Analysis of Usefulness of a Fuzzy Transform for Industrial Data Compression

Anna Sztyber

Warsaw University of Technology, Faculty of Mechatronics, Institute of Automatic Control and Robotics, Boboli 8, 02-525 Warsaw, Poland

E-mail: a.sztyber@mchtr.pw.edu.pl

Abstract. This paper presents the first part of an ongoing work on detailed analysis of compression algorithms and development of an algorithm for implementation in a real industrial data processing system. Fuzzy transforms give promising results in an image compression. The main aim of this paper is to test the possibility of an application of the fuzzy transforms to the industrial data compression. Test are carried out on the data from DAMADICS benchmark. Comparison is provided with a piecewise linear compression, which is nowadays the standard in the industry. The last section contains discussion of the obtained results and plans for the future work.

1. Introduction

Nowadays, in industrial plants huge amount of data is measured, processed, and archived, so an issue of data compression arises. The question is not only how to compress collected sensor data, but also whether they should be compressed at all?

The cost of storage devices for data archiving is decreasing, but it is still nonzero. There are also problems of processing large data sets and network traffic.

On the other hand, advanced techniques of data mining and data processing are becoming more common, so there is an issue of the impact of the data compression on the results of calculations.

The archival process data can be used for:

- calculation of statistics, like means and standard deviations,
- balance calculations,
- feature extraction,
- control loop monitoring,
- building cause-effect process models,
- tuning diagnostic systems,
- identification of models (fuzzy, neural, parametric) for fault detection,
- synthesis of control algorithms, like model predictive controller.

Due to the author specialisation the above list contains mainly tasks connected with fault diagnosis and process monitoring and is probably not complete.

Analysis of the impact of the compression on data driven process analyses was presented in [1]. Authors point out that standard compression methods change means, standard deviations, non-linearity indices, and Harris index. Moreover the compression can change signal spectrum, introduce nonlinearities, and dump higher frequency oscillations. Even a noise reduction can be a flaw, because it changes the variation and Harris index. Authors of [1] insist, that calculations should be carried out using the original data. There are also analyses of the impact of the data compression on control loop monitoring [2], pattern matching [3], and PCA models [4].

All above considerations regards the standard industrial compression algorithms. Nowadays, the standard is a piecewise linear trending. Among the most widespread solutions there are:

- Boxcar [5]–archived are only data points, which differ more than the predefined threshold from previous recorded value;
- Backward Slope [5]–signal is approximated by a piecewise linear function, the line slope is determined by the two previously recorded values. Archived are only data points, which differ more than the predefined threshold from the calculated slope;
- Boxcar-Backward Slope–combination of the two methods, data point must exceed both thresholds to be archived;
- Swinging Door [6]–signal is approximated by a piecewise linear function, data points are limited by two line segments with specified maximal distance;

In all methods data points are archived with a time stamp. All the standard methods are very simple and easy to implement, but not very sophisticated. Nowadays, data compression is a large scope of research, mainly due to the multimedia applications. So the question arises, if some solutions can be adapted to improve the industrial data compression. In this paper a fuzzy transform is analysed.

The paper is structured as follows. Definitions of fuzzy transforms are recalled in Sect. 2. Results of calculations are described in Sect. 3, and results for piecewise linear compression in Sect. 3.1. Discussion of the results is presented in Sect. 4.

2. Fuzzy Transform

Fuzzy sets were defined to represent uncertain information [7]. The fuzzy logic is used in the industry for synthesis of fuzzy controllers, fuzzy modelling (Wang-Mendel or Takagi-Sugeno-Kanga models), and to deal with uncertain information in diagnostic systems.

The fuzzy transforms have applications in the image compression in order to avoid JPEG blocking artifacts [8–10]. Main idea of a fuzzy compression is to cover function domain by the set of fuzzy sets, called a fuzzy partition, and represent function by a smaller set of transform coefficients. There is one coefficient for each fuzzy set in the partition. The coefficients are the best in the mean square sense with respect to the selected partition.

Definitions of fuzzy partitions and fuzzy transforms will be cited after [8].

