Log-Normal Merging for Distributed System Identification

Václav Šmídl and Miroslav Kárný

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

Abstract Growing interest in applications of distributed systems, such as multi-agent systems, increases demands on identification of distributed systems from partial information sources collected by local agents. We are concerned with fully distributed scenario where system is identified by multiple agents, which do not estimate state of the whole system but only its local 'state'. The resulting estimate is obtained by merging of marginal and conditional posterior probability density functions (pdf) on such local states. We investigate the use of recently proposed non-parametric log-normal merging of such 'fragmental' pdfs for this task. We derive a projection of the optimal merger to the class of weighted empirical pdfs and mixtures of Gaussian pdfs. We illustrate the use of this technique on distributed identification of a controlled autoregressive model.

Keywords: system identification, distributed-parameter systems, distributed artificial intelligence, decentralized systems

1. INTRODUCTION

The field of estimation of distributed systems was established by early works [Speyer, 1979] where distributed Kalman filter was developed. Nowadays, this area is receiving a lot of attention due to growing interest in networked and distributed systems, such as multi-agent systems or sensor networks. Traditionally, distributed system identification is performed by a group of agents with identical description of the state, which is estimated from observations available to each agent [Alriksson, 2008]. Each agent provides posterior probability density function (pdf) on the common state variable. The pdfs are then combined using various merging techniques. This approach is suitable for applications in distributed target tracking or localization [Rosencrantz et al., 2003]. However, this arrangement does not scale well to systems with large state such as traffic networks [Šmídl and Přikryl, 2006]. Moreover, modelling of the full common state may be unnecessary in situations when only local state of an agent is required to make decisions. In such applications, it is sufficient to improve only posterior pdf on the local state using information from the neighbors.

We are concerned with fully distributed scenario, where each agent is building posterior pdf on its local state and can communicate this pdf to its neighbors. The communicated pdfs can be of any type and shape, can be defined on any variable and conditioned on any variable. They have to be only weakly compatible: the underlying joint pdfs have to have common support. Up to our knowledge, this task has been addressed in the literature only very recently by Kárný et al. [2008], where a methodological solution for fully decentralized decision-making with a flat structure of cooperation was introduced. This approach naturally embraces distributed identification and has the following features of interest:

- the proposed methodology allows to combine fragmental information such as pdfs of overlapping variables, or merging of pdfs characterized by moments only,
- the used non-parametric evaluation scheme allows a unified handling of different types of pdfs,
- the numerical implementation of the resulting combination was elaborated only for a rectangular grid support.

While the first two features are of practical interest, the third represents a strong limitation for evaluation of high-dimensional pdfs or pdfs varying on wide support.

In this paper, we present a new way of evaluation of the non-parametric results. Specifically, we project the non-parametric pdfs into the class of weighted empirical pdfs. Merging of the source pdfs projected into this class is almost identical to the grid-based version. However, decomposition of the pdf into conditional and marginal pdfs—which is needed for merging of fragmental pdfs—is difficult in this class. This problem is addressed by projecting the empirical pdfs into a mixture of pdfs from exponential family. The resulting algorithm provides a universal tool for merging of pdfs. It can be used in various structures of decentralized identification such as peer-to-peer structures, hierarchical structures or structures with special-purpose mediator agents.

The paper is organized as follows. In Section 2, we shortly review the non-parametric probability merging of [Kárný et al., 2008]. In Section 3, we provide the proposed evaluation scheme based on importance sampling and probabilistic mixtures. Finally, the resulting algorithm is applied to distributed estimation of a controlled regressive model in Section 4.

2. BAYESIAN MERGING

We are concerned with two scenarios: (i) merging of source pdfs describing a common variable x, (ii) and merging of sources with only fragmental information about x. Solution of the former scenario will be used as a subroutine for solution of the latter scenario. If not stated otherwise, all results in this Section are from [Kárný et al., 2008].

