REGULARIZED KERNEL-BASED WIENER FILTERING. APPLICATION TO MAGNETOENCEPHALOGRAPHIC SIGNALS DENOISING

Ibtissam CONSTANTIN[(1)], *Cédric RICHARD*[(2)], *Régis LENGELLE*[(2)], *Laurent SOUFFLET*[(1)]

[(1)]Formation et recherche en neurosciences appliquées à la psychiatrie (FORENAP)
Centre Hospitalier de Rouffach, 27 rue du 4. RSM, BP 29, 68250 Rouffach - France
[(2)]Institut des Sciences et Technologies de l'Information de Troyes (ISTIT)
Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060, 10010 Troyes cedex - France
ibtissam.constantin@forenap.asso.fr

ABSTRACT

In this paper we proceeded to take up a new approach of nonlinear Wiener filtering. This approach is based on the theory of reproducing kernel Hilbert spaces (RKHS). By means of the well-known "kernel trick", the arithmetic operations are carried out in the initial space. We show that the solution is given by solving a linear system which may be ill-conditioned. To find a solution for such problem, we resorted to kernel principal component analysis (KPCA) method to perform dimensionality reduction in RKHS. A new reduced-rank Wiener filter based on KPCA is thus elaborated. It is applied on magnetoencephalographic (MEG) data for cardiac artifacts extraction.

1. INTRODUCTION

We consider a *filter*, every material or programmed structure applied to a quantity of interest in order to extract significant information in the meaning of a given criterion. These data may for example come from noisy sensing devices, or also be issued from communication channels subject to perturbations. The canonical form of the filtering problem is proposed in Fig. 1. It comprehends an input $x(n)$ and a desired output $d(n)$, supposed to be centered and real without loss of generality. We note $e(n) = d(n) - y(n)$ the committed error. The requirement is the selection of a model w and the implementation of an operational technique enabling the determination of its parameters by the optimization of a performance criterion. Wiener theory is applied to jointly wide-sense stationary processes and consists in finding the linear structure [1]

$$y(n) = \sum_{i=0}^{N-1} w_i \, x(n-i) \tag{1}$$

minimizing the variance of the error. By introducing the following vectorial notations

$$\boldsymbol{x}_n = [x(n), \ldots, x(n-N+1)]^t$$
$$\boldsymbol{w} = [w_0, \ldots, w_{N-1}]^t,$$

the solution of the problem is obtained by solving the equations $\mathcal{E}\{\boldsymbol{x}_n(d(n) - \boldsymbol{w}^t\boldsymbol{x}_n)\} = 0$, where $\mathcal{E}\{.\}$ denotes the mathematical expectation. The previous expression may be written in the compact matrix form

$$\boldsymbol{R}_{\boldsymbol{x}} \, \boldsymbol{w} = \boldsymbol{R}_{\boldsymbol{x}d} \tag{2}$$

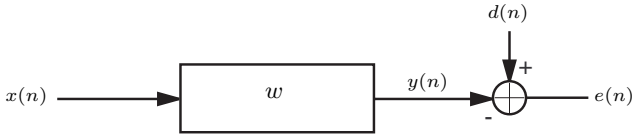

Fig. 1. Block diagram and notations.

with $\boldsymbol{R}_{\boldsymbol{x}}$ and $\boldsymbol{R}_{\boldsymbol{x}d}$ designating $\mathcal{E}\{\boldsymbol{x}_n \, \boldsymbol{x}_n^t\}$ and $\mathcal{E}\{d(n) \, \boldsymbol{x}_n\}$.

We must have noticed that the linear character of the filter facilitates its elaboration, to the detriment of its capacity to give satisfactory solution to every problem. To overcome such limit, while keeping a comparable structure to (1), we can map the observation $\boldsymbol{x}_n$ in a high-dimensional feature space by means of a nonlinear application $\phi(\cdot)$ and then consider the filter $y(n) = \boldsymbol{w}^t\phi(\boldsymbol{x}_n)$. As previously, $\boldsymbol{w}$ can be determined by solving Wiener-Hopf equations defined from

$$\mathcal{E}\{\phi(\boldsymbol{x}_n)(d(n) - \boldsymbol{w}^t\phi(\boldsymbol{x}_n))\} = 0. \tag{3}$$