Definition 1 (Fuzzy partition [8]) *Let $x_1 < \ldots < x_n$ be fixed nodes within $[a, b]$, such that $x_1 = a$, $x_n = b$ and $n \geq 2$. We say that fuzzy sets $A_1, \ldots, A_n$, identified with their membership functions $A_1(x), \ldots, A_n(x)$, defined on $[a, b]$, form a* ***fuzzy partition*** *of $[a, b]$ if they fulfil the following conditions for $k = 1, \ldots, n$:*

(i) $A_k : [a, b] \to [0, 1], A_k(x_k) = 1,$

(ii) $A_k(x) = 0,$ *if* $x \notin (x_{k-1}, x_{k+1}),$

(iii) $A_k(x)$ *is continuous,*

(iv) For $k = 2, \ldots, n$ $A_k(x)$ monotonically increases on $[x_{k-1}, x_k]$, and for $k = 1, \ldots, n-1$ $A_k(x)$ monotonically decreases on $[x_k, x_{k+1}]$,

(v) $\forall x \in [a, b] \Sigma_{k=1}^{n} A_k(x) = 1.$

The partition is called uniform if points $x_1, \ldots, x_n$ are equidistant. Exemplary uniform fuzzy partitions are shown in Fig. 1 and Fig. 2.

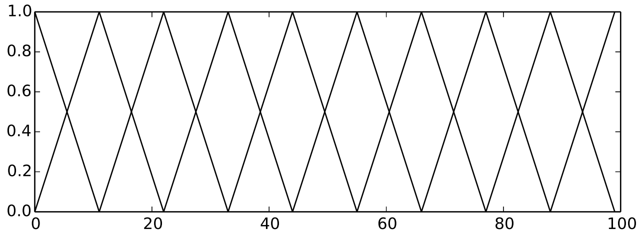

Figure 1. Uniform fuzzy partition of $[0, 100]$ with triangular shaped functions.

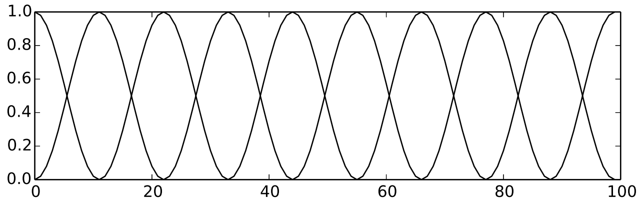

Figure 2. Uniform fuzzy partition of $[0, 100]$ with sinusoidal shaped functions.

The fuzzy transform, called also F-transform, is a linear transformation. The transform coefficients are weighted mean values of the transformed function with respect to the fuzzy partition. Discrete fuzzy transform and inverse discrete fuzzy transform are defined as follows.

Definition 2 (Discrete F-transform [8]) *Let a function f be given at nodes $p1, \ldots, p_m \in [a, b]$, $A_1, \ldots, A_n$ be basic functions which form a fuzzy partition of $[a, b]$. We say that the n-tuple of real numbers $[F_1, \ldots, F_n]$ is the* ***discrete F-transform*** *of f with respect to $A_1, \ldots, A_n$ if:*

$$F_k = \frac{\Sigma_{i=1}^{m} f(p_i) A_k(p_i)}{\Sigma_{i=1}^{m} A_k(p_i)}.$$

Definition 3 (Inverse discrete F-transform [8]) *Let function $f(x)$ be given at nodes $p1, \ldots, p_m \in [a, b]$, and $[F_1, \ldots, F_n]$ be the discrete F-transform of f w.r.t. $A_1, \ldots, A_n$. Then the function*

$$f_{F,n}(p_i) = \Sigma_{k=1}^{n} F_k A_k(p_i),$$

defined at the same nodes, is the ***inverse discrete F-transform****.*

For each data point p_i only two membership functions have nonzero values, so the fuzzy transform can be calculated in time $O(m)$, were m denotes the number of data points.

There is also a possibility of using some t-norm instead of weighted average [9, 10], but it introduces additional artifacts into the recovered signal.

