Density-based Nonlinear Estimation - Sum of Gaussians Filter

Bruno Miguel Gil Rosa[1]

[1]Instituto Superior Técnico
Av. Rovisco Pais 1
1049-001 Lisboa

email: brunomiguel85@hotmail.com

Abstract: *A nonlinear filter is associated to the estimation of the state of a dynamical system from noisy measurements when either the system dynamics is a nonlinear function of the state and/or the disturbances do not follow normality conditions. Until now, there is no closed-form solution to the nonlinear estimation problem and one is forced to use approximation techniques. The Gaussian Sum filter (GSF) is a density-based approximation technique which involves a weighted average process over a collection of Extended Kalman filters (EKF). Since its introduction in the early 70's, the GSF has been widely applied in many dynamical stochastic problems but still with no overall convergence results accepted.*

In this paper, one derives the theoretical calculations behind the GSF, taking as starting point the Bayes' theory. Then, some numerical examples are provided that test the performance of the GSF. In the end, a discussion about the merits and drawbacks of the GSF is also presented.

Keywords: *Nonlinear Estimation, Stochastic Dynamical Systems, Bayesian Framework, Density-based Estimates, Gaussian Sum Filter.*

Introduction

In estimation theory, some sort of estimator is used to obtain the state of a stochastic dynamical system from noisy measurements. This is so, because it is assumed that the desired information is, somehow, embedded in a noisy signal: noise adds uncertainty without which the problem would become deterministic and the estimation process would not be needed.

The system's dynamics in the estimation procedure can be modelled as linear if both transition and measurement equations are linear functions of the state-variables involved or nonlinear otherwise, while the noise can be modelled by any probability density function, being the normal or gaussian one the most used due to the simpler statistics involved [1].

The Bayes' theory has been used ever since in the estimation process by providing a systematic method on how to include prior knowledge to the system being modelled [6]. From this point on, a separation has been made on the bayes' framework for handling linear and nonlinear estimation problems. In the linear case, the optimal solution can be explicitly written down, resulting in the celebrated Kalman-Bucy filter (1960) [2]. However, for the nonlinear case, no closed-form solution has been yet purposed and one is forced to seek approximation techniques.

The function-based techniques generally approximate the nonlinear model by a linear one (using 1st or 2nd order Taylor's series approximation) and then use the optimal Kalman filter for the linearized model. This works well if the nonlinearities are not too severe but, if one looks deeply in the derived system, one will realise that is solving the wrong problem: by keeping the original nonlinear model, then only this way can one pose the optimal solution even if it does not exist or it is impossible to solve the equations analytically.

The density-based approximation techniques had emerged to tackle the difficulties encountered in the function-based approximations in order to obtain less biased filtering estimates of the state-variables. Within this approach, several methods have been proposed: the density approximation as a sum of a gaussian distribution, the density approximation through numerical integration, the simulation-based density approximation using Monte-Carlo integration with importance sampling and a recursive algorithm of random draws with use of rejection sampling.

The first publications to pose a general solution to the nonlinear estimation problem date back to 1970 by Jazwinski [2]. By its turn, the main references to the Gaussian Sum filter are presented by Sorenson and Alspach (1971), Alspach and Sorenson (1972) and Anderson and Moore (1979).

Methods

A. Recursive Bayesian Estimation

The objective of the estimation problem is to gather information about the parameter x given an observation of an experimental outcome z. In the Bayesian framework, both vectors x and z are treated as random variables [7]. The vector x is assumed having a known prior density function $p(x)$ that encompasses everything known (and unknown) about the parameters before the experimental outcome. By its turn, the observation vector is assumed to have a probability density function belonging to a class indexed by the

parameters, $p(z \mid x)$.

After observing the event, the knowledge about the parameter is altered, so that, its new value it's now given, in accordance to the Bayes' rule, by:

$$p(x \mid z) = \frac{p(z \mid x)p(x)}{p(z)} \tag{1}$$

A purely bayesian view on the estimation problem states that the *a posteriori* density function $p(x \mid z)$ describes everything about the parameter x after the experimental outcome has been obtained [6, 7]. The denominator of expression (1) is just a scalar positive constant and it can be found by marginalization as,

$$p(z) = \int_{\mathbb{R}^n} p(z \mid x)p(x)dx \tag{2}$$

and one needs only to specify the product inside the integral.

