

Tracking Analysis of Minimum Kernel Risk-Sensitive Loss Algorithm Under General Non-Gaussian Noise

Amir Rastegarnia, Parnian Malekian, Azam Khalili, Wael M. Bazzi, and Saeid Sanei, *Senior Member, IEEE*

Abstract—In this paper the steady-state tracking performance of minimum kernel risk-sensitive loss (MKRSL) in a non-stationary environment is analyzed. In order to model a non-stationary environment, a first-order random-walk model is used to describe the variations of optimum weight vector over time. Moreover, the measurement noise is considered to have non-Gaussian distribution. The energy conservation relation is utilized to extract an approximate closed-form expression for the steady-state excess mean square error (EMSE). Our analysis shows that unlike for the stationary case, the EMSE curve is not an increasing function of step-size parameter. Hence, the optimum step-size which minimizes the EMSE is derived. We also discuss that our approach can be used to extract steady-state EMSE for a general class of adaptive filters. The simulation results with different noise distributions support the theoretical derivations.

Index Terms—Adaptive filter, Kernel risk-sensitive loss, Non-stationary, tracking analysis.

I. INTRODUCTION

ADAPTIVE filters have been successfully applied in a variety of fields [1]. Generally, in the adaptive filters, the filter coefficients are adjusted to minimize (or maximize) a cost function, which is indeed a statistical similarity measure among the desired and actual filter outputs. Among the available measures, those based on the second-order statistics, such as mean-square error (MSE) are more popular due to their simplicity and optimality under the linearity and Gaussian data. Unfortunately, adaptive filtering algorithms that rely only on the second-order similarity measure perform poorly when dealing with nonlinear models and non-Gaussian signals [2], [3]. Therefore, to enhance the performance of algorithms in such situations it is necessary to consider similarity measures that use high-order statistics (such as kurtosis and skewness) [3]. As another alternative to the MSE, the risk-sensitive loss (RSL) has been reported in [2] which can deliver suitable performance when the data model deviates from the Gaussian distribution. However, as the error surface of RSL tends to be super-convex, it is not robust to outliers or impulsive noises,

Manuscript received 2018. A. Rastegarnia, P. Malekian and A. Khalili are with the Department of Electrical Engineering, Malayer University, Malayer 65719-95863, Iran (email: rastegarnia@malayeru.ac.ir; parnian.malekian@stu.malayeru.ac.ir; a-khalili@tabrizu.ac.ir).

W. M. Bazzi is with the Electrical and Computer Engineering Department, American University in Dubai Dubai, United Arab Emirates, (email: wbazzi@aud.edu).

S. Sanei is with the School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, U.K. (e-mail: saeid.sanei@ntu.ac.uk) Digital Object Identifier 2018/XX

specially when it is used as cost function in stochastic gradient based algorithms.

In recent years, information theoretic learning (ITL) has been developed as an alternative learning framework [3]. In the ITL framework, learning metrics (such as entropy, divergence) are defined as functions of probability density function (pdf) of the data, meaning that they contain all the statistical information (e.g. the higher-order statistics) of the data. Using Parzen window density estimator, the PDF can be elegantly estimated from the data samples which makes the ITL metrics suitable for applications where on-line learning is needed [4]. Therefore, ITL has been successfully used in many applications such as adaptive filtering, blind source separation, classification, neural networks training, matched filtering, and feature extraction [3], [5]–[7].

An important result of the ITL framework is the connection between the ITL and kernel method. More specifically, most similarity measures that are originally developed in ITL framework can be described in a new space, namely reproducing kernel Hilbert space (RKHS) [3]. For example, the correntropy is defined in ITL as a local similarity criterion which indicates the similarity between two random variables over an observation window (kernel bandwidth) [8]. In the kernel space, the correntropy can be interpreted as a correlation metric. The correntropy captures the higher-order statistics of data. Moreover, the kernel bandwidth can be used to reduce the effect of outliers and out-of-range data. These properties motivated using the correntropy as a cost function in adaptive filters [9]–[13].