3. Results

Numerical experiments with fuzzy compression were carried out on real industrial data from DAMADICS benchmark [11]. Measurements of juice pressure at valve inlet ($P51\_05$, in the following abbreviated to $P1$) were used. The range of values was 0-1000kPa. The data was collected each second, overall batch from one day of 86400 points was processed.

Firstly, tests were carried out with uniform fuzzy partitions. Results are shown in Fig. 3. Original signal is shown in the upper plot ($P1$). The signal was sampled with high speed, so large compression ratios can be considered. Signals $P1_{FTT1}$, $P1_{FTT2}$, $P1_{FTT3}$ were reconstructed respectively from 200, 2000 and 8000 fuzzy transform coefficients.

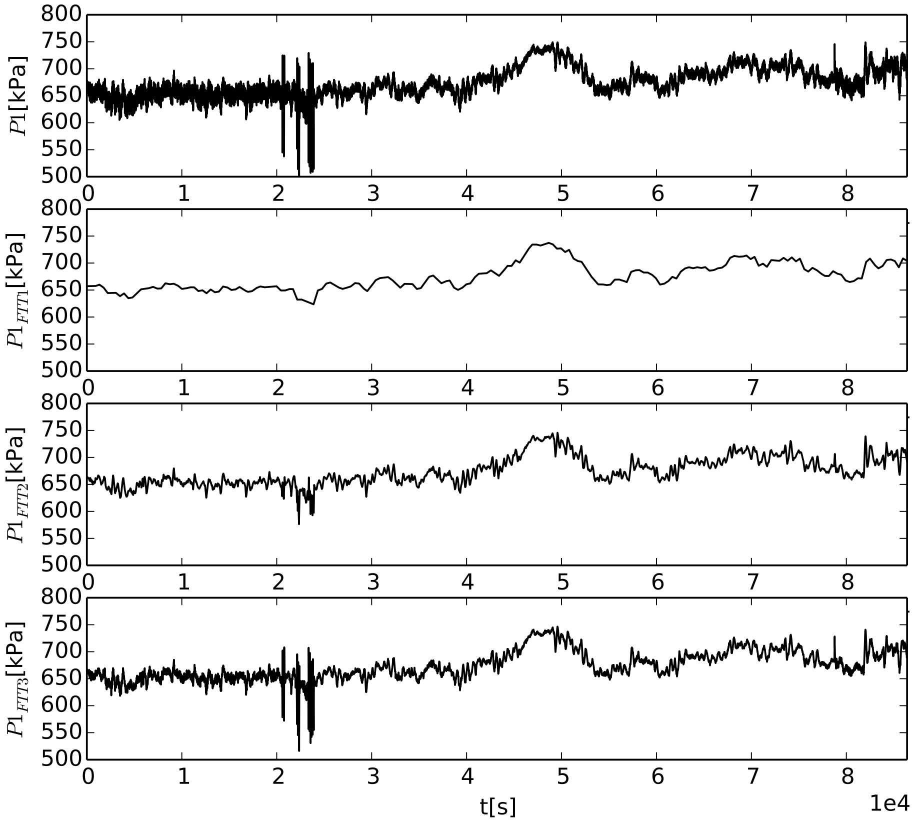

Figure 3. Fuzzy compression with uniform, triangular shaped fuzzy partition.

Results of compression are summarised in Tab. 1. The results contain mean squared error mse, maximal error $maxe$, and difference of a mean of original signal and mean of recovered signal $emean$. All errors are in kPa. Number of the fuzzy transform coefficients is denoted by n. Results of transformation with sinusoidal shaped functions are included in the table, denoted respectively by $P1_{FTS1}$, $P1_{FTS2}$, and $P1_{FTS3}$. Results for sinusoidal shaped membership functions are slightly better.

Table 1. Results of fuzzy compression with uniform fuzzy partition.

