ROBUST MATCHED FILTERING IN THE FEATURE SPACE

Ignacio Santamaría[1], *Deniz Erdogmus*[2], *Rati Agrawal*[3] *and Jose C. Principe*[3]

[1]Dept. Ingeniería de Comunicaciones, Universidad de Cantabria, 39005 Santander, Spain,
[2]CSEE Department, Oregon Health and Science University, Portland, OR, USA,
[3]Computational NeuroEngineering Laboratory (CNEL), University of Florida, Gainesville, FL 32611 USA,
[1]e-mail:nacho@gtas.dicom.unican.es, [2]e-mail:derdogmus@ieee.org, [3] e-mail:{ratiag,principe}@cnel.ufl.edu

ABSTRACT

In this paper the problem of detecting a known waveform in noise is solved in a high dimensional transformed (feature) space. The proposed test statistic is the inner product between two hyperplanes constructed using the nonlinearly transformed template and observations, which becomes a simple quadratic form after applying the *kernel trick*. To obtain the optimal projections for the template and the observations we maximize the Fisher discriminant analysis (FDA) criterion in the feature space. Under the white Gaussian noise assumption, closed-form expressions for the means and the variances under each hypothesis are obtained, and an iterative procedure to get the optimal projections is proposed. Interestingly, the analysis of the results shows that the optimal projections preserve the information about the original waveform shape. This can be used to simplify the optimization procedure since one of the projectors can be fixed in advance. Some simulation results indicate that the proposed test statistic achieves the optimal performance of the linear matched filter under Gaussian noise, but shows an increased robustness against impulsive noise distributions.

1. INTRODUCTION

The detection of a known waveform in noise is a fundamental problem with a wide range of applications such as communications, radar and biomedical signal processing [1], [2]. Under the assumption of additive Gaussian noise, the optimal solution is given by the matched filter, which is the linear filter that maximizes the signal-to-noise ratio at its output. However, when the interference is non-Gaussian the optimum detector is, in general, nonlinear and depends on the noise distribution [3]. In addition, if the waveform suffers some nonlinear distortion the matched filter is not optimal anymore. In this situation a test statistic that extracts all the higher order moments, such as the recently proposed quadratic mutual information (QMI) [4], outperforms the linear matched filter.

In this paper we describe a new solution for matched filtering in a feature (transformed) space. The idea of using a nonlinear transformation to a high dimensional feature space where a solution can be found was properly motivated by statistical learning theory [5], and has been successfully applied to a number of applications ranging from face identification, bioinformatics, marketing, data mining and communications [6], [7].

This work was supported by MEC (Ministerio de Educación y Ciencia) under grant PR2004-0364, by MCYT (Ministerio de Ciencia y Tecnología) under grants TIC2001-0751-C04-03 and TEC2004-06451-C05-02/TCM, and by NSF grant ECS-0300340.

In feature space, a discriminant test statistic is formed as the inner product between two hyperplanes constructed using the nonlinearly transformed template and observations, respectively. After applying the *kernel trick* the test statistic turns out to be a simple quadratic form. The optimal linear projections are obtained by maximizing the Fisher discriminant analysis (FDA) criterion [8]. Under the Gaussian noise assumption it is possible to derive closed-form expressions for the means and variances under each hypothesis, then an iterative technique to maximize the kernel FDA (K-FDA) criterion is applied.

Interestingly, the analysis of the results show that the optimal projections obtained through FDA preserve in the feature space the temporal structure of the waveform, which is a crucial information for this problem. An additional advantage of solving this type of problems in the feature space induced by the reproducing Gaussian kernel is an increased robustness against impulse noise.

