Bayesian Transfer Learning between Uniformly Modelled Bayesian Filters *

Ladislav Jirsa[1], Lenka Kuklišová Pavelková[1], and Anthony Quinn[1,2]

[1] The Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, Czech Republic {pavelkov, jirsa, aquinn}@utia.cas.cz
http://www.utia.cas.cz
[2] Trinity College Dublin, the University of Dublin, Ireland

Abstract. We investigate sensor network nodes that sequentially infer states with bounded values, and affected by noise that is also bounded. The transfer of knowledge between such nodes is the principal focus of this chapter. A fully Bayesian framework is adopted, in which the source knowledge is represented by a bounded data predictor, the specification of a formal conditioning mechanism between the filtering nodes is avoided, and the optimal knowledge-conditioned target state predictor is designed via optimal Bayesian decision-making (fully probabilistic design). We call this framework Bayesian transfer learning, and derive a sequential algorithm for pairs of interacting, bounded filters. To achieve a tractable, finite-dimensional flow, the outputs of the time step, transfer step and data step are locally projected onto parallelotopic supports. An informal variant of the transfer algorithm demonstrates both strongly positive transfer of high-quality (low variance) source knowledge—improving on a former orthotopically supported variant—as well as rejection of low-quality (high variance) source knowledge, which we call robust transfer.

Keywords: Bayesian transfer learning · fully probabilistic design · Bayesian filtering · uniform noise · parallelotopic bounds · robust transfer.

1 Introduction

Transfer learning relates to contexts where a computer-based learning task undertaken by a processing node—called the *target node*—is enhanced in performance and/or efficiency if the results of a related task—undertaken by a distinct *source task*—are made available to it. Within this generic formulation, there is significant current interest in the defining and implementation of transfer learning schemes for a range of machine intelligence [34] and big data problems [42]. One such important focus is in distributed knowledge representation and processing in sensor networks, addressing IoT and Industry 4.0 priorities [6]. Local inferences, driven by data in the individual sensors, are merged or fused in order to design a global inference [21]. Applications range from health monitoring [38]

* This research has been supported by GAČR grant 18-15970S.

and environmental sensing [43], to cooperative systems for indoor tracking [5] and urban vehicular localization [26]. A multi-disciplinary literature on multi-modal and multi-sensor data fusion has evolved, with overviews available in [23] and [13], for example. Many of those methods and applications can be well characterized as transfer learning for data-driven inference tasks.

The technical approaches adopted for problems of data-driven knowledge transfer and fusion are as diverse as the applications themselves. Wiener-type moment estimation for knowledge representation is adopted in the measurement vector fusion method of [39], as well as in sequential Kalman filter variants for fusion of measurements [7] and states [41]. A Dempster-Shafer evidence-based approach to data fusion has recently been outlined [40]. Traditional machine learning and artificial intelligence tools have, of course, also been deployed, such as support vector machines and neural networks [35], clustering methods [38], functional regression [32], and expert system approaches [25].

Our focus in this chapter is on the Bayesian framework for data-driven knowledge representation and decision-making for transfer learning in sensor networks. In this context, sensors convert local data harvests into probability distributions of quantities of interest, and the computed distributions are transferred from source to target tasks (i.e. sensor nodes), to improve inference in the latter. This objective is related to distributional pooling [1], insofar as local knowledge representations are probabilistic, although pooling criteria are not generally deducible from formal Bayesian-theoretic principles. More recently, fully Bayesian approaches to distributional pooling have been derived, for instance in the context of centralized pooling [2]. A Bayesian transfer learning strategy has been reported for static [10] and dynamic [28] transfer between Kalman filters. In those works, the transfer step is induced between the time- and data-steps of isolated Kalman filtering. Our aim in this chapter is to make progress with the extension of these methods—which we call fully probabilistic design (FPD) [30]—to important classes of filtering tasks beyond the linear-Gaussian case.

In particular, there is considerable motivation in sensor network contexts in investigating *bounded* innovations and/or observation processes [11, 14, 36]. Under Gaussian distributional assumptions, one approach has been to project state estimates onto the constraint surface [9], or to truncate the distribution [33]. Using sequential Monte-Carlo sampling methods, constraints are imposed via the accept/reject steps of the algorithm [24]. Meanwhile, non-probabilistic techniques for handling "unknown-but-bounded errors" seek to design an approximate set containing the state estimates [3, 4].

The current authors have adopted a strictly Bayesian framework for linear state filtering with *uniform* stochastic drivers (LSU filtering). Specifically, in [29], the filtering distribution is derived via local distributional approximations which sequentially project the output of the data step and time step into the class of uniform distributions on orthotopic supports (UOS). We will recall this setting in Section 2. In recent work, we have developed a Bayesian knowledge transfer mechanism between pairs of these UOS filters, adopting the axiomatically optimal approach of fully probabilistic design (FPD) [30], as already mentioned

above. An important characteristic of this approach is that it does not require specification of a complete stochastic dependence model between the source and target tasks. While complete modelling is the conventional approach, allowing knowledge transfer via standard Bayesian conditioning [18], it can be difficult to elicit a complete (hierarchical) model of inter-task dependence, leading to model robustness issues. Instead, in our work, the source-knowledge-conditional model is designed in an optimal Bayesian decision-theoretic sense. We refer to this FPD-optimal approach to conditioning for knowledge transfer as Bayesian transfer learning [27].

The knowledge transferred from the source to target filters is represented via the one-step-ahead (data) predictor [31]. This was derived for the UOS filter in [15]. The derivation of the approximate sequential statistics, and the performance of the FPD-optimal transfer algorithm, between UOS filters has recently been reported [17].

In this chapter, we report the following progress with FPD-optimal Bayesian transfer learning between LSU filtering nodes:

(i) The local projections into the UOS class (after each data and time step) have no guarantees bounding the sequential model approximation error. In particular, the projection of the filtering distribution and the state predictor into the orthotopic support is loose. We have recently derived an algorithm for closure of the *isolated* LSU distributions (state filtering and prediction) within the class of uniform distributions on (more flexible) parallelotopic supports (UPS), which we will also review in Section 2. In this chapter, we will extend the UOS-closed Bayesian transfer learning algorithm of [17] to UPS-closure. We anticipate tighter modelling approximations and correspondingly improved filtering performance.

(ii) For the purpose of (i), we must derive the one-step-ahead data predictor via which the UPS-closed (source) filter transfers its knowledge sequentially to the target filter. We therefore adapt the (scalar) data predictor derived for the UOS-closed filter [15] to this more flexible class.

(iii) In [17], our simulations indicated only very modest improvements in filtering (tracking) accuracy of the UOS-closed target filter under FPD-optimal transfer from the UOS source filter. By examining the geometric nature of this data-predictive transfer, we were able to propose an informal transfer method involving intersections between the interval supports of the (scalar) source and target data predictors [17]. In this chapter, we will again develop this alternative—now in the context of the UPS-closed filters—as a variant for comparison with the formal (FPD-optimal) transfer method. In Section 5, we will present simulation evidence of the comparative performance of these transfer learning algorithms.

