Computer Assisted Mechanics and Engineering Sciences, **14**: 353–360, 2007.

Data filtering using dynamic particles method

Łukasz Rauch, Jan Kusiak
Department of Applied Computer Science and Modelling
AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Kraków, Poland

(Received in the final form April 23, 2007)

The identification of the industrial processes is a complex problem, especially in the case of signals denoising. The holistic approaches used for signal denoising processes are recently considered in various types of applications in the domain of experimental simulations, feature extraction and identification. A new signal filtering method based on the dynamic particles (DP) approach is presented. It employs physics principles for the signal smoothing. The presented method was validated in the identification of two kinds of input data sets: artificially generated data according to a given function $y = f(x)$ and the data obtained in laboratory mechanical tests of metals. The algorithm of the DP method and the results of calculations are presented. The obtained results were compared with commonly used denoising techniques including weighted average, neural networks and wavelet analysis. Moreover the assessment of the results' quality is introduced.

1. Introduction

The analysis of the experimental measurement data obtained in the identification process is often difficult and sometimes the results in their rough version are useless, because of superimposed noise. In most cases observed noise is a result of external factors like sensitivity of the industrial measuring sensors. Properly performed analysis based on the denoising techniques allows extracting the vital part of the measurements. Due to the denoising process, which is often very expensive and time-consuming, the experimental data can be restored and used in further calculations [9]. Commonly used denoising methods have some advantages and disadvantages, but no one can be treated as the unified denoising and smoothing method. The unification of such techniques shall give a method, which can be applied for different types of measurement data saddled with a noise of different type. Usually methods have to be reconfigured and adapted to the varying conditions even if the analyzed data has the same form but with the different noise. The example of such data can be seen in Fig. 1, where two similar plots can be found [5]. They contain results of a metal compression tests performed with different velocities. Each of these curves is loaded with noise of different frequencies though they describe the same type of tested material. Therefore denoising methods should be designed to obtain similar results independently of the noise character and, what is more important, independently of the curve shape. This would allow the application of the method in the automated way performing denoising process that won't require reconfiguration of input parameters and additional user's interaction.

The process of denoising of measurement curves can be treated in some cases as the problem of data approximation. However, a lot of other objectives can be assumed for the filtering purposes e.g. determination of rise and fall trends in economical data or data smoothing with simultaneous peak preserving during processing of thermomagnetic coefficient plots. There are many of the algorithms, which support such activities, but the most widely known and used can be enumerated as follows:

- polynomial approximations, weighted average,
- wavelet analysis [1], artificial neural networks [4],

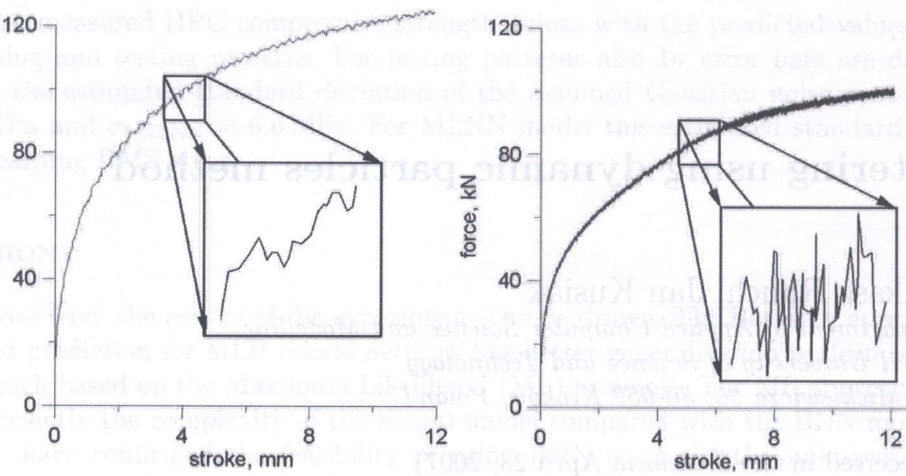

Fig. 1. Example of two different noised measurement data

- large family of convolution methods and frequency based filters [7],
- huge set of representative methods and algorithms of estimation theory [6],
- Kalman stochastic model processing [8],
- dedicated filtering (used mainly in the image filtering processes) e.g. NL-means, neighborhood models [2].

