An Augmented Gaussian Sum Filter Through a Mixture Decomposition

Kostas Tsampourakis [1] Victor Elvira [1,2]

[1]School of Mathematics, University of Edinburgh, United Kingdom [2]The Alan Turing Institute, United Kingdom

Context

State-Space Models and Bayesian Filtering

We consider the nonlinear, additive-Gaussian state-space model:

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}) + \mathbf{q}_t, \qquad \text{where} \qquad \mathbf{x}_t \in \mathbb{R}^{d_x}, \mathbf{y}_t \in \mathbb{R}^{d_y}$$
$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t) + \mathbf{r}_t, \qquad\qquad \mathbf{q}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \mathbf{r}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

The Bayesian filtering (BF) equations, which recursively compute the state posteriors $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ and marginal predictive likelihoods $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$ are:

prediction step:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1},$$

update step:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t}.$$

Extended Kalman filter and Gaussian sum filters

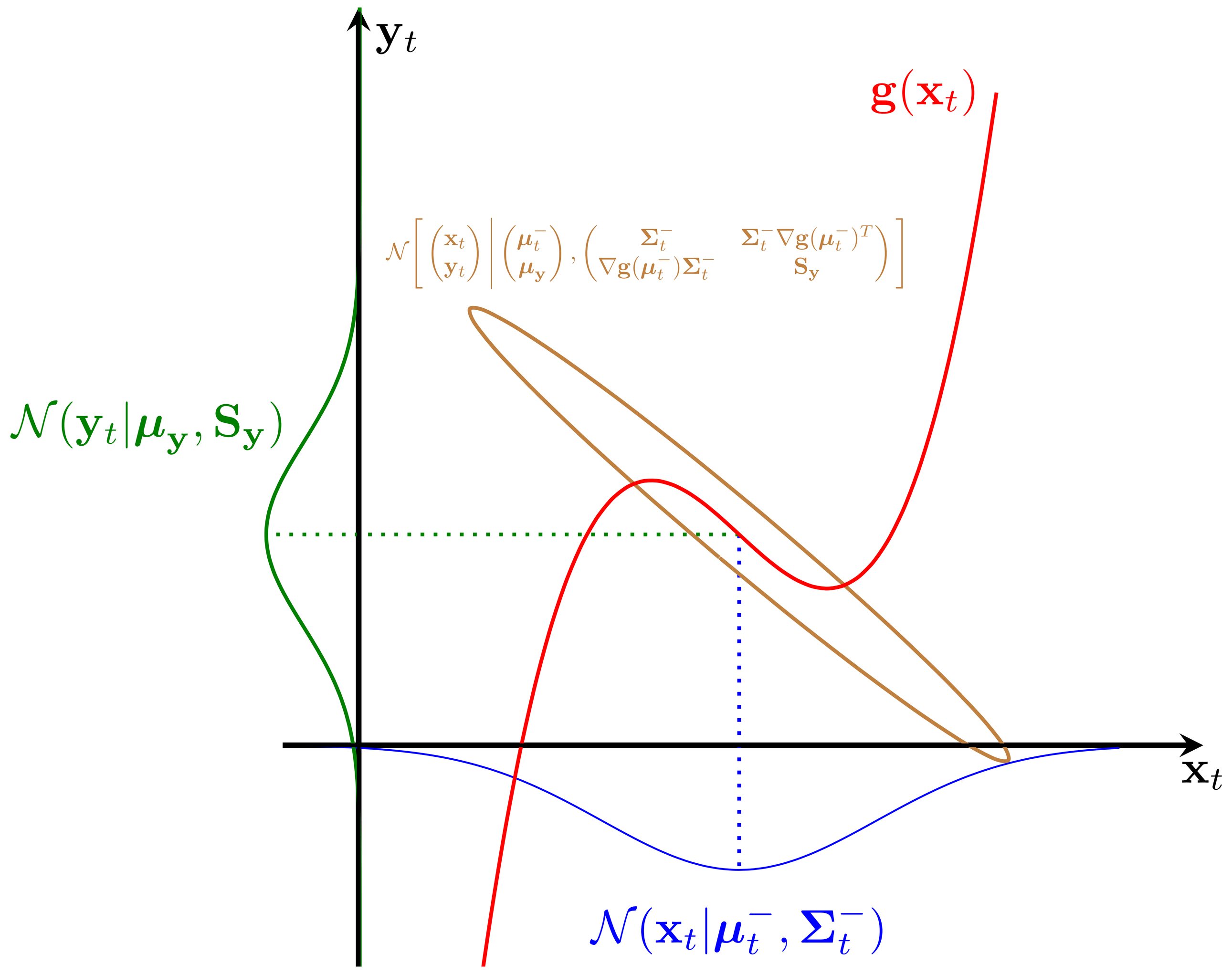
- The Extended Kalman filter (EKF) approximates the solution of the BF equations via *recursive local linearization* and *Gaussian approximation*.
- It makes a Gaussian approximation the joints of $(\mathbf{x}_{t-1}, \mathbf{x}_t)$ and $(\mathbf{x}_t, \mathbf{y}_t)$, for prediction and update. It belongs in the family of *Gaussian filters*.
- Gaussian sum filters (GSF) approximate the filtering distribution by a Gaussian mixture. Each component is given by an EKF. It is a bank of weighted EKFs run in parallel, weighted by the likelihood.

