On Adaptation of Loss Functions in Decentralized Adaptive Control

Václav Šmídl

Institute of Information Theory and Automation, Prague, Czech Republic, (e-mail: smidl@utia.cas.cz)

Abstract Decentralized adaptive agents/controller are able to operate in autonomous regimes, where they estimate model of their own neighborhood and design their control strategies with respect to their local aims. When each agent designs its strategy using only its model, the resulting control will be suboptimal since local models do not allow prediction of consequences of actions of the neighbors. Information about expected future trajectory as perceived by the neighbor can be obtained by means of communication. The task is to design a method of control strategy design that makes use of this information to improve the overall performance. In this paper, we propose to use the predictor to modify the loss function of the receiving agent. Informal justification of the approach is presented and it properties are illustrated in simulation experiments with decentralized LQG control.

Keywords: decentralized control, LQG control, fully probabilistic design

1. INTRODUCTION

This work is concerned with cooperative control strategy design for autonomous agents. This work is motivated by decentralized feedback control of traffic light signalization in urban areas, Homolová and Nagy (2005); Šmídl and Přikryl (2006). The current implementation of traffic light control is based hierarchical approach where some parameters (such as time offsets for a green wave) are set centrally but each intersection controller has certain level of autonomy to adjust its control strategy according to sensor readings available at that intersection. Currently, the level of autonomy is rather minor to prevent non-cooperative behavior which might appear since the controllers are not communicating to each other. An example of traffic control solution with higher level of autonomy is represented by the multi-agent approach, Roozemond (2001). In this approach, the agents (intersection controllers and area controllers) design predictors of traffic situation and design certainty equivalent control strategy using expected values of the predictions. The uncertainty of the predictors is mostly ignored, which is inadequate in such highly uncertain system as traffic flow in urban areas.

In order to improve handling of uncertainty, we make use of the multiple-participant decision-making theory, Kárný and Guy (2004); Kracík (2009). In this approach, the control strategy is based on design of the autonomous control strategy, however it is assumed that that communication with the neighbors may provide additional information. This information allows each agent to modify its key elements of the control strategy design (probabilistic model, and the loss function) to decrease its expected loss. In this approach, however, only local loss functions of each agent are considered. This is not a suitable assumption in the traffic control domain where the task is to optimize overall performance of the whole system. However, the overall improvement should not be achieved at the price of excessive loss in some areas. This makes the problem related to the domain of decentralized adaptive control which is well developed for deterministic systems, Ioannou (Apr 1986), or systems with stochastic inputs, Liu et al. (2007). However, models of the traffic network typically do not fit into commonly studied models in this approach.

To address the problem, we seek a combination of these three approaches that allows: detailed probability modeling of the multiple-participant approach, reconfigurability and intuitive appeal to field engineers of the multi-agent approach, and global properties (such as stability) of the decentralized adaptive control. In this paper, we investigate the following combination: (i) we develop an autonomous control strategy design based on probabilistic model of an agent, (ii) we develop a control strategy design method for the global system, and (iii) we modify the autonomous method to match the globally designed one as close as possible. At present, validation of success of the approach is tested only by simulation.

Following the multiple-participant approach, the agents are using Fully Probabilistic Control Design (FPD), Kárný (1996), as their control strategy design methodology. An independently developed variant of this approach was also used in multi-agent setup in van den Broek et al. (2008). The original FPD was extended to multi-agent scenario by exchange of multi-step predictive probability densities between agents in Šmídl and Andrýsek (2008). The multi-step predictors were merged with local predictors and the FPD was performed with respect to the merged density. However, this approach exhibits slow convergence and excessive computational demand as the multi-step predictor in the form of Gaussian density grows quadratically with the control horizon, and the time needed to process this information grows even more rapidly. Therefore, we propose a new scheme which is based on modification of the loss function of the receiving agent.

The paper is organized as follows. Review of the fully probabilistic control design for a centralized solution is presented in Section 2. Extension of this approach to decentralized scenario is presented in Section 3, where the main result is presented. An illustrative example of the new strategy is presented in Section 4.

