OPTIMAL ESTIMATION OF A GAUSSIAN RANDOM-VARIABLE IN GAUSSIAN-MIXTURE NOISE AND ASSOCIATED ERROR ANALYSIS

YONINA ELDAR and ARIE YEREDOR
Department of Electrical Engineering - Systems
Tel-Aviv University, Tel-Aviv, 69978, ISRAEL
e-mail: yonina@eng.tau.ac.il arie@eng.tau.ac.il

Abstract — **In this paper we derive the optimal estimator for a Gaussian random variable contaminated by impulsive noise, which we model as a mixture of zero-mean Gaussian components. We provide an analytical derivation of the resulting mean squared error, and compare the performance to that of the optimal linear estimator. We identify cases of interest, where the improvement is substantial.**

keywords - outliers, Gaussian Mixture, ZMGM noise, optimal estimation.

1 Introduction

Numerous methods exist (e.g. [1], [2]) for dealing with impulsive noise in the context of filtering. A common approach is to use pre-processing hard limiters, thus practically discarding the outliers. However, a hard-limiting operation may be far from optimal. Other approaches use statistical models to describe the impulsive behavior of the noise, such as alpha-stable models (e.g. [2]). The drawback of these approaches is the relative complexity of both the analytical derivations and implementations involved.

Gaussian mixture (GM) modeling is popular in the signal-processing community mainly in the context of speech recognition; however, little research effort has been invested in GM modeling of time-series for the purpose of filtering. In this paper we propose to model the impulsive noise as a GM random variable (r.v.) of zero-mean components (denoted Zero-Mean GM (ZMGM)). This modeling is appealing due to several reasons: the popular zero-mean Gaussian model is a special case of ZMGM; the sum of ZMGM (and/or Gaussian) r.v.'s is also a ZMGM r.v.; when a ZMGM process undergoes linear filtering, the output is also a ZMGM process. In addition, a ZMGM model is most appropriate for describing outlier situations, by assuming mixtures of two components, where the first component occurs with a high probability, and the second component has a significantly larger variance and occurs with a small probability.

We confine the discussion in this paper to the case of independent, identically distributed (iid) time-series (signal and noise), in which the optimal filtering is equivalent to optimal memoryless (scalar) r.v. estimation. We derive the optimal estimator and provide an analytical derivation for evaluating the resulting mean squared error (mse), which we use to identify cases of interest.

We provide an extension of the results to correlated time-series in [3].

2 Optimal Estimation of a Gaussian r.v. in ZMGM Noise

Let x denote a $\mathcal{N}(0,\sigma_x^2)$ r.v.. It is desired to estimate x from the measurement $y=x+v$ where v is an impulsive noise (independent of x), modeled as a Gaussian Mixture with M zero-mean components (i.e. ZMGM) of known variances $\sigma_1^2<\sigma_2^2<\cdots<\sigma_M^2$ appearing with probabilities $p_1,p_2,\ldots p_M$ respectively. The noise probability distribution function (p.d.f.) is given by:

$$f_v(v)=\frac{1}{\sqrt{2\pi}}\sum_{m=1}^{M}\frac{p_m}{\sigma_m}\exp(-\frac{v^2}{2\sigma_m^2}) \qquad (1)$$

where $\sum_{m=1}^{M}p_m=1$.

This p.d.f. can be given the following interpretation: Let I denote an auxiliary r.v. taking values from 1 to M, indicating the Gaussian component from which v is drawn. Thus, $f_{v|I}(v|I=m)$ is $\mathcal{N}(0,\sigma_m^2)$ $(m=1,2,\ldots M)$. Consequently, $f_v(v)$ can be expressed as $\sum_{m=1}^{M}f_{v|I}(v|I=m)P(I=m)$.

Since x is zero-mean Gaussian, the measurement $y=x+v$ is also a ZMGM r.v., taking variances $s_m^2\triangleq\sigma_x^2+\sigma_m^2$ $(m=1,2,\ldots M)$ with probabilities p_m (respectively).