The performance of correntropy-based adaptive filters have been studied in [14]–[16] where the results show that they provide good steady-state performance with slow convergence rate¹. To address the convergence issue of correntropy, in [17] Chen *et al.* have developed a new similarity metric in kernel space, named kernel-risk sensitive loss (KRSL). They have also used KRSL to derive an adaptive filter referred to as minimum KRSL (MKRSL) algorithm. Based on the provided analysis in [17], MKRSL delivers fast convergence and low steady-state error compared to those given by other robust adaptive filters including correntropy-based adaptive filter.

As adaptive filters are used in time-varying environments it is crucial to investigate the tracking performance of the adap-

¹The slow convergence rate of correntropy-based adaptive filter stems from its non-convex error surface, i.e. it is sharp around the optimum point but flat at points which are far from the optimal point

tation algorithms under mentioned conditions. Therefore, in this paper, we study the tracking performance of the MKRSL algorithm in a non-stationary environment where a first-order random walk model is utilized to model the optimal parameter changes. In our analysis we do not impose any specific distribution for the measurement noise. We only consider the basic assumptions that are widely used in the analysis of adaptive filters. In our analysis, we use the energy conservation method [1] and obtain an approximate closed-form expression for the EMSE. We provide numerical examples with different noise distributions, which show that the MKRSL algorithm is able to keep tracking the optimal weight. Moreover, the simulation results support the theoretical findings.

The remainder of this paper is organized as follows. In Section II, we briefly review the MKRSL algorithm. In Section III, we analyze the tracking performance of the MKRSL algorithm. In Section IV, we present the simulation results to verify our theoretical analysis, and we conclude the paper in Section V.

We use small boldface letters for vectors and bold capital letters for matrices. Normal font letters denote scalars. The notation $(\cdot)^*$ denotes conjugation for scalars and Hermitian transpose for matrices conjugate transposition and $(\cdot)^T$ for matrices transposition. The operator $\text{tr}\{\cdot\}$ refers to the trace of its matrix argument.

II. THE MKRSL ALGORITHM

We consider an adaptive filter with in a system identification setup where the filter length is M . We denote by $\mathbf{w}^o \in \mathbb{R}^{M \times 1}$ the unknown model coefficients of the system. The filter is deployed to estimate \mathbf{w}^o using a stream of wide-sense stationary data $\{d_i, \mathbf{u}_i\}$, where $\{\mathbf{u}_i\}$ denotes a sequence of zero-mean real-valued (row) input vectors and $\{d_i\}$ denotes a sequence of zero-mean real-valued output samples. The system is described by a linear model as

$$d_i = \mathbf{u}_i \mathbf{w}^o + v_i \quad (1)$$

where v_i represents the zero-mean measurement noise of variance $\sigma_{v,k}^2$. As it is convenient, we consider the following assumptions for data model given in (1).

Assumption 1.

- (i) The second-order moments of the input regressors vectors are $\mathbb{E}[\mathbf{u}_i^T \mathbf{u}_i] = \mathbf{R}_{\mathbf{u}} > 0$.
- (ii) The measurement noises $\{v_i\}$ are independent of each other and the regression vectors $\{\mathbf{u}_j\}$ for any i and j .

As discussed in [17], an estimate of \mathbf{w}^o can be obtained by minimizing the KRSL cost function defined as

$$\mathcal{J}(\mathbf{w}) = \frac{1}{\lambda} \mathbb{E}[\exp(\lambda(1 - \kappa_\sigma(e_i)))] \quad (2)$$

where $\lambda > 0$ is the risk-sensitive parameter, $e_i = d_i - \mathbf{u}_i \mathbf{w}$ denotes the error at time i , and \mathbf{w} is the estimated weight vector (filter weight vector). Moreover, κ_σ is the Gaussian kernel with kernel bandwidth σ which is defined as:

$$\kappa_\sigma(e_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{e_i^2}{2\sigma^2}\right) \quad (3)$$

In many practical cases, the distributions of data are unknown and only a finite number of samples is available. In these situations, the expectation value can be estimated by using an average over N samples, which results in the empirical KRSL cost function as:

$$\hat{\mathcal{J}}(\mathbf{w}) = \frac{1}{N\lambda} \sum_{i=1}^N \exp(\lambda[1 - \kappa_\sigma(e_i)]) \quad (4)$$

The online (instantaneous) cost function, $N = 1$, under the MKRSL criterion at time i is

$$\hat{\mathcal{J}}(\mathbf{w}) = \frac{1}{\lambda} \exp(\lambda[1 - \kappa_\sigma(e_i)]) \quad (5)$$