2. REVIEW OF CENTRALIZED PROBABILISTIC ADAPTIVE CONTROL

Consider the following probabilistic model of a stochastic system:

$$y_t \sim f(y_t|u_t, d^{t-\partial:t-1}, \theta_t). \quad (1)$$

$$\theta_t \sim f(\theta_t|\theta_{t-1}). \quad (2)$$

Here, vector y_t denotes output of the system, u_t is the vector of inputs, d_t is their aggregation $d_t = [y_t', u_t']'$, $d^{t-\partial:t-1} = [d_{t-\partial}, \ldots, d_{t-1}]$ is a history of past ∂ observations, and θ_t is an unknown time-variant parameter, its evolution being modeled by (2). In the sequel, we replace $d^{t-\partial:t-1}$ by more general term $d^{1:t-1}$ for brevity. However, it is expected that all predictors and target pdfs will have fixed length statistics. The task is to design control strategy $u_t = u_t(d^{t-\partial:t-1})$, application of which yields output as close to the desired values as possible. Since parameters of the system are unknown and time-varying, we will design the control strategy using model-based adaptive control. This approach can be decomposed in three steps: (i) parameter estimation, (ii) prediction, in which the estimated parameters are used to build predictors of future behavior, and (iii) dynamic programing (DP) optimizing the control strategy using predictors from the prediction step.

Estimation: The model can be estimated using Bayesian filtering, Doucet et al. (2001), by recursive update of posterior density on parameters

$$f(\theta_t|d^{1:t}) \propto \int f(y_t|u_t, d^{1:t}, \theta_{t-1}) f(\theta_t|\theta_{t-1}) d\theta_{t-1},$$

where $\propto$ means equality up to normalizing constant.

Prediction: for given control law, the h-step predictor of y_t is obtained by marginalization:

$$f(y^{(h)}, u^{(h)}|d^{1:t}) = \int f(\theta_t|d^{1:t}) \Big[\prod_{\tau=t+1}^{t+h} f(y_\tau|u_\tau, d^{1:\tau}, \theta_\tau) f(u_\tau|d^{1:\tau}) f(\theta_\tau|\theta_{\tau-1}) \Big] d\theta^{t:t+h} \quad (3)$$

Here, $y^{(h)} \equiv y^{t+1:t+h}$ denotes future h-step output trajectory. Evaluation of (3) is often either intractable or difficult to use in the subsequent steps. Hence, it can be replaced by the following approximation

$$f(y^{(h)}|u^{(h)}, d^{1:\tau-1}) \approx \prod_{\tau=t+1}^{h} f(y_\tau|u_\tau, d^{1:\tau-1}, \hat{\theta}_\tau) f(u_\tau|d^{1:t-1}, \hat{\theta}_\tau). \quad (4)$$

Here, $\hat{\theta}_\tau$ is a point estimate of the parameters θ_τ. This approximation is known as certainty equivalence.

DP: Control strategy for system (1) can be designed using Fully Probabilistic Design (FPD) Kárný (1996). FPD minimizes future expected loss in the form of Kullback-Leibler divergence between the predicted and the target pdf of future trajectory.

$$f(u^{(h)}|y^{(h-1)}d^{1:t}) \quad (5)$$

$$= \arg \min_{f(u^{(h)}|y^{(h-1)}, d^{1:t})} \mathcal{D}\Big(f(d^{(h)}|d^{1:t}) \Big|\Big| g(d^{(h)}) \Big). \quad (6)$$

$$= \arg \min_{f(u^{(h)}|y^{(h-1)}, d^{1:t})} E\left[\ln \frac{f(d^{(h)}|d^{1:t})}{g(d^{(h)}|d^{1:t})} \mid d^{(h)} \right]. \quad (7)$$

Here, $E[\cdot \mid \cdot]$ denotes conditional expected value with respect to pdf (3), $\mathcal{D}(\cdot||\cdot)$is the Kullback-Leibler divergence, $f(d^{(h)}|d^{1:t})$ is a predictive pdf of future outputs (3), and $g(d^{(h)}|d^{1:t})$ is the desired (target) pdf of the future outputs. This non-standard technique of control strategy design allows analytical solution in the same backward-processing way as in dynamic programming, Bertsekas (2001).

Solution of (5) can be found explicitly in the form

$$f\left(u_\tau|d^{1:\tau-1}\right) = g(u_\tau|d^{1:\tau-1}) \frac{\exp[-\omega(u_\tau, d^{1:\tau-1})]}{\gamma(d^{1:\tau-1})}, \quad (8)$$

for $\tau = t+h, \ldots, t+1$, where function $\omega(\cdot)$ and $\gamma(\cdot)$ are recursively evaluated as

$$\omega(u_\tau, d^{1:\tau-1}) = E\left[\ln \frac{f(y_\tau|u_\tau, d^{1:\tau-1})}{\gamma(d^{1:\tau}) g(y_\tau|u_\tau, d^{1:\tau-1})} \mid y_\tau \right], \quad (9)$$

$$\gamma(d^{1:\tau-1}) = \int g(u_\tau|d^{1:\tau-1}) \exp[-\omega(u_\tau, d^{1:\tau-1})] \, du_\tau, \quad (10)$$

for $\tau = t+h, \ldots, t+1$, initiated at

$$\gamma(d^{1:t+h}) = 1. \quad (11)$$

For linear Gaussian models model the mean value of (5) is equivalent to the control law of linear quadratic Gaussian (LQG) method.

