DIFFUSION FILTRATION WITH APPROXIMATE BAYESIAN COMPUTATION

Kamil Dedecius[*]

Inst. of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 1143/4
182 08 Prague 8, Czech Republic

Petar M. Djurić [†]

Department of Electrical and Computer Engineering
Stony Brook University
Stony Brook
NY 11794-2350, USA

ABSTRACT

Distributed filtration of state-space models with sensor networks assumes knowledge of a model of the data-generating process. However, this assumption is often violated in practice, as the conditions vary from node to node and are usually only partially known. In addition, the model may generally be too complicated, computationally demanding or even completely intractable. In this contribution, we propose a distributed filtration framework based on the novel approximate Bayesian computation (ABC) methods, which is able to overcome these issues. In particular, we focus on filtration in diffusion networks, where neighboring nodes share their observations and posterior distributions.

Index Terms— Bayesian filtration, diffusion, distributed filtration, approximate Bayesian computation

1. INTRODUCTION

We address the problem of fully distributed filtration of unknown states of state-space models from noisy observations by diffusion networks, allowing communication of available (statistical) information among nodes within one hop distance. The communication protocol distinguishes between two phases: the *adaptation*, where the observations are exchanged, and the *combination* phase, where the nodes provide their estimates [1, 2, 3]. Unlike in the consensus networks, both phases may occur at most once between two consecutive observations and any intermediate iterations are not allowed. The diffusion methods require less communication and computation resources, and yet, they can still outperform the consensus-based methods [4]. We point out, however, that there are "consensus" methods which, as the diffusion methods, operate with only one exchange between consecutive observations. They are known as running consensus methods [5, 6]. On some of their extensions, see [7, 8].

The filtration task can be solved using the diffusion Kalman filter originally proposed by Cattivelli and Sayed [1]. Their method provides means for simplified covariance-independent combine step, reducing the communication burden between nodes. Hu, Xie and Zhang [9] propose an alternative covariance intersection-based combine step, providing better estimation performance at the cost of higher communication requirements. Recently, Dedecius [10] formulated the state-space estimation problem in the Bayesian paradigm, providing a more general method with the previous two filters as special cases.

Although Monte Carlo-based filtration of nonlinear state-space models is a well established discipline in consensus networks, e.g. Hlinka *et al.* [11, 12], the domain of diffusion algorithms was given attention only recently by Bruno and Dias [13].

A common feature of the mentioned algorithms is their assumption of a fully known model. The already disputable assumption of this knowledge becomes even more problematic in the distributed settings, where the nodes generally face more or less heterogeneous conditions. For instance, this is the case of target tracking and distributed image recognition when the observation model may be known only partially, making the parameter (or state) inference with usual methods complicated or even impossible. Another frequent problem is that in many cases the models are highly complex or even intractable, though it is still possible to sample from them. We propose to resolve these issues by using a group of increasingly popular Monte Carlo methods called *approximate Bayesian computation* (ABC) [14]. They consist in simulating pseudo-observations using Monte Carlo samples of estimated states, plugged into the (approximate) observation models. The samples leading to pseudo-observations close to the true observations then represent the posterior distribution of the inferred state. Besides for completely intractable models, this approach is particularly appealing in settings where it is easier to simulate from the model than to compute the true observations likelihood.

The filtration of the state-space models via ABC has become very promising since 2012 due to the seminal paper of Jasra *et al.* [15], proposing a simple sequential Monte Carlo procedure for sampling from the state space. According to the proposed method, the samples used for simulation of pseudo-observations are either accepted or rejected based on their presence in a convex acceptance set around the true observation. The problem of degenerate uniform importance weights was resolved by Calvet and Czellar [16] via kernel-based evaluation of these weights. Furthermore, Martin *et al.* [17] propose ABC filter approaching the exact solutions via asymptotic sufficiency. ABC-based smoothing was discussed by Martin *et al.* in [18].

1.1. Contribution

Our main contribution consists in formulating the ABC filtering algorithm for use in diffusion networks. We propose methods for (i) adaptation, i.e. assimilation of observations from the neighboring nodes and (ii) combination of posterior densities with low communication demands. A particular emphasis is put on generality of the underlying principles, allowing ABC-based distributed filtration

[*]Email: dedecius@utia.cas.cz. The work of K. Dedecius was supported by the Czech Science Foundation, grant No. 14–06678P.

