Lazy Fully Probabilistic Design of Decision Strategies

Miroslav Kárný, Karel Macek, and Tatiana V. Guy

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
P.O.Box 18, Prague 8 182 08, Czech Republic
{school,macek,guy}@utia.cas.cz

Abstract. Fully probabilistic design of decision strategies (FPD) extends Bayesian dynamic decision making. The FPD specifies the decision aim via so-called ideal - a probability density, which assigns high probability values to the desirable behaviours and low values to undesirable ones. The optimal decision strategy minimises the Kullback-Leibler divergence of the probability density describing the closed-loop behaviour to this ideal. In spite of the availability of explicit minimisers in the corresponding dynamic programming, it suffers from the curse of dimensionality connected with complexity of the value function. Recently proposed a lazy FPD tailors lazy learning, which builds a local model around the current behaviour, to estimation of the closed-loop model with the optimal strategy. This paper adds a theoretical support to the lazy FPD and outlines its further improvement.

Keywords: decision making, lazy learning, Bayesian learning, local model.

1 Introduction

A decision maker (artificial or human) forms with its environment a closed decision-making (DM) loop and aims to influence the closed-loop behaviour by a sequence of its actions. The behaviour is characterised by a collection of observed, selected and considered variables. The decision maker can only use incomplete knowledge and faces random dynamic changes of the environment. DM understood in this way is wide spread and covers stochastic and adaptive control, fault detection as well as inference tasks like estimation, filtering, prediction, classification, etc. The mentioned DM importance and width have naturally stimulated a search for widely applicable normative DM theories. A long-term development has singled out the Bayesian DM theory [3, 4, 9, 29] as the most promising candidate.

The Bayesian DM provides well-justified solutions of DM tasks but the "curse of dimensionality" [1] limits its applicability and approximations are mostly inevitable. Approximate non-linear estimation and filtering [7, 8, 27, 30] and approximate dynamic programming [4, 31, 34] are thus unavoidable, permanently-evolving, complements of the basic DM theory. Practically successful techniques

mostly rely on local approximations around the current realisation of the behaviour. This applies to learning, with lazy learning being its typical representative [5, 20], to adaptive control [13, 23] and other techniques like case-based reasoning [10]. Their success or failure strongly depends on a proper specification of the neighbourhood of the current behaviour. The neighbourhood must be narrow to allow a simple and rich modelling containing relevant information learnt within the closed DM loop. To our best knowledge, no established methodology comparable in the width to the underlying DM theory exists. Mostly they either support a subset of DM problems or use a trial and error method.
Fully probabilistic design (FPD) of DM strategies is an extension of the Bayesian DM [12, 16, 17][1] describes the closed-loop behaviour by a joint probability density (pd) of the involved variables, exactly as the Bayesian DM does. It, however, expresses the DM aims via a decision-maker-adopted ideal pd quantifying desirability of behaviours. The strategy design then reduces to a minimisation of the Kullback-Leibler divergence (KLD, [19]) of the involved pds over the optional strategies. The FPD promises simpler approximations of the unfeasible strategy design as it provides an explicit minimiser in dynamic programming. The rare attempts, e.g. [14], only partially exploited the potential offered by this feature. They are still too much of ad hoc nature and cumbersome. A substantial progress towards an approximate FPD has been recently made, [22]. The proposed *lazy FPD* uses the current ideal pd for weighting the past data records when learning a local model of the optimally closed loop. This treatment overcomes weaknesses of the lazy learning, which: a) serves well to prediction but rarely to dynamic DM; b) is sensitive to a measure quantifying the proximity of behaviours, and c) relies on availability of data records close enough to the current one. The present paper adds a theoretical insight into the technique and improves the lazy FPD using Sanov-type analysis [28]. Section 2 recalls the lazy FPD and Section 3 formulates the addressed problems. Section 4 solves them. Section 5 contains illustrative example and Section 6 offers concluding remarks.

Throughout, $\boldsymbol{x}$ is a set of x-values; all sets are subsets of finite-dimensional spaces; $\mathsf{S}, \mathsf{O}, \ldots$ are mappings; $x \in \boldsymbol{x}$ denotes a possible realisation of a random variable X; $\underline{x} \in \boldsymbol{x}$ is a specific realisation of X; probability density (pd) is Radon-Nikodým derivative with respect to a measure $\mathrm{d}\bullet$; pds having different identifiers in arguments are taken as different; $\tau, t \in \boldsymbol{t} \equiv \{1, \ldots, T\}$ label discrete time; $x_m^n = (x_t)_{t=m}^n$ and $x^n = x_0^n$ describe finite sequences.

