UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386

Generalized ensemble transform Kalman filters for outlier robust data assimilation

Researcher: Hans Reimann
Contact: reimann@math.uni-heidelberg.de

SFB 1294 Data Assimilation

Universität Potsdam

Summary: TL;DR

Recent advances in **generalized Bayesian inference** and corresponding generalized posteriors offer curious opportunities **for Bayesian data assimilation**. First advances resulted in deriving varieties of the celebrated Kalman filter with novel, desirable properties while maintaining conjugacy in the analysis step and stability under mild assumptions. This encourages investigating generalizations to more sophisticated and state-of-the-art approaches in data assimilation, e.g. akin to the popular local ensemble transform Kalman filter (LETKF). We present first results of one such adaptation of the LETKF utilizing diffusion score matching (DSM) to capture discrepancy between measures of observation distributions with an appropriate choice of diffusion matrix to assimilate observations online under suspicion of **mis-specification of the observation likelihood**. Additionally, we showcase consistency of the approach with regard to assumption expanding the KF to the LETKF for the DSM KF in relation to the **DSM LETKF**. The resulting algorithm maintains the desirable properties of the LETKF while obtaining robustness to outliers at a quantified cost in increased uncertainty. Moreover, the resulting formula suggests an insightful interpretations of novel dynamics of this filter.

Model Assumptions: LGSS

X_n **cannot be observed directly**, but $Y_n = g_n(X_n, V_n)$ available. Assume the following linear, time discrete, time varying signal evolution equation and linear observation equation:

(S) $X_n = A_n X_{n-1} + C_n W_n$
(O) $Y_n = H_n X_n + \Gamma_n V_n$

- X_n in $\mathbb{R}^d$, Y_n in $\mathbb{R}^p$,
- W_n and V_n - independent standard Gaussian RVs,
- non-singular $Q_n = C_n C_n^T$ and $R_n = \Gamma_n \Gamma_n^T$, and
- $p(x_0) \sim n(x_0; m_0, P_0)$.

Ex.: Partially observed target tracking in two dimensions.

Problem: Outlier Volatility

When does Bayes fail?

Bayesian learning can be derived via **information processing** (see [9]) to obtain

$$p(x_n|y_{1:n}) = p(x_n|y_{1:(n-1)}) \times \exp\left(-\hat{\mathrm{KL}}\left[p(y_n)\|p(y_n|x_n)\right]\right).$$

The KL divergence assigns large weights to the tail behavior of distributions via its probability ratio (see [5]). The resulting issue is best capture by the quote in [6]:
"With a finite amount of samples, that translates into saying that Bayes' posterior is highly sensitive to outliers in the data."

For **Bayesian filtering**, this directly results in high sensitivity of the analysis distribution regarding outliers arising with tail mis-specification in the observation model.

Generalizing Bayes with DSM

In the mis-specified setting, approximate Bayesian methods may yield better performance than exact methods via robustness to outliers.

Generalized Bayesian learning as in[3] introduces a degree of freedom in choice of divergence measure via

$$p_{\mathrm{D}}(x_n|y_{1:n}) \propto p(x_n|y_{1:(n-1)}) \times \exp\left(-\hat{\mathrm{D}}\left[p(y_n)\|p(y_n|x_n)\right]\right).$$

For different contexts, work in [1, 2] successfully employed **diffusion score matching** as an estimator for minimum diffusion Fisher divergence to obtain **robust generalized posteriors** with properties in conjugacy.