EKF prediction step:

$$\boldsymbol{\mu}_t^- = \mathbf{f}(\boldsymbol{\mu}_{t-1}),$$
$$\boldsymbol{\Sigma}_t^- = \nabla\mathbf{f}(\boldsymbol{\mu}_{t-1})\boldsymbol{\Sigma}_{t-1}\nabla\mathbf{f}(\boldsymbol{\mu}_{t-1})^T + \mathbf{Q},$$

EKF update step:

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_t^- + \mathbf{K}_t(\mathbf{y}_t - \boldsymbol{\mu}_\mathbf{y}),$$
$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^- - \mathbf{K}_t\mathbf{S}_\mathbf{y}\mathbf{K}_t^T,$$
$$\boldsymbol{\mu}_\mathbf{y} = \mathbf{g}(\boldsymbol{\mu}_t^-),$$
$$\mathbf{S}_\mathbf{y} = \nabla\mathbf{g}(\boldsymbol{\mu}_t^-)\boldsymbol{\Sigma}_t^-\nabla\mathbf{g}(\boldsymbol{\mu}_t^-)^T + \mathbf{R},$$
$$\mathbf{K}_t = \boldsymbol{\Sigma}_t^-\nabla\mathbf{g}(\boldsymbol{\mu}_t^-)^T\mathbf{S}_\mathbf{y}^{-1},$$



- Accuracy of EKF/GSF depends on width of Gaussian components.
- Agnostic to the characteristics of the nonlinearities.
- No control over the evolution of Gaussian component covariances can result in unstable behaviour s.a. covariance inflation (dynamic uncertainty increases faster than, observations can reduce it)

