FINITE-MEMORY SUBOPTIMAL FILTERING OF GAUSSIAN SIGNALS IN CORRELATED GAUSSIAN-MIXTURE NOISE

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 develop a sub-optimal (nearly optimal) finite-memory filter for estimating a stationary Gaussian signal contaminated by colored impulsive noise. The noise is modeled as the output of an FIR filter, whose input is an independent, identically distributed sequence of Gaussian Mixture random variables with zero-mean components (denoted ZMGM). The resulting filter is a non-linearly weighted combination of linear filters, for which a computationally efficient architecture is proposed.**

keywords - outliers, Gaussian Mixture, ZMGM noise, AR process, filtering.

1 Introduction

A common approach for dealing with impulsive noise in the context of filtering is to use pre-processing hard limiters, thus practically discarding the outliers (e.g. [1]). However, when either signal or noise involved have a substantial correlation length, such 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 filter implementations involved.

Gaussian mixture (GM) modeling is appropriate for describing outlier situations, where large deviations occur with a small probability. In this paper we allow for an impulsive colored noise modeled as a filtered sequence of independent, identically distributed (iid) GM random variables with zero-mean components (denoted ZMGM). A ZMGM random variable (r.v.) has the following probability distribution function (p.d.f.):

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

where M is the number of zero-mean Gaussian components of known variances $\sigma_1^2 < \sigma_2^2 < \cdots < \sigma_M^2$ appearing with probabilities $p_1, p_2, \ldots p_M$ respectively, with $\sum_{m=1}^{M} p_m = 1$. By exploiting the fact that the sum of independent ZMGM (and/or Gaussian) r.v.'s is also a ZMGM r.v., we develop a fixed-memory filter for estimating a stationary Gaussian signal from measurements corrupted by additive ZMGM noise.

2 Filtering of Gaussian Signals in Colored ZMGM Noise

Let $x[n]$ denote the desired signal, which is stationary zero-mean Gaussian with correlation function $R_{xx}[l]$. It is desired to filter (estimate) $x[n]$ from the noisy measurements $y[n] = x[n] + v[n]$, where $v[n]$ denotes the noise, modeled as a correlated ZMGM process, independent of $x[n]$. The correlated ZMGM noise is created by an iid sequence $w[n]$ of ZMGM r.v.'s passing through an FIR of length $K+1$. For simplicity, we shall assume that $w[n]$ is ZMGM with $M = 2$ mixture components of variances $\sigma_1^2 < \sigma_2^2$ appearing with probabilities p_1 and p_2 respectively. For notational brevity, we denote by p the outlier probability, thus $p_2 = p$, $p_1 = 1 - p$.

We restrict the discussion to finite-memory filters of length L, thus it is desired to filter $x[n]$ from the preceding L observations of $y[n]$, i.e. from the vector $\underline{y}_n = [\ y[n] \quad y[n-1] \quad \cdots \quad y[n-L+1]\ ]^T$.

In [3] it is shown that in the case of memoryless estimation with $v[n]$ ZMGM with $M = 2$, the improvement attained by using optimal estimation (over optimal linear estimation) is substantial for small values of p. We therefore focus our discussion on such cases in the context of finite-memory filtering as well. We further assume that $(L+K)p^2 \ll 1$, such that the probability of *multiple* occurrences of outliers in $w[n]$ influencing $\underline{y}_n$ is negligible. Thus, rather than developing the optimal estimator, we shall develop a sub-optimal estimator, which is approximately optimal under the (highly probable) assumption that no more than one outlier occurred in the recent $K+L$ samples of $w[n]$.

To derive our sub-optimal estimator of $x[n]$, we denote by I_n an auxiliary r.v. taking the values 1 or 2, which indicate the Gaussian component from which $w[n]$ is drawn, such that $f_{w[n]|I}(w[n]|I = m)$ is $\mathcal{N}(0, \sigma_m^2)(m = 1, 2)$. We further define a vector of indicators $\underline{I}_n = [\ I_n \quad I_{n-1} \quad \cdots \quad I_{n-(K+L-1)}\ ]^T$ composed of component indicators of $w[n], w[n-1], \ldots w[n-(K+L-1)]$ respectively.