2. DETECTION OF A KNOWN WAVEFORM IN THE FEATURE SPACE

We consider the problem of detecting a known deterministic signal s_k, corrupted by a zero-mean white additive noise n_k with pdf $f_N(n)$, i.e., we have the following binary hypothesis testing problem [1]

$$\begin{aligned} H_1: \quad r_k &= s_k + n_k, \quad k = 1, \cdots N \\ H_0: \quad r_k &= s_k, \quad k = 1, \cdots N. \end{aligned}$$

If the noise is Gaussian, the optimal linear filter for detection is given by the matched filter $h_k = s_{N-1-k}$, and the corresponding test statistic is

$$T_{MF}(\mathbf{r}) = \mathbf{s}^T \mathbf{r} = \sum_{k=1}^{N} s_k r_k$$

where we have defined $\mathbf{r}$ and $\mathbf{s}$ as column vectors containing the observations and the original waveform, respectively.

If the transmitted signal suffers a nonlinear distortion or the noise distribution is white but not Gaussian, the matched filter is not optimal anymore and its performance is expected to degrade. In these situations, it has been recently shown that a nonlinear criterion based on the Cauchy-Schwartz quadratic mutual information (QMI) between the observations and the template signal, outperforms the matched filter [4]. However, the QMI criterion operates on the template and the observations as if they were i.i.d samples drawn from two different distributions. Therefore, QMI does not take into account the time information conveyed by the waveform template, which is crucial for this particular problem. This explains why in the linear case under Gaussian noise the QMI criterion provides either similar or worse results than the linear matched filter [4].

In this paper we explore an alternative approach that looks for optimal detectors in a feature (transformed) space. To this end, the input data space $\mathscr{R}$ is mapped into a much higher dimensional feature space $\mathscr{F}$

$$\Phi : \mathscr{R} \longrightarrow \mathscr{F}, \qquad x \longrightarrow \Phi(x),$$

where the dot product between feature vectors can be computed using a positive definite kernel function $\langle \Phi(x), \Phi(y) \rangle = \kappa(x,y)$. This is the so-called *kernel trick*, which allows us to obtain nonlinear versions of linear algorithms that can be expressed in terms of inner products, without knowing the exact mapping Φ.

Fig. 1 represents the nonlinear mapping applied to the template and the observations. Typically, to solve any problem in the feature space a cost function involving an empirical risk term and a quadratic regularizer is considered. In this situation, the Representer Theorem [11] shows that the optimal solution can be written as an expansion in terms of the input examples. We use this idea to construct a template hyperplane

$$\mathbf{w}_s = \sum_{j=1}^{N} \beta_j \Phi(s_j),$$

as well as an observation hyperplane

$$\mathbf{w}_r = \sum_{k=1}^{N} \alpha_k \Phi(r_k).$$

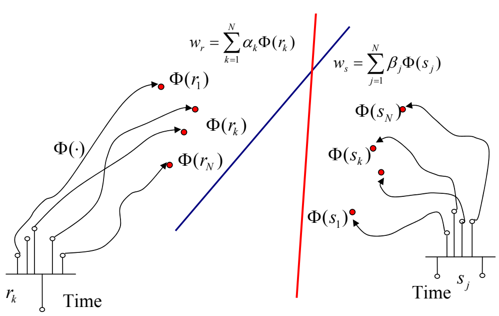

Figure 1: Test statistic in the feature space.

Then, a new statistic for the decision can be formed in the feature space as the inner product between both hyperplanes, i.e.,

$$T(\mathbf{r}) = \sum_{k=1}^{N} \sum_{j=1}^{N} \beta_j \alpha_k \langle \Phi(s_j), \Phi(r_k) \rangle = \sum_{k=1}^{N} \sum_{j=1}^{N} \beta_j \alpha_k \kappa(s_j - r_k), \tag{1}$$

where in the last equality we have used the *kernel trick*. Without loss of generality, here we will only consider a translation-invariant Gaussian kernel with kernel size σ:

$$\kappa_\sigma(x-y) = \exp -\frac{(x-y)^2}{2\sigma^2}.$$

In matrix form, Eq. (1) can be written as

$$T(\mathbf{r}) = \beta^T \mathbf{K} \alpha \tag{2}$$

where $\beta^T = (\beta_1, \ldots, \beta_N)$, $\alpha^T = (\alpha_1, \ldots, \alpha_N)$ and $\mathbf{K}$ is the kernel matrix with elements $K(j,k) = \kappa_\sigma(s_j - r_k)$.

