

A Robust Set-membership Affine Projection Algorithm Based on Outlier Estimation Method

Junfeng Cui¹

*College of Mathematics and Physics, Huaiyin Institute of Technology, Jiangsu Huaiyin, 223001, China
E-mail: junfengcuiedu@gmail.com*

Haijian Shao

*Department of Computer Science and Engineering, Jiangsu University of Science and Technology,
Jiangsu, 212000, Zhenjiang, China
E-mail: shaohaijian2012@gmail.com*

This paper describes an improved set-membership affine projection algorithm by combing the median absolute deviation (MAD) technique and set-membership affine projection (SMAP) algorithm. The quality of given data is usually closely related to the robustness of employed algorithm, in particular, the time series outliers can affect the convergence speed of employed approach. In this paper, the MAD method and SMAP algorithm are applied to eliminate the outliers of given time sequence and data selective adaptive filtering, respectively, and the convergence radius magnitude analysis and spatial structure of generalized algorithm update process of proposed approach are provided. Numerical simulation is given to illustrative the effectiveness of proposed approach.

*ISCC 2015
18-19, December, 2015
Guangzhou, China*

¹Speaker

1. Introduction

Set-membership filtering (SMF) is a Data-Selective adaptive filtering, which is attempting to estimate the feasibility set [1-5]. Usually, an objective function with bounded error constraint is applied in the SMAF such that updates belong to be a feasible solution. The main difference between SMAF and SMF is that the latter takes into account bound associated to the time-domain output error [2]. A method was presented for set-membership identification (SMI) associated to parameter set by Kosut *et al.* [6], and the uncertainty of parametric and nonparametric were discussed. SM-AP algorithm is a new data selective adaptive filtering algorithm, the computational complexity of which is mainly about the number of data reuses [2,7,8]. Based on the least-trace extended set-membership filter (LTESMF), Huang Y *et al.* [9] proposed an improved approach for nonlinear stochastic systems, and the simulation results indicated that the performance of proposed results was better than traditional Kalman filter.

This paper is organized as follows. Preliminaries and theoretical analysis such as median absolute deviation, outlier impact for SMF and SMAP algorithm are given in Section 2, and spatial illustration of the outlier impact for employed SMAP algorithm update is provided. The numerical simulation is presented in Section 3 to illustrative the outlier impact for SMAP, and the detailed comparison between SMAP and AP are proposed. We concluded this paper in Section 4.

2. Preliminaries and Theoretical Analysis

2.1 Median Absolute Deviation

Real data usually contain many unreasonable, irrelevant, noisy one and missing value etc. Data with bad quality can seriously affect the convergence speed of employed algorithm [10, 11].

Consider a data sequence (x_1, x_2, \dots, x_N) and its median value x^* , and in terms of Davies's work [13] which has shown that the median absolute deviation (MAD) from the median, is more robust with good performance in the presence of multiple outliers. More precisely, the large samples in normally distributed x_i are approximately equal to the population standard deviation [12,13]. Hampel is sensitive to the presence of outliers and remove the non-robust statistics [14]. The following equation denotes the outlier value filtration principle of Hampel filter,

$$|x_i - x^*| > TS, \quad S = 1.4826 \text{median} \left\{ |x_i - x^*| \right\}_{i=1, \dots, N}.$$

where factor 1.4826 is defined so that the expected value is equal to the standard deviation. T is the threshold which is used to control range of convergence based on the defined Euclid distance, and it can be estimated the sample standard deviation of the distribution via given value in this paper.

2.2 Outliers Impact related to SMF

SMAP algorithm is a data selective adaptive filtering algorithm, which develops the idea of set-membership normalized least mean squares (SMNLS) filter with constraint sets. Set-membership filtering (SMF) is specified as follows [7], let S_k represent the set of all possible input and desired time series pairs (x_k, d_k) , and

$$\Theta = \bigcap_{(x_k, d_k) \in \bigcup_{k \in \Lambda} S_k = S} \left\{ w \in R^N : (d_k - w^T x_k) \in N(g_k, \gamma_k), \gamma = \max_{k \in \Lambda} \{\gamma_k\} \right\} \quad (2.1)$$

where γ_k is the upper bound related to magnitude of the estimation error $e_k = d_k - w^T x_k$, there are maybe existed many w for a given properly bound γ_k . g_k is the center with respect to estimation error in constraint set H_k which contains all the vectors w at time instant k , $N(g_k, \gamma_k)$ represents the neighborhood of g_k via radius γ_k . The membership set φ_k is given by

$$\varphi_k = \bigcap_{k \in \Lambda} H_k = \bigcap_{k \in \Lambda} \left\{ w \in R^N : (d_k - w^T x_k) \in N(g_k, \gamma_k) \right\} \quad (2.2)$$

