# A MUTUAL INFORMATION EXTENSION TO THE MATCHED FILTER

Deniz Erdogmus[1], Rati Agrawal[2], Jose C. Principe[2]
[1] CSEE Department, Oregon Graduate Institute, OHSU, Portland, OR 97006, USA
[2] CNEL, ECE Department, University of Florida, Gainesville, FL 32611, USA

**Abstract.** Matched filters are the optimal linear filters for signal detection under linear channel and white noise conditions. Their optimality is guaranteed in the additive white Gaussian noise channel due to the sufficiency of second order statistics. In this paper, we introduce a nonlinear filter for signal detection based on the Cauchy-Schwartz quadratic mutual information (CS-QMI) criterion. This filter is still implementing correlation but now in a high dimensional transformed space defined by the kernel utilized in estimating the CS-QMI. Simulations show that the nonlinear filter significantly outperforms the traditional matched filter in nonlinear channels, as expected. In the linear channel case, the proposed filter outperforms the matched filter when the received signal is corrupted by impulsive noise such as Cauchy-distributed noise, but performs at the same level in Gaussian noise. A simple nonparametric sample-estimator for CS-QMI is derived for real-time implementation of the proposed filter.

## 1. INTRODUCTION

Detection of known signals transmitted through linear and nonlinear channels is an important fundamental problem in signal processing theory with a wide range of applications, communications, radar, and biomedical engineering to name just a few [1,2]. Traditionally the matched filter, which is based on the assumptions of linear channel with second order optimality criteria, has been used for tackling the signal detection problem. Theoretically, it is known that among all linear filters, the matched filter maximizes the signal-to-noise ratio (SNR) in the case of linear additive white noise (AWN) channels [3]. The known signal shape is utilized to construct an impulse response, hence the name *matched filter*.

The limitations of the matched filter are already clearly defined by the assumptions under which its optimality can be proven. Specifically, if the channel causes nonlinear distortions on the transmitted signal, as the template used to construct the impulse response will not be *optimal* anymore, the matched filter is expected to yield suboptimal signal detection and false alarm performance. Even in the case of a linear channel, if the noise is impulsive (i.e., has infinite variance) then the SNR optimality of the matched filter is not valid anymore, as theoretically the noise power at the output of the matched filter is still infinite.

In order to address these shortcomings of the linear matched filter theory, we propose in this paper a nonlinear filter topology for signal detection based on a mutual information (MI) criterion. In earlier work, we have demonstrated the superior performance of information theoretic measures over second order statistics in signal processing [4]. In this general framework, we assume an arbitrary instantaneous channel and additive noise with an arbitrary distribution. Specifically in the case of linear channels, we will consider Cauchy noise, which is a member of the family of symmetric $\alpha$-stable distributions, and is an impulsive noise source. These types of noise distributions are known to plague signal detection [5, 6].

Specifically, the proposed nonlinear signal detection filter is based on the Cauchy-Schwartz Quadratic Mutual Information (CS-QMI) measure that has been proposed by Principe *et al*. [7]. This definition of mutual information is preferred because of the existence of a simple nonparametric estimator for the CS-QMI, which in turn forms the basis of the nonlinear filter topology that is being proposed here.

The performance of the proposed nonlinear signal detection filter will be compared with that of the traditional matched filter in a variety of scenarios including linear and nonlinear channels, and Gaussian and impulsive noise distributions. The metric for comparison is the receiver operating characteristics (ROC), which is a standard technique for evaluating the performance of a classifier by demonstrating the trade-off between probability of detection and probability of false alarms [1]. For the sake of simplicity, throughout the paper, we will assume discrete-time signals.

## 2. MATCHED FILTER

Consider a signal template $s_k$ existing in the time interval $[0,T]$, corrupted by an AWN $n_k$ with zero mean and variance $\sigma^2$. The received signal is simply $r_k=s_k+n_k$. A matched filter is then defined by the impulse response [1,3]

$$h_k = s_{T-k} \tag{1}$$

The matched filter output is then $y_k=h_k*r_k=h_k*s_k+h_k*n_k$, where * denotes convolution. Thus, the filter output is composed of a signal and a noise component. The output achieves its maximum average value at the time instant $T$, since there is maximum correlation between the matched filter impulse response and the template at this lag, thereby maximizing the SNR. The SNR has been defined as the ratio of the total energy of the signal template to the noise variance [1]:

$$SNR = \frac{1}{\sigma^2}\sum_{k=0}^{T} s_k^2 \tag{2}$$

If the proper lag $T$ to sample the output of the matched filter is known, this output statistic can then compared with a threshold, to detect the presence or absence of the signal $s_k$.