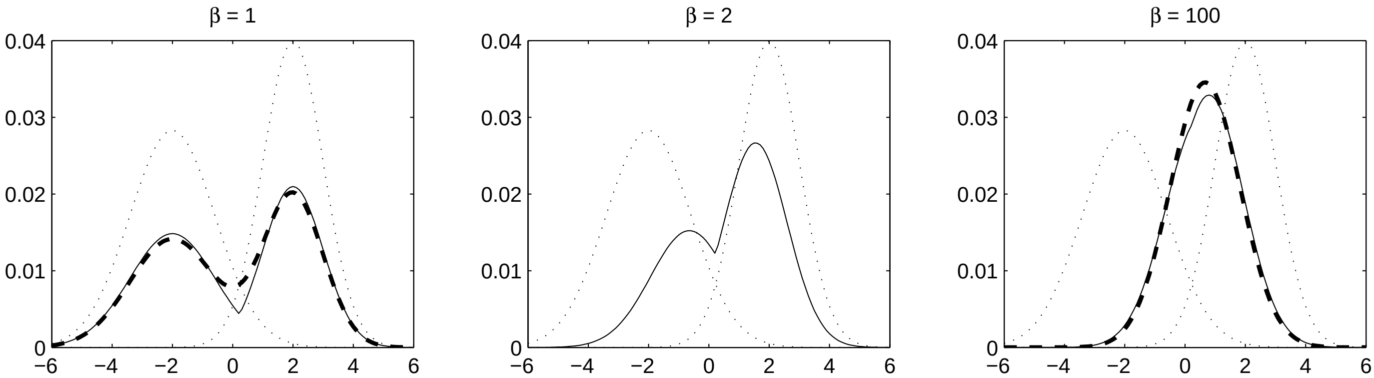

Figure 1. Illustration of log-normal merging for two Gaussian source pdfs. Dotted lines denote original source pdfs, full lines denote result of log-normal merging for $\beta = 1$ (left), $\beta = 2$ (middle) and $\beta = 100$ (right). For comparison, results of arithmetic merging and geometric merging are displayed as thick dashed lines in the left and in the right plot, respectively.

2.1 Merging of pdfs with common variable

The task is to aggregate information from several *source* pdfs, $f_{1:S}(x) \equiv \{f_1(x), \ldots f_S(x)\}$, into one combined pdf, $\tilde{f}(x)$, $\forall x \in x^*$. x^* is common support of all sources and the symbol $\zeta_{1:m} = \{\zeta_1, \ldots, \zeta_m\}$ is used throughout for various ζ.

The considered task is known as probability merging or probability combination. It has been studied in statistics for decades, see e.g. Genest and Zidek [1986] for survey. Many methods were developed for specific problems, however, no generally accepted solution is established. In this paper, we follow so-called *supra Bayesian* approach, where the 'ideally' combined pdf $\tilde{f}(x)$ is interpreted as an unknown parameter and the sources are taken as its 'noisy' observations. Since the 'parameter' $\tilde{f}(x)$ is infinite-dimensional in generic case, the merging becomes the task of non-parametric Bayesian parameter estimation. The task is conceptually solved by the Bayes rule if we choose: (i) likelihood function of sources $f_s(x)$ given $\tilde{f}(x)$, and (ii) prior pdf on the estimated pdf $\tilde{f}(x)$.

The estimation problem is solved in point-wise manner at points $\boldsymbol{x}_i \in x^*$, $i = 1, \ldots, n$, (boldface marks realizations) with the following choices within the supra Bayesian approach:

(1) The likelihood function $f(f_s|\tilde{f}, \sigma, x)$ is chosen as log-normal pdf with mean value $\tilde{f}(x)$ and variance σ. For concise notation, we have dropped explicit mentioning of $\tilde{f}$ being a function of x and replaced it by conditioning on x. This notation will be used in the sequel.

(2) The prior pdf $f(\tilde{f}|\sigma)$ on values $\tilde{f}$ is chosen as improper uniform pdf on positive real axis and prior pdf on the parameter σ as exponential pdf $f(\sigma) \propto \exp(-\beta\sigma)$. The symbol $\propto$ denotes equality up to the normalizing constant.
The choice of a proper prior pdf is necessary for obtaining finite values of the estimate, cf. Remark 1.

The above mentioned model was chosen from several candidates, see discussion of its choice in [Kárný et al., 2008].

The posterior pdf $f(\tilde{f}|f_{1:S}, x)$ is obtained via the Bayes rule:

$$f(\tilde{f}|f_{1:S}, x) \propto \int \prod_{s=1}^{S} f(f_s|\tilde{f}, \sigma, x) f(\tilde{f}) f(\sigma). \quad (1)$$

The above formal posterior pdf specifies distribution on infinite-dimensional pdf $\tilde{f}$ treated in non-parametric way. In majority of practical applications, we seek for a point estimate of $\tilde{f}$, typically, within a class of parametric parametric pdfs, $f(x|\mathscr{V})$, where $\mathscr{V} \in \mathscr{V}^*$ is their parameter. Following the Bayesian approach, the best point estimate within the parametric class $f(x|\mathscr{V})$ is obtained by minimizing expected Kerridge inaccuracy, [Bernardo, 1979]:

$$\hat{\mathscr{V}} = \arg \min_{\mathscr{V} \in \mathscr{V}^*} - \int \hat{f}(x) \ln f(x|\mathscr{V})\, dx \quad (2)$$
$$\hat{f}(x) \equiv \mathsf{E}(\tilde{f}(x)|f_{1:S}, x),$$
$$\mathsf{E}(\cdot|\cdot) \equiv \text{ expectation given by the pdf (1).}$$

The resulting pdf $f(x|\hat{\mathscr{V}})$ will be called the *merger*. For example, the result of minimization (2) for pdfs with support in finite number of points $x \in \boldsymbol{x}_{1:n} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is

$$f(x = \boldsymbol{x}_i|\hat{\mathscr{V}}_{1:S}) \propto \mathsf{E}(\tilde{f}|f_{1:S}, \boldsymbol{x}_i) = \hat{f}(x = \boldsymbol{x}_i) \quad (3)$$
$$\hat{\mathscr{V}}_{1:S} \equiv \{f_{1:S}(x), x \in \boldsymbol{x}_{1:n}\}.$$