Essentially, (2.2) is not easy to be calculated, so adaptive approaches are needed [11,12]. According to the illustration with respect to γ_k , if the given time series contains unreasonable, irrelevant and noisy one, in particular, there exists some relative large value in given time series, because the formula (2.3)

$$\left| d_k - w^T x_k - w^T \eta_k + w^T \eta_k \right| \leq \left| d_k - w^T \eta_k \right| + \left| w^T x_k - w^T \eta_k \right| \leq \gamma_{new}^{(k)} \quad (2.3)$$

where $\gamma_{new}^{(k)} = \gamma_k + w_m (\hat{\sigma} + \delta) \cdot \Phi^{-1}(3/4)$, Φ^{-1} is the

inverse of the cumulative distribution function for the standard normal distribution, δ is a constant which subject to $\max_{k \in \Lambda} |x_k - \eta_k| \leq \hat{\sigma} + \delta$, $\eta_k = \text{median}_{k \in \Lambda} \{x_k\}$. $w_m = \max \{w \in R^N\}$ is the estimated value of the sample variance σ . Then (1) can be denoted as,

$$\Theta = \bigcap_{(x_k, d_k) \in \bigcup_{k \in \Lambda} S_k = S} \left\{ w \in R^N : (d_k - w^T x_k) \in N(g_k, \gamma_{new}^{(k)}), \gamma = \max_{k \in \Lambda} \{\gamma_{new}^{(k)}\} \right\} \quad (2.4)$$

In other words, the convergence range is amplified by the influence of outliers of time series, and the cumulative effects can reduce the convergence speed of set-membership filtering when feasibility set Θ is traversed up to time instant k if there are many outliers in given time series. Similarly, (2.2) is denoted by

$$\varphi_k = \bigcap_{k \in \Lambda} H_k = \bigcap_{k \in \Lambda} \left\{ w \in R^N : (d_k - w^T x_k) \in N(g_k, \gamma_{new}^{(k)}) \right\} \quad (2.5)$$

2.3 Outliers Impact related to SMAP Algorithm

The purpose of general SMAP algorithm is to update the current coefficient related to w_k at the k th iteration such that

$$w_{k+1} \in \varphi_k^p = \bigcap_{i=k-p+1}^k H_i \quad (2.6)$$

where φ_k^p denotes the intersection of the p last constraint sets, without loss of generality, set

$$\left| \frac{\mathbf{w}_{k+1} - \mathbf{w}_k}{\lambda_k} + \boldsymbol{\eta}_k \right| \leq |\mathbf{x}_k - \boldsymbol{\eta}_k| \leq \mathbf{w}_k (\hat{\sigma} + \delta) \cdot \Phi^{-1}(3/4) = \gamma_{new}, \lambda_k \in R^{p+1} \setminus \{0\} \quad (2.13)$$

Combing the (2.9)-(2.13) into (2.14) and (2.15), we have,

$$\mathbf{x}_k^T \lambda_k = \mathbf{x}_k^T (\mathbf{w}_{k+1} - \mathbf{w}_k) \approx \mathbf{x}_k^T (\mathbf{d}_k - \mathbf{x}_k^T \mathbf{w}_{k+1} - \mathbf{g}_k) \leq \mathbf{x}_k^T \gamma_{new} \quad (2.14)$$

$$\mathbf{e}_k - \mathbf{g}_k = \mathbf{d}_k - \mathbf{x}_k^T \mathbf{w}_{k+1} - \mathbf{g}_k \leq \gamma_{new}, \quad \mathbf{e}_k = [\mathbf{e}_k, \mathbf{e}_{k-1}, \dots, \mathbf{e}_{k-p+1}] \in R^{p+1} \quad (2.15)$$

The intersection of feasible region of (2.14) and (2.15) is approximately equal to latter if $|\mathbf{x}_k| \geq 1$, then we obtain,

$$\lambda_k = (\mathbf{e}_k - \mathbf{g}_k) (\mathbf{x}_k^T \mathbf{x}_k)^{-1} \quad (2.16)$$