The optimal minimum mean squared error (mse) estimator of $x[n]$ is well known to be the conditional expectation $E[x[n]|\underline{y}_n] = E[E[x[n]|\underline{y}_n, \underline{I}_n]]$, where $\underline{I}_n$ can take 2^{K+L} values. However, for our sub-optimal estimator we consider only the $K+L+1$ most probable values of $\underline{I}_n$, namely $\underline{I}^0, \underline{I}^1, \ldots \underline{I}^{K+L}$, where $\underline{I}^0 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ and $\underline{I}^k$, $k = 1, 2, \ldots K+L$ are all-ones vectors with a 2 at the k 'th entry (the remaining possibilities will be indexed implicitly as $\underline{I}^k$, $k = K+L+1, \ldots 2^{K+L}-1$).

Given $\underline{I}_n = \underline{I}^k$, $x[n]$ and $\underline{y}_n$ are jointly Gaussian. Thus,

$$E[x[n]|\underline{y}_n, \underline{I}_n = \underline{I}^k] = \underline{r}_{xy}(R^k_{yy})^{-1}\underline{y}_n \triangleq \hat{x}_k[n] \qquad (2)$$

where $R^k_{yy} = R_{xx} + R^k_{vv}$ and $\underline{r}_{xy} = \underline{r}_{xx} = \begin{bmatrix} R_{xx}[0] & R_{xx}[1] & \cdots & R_{xx}[L-1] \end{bmatrix}$. Here R_{xx} is the $L \times L$ correlation matrix of $x[n]$, and R^k_{vv} is the correlation matrix of $v[n]$ given $\underline{I}_n = \underline{I}^k$, which can be derived as follows:

Denote by

$$G = \begin{bmatrix} g[0] & g[1] & \cdots & g[K] & 0 & \cdots & 0 \\ 0 & g[0] & \cdots & & g[K] & \ddots & \vdots \\ \vdots & \ddots & & & & & 0 \\ 0 & \cdots & 0 & g[0] & g[1] & \cdots & g[K] \end{bmatrix} \qquad (3)$$

the $L \times (K+L)$ Toeplitz matrix composed of the noise-generating FIR's impulse response coefficients, $g[l]$, $l = 0, 1, \ldots K$. We now have $\underline{v}_n = G\underline{w}_n$, where $\underline{v}_n = \begin{bmatrix} v[n] & v[n-1] & \cdots & v[n-L+1] \end{bmatrix}^T$ and $\underline{w}_n = \begin{bmatrix} w[n] & w[n-1] & \cdots & w[n-K-L+1] \end{bmatrix}^T$. Thus, $R^k_{vv} = GR^k_{ww}G^T$, where R^k_{ww} is the correlation matrix of $\underline{w}_n$ given $\underline{I}_n = \underline{I}^k$: For $k = 0$ there are no outliers in $\underline{w}_n$, and therefore all the components have equal variance σ_1^2. For $k \neq 0$, there is a single outlier of variance σ_2^2 at the k 'th entry of $\underline{w}_n$. In any event, $\underline{w}_n$'s components are independent. Thus,

$$R^k_{ww} = \begin{cases} \sigma_1^2 I & k = 0 \\ \sigma_1^2 I + \delta^2 \underline{e}_k \underline{e}_k^T & k \neq 0 \end{cases} \qquad (4)$$

where I denotes the identity matrix, $\underline{e}_k$ is the k'th column of I, and $\delta^2 = \sigma_2^2 - \sigma_1^2$. Therefore,

$$R^k_{vv} = \begin{cases} \sigma_1^2 GG^T & k = 0 \\ \sigma_1^2 GG^T + \delta^2 \underline{g}_k \underline{g}_k^T & k \neq 0 \end{cases} \qquad (5)$$

where $\underline{g}_k$ is the k'th column of G.

