On Error Measures for Receding Horizon Control

Farzad Noorian, Philip H. W. Leong

***Abstract*—Time series predictions are utilised to improve regulation in receding horizon control (RHC) problems. For the case of a quadratic objective function in a discrete linear time invariant system we: derive a closed form equation for the increase in cost due to prediction error, ΔJ; demonstrate that dimensionality reduction can be employed to reduce computational costs in testing different prediction techniques; and suggest that ΔJ is a better choice for predictor evaluation than standard measures such as mean square error or mean absolute error.**

***Index Terms*—Receding horizon control, linear quadratic regulator, machine learning, time series, forecasting.**

I. Introduction

A Linear Quadratic (LQ) system's dynamics are described by a set of linear differential equations and a quadratic objective function. Because of their generalisability, LQ systems with shrinking or receding horizon control (RHC) regulators have been widely studied theoretically [1] and are used in many applications including electric power systems [2], inventory management [3] and finance [4].

One relatively recent addition to discrete LQ control is the use of statistical models and machine learning (ML) algorithms, where time varying but measurable disturbance inputs are estimated via a forecasting algorithm [5]–[7]. In time series prediction, error measures such as mean square error (MSE) or mean absolute error (MAE) are used to compare the quality of different prediction techniques [8]. Unfortunately, these assume that the error of each sample is independent of other time-steps, which is not true in general. To obtain full accuracy, one must *validate* a model by full *backtesting*. Alternative approaches exist, such as including the conditional distribution of the measurable disturbances in the dynamic programming problem [9]. None of these approaches are computationally efficient.

The contributions of this note are threefold. Firstly, we derive an exact closed form solution for the increase in performance cost of an LQ RHC system, ΔJ, for any prediction compared to a prescient forecaster. Secondly, we demonstrate that by using dimensionality reduction, the computation of the cost function can be reduced with minimal loss of accuracy, this leading to a significant speed-up in the case where different prediction algorithms and parameters need to be tested for fixed system parameters. Thirdly, we show that using ΔJ to select prediction models results in better predictors than common time series error measures.

F. Noorian and P.H.W. Leong are with the School of Electrical and Information Engineering, University of Sydney, NSW, 2006 Australia. (e-mail: {farzad.noorian, philip.leong}@sydney.edu.au.)

Corresponding author: farzad.noorian@sydney.edu.au.

This research was supported under Australian Research Council's Linkage Projects funding scheme (project number LP110200413) and Westpac Banking Corporation.

II. Background

A discrete linear time invariant (LTI) system can be described using the following state space model [1]:

$$x(t+1) = Ax(t) + Bu(t) + Cv(t) \tag{1}$$

where $x(t)$ is the state at time $t \in [0, 1, \ldots, N]$, $u(t)$ is the controllable input, $v(t)$ is the external disturbance, and A, B and C are the system and input matrices respectively and are controllable.

The cost function is defined by the quadratic equation:

$$J = \sum_{t=1}^{N} x(t)^{\mathrm{T}} Q_t x(t) + \sum_{t=0}^{N-1} u(t)^{\mathrm{T}} P_t u(t), \tag{2}$$

where Q_t and P_t are stage costs of $x(t)$ and $u(t)$ respectively.

These equations can be rewritten in matrix form as

$$\begin{aligned} X =& S_A x(0) + S_B U + S_C V, \\ J =& X^{\mathrm{T}} \bar{Q} X + U^{\mathrm{T}} \bar{P} U, \end{aligned}$$

where $X = [x(1) \ \cdots \ x(N)]$, $U = [u(0) \ \cdots \ u(N-1)]$, $V = [v(0) \ \cdots \ v(N-1)]$, $S_A = [A \ A^2 \ \cdots \ A^N]$,

$$S_B = \begin{bmatrix} B & 0 & \cdots & 0 \\ AB & B & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A^{N-1}B & A^{N-2}B & \cdots & B \end{bmatrix},$$

$$S_C = \begin{bmatrix} C & 0 & \cdots & 0 \\ AC & C & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A^{N-1}C & A^{N-2}C & \cdots & C \end{bmatrix},$$

$\bar{Q} = \text{blockdiag}\{Q_1, Q_2, \cdots, Q_N\}$ and $\bar{P} = \text{blockdiag}\{P_0, P_1, \cdots, P_{N-1}\}$.

