Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Model free probabilistic design with uncertain Markov parameters identified in closed loop

Jianfei Dong and Michel Verhaegen

***Abstract*— In this paper, we present a new probabilistic design approach based on Markov parameters identified via subspace methods from a finite batch of input and output data. This approach not only links closed-loop subspace identification with optimal control; but also directly evaluates parametric uncertainties on the identified Markov parameters. Neither a state-space model nor its stochastic uncertainty has to be realized in this approach. The effects of the parametric uncertainties on the output predictor are analyzed explicitly. Analytic solution to the probabilistic design is derived in a closed form, which avoids computing the empirical mean of a cost function as required by randomized algorithms. The solution hence leads to an easily implementable cautious optimal design, robust to the uncertainties in the identified Markov parameters from a closed-loop plant.**

I. INTRODUCTION

Model-free control design, also known as subspace predictive control (SPC), is recently presented in [1], [2], [3], [4]. These methods circumvent the modeling step, and design an LQG controller directly from the measured inputs and outputs (I/Os) of an unknown system. The key step in an SPC is identifying a future output predictor (in terms of Markov parameters), which maps the past I/Os and future inputs to the future outputs of a system. Then an LQG controller is parameterized by the identified predictor. Since the SPCs skip the realization of the system model and reply only on its Markov parameters, they avoid determining the model order and make it possible to control an infinite-dimensional system, [5].

The existing SPC approaches suffer from one open problem, which hampers their real-life application; i.e. the robustness of the SPC against the identification errors due to a limited number of noisy data samples. It is the purpose of this paper to solve this problem. It will be shown that the estimates of Markov parameters from a finite number of noisy data are subject to two error terms. One is a deterministic bias; and the other is a stochastic one, whose covariance matrix can be estimated from data. The estimates are then random, and hence uncertain.

Robust designs against random parametric uncertainties are known as probabilistic design in the literature, [6], [7], [8], [9], [10], [11]. In these approaches, random model parameters are associated with a probabilistic distribution function (pdf); e.g. a uniform distribution. The parameters are randomly sampled from this pdf, and parameterize a cost function. Either the worst ([9], [11]) or the average (also known as cautious design, see [7] and the references therein) realization of the cost is computed and then minimized for an optimal design. As long as a worst case is concerned, bounded random distributions are usually assumed, [8], [9], [11]; while a cautious design is free of this restriction, [7], [10]. In this sense, it is natural in a cautious design to link the uncertainties with identification errors, due to their unbounded randomness. In the case of a state-space model, the probabilistic designs require identifying not only the parametric matrices (e.g. $\{A, B, C, D\}$), but also their statistics. In fact, these statistics have to be deduced in some way from data. In [12], it is shown that the statistics of $\{A, B, C, D\}$ can be derived from the estimated Markov parameters and their statistics under restrictive assumptions, e.g. consistent order estimation and infinite data horizons. Obviously, the SPCs avoid this step by skipping the realization of the state-space matrices; and hence enables a model-free probabilistic design.

A major drawback with the probabilistic designs is the computational complexity in the randomized algorithms. This can be avoided when an analytic solution can indeed be derived for either the worst or the expected cost. However, it is possible to find such a closed-form solution, only when the uncertainties enter into the model affinely, [6], [10]. Clearly, such an affinity does not exist when the uncertainties in the state-space matrices have to be propagated into the Markov parameters. The SPC, as a model-free probabilistic design, ensures this affinity; and pave the way for a closed-form solution.

The paper is organized as follows. The errors in the identification problem are first evaluated, when the length of the noisy identification data is finite. In Section III, the effects of the bias and stochastic noise in the identified Markov parameters on the output predictor are analyzed. The expectation of a quadratic cost function with respect to the stochastic noise is explicitly derived in Section IV-A. The analytic solution to the model free probabilistic design is given in Section IV-B. The paper concludes with a simulation example and recommendations for future work.

II. IDENTIFICATION OF MARKOV PARAMETERS FROM FINITE DATA SAMPLES

We first formulate the problem of identifying Markov parameters from finitely many data samples.

This research is jointly funded by the grant of the European AI4IA project (Contract No. 514510) under the Marie Curie FP6 Framework, and *SKF ERC* in the Netherlands. The authors would also like to thank Dr. B. Kulcsar for the valuable discussions on this paper.

The authors are with Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands. j.dong@tudelft.nl & m.verhaegen@moesp.org

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

Consider the innovation type state-space model:

$$x(k+1) = Ax(k) + Bu(k) + Ke(k) \quad (1)$$
$$y(k) = Cx(k) + e(k), \quad (2)$$

where $e(k)$ is assumed to be a zero-mean white noise with a non-singular variance matrix of $\mathcal{E}\mathcal{E}^T$. The dimensions are assumed to be $x(k) \in \mathbb{R}^n$, $y(k) \in \mathbb{R}^\ell$, and $u(k) \in \mathbb{R}^m$. We make the following assumption on the plant, which is commonly assumed in subspace identification.

Assumption 1: $D = 0$; i.e. no direct feedthrough.

Assumption 2: $\Phi \triangleq A - KC$ is stable, and the system is minimal.

Assumption 1 is to ensure a one-step delay from the inputs to the outputs, and hence the well-posedness of the closed-loop identification problem. Assumption 2 is actually not restrictive, since any LTI state-space model has such an observer form; where K is the steady-state Kalman gain, and yields a stable Φ.

In the sequel, we denote by s, f respectively the past and future horizon, in both identification and control. N represents the number of columns in the identification data matrices. Let t be the current time instant in the formulation of the identification problem. We shall reserve k for the current time instant in the control formulation.

The identification of an output predictor of the state-space model (1) and (2) is the following least-squares problem,

$$Y_t = C\Phi^s X_{t-s} + \Xi_0 \mathcal{Z}_{[t-s,t)} + E_t. \quad (3)$$

$\Xi_0 \triangleq \begin{bmatrix} C\Phi^{s-1}B \; C\Phi^{s-1}K \; \cdots \; CB \; CK \end{bmatrix}$ contains the Markov parameters of the closed-loop steady-state Kalman filter of the system (1) and (2) with the steady-state gain K. $Y_t = \begin{bmatrix} y(t) \; y(t+1) \; \cdots \; y(t+N-1) \end{bmatrix}$ and $E_t = \begin{bmatrix} e(t) \; e(t+1) \; \cdots \; e(t+N-1) \end{bmatrix}$ are respectively the future output and innovation sequence. The past I/O data are collected in

$$\mathcal{Z}_{[t-s,t)} = \begin{bmatrix} u(t-s) & u(t-s+1) & \cdots & u(t-s+N-1) \\ y(t-s) & y(t-s+1) & \cdots & y(t-s+N-1) \\ u(t-s+1) & u(t-s+2) & \cdots & u(t-s+N) \\ y(t-s+1) & y(t-s+2) & \cdots & y(t-s+N) \\ \vdots & \vdots & & \vdots \\ u(t-1) & u(t) & \cdots & u(t+N-2) \\ y(t-1) & y(t) & \cdots & y(t+N-2) \end{bmatrix}.$$

$X_{t-s} = \begin{bmatrix} x(t-s) \; x(t-s+1) \; \cdots \; x(t-s+N-1) \end{bmatrix}$ is the sequence of the initial states.

Vectorizing the data equation (3) results in

$$\overrightarrow{Y}_t = F \cdot \Theta + \overrightarrow{E}_t + \Delta_Y, \quad (4)$$

where $F \triangleq \mathcal{Z}^T_{[t-s,t)} \otimes I_\ell$, $\overrightarrow{E}_t \triangleq vec(E_t)$, $\Theta \triangleq vec(\Xi_0)$, and $\Delta_Y \triangleq [I_N \otimes (C\Phi^s)]\overrightarrow{X}_{t-s}$. "$\otimes$" stands for Kronecker product. Before giving the solution, we need to assume the closed-loop plant is internally stable, where the data for the identification are collected. This is in fact a necessary condition to ensure the signals are quasi-stationary, as commonly assumed in identification literature, [13]. Specifically, we make the following assumption for the identification.

