Preference Elicitation within Framework of Fully Probabilistic Design of Decision Strategies ⋆

Miroslav Kárný and Tatiana V. Guy *

* *The Czech Academy of Sciences, Institute of Information Theory and Automation, POB 18, 182 08 Prague 8, Czech Republic{school,guy}@utia.cas.cz*

Abstract: The paper proposes the preference-elicitation support within the framework of fully probabilistic design (FPD) of decision strategies. Agent employing FPD uses probability densities to model the closed-loop behaviour, i.e. a collection of all observed, opted and considered random variables. Opted actions are generated by a randomised strategy. The optimal decision strategy minimises Kullback-Leibler divergence of the closed-loop model to its ideal counterpart describing the agent's preferences. Thus, selecting the ideal closed-loop model comprises preference elicitation.
The paper provides a general choice of the best ideal closed-loop model reflecting agent's preferences. The foreseen application potential of such a preference elicitation is high as FPD is a non-trivial dense extension of Bayesian decision making that dominates prescriptive decision theories.
The general solution is illustrated on the regulation task with a linear Gaussian model describing the agent's environment.

Keywords: dynamic decision making, preference elicitation, fully probabilistic design, decision strategy, Kullback Leibler Divergence.

1. INTRODUCTION

Decision making here means any targeted choice among actions at disposal. Quantification of agent's beliefs and preferences is inevitable prerequisite for an optimising decision making under uncertainty and incomplete knowledge Savage (1954); Wallenius et al. (2008).

Knowledge elicitation (beliefs' quantification) is a relatively well-developed area, e.g. Daee et al. (2017); Genest and Zidek (1986); Kadane et al. (1980); O'Hagan et al. (2006); Quinn et al. (2016).

Preference elicitation is less developed in spite of a significant progress. Generally, it is a harder problem Chen and Pearl (2004) than the knowledge-elicitation process Cooke (1994). It has to cope with multivariate (often high dimensional) attributes determining preferences, with their incomplete and informal specifications, with an inherent uncertainty, with the need to elicit group preferences from individual preferences, etc.

The lack of an *unambiguous* way of combining multiple criteria (attributes) Keeny and Raiffa (1978) is the most significant obstacle faced. Even advanced solutions deal with domain specific problems Gajos and Weld (2005), static decision making Boutilier et al. (1997); Dupplaw et al. (2004) and they use quite restrictive assumptions Benabbou et al. (2016). In addition to the multi-attribute problem, the works often care about costs connected with querying about preferences Branke et al. (2017); Drummond and Boutilier (2014). The orientation on repetitive interactions with active human agents limits their use.

⋆ MŠMT ČR LTC18075 and EU-COST Action CA16228 support this research.

Notably, work Chajewska et al. (2000) fosters the preference elicitation based on a probabilistic modelling of utilities (criteria). This way allows to deal systematically with multiple attributes and to adapt a population preference model to an individual. The outlined methodology still deals with static problems and uses the belief model only for stopping of the elicitation process, for the querying end.

Fully probabilistic design (FPD) of decision strategies Guan et al. (2014); Kárný and Guy (2006); Kárný and Kroupa (2012); Todorov (2006) directly describes preferences probabilistically. This conceptually solves the multi-attribute problem and connects probabilistic descriptions of beliefs and preferences. Paper Kárný (2013) exploits FPD for eliciting multiple-attribute-driven preferences in dynamic decision making. It still contains too many arbitrary steps. The current *paper solves the elicitation problem more systematically and provides its unambiguous optimal solution*. It has to be approximated in order to get feasible solution.

Section 2 recalls FPD and introduces the adopted notation and notions. Section 3 provides conceptual solution of the preference-elicitation problem. Section 4 elaborates this in a widely applicable set up. The specialisation to the regulation task with a linear Gaussian model of the environment in Section 5 indicates the solution plausibility. Section 6 adds concluding remarks.

2. FULLY PROBABILISTIC DESIGN

Dynamic decision making concerns an agent[1] that interacts with incompletely known and randomly responding environ-

[1] The agent can be a human, a device or their mixed group.

ment. *Behaviour* $b \in \boldsymbol{b}$ of the closed loop, formed by the pair agent-environment, has the structure [2]

$$b = (g_t, a_t, k_{t-1}) \in \boldsymbol{b}. \tag{1}$$

There, a real time moment at which *action* $a_t \in \boldsymbol{a}$ is applied is labelled by $t \in \boldsymbol{t} = \{1, \ldots, h\}$ with $h < \infty$ being the *decision horizon*. At time t, $g_t \in \boldsymbol{g}$ denotes the considered part of the behaviour unavailable for selecting the action a_t. It consists of future fully observable states $(x_t, \ldots, x_h)$ and actions $(a_{t+1}, \ldots, a_h)$. *Knowledge* $k_{t-1} \in \boldsymbol{k}$ is a part of the behaviour available for selecting the action a_t and consists of a prior knowledge [3] k_0, past states $\underline{x}_1, \ldots, \underline{x}_{t-1}$ and actions $\underline{a}_1, \ldots, \underline{a}_{t-1}$. Underlining marks the realisation of a random variable, whenever this distinction is important.