Let $Y = S_C^{-1}(S_A x(0) + S_C V)$, $J_A = \bar{P} + S_B^{\mathrm{T}} \bar{Q} S_B$, $J_B = S_B^{\mathrm{T}} \bar{Q} S_C$, and $H = -J_A^{-1} J_B$. The optimal control input U^* is given by $U^* = HY$ [1].

III. Problem Formulation

Even assuming system observability, $v(t)$ is not available before time t and must be substituted by $\hat{v}(t) = v(t) + \epsilon(t)$, where $\hat{v}(t)$ is the predicted value and $\epsilon(t)$ is the additive prediction error (i.e., the unmeasurable disturbance).

In a receding (or shrinking) horizon approach, at each time-step τ, $\hat{v}(t)$ is re-estimated for $t \geq \tau$. We denote these re-estimations and their prediction error by $\hat{v}_\tau(t)$ and $\epsilon_\tau(t)$ respectively. Similarly, their matrix counterparts are defined as $\hat{V}_\tau$ and $E_\tau = [\epsilon_\tau(t) \ \cdots \ \epsilon_\tau(N-1)]$.

As derived in Appendix A, U can be computed as an affine function of disturbances:

$$U = \sum_{i=0}^{N-1} \Phi_i (M_Y(i) Y + E_i) \tag{3}$$

where

$$\Phi_i = \sum_{j=0}^{N-1} M_\Phi(j) H_j \Gamma(j, i),$$

$$\Gamma(t,\tau) = \begin{cases} t > \tau & 0_{h_t \times h_\tau} \\ t = \tau & I_{h_t \times h_t} \\ t < \tau & \sum_{i=0}^{t-1} C^{-1} A^{t-i} B M_U(t,h_i) H_i \Gamma(i,\tau) \end{cases},$$

$$M_\Phi(t) = [x_{i,j}]_{N \times h_t}, \begin{cases} i = t+1, j = 1 & x_{i,j} = 1 \\ \text{else} & x_{i,j} = 0 \end{cases},$$

$$M_U(t,n) = [x_{i,j}]_{h_t \times n}, \begin{cases} i = j = 1 & x_{i,j} = 1 \\ \text{else} & x_{i,j} = 0 \end{cases},$$

$$M_Y(t) = [x_{i,j}]_{h_t \times N}, \begin{cases} j - i = t+1 & x_{i,j} = 1 \\ i = 1, j \le t+1 & x_{i,j} = A^{t+1-j} \\ \text{else} & x_{i,j} = 0 \end{cases},$$

$0_{m \times n}$ is an m by n zero matrix, $I_{m \times n}$ is an m by n identity matrix, H_t is $H = -J_A^{-1} J_B$ re-computed for an optimisation at time-step t considering its horizon, and h_t is the horizon length of optimisation at step t, thus allowing the equation to be used in receding, shrinking or a hybrid control scheme.

Using notation $E = [E_0^{\mathrm{T}} \; E_1^{\mathrm{T}} \; \dots \; E_{N-1}^{\mathrm{T}}]^{\mathrm{T}}$, (3) can be reorganised as

$$U = \Psi_Y Y + \Phi E, \tag{4}$$

where $\Phi = [\Phi_0 \; \Phi_1 \; \cdots \; \Phi_{N-1}]$ and $\Psi_Y = \sum_{i=0}^{N-1} \Phi_i M_Y(i)$.

Let $U^* = \Psi_Y Y$ be the optimal input and $J^* = J(U^*)$ the optimal (i.e., minimum) cost using a prescient forecaster. As derived in Appendix B, the increase in cost from optimal due to prediction error E, $\Delta J = J(U) - J^*$, is given by

$$\Delta J = E^{\mathrm{T}} \Theta E + E^{\mathrm{T}} \Omega Y \tag{5}$$

where $\Theta = \Phi^{\mathrm{T}} J_A \Phi$ and $\Omega = 2\Phi^{\mathrm{T}}(J_A \Psi_Y + J_B)$.

IV. Time Series Prediction and RHC

Many multi-horizon forecasting techniques have been proposed [11], which are usually divided into two steps: (1) A learning phase, where a model and its hyperparameters are chosen and the training data is fit to minimise an error p-norm $||v(t) - \hat{v}_t(t)||_p$ [12], [13]; and (2) A *cross validation* or *backtesting* phase, where performance of different models and hyperparameters across a separate set of testing data are compared.