In case of polynomial approximation approach, the algorithms return well-fitting smoothed curves. If the data contain thousands of measured points then the calculation time is very long and the method appears inefficient. However, in case of off-line calculations the time is not a crucial parameter, hence the polynomial approximation approaches are widely used in practice. The weighted average technique allows very fast and flexible data smoothing, but the assessment of obtained results is very difficult and based only on the user's intuition. Thus, the main disadvantage of that method is the problem of a stop criterion of the algorithm. Otherwise the results converge to the straight line joining the beginning and the end of the noised curve. The wavelet analysis is similar to the traditional Fourier method, but is more efficient in the analysis of physical situations, where the signal contains discontinuities and sharp peaks. It allows application of denoising process on different levels of signal decomposition making the solution very precise and controllable. Wavelets are mathematical functions that divide the data into different frequency components. Then the analysis of each component is performed with a resolution matched to the frequency scale. The drawbacks of the method are: the necessity of setting thresholds each time the input data is changing; choosing the quantity of decomposition levels that can be dependent on the noise character. Approach based on the artificial neural networks is also often used. Mainly the Generalized Regression Neural Networks (GRNN) is applied. The results obtained using that technique are smoother than in other methods e.g. wavelet analysis, but the application must also be configured for each calculated data curve. Thus, the neural network approach is suitable for single calculations, but not for the automated application of denoising process.

The main problems of the denoising process are:

- the definition of the stop criterion and the evaluation of the quality results,
- decrease of the computation time of the iterated algorithms run too long in most cases,
- the results are too simplified which makes the further analysis of the denoised data useless.

The objectives of the paper can be divided into following topics:

- Presentation of the algorithm of the DP method as a proposition of unified denoising method that can be applied automatically for different types of measurement data sets, independently of the type of the superimposed noise with discussion on its advantages and disadvantages,
- The analysis of results of various signals using DP method — identification of most important signals' features, evaluation of obtained results,
- Interpretation of results and discussion on the prospectives of designed method in other domains of science as bioengineering, civil engineering or economics.

2. Description of the Dynamic Particles Method

The idea of the Dynamic Particles (DP) method is based on the definition of the particle [3]. The particle can be treated as an object placed in the N-dimensional space. The paper presents the two-dimensional DP method using the particles as two-dimensional vectors related to the measurement data.