Now the problem reduces to obtain the optimal coefficients β and α. In the following section we propose to apply Fisher linear discriminant analysis in the feature space to this end.

3. KERNEL FISHER DISCRIMINANT ANALYSIS

3.1 Introduction

The signal to noise ratio is the criterion maximized in the linear case; however, in the feature space this criterion does not make sense, since obviously $\Phi(s_k + n_k) \neq \Phi(s_k) + \Phi(n_k)$ and therefore it is difficult to define a meaningful *SNR* measure. A more reasonable criterion would be to apply Fisher discriminant analysis (FDA) [8], which seeks a linear projection from the original space into a low dimensional space by maximizing the between-class scatter (the squared difference between the means for the classes), while simultaneously minimizing the within-class scatter (the sum of variances for each class). Therefore, the function to be maximized for FDA is

$$J_{FDA} = \max \frac{(\mu_0 - \mu_1)^2}{\sigma_0^2 + \sigma_1^2}. \tag{3}$$

When applied in the feature space, linear FDA becomes kernel FDA (K-FDA), which was first proposed in [9] and later generalized to the multiclass problem in [10]. The application of K-FDA to our detection problem is somehow different to that of previous approaches, since here we are trying to optimize two projectors: one for the template, β, and one for the observations, α, whereas in the conventional K-FDA only one linear projector is sought [9],[10]. In addition, the conventional K-FDA problem finds the optimal projection using a set of labeled data for each class, while here we only have the template waveform and some information about the noise distribution.

3.2 Fisher's cost function and optimization procedure

Under hypothesis H_0 (only noise is present) the mean value for the test statistic is

$$\mu_0 = \sum_{k,j=1}^{N} \alpha_k \beta_j E\left[\kappa_\sigma(s_j - n_k)\right].$$

If the noise is normally distributed with zero mean and variance σ_n^2, then the following result can be easily obtained

$$E\left[\kappa_\sigma(s_j - n_k)\right] = \frac{\sigma}{\sigma'} \kappa_{\sigma'}(s_j), \tag{4}$$

where $\sigma' = \sqrt{\sigma^2 + \sigma_n^2}$. Therefore, the mean value under H_0 can be written in matrix form as

$$\mu_0 = \beta^T \overline{\mathbf{K}}_0 \alpha,$$

where $\overline{\mathbf{K}}_0$ is an $N \times N$ matrix whose (j,k) element is given by (4).

Similarly, under hypothesis H_1 we obtain

$$\mu_1 = \beta^T \overline{\mathbf{K}}_1 \alpha,$$

where now the elements of $\overline{\mathbf{K}}_1$ are given by

$$\overline{\mathbf{K}}_1(j,k) = \frac{\sigma}{\sigma'} \kappa_{\sigma'}(s_j - s_k).$$

Let us note that, while $\overline{\mathbf{K}}_0$ is a rank-one matrix, $\overline{\mathbf{K}}_1$ will be in general full-rank. The squared difference between the means, which is the numerator of the FDA cost function (3), is given by

$$(\mu_1 - \mu_0)^2 = \underbrace{\alpha^T \left(\overline{\mathbf{K}}_1 - \overline{\mathbf{K}}_0\right)^T \beta}_{\mu_1 - \mu_0} \underbrace{\beta^T \left(\overline{\mathbf{K}}_1 - \overline{\mathbf{K}}_0\right) \alpha}_{\mu_1 - \mu_0}. \quad (5)$$

By switching the order of the multiplicative terms, Eq. (5) can be also written as

$$(\mu_1 - \mu_0)^2 = \beta^T \left(\overline{\mathbf{K}}_1 - \overline{\mathbf{K}}_0\right) \alpha\alpha^T \left(\overline{\mathbf{K}}_1 - \overline{\mathbf{K}}_0\right)^T \beta.$$

