NEURAL NETWORKS FOR PNEUMATIC ACTUATOR FAULT DETECTION

J.F. Gomes de Freitas*, **I.M. MacLeod**** **and J.S. Maltz*****

**University of Cambridge, Department of Engineering Science, UK*

***University of the Witwatersrand, Department of Electrical Engineering, South Africa*

****Berkley University, Electrical Engineering and Computer Science Department, USA*

Abstract. The suitability of artificial neural networks (ANNs) for detecting fault conditions in pneumatic control valve actuators is investigated. Specifically, the ability of a neural network to act as a predictor of correct valve behaviour is examined. Experimental results indicate that standard network architectures are unsuitable for temporal prediction of non-linear system behaviour. An original recurrent network architecture, designed specifically as a predictor and based on autoregressive models and functional approximation is therefore proposed. The performance of this network is evaluated using both measured data and data from simulations based on a mathematical model of the valve. Laboratory implementation of the fault detection system produced encouraging qualitative results, including high success rates for the detection of faults corresponding to valve Coulomb friction changes and input pressure offsets.

Key Words. Fault detection; neural networks; system identification; control valves; pneumatic

1. INTRODUCTION

Pneumatically-actuated control valves occur frequently as a basic component of control systems in many processing and manufacturing plants. The wear and tear to which industrial control valve actuators are subjected leads to degeneration of performance and eventually to failure. In modern automated plants, an unrevealed actuator fault may have serious consequences. Although the detection of sudden failures is usually easily accomplished, this is seldom the case when deterioration is gradual. In fact, the use of feedback control to maintain desirable process operation may compensate for, and thus obscure, a developing fault.

Early detection of incipient faults based on continuous on-line monitoring of system signals can improve safety and efficiency and can help to reduce downtime and plant maintenance requirements. Pioneering efforts in this field involved the use of state and parameter estimators [1]. Such techniques required rigorous and time-consuming mathematical modelling of processes behaviour and plant components. Their reliability was linked directly to the accuracy of the modelling.

The limitations of cumbersome model-based approaches led researchers to consider alternative methods for fault detection and diagnosis. For example, *fault dictionaries* were developed and included in knowledge-based expert systems (KBES) [2]. Ironically, KBES fault-detection systems suffered from similar drawbacks to the model-based systems that they were intended to replace. Fault detection ability was limited by the quality and exhaustiveness of an information base rather than by the accuracy of a model [3].

In the mid-1980s, a resurgence of interest in artificial neural networks (ANNs) for pattern recognition led to their successful application in fault signature recognition systems [4, 2]. However, the present investigation, instead of using the pattern-recognition abilities of neural networks, uses an ANN model-based predictor to delineate faulty and normal actuator operation. This shift of emphasis is motivated by the fact that training samples for faulty valves are difficult to obtain. It is easier to develop a model for the actuator under normal operation and then to use this model to detect faults or abnormalities. The objective is fault detection as opposed to fault diagnosis.

While ANN fault signature recognition methods require additional sensors in order to measure quantities such as vibration and acoustic noise, the system discussed here utilises only normal actuator input and output signals. It is able to detect faults that relate directly to impairment in actuator functionality, including many faults that can not be distinguished using, e.g., vibration analysis.

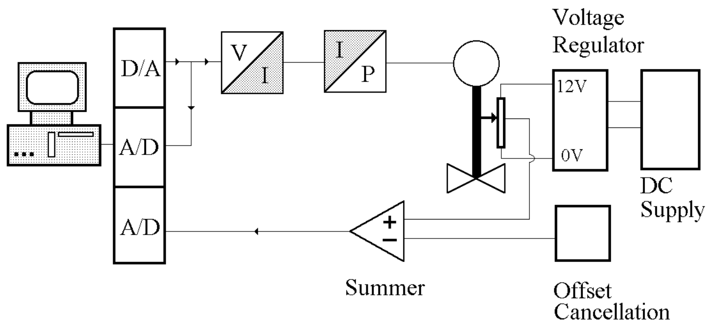

Fig. 1. Experimental system for actuator fault detection

2. EXPERIMENTAL SYSTEM

The experimental arrangement employed to monitor faults is shown in Fig. 1.