Evaluation of the conditional expectation of the unknown pdf as a function of x is non-trivial. An approximate formula for it was given in [Kárný et al., 2008]. Exact formulas for $S = \{2, 3\}$ were derived in [Šmídl, 2008]. It holds:

$$f(\boldsymbol{x}_i|\hat{\mathscr{V}}_{1:2}) \propto \exp\left\{\hat{\mu}_i + \sqrt{2\beta\lambda_i}\left(1 - \sqrt{\tfrac{4\beta-3}{4\beta}}\right)\right\}, \quad (4)$$

$$f(\boldsymbol{x}_i|\hat{\mathscr{V}}_{1:3}) \propto \frac{\mathscr{B}\left(0, \sqrt{2\beta\lambda_i}\sqrt{\frac{3\beta-2}{3\beta}}\right)}{\mathscr{B}\left(0, \sqrt{2\beta\lambda_i}\right)} \exp(\hat{\mu}_i). \quad (5)$$

Here, $\mathscr{B}$ denotes a modified Bessel function of the second kind, $\hat{\mu}_i = \frac{1}{S}\sum_{s=1}^{S} \ln f_s(\boldsymbol{x}_i)$, and λ_i is a remainder after least squares, $\lambda_i = \sum_{s=1}^{S} \ln^2(f_s(\boldsymbol{x}_i)) - S\hat{\mu}_i^2$. Analytical formulas for $S > 3$ can be also obtained using software for symbolic mathematics.

Remark 1. (Choice of β). The parameter β, determining prior pdf on σ occurring in the adopted log-normal model, is the only free parameter of the proposed merging. It can be shown, that $\lim_{\beta\to\infty} \hat{f}(\boldsymbol{x}_i) \propto \exp(\hat{\mu}_i)$. This asymptotic merger

$$\lim_{\beta\to\infty} \hat{f}(x) \propto \exp\left(\tfrac{1}{S}\textstyle\sum_{s=1}^{S} \log f_s(x)\right),$$

coincides with the result of *geometric merging* or (also called logarithmic pooling) [Genest and Zidek, 1986]. It was used in distributed identification by Julier and Uhlmann [1997].

The merger is to be a proper pdf. This determines lower bound on β. At present, just numerical experiments suggest that β should not be chosen below 1. For $\beta \approx 1$, $\hat{f}(x)$ is approaching the result of *arithmetic merging*.

Fig. 1 illustrates influence of β on merging of two Gaussian pdfs, $f_1(x) = \mathcal{N}(-2,1)$ and $f_2(x) = \mathcal{N}(2,2)$. It shows that the gained merger provides a compromise between properties of the two most common arithmetic and geometric merging.

2.2 *Merging of lower dimensional and conditional pdfs*

Consider the case when respective sources inform about parts of multivariate variable x. The sth source provides the pdf $f_s(x_{s,d}|x_{s,c})$ concerning of sub-selection $x_{s,d}$ of x-entries conditioned on another sub-selection $x_{s,c}$. A part of the x, denoted $x_{s,0}$, may not be modelled by the source at all. Each source may have different partitioning of x. In order to combine all sources via the method described in Section 2.1, each source must be extended to the common variable x using the chain rule:

$$f_s(x) = \overline{f}_s(x_{s,0}|x_{s,d},x_{s,c}) f(x_{s,d}|x_{s,c}) \overline{f}_s(x_{s,c}),\ s = 1,\dots,S. \quad (6)$$

The choice of the pdf $\overline{f}(\cdot)$ used for the extensions (6) has significant influence on the resulting combined pdf. Following the Bayesian approach, the unknown extensions are to be optimized by minimization of the Kerridge inaccuracy, in same spirit as in (2). It can be shown that the optimal extension is constructed using the optimal merger. This converts the formula for the merger into an implicit equation. The resulting implicit equation can be solved by successive approximation as follows:

Algorithm 1. (Merging of fragmental sources).

(1) Obtain initial guess of the merger $\hat{f}^0(x)$, set maximum number of iterations $j_{\max}$, and set iterative counter $j = 0$,
(2) For each source, relating $x_{s,d}, x_{s,c}$ via $f_s(x_{s,d}|x_{s,c})$, $s \in \{1\dots S\}$,
– factorize the current merger $\hat{f}^j(x)$ such that

$$\hat{f}^j(x) = \hat{f}^j(x_{s,0}|x_{s,d},x_{s,c}) \hat{f}^j(x_{s,d}|x_{s,c}) \hat{f}^j(x_{s,c}). \quad (7)$$

– Create the sth extended source by complementing it by factors of the current merger on the non-modelled variables and variables in condition:

$$\hat{f}_s^j(x) = \hat{f}^j(x_{s,0}|x_{s,d},x_{s,c})\, f_s(x_{s,d}|x_{s,c})\, \hat{f}^j(x_{s,c}). \quad (8)$$

(3) Combine extended sources $\hat{f}_s^j(x)$ into $\hat{f}^{j+1}(x)$ using (3).
(4) Stop if $\hat{f}^j$ converged or $j \geq j_{\max}$. Otherwise set $j = j+1$ and go to step (2).

