We take as our dictionary the functions

$$g_\alpha(t) = \sqrt{\frac{b}{\pi}} \operatorname{sinc} b(t - \alpha), \tag{13}$$

where $\alpha$ ranges over the set of sampling points. The $(g_\alpha)$ are normalized, and form an orthonormal basis of $PW(b)$ if we make the choice $\alpha_n = n\pi/b$, although this is not necessary for what follows. We have $\hat{g}_\alpha = \sqrt{\frac{\pi}{b}} e^{-i\alpha x}$ for $x \in [-b, b]$ and is zero otherwise, and (12) can be derived by expansion in terms of this orthonormal basis.

Implementation of the MP may be performed exactly and directly using the sinc functions (other dictionaries would not have this property) because of the standard fact (see e.g. Partington [15]) that the sinc function acts as a reproducing kernel, namely

$$\langle f, k_s \rangle = f(s), \qquad \text{for } f \in PW(b), \tag{14}$$

where $k_s(t) = \frac{b}{\pi} \operatorname{sinc} b(t-s)$, for $s,\, t \in \mathbb{R}$. Hence, given $f_n$, we obtain $f_{n+1}$ by choosing $\alpha = \alpha_n$ to maximize $|f_n(\alpha)|$ over $\alpha$ in the set of sampling points, and writing $f_{n+1} = f_n - \lambda f_n(\alpha_n) \operatorname{sinc} b(t - \alpha_n)$. In the case $\lambda = 1$ this means that $\alpha_n$ is chosen so that $f_{n+1}(\alpha_n) = 0$.

Our decomposition becomes for $f$ and $\hat{f}$

$$f(t) \approx \lambda \sum_{n=0}^{N} f_n(\alpha_n) \operatorname{sinc} b(t - \alpha_n),$$

$$\hat{f}(x) \approx \frac{\lambda\pi}{b} \sum_{n=0}^{N} f_n(\alpha_n) \exp(-i\alpha_n x)\,, \qquad \text{for } |x| < b\,. \tag{15}$$

The advantage of this procedure is that we can obtain good estimates of $f$ and $\hat{f}$, respectively in $PW(b)$ and $L_2(-b, b)$ even using a small amount of data (e.g. thirty values per second in the VCG example which follows).

### 3.2. The Vector-Valued Case

In Section 2.2 we showed how to apply the MP in the situation when we have vector-valued data. We now specify to a particular case, when the process is somewhat more intuitive to describe.

We now take as our dictionary the infinite set of all functions $g_\alpha \otimes v$, defined by

$$(g_\alpha \otimes v)(t) = g_\alpha(t) v, \qquad \text{for } t \in \mathbb{R}, \tag{16}$$

where $\alpha$ is a sampling point, $v$ a unit vector in $\mathbb{R}^N$, and the scalar-valued function $g_\alpha$ as given in (13).

At first sight, implementing the MP looks like an intractable numerical problem, since the dictionary is always infinite if $N > 1$, but we note that using (14)

$$|\langle f, g_\alpha \otimes v \rangle| = \sqrt{\frac{\pi}{b}}\, |f(\alpha) \cdot v|, \tag{17}$$

so that to maximize $|\langle f, g_\alpha \otimes v\rangle|$ over all $\alpha$ and $v$ is equivalent to the simpler problem of maximizing $\|f(\alpha)\|$ over all the sampling points $\alpha$ and taking $v = f(\alpha)/\|f(\alpha)\|$.

The iterative step of the algorithm thus reads, setting $v_n = \frac{f_n(\alpha_n)}{\|f_n(\alpha_n)\|}$

$$\begin{aligned} f_{n+1}(t) &= f_n(t) - \lambda\langle f_n, g_{\alpha_n} \otimes v_n\rangle g_{\alpha_n}(t) v_n \\ &= f_n(t) - \lambda f_n(\alpha_n)\,\mathrm{sinc}\, b(t-\alpha_n), \end{aligned} \quad (18)$$

where now $f_n(t)$ and $f_{n+1}(t)$ are vectors in $\mathbb{R}^N$ for each $t$. Since the algorithm is still a special case of the (relaxed) MP, the convergence follows from Theorem 1.