Signal	$P1_{FTT1}$	$P1_{FTT2}$	$P1_{FTT3}$	$P1_{FTS1}$	$P1_{FTS2}$	$P1_{FTS3}$
n	200	2000	8000	200	2000	8000
mse	10.13	6.61	4.24	9.98	6.52	4.14
$emax$	135.6	106.6	44.7	135.3	105.9	42.0
$emean$	0.01	0.0002	0.0002	0.01	0.0001	0.0002

Difference between a signal compressed with triangular and sinusoidal shaped membership functions can only be seen while zooming in (Fig. 4). Both methods introduce their characteristic artifacts. With the triangular shaped membership functions decompressed signal is piecewise linear.

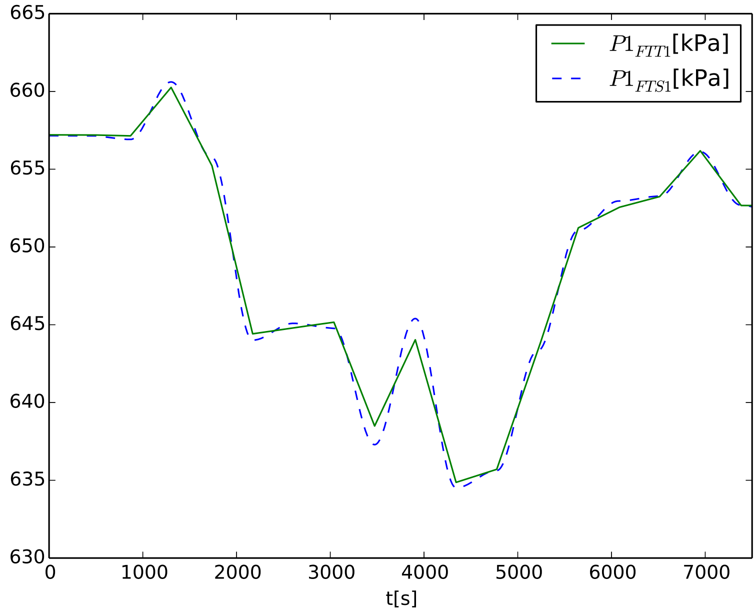

Figure 4. Signals compressed with triangular ($P1_{FTT1}$) and sinusoidal ($P1_{FTS1}$) shaped membership functions.

The fuzzy transform has smoothing properties, therefore some quickly varying signal components are eliminated. This causes large maximal error. On the other hand, distraction in signal mean is negligible. The question arises, if it is possible to better maintain quickly varying signal components. The natural answer is to use a non-uniform partition. This approach was used for the image compression in [9].

Next part of the tests was carried out with the non-uniform fuzzy partitions. The points forming partition $x_1, \ldots, x_n$ were placed with density approximately proportional to the original signal variation. Example of the non-uniform partition is shown in Fig. 5. There is a larger number of fuzzy functions for quickly varying part of the signal. For clarity only 25 sets were drawn.

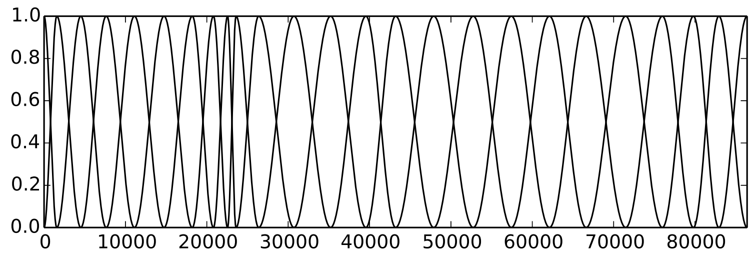

Figure 5. Example of non-uniform fuzzy partition with sinusoidal shaped functions.

Signals compressed with non-uniform partition are shown in Fig. 6. Results are listed in Tab. 2. Method for finding non-uniform partition proportional to signal variance can return slightly different size of the obtained partition then desired, so actual numbers are provided in the table. Signals $P1_{FTT1var}$, $P1_{FTT2var}$, and $P1_{FTT3var}$ were compressed with triangular membership functions, and $P1_{FTS1var}$, $P1_{FTS2var}$, and $P1_{FTS3var}$ with sinusoidal.