Convergence of the algorithm can be checked, for example, by checking if a statistical divergence of $\hat{f}^j(x)$ on $\hat{f}^{j+1}(x)$ is smaller than a chosen threshold. At present, no proof of convergence of the algorithm is known. The influence of the initial guess pdf $\hat{f}^0(\cdot)$ on the result is also not known. Simulation studies of the algorithm, Šmídl [2008], suggest that:

- The algorithm cannot provide unique results when merging sources with important information missing. For example, when merging sources $f_1(a|b,c)$ and $f_2(a)$, the extension on $f(b,c)$ is arbitrary and the solved equation lacks a unique solution. However, even in this case, projection $\hat{f}(a|b,c)$ of the arbitrary merger $\hat{f}(a,b,c)$ typically converges to a single solution.
- The algorithm converges to a unique solution when degenerated cases mentioned above are structurally excluded.

3. PROJECTION TO WEIGHTED EMPIRICAL DENSITIES

In principle, the optimal non-parametric merger can be projected to any class of pdfs using formula (2). The implied optimization problem is, however, difficult to solve for practically important families, such as mixtures of Gaussians. We have already noted that (2) is analytically tractable for pdfs with support on a finite number of discrete points, (3). The challenge is to position these points into areas of high density of the merger. Here, we design a procedure based on importance sampling and mixtures of exponential family pdfs.

3.1 *Importance sampling*

Importance sampling, [Gilks et al., 1996], refers to techniques for generating an empirical approximation of a pdf $f(x)$:

$$f(x) \approx f(x|\boldsymbol{x}_{1:n}) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - \boldsymbol{x}_i), \quad (9)$$

where $\boldsymbol{x}_i$, $i = 1,\dots,n$ are independent identically distributed samples from the pdf $f(x)$ and $\delta(\cdot)$ denotes the Dirac δ-function. The approximation (9) is feasible only if we can sample from the exact pdf $f(x)$. Otherwise, we can draw samples from a chosen proposal pdf (importance function), $q(x)$, yielding approximation in the form of a *weighted* empirical pdf:

$$f(x|\boldsymbol{x}_{1:n}) \approx \kappa^{-1} \sum_{i=1}^{n} w_i \delta(x - \boldsymbol{x}_i), \quad (10)$$

$$w_i = f(\boldsymbol{x}_i)/q(\boldsymbol{x}_i). \quad (11)$$

Under this *importance sampling* procedure, the approximated pdf $f(x)$ need only be evaluated point-wise. Furthermore, normalizing constant of $f(\cdot)$ is not required, since (10) can be normalized trivially via the constant $\kappa = \sum_{i=1}^{n} w_i$.

Note that this procedure alone can be used to approximate the merger $\hat{f}(x)$ when all sources describe the same variable, Section 2.1. Samples $\boldsymbol{x}_i$ can be drawn from any proposal pdf on the same support as the sources, weighted using (11) with the numerator $\hat{f}(\boldsymbol{x}_i)$. The resulting weighted empirical pdf converges to the non-parametric merger for $n \to \infty$. Convergence rate of the procedure depends on closeness of the proposal pdf to the merger.

Extension of this procedure to the case with fragmental sources is challenging since Algorithm 1 requires partitioning of the merger into marginal and conditional pdfs and evaluation of those pdfs at points $\boldsymbol{x}_i$. Due to the involved Dirac functions, these operations are poorly defined and can be evaluated only approximately. Kernel smoothing, [Wand and Jones, 1995], is a typical solution to this problem. It positions a kernel function around each sample and provides the approximation in the form of a mixture model. Drawbacks of this approach are high computational cost for large number of samples and the need for an adequate choice of the kernel. An alternative approach—used e.g. in [Bishop et al., 1998]—is to choose a mixture model with a low number $K \ll n$ of parametric components and fit the mixture to the empirical pdf using one of many available techniques. However, an algorithm for fitting *weighted* empirical pdfs is not available. It is derived below.

3.2 *Fitting mixture models to weighted empirical pdf*

Fitting any parametric model to a non-weighted empirical pdf is equivalent to Bayesian estimation of the model parameters using $\boldsymbol{x}_{1:n}$ as data. Point Bayesian estimation of mixture model parameters can be solved using the EM algorithm and its variants [Titterington et al., 1985]. Here, we outline an algorithm based on the Quasi-Bayes (QB) approach [Kárný et al., 2005].

The algorithm is suitable for mixture of pdfs from exponential family, but for simplicity it will be derived for mixtures of Gaussian pdfs (components)

$$f(x|\mathscr{V}_m) = \sum_{k=1}^{K} \alpha_k \mathscr{N}(\mu_k, \Sigma_k). \tag{12}$$

Here, $\mathscr{V}_m \equiv \{\alpha_k,\, \mu_k,\, \Sigma_k\}_{k=1}^{K}$, where μ_k denote mean value and Σ_k covariance matrix of the kth Gaussian component, α denotes vector of component weights $\sum_{k=1}^{K} \alpha_k = 1,\ \alpha_k \geq 0$.