Then, the following formula can be obtained by combing (2.12) with (2.16),

$$\mathbf{w}_{k+1} = \begin{cases} \mathbf{w}_k + \mathbf{x}_k (\mathbf{e}_k - \mathbf{g}_k) (\mathbf{x}_k^T \mathbf{x}_k)^{-1}, & \text{if } |\mathbf{e}_k| > \gamma_{new} \\ \mathbf{w}_k, & \text{otherwise} \end{cases} \quad (2.17)$$

Remark 1. If value of \mathbf{x}_k is less than one, or even in a small neighbor of origin, (2.16) cannot be established because their convergence points of convergence range are not equal to each other such that the intersection of feasible region of (2.14) and (2.15) are not same. Otherwise, the feasible region of (2.14) can be expanded if the value of \mathbf{x}_k is relative large, moreover, if there are many outliers with large value in the feasible region of (2.14), then the intersection of feasible region of (2.14) and (2.15) become more smaller than before such that convergence speed of employed algorithm is slow.

3. Experimental Results

In this Section, the simulation results are given to verify the effectiveness of the proposed results. The update type is employed by the choice 2 of [7] for the SMAP algorithm which is used to identify a system with order $N-1=254$. The median Hampel filter based on filter window (half windows size was set to 3) is applied to handle the irrelevant, unreasonable and outliers data of the input samples, and the state trajectories of original and filtered data are shown in Fig.2.

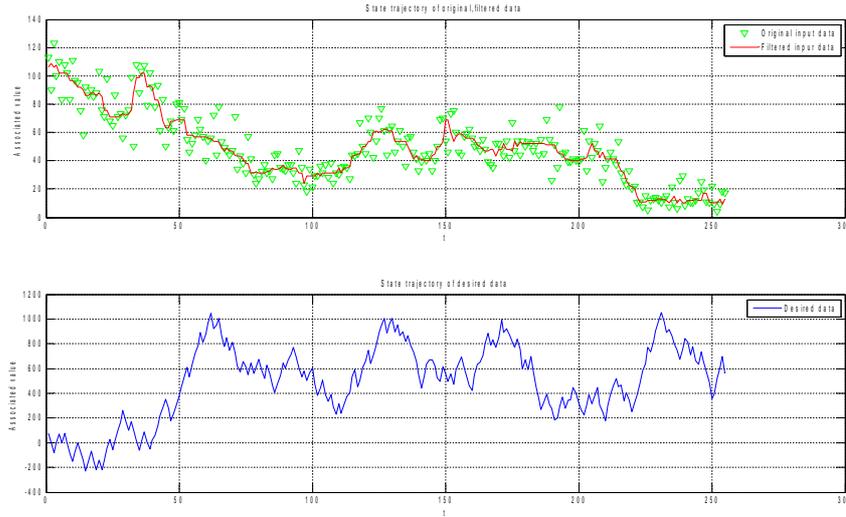


Figure 2: State trajectories of utilized data.

The signal-to-noise ratio (SNR) of SMAP and AP are set to 120dB and 230dB, respectively. The bound γ_{new} is , $\sqrt{5\sigma_k^2} = \sqrt{5*(0.2)^2}$ where σ is the variance of the noise, and it indicates the noise power. The number of iterations and realizations within the ensemble are chosen to 2000 and 255, respectively. The learning curves related to Mean squared error (MSE) for SMAP with data reuse factor $L=10, 20, 30, 60$ and 150 constraint set are shown in Fig. 3.

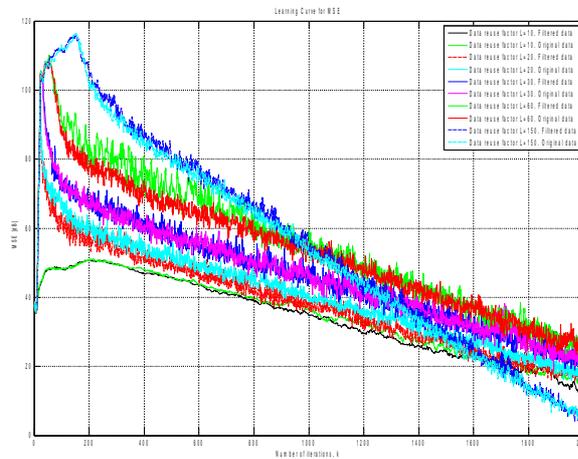


Figure 3: Learning curves for MSE (SMAP)

Moreover, the detailed comparisons between AP and SMAP before and after data filter are given in Tables 1 and 2, respectively.

