

Identification and Predictive Control by p -norm Minimization

J. Pekař, J. Štecha

Real time system parameter estimation from the set of input-output data is usually solved by minimization of quadratic norm errors of system equations – known in the literature as least squares (LS) or its modification as total least squares (TLS) or mixed LS and TLS. It is known that the utilization of the p -norm ($1 \leq p < 2$) instead of the quadratic norm suppresses the wrong measurements (outliers) in the data. This property is shown for different norms, and it is shown that the influence of outliers is suppressed if $p \rightarrow 1$. Also optimal predictive control utilizing p -norm minimization of the criterion is developed, and the simulation results show the properties of such control.

Keywords: identification, predictive control, p -norm, ARX model, iteratively reweighted least squares, linear programming.

1 Introduction

The dynamic properties of a real plant are usually identified by making a model – choosing a model structure and estimating the unknown parameters of the model using data measured on the real plant [1]. The first goal of this paper is to compare a set of parameter estimations of an ARX model where each estimation is obtained by minimizing the p -norm ($1 \leq p \leq 2$). The measurement of the system output is considered to be damaged by a number of outliers.

Another problem is optimal control of dynamic systems. Model predictive control (MPC) strategies are very popular [2, 4]. Optimal predictive control of an ARX or state space model is usually obtained by minimizing the quadratic criterion. If a non-quadratic norm is used in the optimality criterion, different results are obtained. For example, for $p \rightarrow 1$ dead beat control is obtained. Minimizing the l_1 norm using linear programming in MPC control has been considered by many authors (e.g. [5, 6, 7]). A connection between linear programming and optimal control is shown for example in [8, 9]. In this paper, optimal predictive control utilizing p -norm minimization of the criterion is shown, and the results are illustrated by simple examples.

The paper is organized as follows: The second section shows identification of an ARX model using the p -norm. The algorithm for minimizing the p -norm where $1 < p < 2$ is recapitulated in section 3. This algorithm is known as Iteratively Reweighted Least Squares (IRLS). In section 4, the predictive control strategy for state space models is reviewed. In sections 5 and 6, examples of system identification and control in p -norm are shown.

2 Identification of an ARX model using p -norm

The ARX model of a dynamic system can be described by the equation

$$y(t) = a_1 y(t-1) + \dots + a_{n-1} y(t-n-1) + b_0 u(t) + \dots + b_{n-1} u(t-n-1) + e(t) \quad (1)$$

where $y(t)$ is the system output in time t and $u(t)$ is the system input, $e(t)$ is an equation error and a_i, b_j are system parameters. Note that only a single input – single output (SISO) system is considered in this paper. The parameter estimation problem from a given set of data

$$\mathcal{D}^t = \{y(t), y(t-1), \dots, u(t), u(t-1), \dots\}$$

leads to minimization of error vector $\varepsilon = [e(t), e(t-1), \dots]^T$. We are looking for the parameter vector

$$\mathbf{x} = [a_1, a_2, \dots, b_0, b_1, \dots]^T$$

that ensures the best approximation of the output vector $\mathbf{b} = [y(t), y(t-1), \dots, y(t-m-1)]^T$ by a vector $A\mathbf{x}$, where

$$A = \begin{bmatrix} y(t-1) & y(t-2) & \dots & u(t) & \dots \\ y(t-2) & y(t-3) & \dots & u(t-1) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \quad (2)$$

Both vector \mathbf{b} and matrix A are formed from the measured data \mathcal{D}^t . The most usual solution of the parameter estimation problem involves minimization of the quadratic norm of the error vector, known as the Euclidean norm. In such a case, the optimization problem is

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2. \quad (3)$$

The solution is known as Least Squares (LS) [10]. In reality the noise of the output measurement is also found in elements of data matrix A . The solution of such a problem leads to Total Least Squares (TLS) [11]. If the input measurement is noise free, the problem can be solved by Mixed LS and TLS [11]. In some applications, we can use the more general p -norm instead of the quadratic norm [10, 12]. The p -norm is defined as

$$\|y\|_p = \left(|y_1|^p + \dots + |y_n|^p \right)^{\frac{1}{p}}, \quad 1 \leq p < \infty \quad (4)$$

$$\|y\|_\infty = \max_i (|y_1|, |y_2|, \dots).$$

Then the problem of parameter estimation is defined as follows

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_p^p. \quad (5)$$

For $1 < p \leq 2$ p -norm minimization can be done by iterative solution of the LS problem. For $p = 1$ and $p = \infty$ the solution can be obtained by linear programming.

3 p -norm minimization

The general p -norm can be used in the optimization problem defined by (5). If the p -norm is restricted by $1 < p < 2$ the

minimization problem is convex and the solution is unique. This problem can be solved by an iterative algorithm known as Iteratively Reweighted Least Squares (IRLS) [10]. Note that the optimization problem with 1-norm or ∞ -norm can be solved by linear programming.