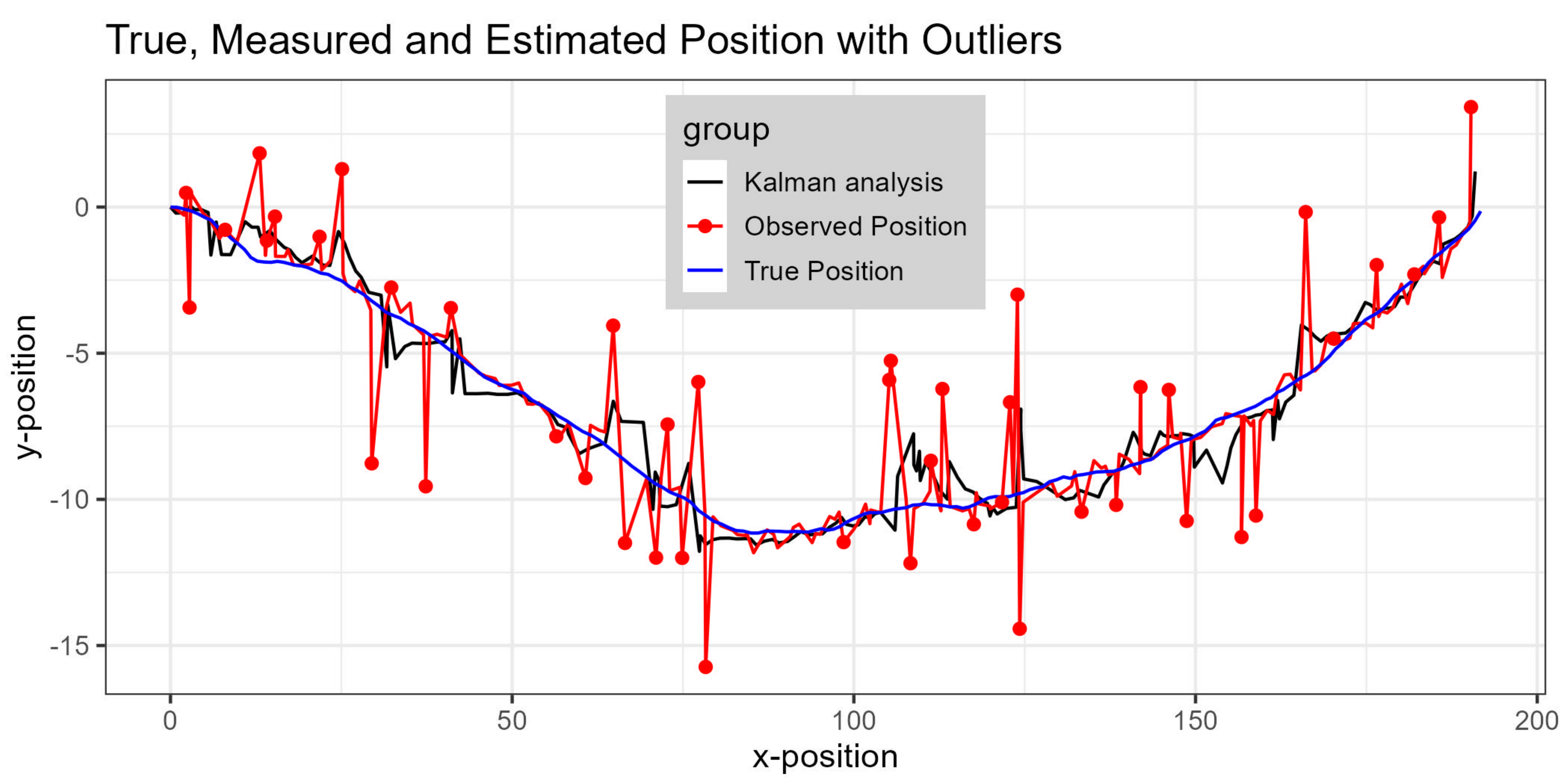

Figure 1: Simulation experiment of 2D target tracking with unobserved velocity, contaminated observations and estimated position via the mean of the **regular posterior** of the Kalman filter.