Assumption 3: The identification data are quasi-stationary and bounded by the positive numbers, $\bar{x}, \sigma_{ex}, \sigma_{xx}$, as

$$\|x(\tau)\|_2 \le \bar{x}, \; \forall t-s \le \tau \le t+N-1,$$
$$\|\mathbb{E}[\overrightarrow{E}_t \cdot \overrightarrow{X}^T_{t-s}]\|_2 \le \sigma_{ex} < \frac{\lambda_{min}(\mathcal{E}\mathcal{E}^T)}{\|C\Phi^s\|_2}, \text{ and}$$
$$\|Cov(\overrightarrow{X}_{t-s})\|_2 \le \sigma_{xx} < \frac{\lambda_{min}(\mathcal{E}\mathcal{E}^T)}{\|C\Phi^s\|_2^2},$$

where λ_{min} stands for the smallest eigenvalue.

Note that the upper bounds, $\bar{x}, \sigma_{ex}, \sigma_{xx}$, are not necessarily known, which are in fact hard to estimate. They are only needed to establish the following lemma.

Lemma 1: The least-squares solution of (4) is

$$\hat{\Theta} \triangleq vec(\hat{\Xi}_0) = vec(Y_t \cdot \mathcal{Z}^\dagger_{[t-s,t)}), \quad (5)$$

where"†" stands for pseudo-inverse. Under Assumption 3, the estimation error can be represented as

$$\hat{\Theta} - \Theta = \delta\Theta + \Sigma^{1/2}\epsilon, \text{ with} \quad (6)$$
$$\Sigma = [\mathcal{Z}_{[t-s,t)}\mathcal{Z}^T_{[t-s,t)}]^{-1} \otimes (\mathcal{E}\mathcal{E}^T), \quad (7)$$

where Σ is the covariance matrix; while $\delta\Theta$ is the mean of the bias, with a bounded 2 norm. $\epsilon \in \mathbb{R}^{M\ell \times 1}$ is a white noise with zero mean and identity covariance matrix, where $M = s(m+\ell)$.

Proof: As a sketch of proof, note that the initial states X_{t-s} and their correlation with the noise E_t are generally unknown. If Assumption 3 holds, then the contribution of the initial states to the covariance matrix is smaller than that of the measurement noise; and hence can be neglected by choosing a sufficiently large s. It is then straightforward to derive the covariance matrix, (7), based on the properties of least squares, [13]. ■

To ensure the existence of Σ and the solvability of $\mathcal{Z}^\dagger_{[t-s,t)}$, we assume that $\mathcal{Z}_{[t-s,t)}$ has full row rank.

Remark 1: Σ is related to the inverse of the signal to noise ratio (SNR); i.e. the higher the SNR of the identification signals, the lower the stochastic uncertainty on the estimated Markov parameters.

Note that it is hard to exactly estimate $\delta\Theta$. Hopefully, this is not necessary. We will treat its effect on the output predictor as a deterministic perturbation with bounded 2 norm in the next section.

III. The Output Predictor Based on the Uncertain Identified Markov Parameters

Form the above section, we see that the true Markov parameters of the plant are random in terms of the estimate,

$$\Theta = \hat{\Theta} + \delta\Theta + \Sigma^{1/2}\epsilon, \quad (8)$$

where the "$-$" sign before $\delta\Theta$ is changed to "$+$" for clarity. Ideally, the output predictor is built from the true Θ. But since it is estimated, we have to use the random expression as an alternative. First, with $\hat{\Theta}$ we can build an f-step ahead nominal deterministic predictor as follows.

Denote the deterministic prediction of the future outputs as $\hat{\mathbf{y}}^d_{[k,k+f)} \triangleq \begin{bmatrix} (\hat{y}^d(k))^T \; \cdots \; (\hat{y}^d(k+f-1))^T \end{bmatrix}^T$; and denote

$\mathbf{u}_{[k,k+f-1)}$ by $\left[u^T(k)\ \cdots\ u^T(k+f-2)\right]^T$ the future control inputs. To distinguish with $\mathcal{Z}_{[t-s,t)}$ in the identification problem, we use $\bar{Z}_{[k-s,k)} = \left[u(k-s)^T\ y(k-s)^T\ \cdots\ u(k-1)^T\ y(k-1)^T\right]^T$ to represent the past I/Os in the control problem. Then the nominal deterministic predictor reads, [1],

$$\hat{\mathbf{y}}^d_{[k,k+f)} = \Gamma \bar{Z}_{[k-s,k)} + \bar{\Lambda}\cdot \mathbf{u}_{[k,k+f-1)},$$
$$\Gamma = \begin{bmatrix} \Gamma_0 \\ \vdots \\ \Gamma_{f-1} \end{bmatrix}, \bar{\Lambda} = \begin{bmatrix} 0 \\ \Lambda \end{bmatrix}, \Lambda = \begin{bmatrix} \Lambda_1 & & \\ \vdots & \ddots & \\ \Lambda_{f-1} & \cdots & \Lambda_1 \end{bmatrix}. \tag{9}$$

The parameters, $\{\Gamma_i,\ \Lambda_j | i,j = 1,\cdots,f-1\}$, are computed by,

$$\begin{aligned}\Gamma_i &= \hat{\Xi}_i + \textstyle\sum_{\tau=0}^{i-1} \overline{C\Phi^{i-\tau-1}K}\cdot\Gamma_\tau, \Gamma_0 = \hat{\Xi}_0,\\ \Lambda_j &= \overline{C\Phi^{j-1}B} + \textstyle\sum_{\tau=1}^{j-1}\overline{C\Phi^{j-\tau-1}K}\cdot\Lambda_\tau, \Lambda_1 = \overline{CB},\end{aligned} \tag{10}$$

here $\overline{C\Phi^\tau B}$ and $\overline{C\Phi^\tau K}$, $\tau = 0,\cdots,f-2$, are the block elements of $\hat{\Xi}_0$; and $\hat{\Xi}_i = \left[0_{\ell\times i(m+\ell)}\ \overline{C\Phi^{s-1}B}\ \overline{C\Phi^{s-1}K}\ \cdots\ \overline{C\Phi^i B}\ \overline{C\Phi^i K}\right]$ is simply a right-shifted and zero-padded version of $\hat{\Xi}_0$. We denote $0_{m\times n}$ as an m-by-n zero matrix; and I_m as an m-dimensional identity matrix, in the sequel.

In the case that N, s, f are finite, the error terms, $\delta\Theta$ and $\Sigma^{1/2}\epsilon$, cannot be neglected. In fact, one needs to substitute the true Θ expressed as a random variable in (8) into the exact mapping from the past I/Os and states to the future outputs with true system parameters, [1],

$$\hat{\mathbf{y}}_{[k,k+f)} = \begin{bmatrix} \Xi_0 \\ \Xi_1 \\ \vdots \\ \Xi_{f-1}\end{bmatrix} \bar{Z}_{[k-s,k)} + \underbrace{\begin{bmatrix} C\Phi^s x(k-s) \\ C\Phi^s x(k-s+1) \\ \vdots \\ C\Phi^s x(k-s+f-1)\end{bmatrix}}_{\mathbf{b}_x}$$
$$+ \begin{bmatrix} 0 & & & \\ \Psi_1 & 0 & & \\ \vdots & \ddots & \ddots & \\ \Psi_{f-1} & \cdots & \Psi_1 & 0 \end{bmatrix} \cdot \begin{bmatrix} u(k) \\ y(k) \\ \vdots \\ u(k+f-1) \\ y(k+f-1) \end{bmatrix} + \mathbf{e}_{[k,k+f)}, \tag{11}$$

where $\Psi_\tau \triangleq C\Phi^{\tau-1}\left[B\ K\right]$, $\tau = 1,\cdots,f-1$; and Ξ_i is defined in the same way as $\hat{\Xi}_i$ but with true parameters. $\mathbf{b}_x$ depends on the unknown states $\{x(k-s),\cdots,x(k-s+f-1)\}$. We will treat $\mathbf{b}_x$ as a deterministic perturbation in the sequel. $\mathbf{e}_{[k,k+f)} = \left[e^T(k)\ \cdots\ e^T(k+f-1)\right]^T$ are the future unknown measurement noise.