3.1 Iteratively reweighted least squares

Let us solve the approximation problem

$$\min_x \left\{ \Psi(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_p^p \right\} \quad 1 < p < 2. \quad (6)$$

We assume, that all coordinates of the residuum $\varepsilon(\mathbf{x}) = \mathbf{b} - \mathbf{Ax}$ are nonzero. Then the function $\Psi(\mathbf{x})$ can be defined as

$$\Psi(\mathbf{x}) = \sum_{i=1}^m |\varepsilon_i(\mathbf{x})|^p = \sum_{i=1}^m |\varepsilon_i(\mathbf{x})|^{p-2} \varepsilon_i(\mathbf{x})^2. \quad (7)$$

The previous problem is weighted Least Squares:

$$\min_x \left\| \mathbf{D}(\varepsilon) \frac{p-2}{2} (\mathbf{b} - \mathbf{Ax}) \right\|_2, \quad \mathbf{D}(\varepsilon) = \text{diag}(|\varepsilon|). \quad (8)$$

Because of the dependency of the diagonal weighting matrix $\mathbf{D}(\varepsilon)$ on the unknown solution \mathbf{x} , the problem must be solved by an iterative algorithm:

1. $\mathbf{x}^{(0)}$ is an initial solution, set the iteration counter $k=0$.
2. calculate $\varepsilon^{(k)}$ by

$$\varepsilon^{(k)} = \mathbf{b} - \mathbf{Ax}^{(k)}, \mathbf{D}^{(k)} = \text{diag} \left(\left| \varepsilon^{(k)} \right|^{\frac{p-2}{2}} \right). \quad (9)$$

3. utilizing the weighted LS algorithm, $\delta \mathbf{x}^{(k)}$ is obtained by solving

$$\delta \mathbf{x}^{(k)} = \arg \min_{\delta \mathbf{x}} \left\| \mathbf{D}^{(k)} (\varepsilon^{(k)} - \mathbf{A} \delta \mathbf{x}) \right\|_2. \quad (10)$$

4. the next iteration $\mathbf{x}^{(k+1)}$ is obtained as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta \mathbf{x}^{(k)}. \quad (11)$$

5. if the convergence criterion is satisfied, then stop and $\mathbf{x}^{(k+1)}$ is the solution, else set $k \leftarrow k+1$ and go to 2.

3.2 Minimization of 1-norm by linear programming

In order to minimize the norm $p=1$ linear programming (LP) can be used. The following problems are equivalent

$$\min_x \|\mathbf{Ax} - \mathbf{b}\|_1 \Leftrightarrow \min_y \{ \mathbf{1}^T \mathbf{y} : \mathbf{Ax} - \mathbf{b} \leq \mathbf{y}, \mathbf{Ax} - \mathbf{b} \geq -\mathbf{y} \}. \quad (12)$$

Introducing the augmented vector $\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$, the standard

form of the LP problem is obtained

$$\min_z \{ \mathbf{c}^T \mathbf{z} : \bar{\mathbf{A}} \mathbf{z} \leq \bar{\mathbf{b}} \}, \quad (13)$$

$$\mathbf{c}^T = [0, \dots, 0, 1, \dots, 1],$$

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & -\mathbf{I} \\ -\mathbf{A} & -\mathbf{I} \end{bmatrix}, \quad \bar{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix}.$$

The only drawback of such computation is that the LP problem can have more than one solution.

3.3 Minimization of ∞ -norm by linear programming

In order to minimize the norm $p=\infty$, linear programming (LP) can also be used. Similarly as for 1-norm, the following two problems are equivalent

$$\min_x \|\mathbf{Ax} - \mathbf{b}\|_\infty \Leftrightarrow \min_y \{ \mathbf{y} : \mathbf{Ax} - \mathbf{b} \leq \mathbf{y} \mathbf{1}, \mathbf{Ax} - \mathbf{b} \geq -\mathbf{y} \mathbf{1} \}, \quad (14)$$

where $\mathbf{1} = [1, 1, \dots, 1]^T$ is the unit vector. Introducing the augmented vector $\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$, the standard form of the LP problem is

obtained:

$$\min_z \{ \mathbf{c}^T \mathbf{z} : \bar{\mathbf{A}} \mathbf{z} \leq \bar{\mathbf{b}} \}, \quad (15)$$

where

$$\mathbf{c} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \bar{\mathbf{A}} = \begin{bmatrix} & & & -1 \\ & \mathbf{A} & & \vdots \\ & & & -1 \\ & & & -1 \\ & -\mathbf{A} & & \vdots \\ & & & -1 \end{bmatrix}, \quad \bar{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix}.$$