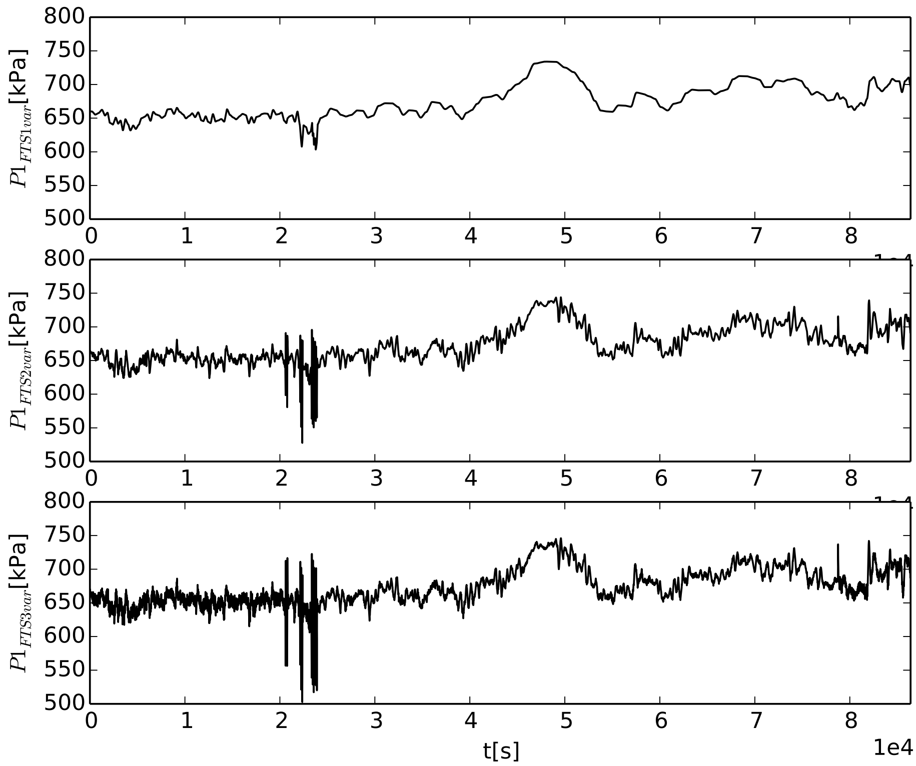

Figure 6. Fuzzy compression with non-uniform, sinusoidal shaped fuzzy partition.

Table 2. Results of fuzzy compression with non-uniform fuzzy partition.

Signal	$P1_{FTT1var}$	$P1_{FTT2var}$	$P1_{FTT3var}$	$P1_{FTS1var}$	$P1_{FTS2var}$	$P1_{FTS3var}$
n	201	2002	7995	201	2002	7995
mse	9.99	5.23	3.83	9.87	5.12	3.76
$emax$	116.4	64.0	26.1	116.9	60.8	25.2

Results for the non-uniform partition are clearly better then for the uniform partition, but there is one drawback. For uniform partition only the transform coefficients must be written into the archived file. In case of the non-uniform partition time stamp for each coefficient is needed.

3.1. Standard compression

The obtained results should be compared with the standard compression methods. For this purpose program was written to simulate a compression of a system OSISoft PI. The compression has two steps: the first is a Boxcar algorithm, and the second is described as a Swinging Door algorithm, however the provided description [12, 13] is not exactly consistent with the original algorithm [6]. Purpose of the program was not to precisely mimic PI compression, but to have some representation of the piecewise linear compression.

Algorithm needs two parameters: $ExDev$ - threshold for the Boxcar algorithm, and $ExComp$ - threshold for the Swinging Door algorithm. It is not straightforward to obtain desired compression ratio using this two parameters, so numbers of recorded points will be slightly different then numbers of transform coefficients.

Results are presented in Fig. 7. Signals $P1_{PI1}$, $P1_{PI2}$, and $P1_{PI3}$ were recovered respectively from 201, 2002 and 7995 data points. Numerical results are summarised in Tab. 3. The number of recorded points is denoted by n.