2.1 Special case of Linear Quadratic design

LQG control arise as a special case of FPD (8)–(11), when both the model and the target pdfs are Gaussian with linear function of their mean value:

$$f(y_t|u_t, d^{1:t-1}) = \mathcal{N}(\theta\psi_t, R),$$

$$g(y_t, u_t|d^{1:t-1}) = \mathcal{N}\left(\begin{bmatrix} \overline{y}_t \\ \overline{u}_t \end{bmatrix}, \begin{bmatrix} Q_y & \\ & Q_u \end{bmatrix} \right). \quad (12)$$

Here, θ is a vector of known (estimated) parameters, and ψ is a function of time-delayed values of y_t and u_t.

Substituting (12) into (9) for $\gamma(d^{1:t+h}) = 1$ yields:

$$\omega(u_{t+h}, d^{1:t+h-1}) = \frac{1}{2}\big[\ln(Q_y R^{-1}) - dim(y) + tr(RQ_y^{-1}) + (\theta\psi_t - \overline{y}_t)' Q_y^{-1} (\theta\psi_t - \overline{y}_t) \big], \quad (13)$$

$$\gamma(d^{1:t+h-1}) = \int \exp\left(-\frac{1}{2} (\theta\psi_t - \overline{y}_t)' Q_y^{-1} (\theta\psi_t - \overline{y}_t) \right) \int \exp\left(-\frac{1}{2} (u_t - \overline{u}_t) Q_u^{-1} (u_t - \overline{u}_t) \right), \quad (14)$$

where the first three terms in $\omega()$ are independent of u_t and y_t making them irrelevant. The recursion from $t+h$

Autonomous:

$$\omega(u_{i,\tau},d^{1:\tau-1}) = E\Big[\ln\frac{f(y_{i\rceil,\tau}|y_{\cap,\tau},u_{i,\tau},d_{[i]}^{1:\tau-1})}{g(y_{i\rceil\tau}|y_{\cap,\tau},u_{i,\tau},d_{[i]}^{1:\tau-1})} + \ln\frac{f(y_{\cap,\tau}|u_{i,t},d_{[i]}^{1:\tau-1})}{g(y_{\cap,\tau}|u_{i,t},d_{[i]}^{1:\tau-1})} - \ln(\gamma(y_{[i],\tau},d_{[i]}^{1:\tau-1}))\Big| y_{[i]\tau}\Big]. \quad (15)$$

Global:

$$\omega(u_\tau,d^{1:\tau-1}) = E\Big[\ln\frac{f(y_{j\rceil,\tau}|y_{i\rceil,\tau}y_{\cap,\tau},u_\tau,d^{1:\tau-1})}{g(y_{j\rceil\tau}|y_{i\rceil,\tau},y_{\cap,\tau},u_\tau,d^{1:\tau-1})} + \ln\frac{f(y_{i\rceil,\tau}|y_{\cap,\tau},u_\tau,d^{1:\tau-1})}{g(y_{i\rceil\tau}|y_{\cap,\tau},u_\tau,d^{1:\tau-1})} + \ln\frac{f(y_{\cap,\tau}|u_\tau,d^{1:\tau-1})}{g(y_{\cap,\tau}|u_\tau,d^{1:\tau-1})} - \ln(\gamma(y_\tau,d^{1:\tau-1}))\Big| y_\tau\Big]. \quad (16)$$

Cooperative:

$$\omega(u_{i,\tau},d^{1:\tau-1}) = E\Big[\ln\frac{f(y_{i\rceil,\tau}|y_{\cap,\tau},u_{i,\tau},d_{[i]}^{1:\tau-1})}{g(y_{i\rceil\tau}|y_{\cap,\tau},u_{i,\tau},d_{[i]}^{1:\tau-1})} + \ln\frac{f(y_{\cap,\tau}|u_{i,\tau},d_{[i]}^{1:\tau-1})}{g^{\alpha}(y_{\cap i,\tau}|u_{i,\tau},d_{[i]}^{1:\tau-1})f^{1-\beta}(y_{\cap,\tau}|u_{j,\tau},d_{[j]}^{1:\tau-1})} - \ln(\gamma(y_{[i],\tau},d_{[i]}^{1:\tau-1}))\Big| y_{[i]\tau}\Big]. \quad (17)$$

to t yields reveals that $\overline{u}_t$ is equivalent to LQG designed strategy with loss function given by the quadratic form in exp of (13), and $\ln\gamma()$ in the role of the Bellman function Kárný (1996).