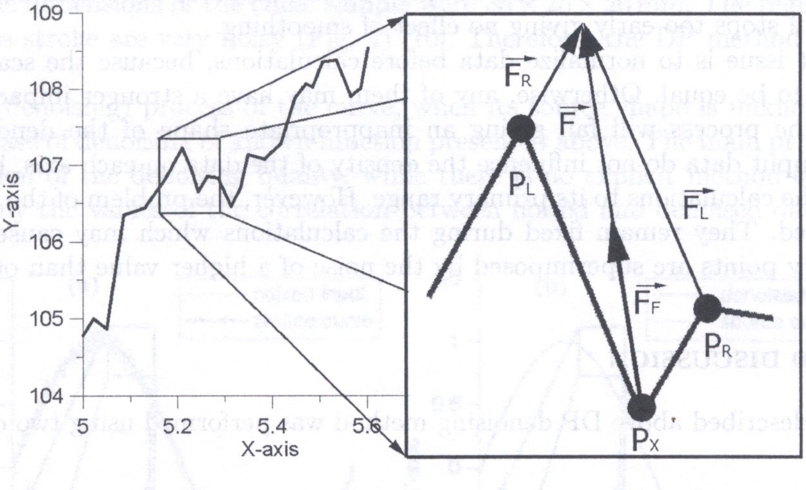

Fig. 2. Example of particles acting with reverse forces F on their neighbors

In case of registered noisy signal $y = f(x)$ of one variable x (see Fig. 2), each peak can be treated as a single dynamic particle. Thus, the whole measurement curve can be seen as a set of connected particles in the 2D Euclidean space. The main idea of the denoising DP algorithm consists in the appropriate move of each particle according to the calculated direction. It is assumed, that the first and the last particles (the beginning and the end of the curve) remain fixed during the running of the denoising algorithm. Every other particle (peak of the noise) is influenced by two neighboring particles. The considered particle moves in the direction that is calculated with respect to the position of two adjacent particles. The algorithm is running in the iterated manner where the number of iterations depends on the complexity of the analyzed measurement data and superimposed noise.

The main reverse force F is the resultant of two force vectors from the left and right particle's neighbors i.e. F_L and F_R. These force vectors are parallel to the opposite slopes of the peaks of the left and right particles and their magnitudes are equal to the lengths of slopes (see Fig. 2). The average force acting currently on the chosen particle can be treated as a particle's potential V_i.

The gradient of this potential is mainly responsible for the movement of the particle in each step of calculations. Thus, the differential equations of particles movement can be written as follows,

$$\begin{cases} m_i \dfrac{d\vec{v}_i}{dt} = -\nabla V_i - f_c \vec{v}_i\,, \\ d\vec{r}_i = \vec{v}_i dt\,, \end{cases} \tag{1}$$

where $\vec{v}_0 \equiv 0$ and $\forall_i m_i = 1$. The magnitude of the force that causes the movement of considered particle is reduced by the friction coefficient $f_c < 1$. It has been found that according to the Newtonian laws of motion if all pre-conditions would be fixed properly, the whole system will remain stable and convergent to the expected results. The friction coefficient f_c can be also modified with the potential of each particle during the calculations of the algorithm, which has a great impact on the final smoothness of the obtained results.

The reductions of the forces, and the reduction of the particles' movements, are the main issues of the stop criterion of the algorithm. If the force acting on the single particle is less than the threshold defined at the beginning then the particle does not move. The whole algorithm reaches the end of the run when all particles stop moving. However, the threshold responsible for the motion of the particles defines also the smoothness of the expected results. If it is set as the small value, then the algorithm is running till all forces on the curve reach the threshold and the differences between positions of two adjacent particles are very low. Otherwise, the plot of new curve is sharper sustaining all most important peaks. The value of this parameter can vary between 10^{-5} and 10^{-20}. If its value is too small, then it has no more impact on the shape of the curve. Otherwise, if is too high, the algorithm stops too early giving no effect of smoothing.

Very important issue is to normalize data before calculations, because the scales on each axis of the curve have to be equal. Otherwise, any of them may have a stronger impact on the motion of particles and the process will fail giving an inappropriate shape of the denoised curve. The normalization of input data do not influence the density of the data on each axis, because the data is re-scaled after the calculations to its primary range. However, the problem of the boundary points has not been solved. They remain fixed during the calculations which may cause some problems while the boundary points are superimposed by the noise of a higher value than other particles.

3. Results and discussion

The validation of described above DP denoising method was performed using two different types of the input data:

- data set generated according to a given equation. The generated data were the values of a chosen function $y = f(x)$ for the values of $x \in [-a, +a]$, with step s. Next, each calculated point $y = f(x)$ of the plot was superimposed by a noise with the random value from the range $[-r, +r]$. The r parameter was set independently for each function, allowing the original generated function to stay unchanged. Thus the expected results returned by the denoising procedure should match the source curve,
- experimental data sets of variables measured in real conditions:
  - load vs. stroke curves measured in the compression tests of the steel samples,
  - thermomagnetic material properties curves obtained during the heat treatment of the magnetic samples,
  - financial series based on the timeline — Euro–USD exchange rates.