Estimation of the mixture model via the QB algorithm is based on defining internal (latent) variable $l_{k,i}$ such that $l_{k,i} = 1$ if the ith data record, $\boldsymbol{x}_i$, was generated by the kth component of the mixture and $l_{k,i} = 0$ otherwise. Assuming that we know that the $\bar{k}$th component was used to generate ith data sample, likelihood of the ith sample is the Gaussian pdf: $f(\boldsymbol{x}_i|\alpha, \mu, \Sigma, l_{\bar{k},i} = 1) = f(\boldsymbol{x}_i|\mu_{\bar{k}}, \Sigma_{\bar{k}})$. The advantage of this parameterization is that estimation of parameters μ_k, Σ_k is trivial if $l_{k,i}$ is known for each i. Conjugate pdf to a Gaussian likelihood with unknown mean and covariance matrix is Gauss-inverse-Wishart pdf. Hence, with appropriate choice of prior, the posterior pdf on parameters of the kth component, $f(\mu_k, \Sigma_k|\boldsymbol{x}_{1:n})$ is of the Gauss-inverse-Wishart type [Peterka, 1981] with sufficient statistics:

$$V_k = \sum_{i=1}^{n} l_{k,i} \begin{bmatrix} \boldsymbol{x}_i \\ 1 \end{bmatrix} [\boldsymbol{x}_i{}'\ 1] + V_{k,0},\ \ \nu_k = \sum_{i=1}^{n} l_{k,i} + \nu_{k,0}, \tag{13}$$

where $V_{k,0}$, $\nu_{k,0}$ parameterize prior Gauss-inverse-Wishart pdf.

The QB algorithm substitutes the unknown $l_{k,i}$ in (13) by their expected value, $l_{k,i} \equiv \hat{l}_{k,i}$, re-estimates posterior pdf μ_k, Σ_k, then re-estimates $\hat{l}_{k,i}$ and so on until convergence. The same idea applied to the case of *weighted* empirical pdfs yields posterior pdfs on the mixture component parameters in the form of Gauss-inverse-Wishart pdf with sufficient statistics:

$$V_k = \sum_{i=1}^{n} \hat{l}_{k,i} w_i \begin{bmatrix} \boldsymbol{x}_i \\ 1 \end{bmatrix} [\boldsymbol{x}_i'\ 1] + V_{k,0},\ \ \nu_k = \sum_{i=1}^{n} \hat{l}_{k,i} w_i + \nu_{k,0}. \tag{14}$$

This formula achieves the same effect as mixture estimation on non-weighted empirical pdf obtained by re-sampling operation [Doucet et al., 2001], see Šmídl [2008] for discussion.

The parameter $\hat{\mathscr{V}}_m = \{\hat{\alpha}_k, \hat{\mu}_k, \hat{\Sigma}_k\}_{k=1}^{K}$ minimizing (approximately) the inaccuracy (2) within the mixture class (12) is composed of expected values, which are computed from the statistics (14), [Peterka, 1981].

3.3 Final Merging Algorithm

Using building blocks designed and reviewed above, the proposed log-normal merging of fragmental information applicable to high-dimensional case is described by the following algorithm.

Algorithm 2.

(1) Obtain all sources, $f_s(x_{s,d}|x_{s,c})$, and specify statistics $\hat{\mathscr{V}}_m^0$ characterizing initial guess of the mixture merger (12). Set maximum iteration count $j_{\max} = 1$.
(2) Generate n samples $\boldsymbol{x}_{1:n}$ from the mixture $f(x|\hat{\mathscr{V}}_m^j)$ and store the values $f(\boldsymbol{x}_i|\hat{\mathscr{V}}_m^j)$, $i = 1, \ldots, n$.
(3) Call Algorithm 1 with $\hat{f}^j(x) = f(x|\hat{\mathscr{V}}_m^j)$, samples $\boldsymbol{x}_{1:n}$ and stopping time $j_{\max}$ to obtain $f(\boldsymbol{x}_i|\hat{\mathscr{V}}_{1:S})$, $i = 1, \ldots n$, (3).
(4) Evaluate weights (11) of the empirical approximation of the merger $w_i \propto f(\boldsymbol{x}_i|\hat{\mathscr{V}}_{1:S})/f(\boldsymbol{x}_i|\hat{\mathscr{V}}_m^j)$.
(5) Obtain $(j+1)$th mixture approximation of the merger, $f(x|\hat{\mathscr{V}}_m^{j+1})$, (12), using the modified QB algorithm, Section 3.2.
(6) Stop, if the algorithm has reached a steady state. Otherwise, goto step (2).

Two unspecified parts of the Algorithm 2 are how to choose the initial proposal pdf and how to stop the algorithm. In practical application, the initial proposal may be designed using expert knowledge, or alternatively a Gibbs-sampling-like scheme can be designed [Šmídl, 2008]. The stopping rule of the algorithm is based on comparing divergence between the current and previous approximating mixture.

