Information-seeking polynomial NARX model-predictive control through expected free energy minimization

Wouter M. Kouw

***Abstract*— We propose an adaptive model-predictive controller that balances driving the system to a goal state and seeking system observations that are informative with respect to the parameters of a nonlinear autoregressive exogenous model. The controller's objective function is derived from an expected free energy functional and contains information-theoretic terms expressing uncertainty over model parameters and output predictions. Experiments demonstrate that the proposed controller dynamically balances information-seeking and goal-seeking behaviour based on parameter uncertainty.**

I. Introduction

Nonlinear autoregressive exogenous (NARX) models are useful black-box descriptions of dynamical systems [1]. However, a model-predictive controller that relies on a NARX model that has yet to accurately capture dynamics can behave erratically, which may lead to damage during operation [2]. To avoid such risks, in practice, model parameters are often first estimated using a designed control signal (i.e., offline system identification). However, this is time-consuming. Here we present a model-predictive controller (referred to as an agent) that seeks control inputs that produce system outputs that are informative for online system identification, when its parameter uncertainty and predictive variance are high. When these are low, it seeks control inputs that lead to the goal output.

The proposed agent estimates a sequence of control inputs based on variational Bayesian inference, a form of statistical inference that optimizes an analytic approximation to posterior distributions [3]. Specifically, it minimizes an expected free energy (EFE) functional, a quantity proposed in the neuroscience community that describes perception and action in terms of information-theoretic quantities (also known as active inference) [4]. Free energy minimization is a relatively new framework for control, but it has strong ties to stochastic optimal control and encompasses many classical methods, such as proportional-integrative-derivative and linear-quadratic Gaussian control [5], [6]. So far, EFE-based controllers have relied on (partially) known dynamics or neural networks [7], [8], [9]. We propose an EFE minimizing agent based on a polynomial NARX model. NARX models do not require knowledge of system dynamics and require far fewer parameters than neural networks. Furthermore, free energy minimization in autoregressive models has been shown to produce superior estimates at small sample sizes and more stable k-step ahead predictions [10], [11], [12]. Our work falls within the scope of information-theoretic model-predictive control [13], [14], but we focus on information gathering specifically for NARX model identification.

Our contributions include:

- The derivation of parameter update rules for a conjugate prior to the Gaussian NARX likelihood.
- The derivation of a location-scale T-distributed posterior predictive distribution for outputs given control inputs.
- The derivation of an objective function for an adaptive model-predictive controller that dynamically balances information-seeking and goal-seeking behaviour.

II. Problem statement

Consider an agent, operating in discrete time, that receives noisy outputs $y_k \in \mathbb{R}$ from a system and sends control inputs $u_k\mathbb{R}$ back. It must drive the system to a desired output y_* without knowledge of the system's dynamics.

III. Model specification

We propose an agent with a probabilistic model that factorizes according to:

$$p(y_{1:N}, u_{1:N}, \theta, \tau) = p(\theta, \tau) \prod_{k=1}^{N} p(y_k \mid u_k, \theta, \tau) p(u_k) \quad (1)$$

for N time steps. The likelihood is auto-regressive in nature,

$$p(y_k \mid u_k, \theta, \tau) \triangleq \mathcal{N}(y_k \mid \theta^\top \phi(x_k, u_k), \tau^{-1}) \,, \quad (2)$$

with coefficients $\theta \in \mathbb{R}^D$ and output noise precision $\tau \in \mathbb{R}^+$ [11]. The variable x_k represents the collection of M_y past outputs and M_u past control inputs, $x_k \triangleq (y_{k-1}, \dots y_{k-M_y}, u_{k-1}, \dots u_{k-M_u})$. The function ϕ performs a polynomial basis expansion, $\phi : \mathbb{R}^M \rightarrow \mathbb{R}^D$, where $M = M_y + M_u + 1$ and D is the number of terms in the polynomial. For example, a second-order polynomial basis without cross-terms with $M_y = M_u = 1$ would produce $\phi(y_{k\text{-}1}, u_{k\text{-}1}, u_k) = [1 \ y_{k\text{-}1} \ u_{k\text{-}1} \ u_k \ y_{k\text{-}1}^2 \ u_{k\text{-}1}^2 \ u_k^2]$.

The prior distribution on the parameters is a multivariate Gaussian - univariate Gamma distribution [15, ID: D5]:

$$p(\theta, \tau) \triangleq \mathcal{NG}\big(\theta, \tau \mid \mu_0, \Lambda_0, \alpha_0, \beta_0\big) \quad (3)$$

$$= \mathcal{N}(\theta \mid \mu_0, (\tau\Lambda_0)^{-1}) \, \mathcal{G}\big(\tau \mid \alpha_0, \beta_0\big) \quad (4)$$

This choice of parameterization allows for inferring an exact posterior distribution (Lemma III.1) and an exact posterior predictive distribution (Lemma III.2).

The prior distributions over control inputs are assumed to be independent over time:

$$p(u_k) \triangleq \mathcal{N}(u_k \mid 0, \eta^{-1}) \,, \quad (5)$$

WM Kouw is with the Bayesian Intelligent Autonomous Systems laboratory, Department of Electrical Engineering, TU Eindhoven, Postal box 513, 5600 MB Eindhoven, Netherlands. @: w.m.kouw@tue.nl

with precision parameter η. This choice has a regularizing effect on the inferred controls (Sec. IV-B).

IV. Inference

A. Parameter estimation

First, we note that, at time k, the control u_k has been executed and is known to the agent. We use $\hat{u}_k$ and $\hat{y}_k$ to differentiate observed variables from unobserved ones, i.e., u_k and y_k.