The agent has to construct an (optimal) *randomised decision strategy* (policy) $\mathsf{s} \in \mathbf{s}$ consisting of a sequence of informationally causal *decision rules* r_t, i.e.

$$\mathsf{r}_t : \boldsymbol{k} \to \boldsymbol{a}, \quad \mathsf{s} = (\mathsf{r}_t)_{t \in \boldsymbol{t}}. \tag{2}$$

The resulting closed-loop model $\mathsf{c}^{\mathsf{s}}(b)$ describes closed-loop behaviours $b \in \boldsymbol{b}$ and it is influenced by the used decision strategy s. Behaviour b is a multivariate random variable, Savage (1954), completely described by its joint *probability density* (pd) $\mathsf{c}^{\mathsf{s}}(b)$, $b \in \boldsymbol{b}$. It can be factorised via the chain rule for pds Peterka (1981). For presentation simplicity, Markovian environment models are considered and consequently Markovian decision rules assumed Mine and Osaki (1970). This gives

$$\mathsf{c}^{\mathsf{s}}(b) = \prod_{t \in \boldsymbol{t}} \overbrace{\mathsf{m}(x_t|a_t, x_{t-1})}^{\text{environment model}} \times \overbrace{\mathsf{r}(a_t|x_{t-1})}^{\text{decision rule}} \tag{3}$$

Within the adopted FPD framework, the *optimal decision strategy* $\mathsf{s}^{\mathsf{o}} \in \mathbf{s}$ minimises Kullback-Leibler divergence (KLD, Kullback and Leibler (1951)) of the closed-loop model c^{s} to its *ideal* counterpart describing the agent's preferences , i.e.

$$\mathsf{s}^{\mathsf{o}} \in \operatorname{Arg}\min_{\mathsf{s} \in \mathbf{s}} \mathsf{D}(\mathsf{c}^{\mathsf{s}}||\mathsf{c}^{\mathsf{i}}) \tag{4}$$

$$\mathsf{D}(\mathsf{c}^{\mathsf{s}}||\mathsf{c}^{\mathsf{i}}) = \int_{\boldsymbol{b}} \mathsf{c}^{\mathsf{s}}(b) \ln\left(\frac{\mathsf{c}^{\mathsf{s}}(b)}{\mathsf{c}^{\mathsf{i}}(b)}\right) \mathrm{d}b.$$

The ideal pd $\mathsf{c}^{\mathsf{i}}(b)$, $b \in \boldsymbol{b}$, models the *desired* closed-loop behaviour and can be factorised similarly to (3)

$$\mathsf{c}^{\mathsf{i}}(b) = \prod_{t \in \boldsymbol{t}} \overbrace{\mathsf{m}^{\mathsf{i}}(x_t|a_t, x_{t-1})}^{\text{ideal environment model}} \times \overbrace{\mathsf{r}^{\mathsf{i}}(a_t|x_{t-1})}^{\text{ideal decision rule}} \tag{5}$$

The next proposition provides the FPD-optimal strategy. The evaluations are close to dynamic programming but give the optimal decision rules $(\mathsf{r}^{\mathsf{o}}_t)_{t \in \boldsymbol{t}}$ in an explicit form.

Proposition 1. (FPD, proof in Kárný et al. (2006)). Optimal decision strategy $\mathsf{s}^{\mathsf{o}}(b) = (\mathsf{r}^{\mathsf{o}}(a_t|x_{t-1}))_{t \in \boldsymbol{t}}$ is given by the following optimal decision rules ($\propto$ denotes proportionality)

$$\mu(a_t, x_{t-1}) = \int_{\boldsymbol{x}} \mathsf{m}(x_t|a_t, x_{t-1}) \ln\left(\frac{\mathsf{m}(x_t|a_t, x_{t-1})}{\gamma(x_t)\mathsf{m}^{\mathsf{i}}(x_t|a_t, x_{t-1})}\right) \mathrm{d}x_t$$

$$\mathsf{r}^{\mathsf{o}}(a_t|x_{t-1}) = \frac{\mathsf{r}^{\mathsf{i}}(a_t|x_{t-1}) \exp[-\mu(a_t, x_{t-1})]}{\underbrace{\int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a_t|x_{t-1}) \exp[-\mu(a_t, x_{t-1})] \mathrm{d}a_t}_{\gamma(x_{t-1})}} \tag{6}$$

$$\propto \mathsf{r}^{\mathsf{i}}(a_t|x_{t-1}) \exp[-\mu(a_t, x_{t-1})].$$

The backward evaluations start with

$$\gamma(x_h) = 1 \geq \gamma(x_t) \geq \gamma^{\mathsf{i}}(x_t), \text{ see (9) } \forall t \in \boldsymbol{t}. \tag{7}$$

The reached minimum is

$$\min_{\mathsf{s} \in \mathbf{s}} \mathsf{D}(\mathsf{c}^{\mathsf{s}}||\mathsf{c}^{\mathsf{i}}) = -\ln(\gamma(x_0)). \tag{8}$$

Value function $-\ln(\gamma(x_t))$ is bounded from above by its greedy analogy $-\ln(\gamma^{\mathsf{i}}(x_t))$ with

$$\gamma^{\mathsf{i}}(x_{t-1}) = \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a_t|x_{t-1}) \tag{9}$$

$$\times \exp\Big[\underbrace{-\int_{\boldsymbol{x}} \mathsf{m}(x_t|a_t, x_{t-1}) \ln\left(\frac{\mathsf{m}(x_t|a_t, x_{t-1})}{\mathsf{m}^{\mathsf{i}}(x_t|a_t, x_{t-1})}\right) \mathrm{d}x_t}_{\mu^{\mathsf{i}}(a_t, x_{t-1})}\Big] \mathrm{d}a_t.$$

3. CONCEPT OF OPTIMAL PREFERENCE ELICITATION

The need for preference elicitation arises whenever the agent's preferences do not specify uniquely the ideal pd (5), i.e. when the set $\mathbf{c}^{\mathbf{i}}$ of prospective ideal pds

$$\mathbf{c}^{\mathbf{i}} = \left\{\mathsf{c}^{\mathsf{i}}(b) : \text{ pds on } \boldsymbol{b} \text{ reflecting the agent's preferences}\right\} \tag{10}$$

contains at least two different ideal pds. Then, the agent faces an additional decision problem: *How to select the most suitable and informative ideal pd from* set $\mathbf{c}^{\mathbf{i}}$?