[†]Email: petar.djuric@stonybrook.edu. The work of P.M. Djurić was supported by NSF under award CCF-1320626.

regardless of which particular Monte Carlo method is employed. The simulation example, exploiting sequential importance sampling, shows that the proposed diffusion filter is (even without tedious tuning) close in performance to the more demanding particle filter (PF) in terms of mean square error performance.

2. FILTRATION WITH ABC

We assume a discrete-time state-space model with measurement and transition distributions in the form

$$\begin{aligned} Y_t|X_t &\sim g_t(y_t|x_t), \\ X_t|X_{t-1} &\sim q_t(x_t|x_{t-1}), \end{aligned} \tag{1}$$

where the likelihood (1) is either too complex to be evaluated analytically or numerically (*but it is possible to sample from it*), or it is a rough approximation of the true model. Jasra *et al.* [15] propose to infer its parameters via a sequential Monte Carlo-based ABC algorithm replacing the posterior density

$$\pi_t(x_{0:t}|y_{1:t}) \propto \pi(x_0) \prod_{\tau=1}^{t} g_\tau(y_\tau|x_\tau) q_\tau(x_\tau|x_{\tau-1})$$

by an approximate density represented with samples $x_\tau^{(j)}$ from $q_\tau(x_\tau|x_{\tau-1})$, that produce pseudo-observations $u_\tau^{(j)} \sim g_\tau(y_\tau|x_\tau^{(j)})$ within some admissible set around the true y_τ,

$$\begin{aligned} \pi_t^\varepsilon(x_{0:t}|y_{1:t}) &= \pi(x_0) \int \pi_t^\varepsilon(x_{1:t}, u_{1:t}|y_{1:t}) du_{1:t} \\ &\propto \pi(x_0) \prod_{\tau=1}^{t} \left[\int \tilde{g}_{\tau,\varepsilon}(y_\tau, u_\tau) g_\tau(u_\tau|x_\tau) du_\tau \right] q_\tau(x_\tau|x_{\tau-1}), \end{aligned} \tag{2}$$

where the selection function $\tilde{g}_{\tau,\varepsilon}(y_\tau, u_\tau)$ determines this set. The SMC samples $x_t^{(j)}$ after each ABC update in (2) have weights

$$W_t^{(j)} \propto W_{t-1}^{(j)} \tilde{g}_{t,\varepsilon}\left(y_t, u_t^{(j)}\right). \tag{3}$$

Most of the ABC literature uses the characteristic function

$$\tilde{g}_{t,\varepsilon}\left(y_t, u_t^{(j)}\right) = \mathbb{1}_{A_{\varepsilon,y_t}}\left(u_t^{(j)}\right),$$

where

$$A_{\varepsilon,y_t} = \{u : \rho(u, y_t) < \varepsilon\}.$$

The metric $\rho(u, y_t)$ measures the distance of the pseudo-observation u from the true observation y_t. The authors in [15] show that under fixed ε the filter converges to a biased estimator as the number of particles tends to infinity and that the bias itself tends to zero as ε goes to zero.

The characteristic function in the role of $\tilde{g}_{t,\varepsilon}$ is rather inappropriate in dynamic scenarios, since the set A_{ε,y_t} may become empty if the noise realization exceeds the set radius ε. Moreover, from (3) it is obvious that this choice produces simplistic uniform particle weights. In order to resolve this issue, we adhere to a kernel-based $\tilde{g}_{t,\varepsilon}$ mentioned also by Jasra *et al.* [15] and only recently studied by Calvet and Czellar [16] (currently in preprint). Some examples of kernels are:

Rational quadratic kernel

$$\tilde{g}_{t,\varepsilon}\left(y_t, u_t^{(j)}\right) = 1 - \frac{||y_t - u_t^{(j)}||^2}{||y_t - u_t^{(j)}||^2 + \varepsilon^2}$$