In many applications of estimation theory, the parameters admit some dynamical properties that make them change with time (or the estimate of the parameters is needed *online*) so, the information about the parameters needs to be updated with each measurement. This is true in recursive estimation problems, where the parameters of a dynamical system working in a closed-loop are tracked using measurements of system input and output.

Let x_k denote the state at time index k ($k \in \mathbb{N}$). Recursive estimation is about obtaining an estimate of the current state given information about the last estimate and the current measurement, z_k. A useful notation is z^k which denotes all measurements up to time index k, i.e., $\{z_i\}_{i=0}^{k} = \{z_1, z_2, ...z_k\}$. Hence, conditioned on the present state, no additional information about the future states should be available in past observations. This implies that the states evolve in time according to a Markov process with an initial state $x_0 \sim p(x_0)$ and transition kernel of the form:

$$p(x_{k+1} \mid x_0, x_1, ..., x_k) = p(x_{k+1} \mid x_k) \tag{3}$$

The measurement observed at the time instant k is conditionally independent of the previously observed measurements given the current state value:

$$p(z_k \mid x_0, x_1, ..., x_k) = p(z_k \mid x_k) \tag{4}$$

A recursive estimation problem is uniquely defined when the $prior\, p(x_0)$, the transition kernel $p(x_{k+1} \mid x_k)$ and the likelihood $p(z_k \mid x_k)$ are given. One is interested on how to update the *a posteriori* or filtering density function as new measurements arrive. Resorting to the Jazwinski's work (1970), this density is recursively computed from the following relations, where the aforementioned Bayes' rule and Markov property have been used:

$$p(x_k \mid z^k) = \frac{p(z_k \mid x_k)p(x_k)}{p(z^k)} = \frac{p(z_k, z^{k-1} \mid x_k)p(x_k)}{p(z_k \mid z^{k-1})} =$$

$$= \frac{p(z_k \mid z^{k-1}, x_k)p(z^{k-1} \mid x_k)p(x_k)}{p(z_k \mid z^{k-1})p(z^{k-1})} =$$

$$= \frac{p(z_k \mid x_k)p(x_k \mid z^{k-1})p(z^{k-1})p(x_k)}{p(z_k \mid z^{k-1})p(z^{k-1})p(x_k)} =$$

$$= \frac{p(z_k \mid x_k)p(x_k \mid z^{k-1})}{p(z_k \mid z^{k-1})} \tag{5}$$

Moreover, the denominator of the *a posteriori* density function can be expressed in the same way as relation (2), i.e., by marginalization, one gets:

$$p(z_k \mid z^{k-1}) = \int_{\mathbb{R}^n} p(z_k \mid x_k)p(x_k \mid z^{k-1})dx_k \tag{6}$$

The filtering density function $p(x_k \mid z^k)$ just derived is used to obtain the one-step prediction or time-update $p(x_{k+1} \mid z^k)$ as:

$$p(x_{k+1} \mid z^k) = \int_{\mathbb{R}^n} p(x_{k+1}, x_k \mid z^k)dx_k =$$

$$= \int_{\mathbb{R}^n} p(x_{k+1} \mid x_k, z^k)p(x_k \mid z^k)dx_k =$$

$$= \int_{\mathbb{R}^n} p(x_{k+1} \mid x_k)p(x_k \mid z^k)dx_k \tag{7}$$

The integrals on the previous expressions only permit an analytical solution in a few special cases. The most important special case is when the dynamical model is linear and the initial conditions and noises are normally distributed. One of the attempts to apply the Bayesian framework to more general models is presented next, where the nonlinearities are tackled by using Gaussian sum approximations.

B. Gaussian Sum Approximations

The Gaussian sum filter is a nonlinear algorithm that involves computing collections of extended Kalman filters. Since it is impossible to obtain exact representations of the *a posteriori* density for nonlinear systems, it is reasonable to seek approximations that will make expression (5) more tractable [3].