Augmented Gaussian Sum Filter

Augmentation of a Gaussian random variable

We construct a KDE of a Gaussian density by leveraging the identity,

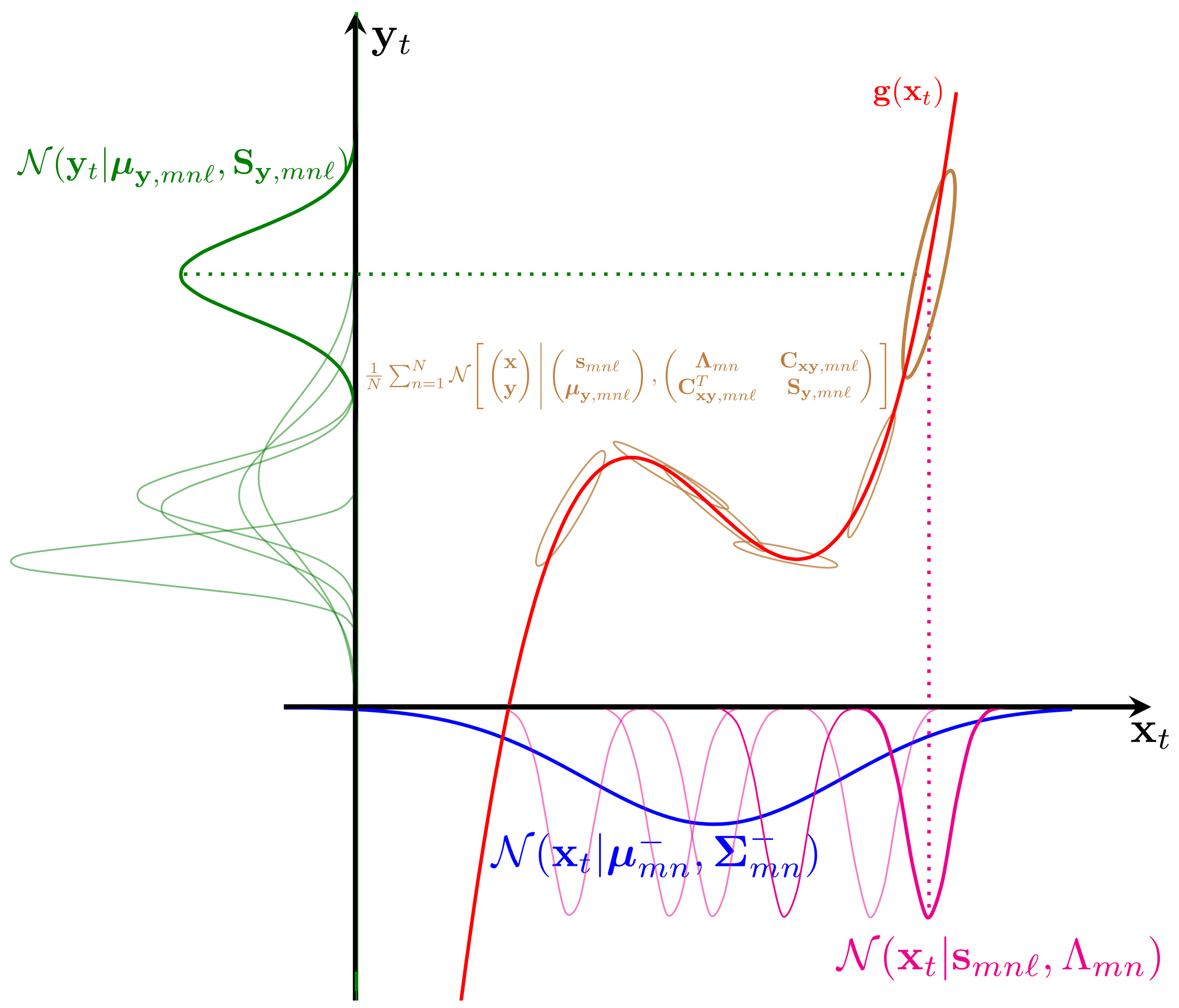
$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \mathcal{N}(\mathbf{x}|\boldsymbol{\Gamma}\mathbf{s} + \mathbf{c}, \boldsymbol{\Delta})\mathcal{N}(\mathbf{s}|\boldsymbol{\mu}_\mathbf{s}, \boldsymbol{\Sigma}_\mathbf{s})d\mathbf{s},$$

where $\mathbf{s} \in \mathbb{R}^{d_s}$ is an auxiliary random variable. For the identity to hold, we must have $\boldsymbol{\mu} = \boldsymbol{\Gamma}\boldsymbol{\mu}_\mathbf{s} + \mathbf{c}$, $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Sigma}_\mathbf{s}\boldsymbol{\Gamma}^T + \boldsymbol{\Delta}$ and also $\boldsymbol{\Delta} \succeq \mathbf{0}, \boldsymbol{\Sigma}_\mathbf{s} \succeq \mathbf{0}$. We use the above identity to make the Monte-Carlo based, Gaussian mixture approximation, by sampling the values of $\mathbf{s}$,