Typically, when applying cross validation, MSE error $E^T E$ is used. This ignores possible dependencies between errors and also between the errors and the inputs. To include this knowledge of the RHC controller, a full step-by-by simulation of RHC has to be performed, or equivalently (5) can be used to compare the final cost of different models.

V. Cost matrix dimensionality reduction

Computing ΔJ using (5) is of $O(N^2)$ time complexity. Assuming repeated evaluation on a fixed system one can precompute Θ and ΩY, and calculate ΔJ efficiently for different values of E, the latter coming from different prediction models.

With this assumption, matrix decomposition can also be used to further reduce computation complexity by approximating matrix $\Theta \in \mathbb{R}^{N \times N}$ with a matrix of lower rank [10].

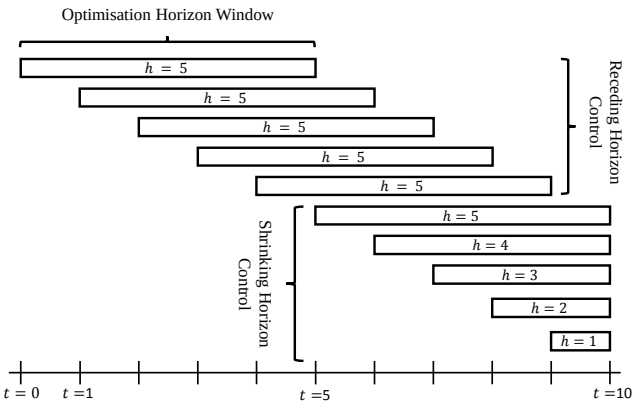

Fig. 1. A hybrid receding-shrinking horizon control scheme, with a maximum horizon of 5.

Let $\Theta = V \Sigma V^{\mathrm{T}}$, where Σ is the diagonal matrix of eigenvalues and V is the matrix of eigenvectors of Θ. By only keeping the $0 < L < N$ largest eigenvalues of Θ and their corresponding eigenvectors, Θ can be approximated with $\Theta' = V' \Sigma' V'^{\mathrm{T}}$, where $V' \in \mathbb{R}^{N \times L}$ and $\Sigma' \in \mathbb{R}^{L \times L}$. Let $W = V' \sqrt{\Sigma'}$. An approximation to (5) is

$$\Delta J' = E^{\mathrm{T}} W' W'^{\mathrm{T}} E + E^{\mathrm{T}} \Omega Y. \tag{6}$$

Exploiting the symmetric structure of the new cost function, diagonality of Σ, and assuming availability of W' and ΩY, evaluating (6) is reduced to time complexity $\mathrm{O}(NL)$.

The choice of L is problem dependent. A general guideline is to select L such that $\mathrm{tr}(\Sigma') \ge \lambda \mathrm{tr}(\Sigma)$, where tr is the trace operator and $0 \le \lambda \le 1$ determines how much of the matrix's *energy* is to be conserved. In practice, $\lambda > 0.99$ is commonly used.

VI. Numerical Example and Simulation Results

An LQ system with $A = 1$, $B = 1$, $C = 1$, $P_t = 1$, $Q_t = 1$ and $x(0) = 0$ is considered. The measurable disturbance v_t is generated via a 5th order auto-regressive (AR) model, $v_{t+1} = 2.76 v_t - 3.13 v_{t-1} + 1.79 v_{t-2} - 0.50 v_{t-3} + 0.05 v_{t-4} + \epsilon_t$. A receding/shrinking horizon controller governs the system for 10 steps, with 5 receding windows of length 5 followed by 5 windows with shrinking lengths (Fig. 1).

For predicting v_t, 5 different linear models were used to predict each future horizon h, using $\hat{v}_t(t+h) = \sum_{i=0}^{m_h} \beta_{h,m} v(t-m)$. An exhaustive search was performed to cross validate the model orders $m_h \in [2, \dots, 8]$, using MAE, MSE and ΔJ for each of the 5 horizons.

For each test, 100 values were generated with the AR model. 80% of values were assigned to *in-sample* testing, with 60 values used for training the predictor and the next 20 values for model validation. The last 20% were used to report *out-of-sample* prediction errors. The test was repeated 10 times.