As an example of the first data set the sinus function was chosen within the range $[0, 2\pi]$. Then, each calculated value of the $\sin(x)$ was noised with random numbers within the range of $[-0.1, 0.1]$. The expected results of the denoising process should be simply the sinus curve itself. Therefore, working on the generated data the coefficient of the denoising quality can be evaluated by calculating the differences between the source curve, noised input and denoised output as follows,

$$D_q = \frac{calc\_diff\,(S_c\,,\,N_i)}{calc\_diff\,(S_c\,,\,D_o)}, \tag{2}$$

where D_q is denoising quality coefficient; S_c — source curve; N_i — noised input; D_o — denoised output. The *calc_diff* function used in Eq. (2) is defined as the modified standard deviation

$$calc\_diff = \sqrt{\frac{\sum d_i^{\,2}}{n-1}} \tag{3}$$

where d_i is the Euclidean distance between corresponding particles on both curves (various measures of distance, dependent on chosen metrics, can be applied) and n is the number of points (particles). The denoising quality coefficient D_q should be greater than 1. The higher value means the better denoising result. Additional parameter that would allow evaluation of denoise quality is correlation coefficient, which illustrates how the trends of the data are similar for both plots. The results obtained from the calculations performed on the generated data are shown in Fig. 3.

The value of the denoising quality coefficient (Eq. (2)) is equal 5.8854 which indicates that in case of generated function $y = \sin(x)$, the denoised results match to the source curve almost six times better than the noised input (Table 1).

The second analyzed data set consisted of the industrial measurements of compression tests of steel samples. The dimensions of the cubic sample were $35 \times 25 \times 20$ mm. The registered plots of the loads versus press stroke are very noisy (Fig. 4) [10]. Therefore, the DP method was used to filter these plots.

The filtering (denoising) process of the curve, when its source shape is unknown, is much more difficult than in case of denoising of known function presented above. The main problem is connected with the evaluation of the denoising quality, while there is no explicit method of the quality estimation. Thus, only the values of the correlation between noised and denoised data can be counted

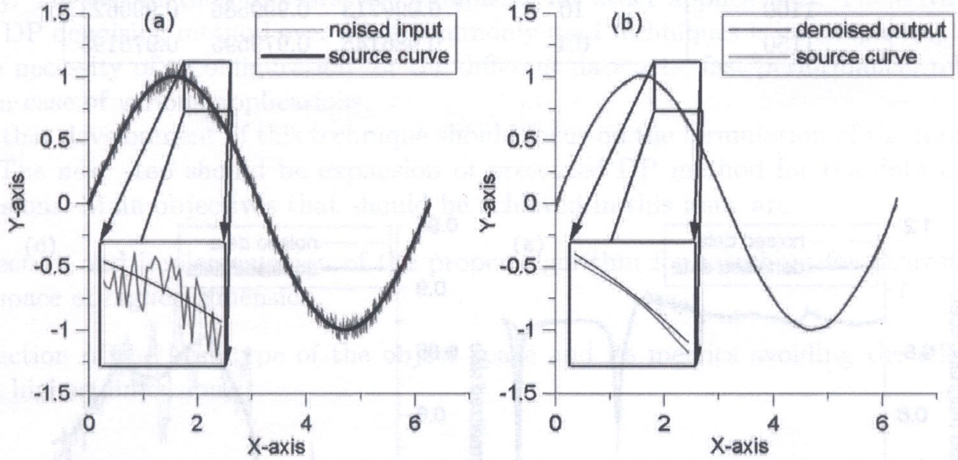

Fig. 3. Comparison of the source and noised curves (a) and denoised output (b)

Table 1. The standard deviations of the analyzed $y = \sin(x)$ noised curve

	Modified standard deviation
source curve – noised input $calc\_diff\,(S_c - N_i)$	**0.0565**
source curve – denoised output $calc\_diff\,(S_c - D_o)$	**0.0096**
denoise quality D_q	**5.8854**

