2015 International Symposium on Nonlinear Theory and its Applications
NOLTA2015, Kowloon, Hong Kong, China, December 1-4, 2015

Design Framework for practical optimum nonlinear filter for unknown noise characteristics

Yukihiro Tadokoro†

†TOYOTA Central R&D Labs., Inc.
41-1, Nagakute, Aichi 480-1192, JAPAN
Email: tadokoro@mosk.tytlabs.co.jp

Abstract—A nonlinear filtering method was previously proposed that can estimate a weak signal buried in strong non-Gaussian noise. This method is useful in signal processing because it can realize the Cramér-Rao lower bound in the estimation. To determine the filter characteristics, a mathematical expression for the probability density function (PDF) of the noise is necessary. The original study assumed that the PDF was known, but in practical situations, this assumption is not true. The present study considers the use of the filter in a situation where the PDF is unknown. A method for estimating the PDF is presented, together with the corresponding filter function. Kernel density estimation is employed using Epanechnikov kernel in order to reduce the computational complexity, and the optimum bandwidth (at which the estimation performance is maximized) is derived. Through numerical evaluation, the proposed method is confirmed to be effective.

1. Introduction

The estimation of a weak signal buried in strong noise has been well discussed in the signal processing field [1,2]. The theoretical limit, which is called the Cramér-Rao lower bound, states that the estimation error is not less than the inverse of the Fisher information [3, 4]. Our previously proposed method [5, 6] is attractive since it can achieve the bound in the case of non-Gaussian noise cases. From the input signal x, which contains the weak signal and the noise, a device/algorithm with the following in-out function can output the signal component:

$$F_{opt}(x) = a - b\left\{\frac{\partial}{\partial n}\log\rho(n)\right\}_{n=x}. \quad (1)$$

Note that $\rho(n)$ is the probability density function (PDF) for the noise, and a is constant and b is non-zero constant.

Since Eq. (1) involves a derivative of the noise PDF, an estimated mathematical expression for the noise PDF is required. In our previous studies, the PDF was assumed to be known. However, in practical situations, the noise characteristics depend on the surrounding environment, and hence should be estimated from noise samples before filtering the noisy input.

The present paper proposes an estimation framework for the noise PDF and the corresponding in-out function. The estimation method is optimized for the in-out function $F_{opt}(x)$ in order to satisfy the following requirements: 1) the estimation performance should be high enough to maximize signal-to-noise ratio (SNR) at the output, 2) the proposed method can be applied to any type of white noise, and 3) the estimated PDF is differentiable. The function in Eq. (1) was derived for the purpose of maximizing the output SNR. In this sense, the method proposed in present study should also yield a maximum SNR. To improve its validity, we focus on non-parametric estimation because the method should be applicable to any type of white noise. The third point listed above is obvious since the function is obtained by taking a differential.

Many types of non-parametric estimation methods have been discussed, including those based on the traditional histogram, kernel density estimation (KDE) [7, 8], and the characteristic kernel [9–11]. The method in the present study is based on the kernel density method because A) the estimated function is differentiable if the Epanechnikov kernel is employed, B) the computational complexity using the Epanechnikov kernel is reduced to half that for other kernels, and C) the optimum bandwidth, which gives the maximum output SNR, is analytically derived. It is well known that the performance of KDE depends on the "bandwidth". The optimum value has been theoretically derived [12], but it cannot be calculated because the original PDF $\rho(n)$ is required. Sub-optimum methods have also been proposed. In many cases, a Gaussian kernel is assumed, and/or additional cost is required (e.g. the plug-in method) [13, 14]. As will be discussed in Sec. 3, the proposed method exploits the characteristic of the in-out function $F_{opt}(x)$, which gives a simple solution for optimum parameter tuning. The numerical evaluation enhances the effectiveness of the proposed method.

