Knowledge Elicitation via Extension of Fragmental Knowledge Pieces

Miroslav Kárný

Abstract— **The paper describes an advanced methodology of automatic knowledge elicitation. It merges fragmental uncertain knowledge pieces into the prior distribution of unknown parameter of a probabilistic model of a dynamic system.**

Careful knowledge elicitation helps in achieving as bump-less start of model-based controllers as possible. It is also important when observed data are poorly informative, which is a typical situation in closed control loops.

Rigorous use of the Bayesian paradigm to the knowledge elicitation forms the essence of the methodology. Unlike former solutions, it can handle fragmental and incompletely compatible knowledge pieces in a systematic way. The description of the methodology and of the uniform model relating knowledge pieces to the ideal merger dominate the paper. An illustrative example is presented.

I. INTRODUCTION

Model-based adaptive controllers rely on a recursive parameter estimation. Its various aspects (model structure selection, design and analysis of estimation algorithms etc.) are relatively well developed. A systematic handling of transient behavior is developed less. In the control context, transients are predominantly treated by specific control strategies, e.g. [1], [2]. It is a common knowledge that transients of the recursive estimation and consequently transients of the controller using them are significantly influenced by the estimation initiation. It can be shown that proper choice of initial conditions, made in off-line mode, helps in achieving (almost) bump-less start.

The conversion of available prior knowledge about unknown parameter, known as knowledge elicitation [3], is a well elaborated art in Bayesian statistics. However, as the excellent review [4] confirms, the majority of the developed techniques rely on a facilitator, who guides the expert providing knowledge and quantifies gathered knowledge. Excessive expenses make these techniques hardly applicable in control domain. Moreover, human being can cope only with relatively simple cases. This justifies the search for knowledge elicitation techniques weakly dependent on a facilitator, [5], [6], [7], [8], [9]. The results confirmed usefulness and feasibility of this direction but the developed techniques still lack sufficient universality and contain too many ambiguous steps. This paper contributes to decreasing this ambiguity. It adopts recently revived [10] formulation of knowledge elicitation as a Bayesian estimation task [11]. The paper proposes uniform model relating the ideal knowledge merger to respective knowledge pieces. Unlike the logarithmic model presented in [10], the obtained merger assigns non-zero probability to events for which there is at least one evidence that they may happen.

The adopted methodology is recalled Section II. Section III provides a plausible uniform model relating knowledge pieces to the unknown ideal merger of prior knowledge and applies Bayesian estimation to it. Section IV extends fragmental knowledge pieces to probability density functions (pdf) needed for the described merging. Illustrative example indicates use of the methodology, Section V. Concluding remarks in Section VI are dominated by a list of open research problems.

Throughout x^* denotes the set of x values, $\equiv$ is used as defining equality.

II. PRELIMINARIES

Bayesian estimation [12] has the following simple logical structure. Given data $d \in d^*$ and unknown parameter $\theta \in \theta^*$ ($\equiv$set of parameter values), their joint probability density function (pdf) $f(d, \theta)$

$$f(d,\theta) = \underbrace{f(d|\theta)}_{parametric\,pdf} \times \underbrace{f(\theta)}_{prior\,pdf} \tag{1}$$

is constructed. The parametric pdf (model) relates the observed data d to the unknown parameter θ. The prior pdf quantifies prior knowledge on the unknown parameter θ. Any Bayesian inference is based on characteristics of the posterior pdf $f(\theta|d)$

$$f(\theta|d) = \frac{f(d,\theta)}{f(d)} \propto f(d|\theta)f(\theta) \tag{2}$$

with the data realization inserted into it.

Let us outline application of this methodology to the knowledge elicitation.