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \simeq \frac{1}{N}\sum_{n=1}^{N}\mathcal{N}(\mathbf{x}|\mathbf{s}_n, \boldsymbol{\Delta}).$$

Selection of $\boldsymbol{\Delta}$:

- The augmentation allows us to select the 'kernel width' $\boldsymbol{\Delta}$.
- We can use $\boldsymbol{\Delta}$ to control the linearization error of the EKF.
- Limiting case, $\boldsymbol{\Delta} = \mathbf{0}$ corresponds to Monte-Carlo sampling
- $\boldsymbol{\Delta} = \boldsymbol{\Sigma}$ corresponds to simple linearization (EKF).



AGSF algorithm

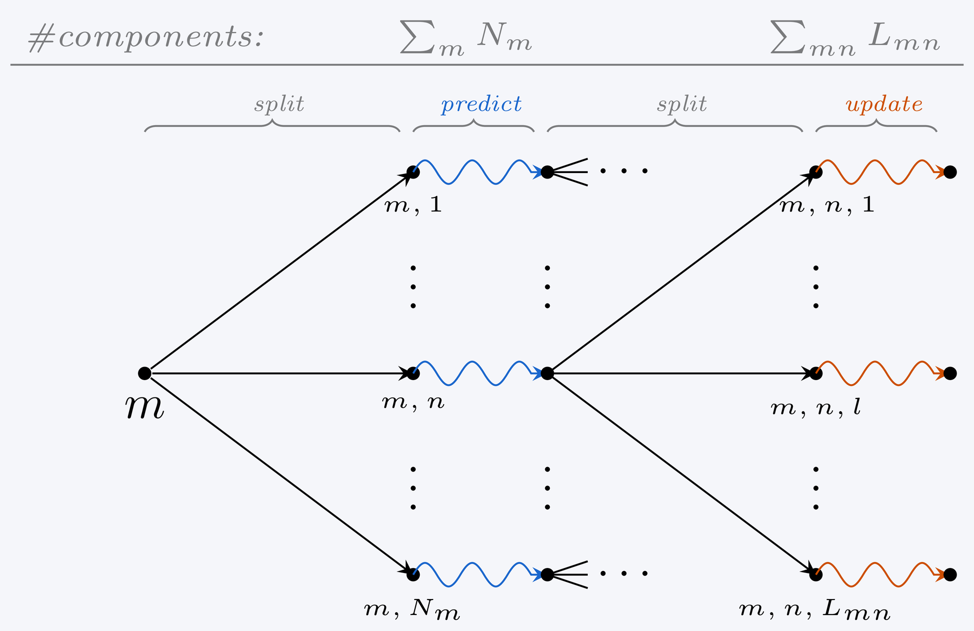


Figure 1. Splitting, prediction and update steps for a component of the AGSF.

$$\widehat{p}(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) = \sum_{m=1}^{M} w_{t-1}^{(m)}\mathcal{N}(\mathbf{x}_t|\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1}),$$

AGSF prediction step:

$$\widehat{p}(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \sum_{m=1}^{M}\sum_{n=1}^{N_m} w_{mn}\mathcal{N}(\mathbf{x}_t|\boldsymbol{\mu}_{mn}^-, \boldsymbol{\Sigma}_{mn}^-),$$

where

$$\boldsymbol{\mu}_{mn}^- = \mathbf{f}(\mathbf{z}_{mn}),$$
$$\boldsymbol{\Sigma}_{mn}^- = \nabla\mathbf{f}(\mathbf{z}_{mn})\boldsymbol{\Delta}_m\nabla\mathbf{f}(\mathbf{z}_{mn})^T + \mathbf{Q},$$
$$\mathbf{z}_{mn} \sim \mathcal{N}(\boldsymbol{\mu}_{t-1}^{(m)}, \boldsymbol{\Sigma}_{t-1}^{(m)} - \boldsymbol{\Delta}_m).$$