For the elaboration of a 2nd order polynomial filter for example, this basic concept advocates the implementation of a linear filter operating on the observation $\phi(\boldsymbol{x}_n)$, constituted of the components of $\boldsymbol{x}_n$ and their 2nd order products. The number of parameters to be estimated, equal to $N(N+3)/2$ and already prohibitive with regard to the relative simplicity of the filter, suggests practical difficulties when such strategy is adopted without precautions. In the field of pattern recognition, several results on reproducing kernel Hilbert spaces (RKHS) made however analogous practices possible. By authorizing the synthesis of generalized linear structures without ever explicitly evaluating the map $\phi(\boldsymbol{x}_n)$, these findings reflected new perspectives within the framework of kernel methods, particularly with support vector machines [2], for classification and regression problems.

The present paper studies the problem of nonlinear Wiener filtering in RKHS. We show that the solution is provided by the resolution of a linear system that may be ill-conditioned. In order to overcome such difficulty, we had recourse to kernel principal component analysis (KPCA) method [3] which effects a dimensionality reduction in RKHS. KPCA is a nonlinear extension of principal component analysis (PCA) [4]. It has been successfully applied in various fields, for instance in the context of object recognition and text categorization. It is used in this paper, in a filtering

context, to derive a reduced-rank Wiener filter. The efficiency of this algorithm is demonstrated on the problem of cardiac artifacts extraction from magnetoencephalographic (MEG) data. However, before getting on with these themes, some necessary prerequisites on the theory of RKHS are to be firstly outlined.

2. RKHS AND MERCER'S CONDITION

Let $\mathcal{H}$ be a reproducing kernel Hilbert space consisting of mappings ψ from a signal space $\mathcal{X}$ to $\mathbb{R}$ and let $< \cdot, \cdot >_{\mathcal{H}}$ denote the inner product defined on $\mathcal{H}$. As the Riesz representation theorem states, there is a unique function $\kappa(\cdot, \boldsymbol{y})$ of $\mathcal{H}$ which verifies the following reproducing property:

$$\psi(\boldsymbol{y}) = <\psi\,, \kappa(\cdot, \boldsymbol{y})>_{\mathcal{H}}, \qquad \forall \psi \in \mathcal{H}, \tag{4}$$

for every fixed $\boldsymbol{y} \in \mathcal{X}$. A proof of this may be found in [5]. Here $\kappa(\cdot, \boldsymbol{y})$ is the representer of evaluation at $\boldsymbol{y}$ and $\kappa(\cdot, \cdot)$ is the reproducing kernel associated with $\mathcal{H}$. In particular $\{\kappa(\cdot, \boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$ spans $\mathcal{H}$ and the inner product $< \cdot, \cdot >_{\mathcal{H}}$ has just to be defined on it. Denoting the function $\kappa(\cdot, \boldsymbol{x})$ by $\Phi(\boldsymbol{x})$, equation (4) implies

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = <\Phi(\boldsymbol{x})\,, \Phi(\boldsymbol{y})>_{\mathcal{H}}, \tag{5}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$. The kernel κ then evaluates the inner product of all the pairs of elements of $\mathcal{X}$ mapped to $\mathcal{H}$, without any explicit knowledge of either Φ or $\mathcal{H}$. The key idea of the kernel technique used in this paper, commonly known as the "kernel trick", is to choose a kernel κ rather than a mapping Φ for designing joint representations in signal analysis. Of course, not every function κ can serve as a kernel. According to the Hilbert-Schmidt theory [6], any continuous symmetric function κ can be expanded as follows:

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=0}^{\infty} \lambda_i\, \psi_i(\boldsymbol{x})\, \psi_i(\boldsymbol{y}), \tag{6}$$

where λ_i and ψ_i are eigenvalues and eigenfunctions given by

$$\int \kappa(\boldsymbol{x}, \boldsymbol{y})\, \psi_i(\boldsymbol{x})\, d\boldsymbol{x} = \lambda_i\, \psi_i(\boldsymbol{y}). \tag{7}$$