Generally, the available pieces of prior knowledge can be expressed as characteristics of pdfs $f_k(\theta)$, $k \in k^* \equiv \{1, 2, \ldots, \mathring{k}\}$. The number $\mathring{k}$ of these *elicitation data* is finite. The prior pdf $f(\theta)$ that respects (merge) them in the best way is *unknown elicitation parameter*. The adopted wording indicates that the desired construction of the prior pdf can be cast into the Bayesian framework with the following correspondence

$$\text{data}\, D \equiv \{f_k(\theta)\}_{k\in k^*}, \text{unknown parameter}\, \Theta \equiv f(\theta),\ \theta \in \theta^*. \tag{3}$$

The merger of knowledge pieces is then a point estimate $f(\theta|\hat{\mathcal{V}})$ of $f(\theta)$ selected on the basis of the posterior distribution $F(\Theta|D)$ computed from the joint distribution $F(D, \Theta)$. The point estimate $f(\theta|\hat{\mathcal{V}})$ of $f(\theta)$ is searched within a

This work was partially supported GAČR 102/08/0567 and MŠMT 2C06001

M. Kárný is with Department of Adaptive Systems, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, POB 18, 182 08, Prague 8, Czech Republic school@utia.cas.cz

ISBN 978-963-311-369-1

class of computationally plausible pdfs $f(\theta|\mathcal{V})$, $\mathcal{V} \in \mathcal{V}^*$. The introduced pointer $\mathcal{V}$ "points" to respective members of the class and has to be finite-dimensional as it has to be stored in a computer.

This construction requires: (i) a sound choice of the joint distribution $F(D, \Theta)$, (ii) an extension of all offered knowledge pieces to the pdfs $\{f_k(\theta)\}_{k \in k^*}$, and (iii) evaluation of a well-justified point estimate $\hat{f}(\theta)$ of $f(\theta)$, $\theta \in \theta^*$, based on $\{f_k(\theta)\}_{k \in k^*}$.

The step (i) is by no means unique but it seems that for a group models studied up to now the results are relatively close each other. The steps (ii) and (iii) are uniquely driven by the Bayesian paradigm. It means that poor modeling or lack of prior knowledge can be blamed for possible non-satisfactory results.

III. Modeling and Estimation

Both the processed data $D \equiv \{f_k(\theta)\}_{k \in k^*}$ and the estimated parameter $\Theta \equiv f(\theta)$ are generally infinite-dimensional. Thus $F(D, \Theta)$ is a probabilistic measure on the modeled stochastic processes indexed by $\theta \in \theta^*$.

A. Modeling assumptions and general point estimate

The following assumptions are believed to be commonly acceptable for the considered knowledge-elicitation task. Moreover, they allow us to avoid technically difficult work with stochastic processes.

Assumption 1 (Localness principle): For a given $\theta \in \theta^*$, The values $f_k(\theta)$, $k \in k^*$, matter in estimation of $f(\theta)$ and in construction of $f(\theta|\hat{\mathcal{V}})$.

Within the Bayesian paradigm, this assumption and consideration of smooth loss functionals for point estimation imply, [13], that the point estimate is to minimize the Kerridge inaccuracy, [14]. It defines the best pointer $\hat{\mathcal{V}}$

$$\hat{\mathcal{V}} \in \mathrm{Arg} \min_{\mathcal{V} \in \mathcal{V}^*} - \int \mathsf{E}\left[f(\theta)|f_k(\theta), k \in k^*\right] \ln(f(\theta|\mathcal{V}))\,\mathrm{d}\theta, \tag{4}$$

where the expectation E is taken over the posterior distribution of the value $f(\theta)$ given by knowledge pieces $f_k(\theta), k \in k^*$. Thus, the choice of $\hat{\mathcal{V}}$ can be made after evaluating the conditional expectation

$$\hat{f}(\theta) \equiv \mathsf{E}\left[f(\theta)|f_k(\theta), k \in k^*\right]. \tag{5}$$

In evaluating the posterior distribution over the values $f(\theta)$ of the unknown prior pdf, we adopt

Assumption 2 (Independence assumptions):

1) Deviations of the processed values $f_k(\theta)$ from the values of the unknown prior pdf $f(\theta)$ are mutually conditionally independent for the given pdf $f(\theta)$.
2) The values $f_k(\tilde{\theta})$, $f_k(\theta)$, $\tilde{\theta} \neq \theta$, $\tilde{\theta}, \theta \in \theta^*$, are independent for the given $f(\tilde{\theta})$, $f(\theta)$.
3) All prior knowledge pieces are trustworthy to the same degree.