The control output signal to the valve is generated by software and is passed to a D/A converter. A voltage to current converter produces a corresponding current signal in the standard 4–20 mA range. This, in turn, is converted into a valve-actuating pressure signal in the 20–100 kPa range by a current to pressure converter. A type 513 RGS control valve (13 mm body, 19 mm linear stroke), manufactured by Fisher Controls Co. is used in the experiments. A precision linear potentiometer, which senses valve stem position, is the only transducer employed.

In order to evaluate a fault-detection system, one would ideally like to introduce physical modifications that affect actuator operation. Unfortunately, it is very difficult to modify an industrial control valve in order to introduce subtle incipient faults of a quantifiable nature. A mathematical model that describes valve dynamic behaviour was therefore used to evaluate the effectiveness of the fault-detection system in such cases. Details of this model are discussed in Appendix A. The actuator faults simulated using the mathematical model were increases in friction and input pressure offsets. Such pressure offsets could, e.g., represent calibration errors in the current to pressure converter, degradation of the actuator spring or partial blockage of the tubing supplying air to the actuator.

3. THE FAULT DETECTION SYSTEM

It is desirable for a fault-detection system to detect faults during transients because it is during sudden changes of the inputs that many kinds of fault are revealed. The proposed fault detection scheme performs transient fault detection by making use of an ANN model of the control valve. This neural net model is trained to behave like a properly-functioning control valve. After training, the model is operated in parallel with the control valve as shown in Fig. 2.

If a fault occurs, the control valve output and the output of the neural net model will differ, giving rise to an error signal that can be used to signal that a fault has occurred. During the experimental work, both the real valve and the mathematical valve model were used as the “control valve” in Fig. 2.

It could be argued that a dynamic mathematical model should be used in place of the neural net model. However, such a model has the inherent problem of being difficult to develop, especially, if the system being modelled is nonlinear. Furthermore, if there is a change in the plant being modelled, e.g., a change in the pressure or viscosity of the fluid flowing through the valve, a first-principles model might have to be re-calibrated or replaced.

Neural net models, on the other hand, can readily be arranged to learn from real data and therefore the discrepancies between the real world and the model can be reduced. It is important though to train the neural models properly. That is, the neural models should not only be able to predict well using training data, but also when presented with data not encountered in the training set. The term *generalisation* is often used to assess the ability of the neural model to predict accurately when presented with novel data [5]. Good generalisation simply means good multivariate interpolation. For the pneumatic actuator it was found to be difficult to achieve acceptable generalisation. Fortunately, theoretical insights into the trade-offs involved have recently become available—this im-

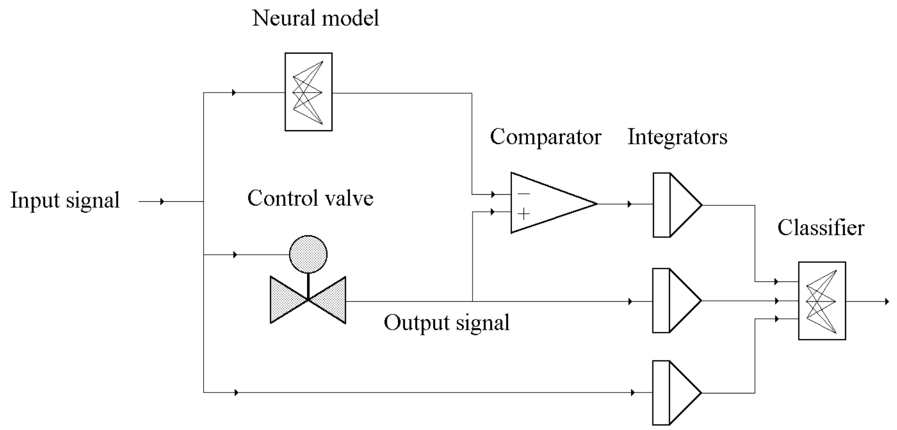

Fig. 2. Proposed fault detection system

portant topic is discussed further in section 3.1.

In Fig. 2, the squared values of the input, output and error signals are integrated over a finite number of samples so as to obtain their "energies". The reason for using these energies and not the signals themselves is that the energy of a signal is invariant with respect to the waveform of the signal. In other words, the classifier does not need to know the shape of the signals. Using the energy in the signals is in fact a form of feature extraction.