## 3. MUTUAL INFORMATION

Mutual information indicates the amount of shared information between two or more random variables. In information theory, the MI between two random variables $X$ and $Y$ is traditionally defined by Shannon as [8]

$$I_S(X;Y) = \iint p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) p_Y(y)} dxdy \qquad (3)$$

where $p_{XY}(x,y)$ is the joint probability density function (pdf) of $X$ and $Y$, and $p_X(x)$ and $p_Y(y)$ are the marginal pdfs. The crucial property of mutual information for our purposes is the fact that if there is a nonlinear function between $X$ and $Y$, such that $Y=f(X)$, the MI achieves its maximum value [8].[1] On the other hand, if $X$ and $Y$ are independent MI becomes zero. In a sense, MI can be considered a generalization of correlation to nonlinear dependencies; that is MI can be used to detect nonlinear dependencies between two random variables, whereas the usefulness of correlation is limited to linear dependencies.

Although Shannon's MI is the traditionally preferred measure of shared information, essentially it is a measure of divergence between the variables $X$ and $Y$ from independence. Based on this understanding, a different, but qualitatively similar measure of independence can be obtained using the Cauchy-Schwartz inequality for inner products in vector spaces: $\langle \mathbf{x},\mathbf{y} \rangle \le \|\mathbf{x}\| \cdot \|\mathbf{y}\|$. The following expression is defined as the CS-QMI between $X$ and $Y$ [7]:

$$I_{CS}(X,Y) = \frac{1}{2} \log \frac{\iint p_{XY}^2(x,y)dxdy \iint p_X^2(x) p_Y^2(y) dxdy}{\left( \iint p_{XY}(x,y) p_X(x) p_Y(y) dxdy \right)^2} \qquad (4)$$

This measure evaluates to zero when the variables are independent. The inverse of the argument of the log is bounded between one (independence) and zero (maximal dependence). Computation of CS-QMI involves the estimation of the joint and marginal pdfs of the random variables. However using the Parzen window density estimate, CS-QMI can be estimated nonparametrically in a straightforward manner as demonstrated below.

The Parzen window estimate for the pdf $p_X(x)$ of a random variable $X$ is given in terms of the iid samples $\{x_1,\ldots,x_N\}$ as [9]

$$\hat{p}_X(x) = \frac{1}{N} \sum_{i=1}^{N} K_\sigma(x - x_i) \qquad (5)$$

where $K_\sigma(.)$ represents the kernel (or the window) function. Parzen windowing, also referred to as kernel density estimation, is known to have very good convergence properties [10]. The parameter $\sigma$ determines the kernel size (sometimes called the window size or the bandwidth) and it presents a trade-off between estimation bias and variance. On average, the estimated density is the convolution of the true underlying density with the kernel function. Consequently, decreasing the kernel size towards zero, such that the kernel function approaches a delta function, reduces estimation bias. On the other hand, larger kernel sizes lead to smaller estimation variance. Nevertheless, it is possible to obtain an asymptotically unbiased and consistent density estimate, as the number of samples tends to infinity, if a suitable kernel function, such as Gaussian, is utilized.

In addition, a Gaussian kernel facilitates the derivation of a simple expression for the CS-QMI measure.[2] Specifically, if we assume a separable kernel $K(x,y)=G_1(x)G_2(y)$, where

$$G_i(\xi) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\xi^2/(2\sigma_i^2)} \qquad (6)$$

the corresponding nonparametric estimator for CS-QMI can be obtained after some calculations to be

$$\begin{aligned} \hat{I}_{CS}(X;Y) &= \frac{1}{2}\log\left( \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} G_1(x_j - x_i) G_2(y_j - y_i) \right) \\ &+ \frac{1}{2}\log\left( \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} G_1(x_j - x_i) \right) \\ &+ \frac{1}{2}\log\left( \frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} G_2(y_j - y_i) \right) \\ &- \log\left( \frac{1}{N^3} \sum_{k=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} G_1(x_k - x_j) G_2(y_k - y_i) \right) \end{aligned} \qquad (7)$$

This nonparametric estimator for CS-QMI also facilitates a deeper understanding of the measure itself. Consider the argument of the log in the original measure defined in (4). This is simply the inverse-square of the inner product between the joint pdf of $X$ and $Y$, and the product of their marginal pdfs. Consequently, the denominator of this term, being the *correlation* of the joint pdf with the product of marginal pdfs, indicates some sort of link to second order statistics of the signals $X$ and $Y$ after the nonlinear transformation produced by the kernels.