In a parallel way, the variance under hypothesis H_i, for $i = 0, 1$ can be obtained as

$$\sigma_{H_i}^2 = \alpha^T E\left[(\mathbf{K}_i - \overline{\mathbf{K}}_i)^T \beta\beta^T (\mathbf{K}_i - \overline{\mathbf{K}}_i)\right] \alpha = \alpha^T \mathbf{Q}_{H_i}^{\beta} \alpha, \quad (6)$$

where we have used the notation $\mathbf{Q}_{H_i}^{\beta}$ to stress the dependence on β. Alternatively, the variance could have been obtained as

$$\sigma_{H_i}^2 = \beta^T E\left[(\mathbf{K}_i - \overline{\mathbf{K}}_i)\alpha\alpha^T (\mathbf{K}_{H_i} - \overline{\mathbf{K}}_i)^T\right] \beta = \beta^T \mathbf{Q}_{H_i}^{\alpha} \beta.$$

Under the assumption of white Gaussian noise, the matrices $\mathbf{Q}_i^{\beta}$ and $\mathbf{Q}_i^{\alpha}$ can be computed in closed form (due to the lack of space we do not include here the derivation).

After substituting the means and variances in (3), the cost function for K-FDA becomes a nonlinear function of α and β, which must be iteratively solved to get the optimal solution. More specifically, if we consider β fixed we can write

$$J_{K-FDA}(\alpha, \beta) = \frac{\alpha^T \mathbf{S}^{\beta} \alpha}{\alpha^T \left(\mathbf{Q}_1^{\beta} + \mathbf{Q}_0^{\beta}\right) \alpha} \quad (7)$$

where we have defined $\mathbf{S}^{\beta} = \left(\overline{\mathbf{K}}_1 - \overline{\mathbf{K}}_0\right)^T \beta\beta^T \left(\overline{\mathbf{K}}_1 - \overline{\mathbf{K}}_0\right)$. It is easy to show that $\mathbf{S}^{\beta}$ is positive definite and $\left(\mathbf{Q}_1^{\beta} + \mathbf{Q}_0^{\beta}\right)$ is semidefinite positive, therefore (7) is a generalized Rayleigh quotient, whose maximum is given by the eigenvector corresponding to the maximum eigenvalue of the following generalized eigenvalue (GEV) problem

$$\left(\mathbf{Q}_1^{\beta} + \mathbf{Q}_0^{\beta} + \mu\mathbf{I}\right) \alpha = \lambda \mathbf{S}^{\beta} \alpha, \quad (8)$$

where μ is a small positive constant which has been introduced to avoid numerical issues in the solution of the GEV problem as well as to impose additional capacity control in the space of solutions.

The value of α obtained in this way, is used to estimate $\mathbf{Q}_0^{\alpha}$ and $\mathbf{Q}_1^{\alpha}$ and a new solution for β is now obtained by solving the following GEV problem

$$\left(\mathbf{Q}_1^{\alpha} + \mathbf{Q}_0^{\alpha} + \mu\mathbf{I}\right) \beta = \lambda \mathbf{S}^{\alpha} \beta \quad (9)$$

where $S^{\alpha} = \left(\overline{\mathbf{K}}_1 - \overline{\mathbf{K}}_0\right) \alpha\alpha^T \left(\overline{\mathbf{K}}_1 - \overline{\mathbf{K}}_0\right)^T$. The procedure is repeated until convergence.

3.3 A simplified procedure

Let us consider that the transmitted waveform is a Gaussian pulse $s = \exp(-t^2)$, where $t = -5 : 0.25 : 5$. This signal is received in AWGN and the signal to noise ratio was set to $SNR = 10$dB. Fig. 2 shows the optimal β and α obtained by applying the proposed K-FDA detector to this example (the regularization parameter was $\mu = 1e-5$ and the kernel size was $\sigma^2 = 5$).

We can see that both curves look like the original Gaussian pulse; in fact, α (i.e. the projector for the observations) is indistinguishable (up to a scale factor) from the original waveform. This observation, which has been corroborated in a number of examples, allows us to simplify the optimization procedure to get α and β. In particular, we can fix $\alpha = \mathbf{s}$ and then the optimal β can be obtained in a single step by solving the GEV problem (9). With this simplification, only matrices $\mathbf{Q}_i^{\alpha}$ for $i = 0, 1$ (see Appendix A) are needed in the optimization procedure.