## 4. Approximation of VCG Data

Without going into details of electrocardiography, let us just recall that an electrical activity results from heart excitation. While this excitation front progresses through the cardiac muscle, electric potentials of different modulus and directions are created. The equivalent dipole of the heart's electrical activity (see Panfilov et al [14]) is the sum vector of these potentials. It varies in length and direction during the excitation cycle, describing a curve in $\mathbb{R}^3$ (the VCG) which might be identified by placing several electrodes on patients at appropriate points in the body. In fact, placing four electrodes will allow one to obtain three scalar-valued signals called ECG, which can be considered as being the orthogonal projections of the VCG if the electrodes are suitably placed.

Only two electrodes are needed to obtain a single ECG. The analysis of cardiac frequency is actually done from this ECG by identifying the most important peaks of the signal (arising at each heart beat) and constructing the so-called RR interval. The instantaneous frequency is given by the inverse of the interval between two peaks.

Although there are still improvements in detecting the RR interval, it is now rather "routine" for cardiologists. However detection of smaller events in the ECG, as the T-wave peak and T-wave end for example (see figure 1), is an actual medical challenge (see for example Daskalov et al [5]). The T-wave is in fact involved in the diagnostic of long QT syndrome which can lead to a fatal issue (see Friedman et al [10] for long QT congenital syndrome and Lancellotti et al [12] for acquired long QT syndrome - due to myocardial infarction) and cardiologists are interested in recognition as there exists an appropriate treatment which may prevent the fatal consequences.

Obviously, an analysis of the 3D signal should give more information. In particular, compared to a 1D analysis, it should allow an easier detection of small peaks and a better estimation of some inter-peak durations.

To illustrate the method, we work with 1500 data points, each comprising 3 coordinates, which were kindly provided by Dr. Laurence Mangin; these represent 3 seconds' VCG measurements sampled 500 times per second.

As we can see on Figure 1, the VCG has three main loops with diameters $PQ$, $RS$, $TU$ corresponding after projection to the usual peaks of the ECG ($P_X$, $Q_X$,$R_X$, $S_X$, $T_X$, $U_X$). For example, it may be easier to detect $P$, $Q$, $T$ in the VCG than in the ECG allowing easier estimation of the duration of the QT interval (ventricular depolarization-repolarization).

*Approximation of the 3D-signal*: It was found convenient to subtract a constant from the data, to give it mean value zero, in order to avoid the behaviour of $\hat{f}$ being dominated by its DC value (the value at 0), but this is not essential. Also the Gram–Schmidt process was used to pre-process the data and make the three channels orthonormal.

Taking a bandwidth of $b = 200$ and $\lambda = 1$ (respectively $\lambda = 0.8$), the approximation produced by 100 iterations of the vectorial (respectively vectorial relaxed) MP gives a relative vectorial $L_2$-error norm equal to 0.009 (respectively 0.003) as the approximation produced using the scalar spline algorithm (available in scilab or matlab) on each coordinate on 100 equidistant points among the full data gives a relative error equal to 0.061. Note that if we were to approximate each coordinate by 100 iterations of the scalar MP, then in general we would arrive at a much more complex model, since we would not be making use of the vectorial structure of the signal.

**Acknowledgments**

The authors would like to thank Claire Médigue for her helpful suggestions on the QT-problem. They also gratefully acknowledge useful discussions with Julie Bestel and Jean Clairambault.

## References

[1] M. Akay, E.J.H. Mulder, Investigating the effect of maternal alcohol intake on human fetal breathing rate using adaptive time-frequency analysis-methods, *Early Human Develop.*, **46** (1996), 153-164.