A sufficient condition to ensure that κ is an inner product in a Hilbert space $\mathcal{H}$ is that all the λ_i's in (6) are positive. According to Mercer's theorem [7], this condition is achieved if, and only if,

$$\iint \kappa(\boldsymbol{x}, \boldsymbol{y})\, f(\boldsymbol{x})\, f(\boldsymbol{y})\, d\boldsymbol{x}\, d\boldsymbol{y} \geq 0 \tag{8}$$

for all f fulfilling $\int f(\boldsymbol{x})^2\, d\boldsymbol{x} < \infty$. From (6), it is straightforward to construct a map Φ into a potentially infinite-dimensional space which satisfies (5). For instance we may use $\Phi(\boldsymbol{x}) = [\sqrt{\lambda_0}\psi_0(\boldsymbol{x}), \sqrt{\lambda_1}\psi_1(\boldsymbol{x}), \ldots]^t$. In the following, we give some classical kernel examples, a more complete list may be viewed in [2, 8].

Polynomial kernels: The polynomial kernel $\kappa(\boldsymbol{x}, \boldsymbol{y}) = (\beta_0 + \boldsymbol{x} \cdot \boldsymbol{y})^q$, with $\beta_0 \geq 0$ and $q \in \mathbb{N}^*$, verifies Mercer's condition. This induces that it gives the inner product of the elements $\boldsymbol{x}$ and $\boldsymbol{y}$ mapped by a function $\Phi(\cdot)$. It can be shown that the components of $\Phi(\boldsymbol{x})$ are monomials of degree less or equal to q, constituted of the components of $\boldsymbol{x}$. Because they are function of the inner product $\boldsymbol{x} \cdot \boldsymbol{y}$ of the observations, such kernels are said to be projective.

Radial kernels: Radial kernels depend on $\|\boldsymbol{x} - \boldsymbol{y}\|$. They have received significant attention in the statistical and machine learning communities [2]. We count among them the gaussian kernel defined by $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\|\boldsymbol{x} - \boldsymbol{y}\|^2/\beta_0\right)$, where β_0 is the kernel bandwidth. This kernel is characterized by a continuum of eigenvalues which means that the components of $\Phi(\boldsymbol{x})$ are not in limited number as for the polynomial kernel. Finally, we can also mention in the family of radial kernels the Laplace kernel $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\|\boldsymbol{x} - \boldsymbol{y}\|/\beta_0\right)$.

Sigmoid kernels: We can construct a radial basis function network by selecting the sigmoid kernel $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \tanh(\alpha_0\, (\boldsymbol{x} \cdot \boldsymbol{y}) + \beta_0)$. Unlike the above-mentioned polynomial and radial kernels, the reproducing property of κ depends on the selected parameters α_0 and β_0. If they are not carefully chosen, this kernel surely does not fulfill Mercer's condition [3].

3. WIENER FILTERING IN RKHS

Let $\mathcal{H}$ be a reproducing kernel Hilbert space defined by a kernel κ. We note $\Phi(\boldsymbol{x})$ the mapping function from $\mathcal{X}$ to $\mathcal{H}$. We seek for a function ψ of $\mathcal{H}$ minimizing the variance of the error between the desired output $d(n)$ and the filter output, presently defined by

$$y(n) = <\psi\,, \Phi(\boldsymbol{x}_n)>_{\mathcal{H}}\,. \tag{9}$$

Since it belongs to $\mathcal{H}$, the function ψ can be written in the form

$$\psi = \sum_i a_i\, \Phi(\boldsymbol{x}_i), \quad a_i \in \mathbb{R}, \tag{10}$$

where $\mathcal{X} \triangleq \{\boldsymbol{x}_i\}_i$ represents the filter inputs. Using (10), expression (9) can be easily expressed as a linear combination of the reproducing kernels $\kappa(\boldsymbol{x}_n, \boldsymbol{x}_i)$, which enables directly the application of Wiener theory to determine the unknown parameters a_i. However, the serious difficulty found in formulation (10) is that this latter is based on an unlimited sum of terms as well as an explicit knowledge of the space of observations $\mathcal{X}$. This problem may be bypassed by restricting our search for a solution ψ to a subspace $\mathcal{H}_M \subset \mathcal{H}$ of dimension M, thus spanned by the functions $\{\Phi(\boldsymbol{x}_i)\}_{0 \leq i \leq M-1}$. The elements $\boldsymbol{x}_i$ are M observations of the input $\boldsymbol{x}$, playing in consequence the role of training basis. In these conditions we have