AGSF update step:

$$\widetilde{p}(\mathbf{x}_t|\mathbf{y}_{1:t}) = \sum_{m=1}^{M}\sum_{n=1}^{N_m}\sum_{\ell=1}^{L_{mn}} w_{mnl}\mathcal{N}(\mathbf{x}_t|\boldsymbol{\mu}_{mn\ell}, \boldsymbol{\Sigma}_{mn\ell}), \quad (3)$$

where

$$\mathbf{s}_{mn\ell} \sim \mathcal{N}(\mathbf{s}|\boldsymbol{\mu}_{mn}^-, \boldsymbol{\Sigma}_{mn}^- - \boldsymbol{\Lambda}_{mn}),$$
$$w_{mnl} \propto (w_{mn}/L_{mn})\mathcal{N}(\mathbf{y}_t|\boldsymbol{\mu}_{\mathbf{y},mn\ell}, \mathbf{S}_{\mathbf{y},mn\ell}),$$
$$\boldsymbol{\mu}_{mn\ell} = \mathbf{s}_{mn\ell} + \mathbf{G}_{mn\ell}(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y},mn\ell}),$$
$$\boldsymbol{\Sigma}_{mn\ell} = \boldsymbol{\Lambda}_{mn} - \mathbf{G}_{mn\ell}\mathbf{S}_{\mathbf{y},mn\ell}\mathbf{G}_{mn\ell}^T,$$
$$\boldsymbol{\mu}_{\mathbf{y},mn\ell} = \mathbf{g}(\mathbf{s}_{mn\ell}),$$
$$\mathbf{S}_{\mathbf{y},mn\ell} = \nabla\mathbf{g}(\mathbf{s}_{mn\ell})\boldsymbol{\Lambda}_{mn}\nabla\mathbf{g}(\mathbf{s}_{mn\ell})^T + \mathbf{R},$$
$$\mathbf{G}_{mn\ell} = \boldsymbol{\Lambda}_{mn}\nabla\mathbf{g}(\mathbf{s}_{mn\ell})^T\mathbf{S}_{\mathbf{y},mn\ell}^{-1},$$

Resampling: For $m' = 1, \ldots, M$, sample a triplet $(mn\ell)$ with probability $w_{mn\ell}$ and set $w_t^{(m)} = 1/M$

Numerical Example

We consider the one-dimensional, nonlinear state space model:

$$x_t = \sin(10x_{t-1}) + q_t$$
$$y_t = ax_t^2 + r_t$$

Table 1. *RMSE , RT ± one standard deviation. Lower is better.*

	$a = 0.01$		$a = 0.1$		$a = 1$	
	RMSE	time(s)	RMSE	time(s)	RMSE	time(s)
EKF	154.24±61.27	0.40±0.03	15.77±6.55	0.40±0.05	4.30±1.59	0.48±0.12
GSF	150.43±74.67	1.06±0.05	17.57±13.60	1.06±0.08	3.66±3.19	1.23±0.24
AGSF	**0.35±0.02**	3.33±0.25	**0.38±0.03**	3.32±0.26	**0.52±0.06**	3.88±0.80

Conclusions

- We present a novel Gaussian filter, based on a Gaussian identity.
- The AGSF allows to control the linearization error by tuning a covariance parameter $\boldsymbol{\Delta}$.
- We have showed in a numerical example, how $\boldsymbol{\Delta}$ can be used to mitigate the numerical instability of the GSF.
- The AGSF has as limiting cases the GSF and the BPF for different settings $\boldsymbol{\Delta}$. By having $\boldsymbol{\Delta}$ change adaptively, the AGSF can behave more like a PF or a GSF according to the degree of nonlinearity.

kostas.tsampourakis@ed.ac.uk