3. EXTENSION TO DECENTRALIZED STOCHASTIC CONTROL

In decentralized control, the global control task is split between agents where each agent is assigned to control only a sub-set of all considered inputs using only locally available data and models. For simplicity, we will consider only two agents, A_1 and A_2. Data space of the ith agent, $d_{[i],t}$, is divided into its 'private' variables $d_{i\rceil,t}$ and commonly available variables $d_{\cap,t}$. The full data set is then $d_t = [d_{1\rceil t}, d_{\cap t}, d_{2\rceil t}]$. Input spaces of both agents are non-overlapping, $u_t = [u_{1,t}, u_{2,t}]$.

The task is to design for each agent: (i) an autonomous control strategy, and (ii) a cooperative cooperative strategy that outperforms the autonomous one in terms of global loss. Moreover, the cooperative control strategy must remain of the same complexity as the autonomous one. We will address the task as follows:

(1) Autonomous control strategy design for both agents and the global control strategy is developed.
(2) Each local autonomous strategy is interpreted as an approximation within the global strategy.
(3) The global control strategy is restructured such that application of the same approximation as in step 2. yields cooperative strategy.
(4) Performance of the new cooperative strategy is studied in simulation.

3.1 Autonomous regime

In this paper, we consider only agents, A_1 and A_2, that model only variables they explicitly observe, i.e. $y_{[1]}$ and $y_{[2]}$, respectively. The reason for this is twofold: first, by considering some variables of the neighbors in its model, the agent would significantly increase the complexity of its models and the imposed computational load; second, when reconfiguration is allowed, the structure of the neighborhood would change and it would be necessary to rebuild the internal model.

We assume that each agent has its model and is able to construct its own predictors:

$$f(y_{[i]}^{(h)},u_i^{(h)}|d^{1:t}) = \prod_{\tau=t+1}^{t+h} f(y_{[i],\tau},u_{i,\tau}|d_i^{1:\tau-1}),$$
$$= \prod_{\tau=t+1}^{t+h} f(y_{i\rceil,\tau}|y_{\cap,\tau},u_{i,\tau},d_{[i]}^{1:\tau-1}) f_i(y_{\cap,\tau}|u_{i,\tau},d_{[i]}^{1:\tau-1}) f(u_{i,\tau}|d^{1:t-1}). \quad (18)$$

Each agent is supposed to have its local aims

$$g(y_{[i]}^{(h)},u_i^{(h)}|d^{1:\tau}) = \prod_{\tau=t+1}^{t+h} g(y_{i\rceil,\tau}|y_{\cap,\tau},u_{i,\tau},d_i^{1:\tau-1}) g(y_{\cap,\tau}|u_{i,\tau},d_i^{1:\tau-1}) g(u_{i,\tau}|d_i^{1:\tau-1}). \quad (19)$$

The autonomous control strategy is be designed using (8)–(11), where $\omega()$ for this case is displayed in (15).

3.2 Centralized control

A hypothetical agent $\overline{A}$ constructs its global predictor, designs global control strategy using (16) and its subsequent integration via (10). The result is a globally optimal strategy. This strategy can not be designed under the incomplete information structure outlined in the introduction and must be approximated.

The autonomous strategy is can be interpreted as an approximation of the global design in (16). Note that the autonomous control arise when all densities model ling private data of a neighbor are dropped, and predictors that are conditioned on the data of the neighbor are approximated as follows:

$$\ln\frac{f(y_{i\rceil,\tau}|y_{\cap,\tau},u_\tau,d^{1:\tau-1})}{g(y_{i\rceil\tau}|y_{\cap,\tau},u_\tau,d^{1:\tau-1})} \approx \ln\frac{f(y_{i\rceil,\tau}|y_{\cap,\tau},u_{i,\tau},d_{[i]}^{1:\tau-1})}{g(y_{i\rceil\tau}|y_{\cap,\tau},u_{i,\tau},d_{[i]}^{1:\tau-1})}. \quad (20)$$

The same approximation arise for the jth agent under the use of different order of the chain rule in (16). The parallel run of both autonomous strategies may result in poor performance since predictors in each strategy may yield different predictions. One way to achieve cooperative behavior is resolution of the conflict in predictions. A solution offered in Šmídl and Andrýsek (2008) is to harmonize the incompatible predictors using tools of probability combination, Genest and Zidek (1986).