One begins by writing down the general expression for the nonlinear dynamical systems. It is composed by the plant and measurement equations as,

$$\begin{cases} x_{k+1} = f_{k+1}(x_k) + w_k \\ z_k = h_k(x_k) + v_k \end{cases} \tag{8}$$

where w_k and v_k are both Gaussian white-noise sequences with zero means and covariances Q_k and R_k, respectively. Then, a filtering procedure based on the bayesian framework is developed in order to obtain $p(x_k \mid z^k)$ in conjunction with the linearization step that has been used in some Kalman filter incursions into the nonlinear model field. The development starts before acquiring the new measurement so that, the density function is given by a weighted sum of gaussian terms with small covariance matrix involved:

$$p(x_k \mid z^{k-1}) = \sum_{i=1}^{\varepsilon_k'} \alpha_{ki}' N(x_k - a_{ki}, P_{ki}') \tag{9}$$

where,

$$N(x_k - a_{ki}, P_{ki}') =$$

$$= (2\pi)^{-n/2} \left|P_{ki}'\right|^{-1/2} \exp\left(-\frac{1}{2}(x_k - a_{ki})^T \left(P_{ki}'\right)^{-1} (x_k - a_{ki}) \right)$$

$$\sum_{i=1}^{\varepsilon_k'} \alpha_{ki}' = 1, \quad \alpha_{ki}' \geq 0 \quad \forall i$$

It can be shown that expression (9) converges uniformly to any density function of practical interest as the number of terms in the sum increases and the covariance approaches the zero matrix. Inserting expression (9) into (5), one gets the following general relation for the *a posteriori* density function that one seeks:

$$p(x_k \mid z^k) = c_k \left[\sum_{i=1}^{\varepsilon_k'} \alpha_{ki}' N(x_k - a_{ki}, P_{ki}') p(z_k - h_k(x_k)) \right] \tag{10}$$

with c_k being a scalar positive quantity obtained in a similar way as expression (2).

As in the of the Extended Kalman filter, the linearization of h_k relative to a_{ki} allows that the term $p(z_k - h_k(x_k))$ can be approximated by a Gaussian in the relative small region nearby a_{ki}. Hence, last expression can be rewritten as:

$$p(x_k \mid z^k) = c_k \left[\sum_{i=1}^{\varepsilon_k'} \alpha_{ki}' N(x_k - a_{ki}, P_{ki}') N(\xi_i, R_k) \right] \tag{11}$$

where,

$$\xi_i = z_k - h_k(a_{ki}) - H_{ki}\left[x_k - a_{ki}\right]$$

$$H_{ki} = \left.\frac{\partial h_k}{\partial x_k}\right|_{a_{ki}}$$

Expression (11) is reduced to a more compact way in order to resemble the relations derived for the extended Kalman filter in what concerns gain and updated covariances:

$$p(x_k \mid z^k) = \sum_{i=1}^{\varepsilon_k} \alpha_{ki} N(x_k - y_{ki}, P_{ki}) \tag{12}$$

with,

$$y_{ki} = a_{ki} + K_{ki}(z_k - h_k(a_{ki})) \quad \text{(mean values)}$$

$$P_{ki} = P_{ki}' - K_{ki} H_{ki} P_{ki}' \quad \text{(covariance matrix)}$$

$$K_{ki} = P_{ki}' H_{ki}^T \left[H_{ki} P_{ki}' H_{ki}^T + R_k \right]^{-1} \quad \text{(filter gains)}$$

$$\alpha_{ki}' = \frac{\alpha_{ki}' \beta_{ki}}{\sum_{j=1}^{\varepsilon_k'} \alpha_{ki}' \beta_{kj}}$$

$$\beta_{kj} = N(z_k - h_k(a_{ki}), H_{ki} P_{ki}' H_{ki}^T + R_k)$$

Hence, the Gaussian sum representation is formed as the convex combination of the output of several extended Kalman filters in parallel [4].

Now, let's focus attention in the problem of obtaining the Gaussian sum of the one-step prediction density, $p(x_{k+1} \mid z^k)$. According to [4], two cases must be considered: firstly, when there is little plant noise corrupting the system dynamics and, secondly, when there is a significant amount of noise.