Choosing an ideal pd from a proper subsets of the specified set $\mathbf{c}^{\mathbf{i}}$ (10) enforces additional preferences that *were not* part of the original agent's preferences. In this case a minimum reached, see (4) of form (8), may only increase. Respecting this simple observation, we propose to select *the most suitable ideal from set* $\mathbf{c}^{\mathbf{i}}$ *(10) as a minimiser of the reached minima in (4) over* $\mathbf{c}^{\mathbf{i}}$.

$$\mathsf{c}^{\mathsf{oi}} \in \operatorname{Arg}\min_{\mathsf{c}^{\mathsf{i}} \in \mathbf{c}^{\mathbf{i}}} \min_{\mathsf{s} \in \mathbf{s}} \mathsf{D}(\mathsf{c}^{\mathsf{s}}||\mathsf{c}^{\mathsf{i}}). \tag{11}$$

The resulting pd c^{oi} is called *optimal ideal* closed-loop model c^{oi} onward. Similarly to (5), (11) can be factorised

$$\mathsf{c}^{\mathsf{oi}}(b) = \prod_{t \in \boldsymbol{t}} \mathsf{m}^{\mathsf{oi}}(x_t|a_t, x_{t-1}) \mathsf{r}^{\mathsf{oi}}(a_t|x_{t-1}).$$

The factors $\mathsf{m}^{\mathsf{oi}}(x_t|a_t, x_{t-1})$, $\mathsf{r}^{\mathsf{oi}}(a_t|x_{t-1})$, $t \in \boldsymbol{t}$, are referred as the *optimal* ideal environment model and the *optimal* ideal decision rule.

The optimisation (11) is generically infeasible as the optimised ideal environment model and the ideal decision rules enter $-\ln(\gamma(x_0)) = \min_{\mathsf{s} \in \mathbf{s}} \mathsf{D}(\mathsf{c}^{\mathsf{s}}||\mathsf{c}^{\mathsf{i}})$ in a very complex way, see Proposition 1. This makes us to search for an approximate but feasible solution.

Set $\mathbf{c}^{\mathbf{i}}$ may be empty if the agent's preferences are internally inconsistent. This possibility is here untreated.

[2] Throughout, $\boldsymbol{x}$ denotes a set of xs. Its nature is described only when needed.
[3] The prior knowledge is fixed, includes $\underline{x}_0$, and it is present implicitly.

4. PREFERENCE ELICITATION

This section elaborates the concept presented in Section 3 for a relatively wide class of decision tasks. The solution focuses on one-stage-ahead FPD in order to get a feasible solution. This simplification corresponds with the minimisation of the upper bound on $\min_{\mathsf{s}\in\mathbf{s}} \mathsf{D}(\mathsf{c}^{\mathsf{s}}||\mathsf{c}^{\mathsf{i}})$, cf. (7), (9). The gained optimal ideal closed-loop model is then used in the multi-step FPD described by Proposition 1. Thus, the greedy (one-stage-ahead, myopic) choice of c^{oi} does not cause the decision-quality loss that is connected with it, see e.g. Mayne (2014).

For a fixed $t \in \boldsymbol{t}$, the minimisation over a set of ideal closed-loop models $\mathbf{c}^{\mathsf{i}}$ is a static task. Thus, time index t and fixed condition $\underline{x}_{t-1}$ can be dropped in notation. The considered minimisation is then equivalent to the maximisation of γ^{i} in (9) over the set (10). For $\mathsf{c}^{\mathsf{i}}(x,a) = \mathsf{m}^{\mathsf{i}}(x|a)\mathsf{r}^{\mathsf{i}}(a)$ (5) and $\mu^{\mathsf{i}}(a)$ in (9)

$$\begin{aligned}\mathsf{c}^{\mathsf{oi}} &= \mathsf{m}^{\mathsf{oi}}\mathsf{r}^{\mathsf{oi}} \\ &\in \mathrm{Arg}\max_{\{(\mathsf{m}^{\mathsf{i}},\mathsf{r}^{\mathsf{i}})\in(\mathbf{m}^{\mathsf{i}},\mathbf{r}^{\mathsf{i}})\}} \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a)\exp\left(-\mu^{\mathsf{i}}(a)\right)\mathrm{d}a.\end{aligned} \tag{12}$$

Optimisation in (12) is performed with respect to the ideal decision rule r^{i} first, then with respect to the ideal environment model m^{i}.

Optimal ideal decision rule r^{oi} (12) must select an action in the given action set $\boldsymbol{a}$. The form of optimal decision rule (6) implies that this requirement is fulfilled iff

$$\mathrm{supp}[\mathsf{r}^{\mathsf{i}}] = \{a : \mathsf{r}^{\mathsf{i}}(a) > 0\} \subset \boldsymbol{a}, \tag{13}$$

where $\mathrm{supp}[\mathsf{r}^{\mathsf{i}}]$ is a support of r^{i}.