The first assumption excludes even a partial repetition of the same knowledge pieces. Often, it can be respected.

The second assumption maximizes modeling freedom. It is violated due to the normalization of the modeled pdfs $f_k(\theta)$, $f(\theta)$. If θ^* has finite cardinality, the dependence caused by the normalization can be explicitly removed . In the generic infinite-dimensional case, this dependence is weak and can be neglected. It suffices to normalize the final estimate $f(\theta|\hat{\mathcal{V}})$ of $f(\theta)$.

The third assumption is justified by insufficient reasons for an alternative view. It can be simple weakened if such reasons exist.

The localness and independence assumptions imply that the joint distribution $F(D, \Theta)$ is determined by its marginal projections. We assume that they are given by pdfs $F(f_k(\theta), f(\theta)) = F(f_k(\theta)|f(\theta))F(f(\theta))$, $k \in k^*$, $\theta \in \theta^*$.

The constructed parametric model $F(f_k(\theta)|f(\theta))$ is not unique. But the following assumptions seem to be inevitable for any reasonable model.

Assumption 3 (Common modeling assumptions):

1) The joint pdf $F(f_k(\theta), f(\theta))$ has the support on non-negative arguments.
2) Finite values of knowledge pieces are a priori expected

$$\mathsf{E}[f_k(\theta)] = \int f_k(\theta) F(f_k(\theta))\,\mathrm{d}f_k(\theta) < \infty,\ k \in k^*,\ \theta \in \theta^*. \tag{6}$$

3) The individual knowledge pieces are unbiased, i.e.,

$$\begin{aligned}&\mathsf{E}[f_k(\theta)|f(\theta)] \\ &= \int f_k(\theta) F(f_k(\theta)|f(\theta))\,\mathrm{d}f_k(\theta) = f(\theta)\end{aligned} \tag{7}$$

for all $k \in k^*$, $\theta \in \theta^*$.

4) The parametric model $F(f_k(\theta)|f(\theta))$ has the highest entropy among those meeting conditions put on them.

The first assumption reflects the firm prior knowledge available for the model building.

The second assumption excludes over-confident prior knowledge (represented by Dirac delta functions) and guarantees existence of the finite conditional expectation.

The third assumption is intuitively plausible. It can be also interpreted as a characterization of the unknown prior pdf $f(\theta)$ that respects all available knowledge pieces $f_k(\theta)$, $k \in k^*$.

The fourth assumption reflects our tendency to build into the model *just* available knowledge and let it to be as spread as possible around it.

B. Uniform parametric model

In [10], the above general assumptions were complemented so that log-normal parametric model $F(f_k(\theta)|f(\theta))$ was obtained. The corresponding merging of knowledge pieces is plausible but a single opinion $f_{\underline{k}}(\underline{\theta}) = 0$ for a $\underline{\theta} \in \theta^*$ and $\underline{k} \in k^*$ implies that the merger $\hat{f}(\underline{\theta}) = 0$. Here, we inspect the model implied by the above assumptions complemented by requirement on finiteness of the model support. This parametric model that removes the drawback of the log-normal model is characterized by the following simple proposition.

Proposition 1 (Uniform parametric model): Let the constructed parametric model have a finite support included into the interval $[0, f(\theta) + w(\theta)]$ for some $0 < w(\theta) < \infty$ and meet Assumptions of Section III-A. Then, it is determined by the uniform marginal pdf

$$F(f_k(\theta)|f(\theta), w(\theta)) = \mathcal{U}_{f_k(\theta)}(f(\theta) - w(\theta), f(\theta) + w(\theta)) \quad (8)$$

with $\mathcal{U}_x(\underline{x}, \overline{x}) = \frac{\chi(x \in (\underline{x}, \overline{x}))}{\overline{x} - \underline{x}}$ with $\chi(\bullet)$ being indicator of the set in its argument.