$$\psi = \sum_{i=0}^{M-1} a_i\, \Phi(\boldsymbol{x}_i). \tag{11}$$

Combining (9) and (11), we get the dual formulation $y(n) = \sum_{i=0}^{M-1} a_i\, \kappa(\boldsymbol{x}_n, \boldsymbol{x}_i)$, where the parameters a_i are to be determined in order to minimize the variance of the error $e(n)$. We notice that this expression is not explicitly based on the analytical expression of $\Phi(\cdot)$. It is implicitly defined by the choice of a reproducing kernel κ. For example, the use of the polynomial kernel $\kappa(\boldsymbol{x}, \boldsymbol{y}) = (1 + \boldsymbol{x} \cdot \boldsymbol{y})^q$ leads to a Volterra filter of degree q, whose dual formulation is provided by $y(n) = \sum_{i=0}^{M-1} a_i\, (1 + \boldsymbol{x}_n \cdot \boldsymbol{x}_i)^q$. This latter does not need the evaluation of $\Phi(\boldsymbol{x}_n)$. Remember that the components of $\Phi(\boldsymbol{x}_n)$ are products of degree less or equal to q composed of the components of $\boldsymbol{x}_n$. Finally, the solution of the problem is obtained by solving an analogous system to (2) in which $\tilde{\boldsymbol{\kappa}}(\boldsymbol{x}_n) = [\kappa(\boldsymbol{x}_n, \boldsymbol{x}_0), \ldots, \kappa(\boldsymbol{x}_n, \boldsymbol{x}_{M-1})]^t$ is substituted to $\boldsymbol{x}_n$. It is noted

$$\boldsymbol{R}_{\tilde{\boldsymbol{\kappa}}}\, \boldsymbol{a} = \boldsymbol{R}_{\tilde{\boldsymbol{\kappa}} d}. \tag{12}$$

We should notice that the number of parameters to be estimated is independent from the complexity of the filter, which is characterized by the polynomial degree in the particular case

of Volterra filter. The advantage of linear-in-the parameters estimation remains. In practice, the correlation matrices $\boldsymbol{R}_{\tilde{\boldsymbol{\kappa}}}$ and $\boldsymbol{R}_{\tilde{\boldsymbol{\kappa}}d}$ are unknown. They must be replaced by their estimates, which can be computed easily from the vectors $\{\tilde{\boldsymbol{\kappa}}(\boldsymbol{x}_i)\}_{0\leq i\leq M-1}$ and the corresponding desired responses $\{d(i)\}_{0\leq i\leq M-1}$ by the relations: $\widehat{\boldsymbol{R}}_{\tilde{\boldsymbol{\kappa}}} = \frac{1}{M}\sum_{i=0}^{M-1}\tilde{\boldsymbol{\kappa}}(\boldsymbol{x}_i)\tilde{\boldsymbol{\kappa}}^t(\boldsymbol{x}_i) = \frac{1}{M}\boldsymbol{K}^t\boldsymbol{K}$ and $\boldsymbol{R}_{\tilde{\boldsymbol{\kappa}}d} = \frac{1}{M}\sum_{i=0}^{M-1} d(i)\tilde{\boldsymbol{\kappa}}(\boldsymbol{x}_i) = \frac{1}{M}\boldsymbol{K}^t\boldsymbol{d}$, where $\boldsymbol{K} = [\tilde{\boldsymbol{\kappa}}(\boldsymbol{x}_0),\ldots,\tilde{\boldsymbol{\kappa}}(\boldsymbol{x}_{M-1})]^t$ and $\boldsymbol{d} = [d(0),\ldots,d(M-1)]^t$. Consequently we have

$$\boldsymbol{a} = (\boldsymbol{K}^t\boldsymbol{K})^{-1}\boldsymbol{K}^t\boldsymbol{d} = \boldsymbol{K}^{-1}\boldsymbol{d}. \tag{13}$$

The last equality results from the fact that $\boldsymbol{K}$ is symmetric.