4 Predictive control strategy

Suppose the state and output development model is given in the form

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{Ax}(t) + \mathbf{Bu}(t) + \mathbf{v}(t), \\ \mathbf{y}(t) &= \mathbf{Cx}(t) + \mathbf{Du}(t) + \mathbf{e}(t). \end{aligned} \quad (16)$$

where $\mathbf{x}(t)$, $\mathbf{y}(t)$ and $\mathbf{u}(t)$ are the state, output and input of the system and \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{D} are matrices of appropriate dimensions, $\mathbf{v}(t)$ and $\mathbf{e}(t)$ are state and measurement noises respectively with zero means and covariances \mathbf{R}_v and \mathbf{R}_e , independent of the state and input of the process. For the optimal predictive strategy the quality criterion is usually given in the quadratic form

$$J = \mathcal{E} \left\{ \sum_{t=1}^K (y(t) - w(t))^2 + r(u(t))^2 \mid \mathbf{x}(1) \right\}, \quad (17)$$

where K is the horizon of the predictive control strategy, $w(t)$ is the reference and r is the weighting coefficient. Let us introduce the augmented vectors

$$\begin{aligned} \mathbf{Y} &= [y(1), \dots, y(K)]^T, \quad \mathbf{W} = [w(1), \dots, w(K)]^T, \\ \mathbf{U} &= [u(1), \dots, u(K)]^T, \quad \mathbf{V} = [v(1), \dots, v(K)]^T, \\ \mathbf{E} &= [e(1), \dots, e(K)]^T \text{ and augmented matrices,} \\ \mathbf{P} &= [\mathbf{C}^T, \mathbf{A}^T \mathbf{C}^T, \dots, (\mathbf{A}^T)^{K-1} \mathbf{C}^T]^T, \end{aligned}$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{K-2} & \mathbf{CA}^{K-3} & \dots & \mathbf{D} \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{C} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{K-2} & \mathbf{CA}^{K-3} & \dots & \mathbf{0} \end{bmatrix}.$$

Then the criterion with the quadratic cost function can be written in the form

$$J = \mathcal{E} \left\{ (\mathbf{Y} - \mathbf{W})^T (\mathbf{Y} - \mathbf{W}) + r \mathbf{U}^T \mathbf{U} \mid \mathbf{x}(1) \right\}, \quad (18)$$

where the augmented vector $\mathbf{Y} = \mathbf{Px}(1) + \mathbf{SU} + \mathbf{QV} + \mathbf{E}$. The criterion can be minimized by completing the squares if the

constraints are not considered. After simple manipulation the optimal control results

$$U^* = -(S^T S + rI)^{-1} S^T (P\mathbf{x}(1) - W) \quad (19)$$

The state and output noises only enlarge the optimal value of the criterion. Often only differences of input signal are considered. The criterion then has the form

$$J = \mathcal{E} \left\{ \sum_{t=1}^K (y(t) - w(t))^2 + r(\Delta u(t))^2 | \mathbf{x}(1) \right\}, \quad (20)$$

where $\Delta u(t) = u(t) - u(t-1)$. In such a case, the optimal predictive control strategy is

$$U^* = -(S^T S + rR^T R)^{-1} S^T [(P\mathbf{x}(1) - W) - rR^T U(0)], \quad (21)$$

where $U(0) = [u(0), 0, \dots, 0]^T$ and matrix R equals

$$R = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & \dots & 0 & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}.$$

This way achieves integral action of control. Also for predictive control a different norm minimization of the quality criterion can be used

$$J = \sum_{k=p-1}^k |y(i) - w(i)|_p^p + r|\Delta u(i)|_p^p \quad (22)$$

According to weighting coefficient r , the quadratic norm suppresses large control errors and large input signals. If the norm $p < 2$ is used, small members in the criterion have greater influence, and if $p \rightarrow 1$ the control law approaches to dead beat control.

5 Example I – suppression of outliers by p -norm minimization

This section shows an example of system parameter identification. The experiment clearly shows the influence of outliers on the results of identification for utilization of different norms. Consider a discrete-time system described by the transfer function

$$Y(z) = \frac{0.002870z^2 + 0.009882z + 0.002126}{z^3 - 2.444z^2 + 2.008z - 0.5488} U(z) + \frac{1}{z^3 - 2.444z^2 + 2.008z - 0.5488} E(z). \quad (23)$$

Using this model, the data for our experiments was generated. Fig. 1 shows the system input and output time trajectories. 1000 samples of input-output data are corrupted by 11 outliers (completely wrong measurements).

Fig. 2 and Fig. 3 show the step responses of the identified models together with the step responses of the nominal model, where the each identification was done by minimizing a different norm.

It is not convenient to use the $p = \infty$ norm for system parameter identification from data corrupted by outliers – see Fig. 4.