Using the empirical cost function, a stochastic gradient based adaptive algorithm can be derived as

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} - \mu \nabla_{\mathbf{w}} \hat{\mathcal{J}}(\mathbf{w}_{i-1}) \\ &= \mathbf{w}_{i-1} + \eta \mathbf{u}_i^T \exp(\lambda[1 - \kappa_\sigma(e_i)]) \kappa_\sigma(e_i) e_i \end{aligned} \quad (6)$$

where $\eta = \frac{\mu}{\sigma^2}$. It is noteworthy that as $\sigma \rightarrow \infty$ the MKRSL becomes the LMS algorithm.

III. NON-STATIONARITY ANALYSIS

A. Assumptions and Definitions

In this paper, we investigate the tracking performance of the MKRSL algorithm under non-stationary condition where the optimal weight \mathbf{w}^o varies over time according to a first-order random walk model as

$$\mathbf{w}_i^o = \mathbf{w}_{i-1}^o + \mathbf{q}_i \quad (7)$$

where $\{\mathbf{q}_i\}$ is a zero-mean random process vector with positive-definite covariance matrix $\mathbf{Q} = \mathbb{E}[\mathbf{q}_i \mathbf{q}_i^T]$. In our analysis, we consider the following assumption

Assumption 2. The sequence $\{\mathbf{q}_i\}$ are independent initial value \mathbf{w}_0^o and $\{\mathbf{u}_\ell, v_j, \mathbf{q}_j\}$ for all i and j .

Remark 1. This assumption is the same as the one considered in e.g. [18], [19]. In [1, pp. 271–273] the suitability of random walk model for the tracking performance of adaptive filters has been discussed in more detail. Moreover, in [1, pp. 325–327] the problem of tracking a Rayleigh fading channel in a wireless communication environment has been discussed, which shows how a random walk model arises in practical applications.

As the performance metric, we consider the steady-state excess mean square error which is dened as

$$\xi = \lim_{i \rightarrow \infty} \mathbb{E}[e_{a,i}^2] \quad (8)$$

In (8) $e_{a,i}$ denotes the *a priori* error signal which is defined in terms of the weight-error vector $\tilde{\mathbf{w}}_i$ as

$$e_{a,i} \triangleq \mathbf{u}_i \tilde{\mathbf{w}}_i = \mathbf{u}_i (\mathbf{w}_i^o - \mathbf{w}_i) \quad (9)$$

B. Analysis

We consider the KRSL recursion (6) in a compact form as

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \eta \mathbf{u}_i^T f(e_i) \quad (10)$$

where $f(e_i)$ is given by

$$f(e_i) = \exp(\lambda(1 - \kappa_\sigma(e_i))) \kappa_\sigma(e_i) e_i \quad (11)$$

The first step in the energy-based method is to derive an energy-conservation relation for the MKRSL algorithm. The obtained energy-conservation relation is used to arrive at expression which reveals how $\mathbb{E}[\|\tilde{\mathbf{w}}_i\|^2]$ evolves over time. By subtracting \mathbf{w}_i^o from both sides of (10) and replacing \mathbf{w}_i^o with $\mathbf{w}_{i-1}^o + \mathbf{q}_i$ in the right hand side (r.h.s), we obtain

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} + \mathbf{q}_i - \eta \mathbf{u}_i^T f(e_i) \quad (12)$$

By equating the squared Euclidean norm of (12) the energy-conservation relation for MKRSL algorithm is obtained as:

$$\begin{aligned} \|\tilde{\mathbf{w}}_i\|^2 &= \|\tilde{\mathbf{w}}_{i-1}\|^2 + \tilde{\mathbf{w}}_{i-1}^T \mathbf{q}_i - \eta \tilde{\mathbf{w}}_{i-1}^T \mathbf{u}_i^T f(e_i) + \|\mathbf{q}_i\|^2 \\ &\quad + \mathbf{q}_i^T \tilde{\mathbf{w}}_{i-1} - \eta \mathbf{q}_i^T \mathbf{u}_i^T f(e_i) - \eta \mathbf{u}_i \tilde{\mathbf{w}}_{i-1} f(e_i) \\ &\quad - \eta \mathbf{u}_i \mathbf{q}_i f(e_i) + \eta^2 \|\mathbf{u}_i\|^2 f^2(e_i) \end{aligned} \quad (13)$$