We express the parameter posterior distribution as a Bayesian filtering procedure [16]:

$$\underbrace{p(\theta, \tau \mid \mathcal{D}_k)}_{\text{posterior}} = \frac{\overbrace{p\big(\hat{y}_k \mid \hat{u}_k, \theta, \tau\big)}^{\text{likelihood}}}{\underbrace{p\big(\hat{y}_k \mid \hat{u}_k, \mathcal{D}_{k\text{-}1}\big)}_{\text{evidence}}} \underbrace{p\big(\theta, \tau \mid \mathcal{D}_{k\text{-}1}\big)}_{\text{prior}} . \quad (6)$$

where $\mathcal{D}_k = \{\hat{y}_i, \hat{u}_i\}_{i=1}^k$ is the data up to time k. The evidence (a.k.a. marginal likelihood) is

$$p\big(\hat{y}_k | \hat{u}_k, \mathcal{D}_{k\text{-}1}\big) = \int p\big(\hat{y}_k | \hat{u}_k, \theta, \tau\big) p\big(\theta, \tau | \mathcal{D}_{k\text{-}1}\big) \, \mathrm{d}(\theta, \tau). \quad (7)$$

This integral is typically intractable. In earlier work, we obtained an approximate posterior distribution through mean-field variational inference [11]. In this model we obtain an exact posterior distribution because the multivariate Gaussian - univariate Gamma prior distribution specified in Eq. 3 is conjugate to the NARX likelihood.

Lemma IV.1. *The Gaussian distributed NARX likelihood in Eq. 2 combined with a multivariate Gaussian - univariate Gamma prior distribution over AR coefficients θ and measurement precision τ,*

$$p(\theta, \tau \mid \mathcal{D}_{k\text{-}1}) = \mathcal{NG}(\theta, \tau \mid \mu_{k\text{-}1}, \Lambda_{k\text{-}1}, \alpha_{k\text{-}1}, \beta_{k\text{-}1}) , \quad (8)$$

yields a multivariate Gaussian - univariate Gamma posterior distribution

$$p(\theta, \tau \mid \mathcal{D}_k) = \mathcal{NG}(\theta, \tau \mid \mu_k, \Lambda_k, \alpha_k, \beta_k) . \quad (9)$$

The parameters of the posterior distribution are

$$\begin{aligned} \mu_k &= \big(\phi_k \phi_k^\intercal + \Lambda_{k\text{-}1}\big)^{-1} \big(\phi_k \hat{y}_k + \Lambda_{k\text{-}1} \mu_{k\text{-}1}\big), \quad (10) \\ \Lambda_k &= \phi_k \phi_k^\intercal + \Lambda_{k\text{-}1}, \qquad \alpha_k = \alpha_{k\text{-}1} + \frac{1}{2}, \\ \beta_k &= \beta_{k\text{-}1} + \frac{1}{2}\Big(\hat{y}_k^2 + \mu_{k\text{-}1}^\intercal \Lambda_{k\text{-}1} \mu_{k\text{-}1} + \mu_k^\intercal \Lambda_k \mu_k\Big), \end{aligned}$$

where $\phi_k \triangleq \phi(\hat{x}_k, \hat{u}_k)$.

The proof is in Appendix A. The marginal posterior distributions are Gamma distributed and multivariate location-scale T-distributed [15, ID: P36]:

$$p(\tau \mid \mathcal{D}_k) = \int p(\theta, \tau | \mathcal{D}_k) \mathrm{d}\theta = \mathcal{G}(\tau \mid \alpha_k, \beta_k) \quad (11)$$

$$p(\theta \mid \mathcal{D}_k) = \int p(\theta, \tau | \mathcal{D}_k) \mathrm{d}\tau = \mathcal{T}_{2\alpha_k}\big(\theta^* | \mu_k, \frac{\beta_k}{\alpha_k} \Lambda_k^{-1}\big). \quad (12)$$

The subscript under $\mathcal{T}$ refers to its degrees of freedom.

B. Control estimation

In order to effectively drive the system to the goal output, the agent must make accurate predictions for future outputs. We express the probability of the outputs and parameters given a control input at $t = k + 1$ as

$$p(y_t, \theta, \tau \mid u_t) = p(y_t \mid \theta, \tau, u_t) \, p(\theta, \tau \mid \mathcal{D}_k). \quad (13)$$

Note that at time t, $\hat{x}_t$ is still observed. This probability distribution does not yet include a goal output and without it, the agent will drive the system towards *any* output that leads to minimal prediction error. To include the goal output, we first decompose the joint distribution above to:

$$p(y_t, \theta, \tau \mid u_t) = p(\theta, \tau \mid y_t, u_t) \, p(y_t) . \quad (14)$$

The right-hand side contains a marginal prior distribution over future output, $p(y_t)$, which we may constrain to a specific functional form:

$$p(y_t \mid y_*) \triangleq \mathcal{N}(y_t \mid m_*, v_*) . \quad (15)$$

This constrained prior distribution on the future output is known as a *goal prior* distribution [4].

Using the goal prior, the posterior distribution for the controls can be described as

$$p(u_t \mid y_t, \theta, \tau) = \frac{p(\theta, \tau \mid y_t, u_t) p(y_t \mid y_*) p(u_t)}{\int p(\theta, \tau \mid y_t, u_t) p(y_t \mid y_*) p(u_t) \mathrm{d}u_t} . \quad (16)$$

The integral in the denominator is challenging to solve due to the basis function applied to u_t. Instead, we introduce a variational distribution $q(u_t)$ to approximate the posterior. The approximation error is characterized with an expected free energy functional [4], [17],

$$\mathcal{F}_t[q] \triangleq \mathbb{E}_{q(y_t, \theta, \tau, u_t)} \Big[\ln \frac{q(u_t) p(\theta, \tau \mid \mathcal{D}_k)}{p(\theta, \tau | y_t, u_t) p(y_t | y_*) p(u_t)} \Big], \quad (17)$$

where the expectation is over

$$q(y_t, \theta, \tau, u_t) \triangleq p(\theta, \tau \mid y_t, u_t) p(y_t \mid y_*) q(u_t) . \quad (18)$$