The effects of the errors in $\hat{\Theta}$ on the output predictor (11) are summarized in Lemma 2. Before giving the statement, we first need to define the stochastic error terms as in (13) and (14), on the next page. The matrix Ψ_u in (13) denotes

$$\left[\begin{array}{c|c} O & \underbrace{\begin{bmatrix} I & 0 & 0 & \Lambda_1^T & 0 & \Lambda_2^T & 0 & \cdots & 0 & \Lambda_{f-2}^T \\ 0 & 0 & I & 0 & 0 & \Lambda_1^T & 0 & \cdots & 0 & \Lambda_{f-3}^T \\ 0 & 0 & 0 & 0 & I & 0 & 0 & \cdots & 0 & \Lambda_{f-4}^T \\ 0 & 0 & 0 & 0 & 0 & 0 & I & \cdots & 0 & \Lambda_{f-5}^T \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \Lambda_1^T \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & I & 0 \end{bmatrix}}_{\mathcal{K}\in\mathbb{R}^{(f-1)m\times(f-1)(m+\ell)}} \end{array}\right], \tag{12}$$

with O denoting a $(f-1)m\times(s-f+1)(m+\ell)$ zero matrix.

Lemma 2: The f-step ahead output predictor based on the biased and noisy identified Markov parameters in (5), (6), and (7) is given by

$$\begin{aligned}\hat{\mathbf{y}}_{[k,k+f)} &= \hat{\mathbf{y}}^d_{[k,k+f)} + \begin{bmatrix} 0 \\ \mathcal{L}\end{bmatrix}\cdot\mathbf{e}_u + \tilde{\mathcal{L}}\cdot\mathbf{e}_z + \tilde{\mathcal{L}}\cdot\mathbf{b} + \\ &\tilde{\mathcal{L}}\cdot\mathbf{e}_{[k,k+f)} + o\big(\max\{\|\Sigma^{1/2}\|_2^2, \|\delta\Theta\|_2^2, \\ &\quad \|\Sigma^{1/2}\|_2\cdot\|C\Phi^s\|_2, \|\Sigma^{1/2}\|_2\cdot\|\delta\Theta\|_2\}\big),\ \text{where}\end{aligned} \tag{15}$$

$$\mathcal{L} = \begin{bmatrix} L_1 & & & \\ L_2 & L_1 & & \\ \vdots & \vdots & \ddots & \\ L_{f-1} & L_{f-2} & \cdots & L_1 \end{bmatrix}, \tilde{\mathcal{L}} = \left[\begin{array}{c|c} L_1 & 0 \\ \hline L_2 & \\ \vdots & \mathcal{L} \\ L_f & \end{array}\right]. \tag{16}$$

$L_1 = I_\ell$ and $L_j = \sum_{\tau=1}^{j-1}\overline{C\Phi^{j-1-\tau}K}\cdot L_\tau,\ j = 2,\cdots,f.$ The symbol "$o(\alpha)$" stands for the terms with magnitude equal to or below α. The stochastic perturbations, $\mathbf{e}_u$ and $\mathbf{e}_z$, are given by (13) and (14). The perturbations due to the deterministic error $\delta\Theta$ are collected in $\mathbf{b}$ as

$$\mathbf{b} = \mathbf{b}_x + \mathbf{b}_u + \mathbf{b}_z,$$

where $\mathbf{b}_u = \begin{bmatrix} 0_{\ell\times 1} \\ ((U_k\cdot\Psi_u)\otimes I_\ell)\cdot\delta\Theta \end{bmatrix}$ and $\mathbf{b}_z = [V_k\otimes I_\ell]\cdot\delta\Theta$.

Proof: See Appendix. ■

Note that the higher-order terms in $o\big(\max\{\|\Sigma^{1/2}\|_2^2, \|\delta\Theta\|_2^2, \|\Sigma^{1/2}\|_2\cdot\|C\Phi^s\|_2, \|\Sigma^{1/2}\|_2\cdot\|\delta\Theta\|_2\}\big)$ are negligible if the following assumption holds.

Assumption 4: $\|\Sigma^{1/2}\|_2, \|C\Phi^s\|_2, \|\delta\Theta\|_2 \ll 1.$

In fact, the number of these terms grows exponentially with the prediction horizon f, yielding an untractable problem unless they are ignored. Note that this is not an unrealistic assumption, since $\|\Sigma^{1/2}\|_2 \ll 1$ can be achieved by a large enough (but finite) N; and $\|C\Phi^s\|_2, \|\delta\Theta\|_2$ can be made relatively small by choosing a sufficiently large (but finite) s.

Note that the perturbations, $\mathbf{b}_z$ and $\mathbf{b}_x$, depend on the information up to s steps prior to the current time k. Besides, as made clear in Section IV-B, the future control inputs depend linearly on $\bar{Z}_{[k-s,k)}$. It is then easy to verify that the following statement holds.

Proposition 1: If the closed-loop system is internally stable before the time instant k and Assumption 3 holds, then the perturbation $\mathbf{b}$ has a bounded 2 norm.

This proposition allows us to treat $\mathbf{b}$ as a norm-bounded perturbation, instead of exactly computing its value.

With the higher-order terms ignored, the output predictor (15) reduces to

$$\hat{\mathbf{y}}_{[k,k+f)} \approx \hat{\mathbf{y}}^d_{[k,k+f)} + \begin{bmatrix} 0 \\ \mathcal{L}\end{bmatrix}\cdot\mathbf{e}_u + \tilde{\mathcal{L}}\cdot\mathbf{e}_z + \tilde{\mathcal{L}}\cdot\mathbf{b} + \tilde{\mathcal{L}}\cdot\mathbf{e}_{[k,k+f)}. \tag{17}$$

Based on (17), we are then able to proceed to formulate an $\mathcal{H}_2$ optimal control problem in the following section.

IV. CAUTIOUS $\mathcal{H}_2$ OPTIMAL CONTROL

A. The $\mathcal{H}_2$ performance criterion

We consider the general trajectory tracking problem. Suppose the future references, $\mathbf{r}_{[k,k+f)} = [r^T(k)\ r^T(k+$

$$\mathbf{e}_u = \left[(U_k \cdot \Psi_u) \otimes I_\ell\right] \cdot \Sigma^{1/2} \cdot \epsilon, \text{ with } U_k = \begin{bmatrix} 0 & 0 & \cdots & 0 & u^T(k) \\ 0 & 0 & \cdots & u^T(k) & u^T(k+1) \\ \vdots & \vdots & & \vdots & \vdots \\ u^T(k) & u^T(k+1) & \cdots & u^T(k+f-3) & u^T(k+f-2) \end{bmatrix}. \tag{13}$$

$$\mathbf{e}_z = \left(V_k \otimes I_\ell\right) \cdot \Sigma^{1/2} \cdot \epsilon, \text{ with } V_k = \begin{bmatrix} \bar{Z}^T_{[k-s,k)} \\ \bar{Z}^T_{[k-s+1,k)} \quad 0_{1\times m} \quad \bar{Z}^T_{[k-s,k)}\Gamma_0^T \\ \bar{Z}^T_{[k-s+1,k)} \quad 0_{1\times m} \quad \bar{Z}^T_{[k-s,k)}\Gamma_0^T \quad 0_{1\times m} \quad \bar{Z}^T_{[k-s,k)}\Gamma_1^T \\ \vdots \\ \bar{Z}^T_{[k-s+f-1,k)} \quad 0_{1\times m} \quad \bar{Z}^T_{[k-s,k)}\Gamma_0^T \quad \cdots \quad 0_{1\times m} \quad \bar{Z}^T_{[k-s,k)}\Gamma_{f-2}^T \end{bmatrix}. \tag{14}$$