The half-width $w(\theta) \in [0, f(\theta)]$ is an additional optional parameter distinguishing various finite supports and thus various parametric models.

Proof: Let $f(x)$ be a pdf with its support within a finite interval $[0, \overline{x}]$. Then, $\mathsf{E}[x] = \int_0^{\overline{x}} x f(x)\,\mathrm{d}x \le \overline{x}$. Thus, the unbiasedness implies that the expression of the upper bound on the support in the form $f(\theta) + w(\theta)$ with $w(\theta) \in [0, \infty)$ is always possible.

Let $A(\theta) \subset [0, f(\theta) + w(\theta)]$ be the support of the constructed parametric model $F(f_k(\theta)|f(\theta), w(\theta))$. Then, the necessary condition for the extreme of the entropy under unbiasedness reads $F(f_k(\theta)|f(\theta), w(\theta)) = \frac{\chi(F(f_k(\theta)|f(\theta), w(\theta)) \in A)}{volume(A)}$. The entropy is decreasing function of the $volume(A)$. Thus, we are searching for the support included in $[0, f(\theta) + w(\theta)]$ with the largest volume while guaranteeing unbiasedness. The interval $[f(\theta) - w(\theta), f(\theta) + w(\theta)]$ is of this type. If $f(\theta) - w(\theta) \ge 0$ then the inclusion $A(\theta) \subset [0, f(\theta) + w(\theta)]$ is guaranteed. In the opposite case, we have to replace $w(\theta)$ by the largest value $\tilde{w}(\theta)$, smaller than $w(\theta)$ that allows us to embed the whole symmetric interval $[f(\theta) - \tilde{w}(\theta), f(\theta) + \tilde{w}(\theta)]$ into $[0, f(\theta) + w(\theta)]$. As neither $f(\theta)$ nor $w(\theta)$ are known we can simply rename $\tilde{w}(\theta)$ to $w(\theta)$. ■

C. Prior pdf

It remains to select the prior pdf $F(f(\theta), w(\theta))$. Its choice is driven by the wish to introduce as little pre-judice as possible by using a flat improper prior pdf. At the same time, simple evaluations and existence of a proper posterior pdf with a finite expectation have to be guaranteed. The following choice meets both these wishes.

$$F(f(\theta), w(\theta)) \propto \frac{\chi(f(\theta) \ge w(\theta)) \ge 0)}{f(\theta) w(\theta)}, \quad (9)$$

i.e., the prior probability of large values of the merger $f(\theta)$ as well as of large deviations of knowledge pieces (bounded by $w(\theta)$) falls to zero slowly.

D. Evaluation of the posterior expectation

Solution of the task (4) depends on the form of pdfs $f(\theta|\mathcal{V})$ but always the posterior expectation $\mathsf{E}[f(\theta)|f_k(\theta), k \in k^*]$ is always needed. It is evaluated here. During evaluations, a fixed argument θ is considered so that it can be temporarily dropped.

For the chosen model (8) and the prior pdf (9), we get the posterior pdf

$$\begin{aligned} & F(f, w|f_k, k \in k^*) \\ & \propto \frac{\chi(0 \le f - w \le L \le U \le f + w)\chi(f \ge w \ge 0)}{f w^{\mathring{k}+1}} \\ & L \equiv \min_{k \in k^*} f_k, \quad U \equiv \max_{k \in k^*} f_k. \end{aligned} \quad (10)$$