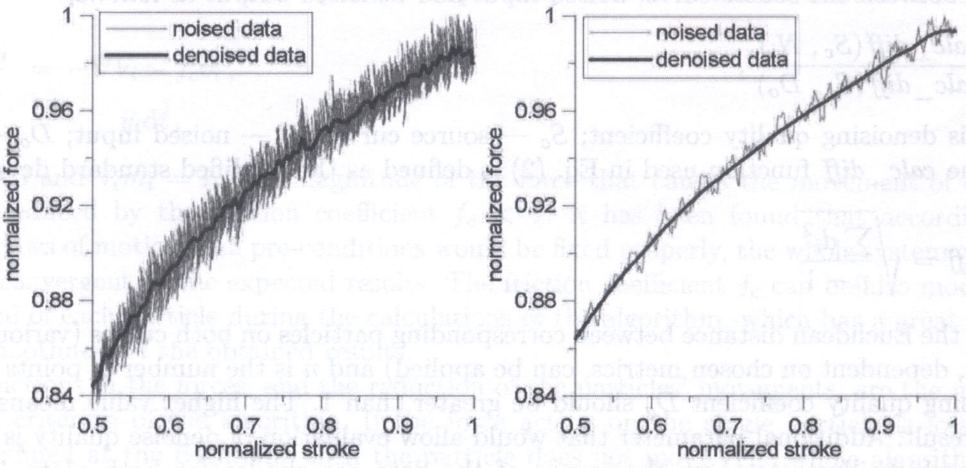

Fig. 4. The results of the DP denoising process of the compression test curves

Table 2. Comparison of denoising results obtained using different methods for load vs. stroke curves

Temperature	Stroke velocity	Correlation coefficient		
°C	s^{-1}	DP	WA	ANN
800	0.1	0.999762	0.999836	0.999821
950	0.1	0.999472	0.999453	0.999520
950	1	0.999521	0.999677	0.999744
950	10	0.999156	0.999825	0.998939
1100	10	0.999213	0.999586	0.999623
1150	0.1	0.986145	0.979595	0.976195

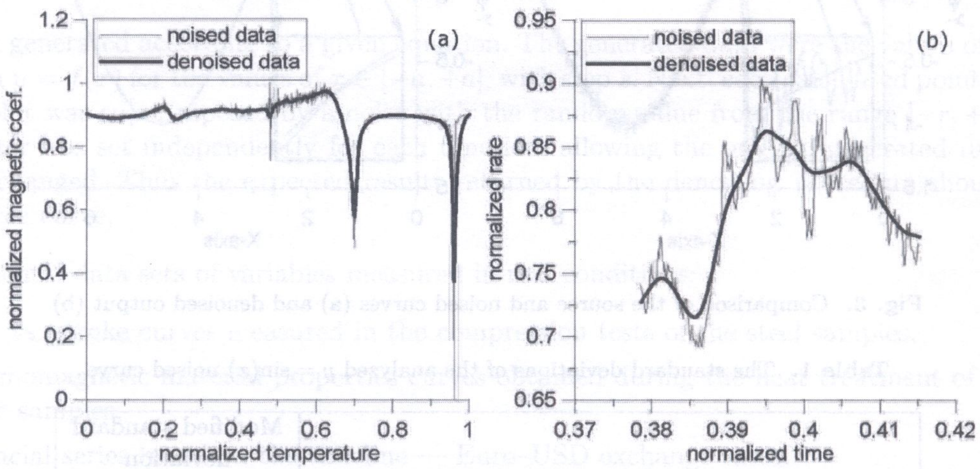

Fig. 5. The results of denoising process for the magnetic material data (a) and currency exchange rate (b)

for, and then compared with the results of other methods, for example: artificial neural networks or wavelet analysis.

The following wavelets families: Coiflet5, BiorSplines3.9 available in the Matlab6.5 environment; were chosen in the work to the filtering process. Also, the filtering using the different types of neural networks was performed. The Multilayer Perceptron Networks, generalized regression (GRNN) and RBF (Radial Basis Function) neural networks were tested. The data set of available input data was divided into three independent subsets as required in the ANN approach: learning (50%), verification (25%) and testing (25%).