Inferring the optimal control at time t refers to minimizing the free energy functional with respect to the variational distribution $q(u_t)$:

$$q^*(u_t) = \underset{q \in Q}{\arg\min} \ \mathcal{F}_t[q] . \quad (19)$$

where Q represents the space of candidate distributions. We can re-arrange the free energy functional to simplify the variational minimization problem. First, note that

$$\begin{aligned} &\mathbb{E}_{q(y_t, u_t, \theta, \tau)} \Big[\ln \frac{p(\theta, \tau \mid \mathcal{D}_k) \, q(u_t)}{p(\theta, \tau \mid y_t, u_t) p(y_t \mid y_*) p(u_t)} \Big] = \quad (20) \\ &\mathbb{E}_{q(u_t)} \Big[\mathbb{E}_{p(y_t, \theta, \tau | u_t)} \big[\ln \frac{p(\theta, \tau \mid \mathcal{D}_k)}{p(\theta, \tau | y_t, u_t) p(y_t | y_*)} \big] + \ln \frac{q(u_t)}{p(u_t)} \Big]. \end{aligned}$$

Now we define the expected free energy *function*:

$$\mathcal{J}(u_t) \triangleq \mathbb{E}_{q(y_t, \theta, \tau \mid u_t)} \Big[\ln \frac{p(\theta, \tau \mid \mathcal{D}_k)}{p(\theta, \tau \mid y_t, u_t) p(y_t \mid y_*)} \Big] . \quad (21)$$

Using $\mathcal{J}(u_t) = \ln(1/\exp(-\mathcal{J}(u_t)))$, the expected free energy functional can be expressed as a Kullback-Leibler divergence

$$\mathcal{F}_t[q] = \mathbb{E}_{q(u_t)}\Big[\ln \frac{q(u_t)}{\exp\big(-\mathcal{J}(u_t)\big)p(u_t)}\Big], \quad (22)$$

which is minimal when [3]:

$$q^*(u_t) = \exp\big(-\mathcal{J}(u_t)\big)p(u_t). \quad (23)$$

Thus, we have an optimal approximate posterior distribution over controls.

To proceed, we must solve the expectation in (21). We have access to the parametric forms of all distributions involved except for the distribution over parameters given the future output and control. It can be tied to known distributions through Bayes' rule:

$$p(\theta, \tau \mid y_t, u_t) = \frac{p(y_t \mid \theta, \tau, u_t)\ p(\theta, \tau \mid \mathcal{D}_k)}{\int p(y_t \mid u_t, \theta, \tau)p(\theta, \tau \mid \mathcal{D}_k)\mathrm{d}(\theta, \tau)}. \quad (24)$$

The distribution that results from the marginalization in the denominator is the posterior predictive distribution and can be derived analytically within our model.

Lemma IV.2. *The marginalization of the NARX likelihood (Eq. 2) over the parameter posterior distribution (Eq. 9) yields a location-scale T-distribution:*

$$p(y_t \mid u_t) = \int p(y_t \mid \theta, \tau, u_t)\ p(\theta, \tau \mid \mathcal{D}_k)\,\mathrm{d}(\theta, \tau) \quad (25)$$

$$= \int \mathcal{N}(y_t | \theta^\intercal \phi_t, \tau^{-1})\mathcal{NG}(\theta, \tau | \mu_k, \Lambda_k, \alpha_k, \beta_k)\mathrm{d}(\theta, \tau) \quad (26)$$

$$= \mathcal{T}_{\nu_t}\Big(y_t \mid m_t,\ s_t^2\Big), \quad (27)$$

where $\phi_t \triangleq \phi(\hat{x}_t, u_t)$ *and*

$$\nu_t \triangleq 2\alpha_k\,,\ m_t \triangleq \mu_k^\intercal \phi_t\,,\ s_t^2 \triangleq \frac{\beta_k}{\alpha_k}\big(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1\big). \quad (28)$$

The proof is in Appendix B.

If we replace $p(\theta, \tau \mid y_t, u_t)$ in Eq. 21 with the right-hand side of Eq. 24, then the EFE function can be split into two components:

$$\begin{aligned}\mathcal{J}(u_t) &= \mathbb{E}_{p(y_t, \theta, \tau \mid u_t)}\Big[\ln \frac{1}{p(y_t \mid y_*)}\Big] \quad (29)\\ &\quad + \mathbb{E}_{p(y_t, \theta, \tau \mid u_t)}\Big[\ln \frac{p(\theta, \tau \mid \mathcal{D}_k)p(y_t \mid u_t)}{p(y_t \mid \theta, \tau, u_t)p(\theta, \tau \mid \mathcal{D}_k)}\Big]\\ &= \mathbb{E}_{p(y_t \mid u_t)}\Big[-\ln p(y_t \mid y_*)\Big] \quad (30)\\ &\quad - \mathbb{E}_{p(y_t, \theta, \tau | u_t)}\Big[\ln \frac{p(y_t, \theta, \tau \mid u_t)}{p(\theta, \tau \mid \mathcal{D}_k)p(y_t \mid u_t)}\Big].\end{aligned}$$

One may recognize the first term as a cross-entropy, describing the dissimilarity between the predictive distribution and the goal prior distribution [3]. The second term is the mutual information between the parameter posterior and the posterior predictive distribution. It describes how much information is gained on the parameters upon measuring a system output [3].

Solving the expectations yields a compact objective:

Theorem IV.3. *The expected free energy function in Eq. 30 evaluates to:*

$$\begin{aligned}\mathcal{J}(u_t) = \frac{1}{2v_*}\Big(\big(\mu_k^\intercal \phi_t - m_*\big)^2 + \frac{\beta_k}{\alpha_k}\big(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1\big)\frac{\nu_t}{\nu_t - 2}\Big)\\ - \frac{1}{2}\ln\Big(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1\Big) + \textit{constants}. \quad (31)\end{aligned}$$

The proof is found in Appendix C.

We consider MAP estimation of the control policy, because - although it is informative to quantify uncertainty over controls - we can ultimately only execute one action. The MAP estimate can be expressed as a minimization over a negative logarithmic transformation of $q^*(u_t)$:

$$\hat{u}_t = \underset{u_t \in \mathcal{U}}{\arg\max}\, q^*(u_t) = \underset{u_t \in \mathcal{U}}{\arg\min}\ \mathcal{J}(u_t) - \ln p(u_t). \quad (32)$$

The $\mathcal{U}$ refers to the space of affordable controls and allows for incorporating practical constraints such as torque limits.