The Euclidean norm of the parameter estimation error can be used to demonstrate the accuracy of the estimation. The norm is

$$J = \|\mathbf{x} - \mathbf{x}^*\|_2, \quad (24)$$

where \mathbf{x} is the estimation and \mathbf{x}^* is the vector of the true parameters of the nominal system. Fig. 5 shows the dependency of the norm of the estimation error on the p -norm of the criterion (i.e., on p).

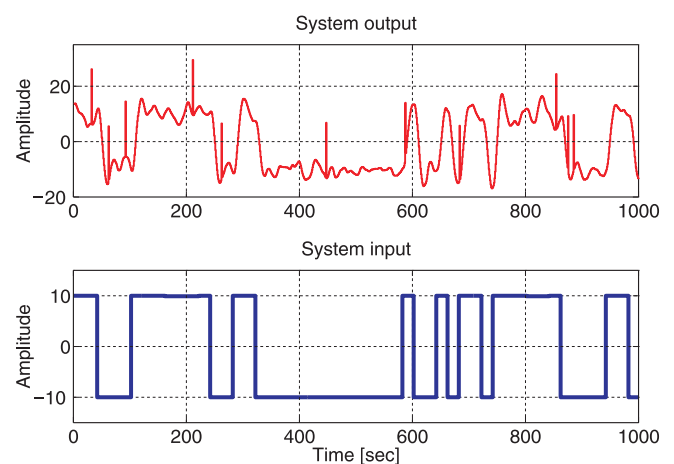


Fig. 1: Data for experiments

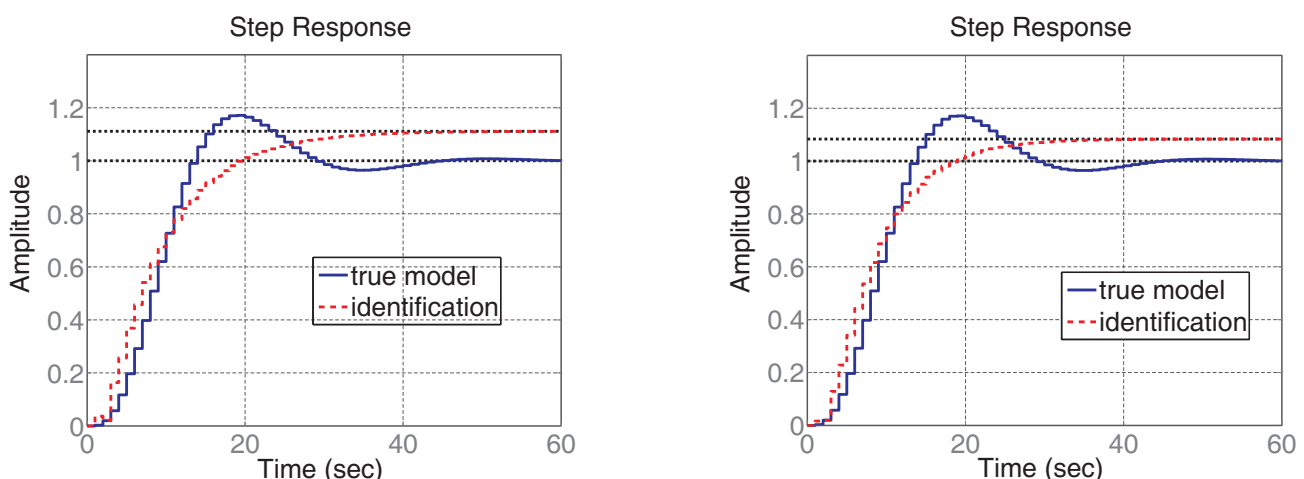


Fig. 2: Results of identification – step responses (left $p = 2$, right $p = 1.7$)