All actions in $\boldsymbol{a}$ are admissible and thus *no action in* $\boldsymbol{a}$ *should be a priori excluded*. This specifies the following general requirement

$$\mathrm{supp}[\mathsf{r}^{\mathsf{i}}] = \boldsymbol{a}. \tag{14}$$

This requirement represents an active constraint. It becomes obvious when writing the explicit form of the optimisation (12) over the unconstrained set $\mathbf{r}^{\mathsf{i}}$ of ideal decision rules

$$\max_{\mathsf{r}^{\mathsf{i}}\in\mathbf{r}^{\mathsf{i}}} \gamma^{\mathsf{i}} = \max_{\mathsf{r}^{\mathsf{i}}\in\mathbf{r}^{\mathsf{i}}} \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a)\rho^{\mathsf{i}}(a)\mathrm{d}a, \quad \rho^{\mathsf{i}}(a) = \exp\left(-\mu^{\mathsf{i}}(a)\right) \geq 0. \tag{15}$$

The optimised ideal decision rule r^{i} enters (15) linearly. Consequently, the unconstrained optimal ideal decision rule is deterministic: it concentrates on an action maximising ρ^{i}. Thus, the unconstrained optimal ideal decision rule violates requirement (14).

The next proposition characterises deterministic decision rules in the way, which allows us to exclude them, i.e. to guarantee that the requirement (14) is met.

Proposition 2. (Deterministic Decision Rules). Let us consider $\mathbf{r} = \{$ set of all pds on $\boldsymbol{a}$, which is a subset of a finite-dimensional real space$\}$. Let us define the constant $\bar{\kappa}$ according to the cardinality of action set $\boldsymbol{a}$

$$\bar{\kappa} = \begin{cases} 1 & \text{if the cardinality of } \boldsymbol{a} < \infty \\ \infty & \text{if the cardinality of } \boldsymbol{a} = \infty \end{cases}.$$

Then, for a fixed rule $\mathsf{r} \in \mathbf{r}$, it holds

$$\bar{\kappa} = \int_{\boldsymbol{a}} \mathsf{r}^2(a)\mathrm{d}a \quad \Leftrightarrow \quad \mathsf{r}(a) \text{ is deterministic.} \tag{16}$$

Proof

The case with a finite cardinality of $\boldsymbol{a}$

To prove the implication $\Leftarrow$ in (16), let $\mathsf{r}(\underline{a}) = 1$ for some $\underline{a} \in \boldsymbol{a}$. Then, $\mathsf{r}(a) = 0$ for $a \in \boldsymbol{a} \setminus \{\underline{a}\} \Rightarrow \sum_{a\in\boldsymbol{a}} \mathsf{r}^2(a) = 1 = \bar{\kappa}$.

The implication $\Rightarrow$ in (16) is proved by contradiction. Let $\sum_{a\in\boldsymbol{a}} \mathsf{r}^2(a) = 1 = \bar{\kappa}$. Let $\bar{\boldsymbol{a}}$ be the non-empty subset of the action set $\boldsymbol{a}$ on which $\mathsf{r}(a) \in (0,1)$ and $\mathsf{r}(a) = 0$ on $\boldsymbol{a} \setminus \bar{\boldsymbol{a}}$. Then, the contradicting inequality arises

$$1 = \sum_{a\in\boldsymbol{a}} \mathsf{r}(a) = \sum_{a\in\bar{\boldsymbol{a}}} \mathsf{r}(a) > \sum_{a\in\bar{\boldsymbol{a}}} \mathsf{r}^2(a) = \sum_{a\in\boldsymbol{a}} \mathsf{r}^2(a) = 1.$$

The case with an infinite cardinality of $\boldsymbol{a}$ is outlined for $\boldsymbol{a}$ being the whole finite-dimensional real space and $\bar{\kappa} = \infty$.

The implication $\Leftarrow$ in (16) is proved directly. A deterministic decision rule giving $\underline{a} \in \boldsymbol{a}$ is the limit of Gaussian pds $\mathcal{G}_a(\underline{a},\varepsilon)$ with a common mean $\underline{a}$ and diagonal covariances with non-zero entries $\varepsilon \to 0^+$. It holds

$$\begin{aligned}\int_{\boldsymbol{a}} \mathsf{r}^2(a)\mathrm{d}a &= \int_{\boldsymbol{a}} \lim_{\varepsilon\to0^+} \mathcal{G}_a^2(\underline{a},\varepsilon)\mathrm{d}a \\ &= \lim_{\varepsilon\to0^+} \int_{\boldsymbol{a}} \mathcal{G}_a^2(\underline{a},\varepsilon)\mathrm{d}a = \infty = \bar{\kappa},\end{aligned}$$

where the last equality can be analytically verified.