The paper is organised as follows. In Section 2, the UOS and UPS classes are introduced. Then, the optimal local distributional projections (i.e. approximations) within both the UOS and UPS class of data and time updates are presented, as well as the relevant predictors. Section 3 defines the problem of

transfer learning in a pair of Bayesian filters and proposes a tractable (i.e. finite-dimensional and recursive) framework for FPD-optimal knowledge transfer of the source data predictor. This framework is then applied to (i) a pair of UOS filters and (ii) a pair of UPS filters. In Section 3.2, the informal knowledge transfer method between a pair of UOS filters—inspired by a support-intersection property of FPD [17]—is recalled, and deployed here, too, to transfer between a pair of UPS filters. Section 5 reports the comparative performances of the proposed transfer learning algorithms in the context of UOS and UPS filter pairs.

The following notation is used: z_t is the value of a column vector, z, at a discrete time instant, $t \in t^\star \equiv \{1, 2, \ldots, \overline{t}\}$; $z_{t;i}$ is the i-th entry of z_t; $z(t) \equiv \{z_t, z_{t-1}, \ldots, z_1, z_0\}$; $\equiv$ denotes equality by definition; $\propto$ denotes equality up to a constant factor. Matrices are capitalized (e.g. A), vectors and scalars are lowercase (e.g. b). A_{ij} is the element of matrix A in the i-th row and j-th column. A_i denotes the i-th row of A. ℓ_z denotes the length of a (column) vector, z, and $\mathbb{Z}$ denotes the defined set of z. I is the identity matrix. $\chi_z(\mathbb{Z})$ is the set indicator, which equals 1 if $z \in \mathbb{Z}$ and 0 otherwise. The p-norm $\|z\|_p = \left(\sum_{i=1}^{\ell_z} z_i^p\right)^{1/p}$, particularly, $\|z\|_2$ is an Euclidean norm and $\|z\|_\infty = \max_i(|z_i|)$. No formal distinction is made between a random variable and its realisation.

2 Bayesian filtering and prediction with uniformly distributed noise processes

In this section, Bayesian filtering and prediction are introduced and applied to a state-space model with uniformly distributed noises (which we call the LSU model). Two cases are distinguished: projection of the states into (i) the class of uniform distributions on an orthotopic support (i.e. the UOS class); and into (ii) the class of uniform distributions on a parallelotopic support (i.e. the UPS class), respectively.

2.1 UOS and UPS class definition

We consider a finite-dimensional vector random variable z with realisations in a bounded subset of R^{ℓ_z}. We now define appropriate support sets in a ℓ_z-dimensional space.

A *polytope* is a bounded set defined (bounded) by a finite number of flat facets. In this paper, we specialise this to the following types of convex polytope.

A *zonotope* $\mathbb{Z}_Z$ is a convex polytope formed by the intersection of k strips, $k \geq \ell_z$. It can be expressed as

$$\mathbb{Z}_Z = \{z : a \leq Vz \leq b\}, \tag{1}$$

where a and b are vectors of length k, of lower and upper bounds, respectively, which are meant entry-wise; V is a matrix of size $k \times \ell_z$ with rank ℓ_z. The i-th

strip is therefore given by the inequality

$$\mathbb{Z}_{S_i} = \{z : a_i \leq V_i z \leq b_i\}. \tag{2}$$

A *parallelotope* $\mathbb{Z}_P$ is a special case of a zonotope (1) with $k = \ell_z$, so that V is a square invertible matrix.

An *orthotope* $\mathbb{Z}_O$ is a special case of the parallelotope with adjacent facets orthogonal. It then holds that $V = I$ in (1). Furthermore, $\underline{z} \equiv a$ and $\overline{z} \equiv b$, being the lower and upper bounds of z, respectively. Then, the orthotope is specified by

$$\mathbb{Z}_O = \{z : \underline{z} \leq z \leq \overline{z}\}. \tag{3}$$

Table 1. The degrees of freedom (dof) and Lebesgue measure, i.e. volume, $\mathcal{V}$, of various convex polytope specialisations.

	dof	volume $\mathcal{V}$
orthotope	$2\ell_z$	$\prod_{i=1}^{\ell_z} (\overline{z}_i - \underline{z}_i)$
parallelotope	$\ell_z(\ell_z + 2)$	$\lvert \det V \rvert^{-1} \prod_{i=1}^{\ell_z} (b_i - a_i)$, see [29]
zonotope	$k(k+2)$, $k > \ell_z$	the sum of the $\mathcal{V}$s of its generating parallelotopes, see [12]

Table 1 presents volumes —equivalent to Lebesgue measure—and degrees-of-freedom (dof) for the above defined support sets. The dof corresponds to the minimal number of geometric parameters that unambiguously defines the mentioned set.

Besides the standard form (1), we introduce another two equivalent descriptions of a parallelotope $\mathbb{Z}_P$ that will be used further. The $[-\mathbf{1}, \mathbf{1}]$-form of parallelotope, equivalent to standard form (1), is defined as

$$\mathbb{Z}_P = \{z : -\mathbf{1}_{(\ell_z)} \leq Wz - c \leq \mathbf{1}_{(\ell_z)}\}, \tag{4}$$

where $\mathbf{1}_{(\ell_z)}$ is a unit vector of length ℓ_z and $W_{ij} = 2V_{ij}/(b_i - a_i)$, $c_i = (b_i + a_i)/(b_i - a_i)$, $i, j = 1, \ldots, \ell_z$. We can transform it back to the standard form (1) using $a = c - \mathbf{1}_{(\ell_z)}$, $b = c + \mathbf{1}_{(\ell_z)}$, $V = W$.

An expression for the parallelotope (4) in terms of its centroid $\hat{z}$ is [37]

$$\mathbb{Z}_P = \{z : z = \hat{z} + T\xi\}, \tag{5}$$

where $T = W^{-1}$, $\hat{z} = Tc$, $\forall \xi$ s.t. $\|\xi\|_\infty \leq 1$.

UPS class — A uniform distribution of z on a parallelotopic support (1), i.e. the UPS distribution, is defined as

$$\mathcal{U}_z(a, b, V) \equiv \mathcal{V}_P^{-1} \chi_z (a \leq Vz \leq b) \tag{6}$$

where $\mathcal{V}$ is given in the second row of Table 1.

The first moment (mean value) of the UPS distribution (6) is

$$\mathsf{E}[z|a,b,V] \equiv \hat{z} = V^{-1}\ \frac{b+a}{2}. \tag{7}$$

The second central moment (covariance) of the UPS is

$$\mathsf{cov}[z|a,b,V] = \frac{1}{3}\ V^{-1}GG'\big(V^{-1}\big)', \tag{8}$$

where $G_{ii} = \frac{b_i - a_i}{2}$ and $G_{ij} = 0$, $i \neq j$, $i,j = 1,\ldots,\ell_z$.

UOS class — A uniform distribution of z on an orthotopic support (3), i.e. the UOS distribution, is defined as

$$\mathcal{U}_z(\underline{z},\overline{z}) \equiv \mathcal{U}_z(a,b,I) = \mathcal{V}_O^{-1}\chi_z\left(\underline{z} \leq z \leq \overline{z}\right) \tag{9}$$

$\underline{z}$=a, $\overline{z}$=b are the lower and upper bound on z, respectively, and $\mathcal{V}$ is given in the first row of Table 1.