2. Proposed filtering system with PDF estimation

A schematic diagram of the proposed method is shown in Fig. 1. The nonlinear filter has an input x_i which contains a weak signal s_i and white noise n_i. The subscript i represents the time index. The filter extracts the weak signal component from the noisy input, and then outputs the signal y_i. The filter function $\hat{F}_{opt}(x)$ is calculated based on estimated noise PDF $\hat{\rho}(n)$. For simplicity, it is assumed that a large

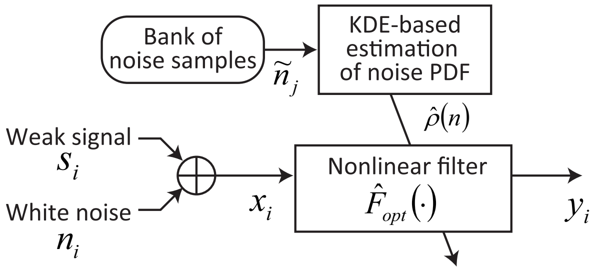

Figure 1: Schematic diagram of proposed system with noise PDF estimation.

number of noise samples $\tilde{n}_j$ are measured in advance, and n_i and $\tilde{n}_j$ have the same PDF $\rho(n)$. In addition, to consider the situation where a weak signal is buried in the noise, the noise intensity is large compared to the signal power.

The KDE method estimates the PDF by using a kernel function $K(u)$ such as,

$$\hat{\rho}(n) = \frac{1}{Nh}\sum_{j=1}^{N} K\left(\frac{n-\tilde{n}_j}{h}\right) \tag{2}$$

where N is the number of the noise samples $\tilde{n}_j$, and h is bandwidth, which is the key parameter for the estimation performance [7]. Substituting Eq. (1) into Eq. (2), we have

$$\hat{F}_{opt}(x_i) = a - b\frac{\sum_{j=1}^{N}\frac{\partial}{\partial x_i}K\left(\frac{x_i-\tilde{n}_j}{h}\right)}{\sum_{j=1}^{N}K\left(\frac{x_i-\tilde{n}_j}{h}\right)}. \tag{3}$$

This expression indicates that to calculate the function, a computation of the order $2N$ is required, which is denoted by $O(2N)$. Each of the summations in Eq. (3) requires N addition operations.

The proposed method employs the Epanechnikov kernel which is expressed as,

$$K(u) = \frac{3}{4}(1-u^2)U(|u| \le 1) \tag{4}$$

where

$$U(|u| \le \alpha) = \begin{cases} 1 & (|u| \le \alpha) \\ 0 & (|u| > \alpha) \end{cases} \tag{5}$$

and α is constant. From Eq. (5), the kernel is discontinuous at the point $|u| = \alpha$; thus, the third requirement given in Sec. 1 is not satisfied. To obtain a differentiable kernel, a small parameter ϵ is now introduced. In the region $\tilde{n}_j - h + \epsilon \le x_i \le \tilde{n}_j + h - \epsilon$, the kernel function $K(\frac{x_i-\tilde{n}_j}{h})$ is continuous, i.e., differentiable. Substituting Eq. (4) into Eq. (3) gives the proposed function as

$$\hat{F}_{opt}(x_i) = a + \frac{b}{h}\left(\frac{1}{1-\frac{x_i-\mu_{x_i}}{h}} - \frac{1}{1+\frac{x_i-\mu_{x_i}}{h}}\right). \tag{6}$$

The variable μ_{x_i} denotes the averaged value of the noise samples $\tilde{n}_j$ included in the region $\tilde{n}_j - h + \epsilon \le x_i \le \tilde{n}_j + h - \epsilon$.

It may be thought that since other kernels such as Gaussians are differentiable, the merit of Eq. (6) is not apparent. However, one of the advantages of using the Epanechnikov kernel is that the computational complexity can be reduced by a factor of two. As previously mentioned, KDE-based functions generally require $O(2N)$ computations. For example, in the case of a Gaussian kernel, the derivative is an exponential function, and then Eq. (3) involves two different summations of exponential functions. Each requires $O(N)$ computations, and the resulting function does not have a simple form like the in Eq. (6). Due to the use of the second-order kernel, the function in Eq. (6) is derived, and the filter output can be obtained only by calculating the average of the noise samples, μ_{x_i}. Such an operation requires $O(N)$ computations, which means that the computational complexity of the proposed method is half that for other kernels.