Exponential kernel

$$\tilde{g}_{t,\varepsilon}\left(y_t, u_t^{(j)}\right) = \exp\left(-\frac{||y_t - u_t^{(j)}||}{\varepsilon^2}\right)$$

Gaussian kernel

$$\tilde{g}_{t,\varepsilon}\left(y_t, u_t^{(j)}\right) = \exp\left(-\frac{||y_t - u_t^{(j)}||^2}{2\varepsilon^2}\right)$$

Cauchy kernel

$$\tilde{g}_{t,\varepsilon}\left(y_t, u_t^{(j)}\right) = \left(1 + \frac{||y_t - u_t^{(j)}||^2}{\varepsilon^2}\right)^{-1}$$

Algorithm 1 summarizes a generic ABC sequential Monte Carlo (SMC) filter.

Algorithm 1 ABC-SMC FILTERING ALGORITHM [15]

Initialize particles $\left\{x_0^{(j)}, W_0^{(j)}\right\}_{j=1,\ldots,J}$ by sampling from a suitable prior $\pi(x_0)$. For $t = 1, 2, \ldots$ do:

1. Obtain observation y_t.
2. Propose $x_t^{(j)} \sim q_t\left(x_t|x_{t-1}^{(j)}\right)$.
3. Simulate pseudo-observation $u_t^{(j)} \sim g_t\left(y_t|x_t^{(j)}\right)$.
4. Update weights $W_t^{(j)} \propto W_{t-1}^{(j)} \tilde{g}_{t,\varepsilon}\left(y_t, u_t^{(j)}\right)$ such that $\sum_{j=1}^{J} W_t^{(j)} = 1$.
5. Resample if the effective sample size drops below a specified threshold.

3. DIFFUSION FILTRATION WITH ABC

This section devises the approximate filtration in diffusion networks consisting of nodes $n = 1, \ldots, N$, taking uni- or multivariate observations $y_{n,t}$ [3]. These observations are driven by an underlying hidden Markov process with *global states* x_t. Each node n exchanges own observations (*adaptation step*) and estimates (*combination step*) with its adjacent neighbors within one hop distance. These neighbors form its *neighborhood* $\mathcal{Q}_n$; note that $n \in \mathcal{Q}_n$, too. The goal is to concurrently estimate x_t using the characterized ABC method.

3.1. Adaptation step

During the adaptation step the nodes $n = 1, \ldots, N$ incorporate the observations $y_{m,t}$ of their neighbors $m \in \mathcal{Q}_n$ into their own local statistical knowledge $\pi_{n,t-1}^\varepsilon(x_{1:t-1}|\tilde{y}_{n,1:t-1})$ as in (2). The term $\tilde{y}_{n,t-1}$ has the meaning of the sequence of observations $\{y_{m,\tau}; m \in \mathcal{Q}_n\}_{\tau=1,\ldots,t-1}$. Instead of a sequential simulation of pseudo-observations for each measurement from the neighbors we propose to surrogate all $y_{m,t}$ of $m \in \mathcal{Q}_n$ by a *summary statistics* $\mathcal{S}_{n,\tilde{y}_t} \equiv \mathcal{S}(y_{m,t}, \forall m \in \mathcal{Q}_n)$, an approach common in the nonsequential ABC. Then, the selection function $\tilde{g}_{t,\varepsilon}$ acts on $(\mathcal{S}_{n,\tilde{y}_t}, u_{n,t}^{(j)})$. For typical state-space models with zero-centered additive measurement noise and without fixed parameters the arithmetic mean is usually a good choice. An alternative method for filtering of state-space models (with constant parameters) via asymptotically sufficient summary statistics was recently proposed by Martin *et al.* [17].

3.2. Combination step

The goal of the combine step is twofold: first, it is necessary to determine the mechanism of information exchange among nodes. Second, it is necessary to merge the information provided by the neighboring nodes in order to obtain their jointly-optimal representation.