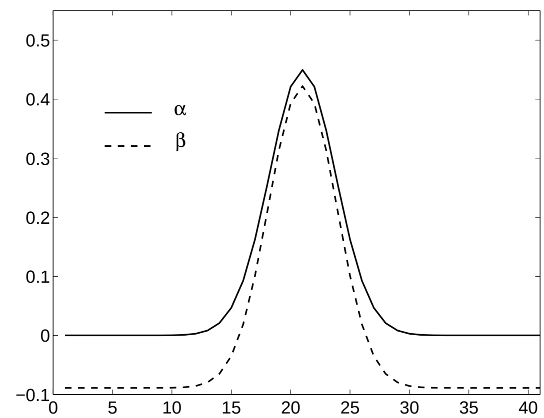

Figure 2: Optimal values of α and β for the Gaussian pulse.

4. SIMULATION RESULTS

In this section we compare the performance of the proposed kernel Fisher discriminant analysis (K-FDA) statistic with that of the conventional linear matched filter (MF) and the recently proposed quadratic mutual information (QMI) estimator [4], for different known signal waveforms and under different noise distributions (Gaussian and impulsive).

The QMI is defined as

$$I_{QMI} = \frac{1}{2} \log \frac{\int\int f_{SR}^2(s,r)dsdr \int\int f_S^2(s) f_R^2(r)drds}{\left(\int\int f_{SR}(s,r) f_S(s) f_R(r)dsdr\right)^2}, \quad (10)$$

and it measures nonlinear dependencies between the observations r_k and the signal s_k. In fact, it measures the correlation between the joint pdf of $\mathbf{R}$ and $\mathbf{S}$ and the product of their marginals [4].

We consider again the Gaussian waveform of Section 3.3. All the results shown for the K-FDA detector have been obtained with $\sigma^2 = 5$ and $\mu = 1e-5$ (regularization parameter) using the simplified procedure described in Section 3.3.

In order to evaluate the performance of K-FDA under impulsive noise, we have considered the following Gaussian mixture model

$$f_N(n) = (1-\varepsilon)N(0, \sigma_1^2) + \varepsilon N(0, \sigma^2)$$

where ε measures the percentage of noise spikes and $\sigma_2^2 >> \sigma_1^2$. In our simulations we have used $\varepsilon = 0.15$ and $\sigma_2^2 = 50\sigma_1^2$.

Fig. 3 shows the receiver operating characteristic (ROC) curves for MF, QMI and K-FDA, when the waveform is distorted by AWG noise or impulsive noise; in both situations the *SNR* was set to 10 dB. Under Gaussian noise, obviously the MF is the optimal test statistic for the problem. We can see that the K-FDA detector, although obtained in a completely different way, provides the same performance. On the other hand the QMI is not able to achieve the optimal performance in the linear case. When the noise is impulsive the proposed K-FDA detector clearly outperforms the linear matched filter. This increased robustness against impulsive noise is attributed to the fact that when an outlier is present, the inner product in the feature space computed via the Gaussian kernel tends to be zero (i.e., $\kappa(s_i - r_k) \approx 0$ when r_k has a large value).

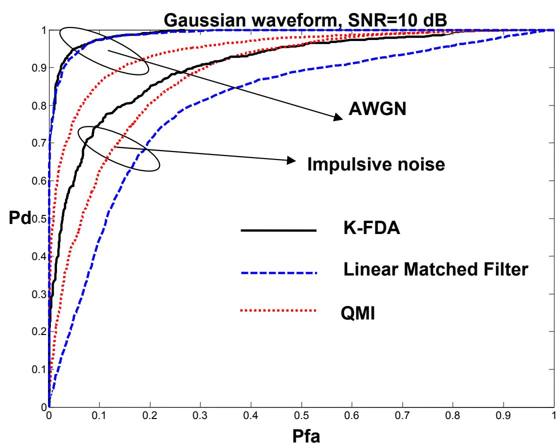

Figure 3: ROC curves for K-FDA, MF and QMI. Gaussian pulse, AWGN, SNR=5 dB.

