ESTIMATION OF NOISY SIGNALS BASED ON LOCAL TRANSFORMS

*Pavel Kisilev[+], Yehoshua Y. Zeevi[+], Hillel Pratt**

[+]Dept. of Electrical Engineering and *Evoked Potentials Laboratory
Technion, Israel Institute of Technology
Haifa, 32000, Israel

ABSTRACT

An algorithm for estimation of unknown signals corrupted by colored noise is proposed. The local Karhunen-Loeve (LKL) basis, derived from the local autocorrelation function of the noisy signal, is used for optimal signal representation (in minimum mean squared error - MMSE - sense). The vector space of the noisy signal is decomposed by the LKL transform into the corresponding complementary orthogonal subspaces, i.e. the signal-plus-noise and the noise only. The desired signal is estimated from the signal-plus-noise subspace by modifying the corresponding LKL components with a Wiener gain function.

1. INTRODUCTION

The problem of estimation of signals corrupted by colored noise arises in processing of event related potentials (ERP) - the signals generated by detectable events. One particular case of ERP is the evoked potential (EP) - the brain response to external stimuli. The main obstacles encountered in extracting a single-trial EP from ongoing EEG activity are a low signal-to-noise ratio (SNR) and a high variability of single EP from trial to trial and from subject to subject.

Methods based on template matching for estimation of a single EP (e.g. [1]) smooth out features of a particular single EP. Methods based on modeling of a single EP as a sum of damped sinusoids (e.g. [2]) require detailed specification of the linear model and the solution of a hard parameter estimation problem, and are not necessarily optimal in any sense. We therefore propose an alternative method, based on the LKL transform of the measured signal, and on modification of its components by a Wiener gain function. This procedure preserves local features of the signal, it is optimal in MMSE sense and is completely driven by data.

2. ESTIMATION METHOD

2.1 Basic assumptions

We assume that the noise is: 1) additive, 2) stationary during the pre- and post-stimulus measurements over a period of one second (i.e. short-term stationary) and 3) uncorrelated with the signal. The noise process can be described by an autoregressive (AR) model of appropriate order. A noisy signal is first preconditioned by a whitening filter. This filter is derived from the AR model of the prestimulus EEG signal.

2.2 Signal subspace approach

The signal subspace approach, popular in the array processing problems, was used in [3] for speech enhancement. We use this technique as a framework for our estimation method and combine it with a local transform ideas.

Let y=s+n be a distorted noisy signal, where s is a distorted noiseless signal, and n is a white Gaussian noise. The covariance matrix of y is given by

$$R_y = R_s + R_n , \tag{1}$$

where $R_n = \sigma_n^2 I$, and σ_n^2 is the noise variance. The vector space of the preconditioned noisy signals is decomposed by the KL transform into the corresponding complementary orthogonal subspaces, i.e. the *signal subspace* (that contains both the signal and the noise components), and the *noise subspace* (that contains the noise components only) [4].

Let the eigendecomposition of the covariance matrix R_y be

$$R_y = V\Lambda_y V^T , \tag{2}$$

where V is an orthonormal matrix of eigenvectors of R_y, Λ_y is a diagonal matrix of eigenvalues of R_y, $\Lambda_y = diag[\lambda_y(1),\ldots,\lambda_y(K),\lambda_y(K+1),\ldots,\lambda_y(M)]$, and K $(K \le M)$ is the signal subspace dimension which is estimated in advance. The estimation of the signal subspace dimension can be performed by the minimum description length (MDL) approach of Rissanen [5]. For the problem under consideration, this criterion is given by:

$$MDL(r) = \log\left(\frac{1}{K} y^T \left(\sum_{i=r+1}^{M} v_i v_i^T\right) y\right)^{M/2} + \frac{1}{2} r \log M , \tag{3}$$

where v_i are the columns of the matrix V. The signal subspace dimension, K, is determined as the value of $r \in \{1,2,\ldots, M\}$ for which the MDL is minimized.

From (1) and (2) we conclude that

$$\lambda_y(k) = \lambda_s(k) + \sigma_n^2, \quad k = 1,\ldots, K, \tag{4}$$

where $\lambda_s(k), k = 1,\ldots,K$ are the eigenvalues of the covariance matrix R_s (which is of rank K). The eigenvalues $\lambda_y(K+1),\ldots,\lambda_y(M)$ are all equal to σ_n^2. Thus, the signal s can be estimated from only the first K eigenvectors of V. In particular, the desired signal is estimated with reference to the signal subspace by modifying the corresponding KL components with a gain function W:

$$\hat{s} = VWV^T y, \tag{5}$$

where V is a $K \times M$ matrix of K eigenvectors, corresponding to the first K eigenvalues of Λ_y, and W is a $K \times K$ diagonal matrix of weights that represents the Wiener gain function. These weights are given by:

$$w_k = (\lambda_y(k) - \sigma_n^2) / \lambda_y(k), k = 1,\ldots, K. \tag{6}$$

The above filtering procedure is optimal in the MMSE sense.

