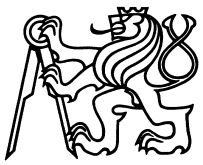
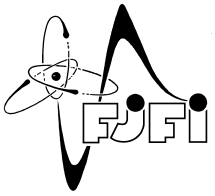


Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering



Distributed Kalman filtration under unknown spatially heterogeneous noise

Master Thesis

Author: Bc. Daniel Hnyk

Supervisor: Ing. Kamil Dedecius, Ph.D.

Academic year: 2017/2018

Acknowledgment: add ack

Declaration:

I hereby certify that this text represents my own work and that all used sources and materials are listed in the bibliography.

Prague, Monday 2nd April, 2018 Daniel Hnyk

Title: Distributed Kalman filtration under unknown spatially heterogeneous noise

Author: Bc. Daniel Hnyk

Branch of study: Mathematical Informatics

Abstract:
add abstract

Key words: Bayesian inference, Kalman filter

Contents

Introduction

Add introduction

Chapter 1

Bayesian probability

In this chapter, I summarize basics and preliminaries of Bayesian interpretation of probability and shortly compare it to frequentist approach. Reader is expected to know basic concepts from probability such as probability event, space, measure, product and sum rule, conditional and joint probability. All of these can be found in [1].

1.1 Probability interpretations

As it is a case with other branches of mathematics, probability theory has multiple interpretations with theirs own strengths and weaknesses. In this particular case, it is a subject of dispute if the term "interpretations" should be used, since there is no single formal system "probability". For example, some of the leading interpretations fail to satisfy the most common Kolmogorov's axiomatization and they still prove themselves useful[2]. Other axiomatization has been proposed for different interpretation [3] and some of the leading interpretations such as [1] don't even derive from *axioms*, but rather *desiderata*[1]. Interpretations don't differ only in the underlying set of axioms, but also in their practical usage or their epistemological status. Mathematicians as well as philosophers has been tackling these issues at least for over five centuries.

1.2 Why Bayesianism

It's hard to fully appreciate and understand main concepts and advantages of Bayesian reasoning if one is not familiar with other interpretations, such as frequentist. Its full description can be found in [4], here we'll cover just the basics arising from the comparison.

The main idea of Bayesian interpretation is its subjective nature, where there is no such thing as *real* probability in the world, but probabilities are in actor's minds. Probability then represents agent's *state of knowledge* or *degree of belief* [5, 6, 7]. This is not a negligible formal difference – it has important consequences, such as the fact that all agents with different state of knowledge can assign different probabilities to the same event. In Bayesian framework, one explicitly convey uncertainty in his models and objects (such as random variables).

Another important aspect worth mentioning is also the fact that Bayes interpretation allows doing predictions for events which haven't occurred yet (where frequentists are unable to do anything). This property is also connected to a fact that one can do partial (*online*) updates based on new evidence.

[1] (I) representations of plausibility are to be given by real numbers; (II) plausibilities are to be in qualitative agreement with common sense; and (III) the plausibilities are to be "consistent", in the sense that anyone with the same information would assign the same real numbers to the plausibilities

1.3 Principles of Bayesian probability

Core idea of Bayesian interpretation is surely a famous Bayes theorem 1.1, which is easily obtained from definition of conditional probability [8].

Theorem 1. Let P be a probability measure, **A** and **B** are events, $p(\mathbf{B}) \neq 0$. Then

$$p(\mathbf{A}|\mathbf{B}) = p(\mathbf{A}) \cdot \frac{p(\mathbf{B}|\mathbf{A})}{p(\mathbf{B})} \tag{1.1}$$

where:

1. $p(\mathbf{A})$ and $p(\mathbf{B})$ are the probabilities of observing A and B independently.
2. $p(\mathbf{A}|\mathbf{B})$ is the probability of observing event A given that B is true.
3. $p(\mathbf{B}|\mathbf{A})$ is the probability of observing event B given that A is true. This factor is called **likelihood**.

