Filtering

Thorsten Schmidt

17. Juli 2015

University of Freiburg, Dep. of Mathematical Stochastics, Eckerstr. 1, 79106 Freiburg, Germany. Email: thorsten.schmidt@stochastik.uni-freiburg.de

Filter theory is a modern area of statistics and probability theory where the goal is to estimate a signal which can only be observed under additional noise. In contrast to classical statistics, the unobserved signal changes dynamically over time. This opens the area for a wide range of applications, and the topics where filtering is applied turns out to be extremly rich: global positiong system (GPS) receivers, advanced driver assistance systems, online monitoring in the intensiv care of patients, model estimation in financial applications, denoising of fotographs or language recordings, and mobile robotics, to just name a few. In particular, for the application to mobile robotics the problem of estimating the robot's position given the available information has been termed

> the most fundamental problem to providing a mobile robot with autonomous capabilities"[1]

which highlights the difficulty of the area as well as the importance of fast and reliable methodologies. A prime example is the kidnapped robot, one or more robots which are carried to an location posing a big challenge on the robot's localization systems.

This article gives a short overview of the historic developments in filtering theory and illustrates the basic concepts of Kalman filtering and particle filtering to grasp some intuition.

To this end, let us state shortly what a (linear) filtering problem is. The aim is to estimate an unboserved and dynamical signal. The signal at time t is denoted by X_t with $t = 0, 1, 2, \dots$. The observer only has access to a noisy observation: the observation at time t is denoted by Y_t. For example, let the signal X_t denote the location of a robot at current time t. Given that the robot was at position X_t it would move to X_{t+1} in the next time step, depending on our steering. However, we do not have full information about the position, so X_t might take different values (for example locations A, B and C or a full interval $[A, B]$), each with different probabilities (respectively a density in the interval case). The observation Y_t could be a GPS signal ore other information which of

[1] See Cox (1991); cited from Fox et al. (2001) with provides a nice introduction to particle filter applications in mobile robotics.

consists of X_t plus noise. Our goal is to optimally use the observation in our estimation procedure. We do so by updating our probabilities for the location which leads to an improved estimate of the actual position.

Depending on the setup and the application in view, the dependence of $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ can be modelled in various ways. The simplest type of dependence is the following linear structure:

$$Y_t = aX_t + \epsilon_t, \quad t = 0, 1, 2, \dots$$

where a is a constant and ϵ_t denotes the noise. We will assume that the noise is independent of X, has mean zero and unit variance. Then the constant a determines the degree of information in the data: clearly, if $a = 0$, no information is included in the data. On the other side, if a is very high, the data has extremly good quality and estimation will be very easy. The more complicated, nonlinear variants of the filtering problem are obtained by replacing aX_t with $a(X_t)$ where a is a nonlinear function.

It is not surprising, that such a fundamental system has been studied for a long time by mathematicians. When $X_t \equiv \theta$ is simply a constant, one arives at the famous linear regression problem in statistics, where Gauss already provided a solution via the least squares approach, see Gauß (1809).

1 History of filtering

The dynamical filtering problem was firstly studied in the 40's by the great architects of probability theory, Andrey N. Kolmogorov and Norbert Wiener. While Kolmogorov studied the problem stated as above in discrete time, Wiener considered already the continuous-time analogue.

It took a while until further developements should appear. Surprisingly, the breakthrough in the field was achieved by studying a important simplification: the *linear* case. It was Rudolf E. Kálmán in 1960 who published a now famous article in an engineering journal studying the linear case with Gaussian noise, Kalman (1960), see Stepanov (2011) for a historical account on the key discoveries in the development of the filter. The main advantage of the simplification to the linear case is that the solution of the filtering problem then can be calculated explicitely. This allowed the article to reach a widespread audience. We will later discuss the conceptual outcomes of this approach shortly. Richard S. Bucy obtained similar results independently, and jointly they solved the linear filtering problem in continuous time which is now called the Kalman-Bucy filter. "The most famous early use of the Kalman filter was in the Appollo mission which took Neil Amstrong to the moon, and (most importantly) brought him back. Today, Kalman filters are at work in every satellite navigation device, every smart phone, and many computer games"[2]. Not very surprising, the most delicate aspect is the return trajectory of the missile towards the earth, where the Apollo command module must enter the atmosphere at a highly precise level. For a historical account on the use of the

[2] cited from Faragher (2012), which also gives a nice intuitive discussion of the Kalman filter.

Kalman filter in the Apollo mission and the various tweaks and adjustements which have been developed while implementing this approach, see Grewal and Andrews (2010).