2 Lazy FPD

The *lazy FPD* selects a decision strategy, which makes a probabilistic description of the closed decision loop close to a pre-specified closed-loop ideal. Instead of the traditional learning of an environment model followed by the strategy optimisation, the lazy FPD uses the currently observed data to estimate, which of simple parametric models provides the closed-loop model near the given ideal.

[1] Re-invented in [33], studied in control [11] and used in brain research [32].

The designed strategy is then a marginal of the found closed-loop model. The next text formalises this.

The inspected DM problem deals with sequences of possible realisations x^T of random environment responses $x_\tau \in \boldsymbol{x_\tau}$, $t \in \boldsymbol{t}$. The realised sequence of responses $\underline{x}^T$ reacts on the realisation $\underline{a}^T$ of actions generated by a randomised strategy, $\mathsf{S}_\tau : \underline{a}^{\tau-1}, \underline{x}^{\tau-1} \rightarrow \underline{a}_\tau$, $\tau \in \boldsymbol{t}$. The action, a_τ, and the environment response, x_τ, forms the data, d_τ, observable at time $\tau \in \boldsymbol{t}$. Pds $\mathsf{S}^T \equiv (\mathsf{S}_t(a_t|a^{t-1}, x^{t-1}))_{\tau=1}^T = (\mathsf{S}_\tau(a_\tau|d^{\tau-1}))_{\tau=1}^T$ describe the strategy. The individual pds in the sequence S^T are decision rules forming the strategy.

Let us consider the current time $t \in \boldsymbol{t}$ delimits the past (when data sequence $\underline{d}^{t-1}$ was observed) and the future, which includes the current inspected DM stage. The current time splits behaviour and all involved pds in their past and future parts. The data considered in the closed DM loop are samples from a closed-loop-describing pd $\mathsf{C}^T = \prod_{\tau=1}^T \mathsf{C}_\tau(d_\tau|d^{\tau-1})$. In the inspected stage, the past and the future closed-loop models are distinguished. The future ideal closed-loop model, given by the joint pd

$$\mathsf{I}_t^T = \mathsf{I}_t\big(d_t^T|\underline{d}^{t-1}\big) = \prod_{\tau=t}^T \mathsf{I}_t(d_\tau|\underline{d}^{t-1}, d_t^{\tau-1}),\ t \in \boldsymbol{t}, \tag{1}$$

quantifies the DM aim. Its factors $\mathsf{I}_t(d_\tau|d^{\tau-1})$ for $\tau \geq t$ may differ from the past ideal factors I_τ for $\tau < t$. Notice that the behaviour evolution within the planning periods starts at the realised $\underline{d}^{t-1}$. The future closed-loop model $\mathsf{C}_t^T = \mathsf{C}_t\big(d_t^T|\underline{d}^{t-1}\big)$ describes the DM loop formed by the environment and the future strategy S_t^T. The strategy making C_t^T close to the future ideal pd $\mathsf{I}_t^T = \mathsf{I}_t(d_t^T|\underline{d}^{t-1})$ (1) is searched for. The lazy FPD uses: i) the observed data realisations $\underline{d}^{t-1}$; ii) the given ideal pd (1); iii) a class of parametric models

$$\mathsf{M}_t\big(d_t^T|\underline{d}^{t-1}, \theta\big) = \prod_{\tau=t}^T \mathsf{M}_t(d_\tau|\underline{d}^{t-1}, d_t^{\tau-1}, \theta),\ \theta \in \boldsymbol{\theta}, \tag{2}$$

serving for extrapolation of the past realised closed-loop behaviour $\underline{d}^{t-1}$. Note that the parametric closed-loop models (2) can be simple as the future closed-loop model $\mathsf{C}_t(d_t^T|\underline{d}^{t-1})$ has to be (approximately) valid only for the behaviours prolonging the past $\underline{d}^{t-1}$.