To relate this simple formula to previous section, consider a scenario where one wants to use Bayesian framework for *inference*. Suppose that **A** is a hypothesis whose probability can be affected by some evidence and **B** is an evidence for this hypothesis. $p(\mathbf{A})$ in formula 1.1 can then be seen as a **prior probability** – a probability one would assign to **A** is true *before* observing the evidence. Factor $p(\mathbf{B}|\mathbf{A})$, the *likelihood*, represents how likely we expect to observe **B** *given* that **A** is true. $p(\mathbf{B})$ is same for all possible hypothesis and serves as a normalizing factor (that is the reason why is often left out in many computations). And finally, $p(\mathbf{A}|\mathbf{B})$, which is the final **posterior probability** which is the factor one usually care about – the probability of **A** after observing **B**.

A factor $\frac{p(\mathbf{B}|\mathbf{A})}{p(\mathbf{B})}$ has also quite useful interpretation and that is the overall impact of evidence **B** on the probability of **A**.

1.4 Conjugate distributions and exponential family

Borrow this part from Mr Dedecius article page 3 - source code and cite it. There is informal definition right before Definition 2 and then consecutive much more detailed definitions. Also use rationale from before definition 2.

The updates can be significantly easier if *conjugate distributions* are used instead.

1.5 Statistical model

Chosen interpretation has a nontrivial influence on what such concept as *statistical model* is and how *inference* is done. I will follow traditional approach taken in [9, 10] and extend it of some necessary concepts from [11].

A statistical model is a pair of (S, M), where S is a set of possible observations (*sample space*) and M is a set of probability distributions on S. Distribution in M are expected to be approximately close to the "true" distribution which generates the data[2]. One of the main distinctions which can be made to separate families of statistical models is how distributions in M are described:

[2]There is a saying: "All models are wrong, but some of them are useful"

1. **Parametric models**: in this case, the set M is parameterized by some parameter θ and can have values from *finite* parameter space$\Theta \subseteq \mathbb{R}^d$. In other words: $M = \{P_\theta | \theta \in \Theta\}$.

2. **Nonparametric models**: this family is an unfortunate misnomer – nonparametric models also have parameters, but they differs from parametric models in a way that the parameter space is Θ is *infinite* – it is not fixed and can grow with the amount of data.

In Bayesian setting, parameters of the model have some prior distributions assigned. These distributions capture prior knowledge one posses before the application. For instance, this can be based on previous research or reliable enough observations. If there is no prior knowledge available, so called *uninformative* priors are being used instead. *Uninformative* is again a misnomer – what is meant by these is the fact that they are not subjectively elicited. Fully Bayesian model is then a model in which *all* parameters have some prior assigned. An interesting property of Bayesian approach arise – random variables (including latent variables) and parameters are treated in the same way.

1.6 Statistical inference

Process of inferring properties of the data underlying distribution is called statistical inference[12]. Formally, it's justification of restricting parameter space based on the data (by e.g. choosing a point estimate for a given parameter).

Unfortunately, for many real world scenarios, the approach of directly inferring the properties is not possible. One of many reasons can be to high dimensionality of the task or the form of the posterior distribution may be to complex, let alone the fact that it does not have to have analytical solution[13]. In these scenarios, approximation methods are being used instead.

1.6.1 Approximation methods

Using a breakdown from [13], approximation methods can be divided to two groups based on their stochastic or deterministic behavior. A notable example of the former is Markov chain Monte Carlo[13, (Chap.11)]. A representative approach of the latter is e.g. variational inference described in detail in section 1.7. The main difference is the fact that stochastic schemes, *given infinite resources*, are guaranteed to get the exact results and the approximation arise from the finite amount of those resources. On contrary, deterministic approximation schemes are based on *analytical approximations* to the posterior distribution and hence are generally not able to generate exact results. Both approaches are then complementary to each other, one being useful for situation where the second is unsuitable and vice versa.

1.7 Variational Bayes inference

In this work, a family of deterministic approximation methods *variational inference (*or *variational Bayes)* is used, excellently described in [11, (p. 463)] and [14]. This technique is based on using solution of an optimization problem to statistical inference, more exactly finding an input which minimize a specific functional. As was described in 1.6.1, this method is guaranteed to get an exact result given some family of possible input functions over which one minimizes. The approximation arise from limiting the possible inputs, for instance by considering only quadratic functions or, as is widely used and is also our case, functions which factorizes in a specific way.