Noting that $R^k_{yy} = R^0_{yy} + \delta^2 \underline{g}_k \underline{g}_k^T$ ($k = 1, 2, \ldots K + L$), where $R^0_{yy} = R_{xx} + \sigma_1^2 GG^T$, and using the Matrix Inversion Lemma we have:

$$(R^k_{yy})^{-1} = (R^0_{yy})^{-1} - \qquad (6)$$
$$- \frac{\delta^2}{1 + \delta^2 \underline{g}_k^T (R^0_{yy})^{-1} \underline{g}_k} [(R^0_{yy})^{-1} \underline{g}_k][(R^0_{yy})^{-1} \underline{g}_k]^T$$

and

$$\det\{R^k_{yy}\} = \det\{R^0_{yy}\}(1 + \delta^2 \underline{g}_k^T (R^0_{yy})^{-1} \underline{g}_k). \qquad (7)$$

Our estimator can be expressed as

$$\hat{x}[n] = \sum_{k=0}^{K+L} q_k(\underline{y}_n)\hat{x}_k[n] \qquad (8)$$

where we denote by $q_k(\underline{y}_n) = P(\underline{I}_n = \underline{I}^k|\underline{y}_n)$ the posterior probability of $\underline{I}_n = \underline{I}^k$ given $\underline{y}_n$,

$$q_k(\underline{y}_n) = \frac{f_{\underline{y}|\underline{I}}(\underline{y}_n|\underline{I}_n = \underline{I}^k)P(\underline{I}_n = \underline{I}^k)}{\sum_{k=0}^{2^{K+L}-1} f_{\underline{y}|\underline{I}}(\underline{y}_n|\underline{I}_n = \underline{I}^k)P(\underline{I}_n = \underline{I}^k)}$$
$$\approx \frac{f_{\underline{y}|\underline{I}}(\underline{y}_n|\underline{I}_n = \underline{I}^k)P(\underline{I}_n = \underline{I}^k)}{\sum_{k=0}^{K+L} f_{\underline{y}|\underline{I}}(\underline{y}_n|\underline{I}_n = \underline{I}^k)P(\underline{I}_n = \underline{I}^k)} \qquad (9)$$

(The $\approx$ should be interpreted in a probabilistic sense: the terms discarded in the denominator are negligible only in the highly probable case where not more than one outlier occurred in $\underline{w}_n$)

where

$$P(\underline{I}_n = \underline{I}^k) = \begin{cases} (1-p)^{K+L} & k = 0 \\ p(1-p)^{K+L-1} & k = 1, 2, \ldots K+L \end{cases} \qquad (10)$$

and

$$f_{\underline{y}|\underline{I}}(\underline{y}_n|\underline{I}_n = \underline{I}^k) = \qquad (11)$$
$$\frac{1}{\sqrt{\det\{2\pi R^k_{yy}\}}} \exp(-\frac{1}{2}\underline{y}_n^T (R^k_{yy})^{-1} \underline{y}_n).$$

Using (6) and (7) we may rewrite (11) as

$$f_{\underline{y}|\underline{I}}(\underline{y}_n|\underline{I}_n = \underline{I}^k) = \qquad (12)$$
$$\frac{1}{\sqrt{\det\{2\pi R^0_{yy}\}}} \cdot$$
$$\exp(-\frac{1}{2}\underline{y}_n^T (R^0_{yy})^{-1} \underline{y}_n) \cdot$$
$$\frac{1}{\sqrt{1 + \delta^2 \underline{g}_k^T (R^0_{yy})^{-1} \underline{g}_k}} \cdot$$
$$\exp(\frac{1}{2} \frac{\delta^2}{1 + \delta^2 \underline{g}_k^T (R^0_{yy})^{-1} \underline{g}_k} (\underline{y}_n^T (R^0_{yy})^{-1} \underline{g}_k)^2)$$