The optimal minimum mean squared error (mse) estimator $\hat{x}(y)$ of x is well-known to be the conditional expectation $E[x|y]=E[E[x|y,I=m]]$. Noting that given $I=m$, x and y are jointly Gaussian, we have

$$E[x|y,I=m]=(\sigma_x^2/s_m^2)y\triangleq h_m y\triangleq\hat{x}_m(y) \qquad (2)$$

$\hat{x}_m(y)$ is actually the optimal (linear) estimator of x assuming that $I=m$, or equivalently, that $v\sim\mathcal{N}(0,\sigma_m^2)$. Denoting by $q_m(y)=P(I=m|y)$ the posterior probability of $I=m$ given y, we have:

$$\hat{x}(y) \quad = \quad E[x|y] \; = \; \sum_{m=1}^{M}q_m(y)\hat{x}_m(y)$$

$$= \sum_{m=1}^{M} \frac{f_{y|I}(y|I=m)P(I=m)}{f_y(y)}\hat{x}_m(y)$$
$$= \frac{\sum_{m=1}^{M} \frac{p_m}{s_m}\exp(-\frac{y^2}{2s_m^2})h_m}{\sum_{k=1}^{M} \frac{p_k}{s_k}\exp(-\frac{y^2}{2s_k^2})}y. \quad (3)$$

Thus, the optimal estimator can be viewed as a weighted combination of the (conditional) optimal linear estimators $\hat{x}_1, \hat{x}_2, \ldots \hat{x}_M$ defined in (2). However, the weighting coefficients introduce non-linearity in the measurement y. To visualize the effect we may compare that estimator to the optimal linear estimator

$$\hat{x}_L(y) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_v^2}y \quad (4)$$

where $\sigma_v^2 = \sum_{m=1}^{M} p_m\sigma_m^2$ denotes the noise variance.

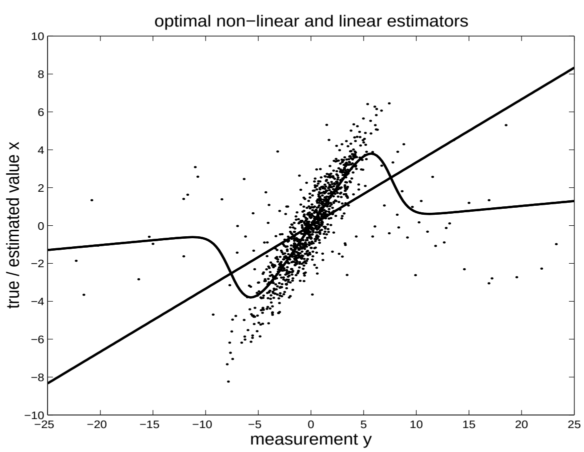

Figure 1: Optimal non-linear and linear estimators of x from $y = x + v$, superimposed on 500 realizations.

Figure 1 demonstrates the behavior of the two estimators superimposed on 500 realizations of true values (x) vs. measurements (y) for the case of $M = 2$ with $p_1 = 0.9, \sigma_1 = 1, p_2 = 0.1, \sigma_2 = 10$ and an SNR of $\sigma_x^2/\sigma_v^2 = 0.5$. It is seen that the linear estimator attempts to compensate for occurrences of outliers by using a moderate slope, which in turn misaligns with the more probable population in the center. The optimal estimator, on the other hand, uses a weighted combination of two linear estimators $\hat{x}_1(y)$ and $\hat{x}_2(y)$, such that the slope of each is properly aligned with the population appearing in its region of domination. For example, the region of domination of $\hat{x}_2(y)$ is the region of large values of y, where $q_2(y) \approx 1$. The resulting optimal estimator is reminiscent of a hard-limiter, which is a popular tool in robust estimation in the presence of impulsive noise. However, while a hard-limiter completely discards suspected outliers, the optimal estimator gradually decreases its sensitivity as the posterior probability of an outlier increases.

3 Estimation Error Analysis

In this section we present an analysis of the mse attained by the optimal estimator $\hat{x}(y)$ for the case of two mixture components ($M = 2$). The mse is given by:

$$E[(\hat{x}(y) - x)^2] = E[E[(\hat{x}(y) - x)^2|y]] = \sigma_x^2 - E[\hat{x}^2(y)]. \quad (5)$$

Thus, evaluating the mse involves the evaluation of $E[\hat{x}^2(y)]$, which in the case $M = 2$ reduces to:

$$E[\hat{x}^2(y)] = \quad (6)$$
$$\int_{-\infty}^{\infty}\left(\frac{\frac{p_1}{s_1}\exp(-\frac{y^2}{2s_1^2})h_1 + \frac{p_2}{s_2}\exp(-\frac{y^2}{2s_2^2})h_2}{\frac{p_1}{s_1}\exp(-\frac{y^2}{2s_1^2}) + \frac{p_2}{s_2}\exp(-\frac{y^2}{2s_2^2})}\right)^2 y^2 f_y(y)dy$$