Implication $\Rightarrow$ in (16) is proved by contradiction. Let r be a non-deterministic decision rule with essential supremum $\alpha = \mathrm{essup}(\mathsf{r}) \in (0,\infty)$. Then, the contradiction arises

$$\begin{aligned}1 &= \int_{\boldsymbol{a}} \mathsf{r}(a)\mathrm{d}a = \alpha \int_{\boldsymbol{a}} \frac{\mathsf{r}(a)}{\alpha}\mathrm{d}a \\ &\overbrace{\geq}^{\frac{\mathsf{r}(a)}{\alpha}\leq 1} \alpha^{-1} \int_{\boldsymbol{a}} \mathsf{r}^2(a)\mathrm{d}a = \alpha^{-1}\bar{\kappa} = \infty.\end{aligned}$$

■

The next proposition provides the optimal ideal decision rule maximising γ^{i} (15) so that the requirement (14) is fulfilled. It uses the set indicator defined as

$$\chi_{\boldsymbol{a}}(a) = \begin{cases} 1 & \text{if } a \in \boldsymbol{a} \\ 0 & \text{if } a \notin \boldsymbol{a} \end{cases}. \tag{17}$$

Proposition 3. (Optimal Ideal Decision Rule). Let an ideal environment model m^{i} be chosen and let almost everywhere (a.e.) on $\boldsymbol{a}$, see (9),

$$\mu^{\mathsf{i}}(a) = \int_{\boldsymbol{x}} \mathsf{m}(x|a) \ln\left(\frac{\mathsf{m}(x|a)}{\mathsf{m}^{\mathsf{i}}(x|a)}\right)\mathrm{d}x < \infty, \tag{18}$$

which is equivalent to the implication

$$\mathsf{m}^{\mathsf{i}}(x|a) = 0 \Rightarrow \mathsf{m}(x|a) = 0, \text{ a.e. on } \boldsymbol{x}. \tag{19}$$

Then, the related optimal ideal decision rule r^{oi} maximising

$$\gamma^{\mathsf{i}} = \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a)\rho^{\mathsf{i}}(a)\mathrm{d}a, \ \rho^{\mathsf{i}}(a) = \exp(-\mu^{\mathsf{i}}(a)) \tag{20}$$

over the set

$$\mathbf{r}^{\boldsymbol{i}} = \left\{\mathsf{r}^{i}(a) : \mathrm{supp}[\mathsf{r}^{i}] \subset \boldsymbol{a} \wedge \int_{\boldsymbol{a}} \left(\mathsf{r}^{\mathsf{i}}\right)^2(a)\mathrm{d}a \leq \kappa < \bar{\kappa}\right\} \tag{21}$$

has the form, cf. (17),

$$\mathsf{r}^{\mathsf{oi}}(a) \propto \chi_{\boldsymbol{a}}(a)\rho^{\mathsf{i}}(a), \ \rho^{\mathsf{i}} = \exp\left[-\mu^{\mathsf{i}}(a)\right]. \tag{22}$$

Its support, $\mathrm{supp}[\mathsf{r}^{\mathsf{i}}] = \boldsymbol{a}$, i.e. this rule meets requirement (14).

Proof The equivalence $\mu^{\mathsf{i}}(a) < \infty$ with (19) is the basic property of KLD. Under (19), the decision rule (22) meets the requirement (14). Thus, it remains to show that this decision rule maximises γ^{i} in (20).

The unconstrained maximiser of γ^{i} (20) reaches bound $\bar{\kappa}$ (16), Proposition 2. Thus, the constraint in (21) is active and the bound κ is attained, $\int_{\boldsymbol{a}} \left(\mathsf{r}^{\mathsf{i}}\right)^2(a)\mathrm{d}a = \kappa \in (0,\infty)$. Moreover, $\beta = \int_{\boldsymbol{a}} \left(\chi_{\boldsymbol{a}}(a)\rho^{\mathsf{i}}(a)\right)^2 \mathrm{d}a \in (0,\infty)$, due to (19) and the definition $\rho^{\mathsf{i}}(a) = \exp(-\mu^{\mathsf{i}}(a))$. Thus, the normalised version of γ^{i}

$$\frac{\int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a)\chi_{\boldsymbol{a}}(a)\rho^{\mathsf{i}}(a)\mathrm{d}a}{\sqrt{\kappa\beta}} \tag{23}$$

can be maximised. Functions r^{i} and ρ^{i} are square-integrable (summable) and the maximised functional (23) is their scalar product normalised by product of their norms. Thus, it is cosine (23) of the angle between the pair of square integrable (summable) functions, which is maximised when they are collinear, Rao (1987). This proves (22).

■

While the requirement (21) on the ideal decision rule is quite general the other preferences are much more variable. A lot of them, however, can be expressed as a constraint on generalised moments. The constraint (24) is given by an agent-specified ℓ_{q}-dimensional vector function $\mathsf{q}(x,a)$, $x \in \boldsymbol{x}, a \in \boldsymbol{a}$,

$$0 = \int_{\boldsymbol{x}}\int_{\boldsymbol{a}} \mathsf{q}(x,a)\mathsf{m}(x|a)\mathsf{r}^{\mathsf{i}}(a)\mathrm{d}x\mathrm{d}a. \tag{24}$$

Importantly, the constraint (24) concerns the closed-loop model with the given — *unchangeable* — environment model and the optional ideal decision rule.

A typical choice of q (used in regulation task, Section 5) is

$$\mathsf{q}(x,a) = x - x^{\mathsf{i}} \tag{25}$$

with x^{i} being the value of x desired by the agent.