Design concept of the lazy FPD: The lazy FPD uses the data realisation for assigning such a posterior pd $\mathsf{P}(\theta|\underline{d}^{t-1})$ to respective parameters $\theta \in \boldsymbol{\theta}$ in (2) so that the model $\mathsf{C}_t(d_t^T|\underline{d}^{t-1}) = \prod_{\tau=t}^T \mathsf{C}_t(d_\tau|d^{\tau-1})$ describes the closed loop with the desired strategy. Its future-describing factors are predictors

$$\mathsf{C}_t(d_\tau|\underline{d}^{t-1}, d_t^{\tau-1}) \equiv \int_{\boldsymbol{\theta}} \mathsf{M}_t(d_\tau|\underline{d}^{t-1}, d_t^{\tau-1}, \theta)\mathsf{P}(\theta|\underline{d}^{t-1})\ \mathrm{d}\theta \tag{3}$$

constructed from the parametric model (2) and the posterior pd $\mathsf{P}(\theta|\underline{d}^{t-1})$. The pd $\mathsf{S}_t\big(a_t|\underline{d}^{t-1}\big) = \int_{\boldsymbol{x}} \mathsf{C}_t\big(d_t|\underline{d}^{t-1}\big)\ \mathrm{d}x_t$ gained from the predictor (3) is the current estimate of the properly tuned decision rule. The action $\underline{a}_t$ is sampled from it and the response $\underline{x}_t$ is observed.

The randomised strategy arising from the lazy FPD cares about the exploration conditioning any successful learning. For a well-peaked $\mathsf{P}\big(\theta|\underline{d}^{t-1}\big)$, the predictors (3) can be approximated by plug in a point estimate of θ into the models $\mathsf{M}_t\big(d_\tau|d^{\tau-1},\theta\big)$ (2).

Neither the local model of the environment working in the closed loop nor the future strategy optimal with respect to the future ideal are known. Thus, the parameters $\theta \in \boldsymbol{\theta}$ pointing to the models (2), which guarantee the closeness of the future closed-loop model (3) to the given future ideal pd (1), are unknown. As such, they should be learned in the Bayesian way. The already observed data realisations $\underline{d}^{t-1}$, however, do not origin from the closed loop tuned with respect to the ideal pd I_t^T (1). The lazy FPD faces this serious obstacle by learning the unknown parameter $\theta \in \boldsymbol{\theta}$ via the weighted Bayes rule. It maps a prior pd $\mathsf{P}(\theta)$ on the posterior pd, $\forall \theta \in \boldsymbol{\theta}$, as follows

$$\mathsf{P}\big(\theta|\underline{d}^{t-1}\big) \propto \mathsf{P}(\theta) \prod_{\tau=1}^{t-1} \mathsf{M}_t^{\mathsf{W}_t(\underline{d}^\tau)}(\underline{d}_\tau|\underline{d}^{\tau-1},\theta) \tag{4}$$

$$\mathsf{W}_t\big(\underline{d}^\tau\big) \propto \mathsf{I}_t(\underline{d}_\tau|\underline{d}^{\tau-1}) \quad \text{and} \quad \propto \text{ denotes proportionality.}$$

After using $\underline{a}_t$ taken from $\mathsf{S}_t(a_t|\underline{d}^{t-1}) = \int_{\boldsymbol{x_t}} \int_{\boldsymbol{\theta}} \mathsf{M}_t(d_t|\underline{d}^{t-1})\mathsf{P}(\theta|\underline{d}^{t-1})\, \mathrm{d}\theta\, \mathrm{d}x_t$ the response $\underline{x}_t$ is observed and the learning step (4) is repeated for time $t+1$.

3 Questions Connected with the Lazy FPD

The weights $\mathsf{W}_t(\underline{d}^\tau)$ chosen in (4) are intuitively plausible. The weight is the higher the more the realised subsequence $\underline{d}^\tau$ fits the ideal factor $\mathsf{I}_t(d_\tau|d^{\tau-1})$ to which closed-loop models (2) with highly probable parameter values should approach. Promising experimental results, partially reported in [22], support this intuition.

The intuition leaves aside the natural questions: i) Is the *use* of the weights W_t in (4) the proper and, ideally, only one? ii) How to *normalise* the weights (4) to get the adequately peaked posterior pd $\mathsf{P}(\theta|\underline{d}^{t-1})$? iii) What happens if the processed data realisations indeed come from the properly tuned closed loop describable by the parametric model (2), i.e. what is the asymptotic behaviour under time-invariant circumstances?

The formal inspection of the weighted Bayes rule (4) with a novel choice of the weights presented in the next section forms the paper core and answers the questions above.

4 Answers to the Formulated Questions

The following normalisation of the weights (4) is inspected

$$\mathsf{W}_t(d^\tau) = \frac{\mathsf{I}_t(d_\tau|d^{\tau-1})}{\mathsf{C}_\tau(d_\tau|d^{\tau-1})}, \tag{5}$$

where $\mathsf{C}_\tau(d_\tau|d^{\tau-1})$ is the pd describing the realisations of the closed-loop behaviour for $\tau \leq t-1$. It can be obtained via the standard Bayesian learning using either a specific parametric model or the model (2). The latter option needs a sort of forgetting [18] coping with the approximate nature of the simple models (2), [15]. For a time-invariant ideal pd, it can alternatively be approximated by the predictors (3) obtained when the planning started at times $\tau < t$.