The first moment (mean value) of the UOS distribution (9) is

$$\mathsf{E}[z|a,b,I] \equiv \hat{z} = \frac{b+a}{2} \equiv \frac{\underline{z}+\overline{z}}{2}. \tag{10}$$

The second central moment (covariance) of the UPS is

$$\mathsf{cov}[z|a,b,I] = \frac{1}{3}\ GG', \tag{11}$$

where $G_{ii} = \frac{b_i - a_i}{2}$ and $G_{ij} = 0$, $i \neq j$, $i,j = 1,\ldots,\ell_z$.

2.2 Bayesian filtering and the LSU model

In the Bayesian framework [10, 19], the state-space system is expressed stochastically, via the following pdfs:

$$\begin{aligned} \text{prior pdf:} \quad & f\left(x_1\right) \\ \text{observation model:} \quad & f\left(y_t|\,x_t\right) \\ \text{time evolution model:} \quad & f\left(x_{t+1}|\,x_t,u_t\right). \end{aligned} \tag{12}$$

Here, y_t is a scalar observable output, u_t is a known system input (optional, for generality), and x_t is an ℓ-dimensional unobservable system state, $t \in t^*$.

We assume that (i) hidden process x_t satisfies the Markov property, (ii) no direct relationship between input and output exists in the observation model, and (iii) the inputs consist of a known sequence $u_1,\ldots,u_{\overline{t}}$.

Bayesian filtering, i.e., state estimation consists of the evolution of the posterior pdf $f(x_t|d(t))$ where $d(t)$ is a sequence of observed data records $d_t = (y_t, u_t)$, $t \in t^*$. The evolution of $f(x_t|d(t))$ is described by a two-steps recursion that starts from the prior pdf $f(x_1)$) and ends by data update at the final time $t = \overline{t}$:

– Data update (Bayes rule)

$$f(x_t|d(t)) = \frac{f(y_t|x_t)f(x_t|d(t-1))}{\int\limits_{x_t^\star} f(y_t|x_t)f(x_t|d(t-1))\mathrm{d}x_t}, \tag{13}$$

– Time update

$$f(x_{t+1}|d(t)) = \int\limits_{x_t^\star} f(x_{t+1}|u_t, x_t)f(x_t|d(t))\,\mathrm{d}x_t. \tag{14}$$

We introduce a linear state space model with a uniform noise (LSU model) in the form

$$\begin{aligned} f(y_t|x_t) &= \mathcal{U}_y(\tilde{y}_t - r, \tilde{y}_t + r) \\ f(x_{t+1}|x_t, u_t) &= \mathcal{U}_x(\tilde{x}_{t+1} - \rho, \tilde{x}_{t+1} + \rho) \end{aligned} \tag{15}$$

where $\tilde{y}_t = Cx_t$, $\tilde{x}_{t+1} = Ax_t + Bu_t$, A, B, C are the known model matrices/vectors of appropriate dimensions, $\nu_t \in (-\rho, \rho)$ is the uniform state noise with known parameter ρ, $n_t \in (-r, r)$ is the uniform observation noise with known parameter r.

Exact state estimation for the LSU model (15)—following (13) and (14)—leads to an intractable form of posterior pdf, with the dimension of the sufficient statistics unbounded with the number, t, of processed data. In [15, 29], approximate Bayesian state estimation (filtering) within the UOS class is proposed. Recently, [16] has relaxed the orthotopic geometry of the distributional supports, instead proposing an approximate Bayesian filtering flow via sequential local projections of the state predictive and filtering distributions into parallelotopic supports. In both cases, the algorithms are based on projections via minimization of Kullback-Leibler divergence (KLD) [22], achieving closure of the filtering distribution, $f(x_t|d(t))$, within the respective class, at each step of filtering.

2.3 Bayesian filtering and prediction within the UOS class

This section summarizes the results of the filtering and prediction within UOS class as presented in [15, 29].

Data update The data update (13) processes $f(x_t|d(t-1))$ together with the $f(y_t|x_t)$ (15) according to the Bayes rule. It starts in $t = 1$ with $f(x_1) = \mathcal{U}_x(\underline{x}_1^+, \overline{x}_1^+)$. The exact pdf is uniformly distributed on a zonotopic support that results from the intersection of an orthotope (9) obtained during previous time update—or $f(x_1)$ in the first step—and strips given by new data

$$\begin{aligned} f(x_t|d(t)) &\propto \chi(\underline{m}_t - \rho \le x_t \le \overline{m}_t + \rho)\chi(Cx_t - r \le y_t \le Cx_t + r) = \\ &= \chi\left(\begin{bmatrix} \underline{m}_t - \rho \\ y_t - r \end{bmatrix} \le \begin{bmatrix} I \\ C \end{bmatrix} x_t \le \begin{bmatrix} \overline{m}_t + \rho \\ y_t + r \end{bmatrix}\right). \end{aligned} \tag{16}$$

In [15], the following approximation is proposed

$$f(x_t|d(t)) \approx \mathcal{U}_{x_t}(a_t,\, V_t,\, b_t). \tag{17}$$

The resulting pdf belongs to UPS class. Nevertheless, the subsequent time update requires pdf from UOS class. Therefore, the support of (17) is circumscribed by the orthotope so that

$$f(x_t|d(t)) \approx \mathcal{U}_{x_t}(\underline{x}_t,\, \overline{x}_t), \tag{18}$$

Time update The time update (14) processes $f(x_t|d(t))$ (18) together with $f(x_{t+1}|x_t, u_t)$ (15). The exact pdf $f(x_{t+1}|d(t))$ is non-uniformly distributed on a zonotopic support. In [15], the following approximation is proposed

$$f(x_t|d(t-1)) \approx \prod_{i=1}^{\ell} \frac{\chi(\underline{m}_{t;i} - \rho_i \le x_{t;i} \le \overline{m}_{t;i} + \rho_i)}{\overline{m}_{t;i} - \underline{m}_{t;i} + 2\rho_i} =$$

$$= \prod_{i=1}^{\ell} \mathcal{U}_{x_{t;i}}(\underline{m}_{t;i} - \rho_i, \overline{m}_{t;i} + \rho_i) = \mathcal{U}_{x_t}(\underline{m}_t - \rho, \overline{m}_t + \rho), \tag{19}$$

where $\underline{m}_t = [\underline{m}_{t;1}, \ldots, \underline{m}_{t;\ell}]'$, $\overline{m}_t = [\overline{m}_{t;1}, \ldots, \overline{m}_{t;\ell}]'$,

$$\underline{m}_{t;i} = \sum_{j=1}^{\ell} \min(A_{ij}\underline{x}_{t-1;j} + B_i u_{t-1}, A_{ij}\overline{x}_{t-1;j} + B_i u_{t-1}), \tag{20}$$

$$\overline{m}_{t;i} = \sum_{j=1}^{\ell} \max(A_{ij}\underline{x}_{t-1;j} + B_i u_{t-1}, A_{ij}\overline{x}_{t-1;j} + B_i u_{t-1}).$$