Now, we proceed to use the energy-conservation relation (13) to estimate $\mathbb{E}[e_{a,i}^2]$. By taking expectations of both sides of (13) and using Assumption 2 we obtain the following variance relation

$$\begin{aligned} \mathbb{E}[\|\tilde{\mathbf{w}}_i\|^2] &= \mathbb{E}[\|\tilde{\mathbf{w}}_{i-1}\|^2] - 2\eta \mathbb{E}[\mathbf{u}_i \tilde{\mathbf{w}}_{i-1} f(e_i)] \\ &\quad + \mathbb{E}[\|\mathbf{q}_i\|^2] + \eta^2 \mathbb{E}[\|\mathbf{u}_i\|^2 f^2(e_i)] \\ &\quad + \mathbb{E}[\tilde{\mathbf{w}}_{i-1}^T \mathbf{q}_i] + \mathbb{E}[\mathbf{q}_i^T \tilde{\mathbf{w}}_{i-1}] \\ &\quad - 2\eta \mathbb{E}[\mathbf{q}_i^T \mathbf{u}_i^T f(e_i)] \end{aligned} \quad (14)$$

Remark 2. Comparing (14) with the variance relation for MKRSL algorithm under stationary condition reveals that (14) contains extra statistical moments that represent the effects of non-stationary environment.

For the third term in the r.h.s of (14) we have

$$\mathbb{E}[\|\mathbf{q}_i\|^2] = \mathbb{E}[\text{tr}\{\mathbf{q}_i \mathbf{q}_i^T\}] = \text{tr}\{\mathbf{Q}\} \quad (15)$$

For the last three terms in (14), firstly it should be noted that $\tilde{\mathbf{w}}_{i-1}$ can be written as

$$\tilde{\mathbf{w}}_{i-1} = \mathbf{w}_{i-1}^o - \mathbf{w}_{i-1} = \mathbf{w}_0^o + \sum_{n=0}^{i-1} \mathbf{q}_n - \mathbf{w}_{i-1}$$

As \mathbf{w}_{i-1} depends on the input data $\{d_i, \mathbf{u}_i\}$ and they are independent of \mathbf{q}_i (see Assumption 2), we have

$$\mathbb{E}[\tilde{\mathbf{w}}_{i-1}^T \mathbf{q}_i] = \mathbb{E}[\mathbf{w}_0^{oT} \mathbf{q}_i] + \sum_{n=1}^{i-1} \mathbb{E}[\mathbf{q}_n^T \mathbf{q}_i] - \mathbb{E}[\mathbf{w}_{i-1}^T \mathbf{q}_i] = 0 \quad (16)$$

Similarly, for the last two terms in (14) we have

$$\mathbb{E}[\mathbf{q}_i^T \tilde{\mathbf{w}}_{i-1}] = \mathbb{E}[\mathbf{q}_i^T \mathbf{u}_i^T f(e_i)] = 0 \quad (17)$$

Using (15)-(17), (14) becomes

$$\begin{aligned} \mathbb{E}[\|\tilde{\mathbf{w}}_i\|^2] &= \mathbb{E}[\|\tilde{\mathbf{w}}_{i-1}\|^2] - 2\eta \mathbb{E}[e_{a,i} f(e_i)] \\ &\quad + \eta^2 \mathbb{E}[\|\mathbf{u}_i\|^2 f^2(e_i)] + \text{tr}\{\mathbf{Q}\} \end{aligned} \quad (18)$$

In the steady-state we have

$$\lim_{i \rightarrow \infty} \mathbb{E}[\|\tilde{\mathbf{w}}_i\|^2] = \lim_{i \rightarrow \infty} \mathbb{E}[\|\tilde{\mathbf{w}}_{i-1}\|^2] \quad (19)$$

Moreover, due to Assumption 2 the third term in the r.h.s of (18) becomes

$$\lim_{i \rightarrow \infty} \mathbb{E}[\|\mathbf{u}_i\|^2 f^2(e_i)] = \text{tr}\{\mathbf{R}_u\} \lim_{i \rightarrow \infty} \mathbb{E}[f^2(e_i)] \quad (20)$$