The normalisation (5) has resulted from the Sanov-type analysis [28] of the posterior pd. It is extended here so that it is applicable to the posterior pd obtained in the closed DM loop with the weighted Bayes rule (4). Its idea is often masked by the focus on difficult but technical problems. The common essence is, however, simple. The posterior pd is re-written as

$$\mathsf{P}(\theta|\underline{d}^{t-1}) \propto \exp[-(t-1) \times \text{sample mean of a data function depending on } \theta]$$

and a law of large number, ergodic arguments or martingale theory [21] are used to show that this sample mean converges to a function bounded from below. Then, it is easy to see that the posterior pd $\mathsf{P}(\theta|\underline{d}^{t-1})$ may concentrate only on $\theta \in \boldsymbol{\theta}$ minimising this function.

The next proposition formalises this way assuming that the time moment $t \in \boldsymbol{t}$ is fixed and the past data d^{t-1} are described by the pd $\prod_{\tau=1}^{t-1} \mathsf{C}_\tau(d_\tau|d^{\tau-1})$.

Proposition 1 (On the Weighted Bayesian Learning) *Let*

$$\ln\left(\frac{\mathsf{I}_t(d_\tau|d^{\tau-1})}{\mathsf{M}_t(d_\tau|d^{\tau-1},\theta)}\right),\ \tau < t, \tag{6}$$

be essentially bounded for all $\theta \in \boldsymbol{\theta}$. Then, the weighted Bayes rule (4) using the weights (5) provides for $t \to \infty$ the same posterior pd as that obtained by the standard Bayes rule applied to data sampled from the closed-loop described by the ideal pd $\mathsf{I}_t(d_\tau|d^{\tau-1})$.

Proof For any $\theta \in \boldsymbol{\theta}$, the posterior pd obtained from (4) can be given the form

$$\mathsf{P}(\theta|\underline{d}^{t-1}) \propto \mathsf{P}(\theta)\exp\Big[-(t-1)\overbrace{\frac{1}{t-1}\sum_{\tau=1}^{t-1}\underbrace{\mathsf{W}_t(\underline{d}^\tau)\ln\left(\frac{\mathsf{I}_t(\underline{d}_\tau|\underline{d}^{\tau-1})}{\mathsf{M}_t(\underline{d}_\tau|\underline{d}^{\tau-1},\theta)}\right)}_{\mathsf{L}_\tau\left(\underline{d}^\tau,\theta\right)}}^{\text{the sample mean } \Omega_{t-1}(\underline{d}^{t-1},\theta)}\Big], \tag{7}$$

exploiting the fact that the proportionality $\propto$ in (7) defines the same posterior pd even when the right-hand side is multiplied by any positive θ-independent factor. The following innovations N_τ are zero-mean, uncorrelated and essentially bounded due to the assumed bounded-ness of (6), [24],

$$\mathsf{N}_\tau(d_\tau,\underline{d}^{\tau-1},\theta) \equiv \mathsf{E}_\tau\big[\mathsf{L}_\tau|\underline{d}^{\tau-1}\big] - \mathsf{L}_\tau\big(d_\tau,\underline{d}^{\tau-1},\theta\big) \quad \text{with}$$

$$\mathsf{E}_\tau\big[\mathsf{L}_\tau|\underline{d}^{\tau-1}\big] \equiv \int_{\boldsymbol{d}_\tau} \mathsf{L}_\tau\big(d_\tau,\underline{d}^{\tau-1},\theta\big)\mathsf{C}_\tau(d_\tau|\underline{d}^{\tau-1})\,\mathrm{d}d_\tau$$

$$= \underbrace{\int_{\boldsymbol{d}_\tau} \mathsf{I}_t(d_\tau|\underline{d}^{\tau-1})\ln\Big(\frac{\mathsf{I}_t(d_\tau|\underline{d}^{\tau-1})}{\mathsf{M}_t(d_\tau|\underline{d}^{\tau-1},\theta)}\Big)\,\mathrm{d}d_\tau}_{\mathsf{H}_\tau(\underline{d}^{\tau-1},\theta)} \geq 0, \text{ due to the Jensen inequality, [26].}$$