Predictive pdf for LSU model The data predictor of a linear state-space model is the denominator of (13), where $f(y_t|x_t)$ is given by (15) and $f(x_t|d(t-1))$ is the result of the approximate time update (19). The approximate uniform predictor as proposed in [15] is

$$f(y_t|d(t-1)) \approx \frac{\chi\left(\underline{y}_t \le y_t \le \overline{y}_t\right)}{\left(\overline{y}_t - \underline{y}_t\right)}, \tag{21}$$

where

$$\underline{y}_t = C\hat{x}_t^+ - \left(r + |C|\left[\frac{\overline{m}_t - \underline{m}_t}{2} + \rho\right]\right) \tag{22}$$

$$\overline{y}_t = C\hat{x}_t^+ + \left(r + |C|\left[\frac{\overline{m}_t - \underline{m}_t}{2} + \rho\right]\right), \tag{23}$$

where $\hat{x}_t^+ = A\hat{x}_{t-1} + Bu_{t-1} \equiv \mathrm{E}\,[x_t|d(t-1)]$ and $|C|$ is absolute value by items.

This predictive pdf is conditioned by $\underline{m}_t$ and $\overline{m}_t$ considered as statistics, provided the parameters A and B be known.

Mean value $\hat{y}_t \equiv \mathrm{E}\left[y_t|d(t-1)\right]$ of (21) is

$$\mathrm{E}\left[y_t|d(t-1)\right] = \frac{\overline{y}_t + \underline{y}_t}{2} = C \underbrace{\left(\frac{\overline{m}_t + \underline{m}_t}{2}\right)}_{\mathrm{E}[x_t|d(t-1)]}. \tag{24}$$

2.4 Bayesian filtering and prediction within the UPS class

This section summarizes the results of the filtering within the UPS class, as presented in [16], and deduces the relevant data predictor.

Data update The data update (13) processes $f(x_t|d(t-1))$ together with the $f(y_t|x_t)$ (15) according to the Bayes rule. It starts in $t = 1$ with $f(x_1) = \mathcal{U}_x(a_1^+, b_1^+, V_1^+)$. The exact pdf is uniformly distributed on a zonotopic support that results from the intersection of a parallelotope (27) obtained during previous time update—or $f(x_1)$ in the first step—and strips (9) given by new data

$$f(x_t|d(t)) \propto \mathcal{U}_x(a_t^+, b_t^+, V_t^+)\mathcal{U}_{y_t}(Cx_t - r, Cx_t + r) \propto$$

$$\propto \chi\left(\begin{bmatrix} a_t^+ \\ y_t - r \end{bmatrix} \leq \begin{bmatrix} V_t^+ \\ C \end{bmatrix} x_t \leq \begin{bmatrix} b_t^+ \\ y_t + r \end{bmatrix}\right). \tag{25}$$

We approximate (25) by a pdf uniformly distributed on a parallelotopic support [16]. Then, the approximate pdf takes the form

$$f(x_t|d(t)) \approx \mathcal{U}_x(a_t, b_t, V_t). \tag{26}$$

Note that the data update step for UPS class is identical to the data update step for UOS class except that last step, i.e, the circumscribing by an orthotope.

Time update The time update (14) processes $f(x_t|d(t))$ (26) together with $f(x_{t+1}|x_t, u_t)$ (15). The exact pdf $f(x_{t+1}|d(t))$ is non-uniformly distributed on a zonotopic support. It has a linear piecewise shape with shaping parameters ρ, a_t and b_t. It is approximated by the uniform pdf, see [16]. The support of $f(x_{t+1}|d(t))$ is computed in two steps. At first, the support $\mathbb{X}_t$ of $f(x_t|d(t))$ (26) is transformed according to the deterministic part, i.e. $\tilde{x}_{t+1}$, of (15). For this, the parallelotope $\mathbb{X}_t$ of form (1) is converted into the form (5) and then the linear transformation $\tilde{x}_{t+1} = Ax_t + Bu_t$ is performed as presented in [16]. The resulting support, $\tilde{\mathbb{X}}_{t+1}$, is then transformed back to the form (1). Secondly, the parallelotope $\tilde{\mathbb{X}}_{t+1}$ is expanded by the set $[-\rho, \rho]$ which corresponds to the stochastic part of (15). The resulting support corresponds to the Minkowski sum of $\tilde{\mathbb{X}}_{t+1}$ and the set $[-\rho, \rho]$ which is a zonotope (1).

The projection of the above mentioned uniform pdf into UPS class is explained in Appendix A, giving

$$f(x_{t+1}|d(t)) \approx \mathcal{U}_x(a_{t+1}^+, b_{t+1}^+, V_{t+1}^+). \tag{27}$$

Predictor The observation predictor of a linear state-space model is the denominator of (13), where $f(y_t|x_t)$ is given by (15) and $f(x_t|d(t-1))$ is the result of the approximate time update (27). Then, equivalently,

$$f(y_t|d(t-1)) \propto \int_{x_t^\star} \underbrace{\chi(y_t - r \leq Cx_t \leq y_t + r)}_{\text{a strip in the } x_t\text{-space}} \underbrace{\chi(a_t^+ \leq V_t^+ x_t \leq b_t^+)}_{\text{a parallelotope in the } x_t\text{-space}} \,\mathrm{d}x_t. \quad (28)$$

The integral (28) is positive, if the intersection of the strip and the parallelotope is nonempty, otherwise zero. Its value, i.e. volume of the intersection, is influenced by y_t. Changing y_t will shift the strip in (28) across the parallelotope, therefore $f(y_t|d(t-1))$ is trapezoidal. Its functional form can be approximated optimally in Bayesian sense by a uniform pdf on the support of the trapezoidal $f(y_t|d(t-1))$, as shown in [29]. Therefore, we need the values $\underline{y}_t$ and $\overline{y}_t$, for which the strip touches the parallelotope from either side. These values will be the bounds of the uniform predictive pdf.

Let the parallelotope $\chi(a_t^+ \leq V_t^+ x_t \leq b_t^+)$ be expressed in the direct form (5). According to [37], T_t^+ must be adjusted so that $CT_{t;i}^+ \geq 0$, $i = 1, \ldots, \ell$, by multiplication of corresponding columns by -1. Such an adjustment does not change the set described by T_t^+. Then, using analysis in [37],

$$\underline{y}_t = C\hat{x}_t^+ - \left(r + \sum_{i=1}^{\ell} CT_{t;i}^+ \right), \quad (29)$$

$$\overline{y}_t = C\hat{x}_t^+ + \left(r + \sum_{i=1}^{\ell} CT_{t;i}^+ \right). \quad (30)$$

and the predictive pdf, approximated by a uniform distribution, is

$$f(y_t|d(t-1)) \approx \mathcal{U}(\underline{y}_t, \overline{y}_t). \quad (31)$$

Note that if V_t^+ is an identity matrix, i.e. the support of $f(x_t|d(t-1))$ is an orthotope then the formulae (29) and (30) correspond to (22) and (23).