Low maximum number of iterations $j_{\max}$ (even 1) in Algorithm 1 supports overall numerical stability. When the initial guess $\hat{f}^j(x)$ is far from the optimal, the values of the merger in many samples may be close to zero and only a few become significant in iterations. In a pathological situation, only one sample receives non-zero weight causing numerical instability. When the initial guess is known to be close to the expected optimum, the value of $j_{\max}$ can be increased, or it can be be estimated using e.g. efficient sample size [Doucet et al., 2001].

Remark 2. (Identification scenarios). For simple scenarios with a limited number of known types of pdfs, an agent evaluating algorithm Algorithm 2 can receive only statistics of the pdfs.

However, in highly heterogeneous environments, it may be advantageous to design a special purpose mediating agent performing Algorithm 2 and distributing evaluation of the third step—i.e. Algorithm 1—between agents. Specifically, the samples generated in step (2) are send to the neighbors which evaluate values of their local pdfs in given points and send the resulting empirical pdf back to the mediator. The mediator use these values within step (3), equation (8) of Algorithm 1. The remaining steps are performed by the mediator. This scenario is illustrated in Figure 2, where the mediating agent, A_3, is merging pdfs from local agents A_1 and A_2.

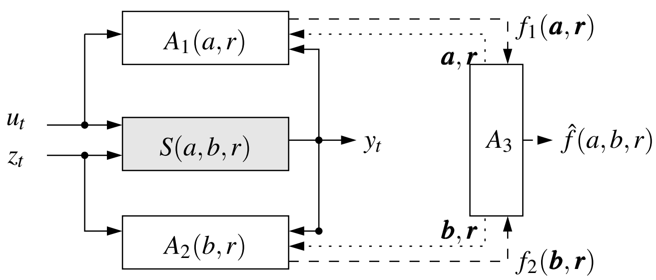

Figure 2. Fully distributed identification. System S with parameters a, b, r is identified by local agents A_1 and A_2 which are capable of observing only part of the data space, y_t, u_t and y_t, z_t, respectively. These agents build their posterior pdfs $f(a, r|y_{1:t}, u_{1:t})$ and $f(b, r|y_{1:t}, z_{1:t})$. Merging of these statistics is mediated by agent A_3, which operates on common support $x = [a, b, r]$. It generates samples $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}$ and sends their appropriate sub-selections to local agents A_1 and A_2 (dotted lines). Local agents evaluates values of their statistics at these points and send them back to A_3 (dashed lines). The merger $\hat{f}(x) = \hat{f}(a, b, r)$ can be provided in either empirical or mixture form.

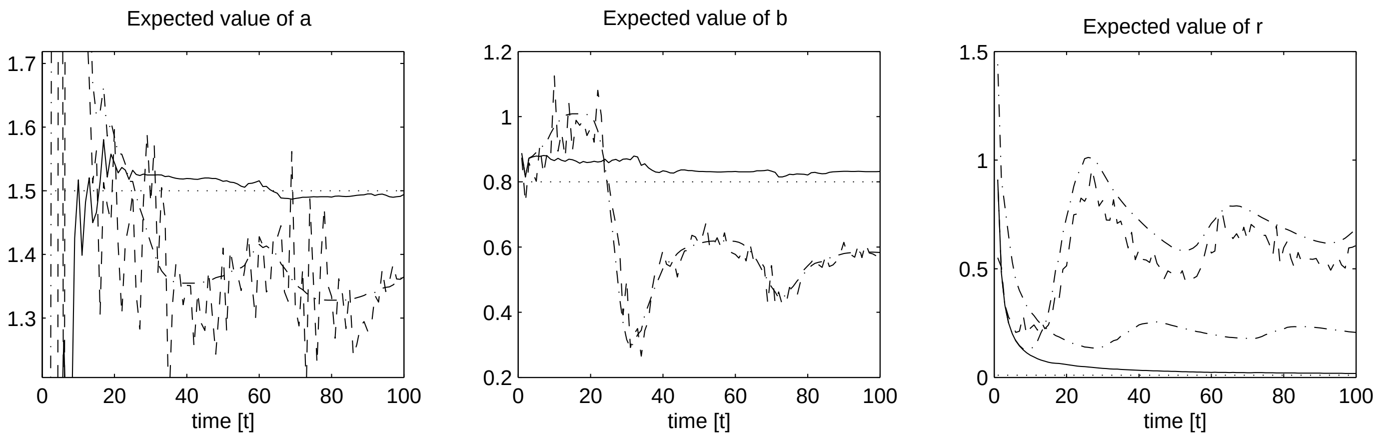

Figure 3. Expected values of the estimated parameters using: centralized approach (solid line), local estimates (dash-dotted line), and supra Bayesian merging of fragmental pdfs with $\beta = 1.5$ (dashed line).