So far, we have only considered a 1-step ahead prediction. Generalizing to $t > 1$ is, in principle, straightforward. Due to the independence assumptions on the prior $p(u_t)$ and the variational control posteriors $q(u_t)$, the joint variational control posterior factorizes over time:

$$q^*(u_t, \dots, u_{t+T}) = \prod_{t=1}^{T} p(u_t)\ \exp\big(-\mathcal{J}(u_t)\big). \quad (33)$$

If we apply the same negative logarithmic transformation as in Eq. 32 to MAP estimation for the joint variational control posterior, then the final control optimization problem becomes:

$$\begin{aligned}\hat{u}^{\text{EFE}} = \underset{u \in \mathcal{U}^T}{\arg\min} \sum_{t=1}^{T} \frac{1}{2v_*}\big(\mu_k^\intercal \phi(\hat{x}_t, u_t) - m_*\big)^2 \quad (34)\\ + \frac{1}{2v_*}\Big(\frac{\beta_k \nu}{\alpha_k(\nu - 2)}\big(\phi(\hat{x}_t, u_t)^\intercal \Lambda_k^{-1}\phi(\hat{x}_t, u_t) + 1\big)\Big)\\ - \frac{1}{2}\ln\Big(\phi(\hat{x}_t, u_t)^\intercal \Lambda_k^{-1}\phi(\hat{x}_t, u_t) + 1\Big) + \eta u_t^2.\end{aligned}$$

where $u = (u_t, \dots, u_T)$ and $\mathcal{U}^T$ is the Cartesian product of input space over T steps. This generalization to $t > 1$ assumes that all elements of x_t have been observed. In non-Bayesian frameworks, the predicted y_t is typically incorporated into x_{t+1} as if it were an observation, thus allowing for recursive predictions arbitrarily far into the future [1], [18]. However, from a Bayesian inference perspective, y_t is not a number but a random variable. This creates a problem because the delay vector x_t would also become a random variable for $t > 1$, which would interfere with the conditioning and marginalization operations described earlier. We avoid this problem by collapsing the posterior predictive distribution for y_t (Eq. 49) into its most probable value $\mu_k^\intercal \phi(\hat{x}_t, \hat{u}_t)$, and incorporating that number into x_{t+1}. This comes at a cost: the predictive variance does not accumulate over the horizon.

The optimization problem (34) can be solved by iterative methods with automatic differentiation for obtaining the gradient. The box constraints over affordable controls $\mathcal{U}$ can be incorporated by utilizing an interior-point method.

V. EXPERIMENTS

A. Seeking informative observations

The more informative a system input-output observation, the faster the agent is able to make accurate predictions. To illustrate this, consider a system that evolves according to $y_k = \theta_1^* y_{k-1} + \theta_2^* u_k$ where $\theta_1^* = 0.5$ and $\theta_2^* = -0.5$. We shall compare two agents, one that estimates controls based on the objective function described in Eq.34 (denoted EFE) and one based on a typical quadratic cost function with regularization (denoted QCR) [2]:

$$\hat{u}^{\text{QCR}} = \underset{u \in \mathcal{U}^T}{\arg\min} \sum_{t=1}^{T} \left(\mu_k^{\intercal}\phi(\hat{x}_t, u_t) - m_*\right)^2 + \eta u_t^2 \qquad (35)$$

Both agents are tasked with driving the system to $m_* = 0.5$ (for EFE: $v_* = 1$) under $\mathcal{U} = [-1\, 1]$ for $T = 1$. The prior distribution's parameters are set to $\alpha_0 = 10$, $\beta_0 = 1$, $\mu_0 = [1\;1]$ and $\Lambda_0 = \frac{1}{2}I$. The top row in Figure 1 compares the two objective functions for the first action u_1, with the QCR objective (left) indicating $u_1 = 0.5$ to be optimal. The EFE objective is affected by the mutual information term, which indicates that u_t's further from 0 are more informative (note Eq. 34 uses *negative* mutual information, meaning the objective is smaller for more informative controls), and selects $u_1 = 0.96$. After executing the respective controls, each system returns a y_1 and we can update the parameter distribution. The middle row in Figure 1 shows two contour levels of the marginal prior and posterior parameter distributions (see Eq. 12) over the $[\theta_1\, \theta_2]$ plane. Note that the posterior under the EFE objective has moved closer to the system coefficients θ^* (i.e., $p(\theta^* \mid \mathcal{D}_1)$ is higher). If we now plot the posterior predictive distributions (bottom row), we see that the EFE predictions are closer to the system's y_2.

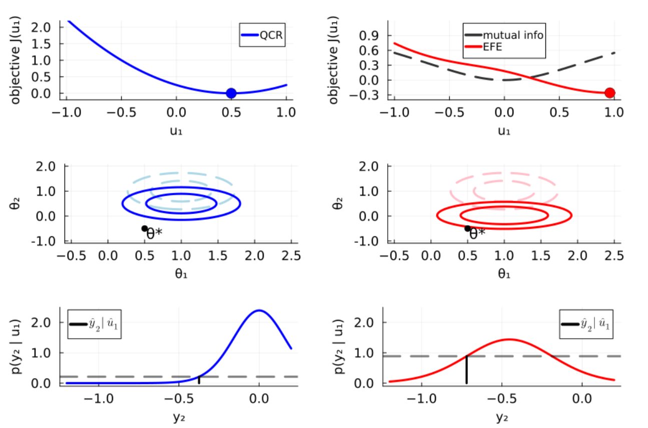

Fig. 1. Example of information-seeking. (Top row) Control objective functions for quadratic cost (QCR, left) vs proposed cost (EFE, right). Black dashed line displays mutual information, which causes $\hat{u}_1^{\text{EFE}} > \hat{u}_2^{\text{QCR}}$. (Middle row) Marginal prior (dashed) and posteriors (solid) for AR coefficients θ obtained after executing $\hat{u}_1^{\text{QCR}}$ (left) and $\hat{u}_1^{\text{EFE}}$ (right), and observing the resulting $\hat{y}_1$'s. Note that the posteriors under EFE controls are closer to the system AR coefficients θ^*. (Bottom row) Posterior predictive distributions for y_2 under QCR (left) and EFE (right) controls. Note that $\hat{y}_2$'s probability is larger under EFE, and thus, the uncertainty-weighted prediction error is smaller.