Solution: The DSM Kalman Filter

Proposition 1 ***DSM Analysis Step***
The analysis step for the Kalman filter variant is given via
$p_{\mathrm{D}}(x_n|y_{1:n}) \sim n(x; m_n^a, P_n^a)$ *with*

$$N_n(y_n)^{-1} = 2k_n^2(y_n)R_n^{-1}$$
$$\tilde{y} = y_n - N_n(y_n)\nabla_{y_n}2k_n^2(y_n)$$
$$\tilde{K}_n(y_n) = P_n^f H_n^T[N_n(y_n) + H_n P_n^f H_n^T]^{-1}$$
$$P_n^a = P_n^f - \tilde{K}_n(y_n)H_n P_n^f$$
$$m_n^a = m_n^f - \tilde{K}_n(y_n)[H_n m_n^f - \tilde{y}_n].$$

Algorithm 1 ***The DSM Kalman Filter***
1. *Forecast Step:*
   - *see regular Kalman filter*
2. *Analysis Step:*
   - *update according to Prop. 1*

Proposition 2 ***Global Bias Robustness***
$p_{\mathrm{D}}(x_n|y_{1:n})$ *is globally bias robust for weight function* $k: \mathcal{Y} \to \mathbb{R}$ *is chosen such that*

$$k_n(y_n) = \left(1 + \frac{\|y_n - H_n m_n^f\|^2_{(H_n P_n^f H_n^T + R_n)^{-1}}}{q^2}\right)^{-\frac{1}{2}}$$

with tuning parameter $q > 0$.

Theorem 1 ***Asymptotic Stability***
Assume finite second moment for the true DGP. Then $N_n(Y_n)$ *has a weak stochastic bound and the resulting DSM Kalman filter is asymptotically stable in that its covariance matrix has a finite steady state.*

The *proof* follows via a Markov inequality utilizing results on stability for Kalman filters with weak stochastically bound components [8].