Using (19) and (20) relation (18) at steady-state becomes

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{E}[e_{a,i} f(e_i)] &= \eta \text{tr}\{\mathbf{R}_u\} \lim_{i \rightarrow \infty} \mathbb{E}[f^2(e_i)] \\ &\quad + \eta^{-1} \text{tr}\{\mathbf{Q}\} \end{aligned} \quad (21)$$

To obtain a closed-form expression for EMSE from (21) we need to calculate the appeared moments $\mathbb{E}[e_{a,i} f(e_i)]$ and $\mathbb{E}[f^2(e_i)]$. As we do not consider any specific distribution for v_i , it is difficult to derive the theoretical expressions for the mentioned moments. To handle this issue, we resort to the Taylor expansion and firstly approximate $f(e_i)$ with respect to $e_{a,i}$ around v_i which yields

$$\begin{aligned} f(e_i) &= f(e_{a,i} + v_i) \\ &\approx f(v_i) + f'(v_i) e_{a,i} + \frac{1}{2} f''(v_i) e_{a,i}^2 + \mathcal{O}(e_{a,i}^2) \end{aligned} \quad (22)$$

where

$$\begin{aligned} f'(v_i) &= \exp(\lambda(1 - \kappa_\sigma(v_i))) \kappa_\sigma(v_i) \\ &\quad \times \left(1 + \lambda \frac{v_i^2}{\sigma^2} \kappa_\sigma(v_i) - \frac{v_i^2}{\sigma^2}\right) \end{aligned} \quad (23)$$

and

$$\begin{aligned} f''(v_i) &= \exp(\lambda(1 - \kappa_\sigma(v_i))) \kappa_\sigma(v_i) \\ &\quad \times \left(\frac{\lambda^2 v_i^3}{\sigma^4} \kappa_{\frac{\sigma}{2}}(v_i) + \frac{3\lambda \sigma^2 v_i - 3\lambda v_i^3}{\sigma^4} \kappa_\sigma(v_i) \right. \\ &\quad \left. + \frac{v_i^3 - 3v_i \sigma^2}{\sigma^4} \right) \end{aligned} \quad (24)$$

Now, (22)-(24) can be used to obtain the approximated moments in (21). In this way for $\mathbb{E}[e_{a,i} f(e_i)]$ we have

$$\begin{aligned} \mathbb{E}[e_{a,i} f(e_i)] &= \mathbb{E}[e_{a,i} f(v_i) + f'(v_i) e_{a,i}^2 + \mathcal{O}(e_{a,i}^2)] \\ &\approx \xi \mathbb{E}[f'(v_i)] \end{aligned} \quad (25)$$

Similarly, for $\mathbb{E}[f^2(e_i)]$ we have

$$\begin{aligned} \mathbb{E}[f^2(e_i)] &\approx \mathbb{E}[(f(v_i) + f'(v_i) e_{a,i} + \frac{1}{2} f''(v_i) e_{a,i}^2)^2] \\ &\approx \mathbb{E}[f^2(v_i)] + \xi \mathbb{E}[f(v_i) f''(v_i) + (f'(v_i))^2] \end{aligned} \quad (26)$$

Replacing (25) and (26) in (21) and solving the resultant equation for ξ we obtain the approximated EMSE for MKRSL algorithm in a non-stationary environment as follows:

$$\xi = \frac{\eta \text{tr}\{\mathbf{R}_u\} \mathbb{E}[f^2(v)] - \eta^{-1} \text{tr}\{\mathbf{Q}\}}{2\mathbb{E}[f'(v)] - \eta \text{tr}\{\mathbf{R}_u\} \mathbb{E}[f(v) f''(v) + (f'(v))^2]} \quad (27)$$

Note that we omit the time index i for brevity.

