Computational Performance of the Parameters Estimation in Extreme Seeking Entropy Algorithm

Jan Vrba, Jan Mareš

University of Chemistry and Technology, Prague

Faculty of Chemical Engineering, Department of Computing and Control Engineering

Czech Republic

Email: jan.vrba@vscht.cz

Abstract—This paper is dedicated to the evaluation of the computational time performance of the algorithms that estimate the parameters of the generalized Pareto distribution, namely Method of Moments, Maximum likelihood estimator and Quasi-maximum likelihood algorithms. The generalized Pareto distribution is utilized by the Extreme Seeking Entropy algorithm to detect novelty in data. The algorithm is evaluating the weight increments of the simple adaptive filter that are obtained via incrementally learning algorithm. The computational time performance is examined in the experiment with the detection of step-change parameters of the signal generator. Its output contains also additive Gaussian noise.

keywords—signal processing, adaptive systems, adaptive algorithms, novelty detection, generalized Pareto distribution

I. INTRODUCTION

The real-time novelty detection is the area that is associated not only with fault detection and diagnostics [1] in industrial processes but also with i.e. detection of anomalies in data streams as those anomalies may be associated with network intrusion or fraud attempts [2], or may be utilized in surveillance security systems [3]. Another wide area of real-time applications is mobile robotics where the computational power is usually limited [4], [5]. The key feature for each novelty detection method that is useful in real-time applications is outstanding computational time performance.

This paper is devoted to comparison of the computational time performance of three different methods for the generalized Pareto distribution [6] (GPD) parameters estimation. Compared methods are a maximum likelihood estimator (ML), method of moments (MOM) and quasi-maximum likelihood estimator (QML). The computational time performance of those methods is evaluated in the experiment of signal generator parameters step-change detection via adaptive novelty detection algorithm Extreme Seeking Entropy (ESE).

II. ADAPTIVE SYSTEM AND ALGORITHM SPECIFICATION

As the aim of the article is the real-time performance of algorithms, that estimate the GPD parameters and not the novelty detection in a complex processes, the simple linear finite impulse response filter is used. The output of the filter y at a discrete time index k is described by the following equation (1) so that

$$y(k) = w_1(k) \cdot x_1(k) + w_2(k) \cdot x_2(k) + \\ + w_3(k) \cdot x_3(k) \quad (1)$$

which is equivalent to vector form

$$y(k) = \boldsymbol{w}^T(k) \cdot \boldsymbol{x}(k) \quad (2)$$

where

$$\boldsymbol{w}^T(k) = [w_1(k), w_2(k), w_3(k)] \in R^3$$

is the vector of filters adaptive weights and

$$\boldsymbol{x}^T(k) = [x_1(k), x_2(k), x_3(k)] \in R^3$$

is the input vector. The adaptive weights are updated using the NLMS algorithm [7] with every new sample obtained. The new value of adaptive weights is given as

$$\Delta \boldsymbol{w}(k) = \frac{\kappa \cdot \boldsymbol{w}(k) \cdot e(k)}{\varepsilon + \boldsymbol{x}(k) \cdot \boldsymbol{x}^T(k)}, \quad (3)$$

where ε is used to avoid division by zero in case of zero input vector, $\kappa \in R$ is the learning rate and $e \in R$ is the output error defined as

$$e(k) = d(k) - y(k), \quad (4)$$

where $d(k) \in R$ is the measured value of the output and k is the discrete time index. For the experiment the value of the learning rate κ was set as $\kappa = 0.8$

III. EXTREME SEEKING ENTROPY ALGORITHM

The ESE algorithm [8] estimates the novelty in data via the probability of adaptive weights increments. Those increments are obtained via incrementally learning algorithm and the probability is estimated via generalized Pareto distribution. The parameters of GPD are obtained using the last n_s adaptive weight increments

on which is applied the peak-over-threshold (POT) method so for parameters estimation only n biggest weight increments are used. The POT method was chosen according to [9] as 10%, so the parameters (denoted n) of GPD are estimated from the 10% of the highest weight increments.