4. REDUCED-RANK WIENER FILTER

The resolution of the previous equation requires a matrix inversion. The solution is numerically unstable if the matrix $\boldsymbol{K}$ is ill-conditioned. In this section, we describe a new algorithm for Wiener filtering in order to tackle such difficulty. The fundamental idea consists of two steps. First we project the function $\Phi(\boldsymbol{x}_n)$ in a reduced-dimensional space and then we perform Wiener filtering in the reduced space. An efficient method for dimensionality reduction in RKHS is KPCA [3]. This method, which is a nonlinear extension of PCA, constructs an orthogonal projection of $\Phi(\boldsymbol{x}_n)$ that preserves most of its important characteristics (measured by the variance). This projection is determined from the eigenvectors of the correlation matrix in $\mathcal{H}$, estimated by

$$\widehat{\boldsymbol{R}}_{\Phi} = \frac{1}{M}\sum_{i=0}^{M-1}\Phi(\boldsymbol{x_i})\Phi^t(\boldsymbol{x_i}) = \frac{1}{M}\boldsymbol{\Phi}^t\boldsymbol{\Phi}, \tag{14}$$

where $\boldsymbol{\Phi} = [\Phi(\boldsymbol{x}_0),\ldots,\Phi(\boldsymbol{x}_{M-1})]^t$. We assume that $\sum_{i=0}^{M-1}\Phi(\boldsymbol{x_i}) = 0$ (A centering method in $\mathcal{H}$ can be found in [3]). Therefore, the problem consists in finding the eigenvalues $\lambda > 0$ and the eigenvectors $\boldsymbol{v}\in\mathcal{H}$ of $\widehat{\boldsymbol{R}}_{\Phi}$ satisfying

$$\lambda\boldsymbol{v} = \widehat{\boldsymbol{R}}_{\Phi}\boldsymbol{v}. \tag{15}$$

Replacing $\widehat{\boldsymbol{R}}_{\Phi}$ in (15) with its expression in (14) leads to

$$\lambda\boldsymbol{v} = \frac{1}{M}\sum_{i=0}^{M-1} <\Phi(\boldsymbol{x_i}),\boldsymbol{v}>_{\mathcal{H}}\Phi(\boldsymbol{x_i}). \tag{16}$$

As can be seen from (16), all eigenvectors with nonzero eigenvalue must be in the span of $\{\Phi(\boldsymbol{x_i})\}_{0\leq i\leq M-1}$. This can be written as

$$\boldsymbol{v} = \sum_{i=0}^{M-1} a_i\Phi(\boldsymbol{x_i}) = \boldsymbol{\Phi}^t\boldsymbol{a}. \tag{17}$$

Using this definition of $\boldsymbol{v}$, expression (15) translates to

$$\lambda\boldsymbol{\Phi}^t\boldsymbol{a} = \frac{1}{M}\boldsymbol{\Phi}^t\boldsymbol{\Phi}\boldsymbol{\Phi}^t\boldsymbol{a}. \tag{18}$$

Premultiplying with $(\boldsymbol{\Phi}\boldsymbol{\Phi}^t)^{-1}\boldsymbol{\Phi}$, equation (18) reads

$$M\lambda(\boldsymbol{\Phi}\boldsymbol{\Phi}^t)^{-1}\boldsymbol{\Phi}\boldsymbol{\Phi}^t\boldsymbol{a} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^t)^{-1}\boldsymbol{\Phi}\boldsymbol{\Phi}^t\boldsymbol{\Phi}\boldsymbol{\Phi}^t\boldsymbol{a}, \tag{19}$$

which is reduced to

$$M\lambda\boldsymbol{a} = \boldsymbol{K}\boldsymbol{a}. \tag{20}$$

We note $\boldsymbol{v}_i$ the i$^{\text{th}}$ eigenvector of $\widehat{\boldsymbol{R}}_{\Phi}$ corresponding to the nonzero eigenvalue λ_i and $\boldsymbol{a}_i$ the associated eigenvector of $\boldsymbol{K}$. To ensure that the eigenvectors $\boldsymbol{v}_i$ have unit norm in the feature space, $\boldsymbol{a}_i$ should be divided by $\sqrt{M\lambda_i}$ (for details see [3]).