Corollary 1. Let's rewrite the EMSE in (27) as

$$\xi = \frac{\eta A + \eta^{-1} B}{C - \eta D} \quad (28)$$

where

$$\begin{aligned} A &= \text{tr} \{ \mathbf{R}_{\mathbf{u}} \} \mathbb{E}[f^2(v)] \\ B &= \text{tr} \{ \mathbf{Q} \} \\ C &= 2\mathbb{E}[f'(v)] \\ D &= \text{tr} \{ \mathbf{R}_{\mathbf{u}} \} \mathbb{E}[f(v)f''(v) + (f'(v))^2] \end{aligned} \quad (29)$$

The first term term in the numerator of (28) decreases with step-size and increases with $\text{tr} \{ \mathbf{Q} \}$ while the second term increases with step-size. So, the EMSE expression is not an increasing function in terms of step-size unlike for the stationary case.

Corollary 2. The optimum step-size value which minimizes the EMSE is obtained by setting the derivation of (27) to zero ($\frac{d\xi}{d\eta} = 0$). Doing so we have

$$\eta^o = \frac{-BD + \sqrt{B^2D^2 + ABC^2}}{AC} \quad (30)$$

Corollary 3. In comparison to stationary environments, there is an increase at the steady-state which is given by

$$\Delta\xi = \frac{\eta^{-1}B}{C - \eta D}$$

Remark 3. The presented approach for derivation of EMSE holds for any adaptive filter which can be described by (10). This is a very useful result in the context of tracking analysis because it allows us to derive the steady-state EMSE for any adaptive filter only by evaluating A , B , C and D for a given function $f(v)$. For example, for the adaptive filtering with the maximum correntropy criterion (MCC), we have $f(v) = v[\exp(v^2/(2\sigma^2))]$. Hence, the approximated EMSE for adaptive filter under MCC is given by (28) with

$$\begin{aligned} A &= \text{tr} \{ \mathbf{R}_{\mathbf{u}} \} \mathbb{E} \left[v^2 \exp \left(\frac{-v^2}{\sigma^2} \right) \right] \\ B &= \text{tr} \{ \mathbf{Q} \} \\ C &= 2\mathbb{E} \left[\left(1 - \frac{v^2}{\sigma^2} \right) \exp \left(\frac{-v^2}{2\sigma^2} \right) \right] \\ D &= \text{tr} \{ \mathbf{R}_{\mathbf{u}} \} \mathbb{E} \left[\left(1 + \frac{2v^4}{\sigma^4} - \frac{5v^2}{\sigma^2} \right) \exp \left(\frac{-v^2}{\sigma^2} \right) \right] \end{aligned}$$

which is the same results as those obtained in [16].

IV. SIMULATION RESULTS

In this section, we present some numerical examples to examine the performance of the MKRSL algorithm in non-stationary condition, and to verify the theoretical analysis. We consider the system identification setup where the filter length is $M = 20$ and the initial vector $\mathbf{w}_0^o = [1, 1, \dots, 1]^T$. The input vector \mathbf{u}_i is generated from a zero-mean Gaussian process with unit variance matrix, and the measurement noise is a zero-mean uniform distribution with unit variance. The covariance matrix in the first-order random walk model is selected as $\mathbf{Q} = 10^{-6}I$. We perform over 500 Monte Carlo simulations, with 20000 iterations in each simulation to ensure that the algorithm reaches the steady-state. A null vector is considered as the initial weight vector \mathbf{w}_0 for MKRSL algorithm.

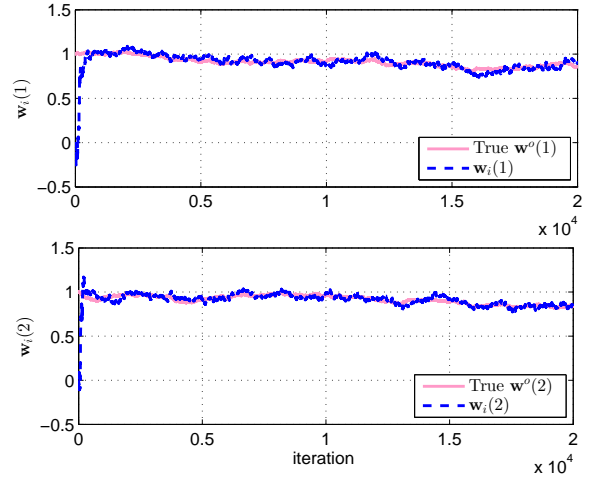


Fig. 1. Time evolution of adaptive filter weights (estimates) over time.