Many other scientists contributed to the problem in those years, amongst them Fujisaki, Kallianpur, Kushner, Liptser, Shiryaev, Striebel, Stratonovich, and Wonham. The researchers started to tackle the general problem and ran into great difficiulties: except for occasional special cases, the filtering problem immediately becomes infinite-dimensional. Compare this to the Kalman-Bucy filter, which in the simplest case allowed to obtain a two-dimensional solution! Not surprisingly, one has tried to stretch the linear problem as far as possible by using locally linear approximations, which is lead to such variants like the extended Kalman filter. An important step was the observation by Moshe Zakai[3] in 1969, that studying the un-normalized densities leads to a linear stochastic partial differential equation for the filtering problem, the so-called Zakai equation. See also the example on particle filtering below, where a similar step increases performance of the algorithm.

The area developed rapidly in the 1980's and 1990's. Many reasearchers tackled the problem from various sides and a suitable level of generality was obtained. In the 1990s, a lot of the research done focussed on numerical algorithms for solving the filtering problem. This is quite natural, recalling that typically no explicit solutions are available. A nowadays very successful numerical method is the so-called particle filter. This approach estimates the distribution of the unkown signal by discretization into finitely many particles and applying the strong law of large numbers. The approach origins from earlier works on sequential Monte Carlo methods and first ideas can be traced back to early works in the 50's, in particular to the work of Alan Turing. The name "particle filter" was first used in Del Moral (1996) and we give some details in an example below.

From the late 90's on, filtering also became increasingly important in the applications in mathematical finance, where one often acknowledges incomplete information and filtering is the ideal tool to treat this. Applications have been in portfolio optimization, calibration of financial models, insider trading, hedging and many more. A prime example is credit risk, where the firm value is a fundamental variable, which is of course not known to market participants. However, there are various sources of information, like quarterly reports, stock, bond and options prices and the goal is to estimate the credit riskiness of the company from this information. Filtering has proved a suitable method to achieve this very effectively, see for example Frey and Runggaldier (2011).

For a historical overview and a comprehensive work on filtering without jumps we refer to Bain and Crisan (2009) and for a treatment from different viewpoints see the compendium Crisan and Rozovskii (2011).

The Kalman filter. For building up intuition on the Kalman filter, we revisit the example from the introducion with the following specialisation: assume that

$$Y_t = aX + \epsilon_t, \quad t = 0, 1, 2, \dots \tag{1}$$

[3] See Zakai (1969).

with X being normally distributed with mean zero and, for simplicity, variance 1 and $\epsilon_1, \epsilon_2, \ldots$ being standard normally distributed, and X and $\epsilon_i, i \geq 1$ are also independent. The aim of filtering is to determine the minimal variance estimate for X given the observation until time t, which we denote by $\hat{X}_t$. It turns out that the conditional expectation $\hat{X}_t = E[X|Y_1, \ldots, Y_t]$ is indeed this optimal estimator. Joint normality implies that X can be written as

$$X = \alpha_1 Y_1 + \cdots + \alpha_t Y_t + \eta$$

where η is again normally distributed with mean zero, but it is also independent from $Y_1, \ldots, Y_t$. Then, $\hat{X}_t = \alpha_1 Y_1 + \cdots + \alpha_t Y_t$, as $E[\eta|Y_1, \ldots, Y_t] = 0$ and the Kalman filter is obtained once the coefficients α_i are computed. The key tool for this is a careful study of the covariances: consider for simplicity $t = 2$. On the one hand,

$$\mathrm{Cov}(X, Y_1) = \mathrm{Cov}(X, X + \epsilon_1) = \mathrm{Var}(X) = 1 = \mathrm{Cov}(X, Y_2).$$

On the other hand,

$$\begin{aligned}\mathrm{Cov}(X, Y_1) &= \mathrm{Cov}(\alpha_1 Y_1 + \alpha_2 Y_2 + \eta, Y_1) = \alpha_1 \mathrm{Var}(Y_1) + \alpha_2 \mathrm{Var}(X) \\ &= \alpha_1(a^2 + 1) + \alpha_2\end{aligned}$$

and, similarily, $\mathrm{Cov}(X, Y_2) = \alpha_1 + \alpha_2(a^2 + 1)$. This gives two linear equations for α_1 and α_2 and we obtain $\alpha_1 = \alpha_2 = (2 + a^2)^{-1}$. In a similar way we obtain

$$\hat{X}_t = \frac{1}{t + a^2} \sum_{i=1}^{t} Y_i.$$

The conditional variance now is easily computed and turns out to be independent of the past, i.e. a constant. The Kalman filter inherits a similar structure: the covariance is a deterministic function which can be computed explicitely, and, the filter has a recursive structure: $\hat{X}_{t+1}$ can be computed from $\hat{X}_t$ and the information update, Y_{t+1}.