3 FPD-optimal transfer learning between Bayesian filters

We consider a pair of Bayesian filters, and distinguish between the primary filter—which we call the *target filter* or node [34]—and the secondary filter, which we call the *source filter* or node. In the latter, all sequences and distributions are subscripted by "s". Each filter processes a local observation sequence, y_t and $y_{s,t}$, $t \in t^\star$, respectively, informative of their local, hidden (state) process, x_t and $x_{s,t}$, respectively.

If transfer learning from the source to the target filter is to be effective (the case known as positive transfer [34]), then we must assume that $y_s(t)$ is informative of $x(t)$. The core technical problems to be addressed are (i) that there is no complete model relating $x(t)$ and $x_\mathrm{s}(t)$, and, therefore, (ii) the standard

Bayesian conditioning does not prescribe the required target distribution conditioned on the transferred knowledge, since it yields this prescription only in the case of complete modelling [18]. However, an insistence on specification of a complete model of inter-filter dependence can be highly restrictive, and incur model sensitivity in applications.

Instead, we acknowledge that the required conditional target state predictor, $\breve{f}(x_t|d(t-1), f_{\mathrm{s}})$, after transfer of the source data predictor, $f_{\mathrm{s}}(y_{\mathrm{s},t}|d_{\mathrm{s}}(t-1))$, is non-unique. Therefore, we solve the incurred decision problem optimally via the fully probabilistic design (FPD) principle [30], choosing between all cases of $\breve{f}(x_t|d(t-1), f_{\mathrm{s}})$ consistent with the knowledge constraint introduced by the transfer of $f_{\mathrm{s}}(y_{\mathrm{s},t}|d_{\mathrm{s}}(t-1))$ [31]. FPD axiomatically prescribes an optimal choice, f^{o}, as a minimizer of the Kullback-Leibler divergence (KLD) from $\breve{f}$ to an ideal distribution, f^I, chosen by the designer [20]:

$$f^o(x_t|d(t-1), f_{\mathrm{s}}) \equiv \underset{\breve{f}\in\mathbf{F}}{\arg\min}\, D(\breve{f}||f_I). \tag{32}$$

Here, the KLD is defined as

$$D(\breve{f}||f_I) = E_{\breve{f}}\left[\ln\frac{\breve{f}}{f_I}\right], \tag{33}$$

and $\mathbf{F}$ in (32) denotes the set of $\breve{f}$ constrained by f_{s} [30],[31]. The ideal distribution is consistently defined as

$$f_I(x_t|d(t-1)) \equiv f(x_t|d(t-1)),$$

being the state predictor of the *isolated* target filter (27 or (19)) (i.e. in the absence of any transfer from a source filter). In [10], the following mean-field operator was shown to satisfy (32), and so to constitute the FPD-optimal target state predictor after the (static) knowledge transfer specified above:

$$f^{\mathrm{o}}(x_t|d(t-1), f_{\mathrm{s}}) \propto f(x_t|d(t-1)) \times$$
$$\times \exp\left[\int_{y_{t^*}} f_{\mathrm{s}}(y_t|d_{\mathrm{s}}(t-1)) \ln f(y_t|x_t)\, \mathrm{d}y_t\right]. \tag{34}$$

Note that the optional input, u_t, is *known* and constant (being a conditioning quantity in $d(t-1)$), and so plays no role in the FPD optimization (32).

3.1 FPD-optimal knowledge transfer within the UOS and UPS classes

In [17], the KLD minimiser (34) is presented with $f_{\mathrm{s}}(y_t|d_{\mathrm{s}}(t-1))$ (21), and $f(y_t|x_t)$ (15) in the form

$$f^{\mathrm{o}}(x_t|d(t-1), f_{\mathrm{s}}) \propto f(x_t|d(t-1)) \times$$
$$\times \exp\left[\int_{y_{t^*}} \chi\left(\underline{y}_{\mathrm{s},t} \le y_t \le \overline{y}_{\mathrm{s},t}\right) \ln \chi\left(Cx_t - r \le y_t \le Cx_t + r\right)\, \mathrm{d}y_t\right]. \tag{35}$$

The first term in the integral indicates the integration limits are $\underline{y}_{\mathrm{s},t}$ and $\overline{y}_{\mathrm{s},t}$. The characteristic function in the logarithm must equal one, which sets the exponent to zero, inducing an additional geometric constraint for x_t represented by the strip (2) in the state space[3]:

$$\overline{y}_{\mathrm{s},t} - r \leq Cx_t \leq \underline{y}_{\mathrm{s},t} + r. \tag{36}$$

The first term $f(x_t|d(t-1))$ in (35) has the form (19) for the UOS class and (27) for the UPS class.

The subsequent time update of the target filter yields a domain of the particular class in the state space. The intersection of this domain and (36) constitutes the transfer learning between a pair of the filters. However, the subsequent data update requires the prior pdf to be on an support of the same particular class, too. Therefore, the constrained set is now circumscribed by another orthotope (UOS) or parallelotope (UPS).

The designs for the UOS and UPS classes are structurally the same, differing only in the detail of the statistics in the data predictors and in state predictive supports. This all is true because of the scalar assumption for y_t.

Note that the similarity of (36) with the observation model (15), and the similarity of their processing flows, point to an interesting correspondence between the constraint step and the data update step. The difference is that the knowledge transfer (36) accepts the interval, i.e. the uniform distribution, $\mathcal{U}_y(\underline{y}_{\mathrm{s},t}, \overline{y}_{\mathrm{s},t})$, whereas the data update is driven by the realized observation, y_t. The latter can be expressed via sifting with the Dirac distribution, $\delta_y(y - y_t)$. In this distributional sense, (34)—and its instantiation in UOS/UPS transfer learning in (35)—is a generalized form of Bayesian conditioning in the incompletely modelled context.

3.2 Informal knowledge transfer

The mechanism (35) for knowledge transfer consists of the state predictor (the first term), modulated by the exponential term, the effect of which is to constrain the support of the FPD-optimal state predictor. This geometric insight—that the effect of the transferred knowledge is to impose a support restriction on the conditional state predictor, $f^{\mathrm{o}}(x_t|d(t-1), f_{\mathrm{s}})$ (35)—led us to propose an informal transfer learning mechanism: the target state predictor is updated under transfer by projecting it onto the intersection of the source and target state predictor support sets. We reported in [17] that this informal variant achieved more positive transfer than the (disappointing) improvement achieved under the FPD-optimal mechanism (35). Though not interpretable as an optimal transfer learning mechanism, the promising performance, and intuitive and computationally efficient nature of this criterion, have together encouraged us to assess its performance in transfer learning between UPS filters, as reported in Section 5.

[3] Note that a typo occurred in [17], in which the lower and upper bounds, $\overline{y}_{\mathrm{s},t}$ and $\underline{y}_{\mathrm{s},t}$, were interchanged. The error is corrected in (36).