3. Optimization of the bandwidth

The estimation performance in KDE depends on the bandwidth. Indeed, the function in Eq. (6) has an implicit bandwidth dependence due to the presence of μ_{x_i}. In this section, a method is proposed for setting the optimum bandwidth in order to obtain a filtering performance close to the theoretical one.

The optimum bandwidth h_{opt} can be derived by minimizing the asymptotic mean square error (AMSE)

$$h_{opt} = \left(\frac{R(k)}{\beta_2^2(k)R(\rho(n))}\right)^{\frac{1}{5}} N^{-\frac{1}{5}}, \tag{7}$$

where $R(z) = \int z^2(v)dv$ and $\beta_2(k)$ is the second moment of the kernel function [7]. In practice, the value of h_{opt} cannot be calculated because the original PDF $\rho(n)$ is required.

A method optimized for the function in Eq. (1) is proposed in this study. This is obtained by exploiting the characteristics of the filter; the optimum function $F_{opt}(x)$ gives the maximum output SNR. From this viewpoint, the optimum bandwidth $\hat{h}_{opt}$ can be derived using the following criterion:

$$\hat{h}_{opt} = \arg\max_{h}\,[\gamma]\,. \tag{8}$$

Note that $\gamma = \frac{\langle \frac{\partial}{\partial x}\hat{F}_{opt}(x)\rangle^2}{\langle \hat{F}^2_{opt}(x)\rangle - \langle \hat{F}_{opt}(x)\rangle^2}$ denotes the normalized output SNR and $\langle z\rangle = \frac{1}{N}\sum_{i=1}^{N} z_i$ [6].

Equation (8) is unfortunately not solvable since it is impossible to take the derivative of the estimated function $\hat{F}_{opt}(x)$ (more precisely, the variable μ_{x_i} is not differentiable). An alternative method is now introduced using the relation between the output SNR and the in-out correlation. The correlation C between the weak signal and the filter output is defined as

$$C = \frac{\frac{1}{N}\sum_{i=1}^{N}(s_i - \langle s\rangle)(y_i - \langle y\rangle)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i-\langle s\rangle)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\langle y\rangle)}}. \tag{9}$$

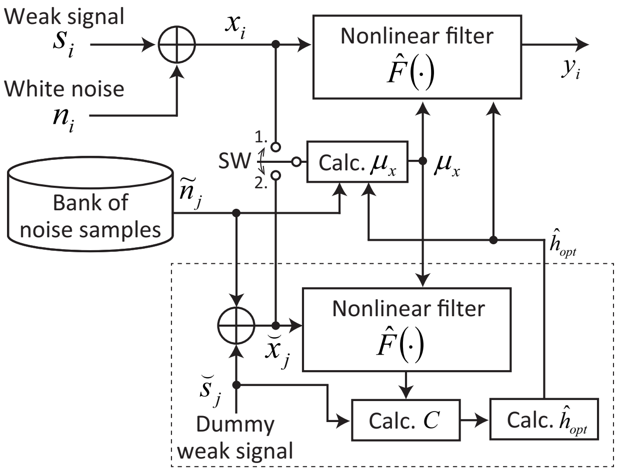

Figure 2: Schematic diagram of proposed filter with optimum bandwidth calculation.