To simplify the exposition, let us define the constants $\gamma_i = p_i/s_i$ $(i = 1, 2)$ and $\frac{1}{s^2} = \frac{1}{s_1^2} - \frac{1}{s_2^2}$.
Noting that the denominator in parenthesis in (6) equals $\sqrt{2\pi}f_y(y)$ and defining $g_i(y) = \gamma_i h_i \exp(-\frac{y^2}{2s_i^2})$ $(i = 1, 2)$, we may break (6) down into three terms denoted $\frac{1}{\sqrt{2\pi}}(T_{11} + 2T_{12} + T_{22})$, where

$$T_{ij} = \int_{-\infty}^{\infty} \frac{g_i(y)g_j(y)}{\sqrt{2\pi}f_y(y)}y^2 dy \triangleq t_{ij}T'_{ij} \quad i, j = 1, 2 \quad (7)$$

here $t_{ij} = \gamma_i h_i \gamma_j h_j/\gamma_2$ and $T'_{ij} = \int_{-\infty}^{\infty} \frac{\exp(-a_{ij}y^2)}{1+b\exp(-cy^2)}y^2 dy$ where $b = \gamma_1/\gamma_2$, $c = \frac{1}{2s^2}$ and $a_{ij} = \frac{1}{2}(\frac{1}{s_i^2} + \frac{1}{s_j^2} - \frac{1}{s_2^2})$ (for $i, j = 1, 2$).

Evaluating T_{ij} amounts to evaluating T'_{ij}, for which there is no known closed-form solution. Nevertheless, we may get rid of the denominator in T'_{ij} by exploiting the relations

$$\frac{1}{1+z} = \begin{cases} \sum_{n=0}^{\infty}(-1)^n z^n & |z| < 1 \\ \sum_{n=1}^{\infty}(-1)^{n+1}z^{-n} & |z| > 1 \end{cases} \quad (8)$$

as follows: Define $z = b\exp(-cy^2)$. Assuming $p_1 > p_2$ (which is a common assumption in an outlier situation), we have $b > 1,\ c > 0$, and we may therefore use $y_{th} \triangleq \sqrt{\frac{1}{c}\log(b)}$ to partition the integration domain into two regions: $D_1 : \{y : \ |y| > y_{th}\}$ and $D_2 : \{y : \ |y| < y_{th}\}$ in which $|z| < 1$ and $|z| > 1$, respectively. We may now evaluate T'_{ij} using

$$\int_{D_k} y^2 \exp(-\alpha y^2)dy = \frac{1}{\alpha\sqrt{\alpha}} \cdot \quad (9)$$
$$\cdot \begin{cases} \frac{\sqrt{\pi}}{2}\mathrm{erfc}(\sqrt{\alpha}y_{th}) + \sqrt{\alpha}y_{th}\exp(-\alpha y_{th}^2) & k = 1 \\ \frac{\sqrt{\pi}}{2}\mathrm{erf}(\sqrt{\alpha}y_{th}) - \sqrt{\alpha}y_{th}\exp(-\alpha y_{th}^2) & k = 2 \end{cases}$$

where $\mathrm{erf}(t) = \frac{2}{\sqrt{\pi}}\int_0^t \exp(-w^2)dw$ and $\mathrm{erfc}(t) = 1 - \mathrm{erf}(t)$.

Let us also define $Q_k(\alpha)$ $k = 1, 2$ as the two possible result of (9). Combining (8) and (9), we may express

T'_{ij} as an infinite sum as follows:

$$T'_{ij} = \sum_{n=0}^{\infty}(-1)^n(\frac{\gamma_1}{\gamma_2})^n Q_1(a_{ij}+nc) - \sum_{n=1}^{\infty}(-1)^n(\frac{\gamma_1}{\gamma_2})^{-n} Q_2(a_{ij}-nc). \quad (10)$$

$a_{ij} - nc$ in the second term of (10) becomes negative for values of n beyond a certain threshold. When $Q_2(\cdot)$ obtains a negative value as its argument, $\text{erf}(\cdot)$ of an imaginary argument has to be computed. The definition for $\text{erf}(t)$ stated below (9) holds true for any complex argument t. However, standard tables and routines for evaluating $\text{erf}(t)$ are usually available for real-valued t's only (e.g. in MATLAB). We therefore provide below the following sum for evaluating $\text{erf}(\tilde{t})$ for imaginary values of $\tilde{t} = it$ (here $i = \sqrt{-1}$ and t is real-valued):

$$\text{erf}(it) = \frac{i}{\pi}[t + 2\sum_{n=1}^{\infty}\frac{1}{n}\exp(-\frac{1}{4}n^2)\sinh(nt)] \quad (11)$$

(see e.g. [4] p. 299).