Any exchange of posterior distributions in terms of raw particles would be highly demanding. This issue is widely studied, e.g., in [12, 19], where the posterior distribution (here $\pi^{\varepsilon}_{n,t}(x_t|\tilde{y}_{1:t})$) represented by weighted particles is approximated by a K-component Gaussian mixture (GM) or a normal distribution, respectively,

$$G_{n,t}(x_t|\cdot) = \sum_{k=1}^{K} \alpha_{n,k,t}\mathcal{N}(\mu_{n,k,t}, \Sigma_{n,k,t}). \tag{4}$$

The exchange of GM parameters dramatically reduces the communication burden, of course at the cost of decreased estimation efficiency and higher computational load.

We propose merging of the posterior distributions (or, for practical reasons their approximations by GMs) exploiting the Kullback-Leibler divergence [20]. This divergence measures the dissimilarity of two probability densities f and g (equivalent in the sense of common support) of a random variable $x \in \mathbb{R}^d$. Its minimization with one argument fixed corresponds to an approximation optimal in the Kullback-Leibler sense. The reason for this particular choice is justified by an information-theoretic argument: the Kullback-Leibler divergence can be interpreted in terms of the Shannon entropy.

The Kullback-Leibler divergence is defined as the nonnegative functional

$$\begin{aligned}\mathcal{D}(f||g) &= \mathbb{E}_f\left[\log\frac{f(x)}{g(x)}\right] = \int_{\mathbb{R}^d} f(x)\log\frac{f(x)}{g(x)}dx \\ &= \mathcal{H}(f,g) - \mathcal{H}(f).\end{aligned}$$

Here $\mathcal{H}(f,g)$ and $\mathcal{H}(f)$ are the cross-entropy and the Shannon entropy, respectively. Obviously, this divergence is a premetric: it is asymmetric, does not satisfy the triangle inequality and is zero if the arguments agree. The merging follows from the following theorem.

Theorem 1. *Fix n. Let $\pi^{\varepsilon}_{m,t} \equiv \pi^{\varepsilon}_{m,t}(x_t|\cdot)$ be the posterior densities of nodes $m \in \mathcal{Q}_n$ and a_{nm} the coefficients taking values in the unit $|\mathcal{Q}_n|$-simplex. The approximate density $\tilde{\pi}^{\varepsilon}_{n,t}$ at node n optimal in the Kullback-Leibler sense*

$$\sum_{m\in\mathcal{Q}_n} a_{nm}\mathcal{D}\left(\pi^{\varepsilon}_{m,t}\middle|\middle|\tilde{\pi}^{\varepsilon}_{n,t}\right) \to \min$$

has the form

$$\tilde{\pi}^{\varepsilon}_{n,t} = \sum_{m\in\mathcal{Q}_n} a_{nm}\pi^{\varepsilon}_{m,t}.$$

Proof. With the help of the definition of the Kullback-Leibler divergence,

$$\begin{aligned}\sum_{m\in\mathcal{Q}_n} a_{nm}\mathcal{D}\left(\pi^{\varepsilon}_{m,t}\middle|\middle|\tilde{\pi}^{\varepsilon}_{n,t}\right) &= \mathcal{D}\left(\sum_{m\in\mathcal{Q}_n} a_{nm}\pi^{\varepsilon}_{m,t}\middle|\middle|\tilde{\pi}^{\varepsilon}_{n,t}\right) \\ &- \sum_{m\in\mathcal{Q}_n} a_{nm}\mathcal{H}\left(\pi^{\varepsilon}_{m,t}\right) + \mathcal{H}\left(\sum_{m\in\mathcal{Q}_n} a_{nm}\pi^{\varepsilon}_{m,t}\right).\end{aligned}$$

Since both the entropies on the right-hand side are independent of $\tilde{\pi}^{\varepsilon}_{n,t}$, the minimum is attained if the arguments of the Kullback-Leibler divergence agree. □

The Kullback-Leibler optimal approximation of the neighborhood's statistical knowledge is hence a mixture density with mixing coefficients a_{nm}. These may be uniform or not, if it is necessary to discriminate among the neighbors, e.g. based on their degrees or noise properties [2]. It is quite natural to expect that the components significantly overlap, as the network nodes observe essentially the same underlying process. More importantly, if each of the components is a GM (as proposed above), the result is again a GM. This intrinsic advantage of the method preserves computationally effective sampling of particles $x^{(j)}_{n,t}$ for the next time step. Additionally, countermeasures can be applied to prevent particles depletion, e.g., the popular local random walk [21].