It is considered that the signal s_i is sufficiently small compared to the noise n_i. In this situation, a Taylor expansion can be applied to the filter output [6], so the correlation is simply expressed as,

$$C \approx \left[1 + \frac{1}{P_s \gamma}\right]^{-\frac{1}{2}}. \tag{10}$$

For a signal power $P_s = 1.0$, the optimum bandwidth is obtained as,

$$\hat{h}_{opt} = \arg\max_h \left[\frac{1}{\frac{1}{C^2} - 1}\right] = \arg\max_h \left[C^2\right]. \tag{11}$$

Figure 2 presents a schematic diagram of the proposed method with an optimum bandwidth estimation. The key point in this approach is that the correlation is calculated using a dummy signal $\breve{s}_j$, which is introduced to the filter instead of the original weak signal s_i. From Eq. (9), the correlation should be calculated using the original weak signal, but this is obviously unknown on the filter side. Owing to the lack of dependence of Eq. (1) on the shape of the weak signal, any weak signal can be chosen to represent s_i. The proposed method consists of two steps. In the first step, the switch “SW” is set to “2.”, and the optimum bandwidth is calculated based on Eq. (11). In the next step, the switch is changed to “1.” and the value of μ_{x_i} corresponding to the input signal x_i is calculated. Then, the filter function Eq. (6) is obtained, and finally, the weak signal component is extracted from the noise input x_i.

4. Numerical examples

The aim of employing the filter is to extract a weak signal buried in strong noise. In this section, the performance of the proposed method is numerically evaluated in terms of the output SNR and the bandwidth.

Table 1: Evaluation parameter settings.

Parameter	Value
White noise	Mixed Gaussian noise
Mean	$\nu_1 = 2.0$, $\nu_2 = -1.0$
Variance	$\sigma_1^2 = 2.0$, $\sigma_2^2 = 1.0$
Weight parameter	$\beta = 0.5$
Weak signal	Sinusoidal
Frequency	$f_c = 100$[Hz]
Amplitude	$A = 0.10$[V]
Num. of noise samples	$N = 30000$

Table 2: Numerical example of optimum bandwidth and output SNR.

	Estimate	Theory
Optimum bandwidth	0.140	0.406
Output SNR [dB]	-2.65055	-2.65040

The settings used for the evaluation are as follows. Mixture Gaussian noise is used for the white noise n_i and $\tilde{n}_j$ because the filter in Eq. (1) is valid for non-Gaussian noise. The PDF is given by $\rho(n) = \beta\Psi(\nu_1, \sigma_1^2) + (1-\beta)\Psi(\nu_2, \sigma_2^2)$, where β is a weight parameter, and $\Psi(\nu, \sigma^2)$ represents a Gaussian PDF with a mean ν and a variance σ^2. The weak signal is sinusoidal with a frequency f_c and an amplitude A. The parameter values are given in Table 1. The power of the weak signal is confirmed to be sufficiently small compared to the white noise. The N noise samples $\tilde{n}_j$ are numerically generated, and the noisy input samples x_i are then filtered out to obtain the output. The filter parameter, a and b are set as 0 and 1, respectively.

The values of the output SNR and the optimum bandwidth are shown in Table 2. These values were obtained by averaging the results of 50 trials. The theoretical bandwidth was calculated based on Eq. (7). From this table, the output SNR for the proposed method is almost the same as the theoretical value, indicating that it is likely to be an effective filtering method.

Although the output SNR is very close to the theoretical value, the estimated bandwidth is different to the theory. Figure 3 shows the dependence of the bandwidth on the output SNR. For reference, the input SNR is also shown. It can be seen that the output SNR does not change very much in the region $0.05 \le h \le 1.00$; an output SNR close to the theoretical value can be achieved for any bandwidth in this region. Even though there is a difference between the estimated and theoretical bandwidths in Table 2, this is not a problem in terms of the filtering performance. Note that the effectiveness of Eq. (1) is confirmed because the output SNR is improved compared to the input.