$1) \ \cdots \ r^T(k+f-1)]^T$, are known. Let $Q, R \succ 0$ be positive definite weighting matrices with appropriate dimensions. Let $0 < N_c < f-1$ be the control horizon; i.e. the control inputs, $u(k+i), N_c \leq i < f-1$, are frozen to $u(k+N_c-1)$. Consider the following $\mathcal{H}_2$ performance criterion,

$$\begin{aligned} J(k) \quad &= \quad \|\hat{\mathbf{y}}_{[k,k+f)} - \mathbf{r}_{[k,k+f)}\|_Q^2 + \|\Delta\mathbf{u}_{[k,k+N_c)}\|_R^2 \\ &\quad -\gamma\|\mathbf{b}_{[k,k+N_c+1)}\|_2^2, \end{aligned} \tag{18}$$

where $\gamma > 0$. $\|v\|_Q^2 \triangleq v^TQv$ denotes a weighted 2 norm. $\Delta u(k+i) = u(k+i) - u(k+i-1)$ is the input change at instant $k+i$. $\mathbf{b}_{[k,k+N_c+1)}$ represents the first $(N_c+1)\cdot\ell$ entries in the bias $\mathbf{b}$. The last term in the cost (18) is to cope with the deterministic disturbance, [14]. The novelty here is that $\mathbf{b}$ is linked with the bias in the identified Markov parameters, instead of being treated as an exogenous disturbance. Note that the future measurement noise, $\mathbf{e}_{[k,k+f)}$, is present in $\hat{\mathbf{y}}_{[k,k+f)}$, (17). Minimization of (18) ignoring the noise is equivalent to minimizing $\mathbb{E}_e J(k)$ over the same set of decision variables, $\Delta\mathbf{u}_{[k,k+N_c)}$. In this sense, $\mathbf{e}_{[k,k+f)}$ can be ignored in the sequel.

Due to the unbounded randomness of the white noise ϵ in the stochastic perturbations $\mathbf{e}_u$ and $\mathbf{e}_z$, we consider the average cost; i.e. $\mathbb{E}_\epsilon J(k)$. Minimizing this cost is known as cautious design in the literature, [7], [10]. The average cost takes not only the range of the parametric uncertainties, but also their likelihood into account; and therefore reduces the conservativeness as in a worst-case cost, [7], [10].

In fact, only the predictor-dependent part of $J(k)$,

$$J_y(k) \triangleq \|\hat{\mathbf{y}}_{[k,k+f)} - \mathbf{r}_{[k,k+f)}\|_Q^2, \tag{19}$$

is a function of ϵ, and needs to be evaluated for $\mathbb{E}_\epsilon J(k)$. This is because $\mathbf{b}_{[k,k+N_c+1)}$ is uncorrelated with the noise ϵ in the identified parameters. Substitute (9) and (17) into (19), take the expectation $\mathbb{E}_\epsilon$, and note that $\mathbf{r}_{[k,k+f)}$, $\hat{\mathbf{y}}^d_{[k,k+f)}$, and $\tilde{\mathcal{L}}\mathbf{b}$ are uncorrelated with ϵ ($\Gamma, \Lambda, \tilde{\mathcal{L}}$ are constructed from $\hat{\Theta}$, which is deterministic). Then it is straightforward to derive

$$\begin{aligned} \mathbb{E}_\epsilon J_y(k) \quad &= \quad \|\hat{\mathbf{y}}^d_{[k,k+f)} + \tilde{\mathcal{L}}\mathbf{b} - \mathbf{r}_{[k,k+f)}\|_Q^2 + \\ &\quad \mathbb{E}_\epsilon\Bigg[\underbrace{\Big\|\begin{bmatrix} 0 \\ \mathcal{L} \end{bmatrix} \cdot \mathbf{e}_u + \tilde{\mathcal{L}} \cdot \mathbf{e}_z\Big\|_Q^2}_{J_\epsilon(k)}\Bigg]. \end{aligned} \tag{20}$$

The following theorem shows that under the operation of $\mathbb{E}_\epsilon$, the quadratic optimization problem with the stochastic perturbations is equivalent to a deterministic problem, with two regularization terms on the decision variables. Before stating the theorem, we need to define three quantities as follows. First, let $\tilde{R}$ be

$$\begin{aligned} &\tilde{R} = q \cdot tr(\mathcal{E}\mathcal{E}^T) \cdot \textstyle\sum_{i=1}^{f-1} W_{ii}, \text{ with} \\ &W_{ii} = \begin{bmatrix} \mathcal{K}^{(f-1)m}_{(i-1)m+1} \cdot \Omega_{22} \cdot (\mathcal{K}^{(f-1)m}_{(i-1)m+1})^T & 0_{(f-i)m\times(i-1)m} \\ 0_{(i-1)m\times(f-i)m} & 0_{(i-1)m\times(i-1)m} \end{bmatrix}, \end{aligned} \tag{21}$$

where "tr" is the trace operator; and Ω_{22} are the last $(f-1)(m+\ell)$ rows and columns of $\Omega \triangleq [\mathcal{Z}_{[k-s,k)}\mathcal{Z}^T_{[k-s,k)}]^{-1}$. $\mathcal{K}$ is defined in (12). Second, define Π_1 as

$$\begin{aligned} &\Pi_1 = q \cdot tr(\mathcal{E}\mathcal{E}^T) \cdot \textstyle\sum_{i=1}^{f-1} \\ &\begin{bmatrix} \begin{bmatrix} \mathcal{I}_{f-i} & \Gamma^{\leftarrow}_{f-i} \end{bmatrix} \cdot \Omega_2 \cdot (\mathcal{K}^{(f-1)m}_{(i-1)m+1})^T & 0_{M\times(i-1)m} \end{bmatrix}, \end{aligned} \tag{22}$$

with $\mathcal{I}_i = \begin{bmatrix} 0_{i(m+\ell)\times(s-i)(m+\ell)} \\ I_{(s-i)(m+\ell)} \end{bmatrix}$, $\Gamma^{\leftarrow}_i = \begin{bmatrix}0 \ \Gamma_0^T \ 0 \ \Gamma_1^T \ \cdots \ 0 \ \Gamma^T_{i-1}\end{bmatrix}$, $\forall 1 \leq i \leq f-1$. Ω_2 are the last $(f-1)(m+\ell)$ columns of Ω. Finally, let Π_2 be

$$\Pi_2 = q \cdot tr(\mathcal{E}\mathcal{E}^T) \cdot \sum_{i=1}^{f} (\Psi_z\Omega\Psi_z^T)_{ii}, \tag{23}$$

where $(\Psi_z\Omega\Psi_z^T)_{ii}$ is the i-th diagonal block of $\Psi_z\Omega\Psi_z^T$, with Ψ_z defined by $\Psi_z = \begin{bmatrix} I_{s(m+\ell)} \\ \begin{bmatrix} \mathcal{I}_1 & \Gamma^{\leftarrow}_1 \\ \vdots & \vdots \\ \mathcal{I}_{f-1} & \Gamma^{\leftarrow}_{f-1} \end{bmatrix} \end{bmatrix}$, with the same $\mathcal{I}_i$ and $\Gamma^{\leftarrow}_i$ as defined in (22).

Theorem 1: Let $q > 0$ be a scalar tuning parameter. And let the weighting matrix Q be chosen as

$$Q = q \cdot (\tilde{\mathcal{L}} \cdot \tilde{\mathcal{L}}^T)^{-1}. \tag{24}$$

Then the expectation of the quadratic function $J_\epsilon(k)$, with respect to the noise ϵ in the identified Markov parameters, takes the following closed form,

$$\mathbb{E}_\epsilon J_\epsilon(k) = \begin{bmatrix} \bar{Z}_{[k-s,k)} \\ \mathbf{u}_{[k,k+f-1)} \end{bmatrix}^T \cdot \begin{bmatrix} \Pi_2 & \Pi_1 \\ \Pi_1^T & \tilde{R} \end{bmatrix} \cdot \begin{bmatrix} \bar{Z}_{[k-s,k)} \\ \mathbf{u}_{[k,k+f-1)} \end{bmatrix}, \tag{25}$$

where the weighting matrices, $\tilde{R}, \Pi_1, \Pi_2$, are respectively given in (21), (22), and (23). Besides, $\tilde{R}, Q \succ 0$.