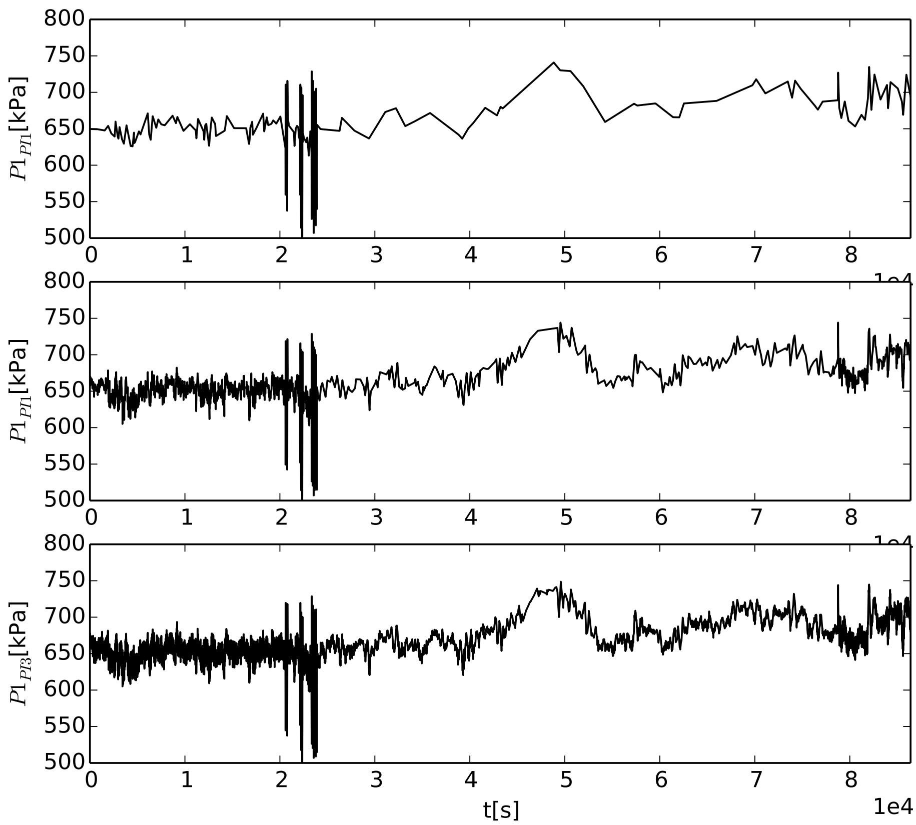

Figure 7. Compression by the piecewise linear approximation.

Table 3. Results of the compression by the piecewise linear approximation.

Signal	$P1_{PI1}$	$P1_{PI2}$	$P1_{PI3}$
n	201	2002	7995
mse	10.86	5.57	3.37
$emax$	44.09	21.2	14.9
$emean$	1.16	0.05	0.02

The piecewise linear compression gives smaller maximal error. Mean square errors are comparable. Fuzzy transforms gives slightly better results for higher compression ratios. Fuzzy compression gives good results in the recovery of the signal mean.

4. Discussion

Fuzzy transform is a possible method of the data compression. In the terms of the mean square error it gives results comparable with standard methods. There is a wide variety of other methods, that can be applied to this problem. In this section characteristic features of fuzzy transform will be summarised, so it can be compared with existing approaches.

The fuzzy compression has following advantages:

- very little error of mean of the decompressed signal, so it is suitable for balance calculations,
- good preservation of lower frequency signal components,
- simple signal transformation,
- during decompression each signal sample is calculated using only two transform coefficients, so there is a possibility to decode only fragment of the signal (advantage over Fourier, Discrete Cosine, and Wavelet Transform),
- compression ratio depends on the number of membership functions in the fuzzy partitions and is known in advance (advantage over piecewise linear trending).

The fuzzy transform has also smoothing and denoising properties. It is a good feature in most applications. Visual appearance of the signal is better. On the other hand, in some applications reduction of the variance can be a drawback [1].

In the fuzzy transform compression ratio is chosen by the user of the system. In the piecewise linear trending user defines thresholds for Boxcar and Swinging Door algorithms. Relation between this parameters and compression ratio depends on the processed signal and is hard to determine, therefore often parameters values are inappropriate, and signal is over-compressed or under-compressed [14, 15].