Columns of V form an orthonormal basis for signal representation. If the Pseudo-KL (PKL) transform [6] is used instead of the regular KL, (5) takes the form:

$$\hat{s} = VWU^T y, \tag{7}$$

where U is the matrix whose columns form a basis biorthonormal to that of V.

Finally, the signal $\hat{s}$ is filtered by the inverse whitening filter to compensate for a signal distortion caused by the whitening operation.

2.3 The local KL and pseudo-KL transforms

The KL transform is known to be the most efficient coordinate system for representation of signals, under the MMSE and the minimum entropy criteria. Nevertheless, features of signals which are localized in the time-frequency domain are not well represented due to the global nature of the eigenvectors. To overcome this problem, the local transform should be used. To perform a local KL transform for time series representation, we divide the initial interval into shorter intervals of equal length which overlap in time [7]. In this case, time-frequency structure which is similar to the short-time Fourier transform is produced. The KL transform (or the PKL transform) along with the filtering procedure are applied to the signal at each interval separately. To avoid edge effects, the signal at a given interval is multiplied by a window function.

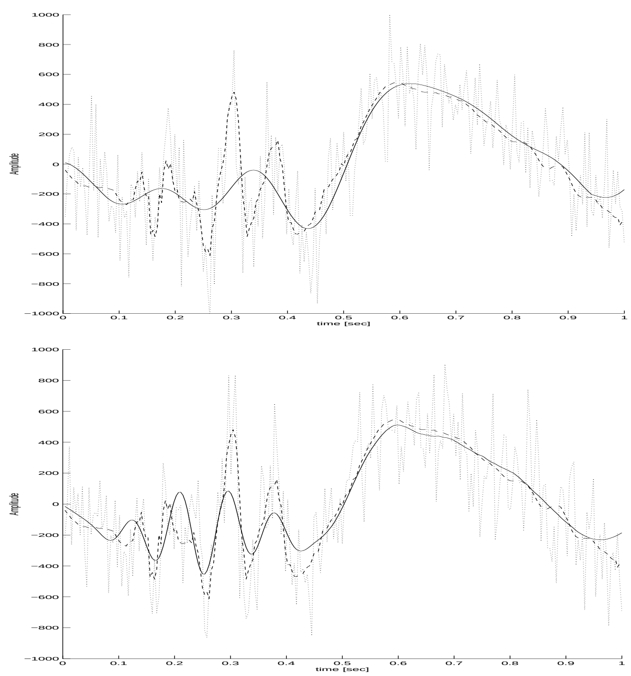

Figure 1. Estimation with the KL (upper) and the LKL (lower) transforms applied to simulated (synthetic) EP signal. The noise-free, noisy (with white Gaussian noise; SNR = 0 dB) and estimated signals are denoted by dashed, dotted and continuous lines, respectively.

If the time-frequency content of the signal is not uniform, then time intervals have to be of varying length. The best basis selection algorithm described in [8] was used for compression and classification purposes and applied to ensemble of signals. This algorithm can be adapted for time series representation and for estimation purposes. First, the original interval is divided into subintervals of equal length $I_1,\ldots,I_K$. Then, starting from the left, two adjacent intervals are examined to see if it is worthwhile (under a particular criterion) to merge them. For instance, if two subintervals I_j, I_{j+1} are deemed worthwhile to merge into the single interval $I_{j,j+1}$, then the next step is to check if it is worthwhile to merge $I_{j,j+1}$ with I_{j+2}. Otherwise, the next step is to check the intervals I_{j+1}, I_{j+2}, and so on.

Let $\lambda_{I_j} = (\lambda_{I_j}(1),\ldots,\lambda_{I_j}(M_j))$ be the set of eigenvalues of the covariance (autocorrelation) matrix of the signal on the interval I_j. Furthermore, let them be arranged in a descending order. The following nonlinear functional can be used as a measure of inefficiency of the subinterval I_j:

$$\mu(\lambda_{I_j}, K) := \sum_{k=1}^{K} (\lambda_{I_j}(k))^p , \qquad (8)$$

where $K \leq Mj$ is the number of features to be used (in our case, the signal subspace dimension), and $0<p<1$. Then, for the intervals I_j, I_{j+1} and $I_{j,j+1}$ we check, if $\mu(\lambda_{I_j}, K) < \mu(\{\lambda_{I_j}, \lambda_{I_{j+1}}\}, K)$. If this inequality holds, then the interval $I_{j,j+1}$ is chosen and $V_{I_j,I_{j+1}}, U_{I_j,I_{j+1}}$ are calculated as global biorthogonal KL bases for $I_{j,j+1}$ (i.e. $V_{I_j,I_{j+1}} = V_{I_{j,j+1}}, U_{I_j,I_{j+1}} = U_{I_{j,j+1}}$). If it does not, then two intervals I_j and I_{j+1} are chosen, $V_{I_j,I_{j+1}} = V_{I_j} \oplus V_{I_{j+1}}$, $U_{I_j,I_{j+1}} = U_{I_j} \oplus U_{I_{j+1}}$, and $\lambda_{I_j,I_{j+1}} = \{\lambda_{I_j}, \lambda_{I_{j+1}}\}$. Note that if the regular KL transform is used instead of the PKL, then $U_j \equiv V_j$. The above algorithm of the best basis selection is similar to the one used in the Wavelet Packets algorithm.

After the LKL basis is constructed, the signal is represented by the transform domain coefficients, and the above estimation procedure is applied to this representation. In this case, the effect of the Wiener filtering procedure is local, since it is applied to the local coefficients.

2.4 Estimation of the local autocorrelation matrix

The autocorrelation (covariance) matrix can be estimated in different ways. The simple autocorrelation estimate is computed from samples $y_j, j=0,\ldots,J-1$ via:

$$\hat{r}_y(m) = \frac{1}{J} \sum_{j=0}^{J-m-1} y(j) y(j+m), \; m = 0,1,\ldots,N-1 . \qquad (9)$$

The total number of samples J, from which the autocorrelation matrix is estimated, is chosen to be smaller than the interval during which the signal is assumed to be stationary. On the other hand, the noisy signal subspace dimension N must be sufficiently large, since the improvement in SNR is proportional to N/K, where K is the noise-free signal subspace dimension.

A slightly better estimate can be achieved by taking a similar approach to that taken by the running autocorrelation method. Suppose that $J=PL$ with $P \in Z^+$ and $L \geq N$. In this case, the sum in (9) can be split into P sums, and the estimate of the autocorrelation matrix is

$$\hat{r}_y(m) = \frac{1}{P} \sum_{k=1}^{P} a_k r'_{y,k}(m), \; m = 0,1,\ldots,N-1 , \qquad (10)$$

where

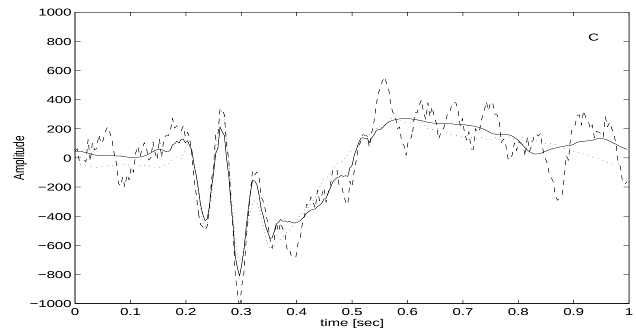

Figure 2. Estimation with the LKL transform applied to simulated (synthetic) EP signal. The noise-free, noisy (with colored Gaussian noise; SNR = 2.5 dB) and estimated signals are denoted by dotted, dashed and continuous lines, respectively.

$$\hat{r}'_{y,k}(m) = \frac{1}{L} \sum_{i=0}^{L-m-1} y(i+(k-1)L) y(i+(k-1)L+m) , \qquad (11)$$

and $\{a_k\}$ are the weights (all being equal to 1 in the simplest case). To perform the weighting operation we use the following formula for calculating a_k:

$$a_k = \begin{cases} b(1-b)^{(P/2-k)}, & k \leq \lfloor P/2 \rfloor \\ 1 & , k = \lceil P/2 \rceil \\ b(1-b)^{(P/2-k-1)}, & k > \lceil P/2 \rceil , \end{cases} \qquad (12)$$

where b is a "forgetting factor" and P is an odd number. Note, that the value $k = \lceil P/2 \rceil$ corresponds to the current frame of the signal, for which the LKLT is calculated.

2.5 Estimation algorithm summary

The estimation procedure (using the orthonormal KLT) can be summarized as follows.

Preconditioning: whiten the noisy signal.

STAGE I: build LKL basis

Step 1: Decompose the global interval I of a preconditioned noisy signal y into subintervals of equal length which overlap in time: $I_1,\ldots,I_N$ (bottom level branches of the binary tree representation)

Step 2: For every I_j above the bottom level and for its two children I_k, I_l , calculate the standard KL bases V_j, V_k, V_l, respectively.