Proof: The derivation of $\tilde{R}, \Pi_1, \Pi_2$ and the proof of this theorem relies heavily on the properties of Kronecker product and matrix trace. Due to the space limitation, we shall refer the details to [15]. ■

Denote by $\mathbb{I}_n$ a column vector with n one's. Define the following structure matrices,

$$S_\Delta = \begin{bmatrix} I_m & & & \\ -I_m & I_m & & \\ & \ddots & \ddots & \\ & & -I_m & I_m \end{bmatrix}, \tilde{S}_\Delta = S_\Delta^{-1},$$

$$\mathcal{I}_u = \left[\begin{array}{c|c} \multicolumn{2}{c}{I_{N_c m}} \\ \hline 0_{(f-N_c-1)m\times(N_c-1)m} & \mathbb{I}_{f-N_c-1}\otimes I_m \end{array}\right],$$

$$S_{uz} = \begin{bmatrix} 0_{m\times(s-1)(m+\ell)} & I_m & 0_{m\times\ell} \\ 0_{(f-2)m\times(s-1)(m+\ell)} & 0_{(f-2)m\times m} & 0_{(f-2)m\times\ell} \end{bmatrix},$$

$$\tilde{S}_{uz} = \begin{bmatrix} 0_{(f-1)m\times(s-1)(m+\ell)} & \mathbb{I}_{f-1}\otimes I_m & 0_{(f-1)m\times\ell} \end{bmatrix},$$

$$S_b = \begin{bmatrix} I_{(N_c+1)\ell} \\ 0_{(f-N_c-1)\ell\times(N_c+1)\ell} \end{bmatrix}.$$

Then, one can write the control changes within the control horizon N_c as

$$\Delta\mathbf{u}_{[k,k+N_c)} = S_\Delta \mathbf{u}_{[k,k+N_c)} - S_{uz}\bar{Z}_{[k-s,k)}; \qquad (26)$$

and similarly the control signals as

$$\begin{aligned} \mathbf{u}_{[k,k+N_c)} &= \tilde{S}_\Delta \Delta\mathbf{u}_{[k,k+N_c)} + \tilde{S}_{uz}\bar{Z}_{[k-s,k)} \text{ and} \\ \mathbf{u}_{[k,k+f-1)} &= \mathcal{I}_u \cdot \mathbf{u}_{[k,k+N_c)}. \end{aligned} \qquad (27)$$

As in [14], we freeze the deterministic perturbations at the end of the control horizon to zero; i.e. $b(k+N_c+1) = \cdots = b(k+f-1) = 0$. Then $\mathbf{b}$ can be written as,

$$\mathbf{b} = S_b \cdot \mathbf{b}_{[k,k+N_c+1)}. \qquad (28)$$

Using (27) and (28), one can find the following relation,

$$\begin{bmatrix} \bar{Z}_{[k-s,k)} \\ \mathbf{r}_{[k,k+f)} \\ \mathbf{u}_{[k,k+f-1)} \\ \mathbf{b} \end{bmatrix} = \underbrace{\begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ \mathcal{I}_u\tilde{S}_{uz} & 0 & \mathcal{I}_u\tilde{S}_\Delta & 0 \\ 0 & 0 & 0 & S_b \end{bmatrix}}_{\mathcal{P}} \cdot \begin{bmatrix} \bar{Z}_{[k-s,k)} \\ \mathbf{r}_{[k,k+f)} \\ \Delta\mathbf{u}_{[k,k+N_c)} \\ \mathbf{b}_{[k,k+N_c+1)} \end{bmatrix}. \qquad (29)$$

Substitute (20), (25), and (29) into $\mathbb{E}_\epsilon J(k)$, where $J(k)$ is defined by (18). By straightforward algebraic manipulations, it is easy to show that $\mathbb{E}_\epsilon J(k)$ takes the compact matrix form in (30). We shall derive a closed-form solution to optimize this $\mathcal{H}_2$ performance criterion in the next section.

B. The solution to the $\mathcal{H}_2$ optimal control problem

So far, we have shown that the bias and noise in the identified Markov parameters due to finite data horizon lead to the deterministic and stochastic perturbations in the output predictor. We have accounted for the stochastic perturbation in the average sense. In this section, we consider the worst-case $\mathcal{H}_2$ performance against the deterministic perturbation; and design optimal control inputs by minimizing this worst performance. Specifically, we have to solve the following problem,

$$\min_{\Delta\mathbf{u}_{[k,k+N_c)}} \max_{\mathbf{b}_{[k,k+N_c+1)}\in\mathcal{L}_2} \mathbb{E}_\epsilon J(k), \qquad (31)$$

where $\mathbf{b}_{[k,k+N_c+1)} \in \mathcal{L}_2$, according to Proposition 1. The main result is summarized in the following theorem.

Theorem 2: Let the following saddle condition be satisfied,

$$\gamma \cdot I \succ \left(1/q \cdot I - \mathcal{F}\cdot\mathcal{G}^{-1}\cdot\mathcal{F}^T\right)^{-1}, \text{ where} \qquad (32)$$

$\mathcal{F} = \tilde{S}_\Delta^T \mathcal{I}_u^T \bar{\Lambda}^T \tilde{\mathcal{L}}^{-T} S_b$ and $\mathcal{G} = -\tilde{S}_\Delta^T\mathcal{I}_u^T \cdot \big(q\cdot\bar{\Lambda}^T\tilde{\mathcal{L}}^{-T}\cdot \begin{bmatrix} 0 & 0 \\ 0 & I_{(f-N_c-1)\ell}\end{bmatrix}\tilde{\mathcal{L}}^{-1}\bar{\Lambda} + S_\Delta^T R S_\Delta + \tilde{R}\big)\cdot\mathcal{I}_u\tilde{S}_\Delta$.

Then the "$\min\max$" optimality of (31) holds. The optimal control changes are solved analytically as

$$\Delta\mathbf{u}^*_{[k,k+N_c)} = -\mathcal{A}(\mathcal{B}\bar{Z}_{[k-s,k)} + \mathcal{C}\mathbf{r}_{[k,k+f)}), \text{ where} \qquad (33)$$

$$\begin{aligned} \mathcal{A} &= [\tilde{S}_\Delta^T\mathcal{I}_u^T\cdot(\bar{\Lambda}^T\bar{Q}\bar{\Lambda} + S_\Delta^T R S_\Delta + \tilde{R})\cdot\mathcal{I}_u\tilde{S}_\Delta]^{-1}, \\ \mathcal{B} &= \tilde{S}_\Delta^T\mathcal{I}_u^T(\bar{\Lambda}^T\bar{Q}\Gamma + \Pi_1^T - S_\Delta^T R S_{uz}) + \\ &\quad \tilde{S}_\Delta^T\mathcal{I}_u^T\cdot(\bar{\Lambda}^T\bar{Q}\bar{\Lambda} + S_\Delta^T R S_\Delta + \tilde{R})\cdot\mathcal{I}_u\tilde{S}_{uz}, \\ \mathcal{C} &= -\tilde{S}_\Delta^T\mathcal{I}_u^T\bar{\Lambda}^T\bar{Q}, \\ \bar{Q} &= q\cdot\tilde{\mathcal{L}}^{-T}\cdot\begin{bmatrix} \frac{\gamma}{\gamma-q}I_{(N_c+1)\ell} & \\ & I_{(f-N_c-1)\ell}\end{bmatrix}\cdot\tilde{\mathcal{L}}^{-1}. \end{aligned}$$

Proof: The proof is based on the saddle point condition for solving "$\min\max$" problems. Similar procedures can be found e.g. in [4], [16]. We shall refer the details to [15]. ■

A simple choice of γ is $\gamma = \alpha\cdot q < 1$, with $q < 1, \alpha > 1$, which satisfies (32) and yields the matrix $\bar{Q} \succ 0$ and hence guarantees the existence of the inverse in $\mathcal{A}$. The online implementation of the control law (33) just boils down to solving the least squares problem (5) and computing a couple of algebraic expressions, (10), (16), (21), (22), and (33).