The information-seeking quality depends on parameter certainty, as can be seen by replicating the previous example with different values of Λ_0. Figure 2 (left) shows that the mutual information term is flatter for $\Lambda_0 = 2I$. It has less of an effect on the overall EFE objective (compare with $\Lambda_0 = \frac{1}{2}I$ shown in Figure 1 top right) and the optimal control input shrinks to $u_1 = 0.75$. For $\Lambda_0 = 100I$ (see Figure 2 right), the mutual information term is so flat that the EFE objective is nearly equal in shape as the QCR objective (see Figure 1 top left) and is also minimal at $u_1 = 0.5$.

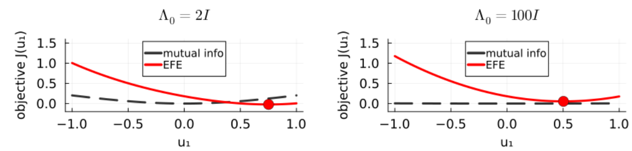

Fig. 2. Examples of how parameter certainty affects the mutual information term and the resulting EFE control objective. For $\Lambda_0 = 2I$ (left; medium certainty) the minimizer shrinks to $u_1 = 0.75$ (compared to $\Lambda_0 = \frac{1}{2}I$, Figure 1 top right). For $\Lambda_0 = 100I$ (right; high certainty), the minimizer shrinks all the way to $u_1 = 0.5$ as in the QCR objective (Figure 1 top left).

B. Control under unknown dynamics

In this experiment, the controllers must swing up a damped pendulum and maintain an upright position[1]. The system parameters are mass $\mathrm{m} = 1.0$, length $\ell = 0.5$, friction $\mathrm{c} = 0.01$ and time-step $\Delta t = 0.1$. The angle ϑ_t is observed under zero-mean Gaussian noise with a standard deviation of 0.001. The agent's NARX model consists of a second-order polynomial basis without cross-terms and with delays $M_y = 2$, $M_u = 2$. The prior distribution's parameters are noise shape $\alpha_0 = 10$, rate $\beta_0 = 0.1$, mean $\mu_0 = 10^{-8} \cdot [1 \ldots 1]$, coefficient precision $\Lambda = \frac{1}{2}I$ and control prior precision $\eta = 0.001$. This would be regarded as weakly informative for noise precision τ and non-informative for coefficients θ. The goal prior has mean $m_* = \pi$ (i.e., upward) and variance $v_* = 0.5$. The top row of Figure 3 visualizes the

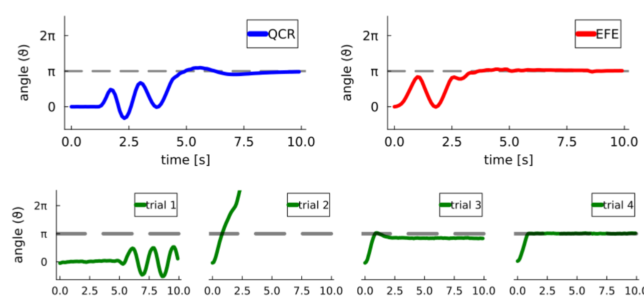

Fig. 3. Pendulum angle over time during swing-up task under unknown dynamics. (Top left) QCR-based controller, (top right) EFE-based controller. (Bottom) Trials for PILCO-based controller.

paths the two controllers took[2]. The QCR-based controller does not initially move much as it cannot predict the effect of the control inputs under the non-informative prior for the

[1] Code at `https://github.com/biaslab/ACC2024-NARXEFE`

[2] We used Optim.jl for optimizing (34) [19].

coefficients. When it starts, it swings up a few times before it reaches the goal and settles. The EFE-based controller moves away from zero earlier and chooses control inputs that let it swing up the pendulum using fewer tries.

For the sake of comparison, we also report the performance of PILCO [20]. It is a Gaussian Process-based identification method and although it is one of the most data-efficient reinforcement learning methods, it requires several trials before it finds a successful policy. The bottom row of Figure 3 shows its progress over the first 4 trials.

VI. Discussion

Close inspection of Eq. 31 reveals that the balance between the mutual information, cross-entropy and prior distribution is also affected by the goal prior variance v_*. It may be possible to optimize this dynamically. Furthermore, note that the optimization problem (34) is quadratic in ϕ_t. This means that for all nonlinear polynomials (i.e., degree larger than 1), the optimization becomes non-convex in u_t.

VII. Conclusion

We have proposed an adaptive NARX model-predictive controller that seeks informative observations for online system identification when its parameter uncertainty is high, and seeks control inputs that drive the system to a goal when its parameter uncertainty is low. This is achieved through a balance between information-theoretic quantities in the control objective, derived from the free energy minimization framework.