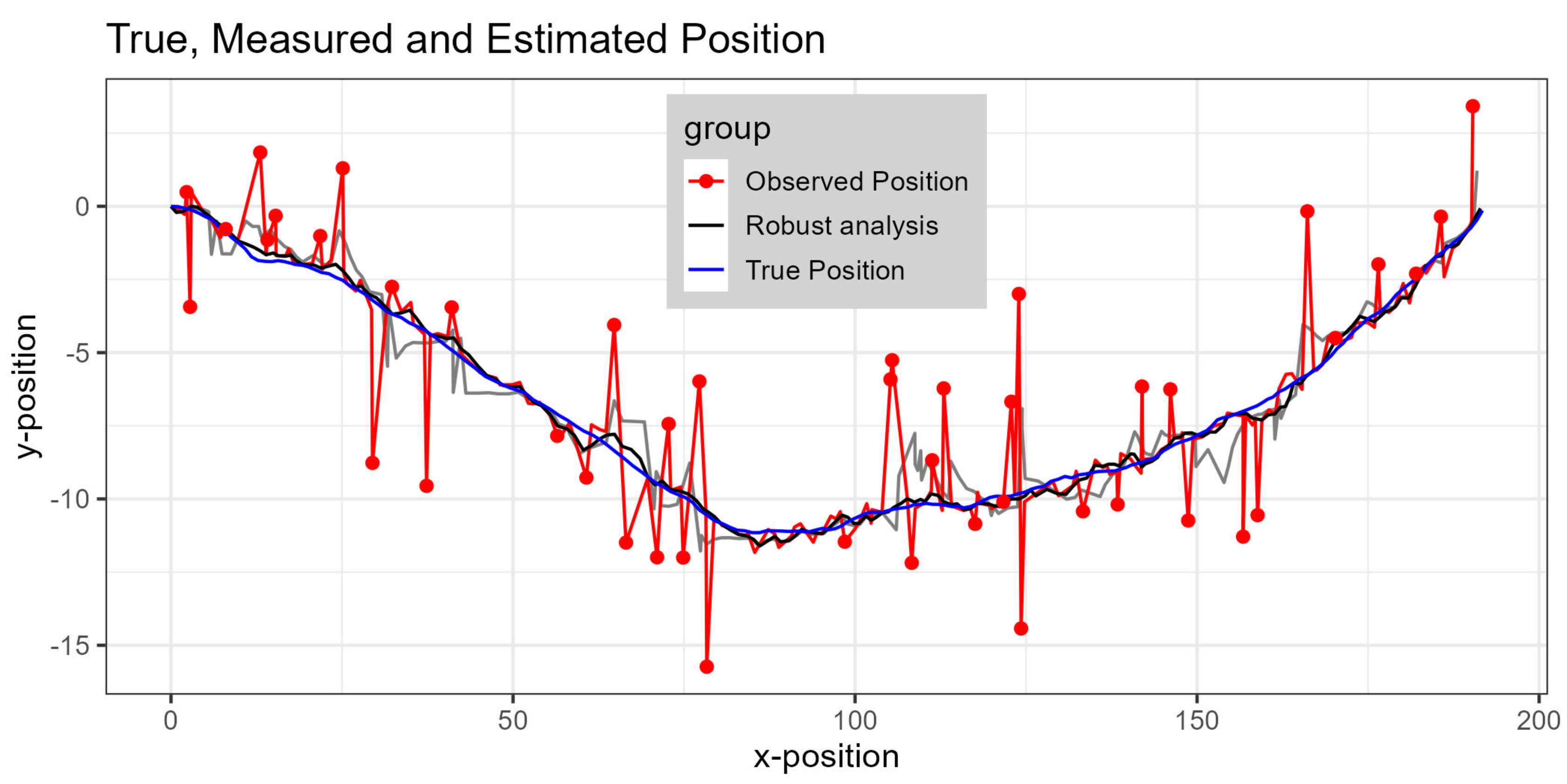

Figure 2: Simulation experiment of 2D target tracking with unobserved velocity, contaminated observations and estimated position via the mean of the **diffusion score matching posterior**.

Non-Linearity: A Family of Ensemble Approximations

While theory is mainly restricted to the linear, Gaussian state space model, we can adapt algorithms to account for non-linearity via usual approaches in ensemble approximation. Following [7], we want to formulate varieties of the DSM Kalman filter via **stochastic** (EnKF) and **deterministic** (ESRF) coupling of forecast and analysis ensemble as well as **local transformation** (LETKF) akin to [4] for non-linear observation operators. We omit basic steps introduced in Alg. 1.

Given a forecast ensemble $\{x_n^{f,(i)}\}_{i=1}^M$ with M ensemble members. The aim is to obtain an analysis ensemble $\{x_n^{a,(i)}\}_{i=1}^M$ conserving first and second moment.

Hereby, assume a linear observation operator H_n for Alg. 2 and 3 and a general observation operator $h_n(\cdot)$ for Alg. 4.

Algorithm 2 ***DSM Ensemble KF with Perturbed Obs.***
1. *Compute the emp. forecast moments* $\bar{x}_n^f$ *and* $\bar{P}_n^f$,
2. *compute the weight* $\bar{k}_n^2(y_n)$ *replacing* m_n^f *with* $\bar{x}_n^f$,
3. *compute emp. Kalman gain*

$$\bar{\tilde{K}}_n(y_n) = \bar{P}_n^f H_n^T[N_n(y_n) + H_n \bar{P}_n^f H_n^T]^{-1},$$

4. *draw the perturbed obs.* $\tilde{y}_n^{\xi,(i)} \sim \mathcal{N}(\tilde{y}_n, N_n(y_n))$,
5. *update individual ensemble members*

$$x_n^{a,(i)} = x_n^{f,(i)} - \bar{\tilde{K}}_n(y_n)[H_n x_n^{f,(i)} - \tilde{y}_n^{\xi,(i)}].$$

The given formulations results in analytic consistency.