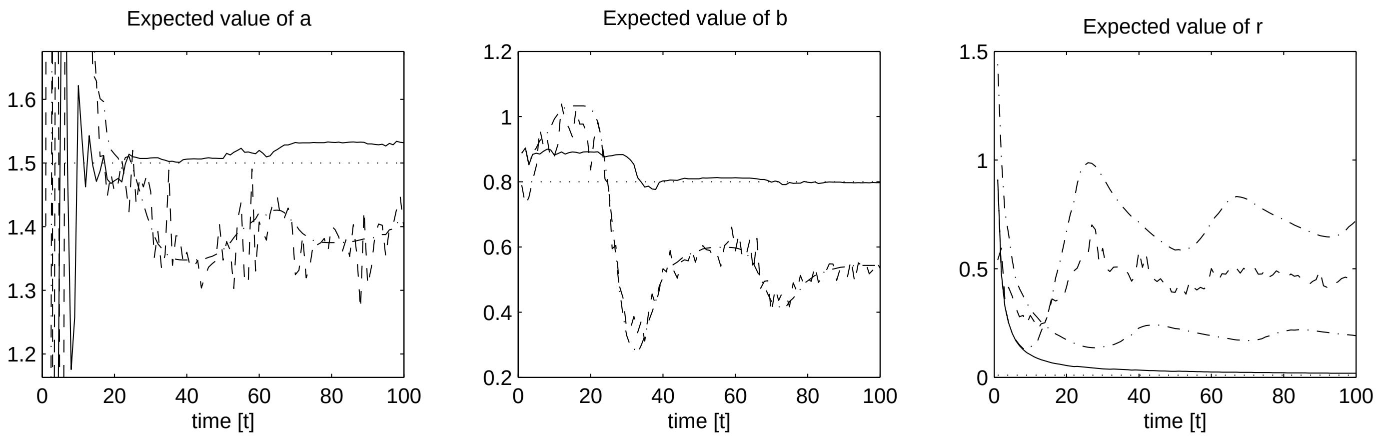

Figure 4. Expected values of the estimated parameters using: centralized approach (solid line), local estimates (dash-dotted line), and supra Bayesian merging of fragmental pdfs with $\beta = 1000$ (dashed line).

4. DISTRIBUTED IDENTIFICATION OF A CONTROLLED REGRESSIVE MODEL

The proposed algorithm was applied to the task of identification of the following controlled regressive model

$$y_t = au_t + bz_t + e_t, \tag{15}$$

where y_t, u_t, z_t are observed data, a, b are unknown regression coefficients and noise e_t is Gaussian with zero mean and unknown variance r. The task is to evaluate posterior pdf $f(a,b,r|y_{1:t}, u_{1:t}, z_{1:t})$. The process is observed by two agents: A_1 is capable of observing y_t and u_t, and A_2 is capable of observing y_t and z_t. They report their (fragmental) posterior pdfs to agent A_3, which merges them using Algorithm 2. In the centralized (full information) setup, the posterior $f(a,b,r|\cdot)$ is of the Gauss-inverse-Wishart form with statistics

$$V_t = \sum_{\tau=1}^{t} \begin{bmatrix} y_\tau^2 & y_\tau u_\tau & y_\tau z_\tau \\ y_\tau u_\tau & u_\tau^2 & u_\tau z_\tau \\ y_\tau z_\tau & u_\tau z_\tau & z_\tau^2 \end{bmatrix} + V_0, \quad \nu_t = \nu_0 + t. \tag{16}$$

Exact marginals $f(a,r|\cdot)$ and $f(b,r|\cdot)$ of $f(a,b,r|\cdot)$ are of the Gauss-inverse-Wishart form with ν_t as in (16) and statistics

$$V_{1,t} = \sum_{\tau=1}^{t} \begin{bmatrix} y_\tau^2 & y_\tau u_\tau \\ y_\tau u_\tau & u_\tau^2 \end{bmatrix} + V_{1,0}, \; V_{2,t} = \sum_{\tau=1}^{t} \begin{bmatrix} y_\tau^2 & y_\tau z_\tau \\ y_\tau z_\tau & z_\tau^2 \end{bmatrix} + V_{2,0}. \tag{17}$$

Thus, agents A_1 and A_2 can collect these statistics and report exact marginals to agent A_3. Note that statistics $\sum_\tau u_\tau z_\tau$ is not collected by any agent and the merging procedure must compensate for that.

A naive approach for A_3 could be to merge marginal pdfs from A_1 and A_2 under the assumption of mutual independence of u_t and z_t. Thus, A_3 can build its statistics $V_{3,t}$ using elements from $V_{1,t}$ and $V_{2,t}$ where appropriate and take the expected value of the missing term. However, the resulting matrix V_3 may be negative definite, disqualifying it from being a statistics of Gauss-inverse-Wishart pdf. Hence, an optimization procedure would have to be derived that optimizes the missing value such that V_3 remains positive definite. The proposed merging scheme achieves this automatically.

A simulation experiment of 100 data generated by model (15) was performed with $a = 1.5$, $b = 0.8$, $r = 0.01$, and

$$u_t = \sin^3\left(\frac{\pi}{40}t\right), \; z_t = \sin^3\left(\frac{\pi}{40}t + \frac{\pi}{10}\right).$$

The point estimates obtained by the merging algorithm for three components and two choices of β, $\beta = 1.5$ and $\beta = 1000$, are displayed in Fig. 3 and Fig. 4, respectively, via posterior expected values of the unknown parameters.