Step 3: Calculate according to (8) the inefficiency measure μ for the subintervals I_j, I_k, I_l.

Step 4: Select the best *local* basis B_j for I_j:

If $\mu(\lambda_j, p) < \mu(\{\lambda_k, \lambda_l\}, p)$,

Then choose the global over I_j PKL basis: $B_j = V_j$, $\{\lambda_j\}$

Else $B_j = V_k \oplus V_l$, $\{\lambda_j\}=\{\lambda_k, \lambda_j\}$

Step 5: Repeat Step 2 - Step 4 for higher level branches of the binary tree

STAGE II: Apply the signal subspace approach, along with the Wiener filtering, to the signal in the LKL domain.

Postconditioning: Apply an inverse whitening filter

3. RESULTS

The algorithm was tested and its performance validated using natural and simulated signals. The real data was sampled from visual EP. The simulated EEG was generated as an AR process of order 6, derived from the AR model of the measured prestimulus EEG signal. The noiseless EP signal was simulated as a stationary deterministic process by a sinusoidal model of order 17, derived from the averaged EP signal. A single-trial EP was simulated as a sum of simulated EP and EEG signals with the desired SNR. 100 such statistically independent single-trials were generated. Figure 1 presents the results of the estimation procedure for a simulated signal, masked by a white Gaussian noise with SNR=0 dB, using the KL and the LKL transforms. The advantage of the LKL over the KL transform for processing signals with time-varying spectrum is evident: the KL transform smoothes out local features, while the LKL does not. Figure 2 presents a typical result of the estimation procedure, applied to a simulated signal masked by a colored Gaussian noise with SNR=2.5 dB, using the LKL transform. A comparison of two estimation procedures is presented in Table 1. The first is based on the wavelet denoising with 'db9' and 'db1' Daubechies wavelet families. The second utilizes the LKL transform along with the Wiener filter. The denoising procedure based on the 'db9' basis gave the best results. The LKL transform along with the Wiener filter gave slightly poorer results than 'db9', but considerably better than the Haar basis. When the Wavelet Packet algorithm was used, the best suited family of wavelets was chosen using a priori knowledge about the noiseless signal. In contrast, no *a priori* knowledge was used in the LKL-transform-based algorithm.

4. DISCUSSION

The proposed method improves the signal-to-noise ratio of a noisy signal without resorting to averaging or requiring any reference signals. The LKL basis provides an additional degree of freedom in representation of signals, as compared to the KL basis. Further relaxation of

Basis	SNR gain, dB	Note
Haar ('db1')	3.2	'worst' basis
Daubechies ('db9')	5.96	best suited basis
LKL (+Wiener)	5.9	no a priori knowledge

Table 1. Comparison of LKL vs. Wavelet Packet efficiency.

constraints can be achieved by using the PKL basis of non-orthogonal functions [6]. The natural extension of the proposed method to the multivariate case should be useful in the analysis of multichannel signals. In this case, spatial distributions of a signal and of noise can impose additional constraints on the derivation of an appropriate spatio-temporal filter. This subject is currently under further investigation.

Although we developed the proposed method primerily for detection of single-trial evoked potentials, the general framework is most suitable for processing a wide variety of signals embedded in colored noise with low SNR and for signals corrupted by white noise with as low as 0 dB SNR.

5. REFERENCES

[1] E.A. Bartnik, K.J. Blinowska and P.J. Durka, "Single evoked potential reconstruction by means of wavelet transform, " Biological Cybernetics, vol. 67, 1992, pp. 175-181.

[2] T. Gansler and M. Hansson, "Estimation of event related potentials," in Proceedings of the 12-th EMBS, vol. 12, 1990, pp. 901-902.

[3] Y. Ephraim, H.L. Van Trees, "A signal subspace approach for speech enhancement," IEEE Trans. Speech, Audio Processing, vol. 3, 1995, pp. 251-266.

[4] L.L Scharf, Statistical Signal Processing: Detection, Estimation, and Time Series Analysis, Addison-Wesley, 1991.

[5] J. Rissanen, "Modeling by shortest data description," Automatica, vol. 14, 1978, pp.465-471.

[6] V. Segalesku, M. Zibulski, Y.Y. Zeevi, "On the role of biorthonormality in representation of random processes," IEEE Trans. Inf. Theory, vol. 42, no.1, 1996, pp. 288-290.

[7] P. Kisilev, Y.Y. Zeevi, H. Prat, "On time-series representation by the local Karhunen-Loeve transform," in preparation.

[8] N. Saito, R. R. Coifman, "The local Karhunen-Loeve bases," in Proc. IEEE-SP Int. TFTS, 1996, pp. 129-132.