In the first case, the covariance of the plant noise is comparable to that of P_{ki} . So, proceeding as previously, one must linearize the plant equation f_{k+1} relative to the mean values of the Gaussian sum, y_{ki} . It follows that,

$$p(x_{k+1} \mid z^k) = \sum_{i=1}^{\varepsilon'_{k+1}} \alpha'_{(k+1)i} N(x_{k+1} - a_{(k+1)i}, P_{(k+1)i}) \tag{13}$$

with,

$$a_{(k+1)i} = f_{k+1}(y_{ki})$$

$$P'_{(k+1)i} = F_{(k+1)i} P_{ki} F^T_{(k+1)i} + Q_k$$

$$\alpha'_{(k+1)i} = \alpha_{ki}$$

$$\varepsilon'_{k+1} = \varepsilon_k$$

$$F'_{(k+1)i} = \left. \frac{\partial f_{k+1}}{\partial x_k} \right|_{y_{ki}}$$

The effect of the plant nonlinearity is to modify the parameter of the Gaussian sum terms. The mean values y_{ki} become equal to $a_{(k+1)i}$ and the covariances P_{ki} are modified to $P'_{(k+1)i}$.

In the second case, when the covariance Q_k is large, then it may be necessary to change the calculation of $p(x_{k+1} \mid z^k)$ since the large plant noise will increase the covariance of each term in the Gaussian sum, creating conditions for a large overlap of the individual terms. In that situation, the linearizations are no longer valid as previous and the next measurement will cause the several terms to have nearly the same mean values.

The approach to follow is to approximate the probability density function for the plant noise, i.e., $p(w_k)$, by a Gaussian sum approximation also and, therefore, it no longer matters to impose imperatively that w_k must be normality distributed. It follows that, now, $p(w_k)$ is of the form:

$$p(w_k) = \sum_{l=1}^{q_k} \gamma_k N(w_k - \omega_{kl}, Q_{kl}) \tag{14}$$

Then, the transition kernel must be linearized relative to y_{ki} , yielding:

$$p(x_{k+1} \mid x_k) =$$

$$= \sum_{l=1}^{q_k} N(x_{k+1} - f_{k+1}(y_{ki}) - F_{(k+1)i}(x_k - y_{ki}) - \omega_{kl}, Q_{kl}) \tag{15}$$

Finally, the one-step prediction is rewritten as:

$$p(x_{k+1} \mid z^k) = \sum_{i=1}^{\varepsilon'_{k+1}} \alpha'_{(k+1)i} N(x_{k+1} - a_{(k+1)i}, P'_{(k+1)}) \tag{16}$$

with,

$$\varepsilon'_{k+1} = \varepsilon_k q_k$$

$$\alpha'_{(k+1)i} = \alpha_{ki} \gamma_{kl}$$

$$a'_{(k+1)i} = f_{k+1}(y_{kl}) + w_{kl}$$

$$P'_{(k+1)i} = F_{(k+1)i} P_{kl} F^T_{(k+1)i} + Q_{kl}$$

The growth of terms could seriously reduce the performance of the Gaussian sum filter. One can alleviate the number of terms by combining some of them together or neglecting the ones whose weight approaches zero.

With all the relations derived above, it is only necessary that the probability density function prescribed for the initial state be also represented by a Gaussian sum approximation as,

$$p(x_0 \mid z^{-1}) = \sum_{i=1}^{\varepsilon'_0} \alpha'_0 N(x_0 - a'_{0i}, P'_{0i}) \tag{17}$$

then, ones uses expressions (12) and (13) or (14) to recursively compute the filtering and one-step prediction estimates.

Results

To test the Gaussian sum filter just derived, one considers a stochastic dynamical system whose plant and measurement equations are not linear functions of the state-variable as follows:

$$\begin{cases} x_{k+1} = 0{,}5x_k + 1 + \sin(0{,}04\pi x_k) + w_k \\ z_k = x_k^2 + v_k \end{cases}$$

For the first set of experiments, one sets the plant and measurement noise distributions as $w_k \sim N(0;10^{-3})$ and $v_k \sim N(0;10^{-3})$.