Note that the merging algorithm preserves expected values of the parameters unique to agents A_1 and A_2, i.e., a and b, respectively, and merges the expected value of the common parameter r. With increasing value of β the merged expectation of r is decreasing. For $\beta = 1.5$, the merged variance almost coincide with its expectation provided by agent A_2. Note that due to the missing part of the statistics, $u_\tau z_\tau$, mutual correlation between u_τ and z_τ is not recognized. Hence, each agent considers the unobserved input as a part of the disturbance e_t, which results in a biased estimates of a and b, and increased expected value of variance σ.

Uncertainty bounds on the estimates are not displayed in Figures for clarity. Terminal variances of the estimates for $\beta = 1000$ at $t = 100$ were:

$$\begin{aligned} A_1 &: \mathrm{var}(a) = 0.006,\ \mathrm{std}(a) = 0.07,\ \mathrm{var}(r) = 0.002, \\ A_2 &: \mathrm{var}(b) = 0.025,\ \mathrm{std}(b) = 0.16,\ \mathrm{var}(r) = 0.007, \\ A_3 &: \mathrm{cov}(a,b) = \begin{bmatrix} 0.016 & 0.003 \\ 0.003 & 0.027 \end{bmatrix},\ \mathrm{var}(r) = 0.008. \end{aligned}$$

Note that standard deviation of the posterior estimate of a is relatively small compared to the bias in the expected value, Figure 4. The merged variance on a is, however, significantly increased, while the merged variance of b was almost preserved. Thus the error caused by the missing statistics was compensated in this way.

Source code for the method and this example are available: `http://mys.utia.cas.cz:1800/trac/bdm`.

5. DISCUSSION AND CONCLUSION

The paper presents a first step development of a fully automatic method of probability merging. Many technical details were necessary to describe the method in full. However, most of those details are routine applications of the probability calculus. The main ideas of the paper are as follows:

- Supra Bayesian approach to merging fragmental probabilistic knowledge pieces is a feasible methodology.
- The best merger is found to be the expectation of the unknown objective pdf describing union of all random variables considered. A fragmental knowledge piece is to be extended in a unique and unambiguous way.
- The evaluating method was designed to be as general as possible. Being a combination of importance sampling with proposal pdf in the form of a mixture of exponential family pdfs, it is ready to be used in various implementations as a black box. The only constraint is the common support of all sources. In case of restricted support, the method has to be extended to accommodate for that e.g. by rejection sampling.

The resulting merging methodology has a wide scope of applications ranging from elicitation of prior pdfs from expert opinions over distributed system identification and data fusion to fully decentralized scenarios of cooperative agents.

The presented application was rather simplistic, its purpose was to demonstrate that the algorithm is capable of merging arbitrary pdfs, such as marginals of Gauss-inverse-Wishart pdf. For this particular example, an algorithm achieving similar performance at lower cost can be designed. The value of our approach is that the merging was achieved *automatically* without the need for analyzing sufficient statistics and their optimization.

The price for generality of the method is its high computational cost. Many simplifications can be made to lower the cost. For example, when the merger is supposed to be in a specific class–e.g. Gaussian–the weighted empirical pdf can be fitted by a Gaussian pdf which is much cheaper operation compared to fitting mixtures.

The methodological foundations of the approach are almost complete, the only missing part is proof of the convergence of the iterative algorithm. Alternatively, a methodology for choosing parameter β can be designed. However, we expect that most of the future work is in application of the method to fully distributed scenarios for which it was developed and designing more efficient approximations of the proposal pdf.

ACKNOWLEDGEMENTS

This research was supported by grants: MŠMT 1M0572, GAČR 102/08/0567, and GAČR 102/08/P250.

REFERENCES

P. Alriksson. *State Estimation for Distributed and Hybrid Systems*. PhD thesis, Lund University, 2008.

J. M. Bernardo. Expected infromation as expected utility. *The Annals of Statistics*, 7(3):686–690, 1979.

C.M. Bishop, M. Svensen, and C.K.I. Williams. GTM: The Generative Topographic Mapping. *Neural Computation*, 10: 215–234, 1998.

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

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

W.R. Gilks, S. Richardson, and DJ Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1996.

S.J. Julier and J.K. Uhlmann. A non-divergent estimation algorithm in the presence of unknown correlations. In *Proceedings of the American Control Conference*, volume 4, 1997.

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.

M. Kárný, T.V. Guy, A. Bodini, and F. Rugeri. Cooperation via sharing of probabilistic elements. Technical Report 9-MI, IMATI Milano, 2008. URL `http://www.mi.imati.cnr.it/iami/abstracts/08-09.html`.

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

M. Rosencrantz, G. Gordon, and S. Thrun. Decentralized sensor fusion with distributed particle filters. In *Uncertainty in Artificial Intelligence*, 2003.

V. Šmídl. Sequential sampling techniques for log-normal combination of probability densities. Technical Report 2244, ÚTIA, AV ČR, 2008.

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

J. Speyer. Computation and transmission requirements for a decentralized linear-quadratic-Gaussian control problem. *Automatic Control, IEEE Transactions on*, 24(2):266–269, 1979.

D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixtures*. John Wiley, New York, 1985.

M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall, 1995.