Now consider the nonparametric estimator in (7). We will focus on the argument of the log in each term of the expression on the right hand side. Since the Gaussian kernel $G_i(.)$ satisfies Mercer's conditions [11], the following eigenfunction expansion exists:[3]

[1] Theoretically, MI can become infinite, however, in practice, when a nonparametric estimator is used to estimate it from samples, this will not happen.

[2] In the rest of this paper, we will assume Gaussian kernels. However, everything that we present is valid for any kernel function that satisfies the conditions of Mercer's theorem. These kernels include, but are not limited to, symmetric and unimodal pdfs with finite variances.

[3] For convenience, we denote the infinite-length inner product summation by a vector-matrix product notation.

$$G(\xi-\eta)=\sum_{l=1}^{\infty}\lambda_l\tilde{\varphi}_l(\xi)\tilde{\varphi}_l(\eta)=\boldsymbol{\varphi}^T(\xi)\boldsymbol{\varphi}(\eta) \quad (8)$$

In (8), $\{\lambda_1>\lambda_2>\ldots\}$ are the eigenvalues of the Gaussian kernel and $\varphi_l(.)=\lambda_l^{1/2}\tilde{\varphi}_l(.)$ are the scaled eigenfunctions, where $\tilde{\varphi}_l(.)$ are the orthonormal eigenfunctions. Substituting this expansion in (7) for each of the corresponding Gaussian kernels and reorganizing the terms, we obtain

$$\begin{aligned}\hat{I}_{CS}(X;Y)=&\frac{1}{2}\log\left(\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\varphi}_1^T(x_j)\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}_1(x_i)\boldsymbol{\varphi}_2^T(y_i)\right)\boldsymbol{\varphi}_2(y_j)\right)\\&+\frac{1}{2}\log\left(\left(\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\varphi}_1(x_j)\right)^T\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}_1(x_i)\right)\right)\\&+\frac{1}{2}\log\left(\left(\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\varphi}_2(y_j)\right)^T\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}_2(y_i)\right)\right)\\&-\log\left(\left(\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\varphi}_1(x_j)\right)^T\left(\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\varphi}_1(x_k)\boldsymbol{\varphi}_2^T(y_k)\right)\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}_2^T(y_i)\right)\right)\end{aligned} \quad (9)$$

Defining the first and second-order moments of the transformed data as

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{\varphi}_1(X)}&=\left(\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\varphi}_1(x_j)\right)\approx E[\boldsymbol{\varphi}_1(X)]\\\boldsymbol{\mu}_{\boldsymbol{\varphi}_2(Y)}&=\left(\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\varphi}_2(y_j)\right)\approx E[\boldsymbol{\varphi}_2(Y)]\\\boldsymbol{\Gamma}_{\boldsymbol{\varphi}_1(X)\boldsymbol{\varphi}_2(Y)}&=\left(\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\varphi}_1(x_k)\boldsymbol{\varphi}_2^T(y_k)\right)\approx E\left[\boldsymbol{\varphi}_1(X)\boldsymbol{\varphi}_2^T(Y)\right]\end{aligned} \quad (10)$$

we obtain a more compact expression for (7) in the so called *feature space* defined by the transformations $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$:

$$\begin{aligned}\hat{I}_{CS}(X;Y)\approx&\frac{1}{2}\log E\left[\boldsymbol{\varphi}_1^T(X)\boldsymbol{\Gamma}_{\boldsymbol{\varphi}_1(x)\boldsymbol{\varphi}_2(y)}\boldsymbol{\varphi}_2(Y)\right]\\&+\frac{1}{2}\log\left\|\boldsymbol{\mu}_{\boldsymbol{\varphi}_1(X)}\right\|^2\\&+\frac{1}{2}\log\left\|\boldsymbol{\mu}_{\boldsymbol{\varphi}_2(Y)}\right\|^2\\&-\log\left[\boldsymbol{\mu}_{\boldsymbol{\varphi}_1(X)}^T\boldsymbol{\Gamma}_{\boldsymbol{\varphi}_1(X)\boldsymbol{\varphi}_2(Y)}\boldsymbol{\mu}_{\boldsymbol{\varphi}_2(Y)}\right]\end{aligned} \quad (11)$$