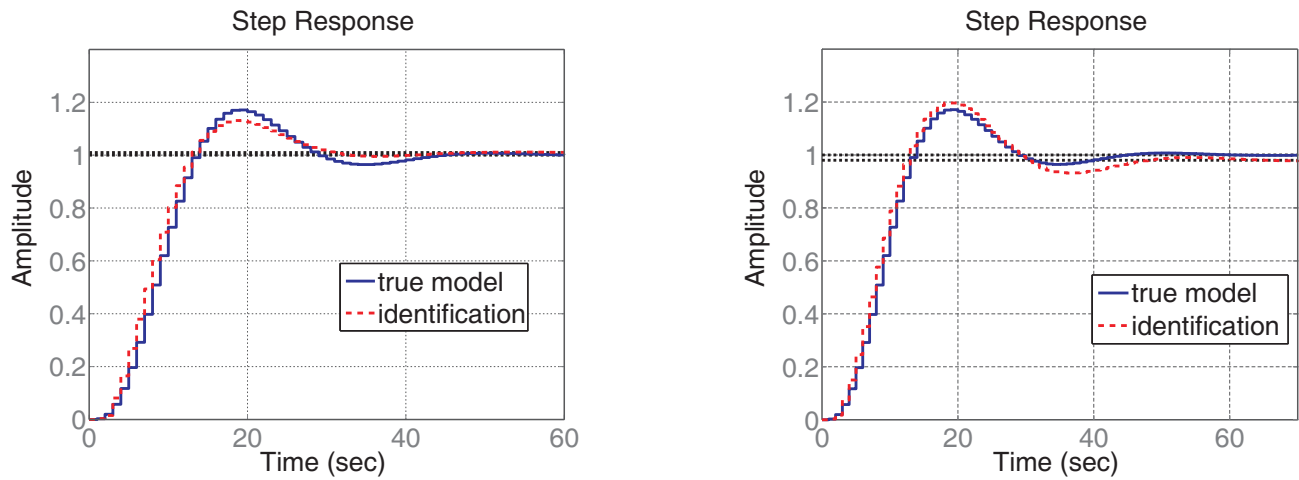


Fig. 3: Results of identification – step responses (left $p = 1.3$, right $p = 1$)

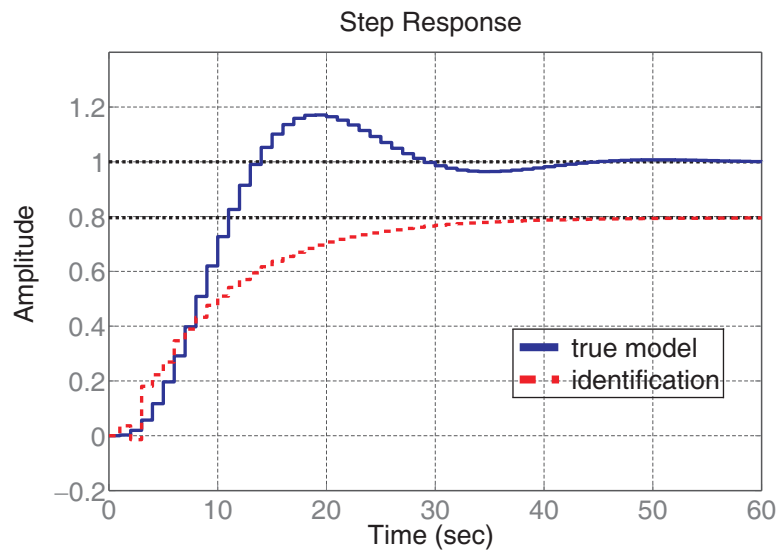


Fig. 4: Results of identification - step response for $p = \infty$

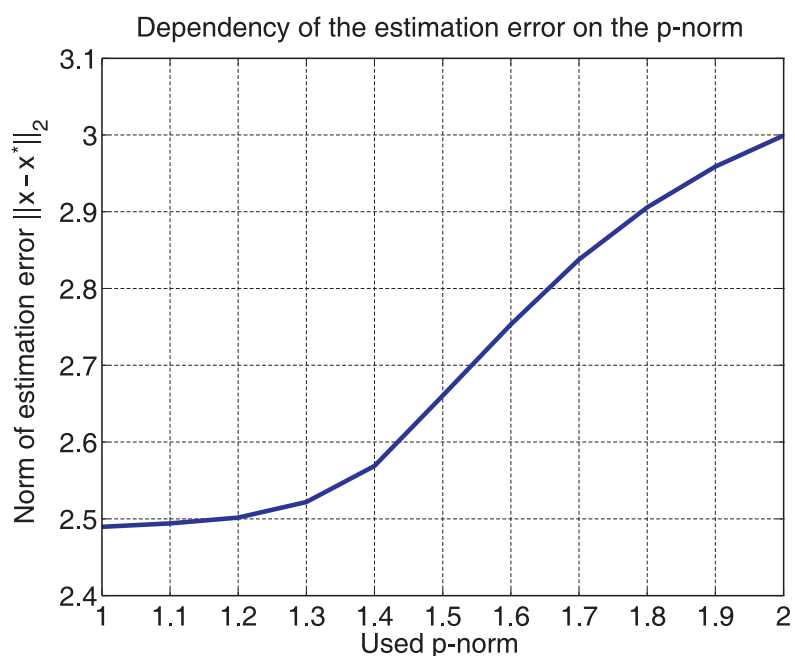


Fig. 5: Dependency of the norm of the estimation error on the p -norm (i.e., on p)