Let $p(\mathbf{X}, \mathbf{Z})$ be a fully Bayesian model, where all prior distributions are given. Let denote the set of all parameters and all latent $\mathbf{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\}$ and the set of all observed variables

$\mathbf{X} = \{\boldsymbol{x_1}, \ldots, \boldsymbol{x}_N\}$. The goal of the inference is to find the posterior distribution $p(\mathbf{Z}|\mathbf{X})$ and the distribution of the model evidence $p(\mathbf{X})$[3]. Unfortunately, in many real world scenarios, $p(\mathbf{Z}|\mathbf{X})$ is almost always intractable (by e.g. trying to integrate all configurations of the hidden variables in denominator). That is where *variational lower bound* comes in. Instead of trying to compute $p(\mathbf{Z}|\mathbf{X})$ directly, we consider $q(\mathbf{Z})$ which is as close approximation as possible to the former and has a convenient and tractable form (their expectations are computable). Furthermore, these approximate distributions can also have their own *variational parameters* which are considered to be in $\mathbf{Z}$ as well. As a measure of closeness between the approximate $q(\mathbf{Z})$ and $p(\mathbf{Z}|\mathbf{X})$, Kullback-Leibler (KL) divergence 1.2 is used[4].

$$\mathrm{KL}(q||p) = -\int q(\mathbf{Z}) \ln \frac{p(\mathbf{Z}|\mathbf{X})}{q(\mathbf{Z})} \, \mathrm{d}\mathbf{Z} \tag{1.2}$$

In a case when $p(\mathbf{X}|\mathbf{Z}) = q(\mathbf{Z})$, Kullback-Leibler divergence is effectively zero. The goal is the following optimization problem:

$$q(\mathbf{Z}) = \arg\min_{q(\mathbf{Z})} \mathrm{KL}(q(\mathbf{Z})||p(\mathbf{Z}|\mathbf{X})). \tag{1.3}$$

The optimization 1.3 cannot be performed directly. To compute $\mathrm{KL}(q||p)$, the unknown evidence $p(\mathbf{X})$ is needed, as can be seen from the following derivation

$$\mathrm{KL}(q||p(\mathbf{Z}|\mathbf{X})) = \mathbb{E}[\ln q] - \mathbb{E}[\ln p(\mathbf{Z}|\mathbf{X})] = \mathbb{E}[\ln q] - \mathbb{E}[\ln p(\mathbf{Z}, \mathbf{X})] + \ln p(\mathbf{X})$$

.

Instead of trying to compute $\mathrm{KL}(q||p)$, we are going to optimize an alternative objective that is equivalent to $\mathrm{KL}(q||p)$ up to an added constant. Let us then define *variational lower bound* (also called *ELBO*) as follows:

$$\mathcal{L}(q) = \int q(\mathbf{Z}) \ln \frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})} \, \mathrm{d}\mathbf{Z} = \mathbb{E}[\ln p(\mathbf{Z}, \mathbf{X})] - \mathbb{E}[\ln q(\mathbf{Z})]. \tag{1.4}$$

It can be shown that$\mathcal{L}(q)$ has an important property being a lower bound of the log probability $\ln p(\mathbf{X}) \geq \mathcal{L}(q)$. Henceforth, when one wishes to maximize marginal $p(\mathbf{X})$, he can instead maximize its variational lower bound $\mathcal{L}(q)$.

Combining 1.4 and 1.2, the following relationship can be derived:

$$\ln p(\mathbf{X}) = \mathcal{L}(q) + \mathrm{KL}(q||p). \tag{1.5}$$

Since $p(\mathbf{X})$ doesn't depend on q, maximizing $\mathcal{L}(q)$ is possible and it is equivalent to minimizing $\mathrm{KL}(q||p)$.

Again, since we don't know the true posterior distribution, we work with some family of distributions $q(\mathbf{Z})$ for which $\mathcal{L}(q)$ becomes tractable and search for the candidate which maximizes it. Obviously, the choice of the distribution family is critical here – while it needs be tractable, it still needs to be flexible enough (as much as possible) to provide *accurate enough* approximation, goals usually going against each other. This is the part from where approximation arise in variational Bayes as has been discussed in 1.6.1.