Hence, we conclude that the CS-QMI estimator that we propose is, in fact, using second order statistical information about the signals $X$ and $Y$ in a transform domain determined by the kernel functions used in (7). The transform domain, called the feature space, is infinite dimensional in the case of Gaussian kernels, however, polynomial kernels for which the feature space is finite dimensional exists. The eigenfunctions of the Gaussian kernel form a basis for the Hilbert space of bounded continuous functions, thus such kernels are also referred to as reproducing kernels in Hilbert spaces (rkhs) [12]. Therefore, the proposed CS-QMI estimator contains higher-order statistical information regarding the original random variable $X$ and $Y$.

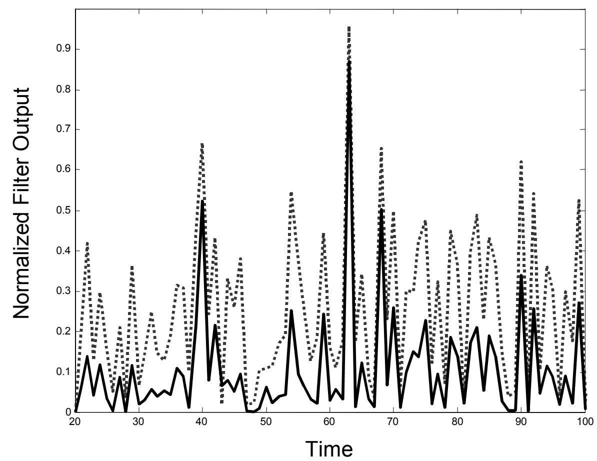


Figure 1. Normalized outputs of the traditional matched filter (dotted) and the proposed nonlinear filter (solid) versus time. The expected time of the filter output peak is indicated by the star. In this illustration, both filters detect the transmitted signal, however the proposed nonlinear filter exhibits better false alarm rejection capabilities due to the reduced background clutter amplitude.

## 4. NONLINEAR SIGNAL DETECTION FILTER

Consider a signal template $s_k$ existing in the time interval $[0,T]$, a nonlinear channel $f(.)$, and AWN $n_k$ with an arbitrary distribution. The received signal is

$$r_k=\begin{cases}f(s_k)+n_k & \text{if signal exists}\\ n_k & \text{otherwise}\end{cases},\ k=0,\ldots,T \quad (12)$$

The samples $\{r_0,r_1,\ldots,r_T\}$ and $\{s_0,s_1,\ldots,s_T\}$ are assumed to be samples drawn from the joint distribution of random variables $R$ and $S$, corresponding to the received signal and the template, respectively. The nonlinear matched filter evaluates $\hat{I}_{CS}(R;S)$ using (7) and its output is compared with a threshold to determine whether the signal exists or not. Large values of $\hat{I}_{CS}(R;S)$ indicate the existence of $f(S)$, a nonlinearly distorted version of the template signal, in the received signal. In contrast, the traditional matched filter simply evaluates the correlation between $R$ and $S$ by $r_0s_0+\ldots+r_Ts_T$ to be compared with a threshold.

In order to illustrate the two filter outputs comparatively, we present in Figure 1 the outputs of both filters in a simple signal detection example, where the template (transmitted signal) is distorted by a sinusoidal channel (described in the next section). In addition, an AWGN corrupts the received signal at 20dB SNR level. Clearly, the proposed nonlinear filter exhibits better peak-to-background clutter ratio, thus resulting in superior detection and false alarm rates as will be demonstrated in the following section extensively.

The size $\sigma_i$ of the Gaussian kernel functions are important parameters that must be carefully chosen in order to achieve the best performance possible. There is a wide literature on selecting the kernel size in the statistics literature [10,13,14] and it is out of the scope of this paper. Here we selected the kernel size experimentally for each case according to $\sigma = \sigma_0 std(x)/\sqrt{(T+1)}$, where $\sigma_0$ is typically in the interval [1,10] and $std(x)$ denotes the sample standard deviation of the template or the received signal at any given time instant.