The decomposition exists due to the essential bounded-ness of $\mathsf{H}_\tau(\underline{d}^{\tau-1},\theta)$ and splits $\Omega_{t-1}(\underline{d}^{t-1},\theta)$ into the mean of non-negative terms $\mathsf{H}_\tau(\underline{d}^{\tau-1},\theta)$ and sample average of innovations $\mathsf{N}_\tau(\underline{d}^{\tau},\theta)$, $\tau \leq t-1$, which almost surely converges for $t \to \infty$ to their zero expectation [21]. Thus, the support of the posterior pd concentrates (quickly due to the factor $-(t-1)$) on minimisers $\theta_{\mathsf{P}} \in \boldsymbol{\theta}$ of $1/(t-1)\sum_{\tau=1}^{t-1}\mathsf{H}_\tau(\underline{d}^{\tau-1},\theta)$: the weighted learning singles out the parametric models as if the data $\underline{d}^{t-1}$ was sampled from the ideally tuned closed loop described by the pd $\prod_{\tau=1}^{t-1}\mathsf{I}_t(d_\tau|d^{\tau-1})$ and processed by the usual Bayes rule [2]. □

Corollary 1 (Asymptotic Optimality of the Lazy FPD) *Let the function (6) be essentially bounded. Then, the predictor of the closed-loop behaviour (3), obtained via the weighted Bayes rule (4) with the weights (5), asymptotically almost surely fulfils the inequality, $\forall\theta \in \boldsymbol{\theta}$:*

$$\int_{\boldsymbol{d_t}} \mathsf{I}_t(d_t|\underline{d}^{t-1})\ln\Big(\frac{\mathsf{I}_t(d_t|\underline{d}^{t-1})}{\mathsf{C}_t(d_t|\underline{d}^{t-1})}\Big)\,\mathrm{d}d_t \leq \int_{\boldsymbol{d_t}} \mathsf{I}_t(d_t|\underline{d}^{t-1})\ln\Big(\frac{\mathsf{I}_t(d_t|\underline{d}^{t-1})}{\mathsf{M}_t(d_t|\underline{d}^{t-1},\theta)}\Big)\,\mathrm{d}d_t. \tag{8}$$

Proof: According to Proposition 1, the support $\boldsymbol{\theta}_{\mathsf{P}} \subset \boldsymbol{\theta}$ of $\mathsf{P}(\theta|\underline{d}_{t-1})$ asymptotically concentrates on minimisers $\theta_{\mathsf{P}} \in \boldsymbol{\theta}$ of

$$\int_{\boldsymbol{d_t}} \mathsf{I}_t(d_t|\underline{d}^{t-1})\ln\Big(\frac{\mathsf{I}_t(d_t|\underline{d}^{t-1})}{\mathsf{M}_t(d_t|\underline{d}^{t-1},\theta)}\Big)\,\mathrm{d}d_t.$$

Thus, for any $\theta_{\mathsf{P}} \in \boldsymbol{\theta}_{\mathsf{P}}$ and any $\theta \in \boldsymbol{\theta}$

$$\int_{\boldsymbol{d_t}} \mathsf{I}_t(d_t|\underline{d}^{t-1})\ln\Big(\frac{\mathsf{I}_t(d_t|\underline{d}^{t-1})}{\mathsf{M}_t(d_t|\underline{d}^{t-1},\theta_{\mathsf{P}})}\Big)\,\mathrm{d}d_t \leq \int_{\boldsymbol{d_t}} \mathsf{I}_t(d_t|\underline{d}^{t-1})\ln\Big(\frac{\mathsf{I}_t(d_t|\underline{d}^{t-1})}{\mathsf{M}_t(d_t|\underline{d}^{t-1},\theta)}\Big)\,\mathrm{d}d_t.$$

Multiplying this inequality by the posterior pd $\mathsf{P}(\theta_{\mathsf{P}}|\underline{d}^{t-1}) > 0$, integrating over its support $\boldsymbol{\theta}_{\mathsf{P}}$, using the Jensen inequality and taking into account that by definition $\mathsf{P}(\theta_{\mathsf{P}}|\underline{d}^{t-1})$ assigns unit probability to $\boldsymbol{\theta}_{\mathsf{P}}$ give the claim (8). □

Even when the function (6) is essentially bounded, the values of $\mathsf{W}_\tau(\underline{d}^{\tau})$ can be too large. Thus, it is reasonable to limit them from above by $\overline{W} \in (1,\infty)$. Corollary 1 implies that it is always possible to select such $\overline{W}$ that the limitation is almost surely inactive. Then, the asymptotic results hold even when using it.