Note that the first two terms are independent of k and are therefore common to the nominator and denominator in (9), which we may consequently rewrite as:

$$q_k(\underline{y}_n) = \frac{\eta_k(\underline{y}_n)}{\sum_{k=0}^{K+L} \eta_k(\underline{y}_n)} \qquad (13)$$

where

$$\eta_k(\underline{y}_n) = \begin{cases} 1-p & k = 0 \\ \gamma_k \exp(\beta_k(\tilde{\underline{h}}_k^T \underline{y}_n)^2) & k \neq 0 \end{cases} \qquad (14)$$

with

$$\tilde{\underline{h}}_k = (R^0_{yy})^{-1} \underline{g}_k \qquad (15)$$
$$\gamma_k = p / \sqrt{1 + \delta^2 \underline{g}_k^T \tilde{\underline{h}}_k} \qquad (16)$$
$$\beta_k = \delta^2 \gamma_k^2 / 2p^2. \qquad (17)$$

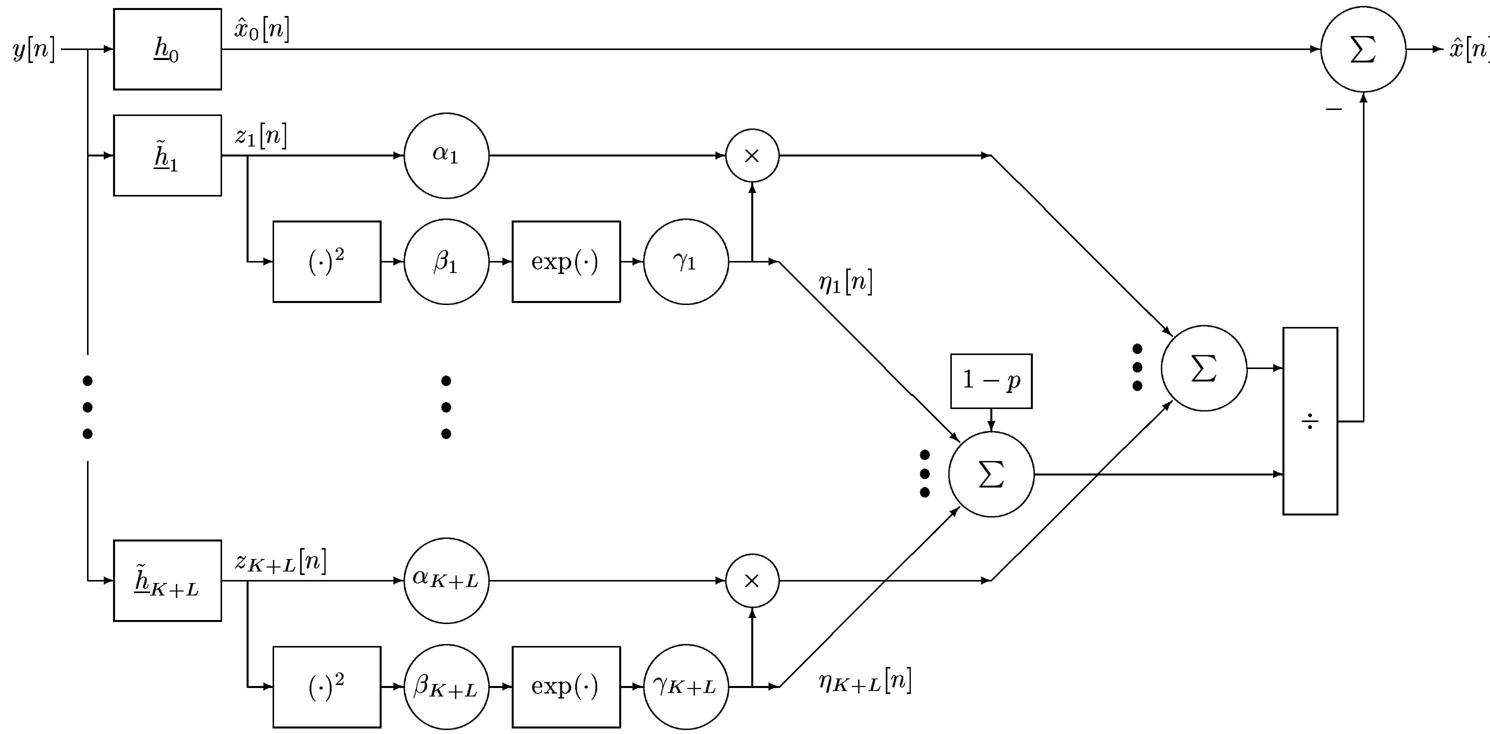

Figure 1: Finite Memory Filter for Gaussian Signals in Colored ZMGM Noise

Substituting $q_k(\underline{y}_n)$ into (8), we observe that $\hat{x}[n]$ can be interpreted as a weighted combination of linear estimators. The first linear estimator, $\hat{x}_0[n]$, is the optimal linear estimator of $x[n]$ assuming that $\underline{v}_n \sim \mathcal{N}(\underline{0}, R_{vv}^0)$, namely that no outlier occurred in $\underline{w}_n$. The other $K+L$ estimators are each optimal linear estimators assuming that $\underline{v}_n \sim \mathcal{N}(\underline{0}, R_{vv}^k)$, namely that a single outlier occurred in the respective location in $\underline{w}_n$. The weights $q_k(\underline{y}_n)$ reflect the posterior probability of the respective events presumed by the estimators.

The data dependent terms $\tilde{\underline{h}}_k^T \underline{y}_n$ in the estimator are common to both the linear estimators $\hat{x}_k[n]$ and to the terms $\eta_k(\underline{y}_n)$ for $k = 1, 2, \ldots K+L$, respectively. This feature can be exploited by the implementation depicted in Figure 1, where $\underline{h}_0 = (R_{yy}^0)^{-1} \underline{r}_{xx}^T$, $\alpha_k = 2\beta_k \underline{r}_{xx} \tilde{\underline{h}}_k$, and $\tilde{\underline{h}}_k$, γ_k and β_k are given above in (15). The estimator is calculated using only $K+L+1$ FIR filters (whose outputs are denoted $\hat{x}_0[n], z_1[n], \ldots z_{K+L}[n]$), scalar non-linearities (composed of simple $(\cdot)^2$ and $\exp(\cdot)$ operations), and one division. This architecture also allows parallel implementation of the FIR and non-linear operations (up to the final division).