It has the marginal pdf

$$F(f|f_k, k \in k^*) \propto \chi(f \ge m(f)) \left[\frac{1}{f m^{\mathring{k}}(f)} - \frac{1}{f^{\mathring{k}+1}} \right] \quad (11)$$

$$\begin{aligned} m(f) &\equiv \max(0, U - f, f - L) \\ &= \begin{cases} U - f & \text{if} \quad f \in [0, M], \; M \equiv 0.5(U + L) \\ f - L & \text{if} \quad f \in [M, \infty) \end{cases}. \end{aligned}$$

The condition $f \ge m(f)$ implies $F(f|f_k, k \in k^*) \propto$

$$\frac{\chi\left(\frac{U}{2} \le f \le M\right)}{f(U - f)^{\mathring{k}}} + \frac{\chi(f \ge M)}{f(f - L)^{\mathring{k}}} - \frac{\chi\left(\frac{U}{2} \le f\right)}{f^{\mathring{k}+1}} \equiv \mathcal{L}(f) \quad (12)$$

The posterior expectation has the form $\hat{f} = \frac{E_1}{E_0}$ with

$$\begin{aligned} E_i &= \int_0^\infty f^k \mathcal{L}(f)\,\mathrm{d}f \underbrace{=}_{(12)} \int_{\frac{U}{2}}^{M} \frac{f^{i-1}}{(U - f)^{\mathring{k}}}\,\mathrm{d}f \quad (13) \\ &+ \int_M^\infty \frac{f^{i-1}}{(f - L)^{\mathring{k}}}\,\mathrm{d}f - \frac{1}{(\mathring{k} - i)(\frac{U}{2})^{\mathring{k}-i}} \\ &= U^{i-\mathring{k}} \int_{0.5}^{0.5(1+\Delta)} \frac{x^{i-1}}{(1 - x)^{\mathring{k}}}\,\mathrm{d}x \\ &+ L^{i-\mathring{k}} \int_{0.5(1+\Delta^{-1})}^{\infty} \frac{x^{i-1}}{(x - 1)^{\mathring{k}}}\,\mathrm{d}x - \frac{(\frac{U}{2})^{i-\mathring{k}}}{(\mathring{k} - i)} \\ \Delta &\equiv \frac{L}{U}. \end{aligned}$$

It gives directly

$$E_1 = \frac{2^{\mathring{k}}}{\mathring{k} - 1} \left[\frac{1}{(U - L)^{\mathring{k}-1}} - \frac{1}{U^{\mathring{k}-1}} \right]. \quad (14)$$

To evaluate E_0, it is sufficient to use the identity

$$\frac{1}{x(1 - x)^{\mathring{k}}} = \frac{1}{x} + \frac{1}{1 - x} + \sum_{j=2}^{\mathring{k}} \frac{1}{(1 - x)^j} \quad (15)$$

Using it, we get

$$\begin{aligned} E_0 &= \ln\left(\frac{U + L}{U - L}\right) \left[U^{-\mathring{k}} - (-L)^{-\mathring{k}}\right] - U^{-\mathring{k}} \sum_{j=1}^{\mathring{k}} \frac{2^j}{j} \\ &+ \sum_{j=1}^{\mathring{k}-1} \frac{2^j}{j(U - L)^j} \left[U^{j-\mathring{k}} - (-L)^{j-\mathring{k}}\right]. \end{aligned} \quad (16)$$

It gives the final formula searched for.

Proposition 2 (Posterior expectation):

$$\begin{aligned} & \hat{f}(\theta) \equiv \mathsf{E}[f(\theta)|f_k(\theta), k \in k^*] \\ &= E_1 \times E_0^{-1} = \frac{2^{\mathring{k}}}{\mathring{k} - 1} \left[\frac{1}{(U - L)^{\mathring{k}-1}} - \frac{1}{U^{\mathring{k}-1}} \right] \quad (17) \\ & \times \left\{ \ln\left(\frac{U + L}{U - L}\right) \left[U^{-\mathring{k}} - (-L)^{-\mathring{k}}\right] - U^{-\mathring{k}} \sum_{j=1}^{\mathring{k}} \frac{2^j}{j} \right. \\ & \left. + \sum_{j=1}^{\mathring{k}-1} \frac{2^j}{j(U - L)^j} \left[U^{j-\mathring{k}} - (-L)^{j-\mathring{k}}\right] \right\}^{-1}. \end{aligned}$$