For the initial state, one chooses a mixture of five Gaussians whose statistics and weighting factors have been established as:

$$p(x_0 \mid z^{-1}) = 0,2N(1,2;0,02) + 0,3N(0,7;0,01) +$$

$$+0,1N(3,4;0,02) + 0,1N(4,37;0,04) + 0,3N(1,9;0,01)$$

The results obtained are plotted in *fig. 1*. From the curves, one can see that the mean estimate is predicting well the evolution of the system dynamics whereas the estimated covariances are bounded and they present small magnitude.

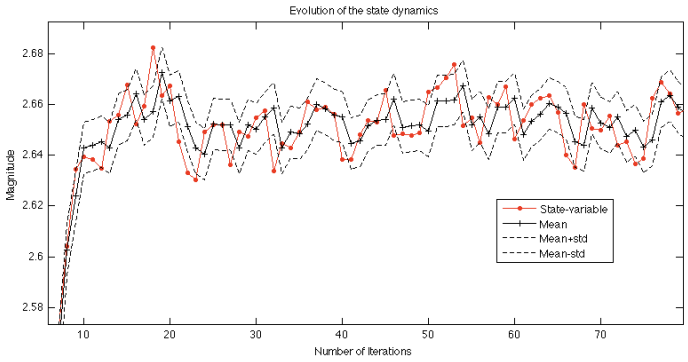

Figure 1 - Evolution of the state-variable and estimates along the iterations of the GSF algorithm.

The relative low level of uncertainty found in this experiment can be attributed to the good performance of the GSF, although the nonlinearities in the plant equation are not too severe and the magnitude of the disturbances are kept at minimum levels.

To see the evolution of the individual means belonging to each term in the Gaussian sum, one plots each tracing in *fig. 2*.

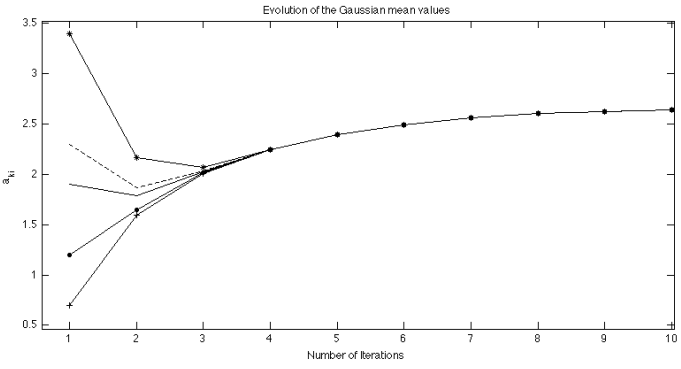

Figure 2 - Evolution of the five means belonging to each term in the Gaussian sum along the iterations.

One sees that the individual means convergence to the same value after a small number of iterations. This also happens every time one changes the initial individual means of the terms and the initial individual weights. Such phenomena suggests that, for the particular dynamic system considered, the initial conditions are not so important in the overall performance of the GSF algorithm and one can start with whatever values for initial means and weights one wishes, as long as one keeps the initial covariances small.

For the second set of experiments, one approximates the plant disturbances as a sum of Gaussian terms as well. Besides this approximation, one also degrades the disturbances' initial statistics making them harden. So now, one deals with a plant noise distribution of the form:

$$p(w_k) = 0,29N(2,14;0,72) + 0,18N(7,45;8,05) +$$

$$+0,53N(4,31;2,29)$$

As previously mentioned, the new conditions will lead to the use of expressions (12) and (14), instead of (13), to recursively compute the filtering and one-step prediction estimates, respectively.

The results obtained are plotted in *fig. 3*. By analysing the tracings, one sees that the variations in amplitude of the state-variable has increased as a consequence of the high plant noise but, the mean estimate continues to predict the evolution of the state-variable very closely. The estimated covariances have also increased in magnitude but, they are still bounded.

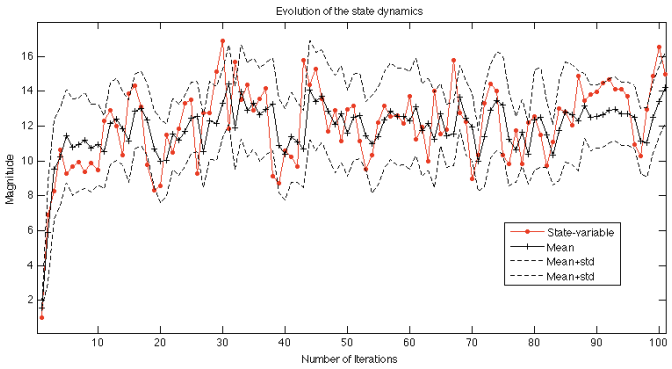

Figure 3 - Evolution of the state-variable and estimates along the iterations for the new GSF algorithm.