It is important to observe, that the filter's memory length L has to be carefully designed: On one hand, the correlation length of the signals involved may require long memory; on the other hand, as L grows, the condition $(L+K)p^2 \ll 1$ may be breached, implying that more than one outlier is likely to occur in $\underline{w}_n$, thus devalidating our filter derivation. It has been verified by simulation (not presented here), that when L is chosen properly, substantial improvement (in terms of mse) over the optimal linear estimator can be attained by the proposed filter. The improvement increases as σ_2/σ_1 increases, but attains an optimum as a function of SNR and p.

It is interesting to note that when it is desired to estimate the noise component, the optimal estimator is the complementary estimator, i.e. $\hat{v}[n] = y[n] - \hat{x}[n]$. This holds true since $y[n] = x[n] + v[n] \Rightarrow E[v[n]|\underline{y}_n] = E[y[n] - x[n]|\underline{y}_n] = y[n] - E[x[n]|\underline{y}_n]$. Consequently, the resulting mse's are the same in both cases, since $\hat{v}[n] - v[n] = -(\hat{x}[n] - x[n])$. In our case, however, the estimators are only nearly optimal, and therefore this property is only approximately apparent.

3 Conclusion

We presented a sub-optimal finite-memory filter for estimating a stationary Gaussian signal from measurements corrupted by colored ZMGM noise. The advantage of the filter is in its ability to deal properly with outlier situations without discarding data on one hand, and without compromising performance in "benign" (no outliers) situations on the other hand. Under the assumption that independent occurrences of outliers are usually sufficiently far apart, the filter is nearly optimal. Adaptive implementations can be pursued, that may exploit the relative simplicity of the proposed architecture.

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] J. Bodenschatz and C. Nikias, "Symmetric Alpha-Stable Filter Theory," *IEEE Trans. Signal Proc.*, vol. 45 no. 9, pp. 2301–2306, Sept. 1997.

[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).