For a pair of knowledge pieces, $\mathring{k} = 2$, the merger becomes

$$\mathsf{E}[f(\theta)|f_k(\theta), k \in k^*] = \frac{4\left[\frac{1}{U-L} - \frac{1}{U}\right]}{\ln\left(\frac{U+L}{U-L}\right)[U^{-2} - L^{-2}] + \frac{2}{U-L}[U^{-1} + L^{-1}] - 4U^{-2}}. \quad (18)$$

The merging is completed by approximating the evaluated posterior expectation $\hat{f}(\theta) \equiv \mathsf{E}[f(\theta)|f_k(\theta), k \in k^*]$ by the pdf $f(\theta|\hat{\mathcal{V}})$ from a plausible class $f(\theta|\mathcal{V})$, $\mathcal{V} \in \mathcal{V}^*$, see (4).

IV. Extension of Fragmental Knowledge Pieces

The posterior expectation $\hat{f}(\theta) \equiv \mathsf{E}[f(\theta)|f_k(\theta), k \in k^*]$ has been constructed assuming that all knowledge pieces are given by pdfs $\{f_k(\theta)\}_{k\in k^*}$. This is a rare situation. Mostly, just a pre-prior pdf is given that roughly delimits support of $f(\theta)$. Possibly some knowledge pieces are determined by complete pdfs $f_k(\theta)$. All other knowledge pieces are given in the form of (generalized) moments

$$\mu_k = \int \phi_k(\theta) f_k(\theta)\, \mathrm{d}\theta, \quad (19)$$

where the finite-dimensional vectors μ_k and the vector function $\phi_k(\theta)$ are the offered knowledge pieces.

We cope with this generic case by extending the respective knowledge pieces μ_k, $\phi_k(\theta)$ to pdfs ${}^e f_k(\theta)$, $\theta \in \theta^*$. The meaningful extensions should naturally preserve the offered knowledge pieces (19), i.e., to fulfill

$$\mu_k = \int \phi_k(\theta)\, {}^e f_k(\theta)\, \mathrm{d}\theta. \quad (20)$$

At the same time, the extensions should be chosen as the best approximation of the constructed merger $\hat{f}(\theta)$. According to the already cited result [13], they have to minimize the expected Kerridge inaccuracy, i.e.,

$${}^e f_k(\theta) \in \mathrm{Arg} \min_{f_k(\theta)\ \text{fulfilling (19)}} - \int \hat{f}(\theta) \ln(f_k(\theta))\, \mathrm{d}\theta. \quad (21)$$

This convex optimization has the unique solution

$${}^e f_k(\theta) \propto \hat{f}(\theta) \exp[-\zeta_k' \phi_k(\theta)] \quad (22)$$

with $'$ denoting transposition and the vector ζ_k chosen so that ${}^e f_k(\theta)$ in (22) meets (20). The formula (22) defines ${}^e f_k(\theta)$ *implicitly* as $\hat{f}(\theta)$ is determined by the formula (17) in which $f_k(\theta)$ are replaced by extensions ${}^e f_k(\theta)$ whenever $f_k(\theta)$ is not a priori given. For reference purposes, we summarize the obtained result into the following proposition.

Proposition 3 (General, uniform model based, merger):

1) Let us adopt Assumptions 1, 2, 3 with the item 2) guaranteed by a finite range of values of $f_k(\theta)$.
2) Let us choose the prior pdf (9).
3) Let us accept the extension principle described in this paragraph.