Let $\beta(\boldsymbol{x}_n)$ be the projection of $\Phi(\boldsymbol{x}_n)$ on the whole set of eigenvectors of $\widehat{\boldsymbol{R}}_{\Phi}$. Denoting by $\boldsymbol{V}$ and $\boldsymbol{A}$ the matrices whose columns are the eigenvectors of $\widehat{\boldsymbol{R}}_{\Phi}$ and $\boldsymbol{K}$ respectively, we have

$$\beta(\boldsymbol{x}_n) = \boldsymbol{V}^t\Phi(\boldsymbol{x_n}) = \boldsymbol{A}^t\tilde{\boldsymbol{\kappa}}(\boldsymbol{x}_n). \tag{21}$$

The last equation is derived using (17). We seek presently to determine the Wiener filter operating on $\beta(\boldsymbol{x}_n)$. The output of the filter is defined by $y(n) = \beta(\boldsymbol{x}_n)^t\boldsymbol{w}$, where $\boldsymbol{w}$ is the unknown parameter vector. According to (13) , $\boldsymbol{w}$ is provided by

$$\boldsymbol{w} = (\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t\boldsymbol{d}, \tag{22}$$

where $\boldsymbol{B} = [\beta(\boldsymbol{x}_0),\ldots,\beta(\boldsymbol{x}_{M-1})]^t$. We notice that $\boldsymbol{B}^t\boldsymbol{B}$ is a diagonal matrix. Indeed we have

$$\boldsymbol{B}^t\boldsymbol{B} = \boldsymbol{V}^t\boldsymbol{\Phi}^t\boldsymbol{\Phi}\boldsymbol{V} = \boldsymbol{\Lambda}, \tag{23}$$

where $\boldsymbol{\Lambda} = diag(M\lambda_0,\ldots,M\lambda_{M-1})$. If we keep the whole set of eigenvectors, the matrices $\boldsymbol{K}$ and $\boldsymbol{\Lambda}$ will have the same rank. However, to avoid numerical problems, eigenvectors associated to the smallest eigenvalues $M\lambda_i$ of $\boldsymbol{K}$ should be discarded.

5. APPLICATION TO MEG SIGNALS DENOISING

We validated the previous approach on the problem of cardiac artifacts extraction from MEG data. MEG recordings are usually contaminated by external undesired interferences. In particular cardiac artifacts, generated by the heart activity, may present serious problems for MEG data analysis. They can be several times stronger in magnitude than MEG signals and may severely interfere with relevant information extraction. MEG were recorded from the temporal area and the electrocardiographic (ECG) signal was measured simultaneously. The sampling frequency was 256 Hz for both signals. ECG was used as a reference signal (input of the kernel reduced-rank Wiener filter). The MEG signal (corrupted by ECG) was used as the desired output, so the residue (whose power is minimized) is the denoised MEG signal. Two implicit data transformations were considered by selecting the gaussian and linear kernels. The latter serves here as a reference by implementing a reduced-rank linear Wiener filter. In order to measure the signal to noise ratio enhancement by direct calculation of the denoised signal power, MEG data were normalized to unit variance. Three databases of 3000 samples were constituted for training, cross validation and testing, respectively. The optimum filter parameters, i.e., the dimension of the input vector N, the kernel bandwidth β_0 and the number of eigenvectors P, were determined so as to minimize the residue power on the cross-validation set. The filter performance was evaluated on the testing set. On Fig. 2 we reported the results obtained on the testing set for both gaussian and linear kernels. The gaussian kernel induces better performance than the linear kernel (the residue power equals 0.887 with $\beta_0 = 36.98$, $N = 16$ and $P = 12$ to be compared with 0.922 with the linear kernel where $N = 15$ and $P = 14$). We depicted in Fig. 3 the residue power as a function of the number of eigenvectors.