Algorithm 2 summarizes the resulting ABC filter for diffusion networks.

4. DISCUSSION OF FILTER PROPERTIES

The thorough analysis of properties of the proposed diffusion filter is, similarly to all ABC methods, quite challenging. In addition to the free parameters ε (bandwidth or scaling parameter) and J (the number of particles), it is necessary to reflect the possible heterogeneity of the network and the GM-based approximation properties. At this point, we give only the following remarks:

Under the (hardly justifiable) assumption that the GM perfectly fits the posterior:

- If the true observation models agree across the network and the noises are zero-centered and symmetric, the adaptation step coincides with the ordinary ABC methods and the estimator is asymptotically consistent [15] but biased if $\varepsilon = const$. With $\varepsilon \to 0$ the asymptotic bias tends to zero.
- If the true observation models differ across the network, the estimator is generally biased. However, under zero-centered and symmetric additive noises, the above-given asymptotics holds as the sum of related variables is again zero-centered. Of course, this does not hold for other moments.

5. SIMULATION RESULTS

The aim of this section is to demonstrate the performance of the proposed ABC diffusion filtering method and compare it to its particle filtering counterpart. Both methods are adopted in their basic forms without any tedious tuning, e.g., of the number of particles and/or the kernel type.

We assume a diffusion network consisting of 15 nodes (depicted in Fig. 1). The nodes process the nonlinear state-space models cooperatively. The models, popular in SMC literature [18, 22], are given by

$$x_{i,t} = \frac{x_{i,t-1}}{2} + \frac{25x_{i,t-1}}{1+x^2_{i,t-1}} + 8\cos(1.2t) + v_{i,t}, \tag{5}$$

$$y_{n,i,t} = \frac{x^2_{i,t}}{20} + w_{n,i,t}, \qquad i = 1,2, \tag{6}$$

where n denotes the node number. The Markov process starts from $x_0 = [0,0]$. The state vectors $x_t \in \mathbb{R}^{d_x}$, $y_{n,t} \in \mathbb{R}^{d_y}$, $d_x = d_y = 2$, the independent identically distributed normal zero-mean noise variables $v_{i,t} \sim \mathcal{N}(0,\sigma^2_x)$ with $\sigma^2_x = 1$ and $w_{n,i,t} \sim \mathcal{N}(0,\sigma^2_{n,i,y})$ with $\sigma^2_{n,i,y} = 0.4n^2$. The series have 100 samples. Each of the nodes employs 1000 particles. The nodes do not know the (heterogeneous) noise covariances and infer the states directly from equations (5) and

Algorithm 2 DIFFUSION QUASI-BAYES ESTIMATION

Initialize nodes $n = 1, \ldots, N$, initialize particles $x_{n,0}^{(j)}, j = 1, \ldots, J$ with uniform weights $W_{n,0}$ by sampling from a suitable prior.

For $t = 1, 2, \ldots$ and each node n do:
Adaptation:

1. Obtain observations $y_{m,t}, m \in \mathcal{Q}_n$ and calculate their summary statistics $S_{n,\bar{y}_t}$.
2. Propose $x_{n,t}^{(j)} \sim q_t(x_t | x_{n,t-1}^{(j)})$.
3. Simulate pseudo-observation $u_{n,t}^{(j)} \sim g_t(y_t | x_{n,t}^{(j)})$.
4. Update weights $W_{n,t}^{(j)} \propto W_{n,t-1}^{(j)} \tilde{g}_{t,\varepsilon}\left(\mathcal{S}_{n,\tilde{y}_t}, u_{n,t}^{(j)}\right)$.
5. Approximate the posterior $\pi_{n,t}^{\varepsilon}$ represented by $\{x_{n,t}^{(j)}, W_{n,t}^{(j)}\}_{j=1,\ldots,J}$ by a GM $G_{n,t}(x_t|\cdot)$.

Combination:

1. Exchange mixtures $G_{m,t}(x_t|\cdot), m \in \mathcal{Q}_n$.
2. Sample new particles $x_{n,t}^{(j)}$ from the combined mixture and set their weights uniform.