Then the conditional expectation $\hat{f}(\theta) \equiv \mathsf{E}[f(\theta)|f_k(\theta), k \in k^*]$ is the solution of the following implicit equation (argument θ suppressed)

$$\hat{f} \propto \frac{2^{\mathring{k}}}{\mathring{k}-1}\left[\frac{1}{(U-L)^{\mathring{k}-1}} - \frac{1}{U^{\mathring{k}-1}}\right] \quad (23)$$
$$\times \left\{ \ln\left(\frac{U+L}{U-L}\right)\left[U^{-\mathring{k}} - (-L)^{-\mathring{k}}\right] - U^{-\mathring{k}} \sum_{j=1}^{\mathring{k}} \frac{2^j}{j} + \sum_{j=2}^{\mathring{k}} \frac{2^{j-1}}{(j-1)(U-L)^{j-1}}\left[U^{j-1-\mathring{k}} - (-L)^{j-1-\mathring{k}}\right] \right\}^{-1},$$
$$L = \hat{f} \min_{k\in k^*} \exp[-\zeta_k' \phi_k], \quad U = \hat{f} \max_{k\in k^*} \exp[-\zeta_k' \phi_k]$$
$$\mu_k \propto \int \phi_k \hat{f} \exp[-\zeta_k' \phi_k]\, \mathrm{d}\theta, \quad k \in k^*.$$

The resulting equation (23) can be well solved by successive approximations both over gradual guesses of extensions for respective k and over all $k \in k^*$.

V. Illustrative Example

To illustrate the presented theory, let us assume that the knowledge piece $k = 1$ is represented by the pdf $f_1(\theta)$ (say pre-prior one). The additional knowledge piece $k = 2$ has the form of the generalized moment (19). The offered information (19) is extended to the pdf ${}^e f_2(\theta)$. The extension ${}^e f_2(\theta)$ is merged with $f_1(\theta)$ according to (23).

Figure 1 shows the result of the merging, obtained, for $\mu_2 = 0.5$ and $\phi_2(\theta) = \chi(\theta > -1)$, i.e., for the offered information on the median of θ.

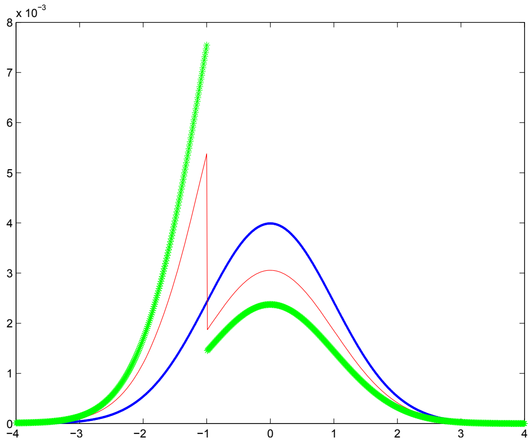

Fig. 1. Parameter values θ are on x-axis, pdf values are y axis. The pre-prior pdf $f_1(\theta)$ (thick line, blue dots in colored version) is modified to the merger $\hat{f}(\theta)$ (thin (red) line) by knowledge about median, which was extended to the pdf ${}^e f_2(\theta)$ (gray (green) stars).

VI. Conclusions

An advanced methodology of prior knowledge elicitation that can cope with a complex, fragmental, partially compatible knowledge pieces has been proposed. Methodologically, it relies on a systematic use of the Bayesian paradigm even to knowledge elicitation. The construction of the merger (elicited prior pdf) is made via the unique, well-justified, extensions of the fragmental knowledge pieces.

Comparing to previous attempts, it (i) offers automatic processing; (ii) unifies treatment of an extreme width of knowledge pieces; and (iii) minimizes the extent of arbitrariness.

The methodological shift in the item (iii) is the most significant one. Practical progress consists of elaboration of the uniform model relating the processed knowledge pieces to the unknown ideal merger.

This paper reports on the work in progress so that the number of open problems to be addressed is relatively large. Let us list and comment them.