5 Illustrative Example

The example illustrates that the proposed weighting indeed improves properties of the lazy FPD. A Markov chain with two states $x \in \boldsymbol{x} \equiv \{1,2\}$ and four actions $a \in \boldsymbol{a} \equiv \{1,2,3,4\}$ is considered. The ideal pd expressing preferability of the state value 1 was selected. The simulated system in given in Table 1. The proposed weighting (5), bounded from above by the value $\overline{W} = 3$, was compared with the standard solution (called un-normalised), which takes the weight W in (4) equal to the ideal pd $\mathsf{I}_t(\underline{x}_\tau, \underline{a}_\tau | \underline{x}_{\tau-1})$. The designed strategy is given in Table 2.

Fig. 1 provides samples of simulated closed-loop behaviour when both weighting variants were applied to the same realisation of the underlying random generator. Fig. 2 provides the corresponding time course of weights. The strategy with the proposed weighting (5) reaches the desirable state $\underline{x}_\tau = 1$ in 91% cases while 65% of units occurred when using un-normalised ideal pd as the weight.

The *limited* simulation experience: i) supports the theoretical arguments; ii) shows that the proposed weighting tends to provide (often significant) improvement; iii) indicates that the proposed weighting substantially speeds up the learning of the optimal decision rule while the presented significant difference in quality diminishes in long run; iv) confirms that the used approximation of the past closed-loop model influences visibly the result quality; v) reveals that very high values of the weight W_τ may occur due to the "practical" violation of assumed essential boundedness; vi) shows that the learning of the closed-loop model with a data-dependent forgetting behaves well.

6 Concluding Remarks

The weights are used properly and no other correct way seems to exist. The proposed normalisation of the weights is conceptually unique – the unambiguous *approximate* choice of the numerator in (5) stays open. The asymptotics, when the time-invariance makes its inspection meaningful, is the correct one: when the ideal situation $\mathsf{I}_t(\underline{d}_\tau | \underline{d}^{\tau-1}) = \mathsf{C}_\tau(\underline{d}_\tau | \underline{d}^{\tau-1})$ occurs, the weight $\mathsf{W}_t = 1$ is reached.

Assumption (6) on the logarithmic ratio excludes parametric models that assign zero probability to data realisations, which are accepted as possible by the selected ideal pd. It can be weakened to the requirement on boundedness of the second moments. Algorithmically, it is connected with the upper bound $\overline{W}$ on weights W_t. Sensitivity to specific values of $\overline{W}$ seems to be low.

If almost no past data can be interpreted as coming from the optimally tuned closed loop, then $\mathsf{W}_\tau << 1$, $\tau \leq t-1$, and the posterior pd becomes flat. This makes the one-step-ahead predictor of the closed-loop behaviour (3) flat, too. This situation enhances the explorative nature of actions generated from it, as desirable.

The choice (5) resembles the trick well-known in Monte Carlo evaluations when a feasible "proposal" pd is used [6]. The past closed-loop model plays its role. The analogy is, however, mechanical and seems to bring no tangible consequences.

Table 1. The simulated system.

$\mathsf{F}(\underline{x}_t \vert a_t, \underline{x}_{t-1})$	$a_t = 1$	$a_t = 2$	$a_t = 3$	$a_t = 4$
$\mathsf{F}(\underline{x}_t = 1 \vert a_t, \underline{x}_{t-1} = 1)$	0.9975	0.0196	0.0196	0.9901
$\mathsf{F}(\underline{x}_t = 2 \vert a_t, \underline{x}_{t-1} = 1)$	0.0025	0.9804	0.9804	0.0099
$\mathsf{F}(\underline{x}_t = 1 \vert a_t, \underline{x}_{t-1} = 2)$	0.0196	0.9901	0.9967	0.0196
$\mathsf{F}(\underline{x}_t = 2 \vert a_t, \underline{x}_{t-1} = 2)$	0.9804	0.0099	0.0033	0.9804

Table 2. The decision rule found.