V. A SIMULATION EXAMPLE

As a case study, we consider the regulating problem of the following system,

$$(1-0.8z^{-1})y(k) = (z^{-1}-0.9z^{-2})u(k) + e(k), \qquad (34)$$

where $e(t)$ is a zero-mean white noise, with a variance of $\mathcal{S}\mathcal{S}^T = 0.25$. 2000 I/O samples are generated for identifying $\hat{\Xi}_0$ from a closed-loop experiment with an initial controller $u(k) = 0.5(r(k)-y(k))$, where the reference signal $r(k)$ is set to a zero-mean white noise with a standard derivation of 0.1 (i.e. small SNR). The past and future horizon are respectively set as $s = 10$ and $f = 10$. In this case, $\|\Sigma\|_2 = 0.07$, $\|C\Phi^s\|_2 \approx 0$, and $\|\delta\Theta\|_2 \approx 0$; i.e. Assumption 4 holds. The other parameters are chosen as $N_c = 7, R = 0.01I, \alpha = 2, q = 0.45$. We compare the optimal solution with the nominal design as proposed in [1], which minimizes the quadratic cost,

$$\underbrace{\|\hat{\mathbf{y}}^d_{[k,k+f)} - \mathbf{r}_{[k,k+f)}\|_Q^2}_{J_y^d(k)} + \|\Delta\mathbf{u}_{[k,k+f-1)}\|_R^2, \qquad (35)$$

where we choose the same Q, R, s, f as in the cautious design. $r(k) = 0$, to regulate the system to the origin. The predictor, $\hat{\mathbf{y}}^d_{[k,k+f)}$, in (35), is formulated from $\hat{\Xi}_0$, without considering its bias and noise. The closed-loop responses using the two control schemes are illustrated in Fig. 1. It is clear that the cautious design drives the system to the origin; while due to the large stochastic errors of $\hat{\Xi}_0$, the nominal design destabilizes the system.

To explain the effectiveness of the cautious design, the following four cost functions are compared; $\mathbb{E}_\epsilon J_y(k)$ as in (20), $J_y^d(k)$ in (35), the cost $J_y(k)$ as computed by using the true Ξ_0 in $\hat{\mathbf{y}}^d_{[k,k+f)}$ as in (35), and the randomized costs

$$\mathbb{E}_\epsilon J(k) = \begin{bmatrix} \bar{Z}_{[k-s,k)} \\ \mathbf{r}_{[k,k+f)} \\ \Delta\mathbf{u}_{[k,k+N_c)} \\ \mathbf{b}_{[k,k+N_c+1)} \end{bmatrix}^T \cdot \mathcal{P}^T \cdot \begin{bmatrix} \Gamma^T Q \Gamma + \Pi_2 + S_{uz}^T Q S_{uz} & -\Gamma^T Q & \Gamma^T Q \bar{\Lambda} + \Pi_1 - S_{uz}^T R S_\Delta & \Gamma^T Q \tilde{\mathcal{L}} \\ -Q\Gamma & Q & -Q\bar{\Lambda} & -Q\tilde{\mathcal{L}} \\ \bar{\Lambda}^T Q \Gamma + \Pi_1^T - S_\Delta^T R S_{uz} & -\bar{\Lambda}^T Q & \bar{\Lambda}^T Q \bar{\Lambda} + S_\Delta^T Q S_\Delta + \tilde{R} & \bar{\Lambda}^T Q \tilde{\mathcal{L}} \\ \tilde{\mathcal{L}}^T Q \Gamma & -\tilde{\mathcal{L}}^T Q & \tilde{\mathcal{L}}^T Q \bar{\Lambda} & \tilde{\mathcal{L}}^T Q \tilde{\mathcal{L}} - \gamma I \end{bmatrix} \cdot \mathcal{P} \cdot \begin{bmatrix} \bar{Z}_{[k-s,k)} \\ \mathbf{r}_{[k,k+f)} \\ \Delta\mathbf{u}_{[k,k+N_c)} \\ \mathbf{b}_{[k,k+N_c+1)} \end{bmatrix}. \quad (30)$$

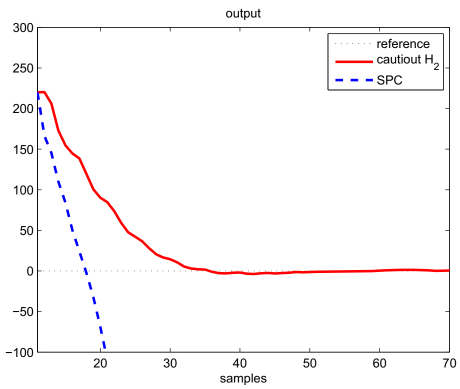

Fig. 1. The outputs of the closed-loop plant with both the cautious $\mathcal{H}_2$ controller and the SPC controller.

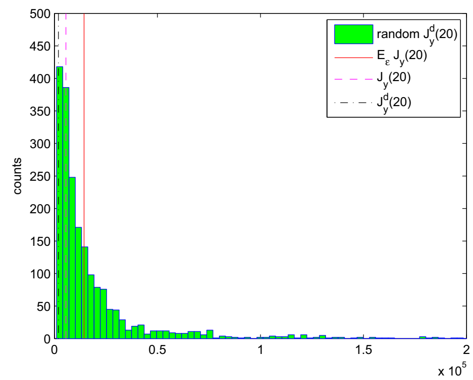

Fig. 2. The four different cost functions evaluated along the prediction horizon $[20, 30)$.

(denoted by $J_y^r(k)$) as computed by replacing $\hat{\Xi}_0$ in $J_y^d(k)$ with the estimates, $vec(\tilde{\Xi}_0) = vec(\hat{\Xi}_0) + \Sigma^{1/2}\epsilon$, from 2000 realizations of ϵ. Since $\|\delta\Theta\|_2 \approx 0$, the bias $\mathbf{b}$ in (20) is ignored. Obviously, $\mathbb{E}_\epsilon J_y(k) > J_y^d(k)$, due to the regularization term, $\mathbb{E}_\epsilon J_\epsilon(k)$. Note that the four cost functions depend also on $\bar{Z}_{[k-s,k)}, \mathbf{u}_{[k,k+f-1)}$, which are different among the four designs, due to the different control strategies. In the following analysis, the same $\bar{Z}_{[k-s,k)}, \mathbf{u}_{[k,k+f-1)}$ are used among the four costs, such that the statistics of the costs depend only on those of the noise ϵ.

The simulation results are shown in Fig. 2, where the costs along the prediction horizon $[20, 30)$ are considered. $\bar{Z}_{[10,20)}, \mathbf{u}_{[20,29)}$ generated by the cautious design are used in the other three cost functions. The 2000 costs, $J_y^r(20)$, are counted in 1000 bins in the histogram, on which the other three cost values are marked by lines. Fig. 2 clearly shows that $\mathbb{E}_\epsilon J_\epsilon(20)$ approximately equals the numerical mean of $J_y^r(20)$, i.e. 1.46×10^4. $\mathbb{E}_\epsilon J_\epsilon(20)$ upper bounds the cost incurred by using the true system parameters, $J_y(20)$. However, $J_y^d(20)$ underestimates the true cost, $J_y(20)$; and hence leads to instability of the closed loop.