A practical consideration for real-time implementation of the matched nonlinear filter is computation complexity. Clearly, the total computational complexity is much more than the matched filter. In fact, a straightforward implementation of (7) in a DSP would require $O(T^2)$ Gaussian function evaluations, which is still reasonable for short templates. Sample evaluation times for the proposed nonlinear filter are presented at the end of Section 5. However, it is also possible to design a specialized circuit (perhaps using an FPGA or a VLSI chip) that would implement the necessary computations in parallel. The second approach aims at distributing the necessary computation load in space rather than in time. Therefore, a simple feedforward circuit structure could be designed to generate an output instantaneously, as soon as the necessary input samples are captured at the outputs of a tap-delay line.

## 5. SIMULATIONS AND RESULTS

Monte-Carlo simulations were performed for detecting a known signal through both linear and nonlinear channels in the presence of different additive noise distributions such as Cauchy, Gaussian, and Laplacian distributions. Specifically, simulations were performed at three different setups:

1. Linear channel, different noise distributions, proper sampling time known.
2. Nonlinear channel, Gaussian noise, proper sampling time known.
3. Nonlinear channel, Gaussian noise, unknown sampling time.

The performance comparisons are presented using the ROC curves of each filter. The ROC curves simply show the probability of detection (vertical axis) versus the probability of false alarm (horizontal axis) for various threshold values. Since the performance in some cases is very similar, we also plot the differences in false alarms and probability of detection between the two detectors in the ROC.

*Experiment 1*: The transmitted signal is a random sequence of length 20 with values selected from {-1,+1} uniformly. The channel is linear ($f(S)=S$) and the noise distribution is varied between Gaussian, Laplacian, and Cauchy. For each noise distribution, 10,000 Monte Carlo

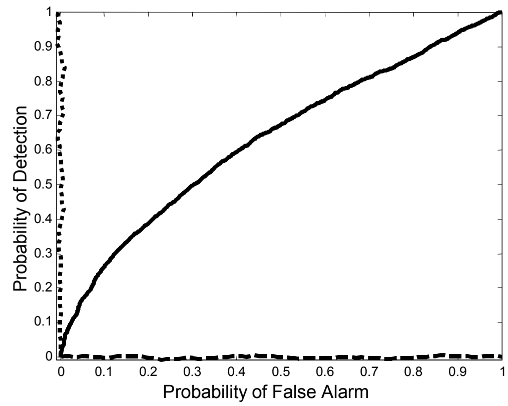


Figure 2. ROC curve for linear channel with additive white Gaussian noise for MI and MF (solid) at 0dB SNR basically coincide. The difference in probability of false alarm of MI and MF (dotted) and difference in probability of detection of MI and MF (dash-dotted) are also shown.

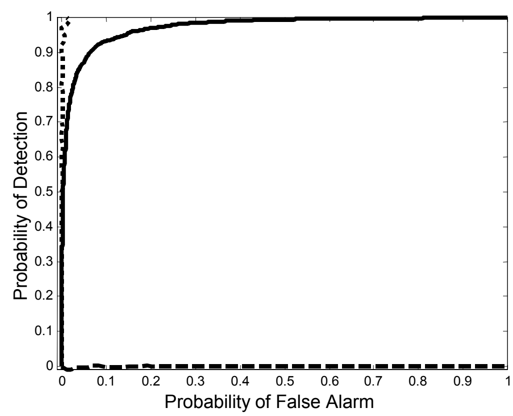


Figure 3. ROC curve for linear channel with additive white Gaussian noise for MI and MF (solid) at 10dB SNR also overlap. The difference in probability of false alarm of MI and MF (dotted) and difference in probability of detection of MI and MF (dash-dotted) are also shown.

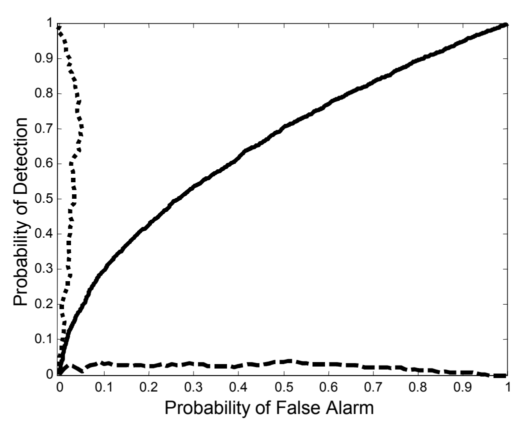


Figure 4. ROC curve for linear channel with additive white Laplacian noise for MI and MF (solid) at 0dB SNR still coincide. The difference in probability of false alarm of MI and MF (dotted) and difference in probability of detection of MI and MF (dash-dotted) are also shown.

detection simulations are performed where in each simulation the probability of the transmitted signal was set at 0.5.