Algorithm 3 ***DSM Ensemble Square Root Filter***
1. *Compute the emp. forecast moments* $\bar{x}_n^f$ *and* $\bar{P}_n^f$,
2. *compute the weight* $\bar{k}_n^2(y_n)$ *replacing* m_n^f *with* $\bar{x}_n^f$,
3. *compute emp. Kalman gain*

$$\bar{\tilde{K}}_n(y_n) = \bar{P}_n^f H_n^T[N_n(y_n) + H_n \bar{P}_n^f H_n^T]^{-1},$$

4. *compute* $\bar{\tilde{x}}_n^a = \bar{x}_n^f - \bar{\tilde{K}}_n(y_n)[H_n \bar{x}_n^f - \tilde{y}_n]$,
5. *compute* $\bar{\tilde{P}}_n^a = \bar{P}_n^f - \bar{\tilde{K}}_n(y_n)H_n \bar{P}_n^f$ *and*
6. *update individual ensemble members via*

$$x_n^{a,(i)} = \bar{\tilde{x}}_n^a - (\bar{\tilde{P}}_n^a)^{\frac{1}{2}}(\bar{P}_n^f)^{-\frac{1}{2}}[x_n^{f,(i)} - \bar{x}_n^f].$$

Algorithm 4 ***DSM Local Ensemble Transform KF***
1. *Compute the emp. forecast mean* $\bar{x}_n^f$, *anomaly matrix* $(X_n^f)_i = x_n^{f,(i)} - \bar{x}_n^f$ *and forecast obs.* $\bar{y}_n^f = h(\bar{x}_n^f)$,
2. *compute the obs. ensemble via* $y_n^{(i)} = h(x_n^{f,(i)})$ *and anomaly matrix* $(Y_n^f)_i = y_n^{(i)} - \bar{y}_n^f$,
3. *employ the DSM KF in obs. anomaly space*
   - $\tilde{P}_n^Y = [(M-1)\mathbf{1} + (Y_n^f)^t N_n^{-1}(y_n) Y_n^f]^{-1}$ *and*
   - $\bar{v}_n^Y = \tilde{P}_n^Y (Y_n^f)^T N_n^{-1}(y_n)[\tilde{y}_n - \bar{y}_n^f]$,
4. *compute* $\bar{\tilde{P}}_n^a = X_n^f \tilde{P}_n^Y (X_n^f)^T$, $\bar{\tilde{x}}_n^a = \bar{x}_n^f + X_n^f \bar{v}_n^Y$ *and*
5. *apply step 6 in Alg. 3 or some other scheme to update the individual ensemble members.*

Future Questions

1. Why the **correction** term in $\tilde{y}_n$?
2. How does a **continuous time** filter variant look like?
3. Can we understand the DSM KF and its proficiency via DSM as a Stein discrepancy?
4. Can we utilize the diffusion matrix as **normalising flow** to improve applicability to general partially known error?

References

[1] M. Altamirano, F.-X. Briol, and J. Knoblauch. Robust and scalable bayesian online changepoint detection. *arXiv preprint arXiv:2302.04759*, 2023.

[2] M. Altamirano, F.-X. Briol, and J. Knoblauch. Robust and conjugate gaussian process regression. *arXiv preprint arXiv:2311.00463*, 2023.

[3] P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5):1103–1130, 2016.

[4] B. R. Hunt, E. J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D: Nonlinear Phenomena*, 230(1-2):112–126, 2007.

[5] J. Jewson, J. Q. Smith, and C. Holmes. Principles of bayesian inference using general divergence criteria. *Entropy*, 20(6):442, 2018.

[6] L. Pacchiardi. Generalizing bayesian inference, Aug 2021. URL `http://www.lorenzopacchiardi.me/blog/2021/generalizedBayes/`.

[7] S. Reich and C. Cotter. *Probabilistic forecasting and Bayesian data assimilation*. Cambridge University Press, 2015.

[8] V. Solo. Stability of the kalman filter with stochastic time-varying parameters. In *Proceedings of 35th IEEE Conference on Decision and Control*, volume 1, pages 57–61. IEEE, 1996.

[9] A. Zellner. Optimal information processing and bayes's theorem. *The American Statistician*, 42(4):278–280, 1988.