3.3 Decentralized control

Note that by running two autonomous strategies in parallel, the common part of the loss, $\ln f(y_\cap) - \ln g(y_\cap)$ in (16), is optimized twice. A heuristic solution would be to split this element in two parts, where each agent optimizes only one of them. This solution may be extended further as follows:

$$\ln \frac{f(y_\cap|d)}{g(y_\cap|d)} = \ln \frac{f(y_\cap|d)f(y_\cap|d)}{g(y_\cap|d)f(y_\cap|d)}$$
$$= \ln \frac{f(y_\cap|d)f(y_\cap|d)}{g^{\alpha}(y_\cap|d)g^{1-\alpha}(y_\cap|d)f^{1-\beta}(y_\cap|d)f^{\beta}(y_\cap|d)} \tag{21}$$

Here, only symbolic condition d was used in place of proper conditions in (16) for brevity. Similarly to the autonomous projection, each element in (21) can be replaced by an approximation of the kind of (20). Here, we make more complex substitutions:

$$\frac{f(y_{\cap,\tau}|u_\tau, d^{1:\tau-1})}{g(y_{\cap,\tau}|u_\tau, d^{1:\tau-1})} \approx$$
$$\approx \ln \frac{f(y_{\cap,\tau}|u_{i,\tau}, d_{[i]}^{1:\tau-1})}{g^{\alpha}(y_{\cap,\tau}|u_{i,\tau}, d_{[i]}^{1:\tau-1})f^{\beta}(y_{\cap,\tau}|u_{j,\tau}, d_{[j]}^{1:\tau-1})}$$
$$+ \ln \frac{f(y_\cap|u_{j,\tau}, d_{[j]}^{1:\tau-1})}{g^{1-\alpha}(y_\cap|u_{j,\tau}, d_{[j]}^{1:\tau-1})f^{1-\beta}(y_\cap|u_{i,\tau}, d_{[i]}^{1:\tau-1})} \tag{22}$$

By a heuristic choice, the ith agent approximate only the first part, yielding (17). The density $f(y_{\cap,\tau}|u_{j,\tau}, d_{[j]}^{1:\tau-1})$ is understood as a fixed quantity obtained from the jth agent by a communication channel. Equation (22) may reveal different scenarios, based on the value of tuning knobs $0 \le \alpha \le 1$ and $0 \le \beta \le 1$.

We note the following:

- Structure of dependence is assigned by values of α, β. For example, for $\alpha = 1, \beta = 0$, the ith agent designs its own autonomous strategy, while the jth agent optimizes its strategy with ith predictors as its target. Their roles are opposite for $\alpha = 0, \beta = 1$.
- For non-zero values of both α and β, equations (17) for i and j form a set of implicit equations. If $f(y_{\cap i}|\cdot)$ is known, $f(y_{\cap j}|\cdot)$ is uniquely determined and vice versa. Therefore a solution can be found via successive approximations yielding a communication mechanism that correspond to negotiation in multi-agent systems.
- When values of α and β are globally assigned, the loss associated with the cooperative solutions should approach that of the global solution. However, we may consider also scenarios, when the agents have freedom to choose α, β to their liking. Indeed, the autonomous strategy (15) correspond to one such choice.

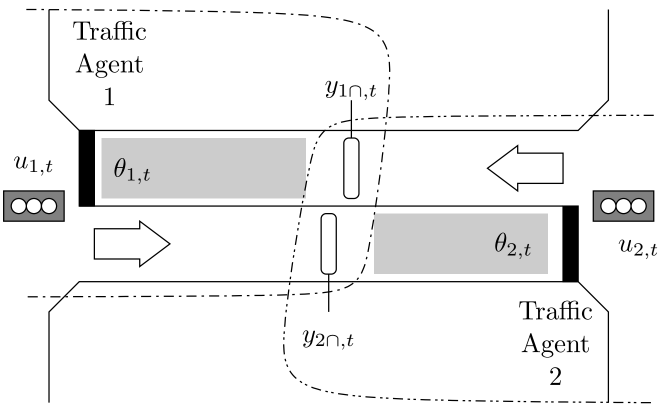

Figure 1. Two traffic intersections with vector of overlapping observations, $y_\cap = [y_{1\cap}, y_{2\cap}]$. Inputs u_1, u_2 are proportions of the green light signal in the control period. Internal variables θ_1, θ_2 are lengths of the queues of waiting cars. These queues are not directly observed and must be estimated.

- Extending agents freedom further, we can consider incompatible targets $g_i(y_\cap) \neq g_j(y_\cap)$. In that case, cooperation may be achieved using the same mechanism as in (21), yielding

$$\ln \frac{f(y_\cap)}{g(y_\cap)} = \ln \frac{f(y_{\cap,i})}{g^{\alpha\varphi}(y_{\cap i})g^{\alpha(1-\varphi)}(y_{\cap j})f^{\beta}(y_{\cap,j})} + \tag{23}$$
$$+ \ln \frac{f(y_{\cap,j})}{g^{(1-\alpha)\varphi}(y_{\cap i})g^{(1-\alpha)(1-\varphi)}(y_{\cap j})f^{1-\beta}(y_{\cap,i})}$$

- Note that the key operation in this scheme is probability combination technique known as geometric combination, Genest and Zidek (1986). Specifically, solution (23) also arise when the target densities are harmonized using geometric combination. However, extending agent's freedom to assign their own φ will depart from that solution.
- Extension to more than two agents will result in geometric combination of more densities.