The results for the linear additive channel are shown in Figures 2-6. It is clear from Figures 2-3 that for the Gaussian noise case, the proposed nonlinear filter (denoted by MI) performs identically to the traditional matched filter (denoted by MF) and cannot outperform it, showing the optimality of the matched filter in this case. We have noticed that the performance of the nonlinear filter becomes worse than the MF for smaller SNR (results not shown here). On the other hand, for Laplacian noise, especially at high SNR, MI slightly outperforms MF as depicted in Figures 4-5. Eventually, for the impulsive Cauchy noise case, MI significantly outperforms MF at both SNR levels, as shown in Figure 6.[4]

*Experiment 2*: The transmitted signal is composed of 21 samples from a Gaussian-shaped waveform. The channel is nonlinear ($f(S)=1/(1+\exp(-10(S-0.3)))-0.5$ or $f(S)=\sin(2\pi S)$) and the noise distribution is Gaussian. For both nonlinear channels, 10,000 Monte Carlo detection simulations are performed where in each simulation the probability of the transmitted signal is set at 1/2. The ROC of the traditional matched filter (MF) and the proposed nonlinear MI-based filter (MI) are estimated as before for different SNR levels. In the nonlinear channel case, the SNR level refers to the SNR measured at the received signal.

The results for these two nonlinear channels are shown in Figures 7-8. In both cases, MI outperformed MF at both SNR levels. The difference in performance is larger for the sinusoidal channel, because it distorts the template more to reduce correlation relative to the sigmoidal channel.

*Experiment 3*: The experimental setup is identical to experiment 2 except that the transmitted signal occurs at unknown lags in a long sequence of received signal values:

$$r_k = \sum_{m=1}^{M} f(s_{k-\Delta_m}) + n_k \,,\; k = 1,\ldots,N \tag{13}$$

Consequently, the streaming outputs of the MF and MI have to be searched for transmitted signals. We assume that there is no overlap between transmitted signals. Furthermore, we know that the filter outputs rise and fall over a period of $2(T+1)+1$ achieving, on average their maximum values at the midpoint of this interval. Hence, by combining the thresholding scheme with the condition that there cannot be two signal detections within any interval of length $2(T+1)+1$, many false alarms can be avoided. The sigmoidal nonlinear channel mentioned earlier is used with Gaussian noise. The ROC curves are shown in Figure 9 only up to a false alarm rate of 0.4 (due to long simulation time required for lower

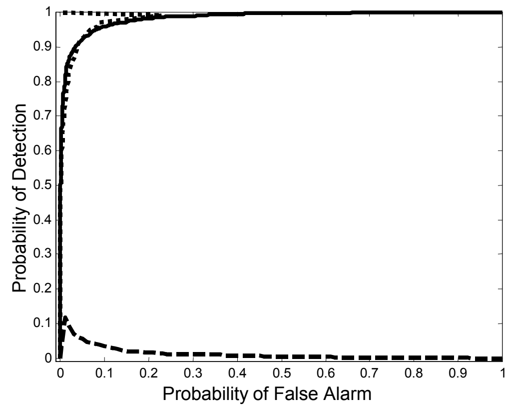


Figure 5. ROC curve for linear channel with additive white Laplacian noise for MI and MF (solid) at 10dB SNR. The difference in probability of false alarm of MI and MF (dotted) and difference in probability of detection of MI and MF (dash-dotted) are also shown for better clarity.

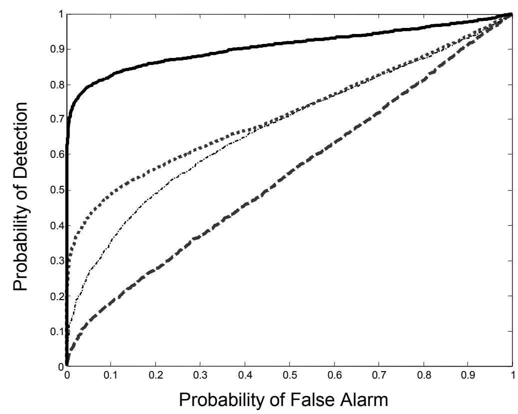


Figure 6. ROC curves for linear channel with additive white Cauchy noise for MI (solid) and MF (dotted) at 20dB SNR and MI (dash-dotted) and MF (dashed) at 10dB SNR.

threshold values) for MI and MF at two SNR levels. The proposed MI approach significantly outperforms the traditional MF. As discussed in section 4, the computational complexity of MI is $O(T^2)$. The average time taken for evaluating the output of the MF and MI in MATLAB, over different sample lengths (# of iterations = 10,000) is tabulated in Table 1. The figures in the table indicate a quadratic increase in the evaluation time of MI with an increase in the sample length and reasonable evaluation times for shorter sample lengths.