- Basic constructions of the model $F(D,\Theta)$ can be made mathematically cleaner.
- Reliability of various knowledge pieces can be differentiated.
- Influence of the model $F(D,\Theta)$ should be studied. Current fragmental experiments indicate that the final mergers do not differ too much in spite of a significant difference between the models used.
- The processed pdf can have form of a marginal pdf or conditional pdf of θ entries. This case can be surely covered [10].
- The implicit equation for merger is now solved by successive approximations. Experiments indicate that it works. But theoretical analysis is missing. It seems that analysis of the minimizer of a mixture of the Kullback-Leibler divergencies made in [15] is more or less directly applicable.
- Numerically, the merging can be sensitive. Indeed, this is the case of the formula obtained for uniform model. A lot of algorithmic work is needed to make the computation safe.
- The a posteriori expectation $\hat{f}(\theta)$ is to be approximated by $f(\theta|\mathcal{V})$ from a suitable class. This is algorithmic and numerical problem on its own.

Promises of the approach and potential benefits make the addressing of this extensive sets of problems worthwhile.

ACKNOWLEDGEMENT

The approach described in this paper was to a significant extent stimulated by critical remarks of Drs' F. Ruggeri, A. Bodini and T.V. Guy to former methodologies of prior knowledge treatment. The discussions were enabled by the joint bi-lateral research project IMATI CNR and ÚTIA AVČR.

REFERENCES

[1] Kosut R.L., “Iterative adaptive control: Windsurfing with confidence”, in *Model Identification and Adaptive Control - From Windsurfing to Telecommunications*, Goodwin G.C., Ed. Springer-Verlag, London, UK, 2001.

[2] Hang C.C., Astrom K.J., and Wang Q.G, “Relay feedback auto-tuning of process controllers a tutorial review”, *Journal of Process Control*, vol. 12, pp. 143–162, 2002.

[3] K.M. Osei-Bryson, “Supporting knowledge elicitation and consensus building for Dempster-Shafer decision models”, *International Journal of Intelligent Systems*, vol. 18, no. 1, pp. 129–148, 2003.

[4] P.H. Garthwaite, J.B. Kadane, and A. O´Hagan, “Statistical methods for eliciting probability distributions”, *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 680–700, Jun 2005.

[5] M. Kárný, “Quantification of prior knowledge about global characteristics of linear normal model”, *Kybernetika*, vol. 20, no. 5, pp. 376–385, 1984.

[6] M. Kárný, N. Khailova, P. Nedoma, and J. Böhm, “Quantification of prior information revised”, *International Journal of Adaptive Control and Signal Processing*, vol. 15, no. 1, pp. 65–84, 2001.

[7] B. Betró and A. Guglielmi, “Methods for global prior robustness under generalized moment conditions”, in *Robust Bayesian Analysis*, D. Rios-Insua and F. Ruggeri, Eds., pp. 273 –294. Springer Verlag, New York, 2000.

[8] M. Kárný, P. Nedoma, N. Khailova, and L. Pavelková, “Prior information in structure estimation”, *IEE Proceedings — Control Theory and Applications*, vol. 150, no. 6, pp. 643–653, 2003.

[9] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař, *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, London, 2005.

[10] Kárný, T.V. Guy, A. Bodini, and F. Ruggeri, “Cooperation via sharing of probabilistic elements”, Tech. Rep., Milano, 2008.

[11] C. Genest and J.V. Zidek, “Combining probability distributions: A critique and annotated bibliography”, *Statistical Science*, vol. 1, no. 1, pp. 114–148, 1986.

[12] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.

[13] J. M. Bernardo, “Expected information as expected utility”, *The Annals of Statistics*, vol. 7, no. 3, pp. 686–690, 1979.

[14] D.F. Kerridge, “Inaccuracy and inference”, *Journal of Royal Statistical Society*, vol. B 23, pp. 284–294, 1961.

[15] J. Kracík, “Cooperation methods in bayesian decision making with multiple participants”, Tech. Rep. 2232, ÚTIA AV ČR, Praha, 2008.