$$n = \lceil 0.1 \cdot n_s \rceil \quad (5)$$

The novelty score is calculated according to following formula (6).

$$ESE(|\Delta\mathbf{w}(k)|) = -\log \prod_{i=1}^{n_f} (1 - f_{cdf_i}(|\Delta w_i(k)|)) \quad (6)$$

where

$$f_{cdf_i}(|\Delta w_i(k)|) =$$
$$= \begin{cases} 0, |\Delta w_i(k)| < \zeta_i \\ F_{(\gamma_i,\mu_i,\sigma_i)}(|\Delta w_i(k)|), |\Delta w_i(k)| \geq \zeta_i. \end{cases}$$

and $F_{(\gamma_i,\mu_i,\sigma_i)}$ is cumulative distribution function of the generalized Pareto distribution with location parameter μ_i, scale parameter σ_i and shape parameter γ_i and n_f is the number of adaptive weights and ζ_i correspond to the minimum of the highest adaptive weight increments that were obtained after proceeding the peak-over-threshold method.

Algorithm 1 Extreme Seeking Entropy Algorithm

1: set n_s and choose POT method
2: initial estimation of the GPDs parameters γ_i, μ_i, σ_i for each adaptable parameter
3: **for** each new $d(k)$ **do**
4: update adaptive model to get $\Delta\mathbf{w}(k)$
5: proceed POT method
6: **if** $|\Delta w_i|(k) > \zeta_i$ **then**
7:   update parameters of GPDs γ_i, μ_i, σ_i
8: **end if**
9: compute ESE according to (6)
10: **end for**

A. Estimation of Generalized Pareto Distribution Parameters

Assume the three parameters form of GPD so that probability density function is defined as follows.

$$f_{(\gamma,\mu,\sigma)}(w) = \begin{cases} \frac{1}{\sigma}\left(1 + \frac{\gamma(w-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}-1} & \text{for } \gamma \neq 0, \\ \exp\left(-\frac{w-\mu}{\sigma}\right) & \text{for } \gamma = 0. \end{cases} \quad (7)$$

Note that parameter $\mu \in (-\infty, +\infty)$ is a location parameter, $\sigma \in (0, \infty)$ is the scaling parameter, and $\gamma \in (-\infty, \infty)$ is a shape parameter. Then the cumulative distribution function then is given as

$$F_{(\gamma,\mu,\sigma)}(w) = \begin{cases} 1 - \left(1 + \frac{\gamma(w-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}} & \text{for } \gamma \neq 0, \\ 1 - \exp\left(-\frac{w-\mu}{\sigma}\right) & \text{for } \gamma = 0. \end{cases} \quad (8)$$

A ML estimator of GPD parameters (σ, γ) has to be found numerically maximizing the log-likelihood function given as

$$\log\mathcal{L}(\sigma,\gamma|w_1,\ldots,w_n) = -n\log\sigma +$$
$$+ \frac{1-\gamma}{\gamma}\sum_n \log(1 - \frac{\gamma}{\sigma}(w_i - \mu)) \quad (9)$$

with $\forall w_i \geq \mu$ (note that in this section the index of adaptive weight w is associated with number of sample that is obtained after POT method processing, not with the number of adaptive weights). The solving problem (9) is difficult, as the optimization method can converge to a sub-optimal solution and the explicit expressions of the ML estimator do not exist. Another issue that is related to the real-time application is that it is not possible to guarantee the maximum computational time dedicated to solving the problem.

A method of moments estimator of parameters (σ, γ) [10] is feasible if the parameter of the GPD $\gamma < 0.5$. Then the first and second raw moments are defined and the parameters of the GPD (σ, γ) are computed as follows

$$\widehat{\sigma} = \frac{1}{2}\overline{w}\left(\frac{\overline{w}^2}{s^2} + 1\right) \quad (10)$$

$$\widehat{\gamma} = -\frac{1}{2}\left(\frac{\overline{w}^2}{s^2} - 1\right) \quad (11)$$

where $\overline{w}$ is sample mean and s^2 is sample variance.