Insertion of the optimal ideal decision rule (22) into the constraint (24) provides the following explicit constraint on admissible ideal environment models

$$\begin{aligned} 0 = &\int_{\boldsymbol{x}}\int_{\boldsymbol{a}} \mathsf{q}(x,a)\mathsf{m}(x|a) \\ &\times \exp\left[-\int_{\boldsymbol{x}} \mathsf{m}(\tilde{x}|a)\ln\left(\frac{\mathsf{m}(\tilde{x}|a)}{\mathsf{m}^{\mathsf{i}}(\tilde{x}|a)}\right)\mathrm{d}\tilde{x}\right]\mathrm{d}x\mathrm{d}a. \end{aligned} \tag{26}$$

Proposition 4. (Optimal Ideal Environment Model). Let the set of ideal closed-loop models $\mathbf{c}^{\mathsf{i}}$ determined by (21), (24) be non-empty [4]. The optimal ideal environment model $\mathsf{m}^{\mathsf{oi}}(x|a)$, maximising γ^{i} in (21) over pairs $(\mathsf{m}^{\mathsf{i}}, \mathsf{r}^{\mathsf{i}})$ in the set $\mathbf{c}^{\mathsf{i}}$, reads ($'$ is transposition) reads

$$\mathsf{m}^{\mathsf{oi}}(x|a) = \frac{\mathsf{m}(x|a)\exp(\lambda'\mathsf{q}(x,a))}{\int_{\boldsymbol{x}} \mathsf{m}(x|a)\exp(\lambda'\mathsf{q}(x,a))\mathrm{d}x} = \frac{\mathsf{L}(x,a)}{\int_{\boldsymbol{x}} \mathsf{L}(x,a)\mathrm{d}x}, \tag{27}$$

where real ℓ_{q}-dimensional vector λ is chosen so that (26) is met.

Proof Consider the following auxiliary optimisation task for a fixed $\mathsf{r}^{\mathsf{i}} \in \mathbf{r}^{\mathbf{i}}$ such that $(\mathsf{m}^{\mathsf{i}}, \mathsf{r}^{\mathsf{i}})$ is in the non-empty set $\mathbf{c}^{\mathbf{i}}$

$$\begin{aligned} &\min_{\mathsf{m}^{\mathsf{i}}\in\mathbf{m}^{\mathbf{i}}} \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a) \\ &\times \left\{\int_{\boldsymbol{x}} \mathsf{m}(x|a)\left[\ln\left(\frac{\mathsf{m}(x|a)}{\mathsf{m}^{\mathsf{i}}(x|a)}\right) + \lambda'\mathsf{q}(x,a)\right]dx\right\}da. \end{aligned}$$

λ is chosen so that the minimising m^{i} meets the constraint (26). It exists as $\mathbf{c}^{\mathbf{i}} \neq \emptyset$. The solution has form (27) and point-wise guarantees that $\mu^{\mathsf{oi}}(a) \le \mu^{\mathsf{i}}(a)$, see (9), for any other $\mathsf{m}^{\mathsf{i}}(x|a)$ among the considered ideals. Thus, it maximises $\rho^{\mathsf{i}}(a) = \exp(-\mu^{\mathsf{i}}(a))$ over $\mathsf{m}^{\mathsf{i}} \in \mathbf{m}^{\mathbf{i}}$ and $\forall a \in \boldsymbol{a}$

$$\rho^{\mathsf{oi}}(a) = \exp(-\mu^{\mathsf{oi}}(a)) \ge \rho^{\mathsf{i}}(a) = \exp(-\mu^{\mathsf{i}}(a)). \tag{28}$$

Consequently,

$$\gamma^{\mathsf{oi}} = \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{oi}}(a)\rho^{\mathsf{oi}}(a)\mathrm{d}a \ge \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a)\rho^{\mathsf{i}}(a)\mathrm{d}a = \gamma^{\mathsf{i}}$$

as it follows from

$$\int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{oi}}(a)\rho^{\mathsf{oi}}(a)\mathrm{d}a \underbrace{\ge}_{\text{Proposition 3}} \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a)\rho^{\mathsf{oi}}(a)\mathrm{d}a \underbrace{\ge}_{(28)} \int_{\boldsymbol{a}} \mathsf{r}^{\mathsf{i}}(a)\rho^{\mathsf{i}}(a)\mathrm{d}a.$$

■

A direct combination of Propositions 3, 4 gives the desired optimal closed-loop ideal.

Proposition 5. (Optimal Ideal Closed Loop Model). The optimal ideal closed-loop model, maximising γ^{i} (21), over the set of closed-loop ideals defined by (21) and (24), reads

$$\begin{aligned} \mathsf{c}^{\mathsf{i}}(x,a) &= \mathsf{m}^{\mathsf{oi}}(x|a)\mathsf{r}^{\mathsf{oi}}(a) \\ &\propto \frac{\mathsf{m}(x|a)\exp[\lambda'\mathsf{q}(x,a)]}{\int_{\boldsymbol{x}} \mathsf{m}(x|a)\exp[\lambda'\mathsf{q}(x,a)]\mathrm{d}x} \\ &\times \chi_{\boldsymbol{a}}(a)\frac{\exp\left(\int_{\boldsymbol{x}} \lambda'\mathsf{q}(x,a)\mathsf{m}(x|a)\mathrm{d}x\right)}{\int_{\boldsymbol{x}} \mathsf{m}(x|a)\exp[\lambda'\mathsf{q}(x,a)]\mathrm{d}x} \end{aligned} \tag{29}$$

where ℓ_{q}-dimensional equation

$$0 = \int_{\boldsymbol{x}}\int_{\boldsymbol{a}} \mathsf{q}(x,a)\mathsf{m}(x|a)\mathsf{r}^{\mathsf{oi}}(a)\mathrm{d}x\mathrm{d}a$$

determines the real ℓ_{q}-dimensional vector λ in (29).