Appendix A

Proof: The posterior is proportional to the likelihood times the prior distribution:

$$\begin{aligned}
&p(\theta, \tau \mid \mathcal{D}_k) \\
&\propto \mathcal{N}\big(\hat{y}_k \mid \theta^\intercal \phi_k, \tau^{-1}\big)\mathcal{NG}\big(\theta, \tau | \mu_{k\text{-}1}, \Lambda_{k\text{-}1}, \alpha_{k\text{-}1}, \beta_{k\text{-}1}\big) \quad (36)\\
&\propto \tau^{1/2} \exp\big(-\frac{\tau}{2}(\hat{y}_k - \theta^\intercal \phi_k)^2\big) \quad (37)\\
&\qquad \tau^{D/2} \exp\Big(-\frac{\tau}{2}(\theta - \mu_{k\text{-}1})^\intercal \Lambda_{k\text{-}1}(\theta - \mu_{k\text{-}1})\Big)\\
&\qquad \tau^{\alpha_{k\text{-}1}-1} \exp\big(-\tau \beta_{k\text{-}1}\big)\\
&= \tau^{\alpha_{k\text{-}1}+(D+1)/2-1} \quad (38)\\
&\exp\Big(-\frac{\tau}{2}\big((\hat{y}_k - \theta^\intercal \phi_k)^2 + (\theta - \mu_{k\text{-}1})^\intercal \Lambda_{k\text{-}1}(\theta - \mu_{k\text{-}1}) + 2\beta_{k\text{-}1}\big)\Big).
\end{aligned}$$

If one expands the squares within the exponential and collects terms involving θ, one obtains

$$\begin{aligned}
&(\theta - \mu_{k\text{-}1})^\intercal \Lambda_{k\text{-}1}(\theta - \mu_{k\text{-}1}) + (\hat{y}_k - \theta^\intercal \phi_k)^2\\
&\quad = \theta^\intercal\big(\Lambda_{k\text{-}1} + \phi_k \phi_k^\intercal\big)\theta - \big(\mu_{k\text{-}1}^\intercal \Lambda_{k\text{-}1} + \hat{y}_k \phi_k^\intercal\big)\theta \qquad (39)\\
&\quad - \theta^\intercal\big(\Lambda_{k\text{-}1}\mu_{k\text{-}1} + \phi_k \hat{y}_k\big) + \mu_{k\text{-}1}^\intercal \Lambda_{k\text{-}1}\mu_{k\text{-}1} + \hat{y}_k^2 .
\end{aligned}$$

Then, using $\Lambda_k \triangleq \Lambda_{k\text{-}1} + \phi_k \phi_k^\intercal$ and $\xi_k \triangleq \Lambda_{k\text{-}1}\mu_{k\text{-}1} + \phi_k \hat{y}_k$, one may isolate a quadratic form:

$$\begin{aligned}
&\theta^\intercal \Lambda_k \theta - \xi_k^\intercal \theta - \theta^\intercal \xi_k \qquad (40)\\
&\quad = \big(\theta - \Lambda_k^{-1}\xi_k\big)^\intercal \Lambda_k \big(\theta - \Lambda_k^{-1}\xi_k\big) - \xi_k^\intercal \Lambda_k^{-1} \xi_k .
\end{aligned}$$

Plugging these back into (38), reveals a multivariate Gaussian - univariate Gamma distribution:

$$\begin{aligned}
&p(\theta, \tau \mid \mathcal{D}_k) \propto \tau^{\alpha_{k\text{-}1}+(D+1)/2-1}\\
&\exp\Big(-\frac{\tau}{2}\big((\theta - \Lambda_k^{-1}\xi_k)^\intercal \Lambda_k \big(\theta - \Lambda_k^{-1}\xi_k\big) \qquad (41)\\
&\qquad - \xi_k^\intercal \Lambda_k^{-1} \xi_k + \mu_{k\text{-}1}^\intercal \Lambda_{k\text{-}1}\mu_{k\text{-}1} + \hat{y}_k^2 + 2\beta_{k\text{-}1}\big)\Big)\\
&= \tau^{D/2} \exp\Big(-\frac{\tau}{2}\big(\theta - \Lambda_k^{-1}\xi_k\big)^\intercal \Lambda_k \big(\theta - \Lambda_k^{-1}\xi_k\big)\Big)\tau^{\alpha_{k\text{-}1}+\frac{1}{2}-1}\\
&\exp\Big(-\tau\big(\beta_{k\text{-}1} + \frac{1}{2}(\hat{y}_k^2 - \xi_k^\intercal \Lambda_k^{-1}\xi_k + \mu_{k\text{-}1}^\intercal \Lambda_{k\text{-}1}\mu_{k\text{-}1})\big)\Big), \quad (42)
\end{aligned}$$

with Λ_k defined as above (40), $\alpha_k \triangleq \alpha_{k\text{-}1} + \frac{1}{2}$ and

$$\begin{aligned}
\mu_k &\triangleq \Lambda_k^{-1}\xi_k = (\Lambda_{k\text{-}1} + \phi_k \phi_k^\intercal)^{-1}(\Lambda_{k\text{-}1}\mu_{k\text{-}1} + \phi_k \hat{y}_k) ,\\
\beta_k &\triangleq \beta_{k\text{-}1} + \frac{1}{2}\big(\hat{y}_k^2 - \xi_k^\intercal \Lambda_k^{-1}\xi_k + \mu_{k\text{-}1}^\intercal \Lambda_{k\text{-}1}\mu_{k\text{-}1}\big). \qquad (43)
\end{aligned}$$

Appendix B

Proof: We start by combining the Gaussian likelihood with the conditional Gaussian distribution for θ:

$$\begin{aligned}
&\mathcal{N}\big(y_t \mid \theta^\intercal \phi_t, \tau^{-1}\big)\mathcal{N}\big(\theta \mid \mu_k, (\tau\Lambda_k)^{-1}\big) = \qquad (44)\\
&\mathcal{N}\Big(\begin{bmatrix}\theta\\ y_t\end{bmatrix} \Big| \begin{bmatrix}\mu_k\\ \mu_k^\intercal \phi_t\end{bmatrix}, \begin{bmatrix}(\tau\Lambda_k)^{-1} & (\tau\Lambda_k)^{-1}\phi_t\\ \phi_t^\intercal(\tau\Lambda_k)^{-1} & \phi_t^\intercal(\tau\Lambda_k)^{-1}\phi_t + \tau^{-1}\end{bmatrix}\Big).
\end{aligned}$$

Marginalizing (44) over θ yields [16]:

$$p(y_t \mid \tau, u_t) = \mathcal{N}\big(y_t \mid \mu_k^\intercal \phi_t, \tau^{-1}(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1)\big) . \qquad (45)$$