[3] $p(\mathbf{X}) = \int p(\mathbf{Z}, \mathbf{X}) \, \mathrm{d}\mathbf{Z}$, which is usually unavailable

[4] We omit arguments of distributions where possible for readability. Most often: $q := q(\mathbf{Z})$ and $p = p(\mathbf{Z}|\mathbf{X})$

1.7.1 Factorized distributions

A common choice for such a family of distributions described in the chapter 1.7 has been *factorized distributions* formalized in physics theory called *mean field theory* [15]. The assumption on the family is quite simple – we treat each variable in $\mathbf{Z}$ as independent, in other words, the family **factorizes**:

$$q(\mathbf{Z}) = \prod_{i=1}^{M} q_i(\mathbf{Z})$$

It is clear that the assumption is often not met. Variables are dependent in the real world – it is the reason why it is so hard to obtain the posterior distribution directly.

Notice that we do not require any specific form of $q_i(\mathbf{Z_i})$. Also, it is worth noting that this is *not* a model of the observed data – it is the ELBO and KL minimization problem, which connects this to the model and data.

1.8 Coordinate ascent mean-field variational

Coordinate ascent mean-field variational (*CAVI*) is a commonly used algorithm for solving optimization problem 1.3[14][13].

1.8.1 Derivation

Optimal $q_i^*(\mathbf{Z_i})$

First, consider the complete conditional of $\mathbf{Z_i}$, which is $p(\mathbf{Z_i}|\mathbf{Z_{-i}}, \mathbf{X})$. If we fix all other variational factors $q_l(\mathbf{Z_l}), l \neq i$, the optimal $q_i(\mathbf{Z_i})$ satisfies the following[5]

$$q_i^*(\mathbf{Z_i}) \propto \exp\{\mathbb{E}_{-i}[\ln p(\mathbf{Z_i}|\mathbf{Z_{-j}}, \mathbf{X})]\} \propto \exp\{\mathbb{E}_{-i}[\ln p(\mathbf{Z_i}, \mathbf{Z_{\{-i\}}}, \mathbf{X})]\} \tag{1.6}$$

where in the last term, we use the fact that $q_{-i}(\mathbf{Z_{-i}}) = \prod_{l \neq i} q_l(\mathbf{Z_l})$.

Obtaining KL

Let us now rewrite $\mathcal{L}(q)$ in 1.4 as a function of $q_i(\mathbf{Z_i})$

$$\mathcal{L}(q_i) = \mathbb{E}_i[\mathbb{E}_{-i}[\ln p(\mathbf{Z_i}, \mathbf{Z_{\{-i\}}}, \mathbf{X})]] - \mathbb{E}_i[\ln q_i(\mathbf{Z_i})] + C \tag{1.7}$$

where in the first term is just an iterated expectation and in the second we retain only product term containing $q_i(\mathbf{Z_i})$, while other parts can be moved to a constant C thanks to using factorized distribution.

Now it's easy to see that 1.7 is a negative KL divergence between $q_i(\mathbf{Z_i})$ and $q_i^*(\mathbf{Z_i})$. That gives as all necessary part for the CAVI algorithm as described in [14]:

Algorithm 2. *CAVI*
Input: *A model $p(\mathbf{X}, \mathbf{Z})$, a dataset $\mathbf{X}$*
Output: *A variational density $q(\mathbf{Z}) = \prod_{i=1}^{m} q_i(\mathbf{Z_i})$*
Initialize: *Variational factors $q_i(\mathbf{Z_i})$*
while *the $\mathcal{L}(q)$ has not converged*
  for *$i \in \{1, ..., m\}$* ***do***
    Set $q_i(\mathbf{Z_i}) \propto \exp\{\mathbb{E}_{-i}[\ln p(\mathbf{Z_i}, \mathbf{Z_{\{-i\}}}, \mathbf{X})]\}$

[5] By $\mathbb{E}_{-i}[...]$, we denote expectation over all q distributions *except* q_i. Similarly, $\mathbf{Z_{\{-i\}}}$ means all latent variables except $\mathbf{Z_i}$.

end
Compute $\mathcal{L}(q) = \mathbb{E}[\ln p(\mathbf{Z}, \mathbf{X})] - \mathbb{E}[\ln q(\mathbf{Z})]$
end
return $q(\mathbf{Z})$

It is guaranteed that this algorithm converge to a local minimum.

1.9 Example

Combination of all of the above in a single example. Ideally with nice visualization.

Let us demonstrate the above with an example from [13].