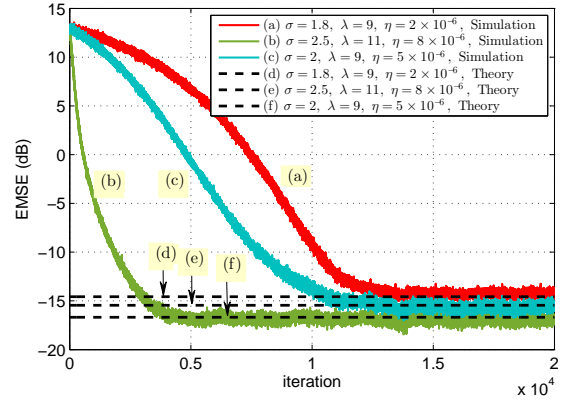


Fig. 2. Theoretical steady-state EMSEs and the simulated convergence curves with different parameter settings under non-stationary environment.

In the first example, we examine the tracking ability of MKRSL algorithm, where the algorithm parameters are selected as $\sigma = 2.5$, $\lambda = 11$, and $\eta = 8 \times 10^{-6}$. Fig. 1 shows how the filter coefficients $\mathbf{w}_i(1)$ and $\mathbf{w}_i(2)$ track \mathbf{w}_i^o . We can observe that estimates $\mathbf{w}_i(1)$ and $\mathbf{w}_i(2)$ approach \mathbf{w}_i^o after few samples and keep tracking of \mathbf{w}_i^o as it changes over time. In the next experiment, we examine the accuracy of expression (27). The theoretical steady-state EMSEs which are calculated by (27) and convergence curves simulations with different parameter setting in the non-stationary environment are shown in Fig. 2. It is obvious that finally the simulated curves converge nearly exactly towards the theoretical values. In Table I, we consider other distributions for the measurement noise including, Gaussian (zero-mean with unit variance which is denoted by $\mathcal{N}(0, 1)$), Binary (± 1 with probability 0.5), Cauchy (with PDF $p(v) = 1/[\pi(1 + v^2)]$) and impulsive noise given by a Gaussian mixture model as

$$p_v(v) = (1 - \epsilon)\mathcal{N}(0, \beta) + \epsilon\mathcal{N}(0, \kappa\beta)$$

where $\beta = 10^{-3}$ is the nominal noise variance, $\epsilon = 0.01$ is the contamination ratio, and $\kappa = 100$. As seen from Table I, the theoretical expression in (27) is valid for both Gaussian and non-Gaussian distributions. In the last example, we investigate

TABLE I
THEORETICAL AND SIMULATED STEADY-STATE EMSEs FOR DIFFERENT NOISE DISTRIBUTIONS.

Noise distribution	Parameter setting	Theory (db)	Simulation (db)
Gaussian	$\lambda = 9, \sigma = 1.2, \eta = 0.000008$	-15.6459	-15.5057 ± 0.6
Binary	$\lambda = 9, \sigma = 1, \eta = 0.00004$	-24.0263	-24.2358 ± 0.9
Cauchy	$\lambda = 7, \sigma = 2, \eta = 0.00003$	-10.0016	-9.7210 ± 0.28
Gaussian mixture model	$\lambda = 8, \sigma = 1.5, \eta = 0.000005$	-10.7955	-10.4037 ± 0.9

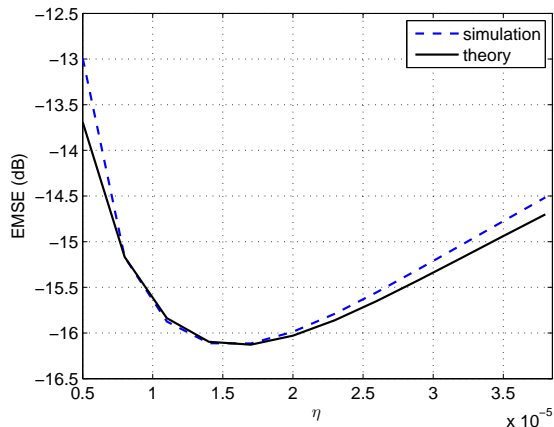


Fig. 3. Theoretical and simulation EMSE versus η .

the effect of η on the steady-state EMSE. To this end, both the theoretical and simulated cases the steady-state EMSE are plotted in terms of η in Fig. 3. The curves are obtained by averaging over the last 200 samples. Fig. 3 supports the claim that in a non-stationary environment the EMSE curve is not an increasing function of step-size parameter. Moreover, this figure shows that the theoretical expression in (27) is valid for a wide range of step-size parameters.

