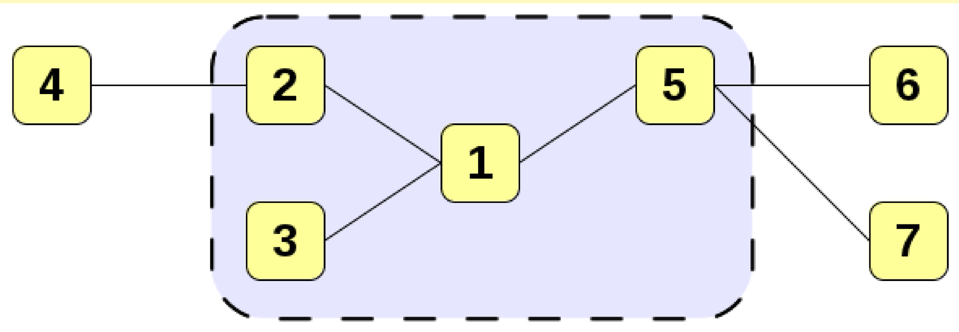
Dynamic Bayesian estimation in diffusion networks

Vladimíra Sečkárová seckarov@karlin.mff.cuni.cz MFF UK, Sokolovská 83, Praha
Kamil Dedecius dedecius@utia.cas.cz UTIA, Pod Vodárenskou věží 4, Praha

Introduction

We deal with the problem of collaborative estimation of unknown environmental parameter from noisy measurements. We focus on a recently formulated diffusion estimation problem, i.e., fully decentralized collaborative estimation in networks allowing the nodes to communicate only with their adjacent neighbours.



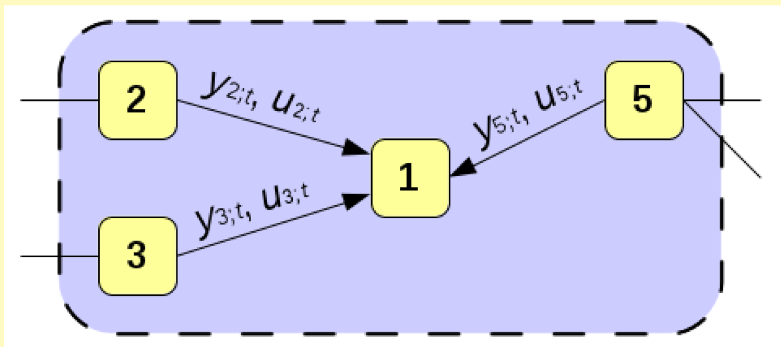
In this field, a couple of non-Bayesian estimation algorithms were proposed. However, these are mostly single problem oriented, e.g., on least-squares estimation [1], recursive least-squares (RLS, Cattivelli et al. [2]), least mean squares (LMS, Lopes and Sayed [3], Cattivelli and Sayed [4]), Kalman filters (Cattivelli et al. [2]) etc. We propose a new method called dynamic Bayesian diffusion estimation, which tackles the problem from the consistent and versatile Bayesian viewpoint and yields rather a methodology applicable to a much wider class of models, including, of course, the mentioned traditional ones.

Dynamic Bayesian Diffusion Estimation

Let there be a distributed network consisting of a set of nodes interacting with their neighbours, which collectively estimate the common parameter of interest using the same model structure. Furthermore, let us impose the following constraint: *the nodes are able to communicate one-to-one only within their closed neighbourhood*. Closed neighbourhood $\mathcal{N}_k$ of the kth node, $1 \leq k \leq M$, is defined as the set consisting of its adjacent nodes and node k. An example of a network including a closed neighbourhood $\mathcal{N}_1 = \{1,2,3,5\}$ of node $k = 1$ is drawn in the end of this section.
The diffusion estimation involves two subsequent steps, the former of which is optional but preferred:

Incremental update – also known as the data update, is a diffusion alternative of (1). The nodes propagate data within their closed neighbourhood and incorporate them into their local statistical knowledge;

Spatial update – the nodes propagate point parameter estimates (i.e. mean values) or posterior pdfs within their closed neighbourhood and correct their local estimates.



Incremental update

To develop the incremental update we use tools such as
– Bayesian decision theory,
– Kullback Leibler divergence,
– minimum cross entropy principle.
These yield a theoretically consistent incremental update

$$g_k(\boldsymbol{\Theta}|\overline{\boldsymbol{d}}(t)) \propto g_k(\boldsymbol{\Theta}|\overline{\boldsymbol{d}}(t-1)) \times \prod_{l\in\mathcal{N}_k} f_l(y_{l,t}|u_{l,t},\boldsymbol{d}_l(t-1),\boldsymbol{\Theta})^{c_{l,k}}, \quad (2)$$

where $\overline{\boldsymbol{d}}(t)$ stands for all data available from sources in $\mathcal{N}_k$ and $c_{l,k}$ are given weights representing the weight of lth node with respect to the kth one, $\sum_{l\in\mathcal{N}_k} c_{l,k} = 1$. The consistency of incremental update is guaranteed by the principle of weighted likelihoods [6, 7].