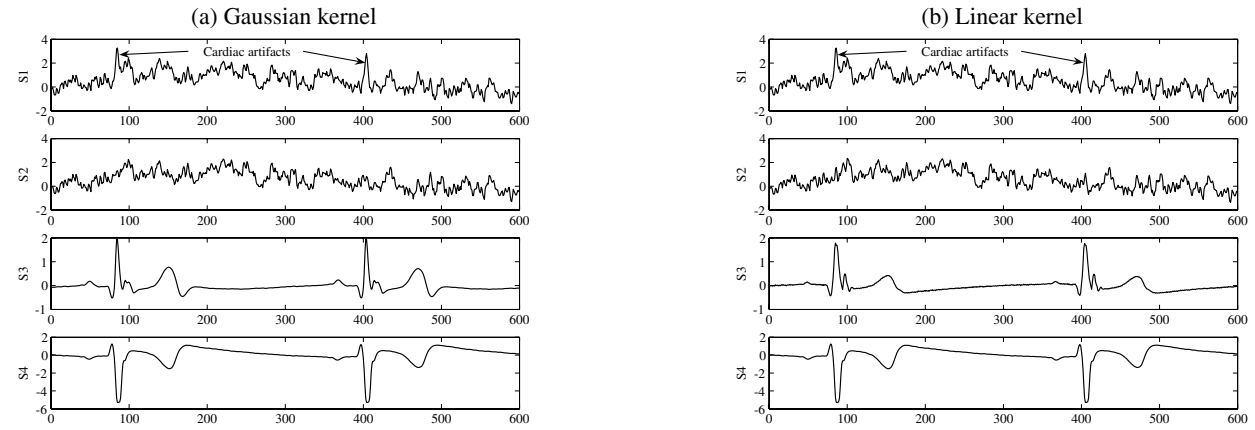

Fig. 2. Comparison of the results obtained by kernel reduced-rank Wiener filter with gaussian kernel (a) and linear kernel (b). S1: Corrupted MEG. S2: Denoised MEG. S3: Estimated ECG contribution in MEG signal. S4: ECG reference signal.

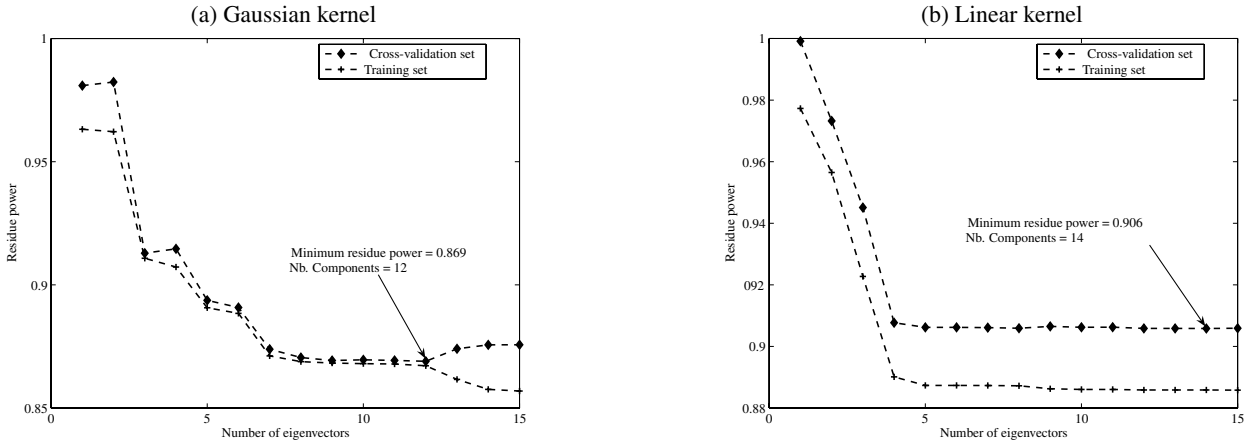

Fig. 3. Residue power as a function of the number of eigenvectors for both gaussian (left curve) and linear (right curve) kernels.

6. CONCLUSION

We highlighted in this paper an efficient approach of nonlinear Wiener filtering that relies on the theory of RKHS. Similarly to the classical Wiener filter, this method leads to the resolution of a linear system that may suffer from ill-conditioning. To solve this problem, we described a new algorithm for Wiener filtering exploiting KPCA method to effectuate a dimensionality reduction in RKHS. The efficiency of this algorithm has been proved on MEG data. To complete this work, we should extend it to nonlinear adaptive filtering.

7. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, New Jersey, 2002.

[2] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[3] B. Schölkopf, C. J. C. Burges, and A. J. Smola, *Advances in Kernel Methods — Support Vector Learning*, MIT Press, Cambridge, MA, 1999.

[4] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.

[5] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, 1950.

[6] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Interscience, New York, 1953.

[7] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philos. Trans. Roy. Soc. London*, vol. A 209, pp. 415–446, 1909.

[8] M. G. Genton, "Classes of kernels for machine learning: A statistics perspective," *Journal of Machine Learning Research*, vol. 2, pp. 299–312, May 2002.