By including the error signal, an extra dimension is added to the input data to the classifier as depicted in Fig. 3. This indicates that with the extra dimension, it is possible for a neural net classifier to fit a hyper surface that divides the total space into two subspaces, namely, the correct-operation subspace and the faulty-operation subspace. A simple classifier composed of sigmoid neurons was employed [6]. As seen in Fig. 3, the clusters indicating normal and faulty operation are well separated and even a common-threshold classifier would have sufficed for the task.

3.1. *Generalisation in Neural Network Models*

Two significant bounds on the generalisation error for sigmoid and radial basis networks have been developed by Barron (in [7]) and Niyogi [8], respectively. These bounds are based on a lemma by Jones on the convergence rate of particular approximation schemes [9–11, 8, 12] and on the Vapnik-Chervonenkis Dimension [13]. Barron's and Niyogi's (with probability $> 1-\alpha$, where α is ideally a small number) generalisation bounds are given by

$$\mathbf{E}_{\text{Barron}} \leq \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}\left(\frac{nd \ln \ell}{\ell}\right) \qquad (1)$$

$$\mathbf{E}_{\text{Niyogi}} \leq \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}\left(\left[\frac{nd \ln(n\ell) - \ln \alpha}{\ell}\right]^{1/2}\right) \qquad (2)$$

where $\mathcal{O}(h)$ represents terms of order h, n is the number of parameters (weights) of the network, ℓ is the length of the training data set and d is the dimension of the input vector.

The first term in (1) and (2) is known as bias, while the second term is referred to as variance. These error terms are intrinsic to all parameterised approximating schemes [5]. The *bias* term measures the distance between the average estimator and the actual function, while the *variance* term quantifies the spread of the estimator with respect to the data distribution. To achieve good performance, both the bias and the variance should be small. Unfortunately, there is a fundamental trade-off between the bias and the variance in that, with increasing number of parameters, the variance term increases while the bias term decreases.

The only way to reduce bias and variance simultaneously is to introduce *a priori* knowledge into the modelling process as discussed in section 4. In the event of variations in the plant being modelled, such as changes in the process fluid, the neural model would simply have to be retrained.

4. DESIGN OF THE NEURAL NET PREDICTOR

The design of the neural net predictor is a system identification problem. The predictor was trained by teaching it what the output time-signal should be for a specific training input time-signal. Band-limited, uniformly-distributed noise signals were chosen as the input training signals because these

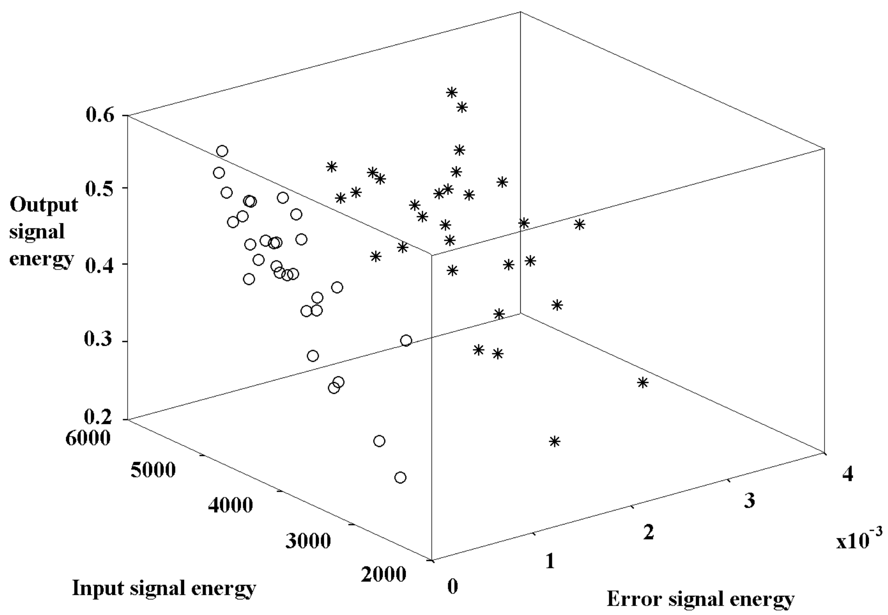

Fig. 3. Signal energies for 60 signals generated at random, [o] Normal valve [*] Friction increased from 5% to 10%

are likely to excite all system modes.