As a consequence of the new conditions for the second experiment, the number of terms in the Gaussian sum has increased along the iterations. This situation is depicted in *fig. 4*.

In the initial iterations, one sees a exponential growth in the number of the terms as expected, followed by a stabilisation around 35. This happens so because one has established a limitation criterion inside the GSF algorithm. If the weight of a term in the Gaussian sum is below a predefined value (for instance less than 10^{-2}), that term is automatically eliminated from the sum in the following iteration. If it wasn't for this

criterion, the number of terms would continue to grow unbounded along the iterations, leading to an overload of the computation requirements.

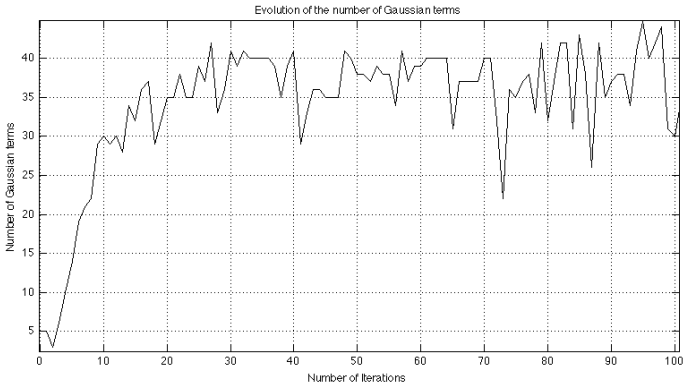

Figure 4 - Evolution of the number of terms present in the Gaussian sum along the iterations.

Within this approach, one can control the magnitude of the estimates by changing the value of the limitation criterion. If the criterion is too narrow one gets the EKF algorithm.

For this second set of experiments, a number of around 35 terms in the Gaussian sum yielded the best results.

Conclusions

A recursive procedure that performs the Gaussian sum filter of $p(x_k \mid z^k)$ and $p(x_{k+1} \mid z^k)$ has been presented. As seen throughout this report, the procedure results in the parallel operation of several Extended Kalman filters: there are as many individual filters as there are terms in the Gaussian sum. So, the computational overhead of the GSF can be significantly greater than those of the EKF.

The nonlinear GSF requires as much as $\varepsilon_k^{'}$ nodes on the initial values of the mean values and covariances. In the case that the number of nodes is equal to one, this algorithm reduces to the EKF. Nonetheless, the filtering estimates produced by the GSF are much less biased than those of the EKF.

In the literature, there are some theoretical bounds on the errors for the approximation made in the calculation of the *a posteriori* density function. However, these bounds are hard to apply to a specific problem. In opposition to the MMAE algorithm, in the GSF one needs to rely on some *ad hoc* rules to track the behaviour of the measurement residual and to keep it consistent with its theoretical properties.

When divergence or inconsistency is detected, one proceeds by reinitialising the filter parameters either by decreasing the magnitude of the covariance or adding extra terms to the sum.

References

[1] H. Tanizaki, *Nonlinear Filters - Estimation and Applications*, Springer Verlag, 1996.

[2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[3] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall Inc., 1979.

[4] H. W. Sorenson and D. L. Alspach, *Recursive Bayesian Estimation Using Gaussian Sums*, Automatica, Vol. 7, July 1971, pp. 465-479.

[5] D. L. Alspach and H. W. Sorenson, *Nonlinear Bayesian Estimation Using Gaussian Sum Approximations*, IEEE Trans. Auto. Control, Vol. AC-17, August 1972, pp. 439-448.

[6] T. Schön, *On Computational Methods for Nonlinear Estimation*, Linköping Studies in Science and Technology, Thesis No. 1047, 2003.

[7] N. Bergman, *Recursive Bayesian Estimation - Navigation and Tracking Applications*, Linköping Studies in Science and Technology, Dissertations No. 579, 1999.