Remark: If the information exchange occurs every time instant t, the posterior weights $W_{n,t}^{(j)} \propto \tilde{g}_{t,\varepsilon}\left(S_{n,\tilde{y}_t}, u_{n,t}^{(j)}\right)$, hence there is no need to initialize them.

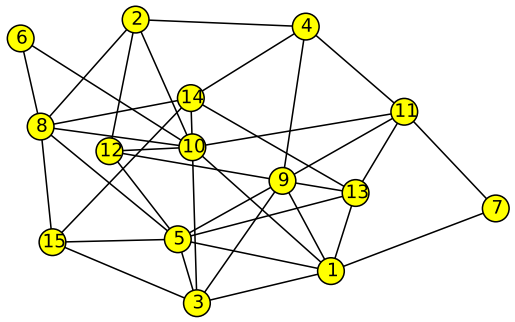

Fig. 1. Diffusion network.

(6) without the noise terms. We fit the posterior with a GM with two components and the coefficients a_{nm} are uniform for neighboring nodes and zero otherwise.

The summary statistic is the arithmetic mean. The selection function is the rational quadratic kernel with a fixed scale parameter $\epsilon = 1.2$. In order to prevent particle depletion, the posterior GMs are evolved with a zero-centered random walk with variance 0.5. The initial population of particles is sampled from a normal distribution centered at the origin and with a diagonal covariance matrix with 10 on the diagonal.

This (indeed tractable) model with multimodal state variables is chosen in order to demonstrate the ability of the framework to approach the properties of the basic PF. The PF is initialized in the same way as the ABC filter. Unlike ABC, the particle filter *requires* fully known observation model of all the neighboring nodes.

Both ABC and PF run two scenarios: (i) no cooperation, where the node do not exchange any information, and (ii) cooperation with the adapt-then-combine strategy. The boxplots of the resulting root

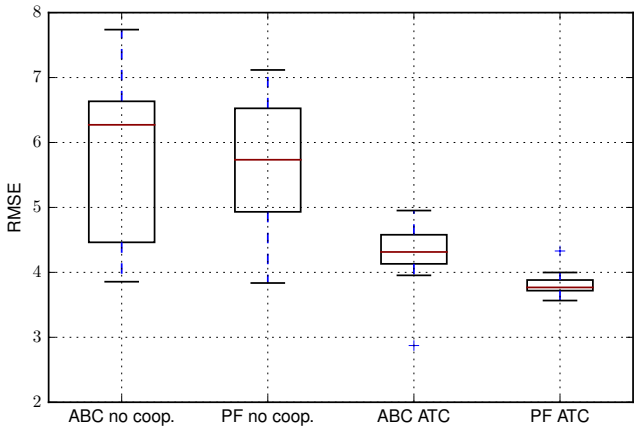

Fig. 2. RMSEs of network nodes. The first two boxplots display values for the ABC and particle filter (PF) without cooperation, the third and fourth boxplots display values for the ABC and particle filter with cooperation (adapt-then-combine, ATC).

mean square errors (RMSE) of the nodes are depicted in Fig. 2. One observes, that the cooperation leads to less diffuse RMSEs and smaller bias both in the diffusion ABC and PF cases. The ABC filtration is generally less efficient than PF as expected; however, in the adapt-then-combine scenario the performance of ABC is reasonably close to PF that has perfect knowledge of the models. In addition, the ABC diffusion filter has a much lower computational burden: no exponentials nor matrix inverses are necessary. Furthermore, local tuning of ε at individual nodes (even to fixed time-invariant values) would probably lead to further improvements. This is postponed for further research.

An important property of the considered distributed filtration framework is its self-stabilization: if all the posterior particles of some node have zero weights due to their location outside the acceptance region driven by $\tilde{g}_{t,\varepsilon}$, other nodes contribute to their resurrection. This may especially happen if the noise distribution is very heavy-tailed.

Our (unpublished) empirical results have shown convergence of the ABC filter to PF if the Gaussian kernel is used. Indeed, this was expected due to the similarity of the normal likelihood (PF) and the Gaussian kernel with reasonable ε (ABC).