4. ILLUSTRATIVE EXAMPLE

Consider the following 3-output 2-input system:

$$f(y_t|\psi_t, \Sigma) = \mathcal{N}(\theta\psi, \Sigma), \tag{24}$$

where

$$y_t = [y_{1,t}, y_{2,t}, y_{3,t}]',$$
$$\psi_t = [y_{1,t-1}, y_{2,t-1}, y_{3,t-1}, u_{1,t}, u_{1,t-1}, u_{2,t}, u_{2,t-1}]',$$
$$\theta = \begin{bmatrix} 0.8 & 0.2 & 0 & -0.3 & 0.4 & 0 & 0 \\ -0.2 & 0.5 & -0.8 & 0.2 & 0.5 & -0.2 & -0.5 \\ 0 & 1.1 & -0.5 & 0 & 0 & -0.2 & 0.3 \end{bmatrix}. \tag{25}$$

The target density is stationary

$$g(y_t)g(u_t) =$$
$$\mathcal{N}\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.1 & & \\ & 0.1 & \\ & & 0.1 \end{bmatrix}\right)\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.1 & \\ & 0.1 \end{bmatrix}\right), \tag{26}$$

corresponding to a standard quadratic loss of LQG (14), with the choice of penalization matrices $10I_3$ and $10I_2$ for y_t and u_t, respectively. Here, I_n denotes identity matrix of dimension n. This example is a greatly simplified model of two intersections with overlapping variables, see Figure 1.

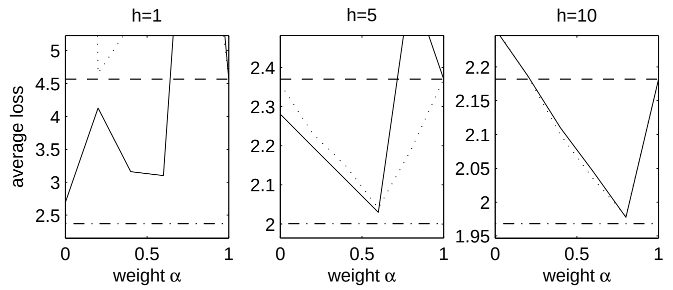

Figure 2. Average loss achieved by different control design approaches for 30 runs of *non-minimum* phase system (24),(25) with $\alpha = 1-\beta$. Strategies are: **Centralized** (dashed-dotted), **Autonomous** (dashed), **Cooperative** with 3 iterations (dotted), **Cooperative** with 10 iterations (full).

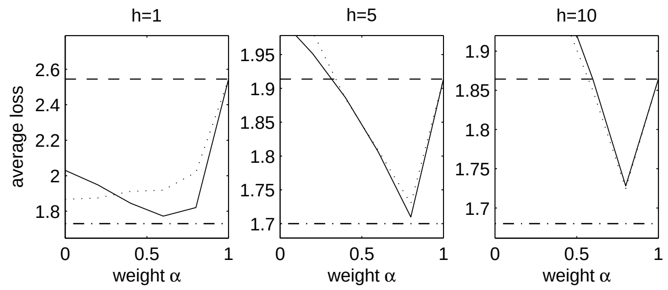

Figure 3. Average loss achieved by different control design approaches for 30 runs of *minimum* phase system (24),(28) with $\alpha = 1-\beta$. Strategies are: **Centralized** (dashed-dotted), **Autonomous** (dashed), **Cooperative** with 3 iterations (dotted), **Cooperative** with 10 iterations (full).

Both centralized and decentralized control strategies were tested with agents whose internal model is first-order linear auto-regressive model estimated using Bayesian method, Peterka (1981). Three different studies were performed: (i) centralized approach: single agent with full model (25) whose control strategy was designed with loss (26), yielding standard LQG solution with receding horizon, (ii) autonomous approach: the systems was estimated using two agents: A_1 observing variables y_1, y_2, u_1 and A_2 observing y_2, y_3, u_2, each modeling its own variables with corresponding first-order auto-regressive model followed by LQG synthesis with loss function 26 being marginal on the modeled variables, and (iii) cooperative approach: the same agents as in (ii) with cooperative design of the control strategy (17).