## 6. CONCLUSION

Signal detection is a fundamental problem in signal processing and traditionally, based on linearity and Gaussianity assumptions, the matched filter has been used for tackling this problem. Clearly, the matched filter, being based on second-order correlations alone, has shortcomings especially when nonlinear distortions are introduced by the channel. In addition, the matched filter performance degrades

[4] Although technically it is not correct to speak of the SNR of a signal corrupted by Cauchy-distributed noise, for the sake of illustration, the actual sample variance of the noise in the simulations can be used as a measure of noise strength in the received signal.

| Sample Length | 10 | 20 | 50 | 100 |
|---|---|---|---|---|
| MF (seconds) | 0.97 | 1.00 | 1.07 | 1.18 |
| MI (seconds) | 9.32 | 9.32 | 42.30 | 44.00 |

Table 1: Speed of computation (in seconds) of MI and MF in 10,000 iterations

in the presence of impulsive noise disturbances at the receiver.

Motivated by these shortcomings of the matched filter and the promise of mutual information as an extension of linear correlation to nonlinear dependencies, we have designed a *nonlinear matched filter*, based on the Cauchy-Schwartz Quadratic Mutual Information measure. A nonlinear filter design based on a nonparametric estimator for CS-QMI has been proposed. Based on the theory of reproducing kernels in Hilbert spaces, it has been shown that the proposed filter is able to capture all essential statistical properties of the nonlinearly transformed signals by virtually creating an infinite dimensional feature space.

Finally, extensive simulation results that demonstrate clearly the superiority of the mutual information based nonlinear filter over the traditional matched filter have been presented in a variety of signal detection conditions including linear and nonlinear channel distortions, as well as various noise distributions including Gaussian, Laplacian, and Cauchy, the latter being a troublesome impulsive noise type for the matched filter. Despite the similarities, the nonlinear matched filter utilizes the information available in the samples differently. According to eq. 11, CS-QMI estimates correlation in the transformed space. But the transformation involves a nonlinear function of pairs of samples and therefore the temporal information about the template is lost. This is one of the reasons the nonlinear filter outperforms the matched filter in nonlinear channels where the template shape changes due to the channel distortion. On the other hand, in linear channels the CS-QMI is bound to loose its performance advantage, especially at low SNR where the noise may take over. Another problem is the variance of the CS-QMI estimate which is dictated by the number of samples of the template (i.e. small templates should degrade performance).

The implementation of the CS-QMI is more computationally demanding when compared with the matched filter, but for small templates it is still manageable. The choice of the Gaussian kernel in the simulations, does not exclude the use of other kernel types, although the authors are of the opinion that there would be insignificant improvement in the performance using other kernel types. Corrections to non-white noise are readily available for the matched filter but they have not been attempted for the CS-QMI. Another area that was not investigated here is the impact in performance when the template suffers time warping distortions, which plague many practical applications of matched filters.

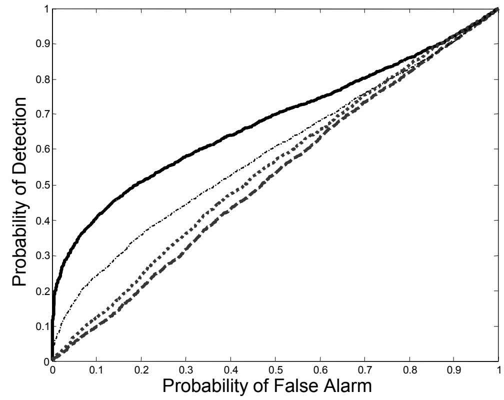


Figure 7. ROC curves for a sigmoidal channel with additive white Gaussian noise for MI (solid) and MF (dotted) at 15dB SNR and MI (dash-dotted) and MF (dashed) at 10dB SNR.