The model Consider the i.i.d dataset $\mathbf{X} = \{x_1, \cdots, x_N\}$ generated by (unknown) Gaussian distribution $\mathbf{X} \sim \mathcal{N}(\mu, \tau^{-1})$. The goal is to infer posterior $p(\mu, \tau|\mathbf{X})$ and the joint probability is

$$p(\mathbf{X}, \mu, \tau) = p(\mathbf{X}|\mu, \tau)p(\mu|\tau)p(\tau)$$

Factorized distribution Let us now approximate $p(\mu, \tau|\mathbf{X})$ by $q(\mu, \tau)$ with the assumption that q factorizes: $q(\mu, \tau) = q(\mu)q(\tau)$.

Conjugated priors The complete data likelihood is Gaussian distribution and hence in exponential family. We set a non-informative conjugated priors for the μ and τ as follows

Is it correct? Is it likelihood to which we set conjugated priors?

$$\begin{aligned}\mu &\sim \mathcal{N}(\mu_0, (\lambda_0\tau)^{-1}) \\ \tau &\sim \text{Gamma}(a_0, b_0)\end{aligned}$$

where the hyperparameters $\mu_0, \lambda_0, a_0, b_0$ are initially set to some small positive number.

Form of $q(\mu)$ and $q(\tau)$ This is usually the hardest part of variational inference. The derivation of the following terms is according to 1.6 and in detail can be found in [13].

The derivations yields following for $q^*_\mu(\mu)$

$$\begin{aligned}q^*_\mu(\mu) &\sim \mathcal{N}(\mu \mid \mu_N, \lambda_N^{-1}) \\ \mu_N &= \frac{\lambda_0\mu_0 + N\bar{x}}{\lambda_0 + N} \\ \lambda_N &= (\lambda_0 + N)\mathbb{E}_\tau[\tau] = (\lambda_0 + N)\frac{a_N}{b_N} \\ \bar{x} &= \frac{1}{N}\sum_{n=1}^{N} x_n\end{aligned}$$

and for $q^*_\tau(\tau)$:

$$
\begin{aligned}
q_\tau^*(\tau) &\sim \text{Gamma}(\tau \mid a_N, b_N) \\
a_N &= a_0 + \frac{N+1}{2} \\
b_N &= b_0 + \frac{1}{2}\mathbb{E}_\mu[\sum_{n=1}^{N}(x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2] = \\
&= b_0 + \frac{1}{2}\left[(\lambda_0+N)\left(\lambda_N^{-1}+\mu_N^2\right) - 2\left(\lambda_0\mu_0 + \sum_{n=1}^{N} x_n\right)\mu_N + \left(\sum_{n=1}^{N} x_n^2\right) + \lambda_0\mu_0^2\right]
\end{aligned}
$$

Perform updates based on CAVI Initially, compute $N, \sum_{n=1}^{N} x_n, \sum_{n=1}^{N} x_n^2$ and set $\lambda_{N,t}$ to some random value. Then iterate these steps until convergence:

1. plug in $\lambda_{N,t}$ to obtain $b_{N,t}$
2. compute new $\lambda_{N,t+1}$ based on $b_{N,t}$ and all other terms needed

Is it correct that a_N, μ_N doesn't update? The "N" suffix here seems to be quite redundant.

Results After m iterations, we are now having new values of all parameters including hyper-parameters. Hence we are able to approximate the posterior distribution. The joint probability is

$$
\begin{aligned}
p(\mathbf{X}, \mu, \tau) &= p(\mathbf{X}|\mu,\tau)p(\mu|\tau)p(\tau) = \\
&= \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{x_n}|\mu_N, \tau_N^{-1})\mathcal{N}(\mu_N, (\lambda_N\tau_N)^{-1})\text{Gamma}(a_N, b_N)
\end{aligned}
$$

Is it correct, or should it be $\mathcal{N}(X|\mu_N, \tau_N^{-1}) \cdot \mathcal{N}(\mu_0, (\lambda_0\tau_N)^{-1}) \cdot \text{Gamma}(a_0, b_0)$. How the hell I get the posterior!?

g

Chapter 2

Distributed Kalman filter

This chapter firstly describes basic (non-distributed) version of Kalman filter. This is later extended to a distributed variant.