Thus, the mse of $\hat{x}(y)$ can be calculated analytically to within arbitrary precision by taking sufficiently many terms in (10). In Fig. 2 we demonstrate (using the analytical results, verified by simulations) the relative improvement in mse attained by the optimal estimator over the optimal linear estimator (whose mse is given by $\sigma_x^2\sigma_v^2/(\sigma_x^2+\sigma_v^2)$) for three SNRs (low (0.5), medium (1.0) and high (5.0)). For each SNR, six curves are presented (for $\sigma_2/\sigma_1 = 2, 5, 10, 15, 20, 25$), showing the relative improvement as a function of the outlier probability p_2. It is evident that the improvement increases as σ_2/σ_1 increases, but attains an optimum as a function of p_2. The peaks become sharper and lower as the SNR increases.

It is interesting to note that in the inverse situation, when it is desired to estimate the noise component, the optimal estimator is the complementary estimator, i.e. $\hat{v}(y) = y - \hat{x}(y)$. This holds true since $y = x + v \Rightarrow E[v|y] = E[y - x|y] = y - E[x|y]$. Consequently, the resulting mse's are the same in both cases, since $\hat{v}(y) - v = -(\hat{x}(y) - x)$. Since this property is also shared by the optimal linear estimator, the relative improvement remains the same. This observation is useful when analyzing the case of an impulsive (ZMGM) iid time-series contaminated by iid Gaussian noise.

4 Conclusion

We presented the optimal (non-linear) estimator and associated error analysis for estimating a Gaussian r.v. from its measurement contaminated by impulsive noise modeled as a ZMGM r.v.. The estimator can be interpreted as a non-linear weighting of optimal linear estimators, each suited to a corresponding mixture component. The weighting reflects the posterior probability of occurrence of the respective components. In extreme outlier situations the optimal estimator resembles a hard-limiter; however, its advantage is in its ability to deal properly with moderate outlier situations without discarding data on one hand, and without compromising performance in "benign" (no outliers) situations (as does the optimal linear estimator) on the other hand.

We demonstrated via error analysis for the case of two mixture components, that as may be expected, the attainable improvement in performance (over the optimal linear estimator) becomes more significant as the variance ratio of the two components increases. However, an optimum is attained as a function of the outlier probability and SNR.
An extension of our results to the case of filtering correlated time-series can be found in [3].

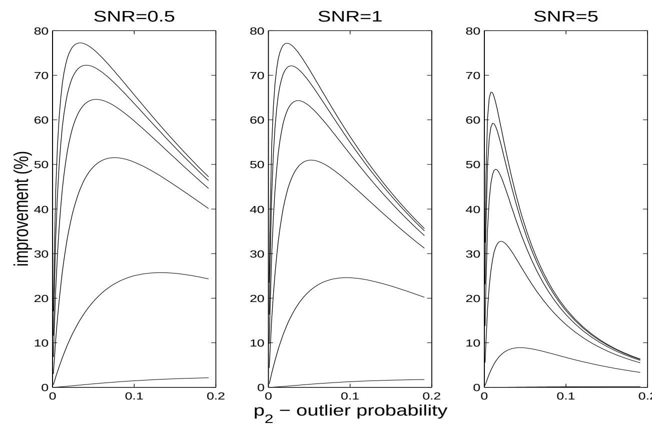

Figure 2: Relative mse improvement [%] attained by the non-linear estimator over the linear for three SNRs : low, medium, high. Curves in each figure show the improvement as a function of the outlier probability p_2, parameterized by $\sigma_2/\sigma_1 = 2, 5, 10, 15, 20, 25$ (curves are monotonically ordered in σ_2/σ_1, with the highest corresponding to $\sigma_2/\sigma_1 = 25$).

References

[1] S. Kassam and V. Poor, "Robust Techniques for Signal Processing: A Survey," *IEEE Proc.*, vol. 73 no. 3, pp. 433–481, Mar. 1985.

[2] M. Shao and C. Nikias, "Signal Processing with Fractional Lower Order Moments: Stable Processes and their applications," *IEEE Proc.*, vol. 81 no. 7, pp. 986–1009, July 1993.

[3] Y. Eldar and A. Yeredor, "Optimal and Sub-Optimal Filtering of Gaussian Signals in Gaussian-Mixture Noise," submitted to *IEEE Signal Processing Letters*, (Mar. 1998).

[4] I. Stegun and M. Abramowitz, *Handbook of Mathematical Functions*, Dover Publications, NY 1972.