The quasi-ML method [11] of estimation GPD parameters (σ, γ) is having following steps, assuming that $(w_1, ..., w_n)$ is ordered sequence such that $w_1 \leq w_2 \leq \cdots \leq w_n$.

1) Compute

$$\widehat{\gamma} = \frac{1}{n-1}\sum_{i=1}^{n-1}\ln\left(1 - \frac{w_i}{max(w_1,...,w_n)}\right) \quad (12)$$

$$Z = 1 - \frac{n^{-1}\sum_{i=1}^n w_i^2}{2\overline{w}^2} \quad (13)$$

2) If $\widehat{\gamma}$ and $Z < 0.2$ compute standard MLEs for $\widehat{\gamma}$ and σ.

3) Otherwise estimate $\widehat{\gamma}$ according to equation 12 and estimate $\widehat{\sigma}$ as follows

$$\widehat{\sigma} = \widehat{\gamma} \cdot max(w_1, ..., w_n). \quad (14)$$

The accuracy of mentioned methods is not aim of this article as it was already evaluated in publications [11], [12], [13].

IV. EXPERIMENTAL RESULTS

The computational time performance was evaluated for the problem of the step-change parameters of the signal generator. Assume that the output of the signal generator $d(k)$ at the discrete time index k is given as

$$d(k) = x_1(k) + x_2(k) + x_3(k) + v(k) \quad (15)$$

for all $1 \leq k \leq 200$. Note that the $v(k)$ represents the additive gaussian noise with zero mean and standard
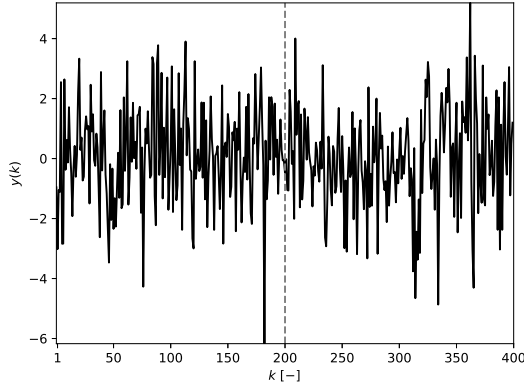
Figure 1. Output of the adaptive filter during the experiment. The step-change of the signal generator parameters at the discrete time index $k = 200$ is highlighted by grey dashed line.
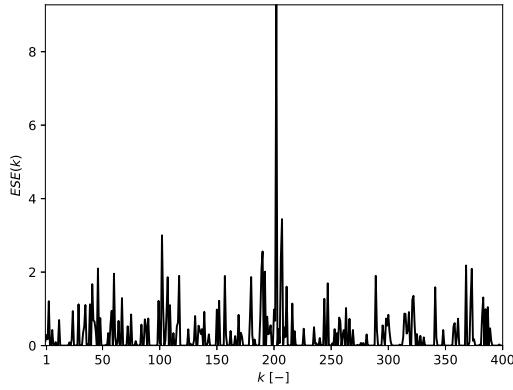

Figure 2. ESE value during the experiment. Note that the peak of ESE corresponds to the step-change of the signal generator.

deviation $\sigma_{noise} = 0.1$. At the discrete time index $k = 201$ the output of the signal generator changes to

$$d(k) = 0.7 \cdot x_1(k) + 1.2 \cdot x_1(k) + 1.1 \cdot x_1(k) + v(k) \quad (16)$$

for $201 \leq k \leq 400$. The value for each input of the signal generator is drawn from the standard normal distribution $\mathcal{N}(0, 1)$ so the ith input $x_i(k) \sim \mathcal{N}(0, 1)$ for $1 \leq k \leq 400$. Note, that the change of the signal generator parameters does not change the mean of the output $d(k)$. The output of the system generator is depicted in the following figure 1. The value of ESE is in the figure 2. The values of the GPD parameters (μ, σ, γ) during the experiment that were estimated by the ML are in the figures 3, 4 and 5.