2.1 Introduction

Kalman filter (KF) is widely used adaptive filtering method firstly described in *A New Approach to Linear Filtering and Prediction Problems* by R.E. Kalman in 1960. The filter is a recursive solution to the discrete-data linear filtering problem [16]. The filter is defined by a set of relatively simple equations and allows for efficient online computation without having to know the whole history of data and it is the optimal linear filter (under several conditions described below). It is one of the solution to one of the most fundamental problems in control theory, the linear quadratic Gaussian control problem. Over the years, numerous variants and extensions have been derived from the original Kalman filter and this thesis is concerned with one such extension – distributed Kalman filtering. See 2.6 for a brief overview of used methods.

Kalman filter has been thoroughly studied over the years and also adopted by multiple industries, having a vast amount of applications wherever an estimation of the state of process is needed and noise in prediction or observation is present. Because of its nature, it is commonly used in navigation and control of robots and vehicles[17], including spacecrafts[18]. It can be applied in time series analysis and hence is used in fields such as econometrics[19] and economics[20] or signal processing[21]. Various applications can be found in a medicine, such as neuroscience[22].

With respect to the current development of the so called Internet of Things and development of cheap measurements sensors and microcomputers, Distributed Kalman Filtering (KDF) is becoming even more relevant. As one of the solution to *sensor fusion* problem, it has been under an intensive research as can be seen in [23]. A network consisting of multiple sensors has many advantages over a single sensor scenario, such as robustness to noise, better field of view or, if designed correctly, not having a single point of failure [24].

If needed, I could add information about the application - where is that predict-update part in each of them.

2.2 Description

The basic Kalman filter is used to estimate the current state of a linear dynamic system in case of noisy measurements. The state can be described by various variables, such as position and velocity. To use a Kalman Filter, one needs to specify system's:

1. control inputs, if any, such as sending a signal to steer a wheel
2. dynamic model, such as physical laws of motion
3. measurements, such as readings from sensors

The filter is useful in situations when dynamic model nor measurements cannot be entirely trusted (otherwise it one would not need to use it at all). It combines the information and while allowing for some noise – uncertainty – during prediction and measurements. The both pieces of information are weighted by so called Kalman's gain and depending on its value favors prediction over measurements and vice versa. The whole process is performed recursively, remembering the last estimate and covariance and hence does not need an entire history, which makes it attractive where memory is expensive. As a result a less noisy (hence a filter) and if used appropriately, better estimates are generated in each step, lying in between measurements and prediction.

Observable parameters coming from measurements does not have to be necessarily complete in a sense of describing the system in its entirety. It is a common case in which one observes only a subset of state variables and the filter can still estimate the entire state.

The distributed extension is easy to describe – instead of having only a single sensor, the system is measured by multiple of them and then somehow combining these measurements together. Despite its simple high level description, multisensor network of sensors must deal with various non-trivial problems and the whole computation is complicated based on the selected parameters of the network's behavior and topology. It also means getting into a distinctive research area of distributed data fusion[25] or consensus theory, which are out of the scope of this thesis.

obrázek:
https://medium.com/@mithi/object-tracking-and-fusing-sensor-measurements-using-the-extended-kalman-filter-algorithm-part-1-f2158ef1e4f0
https://www.google.cz/search?q=kalman+filter&safe=off&source=lnms&tbm=isch&sa=X&ved=0ahUKEv

2.3 Definition

mathematical definition, updates, equations, use that wikipedia chart (and maybe do it nicer if having more time)

Start with basic KF, then extension to (distributed) Diffusion Kalman Filtering. Nice intro with pictures in [24], simple def in [23] (eq. 2.1) linking to https://link.springer.com/article/10.1007/s11424-012-0275-2,
Additional definitions in [26] or (more complex description) in [27].

2.4 Note on optimality

The Kalman filter is the optimal filter if and only if the following conditions are fulfilled, which is rarely the case:

1. the model is exactly reflecting the reality
2. the noise is white
3. the covariances of the noise are exactly known

Optimality for distributed case: Track-to-Track Fusion scheme, e.g. in [24] in Introduction page 3.

why it's the problem, hence the next section ->

2.5 Noise analysis

What is their purpose and what problems we have with them, why it's necessary to nicely estimate them otherwise everything breaks down (described in the main VBAKF article on the beginning) + optimality issue.