$\mathsf{S}(\underline{a}_t \vert x_{t-1})$	$x_{t-1} = 1$	$x_{t-1} = 2$
$\mathsf{S}(\underline{a}_t = 1 \vert x_{t-1})$	0.4607	0.1148
$\mathsf{S}(\underline{a}_t = 2 \vert x_{t-1})$	0.0292	0.3241
$\mathsf{S}(\underline{a}_t = 3 \vert x_{t-1})$	0.0293	0.4463
$\mathsf{S}(\underline{a}_t = 3 \vert x_{t-1})$	0.4808	0.1148

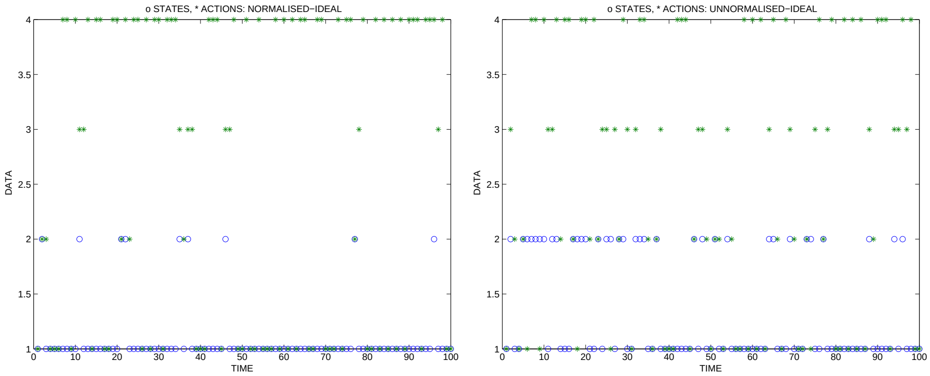

Fig. 1. Simulated behaviour: normalised weight (left), un-normalised weight (right).

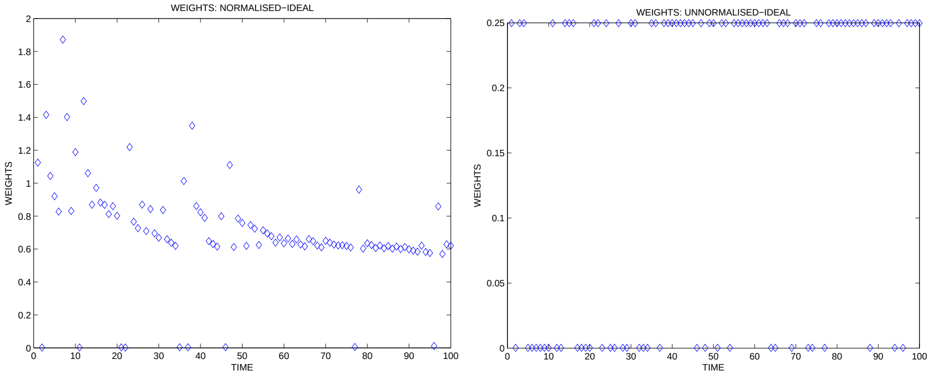

Fig. 2. Time course of the weight: normalised (left) and un-normalised (right).

Open problems: i) A decision, which of mentioned approximations of $\mathsf{C}_\tau(d_\tau|d^{\tau-1})$ is better is to be made or an alternative option found. ii) Closed-loop stability is the major unsolved issue – the approximation of the ideal dynamics $\mathsf{I}_t(d_t|d^{t-1})$ by the closed-loop model $\mathsf{C}_t(d_t|d^{t-1})$ does not guarantee it; iii)The result guarantees that the one-step-ahead predictor of the closed-loop behaviour approximates the one-step-ahead ideal pd. In truly dynamic cases, the receding horizon strategy [25] can be immediately designed: it suffices to handle blocks od decisions. Other, more efficient ways of coping with DM dynamics have to be developed.

Acknowledgements. The reported research was supported by the project GACR13-13502S.