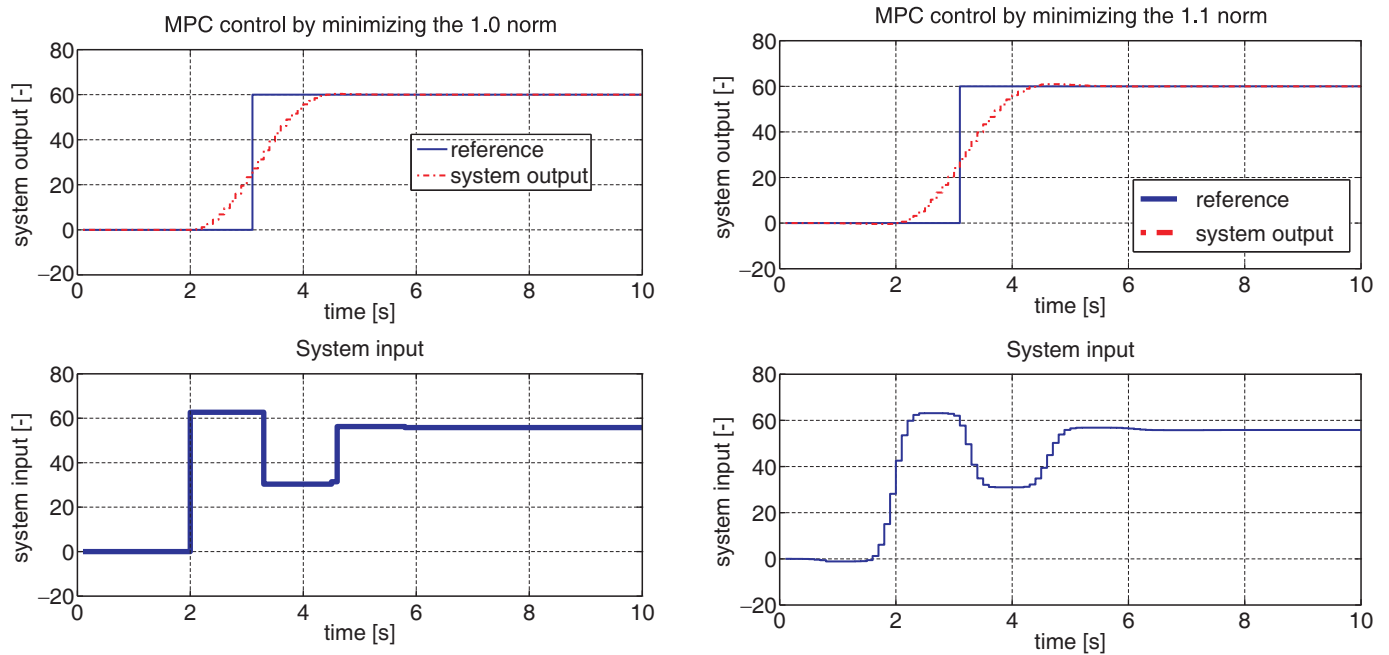


Fig. 6: Simulation results for 1 (left) and 1.1 (right) norm (with weighting coefficient $r = 1$)