All the experiments were performed on PC using Intel(R) Core(TM) i5-7400 4 cores CPU running at 3001 MHz and 32 GB RAM. The running operating system was Windows 10 Pro 64 bit version 10.0.18362 and code was written in Python 3.6.1 using Numpy 1.17.0 and Scipy 1.4.1 libraries. The average computational time $\bar{t}$ of the estimation of all three GPDs parameters (the number of GPDs corresponds to the number of adaptive system parameters) and corresponding standard deviations σ_t are in the following table IV.
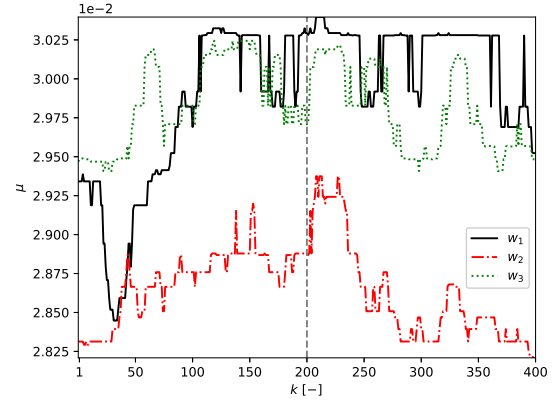

Figure 3. The values of the GPD location parameter μ for all three adaptive weights (w_1, w_2, w_3) during the experiment. The dashed line at the discrete time index k highlights the step-change of the signal generator parameters.
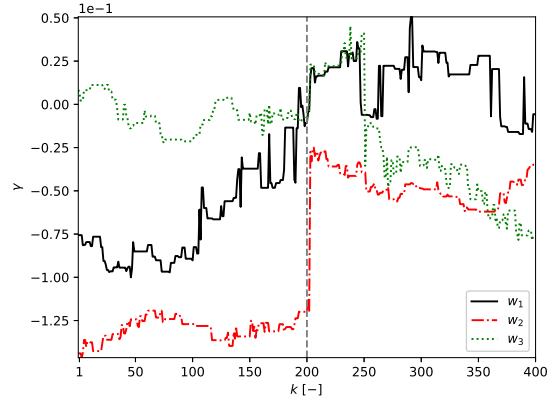

Figure 4. The values of the GPD shape parameter γ for all three adaptive weights (w_1, w_2, w_3) during the experiment. The dashed line at the discrete time index k highlights the step-change of the signal generator parameters.

Figure 5. The values of the GPD scaling parameter σ for all three adaptive weights (w_1, w_2, w_3) during the experiment. The dashed line at the discrete time index k highlights the step-change of the signal generator parameters.
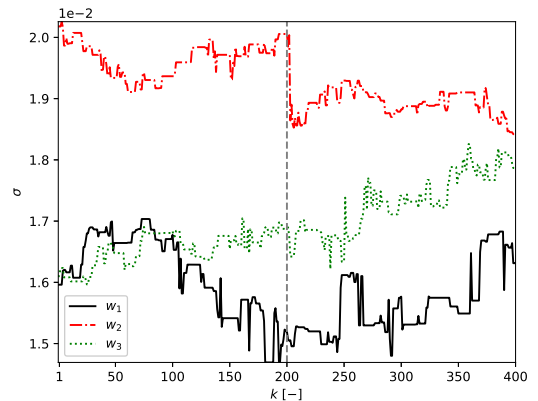
TABLE I
TABLE OF AVERAGE COMPUTATIONAL TIMES AND
CORRESPONDING STANDARD DEVIATIONS FOR SELECTED
METHODS

	Method	$\bar{t}$ [ms]	σ_t [ms]
w_1	ML	26.198	3.396
	QML	0.354	0.478
	MOM	**0.076**	0.264
w_2	ML	26.718	2.302
	QML	0.337	0.471
	MOM	**0.064**	0.244
w_3	ML	24.982	1.964
	QML	0.395	0.489
	MOM	**0.060**	0.238

V. Conclusions

The average computational time and its standard deviations of three different methods (ML, QML, MOM) for estimating the GPD parameters were evaluated in this article. Estimation of those parameters is crucial for successful adaptive novelty detection via the Extreme Seeking Entropy algorithm. The evaluation was done for step-change of signal generator parameters. The output of the signal generator contains additive Gaussian noise with zero mean.