Surprisingly, this simple approach can be generalised in a tedious exercise one can compute the Kalman filter. This perspective allows us to understand the following quote from Kalman himself: "(...) the discovery of the Kalman filter (January 1959) came about through a single, gigantic, persistent mathematical exercise. (...) Just as Newton was lucky having timed his birth so as to have Kepler's laws ready and waiting for him, I was lucky, too."[4]

The interested reader, who wants to further expand this example to the Kalman and the Kalman-Bucy filter, is referred to Frey and Schmidt (2011). This article gives also a more general setup including jumps, applications in mathematical finance, and further references.

Particle filtering Finally, we discuss shortly the particle filter, or sequential Monte Carlo methods, which build an intuitive and flexible numerical scheme to treat nonlinear

[4] Cited from Stepanov (2011).

filtering problems. Of course, there are numerous other numerical approaches and we refer again to Bain and Crisan (2009) for literature and a thorough treatment.

The particle filter is well-suited for the non-linear case of our example. To get intuition, assume for a moment that $X_t = x$ and $Y_t = y$ have positive probability. Then, by Bayes' rule, we have that

$$P(X_t = x|Y_t = y) = \frac{P(Y_t = y|X_t = x)P(X_t = x)}{P(Y_t = y)}. \tag{2}$$

Replacing t P with the conditional probability $P(.|Y_{t-1}, \ldots, Y_1)$ does not violate the argument and we obtained a nice result on discrete filtering. It can be generalized to the continuous case with the used of densities. However, this argument gives the insight that by Bayes' rule the filtering problem can be solved with the knowledge of $P(Y_t \in A|X_t = x)$ and $P(X_t \in B|Y_{t-1}, \ldots, Y_1)$ (the prediction) as well as the normalizing factor $P(Y_t \in B|Y_{t-1}, \ldots, Y_1)$. This means that the filter starts with a prediction on the future state and in a second step incorporates the arriving information.

The particle filter mimicks this procedure. It approximates the conditional distribution by a finite number of particles having different weights. In a simple variant it considers a fixed number of particles which are initially distributed according to the initial distribution. In each iteration, there are two steps: first, the particles move randomly according to the law of X to new positions. Seconds, the weights of the particles are adjusted incorporating the new information with a variant of (2). Each particle gets offsprings proportional to the weights. Different choices for the distributions of the offsprings lead to different algorithms with different properties.

Summarizing, filtering can be applied in a large variety of situations and provides a flexible instruments for various questions. New challenges from applications and upcoming questions stipulate the research and the development of improved algorithms and methodologies.

Literatur

Bain, A. and Crisan, D. (2009), *Fundamentals of Stochastic Filtering*, Springer, New York.

Cox, I. J. (1991), 'Blanche-an experiment in guidance and navigation of an autonomous robot vehicle', *Robotics and Automation, IEEE Transactions on* **7**(2), 193–204.

Crisan, D. and Rozovskii, B., eds (2011), Oxford Univ. Press.

Del Moral, P. (1996), 'Nonlinear filtering: Interacting particle solution', *Markov Processes and Related Fields* **2**(4), 555–580.

Faragher, R. (2012), 'Understanding the basis of the kalman filter via a simple and intuitive derivation', *IEEE Signal Processing Magazine* .

Fox, D., Thrun, S., Burgard, W. and Dellaert, F. (2001), Particle filters for mobile robot localization, *in* N. d. F. A. Doucet and N. Gordon, eds, 'Sequential Monte Carlo Methods in Practice', Springer Verlag. Berlin Heidelberg New York, pp. 499 – 516.

Frey, R. and Runggaldier, W. (2011), Nonlinear filtering in models for interest-rate and credit risk, *in* 'The Oxford handbook of nonlinear filtering', Oxford Univ. Press, Oxford.

Frey, R. and Schmidt, T. (2011), Filtering and incomplete information, *in* T. Bielecki, D. Brigo and F. Patras, eds, 'Credit Risk Frontiers', Wiley.

Gauß, C. F. (1809), *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*, Volume 2.

Grewal, M. S. and Andrews, A. P. (2010), 'Applications of Kalman filtering in aerospace 1960 to the present [historical perspectives]', *Control Systems, IEEE* **30**, 69 – 78.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Transactions of the ASME–Journal of Basic Engineering* **82**(Series D), 35–45.

Stepanov, O. (2011), 'Kalman filtering: past and present. An outlook from Russia. (On the occasion of the 80th birthday of Rudolf Emil Kalman)', *Gyroscopy and Navigation* **2**(2), 99–110.

Zakai, M. (1969), 'On the optimal filtering of diffusion processes', *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **11**, 230–243.