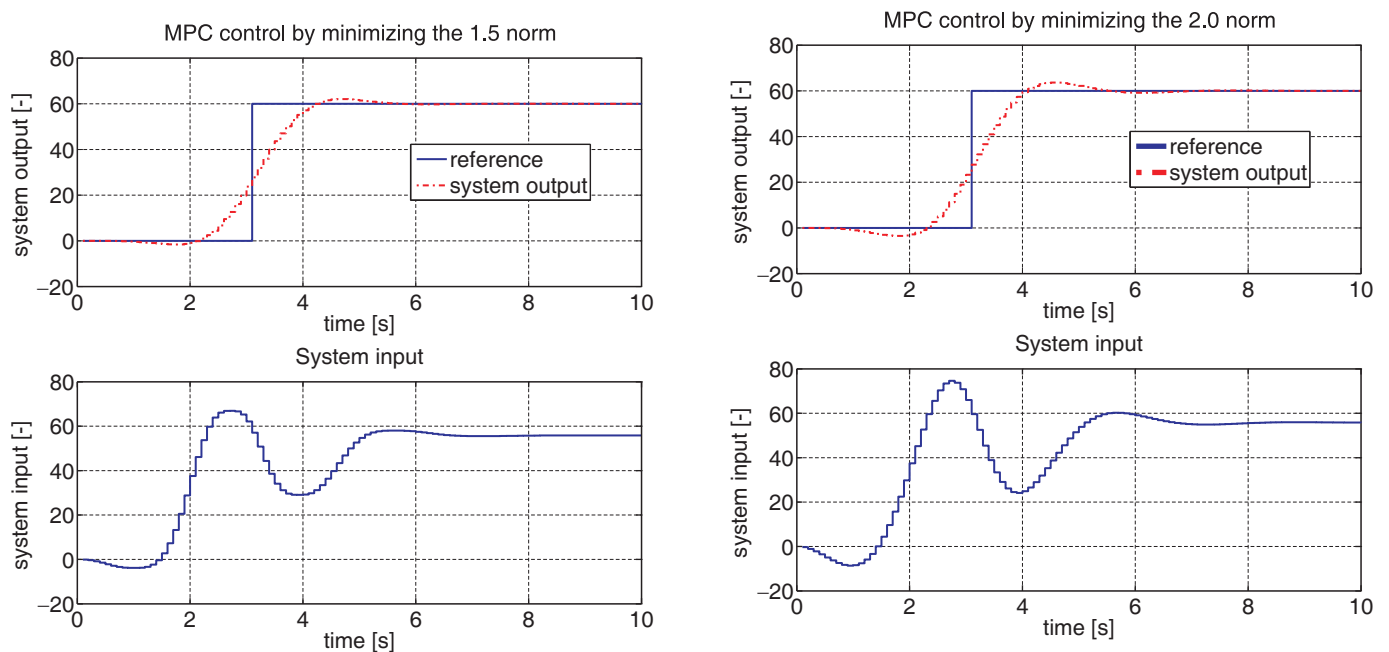


Fig. 7: Simulation results for 1.5 (left) and 2 (right) norm (with weighting coefficient $r = 1$)

6 Example II – predictive control by p -norm minimization

Consider the second order continuous-time system described by a transfer function

$$G(s) = \frac{1}{s^2 + 0.7s + 0.93} \quad (25)$$

that is sampled by a sampling period $T_s = 0.1s$. Note that the discrete-time system model was transformed to the state

space. In the first experiment, three different MPC controllers were designed:

- The first MPC minimizes the 1-norm. Such a minimization problem can be converted to linear programming (see section 3).
- The second MPC minimizes the p -norm where p is considered to be $1 < p < 2$. In this case, the minimization problem can be solved by the IRLS algorithm (see section 3).
- The third case is standard MPC minimizing the quadratic norm of the error vector. This optimization can be solved

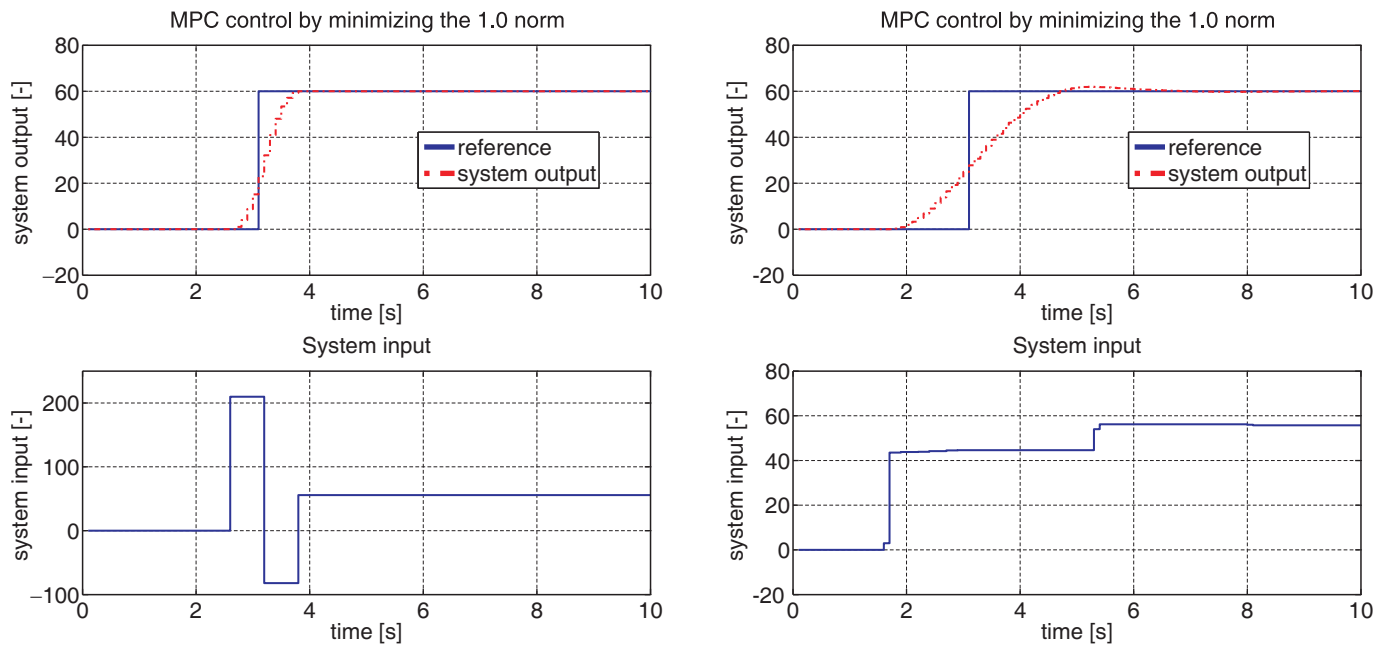


Fig. 8: The simulation results for 1 norm for weighting coefficient $r=0.1$ (left) and $r=10$ (right)

as a least squares problem or, with constraints, as a quadratic programme.