Our first concern is the dimensionality reduction for $\Theta \in \mathbb{R}^{40 \times 40}$. The 10 largest eigenvalues hold more than 99.9% of eigenvalue energy, thus a 4 fold order reduction is possible using $W' \in \mathbb{R}^{40 \times 10}$. Table I shows the time required for computation of each error measure and the value of error. It

TABLE I
COMPARISON OF RUNTIME AND ACCURACY FOR STEP-BY-STEP SIMULATION, FULL CLOSED FORM AND APPROXIMATED ΔJ.

Using	Run-time (s)	Speed-up	Measured ΔJ
Simulator	1427.6640	0.0085	96.8476
Closed Form	12.1360	1.0000	96.8476
Approximated Θ'	2.4320	4.9901	96.8476

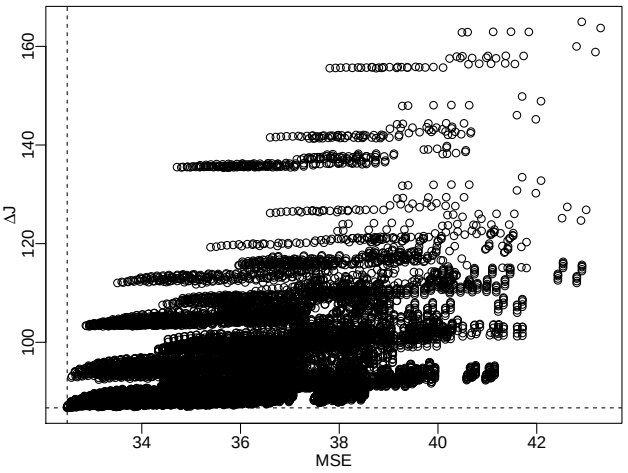

Fig. 2. MSE vs ΔJ for different model orders.

is observed the proposed dimensionality reduction techniques result in a near $5\times$ speed-up, with less than 0.001% loss of accuracy. Comparing run-times of Table I, an overall $587\times$ speed-up over a step-by-step RHC simulation is obtained. This value is mainly due to the fact that the simulator implementation is not vectorised.

Fig. 2 compares in-sample MSE error against ΔJ for different model orders. It is evident that these errors are not strongly correlated. As a result, model selection using MSE does not necessarily improve ΔJ.

Numerical results for models selected using MAE, MSE and ΔJ are summarised in Table II. It can be seen that using ΔJ as an error measure leads to a reduction in ΔJ and hence the controller's cost function J, compared with the standard MSE and MAE measures.

VII. CONCLUSION

This note presented several observations regarding error measures for LQ RHC in presence of a predictable disturbance. Using the techniques presented, computational costs can be reduced, and better results obtained since, in contrast to MSE and MAE, using ΔJ for model selection directly reduces the controller's objective function.

TABLE II
SUMMARY OF MODEL PERFORMANCES FOR OUT-SAMPLE PREDICTION USING MAE, MSE AND ΔJ.

Error measure for selection	In-sample MSE	In-sample ΔJ	Out-sample MSE	Out-sample ΔJ
MAE	32.5053	87.1546	36.2587	95.9280
MSE	**32.4888**	86.9027	**36.3447**	99.4730
ΔJ	32.4892	**86.7015**	38.0916	**94.7870**

APPENDIX

An implementation of the example in this paper, written in the R programming language, is available at http://www.ee.usyd.edu.au/cel/farzad/lqprediction.

A. Derivation of Eq. (4) (U)

Using the definitions of Section II, the system state in presence of prediction error is obtained by

$$X = S_B U + S_C Y + S_C E.$$

Let $\hat{Y} = Y + E = S_C^{-1}(S_A x(0) + S_C \hat{V})$. The solution to $U = \underset{U}{\operatorname{argmin}} J(U, \hat{V})$ is given by $U = H\hat{Y} = H(Y + E)$:

In a receding horizon control, inputs are recomputed at each time-step, e.g., at time-step t,

$$U_t = H_t \hat{\Upsilon}_t = H_t(\Upsilon_t + E_t)$$

where $\Upsilon_t \in \mathbb{R}^{h_t}$ is constructed by accumulating previous disturbances and inputs into the first value of Y_t, the tth horizon *window* of Y. Y_t can be obtained by the affine mapping $Y_t = M_Y(t)Y$.