2.6 Methods

This is what I want: [23]
Survey of currently used methods to deal with unknown noise

2.7 Distributed Kalman Filter

Chapter 3

Distributed Variational Bayes Adaptive Kalman Filtration

V zásadě [28] + naše rozšíření na DKF

Jde vlastně o spojení dvou věcí: vyřešení Noise estimation přes VB a použití v DKF. Nejdříve začít krátkým popisem jejich řešení, pak odvození mých updatů přes exponenciální rovnice (možná rovnou, ať neopisuji to jejich), výměna těch parametrů - tedy teorie. Tady začíná můj vlastní přínos.

3.1 Problem formulation

3.2 Bayesian model

3.2.1 Choosing prior distribution

IW et cetera...

3.2.2 Graphical model

writing the equation for full posterior, picture of graphical model

3.2.3 Deriving update equations

exponential family variant, writing the updates of hyperparameters

3.3 VBAKF algorithm

3.4 Distributed VBAKF algorithm

our novel extension

Chapter 4

Results

Jak si distribuovaná verze vede, jestli to vůbec konverguje a jak si vede vůči ostatním (dle času).

4.1 Implementation

Description of implementation, used libraries, if presentable, then some flow

4.2 Results

charts of comparisons, execution time, ...

Conclusion

Conclusion here

Bibliography

[1] E. T. Jaynes. "Probability Theory: The Logic of Science." In: *The Mathematical Intelligencer* 27.2 (2003), pp. 83–83. ISSN: 0343-6993. DOI: 10.1007/BF02985800. arXiv: arXiv:1011.1669v3.

[2] Edward N ed Hájek Zalta. "Interpretations of Probability". In: *The Stanford Encyclopedia of Philosophy*. 2012. ISBN: 1095-5054. URL: https://plato.stanford.edu/.

[3] Maurice J Dupré and Frank J Tipler. "New Axioms for Rigorous Bayesian Probability". In: *Bayesian Analysis* 1.1 (2004), pp. 1–8. URL: http://dauns.math.tulane.edu/%7B~%7Ddupre/OLDPUBLIC/20090804%20BA%20FNL%203.pdf.

[4] Charles Friedman. "The Frequency Interpretation in Probability". In: *Advances in Applied Mathematics* 23.3 (1999), pp. 234–254. ISSN: 0196-8858. DOI: https://doi.org/10.1006/aama.1999.0653. URL: http://www.sciencedirect.com/science/article/pii/S019688589990653X.

[5] R T Cox. "Probability, Frequency and Reasonable Expectation". In: *American Journal of Physics* 14.1 (1946), pp. 1–13. DOI: 10.1119/1.1990764. URL: https://doi.org/10.1119/1.1990764.

[6] B de Finetti. *Theory of Probability: A critical introductory treatment*. J. Wiley & Sons, Inc., New York, 1975.

[7] E.T. Jaynes. "Bayesian Methods: General Background". In: *Maximum Entropy and Bayesian Methods in Applied Statistics* August 1984 (1986), pp. 1–25. DOI: 10.1017/CBO9780511569678. URL: /chapter.jsf?bid=CBO9780511569678A007%7B%5C&%7Dcid=CBO9780511569678A007.

[8] A Stuart and K Ord. "Kendall's Advanced Theory of Statistics: Volume I—Distribution Theory". In: 1994.

[9] Dirk P. Kroese and Joshua C.C. Chan. *Statistical modeling and computation*. 2014, pp. 1–400. ISBN: 9781461487753. DOI: 10.1007/978-1-4614-8775-3.

[10] Peter McCullagh. "What is a statistical model?" In: *The Annals of Statistics* 30.5 (2002), pp. 1225–1310. ISSN: 00905364. DOI: 10.1214/aos/1035844977.

[11] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Vol. 53. 9. 2013, pp. 1689–1699. ISBN: 978-0-387-31073-2. DOI: 10.1117/1.2819119. arXiv: arXiv:1011.1669v3.

[12] G. Upton and I. Cook. *A Dictionary of Statistics*. 2008, p. 512. ISBN: 978-0-19-954145-4. DOI: 10.1093/acref/9780199541454.001.0001.

[13] Christopher M Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. ISBN: 0387310738.

[14] David M Blei, Alp Kucukelbir, and Jon D Mcauliffe. "Variational Inference: A Review for Statisticians". In: (2017). arXiv: arXiv:1601.00670v6. URL: https://arxiv.org/pdf/1601.00670.pdf.