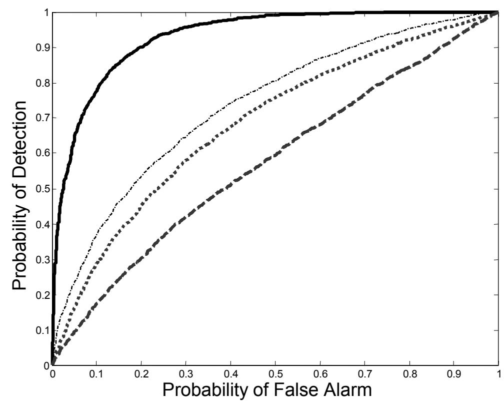


Figure 8. ROC curves for a sinusoidal channel with additive white Gaussian noise for MI (solid) and MF (dotted) at 15dB SNR and MI (dash-dotted) and MF (dashed) at 10dB SNR.

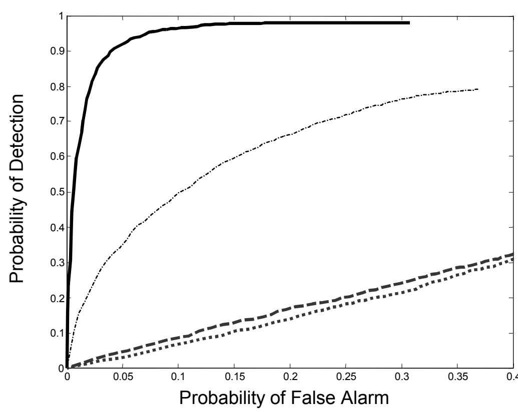


Figure 9. ROC curves for a sigmoidal channel with additive white Gaussian noise in the case of unknown sampling time for MI (solid) and MF (dotted) at 20dB SNR and MI (dash-dotted) and MF (dashed) at 10dB SNR.

The contributions of this paper are not merely limited to the development of a nonlinear signal detection filter. It should be clear that nonlinear techniques such as kernel methods based on information theoretic measures hold the promise of changing the way signal processing is done. These and other recent results with information theoretic learning indicate that many traditional signal processing solutions based on linear systems theory and second order statistics, should be re-examined because relatively straight forward nonlinear extensions exist that show potential in outperforming the conventional signal processing solutions.

**Acknowledgments.** This work is partially supported by NSF grant ECS-0300340.

## REFERENCES

[1] C.W. Helstrom, *Elements of Signal Detection and Estimation*, Prentice Hall, New Jersey, 1995.

[2] J.G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.

[3] G.L. Turin, "An Introduction to Matched Filters," IEEE Transactions on Information Theory, vol. 6, pp. 311-329, 1960.

[4] D. Erdogmus, J.C. Principe, "An Error-Entropy Minimization Algorithm for Supervised Training of Nonlinear Adaptive Systems," IEEE Transactions on Signal Processing, vol. 50, no. 7, pp. 1780-1786, 2002.

[5] S.M. Zabin, G.A. Wright, "Nonparametric Density Estimation and Detection in Impulsive Interference Channels – Part I: Estimators," IEEE Transactions on Communications, vol. 42, pp. 1684-1697, 1994.

[6] S.M. Zabin, G.A. Wright, "Nonparametric Density Estimation and Detection in Impulsive Interference Channels – Part II: Detectors," IEEE Transactions on Communications, vol. 42, pp. 1698-1711, 1994.

[7] J.C. Principe, D. Xu, J. Fisher, "Information Theoretic Learning," in *Unsupervised Adaptive Filtering* (S. Haykin, ed.), Wiley, New York, pp. 265–319, 2000.

[8] T. Cover, J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[9] E. Parzen, "On Estimation of a Probability Density Function and Mode", in *Time Series Analysis Papers*, Holden-Day, Inc., San Diego, California, 1967.

[10] L. Devroye, G. Lugosi, *Combinatorial Methods in Density Estimation*, Springer, New York, 2001.

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice Hall, Upper Saddle River, New Jersey, 1999.

[12] G. Wahba, *Spline Models for Observational Data*, Capital City Press, Montpelier, Vermont, 1992.

[13] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.

[14] R.P.W. Duin, "On the Choice of the Smoothing Parameters for Parzen Estimators of Probability Density Functions," IEEE Transactions on Computers, vol. 25, no. 11, pp. 1175-1179, 1976.