Several attempts have been made to solve the nonlinear identification problem using an extension of the theory of linear autoregressive (ARX) models combined with neural networks [14–20]. In these attempts, however, the basis or activation functions used in each network layer are identical and their relation to the physical system has not been studied.

A two-step procedure is therefore used to model the control valve [21]. The analysis step involves using a network with different basis functions (e.g., polynomials, wavelets, sinusoids, fuzzy membership functions, etc.) in each layer. This makes it possible to determine which basis functions dominate by observing the network output weights, thereby revealing information about the physical system being modelled. Furthermore, by retaining only the dominant basis functions, fewer parameters are required. Consequently, it is possible to reduce error variance (over fitting) and improve generalisation. The bias error term is also decreased because the dominant basis functions are closely linked to the structure of the nonlinear system being approximated.

The synthesis step involves generating a network containing the dominant basis functions and a set of sigmoid basis functions as shown in Fig. 4. The sigmoid basis functions generate a step-wise approximation to the part of the mapping that the dominant basis functions could not approximate. The synthesis step may be viewed as a refinement of the analysis step.

Using more than one hidden layer permits estimators of higher order to be obtained using fewer parameters. For instance, in the case of polynomial basis functions, with a single hidden layer network it is possible to achieve polynomial orders equal to the number of basis functions n. On the other hand, with a two hidden layer network it is possible to achieve polynomial orders of $n^2/2$ if n is even and $(n-1)(n+1)/4$ if n is odd. If the basis functions are differentiable, the network can easily be trained with any gradient-descent technique. Back-propagation with the Levenberg-Marquardt Optimisation [22, 19] was chosen because of its stability and fast (1 to 5 minutes on a 66 MHz 486 PC in this case) convergence rates.

5. RESULTS

After the analysis step, the valve mapping was found to have a strong linear component. In the synthesis step, one linear neuron together with 10 sigmoid functions was used to generate the neural model. This number of sigmoid functions was determined using cross-validation. That is, by equating the approximating errors for training data and validation (novel) data. This method is a practical way of balancing the bias and variance components of the generalisation error. When presented with a random validation signal, the neural model provided an accurate prediction as shown in Fig. 5.