AP	L=10	L=20	L=30	L=60	L=150
OR (ET)	107.74	252.71	329.71	834.20	5959.34
IM (ET)	93.17	209.26	311.03	799.87	5034.06

Table 1: Simulation results related to AP

SMAP		L=10	L=20	L=30	L=60	L=150
	ANUL	1949.67	1949.66	1943.35	1916.71	1854.96
OR	ANULL	97.48	97.48	97.16	95.83	92.74
	ET	42.44	81.66	129.92	286.70	1367.24
	ANUL	1934.64	1930.25	1940.72	1915.27	1849.38
IM	ANULL	96.73	96.51	97.03	95.76	92.46
	ET	39.94	81.54	123.93	281.41	1321.27

Table 2: Simulation results related to Fig.4

OR: Original algorithm; IM: Improved algorithm; ANUP: Average number of updates performed by SMAP Algorithm; ANUPP: Percent of ANUP in a total of 2000 iterations; ET: Elapsed time in seconds.

4.Conclusion

This paper addressed an improved SMAP algorithm by utilizing the MAD method and SMAP algorithm. Firstly, we introduce the MAD technique which is used to eliminate the unreasonable, irrelevant and noisy data such that the given data become relative smooth. Secondly, we give the theoretical analysis for the outlier impact for SMF such as convergence speed and range. Thirdly, we discuss the outlier impact for SM-AP algorithm, and provide the spatial illustration of SMAP algorithm update. Finally, we give an illustrative example through numerical simulation which is used to verify the effectiveness of proposed method.

References

- [1] Fogel, Eli, Huang Yih-Fang. *On the value of information in system identification-bounded noise case*. Automatica, 1982, 18(2): 229-238.
- [2] Paulo S. R. Diniz. *Adaptive filtering: algorithms and practical implementation*. Springer, 2013
- [3] Deller, John R. *Set membership identification in digital signal processing*. ASSP Magazine, IEEE, 1989, 6(4): 4-20.
- [4] Sridhar Gol lamudi, Shirish Nagaraj, Samir Kapoor, Huang Yih-Fang. *Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size*. Signal Processing Letters, IEEE, 1998, 5(5): 111-114.
- [5] Galdino, Juraci F., Jose Apolinario, and Marcello LR De Campos. *A set-membership NLMS algorithm with time-varying error bound*. Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on. IEEE, 2006: 4 pp.-280.
- [6] Kosut, Robert L., Ming K. Lau, and Stephen P. Boyd. *Set-membership identification of systems with parametric and nonparametric uncertainty*. Automatic Control, IEEE Transactions on, 1992, 37(7): 929-941.
- [7] Werner, Stefan, and Paulo SR Diniz. *Set-membership affine projection algorithm*. Signal Processing Letters, IEEE, 2001, 8(8): 231-235.

-
- [8] Werner, Stefan, Paulo SR Diniz, and José EW Moreira. *Set-membership affine projection algorithm with variable data-reuse factor*. Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on. IEEE, 2006: 4 pp.
- [9] Huang, Yi, Zongji Chen, and Chen Wei. *Least trace extended set-membership filter*[J]. Science China Information Sciences, 2010, 53(2): 258-270.
- [10] Famili, Fazel; Shen, W.-M.; Weber, R.; Simoudis, E. *Data pre-processing and intelligent data analysis*. International Journal on Intelligent Data Analysis 1997;1(1).
- [11] Noriega, Gerardo, and Subbarayan Pasupathy. *Application of Kalman filtering to real-time preprocessing of geophysical data*. Geoscience and Remote Sensing IEEE Transactions 1992;30(5):897-910.
- [12] David Ruppert, David S. Matteson. *Statistics and data analysis for financial engineering*. Springer, 2011.
- [13] Davies, Laurie, and Ursula Gather. *The identification of multiple outliers*. Journal of the American Statistical Association 1993;88(423):782-792.
- [14] Allen, David P. *A frequency domain Hampel filter for blind rejection of sinusoidal interference from electromyograms*. Journal of neuroscience methods 2009;177(2):303-310.
- [15] Pearson, Ronald K. *Outliers in process modeling and identification*. Control Systems Technology IEEE Transactions 2002;10(1):55-63.