VI. Conclusions

In this paper, we have introduced a model-free probabilistic design, which is robust to the uncertain Markov parameters identified from a closed-loop plant. The effect of the errors in the identified Markov parameters on the output predictor is explicitly analyzed. An optimal control law is derived in closed form. This approach solves the open problem on the robustness of SPCs against finite-horizon identification. The simulation example verifies the effectiveness of this approach in robustifying the closed-loop performance.

There is still one open problem for SPCs; i.e. the closed-loop stability under finite future prediction horizon. The challenge in ensuring the stability lies in the dynamic model hidden behind the Markov parameters. In fact, it is still an open problem to guarantee the closed-loop stability, when a dynamic model is missing. Other extensions of the proposed approach include handling constraints in the model free probabilistic design framework, and the recursive solutions.

References

[1] J. Dong, M. Verhaegen, and E. Holweg, “Closed-loop subspace predictive control for fault tolerant MPC design,” in *the Proceedings of the 17th IFAC World Congress*, Seoul, South Korea, 2008, pp. 3216–3221.

[2] W. Favoreel, B. de Moor, M. Gevers, and P. van Overschee, “Model-free subspace-based LQG-design,” *Tech. Rep. ESAT-SISTA/TR 1998-34, KU Leuven, Leuven, Belgium*, 1998.

[3] ——, “Closed-loop model-free subspace-based LQG-design,” in *the Proceedings of the 7th Mediterranean Conference on Control and Automation*, Haifa, Israel, 1999.

[4] B. Woodley, *Model free subspace based $\mathcal{H}_\infty$ control, PhD Dissertation*. Stanford University, 2001.

[5] K. Furuta and M. Wongsaisuwan, “Discrete-time LQG dynamic controller design using plant markov parameters,” *Automatica*, vol. 31, pp. 1317–1324, 1995.

[6] G. Calafiore and F. Dabbene, *Probabilistic and randomized methods for design under uncertainty*. Springer-Verlag, 2006.

[7] M. Campi and M. Prandini, “Randomized algorithms for the synthesis of cautious adaptive controllers,” *Systems & Control Letters*, vol. 49, pp. 21–36, 2003.

[8] S. Kanev, B. D. Schutter, and M. Verhaegen, “An ellipsoid algorithm for probabilistic robust controller design,” *Systems & Control Letters*, vol. 49, pp. 365–375, 2003.

[9] S. Kanev and M. Verhaegen, “Robustly asymptotically stable finite-horizon MPC,” *Automatica*, vol. 42, pp. 2189–2194, 2006.

[10] K. Öhrn, A. Ahlén, and M. Sternad, “A probabilistic approach to multivariable robust filtering and open-loop control,” *IEEE Transactions on Automatic Control*, vol. 40, pp. 405–418, 1995.

[11] B. Polyak and R. Tempo, "Probabilistic robust design with linear quadratic regulators," *Systems & Control Letters*, vol. 43, pp. 343–353, 2001.

[12] A. Chiuso, "On the relation between CCA and predictor-based subspace identification," *IEEE Transactions on Automatic Control*, vol. 52, pp. 1795–1812, 2007.

[13] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, 2007.

[14] W. Kwon and S. Han, *Receding horizon control: model predictive control for state models*. Springer-Verlag, 2005.

[15] J. Dong and M. Verhaegen, "Cautious $\mathcal{H}_2$ optimal control using uncertain markov parameters identified in closed loop," *Submitted for publication*, 2008.

[16] K. Furuta and M. Wongsaisuwan, "Closed-form solutions to discrete-time LQ optimal control and disturbance attenuation," *Systems & Control Letters*, vol. 20, pp. 427–437, 1993.

APPENDIX
PROOF OF LEMMA 2

We only derive the stochastic errors, $\begin{bmatrix}0 & \mathcal{L}^T\end{bmatrix}^T \cdot \mathbf{e}_u$ and $\tilde{\mathcal{L}}\mathbf{e}_z$, in (15). For this, we leave out $\delta\Theta$ and $\mathbf{b}_x$ respectively from (8) and (11) for clarity. The inclusion of these two terms leads to the deterministic perturbation, $\tilde{\mathcal{L}} \cdot \mathbf{b}$, which simply follows from the same lines of derivation. The higher order terms, $o(\cdot)$ in (15), are ignored based on Assumption 4. In the sequel, we shall use $(\Sigma^{1/2})_{n_1}^{n_2}$ to denote the partition of the matrix, $\Sigma^{1/2}$, from the n_1-th till the n_2-th row. For instance, as a part of Θ, $\overrightarrow{\overline{CK}}$ can be represented as $\overrightarrow{\overline{CK}} = vec(\overline{CK}) + (\Sigma^{1/2})_{(s-1)(m+\ell)\ell+m\ell+1}^{M\ell} \cdot \epsilon$.

We prove the result by induction as follows.

First, for $\tau = 0$, substitute the true Θ in (8) into the estimate $\hat{y}(k) = \Xi_0 \bar{Z}_{[k-s,k)}$ in (15); and vectorize the equation (note that $vec(\hat{y}(k)) = \hat{y}(k)$); i.e.

$$\begin{aligned}\hat{y}(k) &= (\bar{Z}_{[k-s,k)}^T \otimes I_\ell) \cdot \Theta = (\bar{Z}_{[k-s,k)}^T \otimes I_\ell) \cdot (\hat{\Theta} + \Sigma^{1/2}\epsilon) \\ &= (\bar{Z}_{[k-s,k)}^T \otimes I_\ell) \cdot \hat{\Theta} + (\bar{Z}_{[k-s,k)}^T \otimes I_\ell) \cdot \Sigma^{1/2}\epsilon \\ &= \hat{\Xi}_0 \cdot \bar{Z}_{[k-s,k)} + (\bar{Z}_{[k-s,k)}^T \otimes I_\ell) \cdot \Sigma^{1/2}\epsilon.\end{aligned}$$

Now, suppose the following expression holds,

$$\hat{\mathbf{y}}_{[k,k+i)} = \hat{\mathbf{y}}_{[k,k+i)}^d + \begin{bmatrix} 0 \\ \mathcal{L}_{i-1}\end{bmatrix} \cdot \mathbf{e}_u^{(1:i-1)} + \tilde{\mathcal{L}}_i \cdot \mathbf{e}_z^{(1:i)}, \quad (36)$$

where $\mathbf{e}_u^{(1:i-1)}$ collects the first $(i-1)\ell$ rows of $\mathbf{e}_u$ (same definition for $\mathbf{e}_z^{(1:i)}$). $\hat{\mathbf{y}}_{[k,k+i)}^d$ is defined in (9). $\mathcal{L}_{i-1}, \tilde{\mathcal{L}}_i$ are

$$\mathcal{L}_{i-1} = \begin{bmatrix} L_1 & & \\ \vdots & \ddots & \\ L_{i-1} & \cdots & L_1 \end{bmatrix}, \; \tilde{\mathcal{L}}_i = \begin{bmatrix} L_1 & & \\ \vdots & \ddots & \\ L_i & \cdots & L_1 \end{bmatrix}.$$

We shall then derive the i-th and the $(i+1)$-th block row of respectively $\mathbf{e}_u$ and $\mathbf{e}_z$ (corresponding to $\hat{y}(k+i)$). Based on (11), $\hat{y}(k+i)$ is predicted by,

$$\begin{aligned}\hat{y}(k+i) &= \begin{bmatrix} C\Phi^{i-1}B & \cdots & CB \end{bmatrix} \cdot \mathbf{u}_{[k,k+i)} + \\ &\Xi_i \bar{Z}_{[k-s,k)} + \begin{bmatrix} C\Phi^{i-1}K & \cdots & CK \end{bmatrix} \cdot \hat{\mathbf{y}}_{[k,k+i)} \quad (37) \\ &\triangleq \varphi_1 + \varphi_2 + \varphi_3.\end{aligned}$$