Let $m_t \triangleq \mu_k^\intercal \phi_t$ and $\lambda_t \triangleq (\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1)^{-1}$. Then

$$\begin{aligned}
&p(y_t \mid u_t) = \int p(y_t \mid \tau, u_t) p(\tau) \mathrm{d}\tau \qquad (46)\\
&= \int \mathcal{N}\big(y_t \mid m_t, (\tau\lambda_t)^{-1}\big)\mathcal{G}(\tau \mid \alpha_k, \beta_k)\mathrm{d}\tau \qquad (47)\\
&= \int \Big(\frac{\lambda_t}{2\pi}\Big)^{1/2} \tau^{1/2} \exp\big(-\frac{\tau\lambda_t}{2}(y_t - m_t)^2\big)\\
&\qquad \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)}\tau^{\alpha_k - 1}\exp(-\tau\beta_k)\ \mathrm{d}\tau \qquad (48)\\
&= \Big(\frac{\lambda_t}{2\pi}\Big)^{1/2} \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)}\\
&\quad \int \tau^{\alpha_k + 1/2 - 1} \exp\Big(-\tau\big(\beta_k + \frac{\lambda_t}{2}(y_t - m_t)^2\big)\Big)\ \mathrm{d}\tau \qquad (49)\\
&= \Big(\frac{\lambda_t}{2\pi}\Big)^{1/2} \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} \frac{\Gamma(\alpha_k + \frac{1}{2})}{\Big(\beta_k + \frac{1}{2}\lambda_t(y_t - m_t)^2\Big)^{\alpha_k + 1/2}} \qquad (50)\\
&= \Big(\frac{\lambda_t}{2\pi}\Big)^{1/2} \frac{\Gamma(\alpha_k + \frac{1}{2})}{\Gamma(\alpha_k)}\Big(\frac{\beta_k}{\beta_k + \frac{1}{2}\lambda_t(y_t - m_t)^2}\Big)^{\alpha_k + \frac{1}{2}} \beta_k^{-\frac{1}{2}} \qquad (51)\\
&= \Big(\frac{\alpha_k \lambda_t}{2\alpha_k \beta_k \pi}\Big)^{1/2} \frac{\Gamma(\frac{2\alpha_k + 1}{2})}{\Gamma(\frac{2\alpha_k}{2})}\Big(1 + \frac{\alpha_k \lambda_t (y_t - m_t)^2}{2\alpha_k \beta_k}\Big)^{-\frac{2\alpha_k + 1}{2}} \qquad (52)
\end{aligned}$$

where the integrand in (49) is an unnormalized Gamma distribution and evaluates to its normalization constant in (50). Defining $\nu_t \triangleq 2\alpha_k$ and $s_t^2 \triangleq (\frac{\alpha_k}{\beta_k}\lambda_t)^{-1}$, reveals a location-scale T-distribution:

$$p(y_t \mid u_t) = \frac{1}{\sqrt{\pi \nu_t s_t^2}} \frac{\Gamma(\frac{\nu_t + 1}{2})}{\Gamma(\frac{2\nu_t}{2})}\Big(1 + \frac{(y_t - m_t)^2}{\nu_t s_t^2}\Big)^{-\frac{\nu_t + 1}{2}} . \qquad (53)$$

APPENDIX C

Proof: The cross-entropy is:

$$\mathbb{E}_{p(y_t \mid u_t)}\Big[-\ln p(y_t \mid y_*)\Big]$$
$$= \mathbb{E}_{\mathcal{T}_{\nu_t}(y_t \mid m_t, s_t^2)}\Big[-\ln \mathcal{N}(y_t \mid m_*, v_*)\Big] \tag{54}$$
$$= \frac{1}{2}\ln(2\pi v_*) + \frac{1}{2v_*}\mathbb{E}_{p(y_t \mid u_t)}\Big[y_t^2 - 2y_t m_* + m_*^2\Big] \tag{55}$$
$$= \frac{1}{2}\ln(2\pi v_*) + \frac{1}{2v_*}\Big(m_t^2 + s_t^2\frac{\nu_t}{\nu_t - 2} - 2m_t m_* + m_*^2\Big) \tag{56}$$
$$= \frac{1}{2}\ln(2\pi v_*) + \frac{1}{2v_*}\Big((m_t - m_*)^2 + s_t^2\frac{\nu_t}{\nu_t - 2}\Big). \tag{57}$$

The mutual information can be split into two marginal differential entropies and a joint differential entropy [3]:

$$\mathbb{E}_{p(y_t, \theta, \tau \mid u_t)}\Big[\ln \frac{p(y_t, \theta, \tau \mid u_t)}{p(\theta, \tau \mid \mathcal{D}_k)p(y_t \mid u_t)}\Big] \tag{58}$$
$$= H\Big[p(y_t|u_t)\Big] + H\Big[p(\theta, \tau|\mathcal{D}_k)\Big] - H\Big[p(y_t, \theta, \tau|u_t)\Big].$$

The first marginal entropy is of a location-scale T-distribution. Note that a location-scale T-distributed variable y is a linear transformation of a standard student's T-distributed variable x; $y = m + sx$. In general, differential entropy is invariant to translation, but follows $H[sx] = H[x] + \ln|s|$ with respect to scaling [3]. Using this property, we have:

$$H\Big[p(y_t \mid u_t)\Big] = H\Big[\mathcal{T}_{\nu_t}(y_t \mid 0, 1)\Big] + \frac{1}{2}\ln|s_t^2| \tag{59}$$
$$= \frac{\nu_t + 1}{2}\Big(\psi\big(\frac{\nu_t + 1}{2}\big) - \psi\big(\frac{\nu_t}{2}\big)\Big) + \ln\Big[\sqrt{\nu_t}\, B\big(\frac{\nu_t}{2}, \frac{1}{2}\big)\Big]$$
$$+ \frac{1}{2}\ln\Big|\frac{\beta_k}{\alpha_k}\big(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1\big)\Big|, \tag{60}$$

where $B(\cdot)$ is the beta function. The second marginal entropy is that of a multivariate Gaussian - univariate Gamma distribution, which evaluates to [15, ID: P238]:

$$H\Big[p(\theta, \tau \mid \mathcal{D}_k)\Big] = \frac{D}{2}\ln(2\pi e) - \frac{1}{2}\ln|\Lambda_k|$$
$$+ \alpha_k + \ln\Gamma(\alpha_k) + (1 - \alpha_k)\psi(\alpha_k) - \ln\beta_k, \tag{61}$$

where $\psi(\cdot)$ refers to a digamma function. The joint distribution $p(y_t, \theta, \tau|u_t)$ is the combination of the joint Gaussian in (44) and a marginal Gamma distribution. So its differential entropy is:

$$\mathbb{E}_{p(y_t, \theta, \tau|u_t)}\Big[\ln p(y_t, \theta, \tau \mid u_t)\Big] = -\frac{D+1}{2}\ln(2\pi e) + \tag{62}$$
$$\frac{1}{2}\ln|\Lambda_k| - \alpha_k - \ln\Gamma(\alpha_k) + \frac{D + 2\alpha_k}{2}\psi(\alpha_k) - \frac{D}{2}\ln\beta_k.$$

Given the differential entropies above, the mutual information term is:

$$\mathbb{E}_{p(y_t, \theta, \tau|u_t)}\Big[\ln \frac{p(y_t, \theta, \tau \mid u_t)}{p(\theta, \tau \mid \mathcal{D}_k)p(y_t|u_t)}\Big] = \tag{63}$$
$$\frac{D+2}{2}\psi(\alpha_k) - \frac{D+1}{2}\ln\beta_k + \frac{2\alpha_k + 1}{2}\psi\big(\alpha_k + \frac{1}{2}\big)$$
$$+ \ln\Big[\sqrt{2\alpha_k}B\big(\frac{2\alpha_k}{2}, \frac{1}{2}\big)\Big] + \frac{1}{2}\ln\Big[\frac{\beta_k}{\alpha_k}\big(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1\big)\Big].$$

Note that only the final term depends on u_t (through ϕ_t), and the others are constants. In total:

$$\mathcal{J}(u_t) = \frac{1}{2v_*}\Big((m_t - m_*)^2 + s_t^2\frac{\nu_t}{\nu_t - 2}\Big) \tag{64}$$
$$- \frac{1}{2}\ln\Big(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1\Big) + \text{constants}$$
$$= \frac{1}{2v_*}\Big(\big(\mu_k^\intercal \phi_t - m_*\big)^2 + \frac{\beta_k}{\alpha_k}\big(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1\big)\frac{\nu_t}{\nu_t - 2}\Big)$$
$$- \frac{1}{2}\ln\Big(\phi_t^\intercal \Lambda_k^{-1}\phi_t + 1\Big) + \text{constants}. \tag{65}$$

REFERENCES

[1] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[2] L. Grüne, J. Pannek, L. Grüne, and J. Pannek, *Nonlinear model predictive control*. Springer, 2017.

[3] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge University Press, 2003.

[4] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo, "Active inference and epistemic value," *Cognitive neuroscience*, vol. 6, no. 4, pp. 187–214, 2015.

[5] T. van de Laar, A. Özçelikkale, and H. Wymeersch, "Application of the free energy principle to estimation and control," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4234–4244, 2021.

[6] M. Baioumy, P. Duckworth, B. Lacerda, and N. Hawes, "Active inference for integrated state-estimation, control, and learning," in *IEEE International Conference on Robotics and Automation*, 2021, pp. 4665–4671.

[7] L. Pio-Lopez, A. Nizard, K. Friston, and G. Pezzulo, "Active inference and robot control: a case study," *Journal of The Royal Society Interface*, vol. 13, no. 122, p. 20160616, 2016.

[8] A. A. Meera and M. Wisse, "Free energy principle based state and input observer design for linear systems with colored noise," in *American Control Conference*, 2020, pp. 5052–5058.

[9] J. Huebotter, S. Thill, M. v. Gerven, and P. Lanillos, "Learning policies for continuous control via transition models," in *International Workshop on Active Inference*. Springer, 2023, pp. 162–178.

[10] K. Fujimoto and Y. Takaki, "On system identification for ARMAX models based on the variational Bayesian method," in *Conference on Decision and Control*, 2016, pp. 1217–1222.

[11] W. M. Kouw, A. Podusenko, M. T. Koudahl, and M. Schoukens, "Variational message passing for online polynomial NARMAX identification," in *American Control Conference*, 2022, pp. 2755–2760.

[12] A. Podusenko, S. Akbayrak, İ. Şenöz, M. Schoukens, and W. M. Kouw, "Message passing-based system identification for NARMAX models," in *Conference on Decision and Control*, 2022, pp. 7309–7314.

[13] C. Leung, S. Huang, N. Kwok, and G. Dissanayake, "Planning under uncertainty using model predictive control for information gathering," *Robotics and Autonomous Systems*, vol. 54, no. 11, pp. 898–910, 2006.

[14] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Information-theoretic model predictive control: Theory and applications to autonomous driving," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1603–1622, 2018.

[15] J. Soch, T. J. Faulkenberry, K. Petrykowski, and C. Allefeld, *The Book of Statistical Proofs*, 2021.

[16] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013, vol. 3.

[17] T. van de Laar, M. Koudahl, B. van Erp, and B. de Vries, "Active inference and epistemic value in graphical models," *Frontiers in Robotics and AI*, vol. 9, p. 794464, 2022.

[18] D. Khandelwal, M. Schoukens, and R. Tóth, "On the simulation of polynomial NARMAX models," in *Conference on Decision and Control*, 2018, pp. 1445–1450.

[19] P. K. Mogensen and A. N. Riseth, "Optim: A mathematical optimization package for Julia," *Journal of Open Source Software*, vol. 3, no. 24, p. 615, 2018.

[20] M. Deisenroth and C. E. Rasmussen, "PILCO: A model-based and data-efficient approach to policy search," in *International Conference on Machine Learning*, 2011, pp. 465–472.