References

1. Bellman, R.: Adaptive Control Processes. Princeton U. Press, NJ (1961)
2. Berec, L., Kárný, M.: Identification of reality in Bayesian context. In: Warwick, K., Kárný, M. (eds.) Computer-Intensive Methods in Control and Signal Processing, pp. 181–193. Birkhäuser (1997)
3. Berger, J.: Statistical Decision Theory and Bayesian Analysis. Springer, New York (1985)
4. Bertsekas, D.: Dynamic Programming and Optimal Control. Athena Scientific, US (2001)
5. Bontempi, G., Birattari, M., Bersini, H.: Lazy learning for local modelling & control design. Int. J. of Control 72(7-8), 643–658 (1999)
6. Cappe, O., Godsill, S., Moulines, E.: An overview of existing methods and recent advances in sequential Monte Carlo. Proc. of the IEEE 95(5), 899–924 (2007)
7. Daum, F.: Nonlinear filters: beyond the Kalman filter. Aerospace and Electronic Systems Magazine, IEEE 20(8), 57–69 (2005)
8. Doucet, A., Johansen, A.: A tutorial on particle filtering and smoothing: Fifteen years later. In: Handbook of Nonlinear Filtering. Oxford University Press, Oxford, UK (2011)
9. Feldbaum, A.: Theory of dual control. Autom. Remote Control 21(9) (1960)
10. Gilboa, I., Schmeidler, D.: Case-based decsion theory. The Quaterly Journal of Economics 110, 605–639 (1995)
11. Guan, P., Raginsky, M., Willett, R.: Online Markov decision processes with Kull-backLeibler control cost. IEEE Trans. on Automatic Control (2014)
12. Kárný, M.: Towards fully probabilistic control design. Automatica 32(12), 1719–1722 (1996)
13. Kárný, M.: Adaptive systems: Local approximators? In: Workshop n Adaptive Systems in Control and Signal Processing. pp. 129–134. IFAC, Glasgow (1998)
14. Kárný, M.: On approximate fully probabilistic design of decision making strategies. In: Guy, T., Kárný, M. (eds.) Proceedings of the 3rd International Workshop on Scalable Decision Making, ECML/PKDD 2013. UTIA AV *c*R, Prague (2013), iSBN 978-80-903834-8-7
15. Kárný, M.: Approximate Bayesian recursive estimation. Information Sciences (2014), dOI 10.1016/j.ins.2014.01.048
16. Kárný, M., Guy, T.V.: Fully probabilistic control design. Systems & Control Letters 55(4), 259–265 (2006)

17. Kárný, M., Kroupa, T.: Axiomatisation of fully probabilistic design. Information Sciences 186(1), 105–113 (2012)
18. Kulhavý, R., Zarrop, M.B.: On a general concept of forgetting. Int. J. of Control 58(4), 905–924 (1993)
19. Kullback, S., Leibler, R.: On information and sufficiency. Annals of Mathematical Statistics 22, 79–87 (1951)
20. Li, J., Dong, G., Ramamohanarao, K., Wong, L.: Deeps: A new instance-based lazy discovery and classification system. Machine Learning 54(2), 99–124 (2004)
21. Loeve, M.: Probability Theory. van Nostrand, Princeton, New Jersey (1962), Russian translation, Moscow 1962
22. Macek, K., Guy, T., Kárný, M.: A lazy-learning concept of fully probabilistic decision making. Unpublished manuscript (2014)
23. Martín-Sánchez, J., Lemos, J., Rodellar, J.: Survey of industrial optimized adaptive control. Int. J. of Adaptive Control and Signal Processing 26(10), 881–918 (2013)
24. Peterka, V.: Bayesian system identification. In: Eykhoff, P. (ed.) Trends and Progress in System Identification, pp. 239–304. Pergamon Press, Oxford (1981)
25. Qin, S., Badgwell, T.: A survey of industrial model predictive control technology. Control Engineering Practice 11(7), 733–764 (2003)
26. Rao, M.: Measure Theory and Integration. John Wiley, NY (1987)
27. Roll, J., Nazin, A., Ljung, L.: Nonlinear system identification via direct weight optimization. Automatica 41(3), 475–490 (2004)
28. Sanov, I.: On probability of large deviations of random variables. Matematiceskij Sbornik 42, 11–44 (1957), (in Russian), also in Selected Translations Mathematical Statistics and Probability, I, 1961, 213–244
29. Savage, L.: Foundations of Statistics. Wiley, NY (1954)
30. Schon, T., Gustafsson, F., Nordlund, P.: Marginalized particle filters for mixed linear/nonlinear state-space models. IEEE Tran. on Signal Processing 53(7), 2279–2289 (2005)
31. Si, J., Barto, A., Powell, W., Wunsch, D. (eds.): Handbook of Learning and Approximate Dynamic Programming. Wiley-IEEE Press, Danvers (May 2004)
32. Tishby, N., Polani, D.: Information theory of decisions and actions. In: Cutsuridis, V., Hussain, A., Taylor, J. (eds.) Perception-Action Cycle, pp. 601–636. Springer Series in Cognitive and Neural Systems, Springer, New York (2011)
33. Todorov, E.: Linearly-solvable Markov decision problems. In: Schölkopf, B., et al (eds.) Advances in Neural Inf. Processing, pp. 1369 – 1376. MIT Press, NY (2006)
34. Zhu, C., Zhu, W.: Feedback control of nonlinear stochastic systems for targeting a specified stationary probability density. Automatica 47(3), 539–544 (2006)