[15] Giorgio Parisi. *Statistical Field Theory*. 1988.

[16] Greg Welch and Gary Bishop. "An Introduction to the Kalman Filter". In: (2006). URL: https://www.cs.unc.edu/%7B~%7Dwelch/media/pdf/kalman%7B%5C_%7Dintro.pdf.

[17] Shih-Tin Lin. "Force Sensing Using Kalman Filtering Techniques for Robot Compliant Motion Control". In: *Journal of Intelligent and Robotic Systems* 18.1 (1997), pp. 1–16. ISSN: 09210296. DOI: 10.1023/A:1007946400645. URL: http://link.springer.com/10.1023/A:1007946400645.

[18] David Gaylor and E. Glenn Lightsey. "GPS/INS Kalman Filter Design for Spacecraft Operating in the Proximity of International Space Station". In: *AIAA Guidance, Navigation, and Control Conference and Exhibit*. Reston, Virigina: American Institute of Aeronautics and Astronautics, Aug. 2003. ISBN: 978-1-62410-090-1. DOI: 10.2514/6.2003-5445. URL: http://arc.aiaa.org/doi/10.2514/6.2003-5445.

[19] Gurnain Kaur Pasricha. "M P RA Kalman Filter and its Economic Applications Kalman Filter and its Economic Applications". In: (2006). URL: http://mpra.ub.uni-muenchen.de/22734/.

[20] Ingvar Strid and Karl Walentin. "Block Kalman Filtering for Large-Scale DSGE Models". In: *Computational Economics* 33.3 (Apr. 2009), pp. 277–304. ISSN: 0927-7099. DOI: 10.1007/s10614-008-9160-4. URL: http://link.springer.com/10.1007/s10614-008-9160-4.

[21] S. Gannot. "Speech processing utilizing the Kalman filter". In: *IEEE Instrumentation & Measurement Magazine* 15.3 (June 2012), pp. 10–14. ISSN: 1094-6969. DOI: 10.1109/MIM.2012.6204866. URL: http://ieeexplore.ieee.org/document/6204866/.

[22] Daniel M. Wolpert and Zoubin Ghahramani. "Computational principles of movement neuroscience". In: *Nature Neuroscience* 3.Supp (Nov. 2000), pp. 1212–1217. ISSN: 10976256. DOI: 10.1038/81497. URL: http://www.nature.com/doifinder/10.1038/81497.

[23] Magdi S Mahmoud and Haris M Khalid. "Bibliographic Review on Distributed Kalman Filtering". In: (2013). URL: http://cogprints.org/8906/1/MsM-KFUPM-DCC-2D-One%7B%5C%%7D5B3R%7B%5C%%7D5D.pdf.

[24] Felix Govaers. "Distributed Kalman Filter". In: *Kalman Filters - Theory for Advanced Applications*. InTech, Feb. 2018. DOI: 10.5772/intechopen.71941. URL: http://www.intechopen.com/books/kalman-filters-theory-for-advanced-applications/distributed-kalman-filter.

[25] Mark E. Campbell and Nisar R. Ahmed. "Distributed Data Fusion: Neighbors, Rumors, and the Art of Collective Knowledge". In: *IEEE Control Systems* 36.4 (Aug. 2016), pp. 83–109. ISSN: 1066-033X. DOI: 10.1109/MCS.2016.2558444. URL: http://ieeexplore.ieee.org/document/7515322/.

[26] Ali H Sayed. "Diffusion adaptation over networks". In: (2013). arXiv: arXiv:1205.4220v2. URL: https://arxiv.org/pdf/1205.4220.pdf.

[27] Reza Olfati-Saber. "Distributed Kalman Filtering and Sensor Fusion in Sensor Networks". In: (). URL: https://pdfs.semanticscholar.org/fd77/043a3e588e03028589615b51f51060074a5b.pdf.

[28] Yulong Huang et al. "A Novel Adaptive Kalman Filter With Inaccurate Process and Measurement Noise Covariance Matrices". In: *IEEE Transactions on Automatic Control* 63.2 (Feb. 2018), pp. 594–601. ISSN: 0018-9286. DOI: 10.1109/TAC.2017.2730480. URL: http://ieeexplore.ieee.org/document/8025799/.