V. CONCLUSION

In this paper, we studied the tracking performance of MKRSL algorithm in a non-stationary environment where a first-order random walk model is used to model the variations of optimum weight vector in each iteration. In our analysis, which uses the energy conservation argument, we used a general non-Gaussian distribution for the measurement noise. Our analysis revealed that the EMSE curve is not an increasing function of step-size parameter. Hence, we need to choose an optimum step-size to achieve an acceptable performance. We presented some simulation results which well confirmed the theoretical derivations.

REFERENCES

[1] A. H. Sayed, *Adaptive filters*. John Wiley & Sons, 2011.
 [2] R. K. Boel, M. R. James, and I. R. Petersen, "Robustness and risk-sensitive filtering," *IEEE Transactions on Automatic Control*, vol. 47, no. 3, pp. 451–461, 2002.

[3] J. C. Principe, *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer Science & Business Media, 2010.
 [4] J. C. Principe and D. Xu, "An introduction to information theoretic learning," in *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, vol. 3, 1999, pp. 1783–1787.
 [5] S. Peng, W. Ser, B. Chen, L. Sun, and Z. Lin, "Robust constrained adaptive filtering under minimum error entropy criterion," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 8, pp. 1119–1123, 2018.
 [6] D. Erdogmus and J. C. Principe, "From linear adaptive filtering to nonlinear information processing - the design and analysis of information processing systems," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 14–33, Nov 2006.
 [7] M. Dehmer, M. Drmota, and F. Emmert-Streib, Eds., *Proceedings of the 2008 International Conference on Information Theory and Statistical Learning, ITSL 2008, Las Vegas, Nevada, USA, July 14-17, 2008*. CSREA Press, 2008.
 [8] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.
 [9] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *2009 International Joint Conference on Neural Networks*, June 2009, pp. 2950–2955.
 [10] Y. He, F. Wang, J. Yang, H. Rong, and B. Chen, "Kernel adaptive filtering under generalized maximum correntropy criterion," in *2016 International Joint Conference on Neural Networks (IJCNN)*, July 2016, pp. 1738–1745.
 [11] S. Peng, B. Chen, L. Sun, W. Ser, and Z. Lin, "Constrained maximum correntropy adaptive filtering," *Signal Processing*, vol. 140, pp. 116 – 126, 2017.
 [12] W. Ma, D. Zheng, Y. Li, Z. Zhang, and B. Chen, "Bias-compensated normalized maximum correntropy criterion algorithm for system identification with noisy input," *Signal Processing*, vol. 152, pp. 160 – 164, 2018.
 [13] B. Chen, X. Wang, and J. C. Principe, "Chapter 5 - maximum correntropy criterionbased kernel adaptive filters," in *Adaptive Learning Methods for Nonlinear System Modeling*, D. Comminiello and J. C. Principe, Eds. Butterworth-Heinemann, 2018, pp. 105 – 126.
 [14] M. N. Syed, P. M. Pardalos, and J. C. Principe, "On the optimization properties of the correntropic loss function in data analysis," *Optimization Letters*, vol. 8, no. 3, pp. 823–839, 2014.
 [15] B. Chen, L. Xing, J. Liang, N. Zheng, and J. C. Principe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion," *IEEE signal processing letters*, vol. 21, no. 7, pp. 880–884, 2014.
 [16] A. Khalili, A. Rastegarnia, M. K. Islam, and T. Y. Rezaei, "Steady-state tracking analysis of adaptive filter with maximum correntropy criterion," *Circuits, Systems, and Signal Processing*, vol. 36, no. 4, pp. 1725–1734, 2017.
 [17] B. Chen, L. Xing, B. Xu, H. Zhao, N. Zheng, and J. C. Principe, "Kernel Risk-Sensitive Loss: Definition, Properties and Application to Robust Adaptive Filtering," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2888–2901, 2017.
 [18] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 314–324, Feb 2001.
 [19] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3137–3149, July 2008.