The fuzzy compression has following disadvantages:

- dumping of higher frequencies,
- large maximal error.

Maximal error in the fuzzy compression is not known in advance. This feature is unacceptable in the safety critical measurements. In the piecewise linear trending maximal error is determined by the thresholds.

5. Conclusions

In this paper analysis of the fuzzy transform as a tool for the industrial data compression was presented. The fuzzy transform has a number of promising features, but obtained results, in the author opinion, are not good enough.

Many authors [2–4, 16, 17] indicates that other transforms like Fourier Transform, Discrete Cosine Transform and particularly Wavelet Transform work well in the data compression. Therefore, the next step will be to test this solutions.

Another important issue is a problem of assessment of a compression algorithm. In this paper rather naive or straightforward measures of compression quality were considered. Mean square error is the most popular measure, but it does not take into account signal dynamics. Therefore, there is a need to determine a set of measures, that will cover all spectrum of desired compression algorithm features.

It could also be noted, that very wide variety of compression ratios was considered. In large industrial plants it is impossible to tune compression of each measurement separately, therefore there is a need for a tool for automatic selection of compression parameters.

Acknowledgments
The work was partially financed from the Innotech project of the National Centre for Research and Development, number: Hi3/227511 and by grant of Dean of Faculty of Mechatronics.

References
[1] Thornhill N F, Choudhury S and Shah S L 2004 *Journal of Process Control* **14** 389–398
[2] Sivalingam S and Hovd M 2011 *Control Automation (MED), 2011 19th Mediterranean Conference on* pp 594–599
[3] Singhal A and Seborg D E 2003 *American Control Conference*
[4] Imtiaz S A, Shoukat Choudhury M A A and Shah S L 2007 *Industrial & Engineering Chemistry Research* **46** 481–491
[5] Hale J C and Sellars H L 1981 *Chemical engineering progress.* **77** 3843
[6] Bristol E 1990 *ISA National Conference Proceedings* pp 749–754
[7] Zadeh L 1965 *Information and Control* **8** 338 – 353 ISSN 0019-9958 URL `http://www.sciencedirect.com/science/article/pii/S001999586590241X`
[8] Perfilieva I 2006 *Fuzzy Sets and Systems* **157** 993 – 1023
[9] Nobuhara H, Pedrycz W, Sessa S and Hirota K 2006 *Information Sciences* **176** 2526 – 2552
[10] Hirota K and Pedrycz W 2002 *Fuzzy Sets and Systems* **126** 325 – 335
[11] Bartyś M, Patton R, Syfert M, de las Heras S and Quevedo J 2006 *Control Engineering Practice* **14** 577 – 596 ISSN 0967-0661 a Benchmark Study of Fault Diagnosis for an Industrial Actuator A Benchmark Study of Fault Diagnosis for an Industrial Actuator URL `http://www.sciencedirect.com/science/article/pii/S0967066105001796`
[12] Benvenuto P 2012 Pi 101, module 10: Compression algorithm URL `http://community.osisoft.com/index.php?/topic/1061-pi-101-module-10-compression-algorithm/`
[13] OSIsoftLearning 2012 Osisoft: Exception and compression full details URL `https://www.youtube.com/watch?v=89hg2mme7S0`
[14] Alsmeyer F 2006 *16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering* (*Computer Aided Chemical Engineering* vol 21) ed Marquardt W and Pantelides C (Elsevier) pp 1533 – 1538
[15] Pettersson J and Gutman P O 2004 *Journal of Process Control* **14** 431 – 439 ISSN 0959-1524 URL `http://www.sciencedirect.com/science/article/pii/S0959152403000969`
[16] Watson M, Liakopoulos A, Brzakovic D and Georgakis C 1995 *American Control Conference, Proceedings of the 1995* vol 2 pp 1265–1269
[17] Watson M J, Liakopoulos A, Brzakovic D and Georgakis C 1998 *Industrial & Engineering Chemistry Research* **37** 267–274