6. CONCLUSION

Distributed filtering with approximate Bayesian computations can efficiently solve problems with intractable or unavailable likelihoods. This may be the case of sensor networks deployed over heterogeneous environment, where the true observation models (likelihoods) are only partially known, or simulation from them is easier than their direct computation. The employed framework is independent of the underlying sampling mechanism and hence applicable to any SMC and MCMC algorithms. Numerical simulations show, that the performance of the proposed algorithm is relatively close to the more demanding particle filters.

Future work will focus on more elaborate cooperation, and adaptive setting of the kernel bandwidth during sampling in order to improve convergence with a lower number of particles. The issue of ABC filtration of state-space models with linear substructure (similarly to marginalized particle filters) is currently under investigation.

7. REFERENCES

[1] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2069–2084, 2010.

[2] A. H. Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing*, vol. 3, R. Chellapa and S. Theodoridis, Eds. Academic Press, Elsevier, 2014, pp. 323–454.

[3] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[4] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, 2012.

[5] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Transactions on Signal Processing* vol. 56, no. 7, pp. 3375–3380, 2008.

[6] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypotheses" *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 814–825, 2010.

[7] Y. Wang and P. M. Djurić, "Reaching Bayesian consensus in cognitive systems by decision exchanges," Proceedings of the European Signal Processing Conference, Marrakesh, Morocco, 2013.

[8] Z. Weng and P. M. Djurić, " Efficient estimation of linear parameters from correlated node measurements over networks," *IEEE Signal Processing Letters*, vol. 21, no. 11, pp. 1408–1412, 2014.

[9] J. Hu, L. Xie, and C. Zhang, "Diffusion Kalman filtering based on covariance intersection," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 891–902, Feb. 2012.

[10] K. Dedecius, "Diffusion estimation of state-space models: Bayesian formulation," in *Proc. 2014 IEEE Workshop on Machine Learning for Signal Processing (MLSP2014)*, Reims, France, 2014, IEEE.

[11] O. Hlinka, F. Hlawatsch, and P. M. Djurić, "Distributed particle filtering in agent networks: A survey, classification, and comparison," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 61–81, Jan. 2013.

[12] O. Hlinka, F. Hlawatsch, and P. M. Djurić, "Consensus-based distributed particle filtering with distributed proposal adaptation," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3029–3041, 2014.

[13] M. G. S. Bruno and S. S. Dias, "Collaborative emitter tracking using Rao-Blackwellized random exchange diffusion particle filtering," *EURASIP J. Adv. Signal Process.*, vol. 19, pp. 1–18, Feb. 2014.

[14] J.-M. Marin, P. Pudlo, C.P. Robert, and R.J. Ryder, "Approximate Bayesian computational methods," *Statistics and Computing*, vol. 21, no. 2, pp. 289–291, 2011.

[15] A. Jasra, S. S. Singh, J. S. Martin, and E. McCoy, "Filtering via approximate Bayesian computation," *Statistics and Computing*, vol. 22, no. 6, pp. 1223–1237, May 2012.

[16] L. E. Calvet and V. Czellar, "Accurate methods for approximate Bayesian computation filtering," *Journal of Financial Econometrics*, July 2014. *(to appear)*

[17] G. M. Martin, B. P. M. McCabe, W. Maneesoonthorn, and C. P. Robert, "Approximate Bayesian computation in state space models," *arXiv:1409.8363*, Sept. 2014.

[18] J. S. Martin, A. Jasra, S. S. Singh, N. Whiteley, P. Del Moral, and E. McCoy, "Approximate Bayesian computation for smoothing," *Stochastic Analysis and Applications*, vol. 32, no. 3, pp. 397–420, Apr. 2014.

[19] P. Ramanathan, "Distributed particle filter with GMM approximation for multiple targets localization and tracking in wireless sensor network," in *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.* 2005, pp. 181–188, IEEE.

[20] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[21] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, June 2006.

[22] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *J. R. Stat. Soc. Ser. B (Statistical Methodology)*, vol. 72, no. 3, pp. 269–342, Jun. 2010.