Practically, having $f_{\mathrm{s}}(x_t|d(t-1)) \propto \chi\left(a_{\mathrm{s}t}^{+} \leq V_{\mathrm{s}t}^{+} x_t \leq b_{\mathrm{s}t}^{+}\right)$ for the source filter and $f(x_t|d(t-1)) \propto \chi\left(a_t^{+} \leq V_t^{+} x_t \leq b_t^{+}\right)$ for the target filter, both from the UPS class, their intersection corresponds to a zonotope (1) given by the inequalities $\begin{bmatrix} a_t^{+} \\ a_{\mathrm{s}t}^{+} \end{bmatrix} \leq \begin{bmatrix} V_t^{+} \\ V_{\mathrm{s}t}^{+} \end{bmatrix} x_t \leq \begin{bmatrix} b_t^{+} \\ b_{\mathrm{s}t}^{+} \end{bmatrix}$. Using [37], the zonotope is circumscribed by a parallelotope with shaping parameters $a_t^{\oplus}$, $b_t^{\oplus}$ and $V_t^{\oplus}$ that form a support for the prior pdf after the knowledge transfer.

4 Algorithmic summary

The source and target filters run their state estimation tasks in parallel. The source filter is run in isolation. The latter sequentially transfers its one-step-ahead data predictor, $f_{\mathrm{s}}(y_{\mathrm{s},t}|d_{\mathrm{s}}(t-1))$ (Section 3), to the target filter. Note that the the stochastic mechanism which generates this data-predictive sequence at the source does not need to be specified, in line with the incomplete modelling which characterizes this Bayesian transfer learning framework. The processing of the source knowledge sequence occurs between the time update and data update steps in the target filter.

Initialisation:

- Choose final time $\overline{t} > 1$, set initial time $t = 1$.
- Set prior values $\underline{x}_1$, $\overline{x}_1$, noises ρ, r, r_{s}.

On-line:

1. Get new data u_t, y_t, $y_{\mathrm{s},t}$.
2. *Data update:* add the data strip (15) and approximate the obtained support by a parallelotope (for details see Appendix A.2 in [29]) to obtain the resulting form (26) for each filter.
3. Compute a_t, b_t, V_t (26) for each filter.
4. Compute the point estimate $\hat{x}_t$ (7) for the target filter.
5. If $t = \overline{t}$, go to 10.
6. *Time update:* compute a_{t+1}^{+}, b_{t+1}^{+}, V_{t+1}^{+} in (27) for each filter.
7. *Knowledge transfer (informal):* circumscribe the intersection of the parallelotopes (27) for the source and target state predictors by a parallelotope described by $a_{t+1}^{\oplus}$, $b_{t+1}^{\oplus}$, $V_{t+1}^{\oplus}$ and use it as a prior pdf support for the data update step.
8. Set $t = t + 1$.
9. Go to 1.
10. End.

5 Simulation of Bayesian Transfer Learning for a constrained position-velocity system

We consider the case where the target process is a LSU filter simulating a position-velocity system [8] with uniform drivers, which we also investigated in [17]. We are interested in inferring whether the Bayesian knowledge transfer

mechanism prescribed by the formal FPD-optimal scheme (35), as well as via the informal variant in Section 3.2, enhances the performance of the target filter, when compared to the isolated filter. We also want to consider the influence of the UOS-closed versus UPS-closed variants.

5.1 Simulation details

We simulate $x_t \in \mathbb{R}^2$ and $y_t \in \mathbb{R}$, so that the transferred data predictor in both the UOS and UPS cases are uniform on a line segment, differing only in the bounds, (22) and (23) versus (29) and (30), respectively. The known model parameters are $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $C = [1 \;\; 0]$, $\rho = \mathbf{1}_{(2)} \times 10^{-5}$, where $\mathbf{1}_{(2)}$ is the unit vector of length 2, and $r = 10^{-3}$. Also, $B = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $u_t = 0$, i.e. a system without input/control. The estimation was run for $\bar{t} = 50$ time steps. We investigate the influence of the observation noise, r_{s}, of the source filter (15) on the state estimate precision of the target filter, quanitified by the TNSE (i.e. total norm squared-error), defined as $\mathrm{TNSE} = \sum_{t=1}^{\bar{t}} \|\hat{x}_t - x_t\|_2^2$. For each combination of the parameters, the computation was run 500 times and the TNSEs were averaged.

The *simulated* state processes, x_t (and so the state noise variance, ρ) are equal in the target and source filters, while the observation processes are simulated conditionally independently of x_t, with different noise variances, r and r_{s}, respectively.

5.2 Comparative performance

Holding the target observation variance constant at $r = 10^{-3}$, as stated above, we investigated the TNSE as a function of the source observation variance, r_s. In [17], we reported that positive transfer—i.e. the improved precision of filtering under transfer of high-quality ($r_s < r$) source knowledge—was achieved in the UOS-closed case, for both the FPD-optimal (i.e. formal) transfer (35) and the informal variant (Section 3.2). However, the improvement was very limited—particularly for the formal scheme—because of the coarseness of the orthotopic approximation of the support of $f^{\mathrm{o}}(x_t|d(t-1), f_{\mathrm{s}})$ (35). We speculated that a parallelotopic approximation would yield an improved positive transfer, a hypothesis we now wish to validate experimentally.

For this purpose, we investigated the performance of the informal Bayesian transfer learning variant, comparing the UOS-closed and UPS-closed variants with the isolated target filter performance (Fig. 1). The UOS-closed performance has already been reported in [17], corroborating the main findings: that (i) the informal variant achieves robust transfer when $r_s > r$ (i.e. rejection of poor-quality source knowledge); and (ii) the positive transfer in the $r_s < r$ regime is significant, reducing the TNSE by about an order of magnitude when $r_s \ll r$, compared to the isolated target filter. We supplement these findings with the

performance of the UPS-closed informal transfer scheme. As anticipated in [17], this significantly outperforms the UOS-closed informal scheme in the positive transfer regime, reducing the TNSE by about a further 1.5 orders of magnitude in the $r_s \ll r$ regime.

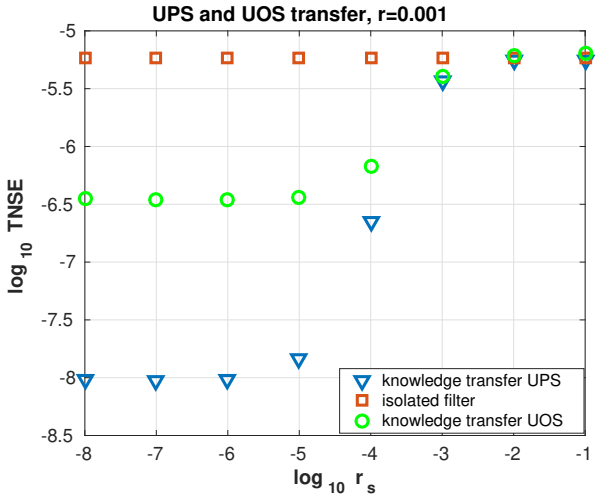

Fig. 1. TNSE of the target LSU filter as a function of the observation noise variance r_s of the source filter. UPS-closed informal knowledge transfer compared to UOS-closed informal knowledge transfer, and to the isolated target filter (no transfer).

5.3 Discussion and Closing Remarks

The impact on the state predictor of the geometric constraint (36) in the transfer step (35) of the FPD-optimal (i.e. formal) scheme requires further study. Cases may exist where higher-variance source predictors may anomalously tighten the target state predictor support, or, indeed, where the strip (36) may not overlap with the support of the pre-transfer target state predictor, $f(x_t|d(t-1))$, at all.