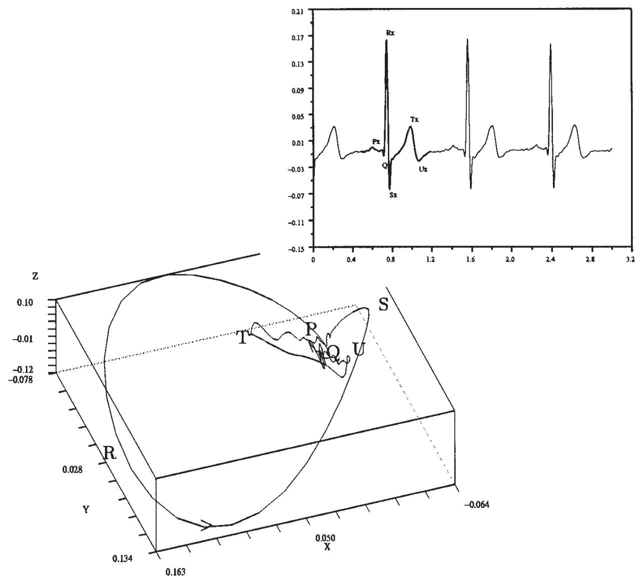


Figure 1: A one beat VCG and a scalar projection (ECG)

[2] R.D. Berger, S. Akselrod, D. Gordon, R.J. Cohen, An efficient algorithm for spectral analysis of heart rate variability, *IEEE Trans. Biomed. Eng.*, **33** (1986), 900-904.

[3] C. Bonnet, J.R. Partington, M. Sorine, A vectorial matching pursuit algorithm applied to the analysis of vecto-cardiogram data, *INRIA Research Report* (August 2002).

[4] I. Daskalov, I. Christov, Improvement of resolution in measurement of electrocardiogram RR intervals by interpolation, *Med. Eng. Phys.*, **19** (1997), 375-379.

[5] I. Daskalov, I. Christov, Automatic detection of the electrocardiogram T-wave end, *Med. Biol. Eng. Comput.*, **37** (1999), 348-353.

[6] G. Davis, S. Mallat, M. Avellaneda, Adaptive greedy approximations, *Constr. Approx.*, **13** (1997), 57-98.

[7] R.W. DeBoer, J.M. Karemaker, J. Strackee, Comparing spectra of a series of point events particularly for heart rate variability data, *IEEE Trans. Biomed. Eng.*, **31** (1984), 384-387.

[8] N.F. Dudley Ward, J.R. Partington, Rational wavelet decompositions of transfer functions in Hardy-Sobolev classes, *Math. of Control, Sig. and Sys.*, **8** (1995), 257-278.

[9] P.J. Durka, K.J. Blinowska, Analysis of EEG transients by means of matching pursuit, *Ann. Biomed. Eng.*, **23** (1995), 608-611.

[10] M.J. Friedman, C.C. Mull, G.Q. Sharieff, N.Tsarouhas, Prolonged QT syndrome in children: an uncommon but potentially fatal entity, *J Emerg. Med.*, **24**, No. 2 (2003), 173-179.

[11] J.R. Higgins, *Sampling Theory in Fourier Analysis and Signal Analysis*, Oxford University Press (1996).

[12] R. Lancellotti, J.B. Mipinda, L.A. Pierard, Electrocardiographic changes during dobutamine stress testing in patients with recent myocardial infarction: relation with residual infarct artery stenosis and contractile recovery, *Acta Cardiol.*, **59**, No. 1 (2004), 11-16.

[13] S.G. Mallat, Z.F. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.*, **41** (1993), 3397-3415.

[14] A.V. Panfilov, A.V. Holden, Ed-s., *Computational Biology of the Heart*, John Wiley (1996).

[15] J.R. Partington, *Interpolation, Identification and Sampling*, Oxford University Press (1997).

[16] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition, In: *Proc. 27-th Asilomar Conference on Signals, Systems, and Computers* (1993).

[17] B.D. Rao, K. Kreutz-Delgado, Basis selection in the presence of noise, In: *Proc. 32-nd Asilomar Conference on Signals, Systems and Computers*, **1** (1998), 752-756.