5. APPLICATION TO LINEAR GAUSSIAN CASE

As an example, let us consider the wide-spread regulation task Meditch (1969) with the linear Gaussian environment model given by known compatible matrices $\mathbb{A}$, $\mathbb{B}$, $\mathbb{R} > 0$. The regulation task is specified by the wish to keep the observed, real finite-dimensional state x at a given fixed, typically zero, value. For this, real finite-dimensional actions a are at disposal.

The regulation aim is expressed by the requirement

$$\mathsf{q}(x,a) = x \text{ and the desired } x^{\mathsf{i}} = 0 \text{ in (25).} \tag{30}$$

It uses Gaussian environment model (recall, at the considered time t, $\underline{x} = \underline{x}_{t-1}$, which is the already observed state)

$$\begin{aligned} \mathsf{m}(x|a) &= \mathcal{G}_x(\mathbb{A}\underline{x} + \mathbb{B}a, \mathbb{R}) \\ &= \frac{\exp[-0.5(x-z)'\mathbb{R}^{-1}(x-z)]}{\sqrt{|2\pi\mathbb{R}|}}, \\ &\text{where} \quad z = \mathbb{A}\underline{x} + \mathbb{B}a. \end{aligned} \tag{31}$$

The application of Proposition 5 gives the optimal ideal environment model, given by (27),

[4] Existence of m^{i} meeting (19) and (26) suffices.

$$\begin{aligned}\mathsf{L}(x,a) &= \mathsf{m}(x|a)\exp[\lambda'\mathsf{q}(x,a)] \\ &= \frac{\exp\left\{-0.5\left[(x-z)'\mathbb{R}^{-1}(x-z)-2\lambda' x\right]\right\}}{\sqrt{|2\pi\mathbb{R}|}} \\ &= \frac{\exp\left\{-\frac{(x-z-\mathbb{R}\lambda)'\mathbb{R}^{-1}(x-z-\mathbb{R}\lambda)-\lambda'\mathbb{R}\lambda-2\lambda' z}{2}\right\}}{\sqrt{|2\pi\mathbb{R}|}}.\end{aligned}$$

This implies the form of the optimal ideal environment model

$$\mathsf{m}^{\mathsf{oi}}(x|a) \overbrace{=}^{\mathsf{q}=x} \mathcal{G}_x(0,\mathbb{R}) = \mathcal{G}_x(z+\mathbb{R}\lambda,\mathbb{R}). \tag{32}$$

This is equivalent to the choice

$$\lambda = -\mathbb{R}^{-1}z = -\mathbb{R}^{-1}(\mathbb{A}\underline{x}+\mathbb{B}a) \;\text{ giving}$$
$$\mathsf{L}(x,a) = \exp\left\{-0.5[x'\mathbb{R}^{-1}x + z'\mathbb{R}^{-1}z]\right\}.$$

This finally gives

$$\mathsf{r}^{\mathsf{oi}}(a) \propto \exp[-0.5(\mathbb{A}\underline{x}+\mathbb{B}a)'\mathbb{R}^{-1}(\mathbb{A}\underline{x}+\mathbb{B}a)],$$

i.e. the optimal ideal decision rule reads

$$\mathsf{r}^{\mathsf{oi}}(a) = \mathcal{G}_a\left((\mathbb{B}'\mathbb{R}^{-1}\mathbb{B})^{-1}\mathbb{B}'\mathbb{R}^{-1}\mathbb{A}\underline{x}, (\mathbb{B}'\mathbb{R}^{-1}\mathbb{B})^{-1}\right). \tag{33}$$

The results (32), (33) have the following interpretation:

- the desired value $x^{\mathsf{i}} = 0$ of the state x, cf. (25), is the mean of the optimal ideal environment model (32) and the non-reducible covariance $\mathbb{R}$ of the environment model (31) is its covariance;
- the optimal ideal decision rule (33) is proportional to the environment model at $x^{\mathsf{i}} = 0$, thus, it prefers the actions that make this desired state the most probable;
- the corresponding FPD-optimal strategy, Proposition 1, preserves its important multi-step character in spite of the greedy construction of the optimal ideal closed-loop model.

6. CONCLUDING REMARKS

The paper solves the preference elicitation problem using the fully probabilistic design of decision strategies that quantifies preferences via the ideal closed-loop pd. The optimal ideal closed-loop pd is derived from:

- the set of admissible actions;
- the (learnable) environment model;
- the agent's incompletely specified preferences expressed via generalised moments.

Methodologically, the proposed preference elicitation complements the minimum KLD principle serving for the knowledge elicitation. Its concept specifies the optimality criterion that respects both knowledge and preferences. Consequently, gradual learning of the environment model directly induces learning of preferences Belda (2009). This is one of yet unfulfilled aim of control theory as well as artificial intelligence Pigozzi et al. (2016).

The proposed methodology quantifies preferences while respecting the used environment model. Thus, it never recommends Gaussian ideal pd (an extension of quadratic loss function) for the preference quantification when dealing with Cauchy states. Importantly, it avoids even less obvious discrepancies in description of beliefs and preferences.

A lot remains to be done, for instance, the requirement (24) and Proposition 5 should be tailored to discrete-valued states and actions in order to support the wide-spread Markov decision processes. For them, the characterisation of *non-empty* sets of prospective ideals (10) is vital.

Generally, other forms of constraints on possible ideal environment models are worth inspecting. Sets specified by inequalities on generalised moments or having the form of unions of KLD balls used in generalisations of minimum KLD principle offer themselves as the first options to be tried.