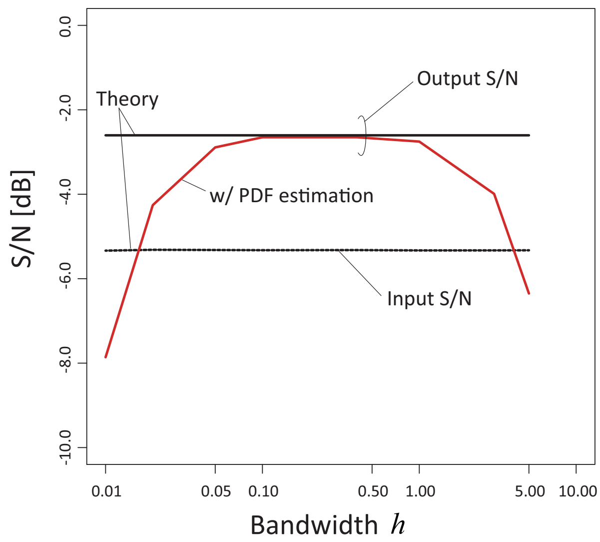

Figure 3: Dependence of output SNR on the bandwidth.

5. Conclusion

The present paper described a method for estimating the noise PDF and the corresponding in-out function. The proposed method is based on KDE, which is a nonparametric estimation method, thus allowing it to be applied to any type of white noise. Due to the use of a second-order Epanechnikov kernel, the computational complexity was reduced to half of that for other kernels. The optimum bandwidth for achieving the maximum output SNR was also derived. The effectiveness of the proposed method was confirmed by a numerical evaluation, which implies that it is likely act as a practical filtering system.

Acknowledgments

The author thanks Assist. Prof. Ichiki in Nagoya University, JAPAN for the beneficial comments.

References

[1] A. V. Oppenheim, R. W. Schafer, and John R. Buck, *Discrete-time signal processing*, Prentice-hall, third edition, 2009.

[2] S. O. Haykin, *Adaptive Filter Theory*, Prentice-hall, fifth edition, 2013.

[3] H. Cramér, *Mathematical Methods of Statistics (PMS-9)*, Princeton Landmarks in Mathematics and Physics Series. Princeton University Press, 1945.

[4] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of Cal. Math. Soc.*, vol. 37, no. 3, pp. 81–91, 1945.

[5] A. Ichiki and Y. Tadokoro, "Relation between optimal nonlinearity and non-Gaussian noise: Enhancing a weak signal in a nonlinear system," *Physical Review E*, vol. 87, no. 1, pp. 012124, Jan. 2013.

[6] Y. Tadokoro and A. Ichiki, "A Simple Optimum Nonlinear Filter For Stochastic-resonance-based Signal Detection," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013, pp. 5760–5764.

[7] B. W. Silverman, *Density estimation for statistics and data analysis*, Springer, 1986.

[8] P. V. Kerm, "Adaptive kernel density estimation," *Stata Journal*, vol. 3, no. 2, pp. 148–156, 2003.

[9] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Scholkopf, and G. R. G. L. Lanckriet, "Hilbert Space Embeddings and Metrics on Probability Measures," *The Journal of Machine Learning Research*, vol. 11, pp. 1517–1561, 2010.

[10] A. Gretton, K. M. Borgwardt, M. Rasch, B. Scholkopf, and A. J. Smola, "A Kernel Method for the Two-Sample-Problem," in *Proceedings of the Advances in neural information processing systems*, 2006, pp. 513–520.

[11] A. Gretton, K. Fukumizu, Z. Harchaoui, and B. K. Sriperumbudur, "A Fast , Consistent Kernel Two-Sample Test," in *Proceedings of the Advances in neural information processing systems*, 2009, pp. 673–681.

[12] M. C. Jones, J. S. Marron, and S. J. Sheather, "A Brief Survey of Bandwidth Selection for Density Estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, Mar. 1996.

[13] Simon J. Sheather, "Density Estimation," *Statistical Science*, vol. 19, no. 4, pp. 588–597, Nov. 2004.

[14] Vikas Chandrakant Raykar and Ramani Duraiswami, "Fast optimal bandwidth selection for kernel density estimation," in *Proceedings of the 2006 SIAM International Conference on Data Mining*. Apr. 2006, Society for Industrial and Applied Mathematics.