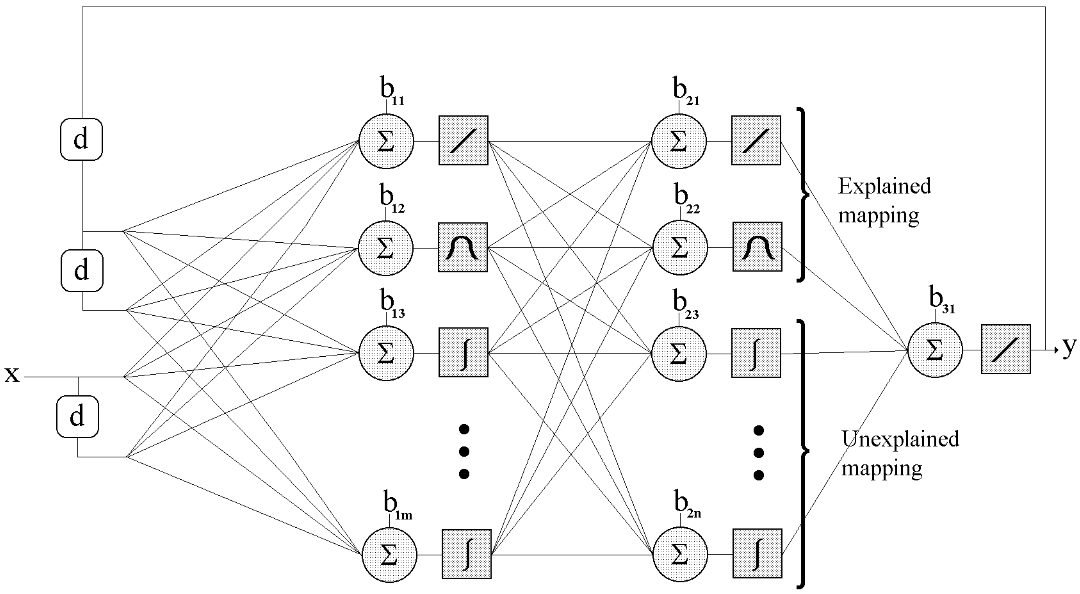

Fig. 4. Synthesis network structure

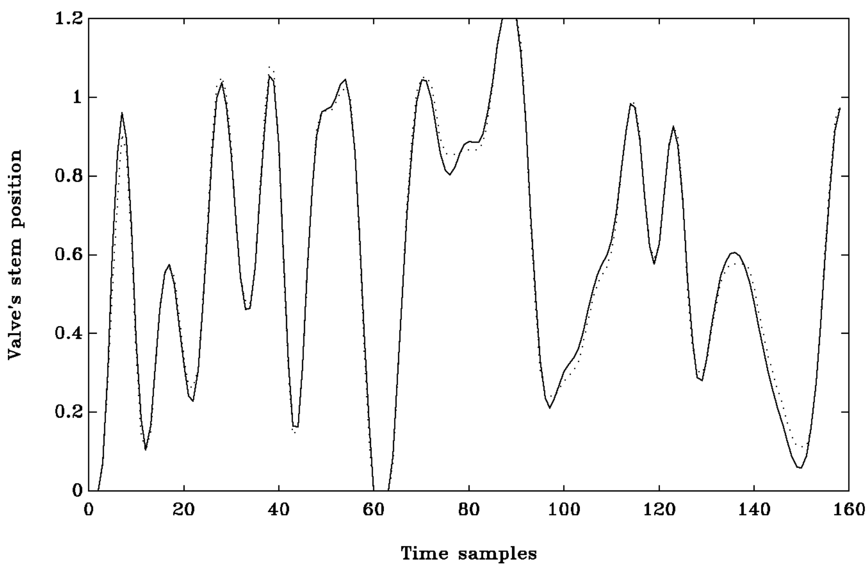

Fig. 5. Network prediction using random validation data, [· · ·] Valve output [—] Neural network output

The performance of the fault detection system was evaluated using both a real pneumatic control valve the valve mathematical model. Table 1 shows the high success rate obtained for the mathematical valve in 100 trials. Of these trials, 50

Table 1 Success rate of fault detection technique

Fault simulated	*Success Rate*
Friction increase to 10%	100%
−5 kPa input pressure offset	99%
−8 kPa input pressure offset	100%

corresponded to a normal valve and the other 50 to a faulty valve. The signals were integrated over 160 samples to obtain their energies. The parameters used in the first-principles valve simulation (Appendix A) were: stroke time $T_s = 2$ s, time constant $T_c = 1$ s and Coulomb friction fraction $F = 0.05$ (5%). These parameters were chosen using experimental measurements [6]. The fault-detection system was also implemented and tested experimentally [6].

6. CONCLUSIONS

An experimental system for detecting faults in pneumatic actuators for industrial control valves has been described. The proposed system continuously monitors readily-available on-line signals and uses a neural network to predict the expected fault-free signals. After suitable signal processing, it processes the predicted and measured signals using a classifier to yield a direct indication of correct or faulty operation. The neural network used for predicting the actuator dynamic behaviour employs a special recurrent architecture that resulted from a careful study of the trade-offs between bias, variance and model complexity.

The investigation shows that neural network concepts have an important role to play in system identification and fault detection. They provide a flexible computational framework for combining approximation theory, statistical estimation and pattern recognition techniques.

Laboratory implementation and testing of the proposed fault detection system produced promising qualitative results and further development and practical testing is recommended.

REFERENCES

[1] R Isermann, "Process fault detection based on modeling and estimation methods—A survey," *Automatica*, vol. 20, pp. 387–404, 1984.

[2] J C Hoskins, K M Kaliyur, and D M Himmelblau, "Fault diagnosis in complex chemical plants using artificial neural networks," *AIChE Journal*, vol. 37, no. 1, pp. 137–141, 1991.

[3] V Venkatasubramanian and K Chan, "A neural network methodology for process fault diagnosis," *AIChE Journal*, vol. 35, no. 12, pp. 1993–2003, 1989.

[4] T Sorsa, H N Koivo, and H Koivisto, "Neural networks in process fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 4, pp. 815–825, 1991.