The obtained results were compared with these of the DP method (Table 2). The DP results seem to be very close to these of the wavelet analysis. Both methods are characterized by a high accuracy. They also do not generalize the results as it is observed during the ANN denoising. The obtained results of the DP denoising process of the compression test curves (load vs. stroke) are presented in the Fig. 4.

The DP method was also tested using the data obtained during the tests of the heat treatment of the magnetic materials [9]. The registered curves are characterized by the noise of different types and frequencies at different stages of a performed test. The objective of the denoising process was to maintain the characteristic peaks (see Fig. 5) for each tested material that is very important in the description of the materials' magnetic properties. Another input data set was related to the finances, i.e. the exchange rate between Euro and USD vs. time [11]. Results of denoising process with the higher smoothing coefficient represent the increasing and decreasing trends of the plot allowing the prediction of exchange rates (Fig. 5).

4. Conclusions

New holistic method for the experimental data denoising process based on the Dynamic Particles (DP) has been presented. The main advantage of this approach is the possibility of automated application (without reconfiguration) for the data influenced by the noise of different type and varying frequencies. Obtained results from the performed calculations are characterized by high accuracy and fidelity. The DP method appears very promising for other applications. The advantage of the developed DP denoising method over other commonly used techniques is its simple implementation, lack of the necessity of reconfiguration for the different data sets, fast performance, reliability and accuracy in case of various applications.

The further development of this technique should focus on the formulation of the filtering quality criterion. The next step should be expansion of presented DP method for the data of more than two dimensions. Main objectives that should be achieved in this issue are:

- the selection and implementation of the proper algorithm for searching for nearest neighbours in the space of higher dimension,
- the selection of the best type of the object space and its metrics avoiding the effect of sparse data in higher dimension.

Acknowledgments

The work is supported by AGH internal grant No. 11.11.110.575

References

[1] R. Adelino, F. da Silva. Bayesian wavelet denoising and evolutionary calibration. *Digital Signal Processing*, **14**: 566–589, 2004.

[2] A. Buades, B. Coll, J.M. Morel. On image denoising methods. Centre de Matematiques et de Leurs Applications, http://www.cmla.ens-cachan.fr, 2004.

[3] W. Dzwinel, W. Alda, D.A. Yuen. Cross-scale numerical simulations using discrete particle models. *Molecular Simulation*, **22**: 397–421, 1999.

[4] J. Falkus, J. Kusiak, P. Pietrzkiewicz, W. Pietrzyk. Filtering of the industrial data for the Artificial Neural Network Model of the Steel Oxygen Converter Process. A chapter in the monograph *Intelligence in Small World – Nanomaterials for the 21th Century*. CRC-PRESS, Boca Raton, Florida, 2003.

[5] J. Gawad, J. Kusiak, M. Pietrzyk, S. Di Rosa, G. Nicol. Optimization methods used for identification of rheological model for brass. *Proc. 6th ESAFORM Conf. on Material Forming*, Salerno, Italy, 359–362, 2003.

[6] R.M. Gray, L.D. Davisson. *An Introduction to Statistical Signal Processing*. Cambridge University Press, 2000.

[7] S. Hara, T. Tsukada, K. Sasajirna. An in-line digital filtering algorithm for surface roughness profiles. *Precision Engineering*, **22**: 190–195, 1998.

[8] M. Piovoso, P.A. Laplante. Kalman filter recipes for real-time image processing. *Real-time Image Processing*, **9**: 433–439, 2003.

[9] Ł. Rauch, J. Talar, T. Zák, J. Kusiak. Filtering of thermomagnetic data curve using artificial neural network and wavelet analysis. *Proc. 7th ICAISC 2004 Conf., Zakopane, Poland.* Springer-Verlag, 2004.

[10] D. Szeliga, P. Matuszyk, R. Kuziak, M. Pietrzyk. Identification of rheological parameters on the basis of various types of plastometric tests. *Journal of Materials Processing Technology*, **125-126**: 150–154, 2002.

[11] Web Site: www.x-trade.biz, 2004.