In order to achieve comparable complexity to the autonomous solution, where the predictive density is unconditional (26), we have projected the multi-step predictor (3) into the same family, i.e. into the product of unconditional normal densities. Then, the geometric combination in (17) is analytically tractable. For example, for the first agent it holds:

$$g(y_{1,\tau}, y_{2,\tau}) = \mathcal{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}q_1 & \\ & q_2\end{bmatrix}\right), f(y_{2,\tau}) = \mathcal{N}(\overline{\mu}_\tau, \overline{q}_\tau),$$

$$g^{\alpha}(y_{1,\tau}, y_{2,\tau}) f^{\beta}(y_{1,\tau}) = \\ = \mathcal{N}\left(\begin{bmatrix}\mu_1\\ \dfrac{\alpha\overline{q}\mu_2 + \beta q_2 \overline{q}}{\alpha\overline{q} + \beta q_2}\end{bmatrix}, \begin{bmatrix}q_1 & \\ & \dfrac{q_2\overline{q}}{\alpha\overline{q} + \beta q_2}\end{bmatrix}\right) \quad (27)$$

where $\overline{\mu}_\tau$, $\overline{q}_\tau, \tau = t+1, \ldots t+h$ is a sequence of mean values and variances of the communicated predictor, and μ_1, μ_2, q_1, q_2 were chosen to be $0, 1, 0.1, 0.1$, respectively, in (26).

A simulation study was run for horizons $h = [1, 5, 10]$ with six settings of $\alpha = 1-\beta = [0, 0.2, \ldots, 1]$. Results of a Monte Carlo simulation are displayed in Figure 2. As expected, for $\alpha = 1$, the loss associated with the cooperative control are identical to that of the autonomous control. Note that for horizon $h = 10$, the cooperative strategy improves the overall loss for $\alpha > 0.2$. For horizon $h = 1$ and some values of α at $h = 3$, the cooperative control does not improve over the autonomous. We conjecture that this is related to

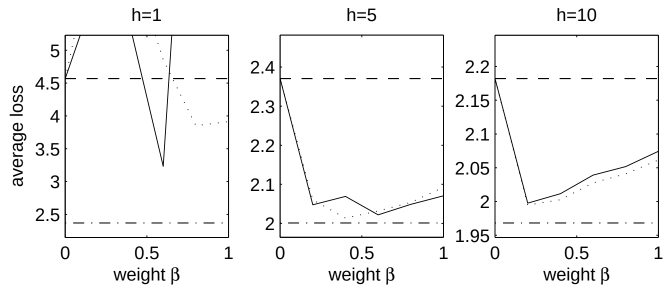

Figure 4. Average loss achieved by different control design approaches for 30 runs of *non-minimum* phase system (24),(25) with $\alpha = 1$, $0 \leq \beta \leq 1$. Strategies are: **Centralized** (dashed-dotted), **Autonomous** (dashed), **Cooperative** with 3 iterations (dotted), **Cooperative** with 10 iterations (full).

the fact that the system (25) is non-minimum phase. For confirmation we have run the same experiment with the following minimum-phase system,

$$\theta = \begin{bmatrix} 0.7 & 0.2 & 0 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0.5 & -0.8 & 0.2 & 0.5 & 0 & 1.0 & 1.0 \\ 0 & 0 & 0 & 0 & 0.7 & 0.2 & 0 & 1.0 \end{bmatrix}, \quad (28)$$

$$\psi_t = [y_{1,t-2}, y_{1,t-3}, y_{2,t-2}, y_{2,t-3}, y_{3,t-2}, y_{3,t-3}, u_{1,t}, u_{2,t}],$$

and the results are displayed in Figure 3. Unexpectedly, the performance is very sensitive to the choice of $\alpha = 1-\beta$.

Note that in (27), the coefficient α has the same role as coefficient q_1 of the local loss. Also, by simultaneous change of the both parameters α and β it is not clear which parameter is affecting the result. Thus, the same Monte Carlo study for model (24) was performed with setting $\alpha = 1$, $\beta = [0, 0.2, \ldots, 1]$ and its results are displayed in Figure 4. Once again, the performance for $h = 1$ is deteriorating for this non-minimum phase system, the same experiment for minimum phase does not suffer from this instability. Note that in this case, the best performance is achieved for low values of β. We conjecture that this is due to the fact that all αs and βs in both agents do not sum to 1.

5. DISCUSSION

The general probabilistic solution (17) was specialized for Gaussian densities. This experiment serves as a proof-of-concept experiment to check feasibility of this approach.

Application to more complex models would require far more work, but the presented approach already demonstrate the following intuitively interesting properties:

(1) The fully probabilistic formulation unifies parametrization of predictors and losses which allows to modify the loss function of the receiving agent by predictor of its neighbor.
(2) The tuning coefficients α and β have rather clear meaning: α is a weight of 'selfish loss' of an agent, and β is a weight of the 'trust' in the neighbor.