The fastest method from evaluated methods is MOM. The drawback of this method is that the usage in real-time applications is limited by the strict restriction on values of the GPD parameters. If the parameters do not meet the restriction, the estimation is inaccurate and the results of the ESE algorithm may be misleading. In general, it is hard to guarantee, that the monitored process is going to satisfy those restrictions.

The least computational efficient method it the ML, which takes on the order four hundred times more computational time than MOM in the tested scenario. On the other side, it provides estimation for a wider range of GPD parameters values. The higher value of computational time is natural for ML method as the optimization of the max-likelihood function is an iterative process where the number of iterations can not be guaranteed. However, the apriori information about the value parameters may speed-up the convergence.

The QML method seems like a good compromise between MOM and ML as it is fast enough and in the first step tests the need for using the ML method. If there is no need of using the ML method, the parameters of GPD are computed directly without the need of using the optimization method. It takes only about five times more computational time than MOM on average and provides the estimation for wider ranges of the GPD parameters. Note that even if the ML estimation has to be done, the computational time between the QML using ML and QML with direct computation of the parameters is not significant.

Based on the performed experiments it seems, that the usage of the ESE algorithm seems promising in real-time application if the computational power is sufficient, and if the data from the monitored process are available at a suitable rate.

References

[1] J. Gertler, *Fault detection and diagnosis in engineering systems*. CRC press, 1998.

[2] K. Yu, W. Shi, N. Santoro, and X. Ma, "Real-time outlier detection over streaming data," in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 125–132, IEEE, 2019.

[3] R. Ramezani, P. Angelov, and X. Zhou, "A fast approach to novelty detection in video streams using recursive density estimation," in *2008 4th International IEEE Conference Intelligent Systems*, IEEE, Sept. 2008.

[4] S. Marsland, U. Nehmzow, and J. Shapiro, "On-line novelty detection for autonomous mobile robots," *Robotics and Autonomous Systems*, vol. 51, pp. 191–206, May 2005.

[5] U. Nehmzow, Y. Gatsoulis, E. Kerr, J. Condell, N. Siddique, and T. M. McGuinnity, "Novelty detection as an intrinsic motivation for cumulative learning robots," in *Intrinsically Motivated Learning in Natural and Artificial Systems*, pp. 185–207, Springer, 2013.

[6] H. T. Davis and M. L. Feldstein, "The generalized pareto law as a model for progressively censored survival data," *Biometrika*, vol. 66, no. 2, pp. 299–306, 1979.

[7] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2005.

[8] J. Vrba and J. Mareš, "Introduction to extreme seeking entropy," *Entropy*, vol. 22, p. 93, Jan. 2020.

[9] W. H. DuMouchel, "Estimating the stable index α in order to measure tail thickness: A critique," *The Annals of Statistics*, vol. 11, pp. 1019–1031, Dec. 1983.

[10] J. R. Hosking and J. R. Wallis, "Parameter and quantile estimation for the generalized pareto distribution," *Technometrics*, vol. 29, pp. 339–349, Aug. 1987.

[11] A. Luceño, "Fitting the generalized pareto distribution to data using maximum goodness-of-fit estimators," *Computational Statistics & Data Analysis*, vol. 51, no. 2, pp. 904–917, 2006.

[12] V. Brazauskas and A. Kleefeld, "Robust and efficient fitting of the generalized pareto distribution with actuarial applications in view," *Insurance: Mathematics and Economics*, vol. 45, no. 3, pp. 424–435, 2009.

[13] V. Singh and H. Guo, "Parameter estimation for 3-parameter generalized pareto distribution by the principle of maximum entropy (pome)," *Hydrological sciences journal*, vol. 40, no. 2, pp. 165–181, 1995.