Spatial update

The spatial update follows after the incremental update. In this step, the nodes exchange information about unknown model parameter $\boldsymbol{\Theta}$, either in the form of its estimates or hyperparameters of its distribution. Formally, for fixed k, the information from all nodes in $\mathcal{N}_k$ describes the finite mixture density

$$g_k(\boldsymbol{\Theta}|\overline{\boldsymbol{d}}(t)) = \sum_{l\in\mathcal{N}_k} a_{l,k} g_l(\boldsymbol{\Theta}|\overline{\boldsymbol{d}}(t)), \quad \sum_{l\in\mathcal{N}_k} a_{l,k} = 1, \quad (3)$$

where $0 \leq a_{l,k} \leq 1$ is the weight of lth node's estimate from kth node's viewpoint.

Future work

The foreseen research activities comprise the analysis of properties of the diffusion estimator and a probabilistic method for dynamic determination of the weighting coefficients $a_{l,k}$ and $c_{l,k}$, $l \in \mathcal{N}_k$.

Bayesian recursive estimation

Let us consider a linear stochastic system with a real input variable u_t and a real output variable y_t, observed at discrete time instants $t = 1,2,\ldots$ The dependence of the output y_t on the previous data $\boldsymbol{d}(t-1) = \{y_0,u_0,\ldots,y_{t-1},u_{t-1}\}$ and the current input u_t can be modelled by a conditional probability density function (pdf)

$$f(y_t|u_t,\boldsymbol{d}(t-1),\boldsymbol{\Theta}),$$

where $\boldsymbol{\Theta}$ is a random potentially multivariate model parameter. By the assumption of natural conditions of control [5] we have

$$g(\boldsymbol{\Theta}|u_t,\boldsymbol{d}(t-1)) = g(\boldsymbol{\Theta}|\boldsymbol{d}(t-1)),$$

i.e., the information about parameter $\boldsymbol{\Theta}$ at time t is conditionally independent of the current input u_t. The Bayesian recursive estimation exploits the Bayes' rule to incorporate new data into the prior pdf of $\boldsymbol{\Theta}$ as follows

$$g(\boldsymbol{\Theta}|\boldsymbol{d}(t)) \propto f(y_t|u_t,\boldsymbol{d}(t-1),\boldsymbol{\Theta})g(\boldsymbol{\Theta}|\boldsymbol{d}(t-1)), \quad (1)$$

where $\propto$ denotes equality up to a normalizing constant. At the next time instant, the posterior pdf on the left-hand side of (1) is used as the prior pdf. The last relation is also known as the dynamic Bayesian data update.

References

[1] L. Xiao, S. Boyd, and S. Lall. A space-time diffusion scheme for peer-to-peer least-squares estimation. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 168–176. ACM, 2006.
[2] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed. Diffusion recursive Least-Squares for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing*, 56(5): 1865–1877, May 2008.
[3] C. G. Lopes and A. H. Sayed. Diffusion Least-Mean squares over adaptive networks: Formulation and performance analysis. *IEEE Transactions on Signal Processing*, 56(7):3122–3136, July 2008.
[4] F. S. Cattivelli and A. H. Sayed. Diffusion LMS strategies for distributed estimation. *IEEE Transactions on Signal Processing*, 58(3):1035–1048, March 2010.
[5] V. Peterka. Bayesian approach to system identification In P. Eykhoff (Ed.) Trends and Progress in System Identification, 1981.
[6] X. Wang. Asymptotic properties of maximum weighted likelihood estimators. *Journal of Statistical Planning and Inference*, 119(1):37–54, January 2004.
[7] X. Wang. Approximating Bayesian inference by weighted likelihood. *Can J Statistics*, 34(2):279–298, 2006.
[8] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley, 1st edition, January 1994.

Funding

The work was supported by the grant SVV 265 315.

Example – Gaussian linear regressive model

Given a regression vector $\boldsymbol{\psi}_t \in \mathbb{R}^n, t = 1,2,\ldots$ and a dependent random variable $y_t \in \mathbb{R}$, the Gaussian linear regressive model takes the form

$$y_t = \boldsymbol{\psi}_t^{\mathrm{T}}\boldsymbol{\theta} + \varepsilon_t, \quad (4)$$

where $\boldsymbol{\theta} \in \mathbb{R}^n$ is the regression coefficient and $\varepsilon_t \sim \mathcal{N}(0,\sigma^2)$ is the Gaussian white noise. This makes $y_t \sim \mathcal{N}(\boldsymbol{\psi}_t^{\mathrm{T}}\boldsymbol{\theta},\sigma^2)$ and the regression model (4) can be expressed by pdf $f(y_t|\boldsymbol{\psi}_t,\boldsymbol{\Theta})$. From the Bayesian viewpoint, the model parameters $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta},\sigma^2\}$ are also random variables. Under ignorance of their values, the proper conjugate prior distribution is the normal inverse-gamma ($\mathcal{N}i\mathit{\Gamma}$) one [8] with pdf:

$$g(\boldsymbol{\theta},\sigma^2|\boldsymbol{V},\nu) = \frac{\sigma^{-(\nu+n+1)}}{\mathcal{I}(\boldsymbol{V},\nu)}\exp\left\{-\frac{1}{2\sigma^2}\begin{bmatrix}-1\\ \boldsymbol{\theta}\end{bmatrix}^{\mathrm{T}}\boldsymbol{V}\begin{bmatrix}-1\\ \boldsymbol{\theta}\end{bmatrix}\right\}$$

where $\mathcal{I}(\cdot)$ is the normalization term, $\boldsymbol{V} \in \mathbb{R}^{N\times N}, N = n+1$, is a symmetric positive definite extended information matrix and $\nu \in \mathbb{R}$ denotes the degrees of freedom. Both $\boldsymbol{V}$ and ν are sufficient statistics [8] representing data $\boldsymbol{d}(t-1) = \{y_{t-1},\psi_{t-1},\ldots,y_0,\psi_0\}$. The Bayesian estimation (1) updates the sufficient statistics $\boldsymbol{V} \in \mathbb{R}^{N\times N}$ and $\nu \in \mathbb{R}$ by real scalar realization y_t and regression vector $\boldsymbol{\psi}_t \in \mathbb{R}^{N-1}$. The multivariate point estimator $\hat{\boldsymbol{\theta}}_t \in \mathbb{R}^{N-1}$ of regression coefficient is the mean value of the $\mathcal{N}i\mathit{\Gamma}$ distribution given by

$$\hat{\boldsymbol{\theta}}_t = \begin{bmatrix}V_{22} & \ldots & V_{2N}\\ \vdots & \ddots & \vdots\\ V_{N2} & \ldots & V_{NN}\end{bmatrix}_t^{-1}\begin{bmatrix}V_{21}\\ \vdots\\ V_{N1}\end{bmatrix}_t$$

In order to derive the dynamic Bayesian diffusion estimator of $\boldsymbol{\Theta}$, we follow the principles given above. Let us consider a network of $M \in \mathbb{N}$ distributed nodes. Each node $k \in \{1,\ldots,M\}$ evaluates a model

$$f(y_{k;t}|\boldsymbol{\psi}_{k;t},\boldsymbol{\Theta},\boldsymbol{V}_{k;t-1},\nu_{k;t-1})$$

and runs the diffusion Bayesian estimation (2) of its parameters in the form

$$g_k(\boldsymbol{\Theta}|\boldsymbol{V}_{k;t},\nu_{k;t}) \propto g_k(\boldsymbol{\Theta}|\boldsymbol{V}_{k;t-1},\nu_{k;t-1}) \times \prod_{l\in\mathcal{N}_k} f_l(y_{l;t}|\boldsymbol{\psi}_{l;t},\boldsymbol{\Theta},\boldsymbol{V}_{l;t-1},\nu_{l;t-1})^{c_{l,k}}.$$

Here $0 \leq c_{l,k} \leq 1$ weights lth node's data with respect to kth node, $l \in \mathcal{N}_k$, where $\sum_{l\in\mathcal{N}_k} c_{l,k} = 1$. Simply put, the kth node updates its prior pdf of $\boldsymbol{\Theta}$ by data from its closed neighbourhood $\mathcal{N}_k$. The incremental update of kth node's prior $\mathcal{N}i\mathit{\Gamma}$ pdf of $\boldsymbol{\Theta}$ by data $[y_{l;t},\boldsymbol{\psi}_{l,t}]^{\mathrm{T}}$, weighted by $c_{l,k}$, from its adjacent neighbours $l \in \mathcal{N}_k$ has the following form:

$$\boldsymbol{V}_{k;t} = \boldsymbol{V}_{k;t-1} + \sum_{l\in\mathcal{N}_k} c_{l,k}\begin{bmatrix}y_{l;t}\\ \boldsymbol{\psi}_{l;t}\end{bmatrix}\begin{bmatrix}y_{l;t}\\ \boldsymbol{\psi}_{l;t}\end{bmatrix}^{\mathrm{T}}$$

$$\nu_{k;t} = \nu_{k;t-1} + 1,$$

The spatial update (3) of the point estimate $\hat{\boldsymbol{\theta}}_{k;t}$ has the form

$$\hat{\boldsymbol{\theta}}_{k;t} = \sum_{l\in\mathcal{N}_k} a_{l,k}\hat{\boldsymbol{\theta}}_{l;t},$$

where $0 \leq a_{l,k} \leq 1$, $\sum_{l\in\mathcal{N}_k} a_{l,k} = 1$, $a_{l,k}$ denotes the weight of lth node's point estimate with respect to kth node.