[5] J F Gomes de Freitas, "Neural network based nonparametric regression for nonlinear system identification and fault detection," Master's thesis, Department of Electrical Engineering, University of the Witwatersrand, 1996.

[6] J F Gomes de Freitas, "Neural networks for actuator fault detection," tech. rep., University of the Witwatersrand, South Africa, 1994.

[7] B Cheng and D M Titterington, "Neural networks: A review from a statistical perspective," *Statistical Science*, vol. 9, no. 1, pp. 2–54, 1994.

[8] P Niyogi and F Girosi, "On the relationship between generalization error, hypothesis complexity, and sample complexity for radial basis functions," Tech. Rep. AIM-1467, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, February 1994.

[9] L K Jones, "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training," *The Annals of Statistics*, vol. 20, no. 1, pp. 608–613, 1992.

[10] L Breiman, "Hinging hyperplanes for regression, classification, and function approximation," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 999–1013, 1993.

[11] A R Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 930–945, 1993.

[12] F Girosi and G Anzellotti, "Convergence rates of approximation by translates," Tech. Rep. AIM-1288, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, April 1995.

[13] V N Vapnik, *Estimation of Dependences Based on Empirical Data*. Berlin: Springer-Verlag, 1982.

[14] B R Pramod and S C Bose, "System identification using ARMA modeling and neural networks," *Journal of Engineering for Industry*, vol. 115, pp. 487–491, 1993.

[15] J Sjöberg, H Hjalmarsson, and L Ljung,

"Neural networks in system identification," in *10th IFAC Symposium on System Identification*, Copenhagen, pp. 187–190, 1994.
[16] O Nerrand, L Personnaz, and G Dreyfus, "Non-linear recursive identification and control by neural networks: A general framework," in *Proceedings of the 2nd European Control Conference*, Groningen, Netherlands, pp. 93–98, 1993.
[17] H J M Telkamp and A A H Damen, "Neural network learning in nonlinear system identification and controller design," in *Proceedings of the 2nd European Control Conference*, Groningen, Netherlands, pp. 798–804, 1993.
[18] K Narendra and K Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, pp. 4–27, 1990.
[19] J Sjöberg *et al.*, "Non-linear black-box modelling in system identification: A unified overview," tech. rep., Linköping University, Sweden, 1995.
[20] A Juditsky *et al.*, "Non-linear black-box models in system identification: Mathematical foundations," tech. rep., Linköping University, Sweden, 1995.
[21] J F Gomes de Freitas, A L Stevens, and J N Ridley, "Generating and using prior information in neural networks for system identification." Submitted to IEEE Transactions on Neural Networks, 1998.
[22] A Cichocki and R Unbehauen, *Neural Networks for Optimization and Signal Processing*. New York: Wiley, 1993, ch. 1, pp. 23–29.
[23] R H Perry and D Green, Eds., *Chemical Engineers' Handbook*, 6th ed. New York: McGraw-Hill, 1984, ch. 22, pp. 84–88.

APPENDICES

A. MATHEMATICAL MODEL OF A PNEUMATIC ACTUATOR

In a typical pneumatic actuator, the input-air pressure provides a sufficient force on a diaphragm or piston to oppose a spring force, thereby producing a displacement of the valve stem [23]. Experiments in the Control Engineering Laboratory at the University of the Witwatersrand have shown that the dynamic motion of the valve stem can be considered to fall into three regions: (1) velocity-limited motion for large changes, (2) first-order exponential motion for small changes and (3) no motion for very small changes (because of Coulomb friction). If the valve stem position is y and the difference between the position demanded by the input pressure and the valve stem position is u, the corresponding differential equations are

$$\frac{dy}{dt} = \begin{cases} \pm 1/T_s & \text{if } |u| - F > T_c/T_s \\ \pm(|u| - F)/T_c & \text{if } 0 < |u| - F \\ & \qquad \leq T_c/T_s \\ 0 & \text{if } |u| - F \leq 0 \end{cases}$$

where T_s is the stroke time (time taken for the valve to move from fully open to fully shut under velocity-limited motion), T_c is the time constant and F is the fraction of total valve stem movement below which input pressure changes do not cause movement. These parameters can be measured experimentally for a given valve.