In contrast, the informal variant of the transfer learning scheme (Section 3.2) avoids these anomalies, and—as demonstrated in the simulations above—shows promise in achieving strongly positive transfer for high-quality source knowledge, and rejection of low-quality knowledge (robust transfer). It is inspired by the behaviour of the formal transfer in this bounded support setting (i.e. support intersection), but a formal justification is required. Another issue is that the informal variant transfers the source state predictor, rather than the source data predictor, implying that this sequential knowledge is made available by a source LSU filter (as in the simulation setting above), an assumption that is not required in the formal scheme.

The isolated LSU target filter performance (TNSE) acts as a reference/datum for assessing all the (informal) transfer algorithms (Fig. 1). It confirms that transfer of a high-quality source predictor reduces the sequential state estimation error in the target LSU filter (i.e. positive transfer). Conversely, low-quality source predictors are rejected by the schemes (i.e. robust transfer), and the target filter performance reverts to that of the isolated filter in this regime. Furthermore, the findings support the hypothesis that the projection of the output of the transfer step (35) into the UPS class is a closer local approximation than projection into the UOS class, owing to the greater number of degrees of freedom in the parallelotopic versus orthotopic support. A similar finding was reported in [16], in the context of isolated UPS-closed LSU filtering, where improved filtering accuracy was reported in comparison to the UOS-closed variant.

Perhaps the most significant finding is the robustness of the informal transfer scheme in the context of LSU filtering with bounded knowledge transfer. Our earlier experience with FPD-optimal conditioning on source data predictors for incompletely modelled pairs of Kalman filters indicated that high-variance (Gaussian) source predictors were not rejected, and that performance can deteriorate relative to the isolated Kalman filter in this regime (i.e. non-robust transfer). This is caused by the loss of the second moment of the Gaussian source predictor under the mean-field operator in (34). The problem was overcome via an informal adaptation of the prescribed algorithm in [10, 28], and via a formal scheme involving a hierarchical model relaxation in [27]. While further work is needed to understand the formal scheme in the current LSU setting, there is clear evidence that robustness is intrinsic in this setting, and that no such moment loss occurs.

Future work will focus on formal understanding of the conditions under which the FPD-optimal Bayesian transfer learning scheme achieves robust transfer, both in interacting LSU filtering pairs, and for more generally specified sensor nodes. As already implied above, the informal transfer scheme may be formally justifiable as an FPD-optimal transfer learning scheme for adapted settings, but this has yet to be demonstrated. Finally, we plan further investigation of the influence of the geometric supports of the local distributional approximations on the quality of positive transfer in the LSU filtering case, seeking (i) to reduce the number of local approximations per step of the algorithm, and (ii) investigating more flexible geometries, particularly low-order zonotopes.

A Appendix

Stochastic expansion of a parallelotope

After the deterministic transformation of the support $\mathbb{X}_t$, according to (15), within the time update, its expansion corresponding to the stochastic effect must be preformed.

For the i-th facet of the parallelotope to be expanded, we define these vectors:

- $\omega^{i\prime}$, which is the i-th row of $\tilde{W}_{t+1}$, orthogonal to the facet. Define $s^i = \frac{\omega^i}{\|\omega^i\|_2}$ a unit vector orthogonal to the i-th facet,

- τ^i, which is the i-th column of the matrix $\tilde{T}_{t+1}$. It points from the centre of the parallelotope to the centre of the i-th facet. Define $v^i = \frac{\tau^i}{\|\tau^i\|_2}$ a unit vector parallel to τ^i,
- $\varrho^i = \left[\alpha_1^i \rho_1, \alpha_2^i \rho_2, \ldots, \ \alpha_\ell^i \rho_\ell\right]'$, where $\rho_1, \ldots, \rho_\ell$ are the components of the state noise parameter ρ and $\alpha_1^i, \ldots, \alpha_\ell^i = \pm 1$ so that its j-th element of the i-th vector $\alpha_j^i \equiv \text{sign}(\varrho_j^i) = \text{sign}(s_j^i)$. This arrangement guarantees s^i and ϱ^i point to the same half-space given by the i-th facet.

Assume the orthotopic set (box) $[-\rho, \rho]$ with its centre on the i-th facet. Then, $d^i = s^{i\prime}\varrho^i$ is the distance of the most distant vertex of the box and the i-th facet of the parallelotope.

Let $\|\Delta\tau^i\|_2 \equiv \Delta\tau^{i\prime}v^i$ be a length increment of the vector τ^i pointing to the centre of the i-th facet shifted by the Minkowski sum. Then, $(\Delta\tau^{i\prime}v^i)'s^i = d^i$, i.e.

$$\Delta\tau^i = \frac{d^i}{v^{i\prime}s^i} = \frac{s^{i\prime}\varrho^i}{s^{i\prime}v^i} = \frac{\omega^{i\prime}\varrho^i}{\omega^{i\prime}v^i}. \tag{37}$$

The i-th column ${\tau^i}^+$ of the new matrix T_{t+1}^+ is

$${\tau^i}^+ = \tau^i \frac{\|\tau^i\|_2 + \|\Delta\tau^i\|_2}{\|\tau^i\|_2}. \tag{38}$$

The expanded circumscribing paralelotope is then described as $x_{t+1} = \hat{x}_{t+1}^+ + T_{t+1}^+\xi$, where $\hat{x}_{t+1}^+ = \hat{\tilde{x}}_{t+1}$ (its centre is preserved). Transforming the shaping parameters T_{t+1}^+ and $\hat{x}_{t+1}^+$ to the form (1), we get the approximate pdf

$$f(x_{t+1}|d(t)) \approx \mathcal{U}_x(a_{t+1}^+, b_{t+1}^+, V_{t+1}^+). \tag{39}$$

Bibliography

[1] Abbas, A.E.: A Kullback-Leibler view of linear and log-linear pools. Decision Analysis **6**(1), 25–37 (2009)

[2] Azizi, S., Quinn, A.: Hierarchical fully probabilistic design for deliberator-based merging in multiple participant systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems **48**(4), 565–573 (2018)

[3] Becis-Aubry, Y., Boutayeb, M., Darouach, M.: State estimation in the presence of bounded disturbances. Automatica **44**, 1867–1873 (2008)

[4] Chisci, L., Garulli, A., Zappa, G.: Recursive state bounding by parallelotopes. Automatica **32**(7), 1049–1055 (1996)

[5] Dardari, D., Closas, P., Djuric, P.M.: Indoor tracking: theory, methods, and technologies. IEEE Transactions on Vehicular Technology **64**(4), 1263–1278 (2015)

[6] Diez-Olivan, A., Del Ser, J., Galar, D., Sierra, B.: Data fusion and machine learning for industrial prognosis: trends and perspectives towards industry 4.0. Information Fusion **50**, 92 – 111 (2019)

[7] Dou, Y., Ran, C., Gao, Y.: Weighted measurement fusion Kalman estimator for multisensor descriptor system. International Journal of Sytems Science **47**(11), 2722–2732 (2016)