The first property offers an interesting alternative for optimization, where new heuristics can be designed to mimic this behavior if different optimization methods are used. The second property allows for tuning of the control strategy when needed, however, it is expected that a default value, say $\beta = 0.2$ for the setup in Figure 4 would be sufficient in most cases.

In situations where computation of predictors for the whole optimization horizon is computationally prohibitive, the method can easily handle predictors on shorter horizons, or even one-step-ahead predictors. This may allow engineers to develop experimental knowledge what needs to be done similar to that developed for LQG control, e.g. Böhm et al. (1989).

The current implementation of traffic control, Homolová and Nagy (2005), is using non-linear state-space model for each intersection, and linear programming for optimization of the control strategy. Application of the presented approach is possible by changing the coefficients of the loss function, however, in that case the uncertainty in the predictions would be neglected. Loss functions which allow incorporation of higher order moment of the predictors are not linear. Nevertheless, they can be handled by general convex optimization methods.

6. CONCLUSION

An approach extending the FPD method of autonomous control strategy design to cooperative control strategy design was presented. The approach was derived by approximation of the global optimal control strategy design, however, it still should be considered as a heuristics. As demonstrated in Section 4, performance of the presented cooperative control of a non-minimum phase system on a short horizon may be worse than that the autonomous control. However, the approach has interesting properties that make it worth of further research. Namely, the cooperative control strategy design has indeed only a minor computational overhead over the autonomous one and allows to make further simplifications. It provides only a minimum of tuning knobs with reasonable defaults.

Further research is necessary to improve stability of the solution—e.g. similar to back-stepping line-search used in approximate DP (Todorov and Tassa (2009))—and to investigate convergence and stability of the approach.

Acknowledgement

Support of grants MŠMT 1M0572 and GAČR 102/08/P250 is gratefully acknowledged.

REFERENCES

Bertsekas, D. (2001). *Dynamic Programming and Optimal Control*. Athena Scientific, Nashua, US. 2nd edition.

Böhm, J., Kárný, M., and Kulhavý, R. (1989). Practically-oriented LQ self-tuners. In *Preprints of the IFAC Workshop on Evaluation of Adaptive Control Strategies in Industrial Applications*. Tbilisi.

Doucet, A., de Freitas, N., and Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Springer.

Genest, C. and Zidek, J.V. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1, 114–148.

Homolová, J. and Nagy, I. (2005). Traffic model of a microregion. In P. Horáček, M. Šimandl, and P. Zítek (eds.), *Preprints of the 16th World Congress of the International Federation of Automatic Control*, 1–6. IFAC, Prague.

Ioannou, P. (Apr 1986). Decentralized adaptive control of interconnected systems. *Automatic Control, IEEE Transactions on*, 31(4), 291–298.

Kárný, M. (1996). Towards fully probabilistic control design. *Automatica*, 32(12), 1719–1722.

Kárný, M. and Guy, T. (2004). On dynamic decision-making scenarios with multiple participants. In J. Andrýsek, M. Kárný, and J. Kracík (eds.), *Multiple Participant Decision Making*, 17–28. Advanced Knowledge International, Adelaide.

Kracík, J. (2009). *Cooperation Methods in Bayesian Decision Making with Multiple Participants*. Ph.D. thesis, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague.

Liu, S., Zhang, J., and Jiang, Z. (2007). Decentralized adaptive output-feedback stabilization for large-scale stochastic nonlinear systems. *Automatica*, 43(2), 238–251.

Peterka, V. (1981). Bayesian approach to system identification. In P. Eykhoff (ed.), *Trends and Progress in System identification*, 239–304. Pergamon Press, Oxford.

Roozemond, D. (2001). Using intelligent agents for proactive, real-time urban intersection control. *European Journal of Operational Research*, 131(2), 293–301.

Šmídl, V. and Andrýsek, J. (2008). Merging of multistep predictors for decentralized adaptive control. In *Proceedings of the American Control Conference*. Seattle, USA.

Šmídl, V. and Přikryl, J. (2006). Distributed Bayesian decision-making for urban traffic control. In *Proceedings of IEEE IECON'06*, 4695–4700. IEEE, Paris, France.

Todorov, E. and Tassa, Y. (2009). Iterative local dynamic programming. In *Proceedings of the 2nd IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, 90 – 95.

van den Broek, B., Wiegerinck, W., and Kappen, B. (2008). Graphical model inference in optimal control of stochastic multi-agent systems. *Journal of Artificial Intelligence Research*, 32(1), 95–122.