We can conclude that under Gaussian noise the K-FDA obtains practically the same results than the optimal MF, but shows a more robust behavior under unknown impulsive noise distributions. On the other hand, the QMI also shows some robustness against impulsive noise, but under Gaussian noise does not achieve the optimal performance of the MF filter.

5. CONCLUSIONS

In this paper we have proposed a new nonlinear test statistic for detecting a known signal in noise. The test statistic computes the inner product between two specific projections in the feature space and the optimal projections for maximal discrimination between both hypotheses are obtained using Fisher discriminant analysis. Under the Gaussian noise assumption we were able to derive closed-form expressions for the means and variances for each hypothesis, and an iterative procedure was proposed to maximize the Fisher's criterion. By means of some simulation results we have shown that the nonlinear test statistic achieves the optimal performance of the linear matched filter under Gaussian noise, while providing an increased robustness against impulsive additive interferences.

Although we have focused this study on the matched filtering problem, we think that similar ideas could be applied to solve prediction or identification problems in the feature space: this seems to be an interesting line for further research.

APPENDIX A

Under the Gaussian noise assumption it can be shown that $\mathbf{Q}_0^\alpha$ is a matrix with elements

$$\mathbf{Q}_0^\alpha(i,j) = ||\alpha||_2^2 \left(\gamma \kappa_{\sigma_1}(s_i, s_j) \kappa_{\sigma_2}\left(\frac{s_i + s_j}{2}, 0\right) - G_0(i,j) \right)$$

and $\mathbf{Q}_1^\alpha$ has elements

$$\mathbf{Q}_1^\alpha(i,j) = \sum_{k=1}^{N} \alpha(k)^2 \left(\gamma \kappa_{\sigma_1}(s_i, s_j) \kappa_{\sigma_2}\left(\frac{s_i + s_j - 2s_k}{2}, 0\right) - G_1(i,j) \right).$$

where $\gamma = \frac{\sigma}{\sqrt{2}\sigma_2}$, $\sigma_1^2 = 2\sigma^2$ and $\sigma_2^2 = \sigma_n^2 + \sigma^2/2$ and . $G_m(i,j =)\overline{\mathbf{K}}_m(i,1)\overline{\mathbf{K}}_m(j,1)$, for $m = 0, 1$.

REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. II, Detection Theory*, Upper Saddle River, NJ: Prentice Hall, 1998.

[2] C. W. Helstrom, *Elements of Signal Detection and Estimation*, Englewood Cliffs, NJ: Prentice-Hall, 1995.

[3] I. Song, S. A. Kassam, "Locally optimum rank detection of correlated random signals in additive noise," *IEEE Trans. Inform. Theory*, vol.38, pp. 1311-1322, 1992.

[4] D. Erdogmus, Rati Agrawal, Jose C. Principe, "A mutual information extension to the matched filter," to appear in *Signal Processing*, 2005.

[5] V. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer Verlag, 1995.

[6] J. Shawe-Taylor, N. Cristiani, *Kernel Methods for Pattern Analysis,*, Cambridge University Press, 2004.

[7] F. Perez-Cruz, O. Bousquet, "Kernel methods and their potential use in signal processing," *IEEE Signal Processing Magazine*, vol. 21, pp. 57-65, May 2004.

[8] R. Duda, P. Hart, D. Stork, *Pattern Classification*, 2nd. Ed., New York: John Wiley, 2001.

[9] S. Mika, G. Rätsch, J. Wetson, B. Schölkopf, K.-R. Müller, "Fisher discriminant analysis with kernels," in *IEEE Int. Workshop on Neural Networks for Signal Processing IX*, pp. 41-48, Madison, WI, USA, Aug. 1999.

[10] G. Baudat, F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, pp. 2385-2404, 2000.

[11] G. S. Kimeldorf, G. Wahba, "Some results on Tchebycheffian spline functions," *J. Math. Anal. Applic.*, vol. 33, pp. 82-95, 1971.