At any time-step, Υ_{t+1} can be recursively computed by adding the observed values of $x(0)$, $v(i)$ and $u(i)$ for $i < t$, to form the new $x_{t+1}(0)$:

$$\begin{aligned}\hat{\Upsilon}_{t+1} &= Y_{t+1} + E_{t+1} + C^{-1}\sum_{i=0}^{t} A^{t+1-i} B M_U(h_{t+1}, h_i) U_i \\ &= Y_{t+1} + E_{t+1} + \sum_{i=0}^{t} C^{-1} A^{t+1-i} B M_U(h_{t+1}, h_i) H_i \hat{\Upsilon}_i\end{aligned}$$

where M_U is used to map the first element of U_i to ith element of U, to serve as the *applied* input control at time i.

This recursive equation can be simplified to

$$\hat{\Upsilon}_{t+1} = \sum_{i=0}^{t+1} \Gamma(t+1, i)(Y_i + E_i).$$

The final U is constructed using $M_\Phi(j)$ by appending first input obtained by each U_j to the jth position of U:

$$\begin{aligned}U &= \sum_{j=0}^{N-1} M_\Phi(j) U_j = \sum_{j=0}^{N-1} M_\Phi(j) H_j \sum_{i=0}^{N-1} \Gamma(j, i)(Y_i + E_i) \\ &= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} M_\Phi(j) H_j \Gamma(j, i)(M_Y(i)Y + E_i) \\ &= \sum_{i=0}^{N-1} \Phi_i (M_Y(i)Y + E_i)\end{aligned}$$

B. Derivation of Eq. (5) (ΔJ)

Let $U_e = U - U^* = \Phi E$ the difference of the optimal input and applied input due to prediction error. The cost difference between minimum cost using U^* and the cost with prediction error using U is given by

$$\begin{aligned}\Delta J &= J(U) - J(U^*) \\ &= U^{\mathrm{T}} J_A U + 2U J_B Y - U^{*\mathrm{T}} J_A U^* - 2U^* J_B Y \\ &= U_e^{\mathrm{T}} J_A U_e + 2U_e^{\mathrm{T}}(J_A \Psi_Y + J_B) Y.\end{aligned}$$

REFERENCES

[1] F. Borrelli, A. Bemporad, and M. Morari, *Predictive Control*. Cambridge University Press, 2014. [Online]. Available: http://www.mpc.berkeley.edu/mpc-course-material

[2] H. S. Ko, J. Jatskevich, G. Dumont, and G. G. Yoon, "An advanced LMI-based-LQR design for voltage control of grid-connected wind farm," *Electric Power Systems Research*, vol. 78, no. 4, pp. 539–546, 2008.

[3] P. Ignaciuk and A. Bartoszewicz, "Linear-quadratic optimal control of periodic-review perishable inventory systems," *Control Systems Technology, IEEE Transactions on*, vol. 20, no. 5, pp. 1400–1407, 2012.

[4] F. Noorian and P. Leong, "Dynamic hedging of foreign exchange risk using stochastic model predictive control," in *IEEE Symposium on Computational Intelligence for Financial Engineering & Economics (CIFEr)*. IEEE, 2014, pp. 441–448.

[5] C. Bordons and J. Cueli, "Predictive controller with estimation of measurable disturbances. Application to an olive oil mill," *Journal of Process Control*, vol. 14, no. 3, pp. 305–315, 2004.

[6] P. Doganis, E. Aggelogiannaki, and H. Sarimveis, "A combined model predictive control and time series forecasting framework for production-inventory systems," *International Journal of Production Research*, vol. 46, no. 24, pp. 6841–6853, 2008.

[7] J. Mattingley, Y. Wang, and S. Boyd, "Receding horizon control: Automatic generation of high-speed solvers," *IEEE control systems*, vol. 31, no. 3, pp. 52–65, 2011.

[8] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[9] J. Skaf, S. Boyd, and A. Zeevi, "Shrinking-horizon dynamic programming," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 17, pp. 1993–2002, 2010.

[10] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[11] S. Ben Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, "A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7067–7083, 2012.

[12] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *European journal of operational research*, vol. 160, no. 2, pp. 501–514, 2005.

[13] L. Cao, "Support vector machines experts for time series forecasting," *Neurocomputing*, vol. 51, pp. 321–339, 2003.