Substituting (36) into φ_3 leads to,

$$\begin{aligned}\varphi_3 = &\begin{bmatrix} C\Phi^{i-1}K & \cdots & CK \end{bmatrix} \cdot \\ &\Big[\hat{\mathbf{y}}_{[k,k+i)}^d + \begin{bmatrix} 0 \\ \mathcal{L}_{i-1}\end{bmatrix} \cdot \mathbf{e}_u^{(1:i-1)} + \tilde{\mathcal{L}}_i \cdot \mathbf{e}_z^{(1:i)}\Big]. \quad (38)\end{aligned}$$

The first term on the right hand side equals

$$vec\big([C\Phi^{i-1}K \cdots CK] \cdot \hat{\mathbf{y}}_{[k,k+i)}^d\big) = \big((\hat{\mathbf{y}}_{[k,k+i)}^d)^T \otimes I_\ell\big) \cdot \Bigg[vec([\overline{C\Phi^{i-1}K} \cdots \overline{CK}]) + \underbrace{\begin{bmatrix} (\Sigma^{1/2})_{(s-i)(m+\ell)\ell+m\ell+1}^{(s-i+1)(m+\ell)\ell} \\ \vdots \\ (\Sigma^{1/2})_{(s-1)(m+\ell)\ell+m\ell+1}^{M\ell} \end{bmatrix}}_{\mathcal{T}_K} \cdot \epsilon\Bigg].$$

The matrix $\mathcal{T}_K$ can be rewritten in a compact form as

$$\Bigg[\underbrace{\left[\begin{array}{c|ccccccccc} & 0 & I_\ell & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & 0 & 0 & 0 & I_\ell & 0 & \cdots & 0 & 0 & 0 \\ O_2 & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ & 0 & 0 & 0 & 0 & 0 & \cdots & I_\ell & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & I_\ell \end{array}\right]}_{\mathcal{S}_\Sigma \in \mathbb{R}^{i\ell \times i(m+\ell)}} \otimes I_\ell\Bigg] \cdot \Sigma^{1/2},$$

where O_2 is an $i\ell \times (s-i)(m+\ell)$ zero matrix. Therefore,

$$\begin{aligned}&\big((\hat{\mathbf{y}}_{[k,k+i)}^d)^T \otimes I_\ell\big) \cdot \mathcal{T}_K \cdot \epsilon = \Bigg[\Big([u^T(k) \cdots u^T(k+i-2)] \cdot \\ &\left[\begin{array}{c|cccc} & \Lambda_1^T & \Lambda_2^T & \cdots & \Lambda_{i-1}^T \\ & & \Lambda_1^T & \cdots & \Lambda_{f-3}^T \\ 0 & & & \ddots & \vdots \\ & & & & \Lambda_1^T \end{array}\right] \cdot \mathcal{S}_\Sigma\Big) \otimes I_\ell\Bigg] \cdot \Sigma^{1/2} \cdot \epsilon + \\ &\Big[\Big(\begin{bmatrix} \bar{Z}_{[k-s,k)}^T\Gamma_0^T & \cdots & \bar{Z}_{[k-s,k)}^T\Gamma_{i-1}^T \end{bmatrix} \cdot \mathcal{S}_\Sigma\Big) \otimes I_\ell\Big] \cdot \Sigma^{1/2} \cdot \epsilon \\ &= \Big[\Big(\begin{bmatrix} u^T(k) & \cdots & u^T(k+i-1) \end{bmatrix} \cdot \Psi_\Lambda\Big) \otimes I_\ell\Big] \cdot \Sigma^{1/2} \cdot \epsilon + \\ &(\Psi_\Gamma \otimes I_\ell) \cdot \Sigma^{1/2} \cdot \epsilon,\end{aligned}$$

where with the zero matrices, $O_3 \in \mathbb{R}^{im \times (s-i)(m+\ell)}$ and $O_4 \in \mathbb{R}^{1 \times (s-i)(m+\ell)}$, $\Psi_\Lambda, \Psi_\Gamma$ are defined as

$$\Psi_\Lambda = \left[\begin{array}{c|cccccccccc} & 0 & 0 & 0 & \Lambda_1^T & 0 & \Lambda_2^T & 0 & \cdots & 0 & \Lambda_{i-1}^T \\ & 0 & 0 & 0 & 0 & 0 & \Lambda_1^T & 0 & \cdots & 0 & \Lambda_{i-2}^T \\ O_3 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \Lambda_1^T \\ & \mathit{0} & \mathit{0} & \mathit{0} & \mathit{0} & \mathit{0} & \mathit{0} & \mathit{0} & \cdots & \mathit{0} & \mathit{0} \end{array}\right]$$

$$\Psi_\Gamma = \left[\begin{array}{c|cccccc} O_4 & 0 & \bar{Z}_{[k-s,k)}^T\Gamma_0^T & 0 & \cdots & 0 & \bar{Z}_{[k-s,k)}^T\Gamma_{i-1}^T \end{array}\right].$$

Note that $u^T(k+i-1)$ is appended to $\mathbf{u}_{[k,k+i-1)}^T$; and correspondingly m rows of zeros (emphasized in Ψ_Λ) are padded on the bottom of $\mathcal{S}_\Sigma$ to get Ψ_Λ.

Ignoring the products of the bias and noise in Θ with $\mathbf{e}_u^{(1:i-1)}, \mathbf{e}_z^{(1:i)}$, the last two terms in (38) reduces to

$$\begin{aligned}&[\overline{C\Phi^{i-1}K}, \cdots, \overline{CK}] \cdot \begin{bmatrix} 0 \\ \mathcal{L}_{i-1}\end{bmatrix} \cdot \mathbf{e}_u^{(1:i-1)} = \\ &[\overline{C\Phi^{i-2}K}, \cdots, \overline{CK}]\mathcal{L}_{i-1}\mathbf{e}_u^{(1:i-1)} = [L_i, \cdots, L_2] \cdot \mathbf{e}_u^{(1:i-1)}, \\ &[\overline{C\Phi^{i-1}K}, \cdots, \overline{CK}]\tilde{\mathcal{L}}_i\mathbf{e}_z^{(1:i)} = [L_{i+1}, \cdots, L_2]\mathbf{e}_z^{(1:i)}.\end{aligned}$$

By similar manipulations, φ_1, φ_2 are derived as follows.

$$\begin{aligned}&\varphi_2 = \hat{\Xi}_i \cdot \bar{Z}_{[k-s,k)} + ([\bar{Z}_{[k-s+i,k)}^T, 0_{1 \times i(m+\ell)}] \otimes I_\ell)\Sigma^{1/2}\epsilon. \\ &\varphi_1 = (\mathbf{u}_{[k,k+i)}^T \otimes I_\ell) \cdot \big[vec([\overline{C\Phi^{i-1}B}, \cdots, \overline{CB}]) + \\ &\quad (\Psi_I \otimes I_\ell) \cdot \Sigma^{1/2} \cdot \epsilon\big];\end{aligned}$$

where, with $O_5 \in \mathbb{R}^{im \times (s-i)(m+\ell)}$,

$$\Psi_I = \left[\begin{array}{c|cccccccc} & I_m & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & 0 & 0 & I_m & 0 & \cdots & 0 & 0 & 0 \\ O_5 & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ & 0 & 0 & 0 & 0 & \cdots & 0 & I_m & 0 \end{array}\right].$$

Assemble $\varphi_1, \varphi_2, \varphi_3$ together into (37). Note that ([1]),

$$\begin{aligned}\hat{y}^d(k+i) &= (\mathbf{u}_{[k,k+i)}^T \otimes I_\ell) \cdot vec([\overline{C\Phi^{i-1}B}, \cdots, \overline{CB}]) + \\ &\hat{\Xi}_i \cdot \bar{Z}_{[k-s,k)} + \big((\hat{\mathbf{y}}_{[k,k+i)}^d)^T \otimes I_\ell\big) \cdot vec([\overline{C\Phi^{i-1}K} \cdots \overline{CK}]).\end{aligned}$$

Then, by straightforward arrangement of the terms, the proof of Lemma 2 is completed.