The presented preference elicitation for regulation task with linear Gaussian environment model indicates that a further development is worth of research effort. Indeed, the considered regulation task is in the root of "classical" modern control theory Meditch (1969) that can be directly extended to tracking problems or applied to economic problems requiring rational inattention Sims (2006).

REFERENCES

Belda, K. (2009). Probabilistic design and tuning of LQ control. *Archives of Control Sciences*, 19(4 (2009)), 355–368.

Benabbou, N., Diodoro, S., Perny, P., and Viappiani, P. (2016). Preference elicitation in multi-attribute domains for choice and ranking with the Borda count. In *The Tenth International Conference on Scalable Uncertainty Management*. Nice, France.

Boutilier, C., Brafman, R., Geib, C., and Poole, D. (1997). A constraint-based approach to preference elicitation and decision making. In *Proceedings of AAAI Spring Symposium on Qualitative Decision Theory*, 19–28. Stanford, CA.

Branke, J., Corrente, S., Greco, S., and Gutjahr, W. (2017). Efficient pairwise preference elicitation allowing for indifference. *Computers & Operations Research*, 88(Suppl. C), 175 – 186.

Chajewska, U., Koller, D., and Parr, R. (2000). Making rational decisions using adaptive utility elicitation. In *Proc. of the 17th Nat. Conf. on AI*, 363–369.

Chen, L. and Pearl, P. (2004). Survey of preference elicitation methods. Technical Report IC/2004/67, HCI Group Ecole Politechnique Federale de Lausanne, Switzerland.

Cooke, N. (1994). Varieties of knowledge elicitation techniques. *International J. of Human-Computer Studies*, 41, 801–849.

Daee, P., Peltola, T., Soare, M., and Kaski, S. (2017). Knowledge elicitation via sequential probabilistic inference for high-dimensional prediction. *Machine Learning*, 106(9), 1599–1620.

Drummond, J. and Boutilier, C. (2014). Preference elicitation and interview minimization in stable matchings. In *Proceedins of Twenty-Eight AAAI Conference on AI*, 645 – 653.

Dupplaw, D., Brunson, D., Vine, A., Please, C., Lewis, S., Dean, A., Keane, A., and Tindall, M. (2004). A web-based knowledge elicitation system (GISEL) for planning and assessing group screening experiments for product development. *J. of Computing and Inf. Science in Engineering*, 4(3), 218–225.

Gajos, K. and Weld, D. (2005). Preference elicitation for interface optimization. In *Proc. of the 18th annual ACM Symposium on User interface and Technology, UIST-2005*, 173 – 182.

Genest, C. and Zidek, J. (1986). Combining probability distributions: A critique and annotated bibliography. *Stat. Sci.*, 1(1), 114–148.

Guan, P., Raginsky, M., and Willett, R. (2014). Online Markov decision processes with Kullback Leibler control cost. *IEEE Trans. on AC*, 59(6), 1423–1438.

Kadane, J., Dickey, J., Winkler, R., Smith, W., and Peters, S. (1980). Interactive elicitation of opinions for normal linear models. *J. of the American Statistical Association*, 75(372), 845–854.

Kárný, M. (2013). Automated preference elicitation for decision making. In T. Guy and et al (eds.), *Decision Making and Imperfection*, volume 474, 65–99. Springer, Berlin.

Kárný, M., Böhm, J., Guy, T., Jirsa, L., Nagy, I., Nedoma, P., and Tesař, L. (2006). *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. Springer.

Kárný, M. and Guy, T.V. (2006). Fully probabilistic control design. *Syst. & Con. Lett.*, 55(4), 259–265.

Kárný, M. and Kroupa, T. (2012). Axiomatisation of fully probabilistic design. *Inf. Sci.*, 186(1), 105–113.

Keeny, R. and Raiffa, H. (1978). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. J. Wiley, NY.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *An. of Math. Stat.*, 22, 79–87.

Mayne, D. (2014). Model predictive control: Recent developments and future promise. *Automatica*, 2967–2986.

Meditch, J. (1969). *Stochastic Optimal Linear Estimation and Control*. Mc. Graw Hill.

Mine, H. and Osaki, S. (1970). *Markovian Decision Processes*. Elsevier.

O'Hagan, A., Buck, C.E., Daneshkhah, A., and et al (2006). *Uncertain Judgement: Eliciting experts' Probabilities*. John Wiley.

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

Pigozzi, G., , Tsoukiàs, A., , and Viappiani, P. (2016). Preferences in artificial intelligence. *Annals of Mathematics and Artificial Intelligence*, 77(3), 361–401. doi:10.1007/s10472-015-9475-5.

Quinn, A., Kárný, M., and Guy, T. (2016). Fully probabilistic design of hierarchical Bayesian models. *Inf. Sci.*, 369, 532–547.

Rao, M. (1987). *Measure Theory and Integration*. John Wiley, NY.

Savage, L. (1954). *Foundations of Statistics*. Wiley.

Sims, C.A. (2006). Rational inattention: Beyond the linear-quadratic case. *The American Economic Review*, 96(2), 158–163.

Todorov, E. (2006). Linearly-solvable Markov decision problems. In B. Schölkopf and et al (eds.), *Adv. in Neur. Inf. Proc.*, 1369 – 1376. MIT Press.

Wallenius, J., Dyer, J., Fishburn, P., Steuer, R., Zionts, S., and Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Sci.*, 54(7), 1336–1349.