[8] Faragher, R.: Understanding the basis of the Kalman filter via a simple and intuitive derivation. IEEE Signal Process. Mag. **29**(5), 128–132 (2012)

[9] Fletcher, R.: Practical Methods of Optimization. John Wiley & Sons (2000)

[10] Foley, C., Quinn, A.: Fully probabilistic design for knowledge transfer in a pair of Kalman filters. IEEE Signal Processing Letters **25**, 487–490 (2018)

[11] Goudjil, A., Pouliquen, M., Pigeon, E., Gehan, O., Targui, B.: Recursive output error identification algorithm for switched linear systems with bounded noise. IFAC-PapersOnLine **50**(1), 14112–14117 (2017)

[12] Gover, E., Krikorian, N.: Determinants and the volumes of parallelotopes and zonotopes. Linear Algebra and its Applications **433**(1), 28 – 40 (2010)

[13] Gravina, R., Alinia, P., Ghasemzadeh, H., Fortino, G.: Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges. Information Fusion **35**, 68–80 (2017)

[14] He, J., Duan, X., Cheng, P., Shi, L., Cai, L.: Accurate clock synchronization in wireless sensor networks with bounded noise. Automatica **81**, 350–358 (2017)

[15] Jirsa, L., Kuklišová Pavelková, L., Quinn, A.: Approximate Bayesian prediction using state space model with uniform noise. In: Gusikhin, O., Madani, K. (eds.) Informatics in Control Automation and Robotics, Lecture Notes in Electrical Engineering, vol. 613, pp. 552–568. Springer International Publishing, Cham (2020)

[16] Jirsa, L., Kuklišová Pavelková, L., Quinn, A.: Bayesian filtering for states uniformly distributed on a parallelotopic support. In: 2019 IEEE Interna-

tional Symposium on Signal Processing and Information Technology (ISSPIT) (ISSPIT2019). Ajman, United Arab Emirates (2019), accepted
[17] Jirsa, L., Kuklišová Pavelková, L., Quinn, A.: Knowledge transfer in a pair of uniformly modelled Bayesian filters. In: Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO,. pp. 499–506. INSTICC, SciTePress (2019)
[18] Karbalayghareh, A., Qian, X., Dougherty, E.R.: Optimal Bayesian transfer learning. IEEE Transactions on Signal Processing **66**(14), 3724–3739 (2018)
[19] Kárný, M., Böhm, J., Guy, T.V., Jirsa, L., Nagy, I., Nedoma, P., Tesař, L.: Optimized Bayesian Dynamic Advising: Theory and Algorithms. Springer, London (2005)
[20] Kárný, M., Kroupa, T.: Axiomatisation of fully probabilistic design. Information Sciences **186**(1), 105–113 (2012)
[21] Khaleghi, B., Khamis, A., Karray, F.O., Razavi, S.N.: Multisensor data fusion: a review of the state-of-the-art. Information Fusion **14**(1), 28 – 44 (2013)
[22] Kullback, S., Leibler, R.: On information and sufficiency. Annals of Mathematical Statistics **22**, 79–87 (1951)
[23] Lahat, D., Adali, T., Jutten, C.: Multimodal data fusion: an overview of methods, challenges, and prospects. Proceedings of the IEEE **103**(9), 1449–1477 (2015)
[24] Lang, L., Chen, W., Bakshi, B.R., Goel, P.K., Ungarala, S.: Bayesian estimation via sequential Monte Carlo sampling — constrained dynamic systems. Automatica **43**(9), 1615–1622 (2007)
[25] Majumder, S., Pratihar, D.K.: Multi-sensors data fusion through fuzzy clustering and predictive tools. Expert Systems with Applications **107**, 165 – 172 (2018)
[26] Nassreddine, G., Abdallah, F., Denoux, T.: State estimation using interval analysis and belief-function theory: application to dynamic vehicle localization. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **40**(5), 1205–1218 (2010)
[27] Papež, M., Quinn, A.: Robust Bayesian transfer learning between Kalman filters. In: 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP). pp. 1–6. Pittsburg (2019)
[28] Papež, M., Quinn, A.: Dynamic Bayesian knowledge transfer between a pair of Kalman filters. In: 2018 28th International Workshop on Machine Learning for Signal Processing (MLSP). pp. 1–6. IEEE, Aalborg, Denmark (2018)
[29] Pavelková, L., Jirsa, L.: Approximate recursive Bayesian estimation of state space model with uniform noise. In: Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO). pp. 388–394. Porto, Portugal (2018)
[30] Quinn, A., Kárný, M., Guy, T.: Fully probabilistic design of hierarchical Bayesian models. Information Sciences **369**(1), 532–547 (2016)
[31] Quinn, A., Kárný, M., Guy, T.V.: Optimal design of priors constrained by external predictors. International Journal of Approximate Reasoning **84**, 150–158 (2017)

[32] Shamshirband, S., Petkovic, D., Javidnia, H., Gani, A.: Sensor data fusion by support vector regression methodology-a comparative study. IEEE Sensors Journal **15**(2), 850–854 (2015)
[33] Simon, D., Simon, D.L.: Constrained Kalman filtering via density function truncation for turbofan engine health estimation. International Journal of Systems Science **41**, 159–171 (2010)
[34] Torrey, L., Shavlik, J.: Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques, pp. 242–264. IGI Global (2010)
[35] Vapnik, V., Izmailov, R.: Knowledge transfer in SVM and neural networks. Annals of Mathematics and Artificial Intelligence **81**(1-2), 3–19 (2017)
[36] Vargas-Melendez, L., Boada, B.L., Boada, M.J.L., Gauchia, A., Diaz, V.: Sensor fusion based on an integrated neural network and probability density function (PDF) dual Kalman filter for on-line estimation of vehicle parameters and states. Sensors **17**(5) (2017)
[37] Vicino, A., Zappa, G.: Sequential approximation of feasible parameter sets for identification with set membership uncertainty. IEEE Transactions on Automatic Control **41**(6), 774–785 (1996)
[38] Vitola, J., Pozo, F., Tibaduiza, D.A., Anaya, M.: A sensor data fusion system based on k-nearest neighbor pattern classification for structural health monitoring applications. Sensors **17**(2) (2017)
[39] Willner, D., Chang, C., Dunn, K.: Kalman filter algorithms for a multisensor system. In: 1976 IEEE Conference on Decision and Control including the 15th Symposium on Adaptive Processes. pp. 570–574 (1976)
[40] Xiao, F.: Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. Information Fusion **46**, 23 – 32 (2019)
[41] Yang, C., Yang, Z., Deng, Z.: Robust weighted state fusion Kalman estimators for networked systems with mixed uncertainties. Information Fusion **45**, 246 – 265 (2019)
[42] Zang, Y., Hu, X.: Heterogeneous Knowledge Transfer via Domain Regularization for Improving Cross-Domain Collaborative Filtering. In: 2017 IEEE International Conference on Big Data. pp. 3968–3974 (2017)
[43] Zou, T., Wang, Y., Wang, M., Lin, S.: A real-time smooth weighted data fusion algorithm for greenhouse sensing based on wireless sensor networks. Sensors **17**(11) (2017)