The only parameter that is different for each MPC is the kind of p -norm. All other parameters are not changed (prediction horizon $N=4s$ and weighting coefficient $r=1$). Fig. 6 and Fig. 7 show the simulation results for the first experiment. The performances of reference tracking and system input are shown for $p=1, 1.1, 1.5, 2$. Minimizing the 1-norm of the quality criterion achieves dead beat control.

The second experiment shows the influence of the weighting coefficient r in the criterion (22) on the time length of dead beat control (for 1 norm). The simulation results for weighting coefficient $r=1$ are shown in Fig. 6 (left), for $r=0.1$ and $r=10$ in the Fig. 8. The relatively large penalty r does not allow big changes in differences of control variable Δu and therefore the time of dead beat control is longer.

7 Conclusions

The purpose of this paper is to show the influence of different p -norm selection on ARX model identification and control. When the measurements of the system output were damaged by outliers, it is shown that the Euclidean norm gives worse results than the p -norm for $1 < p < 2$. The best results were achieved with the p -norm for $p=1$.

Predictive control realized by the p -norm of quality criterion minimization shows interesting results from LQ control (Linear system and Quadratic criterion) to dead beat control.

8 Acknowledgment

This work was partly supported by grants 102/05/0903, 405/05/2173 and 102/05/2075 of the Grant Agency of the Czech Republic.

References

A shorter version of this paper was presented in [3].

- [1] Ljung, L.: *System Identification: Theory for the User*. Prentice Hall, Englewood Cliffs, NJ, 1987.
- [2] Åström, K. J., Wittenmark, B.: *Computer Controlled Systems: Theory and design*. Prentice Hall, Inc., Upper Saddle River, NJ, 1997.
- [3] Pekař, J., Štecha, J., Havlena, V.: "Identification and Predictive Control of ARX Model by p -norm Minimization." Proceedings of the 23rd IASTED International Conference on Modelling, Identification, and Control. Anaheim 2004. ISBN 0-88986-387-3.
- [4] Findeisen, R., Imsland, L., Allgöwer, F., Foss, B. A.: "State and Output Feedback Nonlinear Model Predictive Control: An Overview." *European Journal of Control*, Vol. **9** (2003), p. 190–206.
- [5] Rao, C. V., Rawlings, J. B.: "Linear Programming and Model Predictive Control." *Journal of Process Control*, Vol. **10** (2000), p. 283–289.
- [6] Allwright, J. C., Papavasiliou, G. C.: "On Linear Programming and Robust Model Predictive Control Using Impulse Response." *Sys. Cont. Let.*, Vol. **18** (1992), p. 159–164.
- [7] Genceli, H., Nikolau, M.: "Robust Stability Analysis of Constrained -Norm Model Predictive Control." *AIChE J.*, Vol. **39** (1993), No. 12, p. 1954–1965.
- [8] Change, T. S., Seborg, D. E.: A Linear Programming Approach for Multivariable Feedback Control with Inequality Constraints. *Int. J. Control*, Vol. **37** (1983), p. 583–597.
- [9] Zadeh, L. A., Whalen, J. H.: "On Optimal Control and Linear Programming." *IRE Trans. Auto, Cont.* Vol. **7** (1962), p. 45–46.

- [10] Björck, .: *Numerical Methods for Least Squares Problems*. Siam, Philadelphia, 1996.
- [11] Van Huffel, S., Vandewalle, J.: *The Total Least Squares Problem. Computational Aspects and Analysis*. Siam, Philadelphia, 1991.
- [12] Boyd, S., Vandenberghe, L.: "Introduction to Convex Optimization with Engineering Applications." Lecture notes, Stanford University, 2002.
Available at: isl.stanford.edu/pub/boyd/.

Ing. Jaroslav Pekař, Ph.D.
phone: +420 266 052 325
fax: +420 286 890 555
e-mail: jaroslav.pekar@honeywell.com

Honeywell Prague Laboratory
Pod vodárenskou věží 4
182 08 Prague 8, Czech Republic

Prof. Ing. Jan Štecha, CSc.
phone: +420 224 357 238
fax: +420 224 918 646
e-mail: stech@control.felk.cvut.cz

Department of Control Engineering

Czech Technical University in Prague
Karlovo nám. 13
12 135 Prague 2, Czech Republic