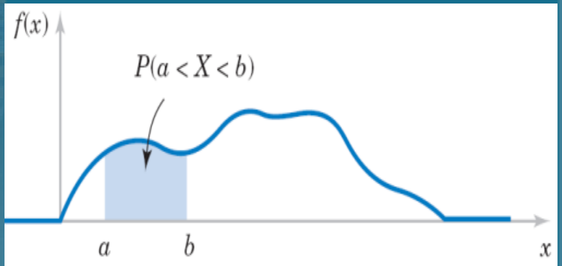
SAND2022-16166C

Sandia National Laboratories

Variational Kalman Filtering with H∞-Based Correction for Robust Bayesian Learning in High Dimensions



Niladri Das, Jed A. Duersch, & Tommie A. Catanach

December 2022

61st IEEE Conference on Decision and Control

U.S. DEPARTMENT OF ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND

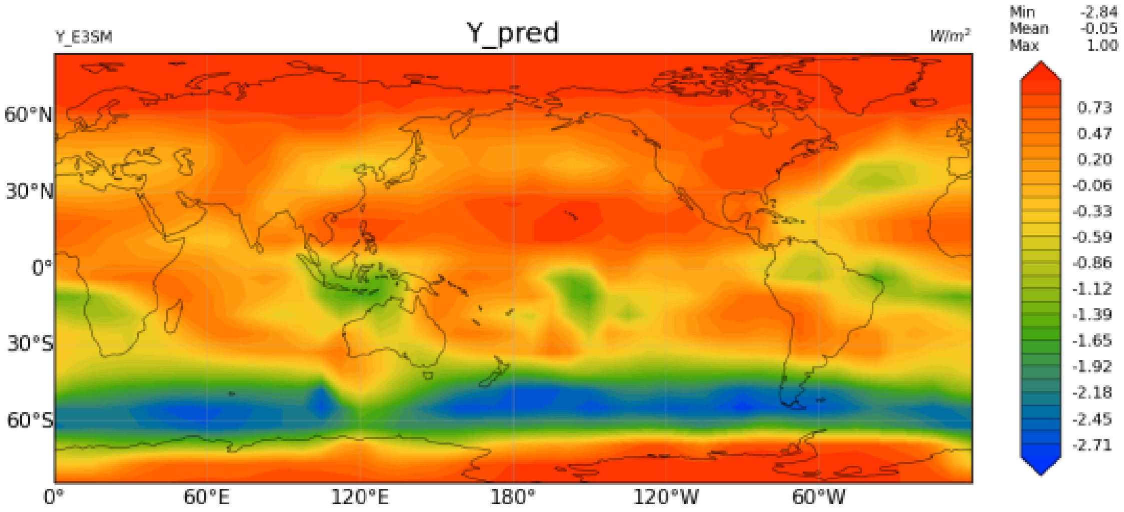
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

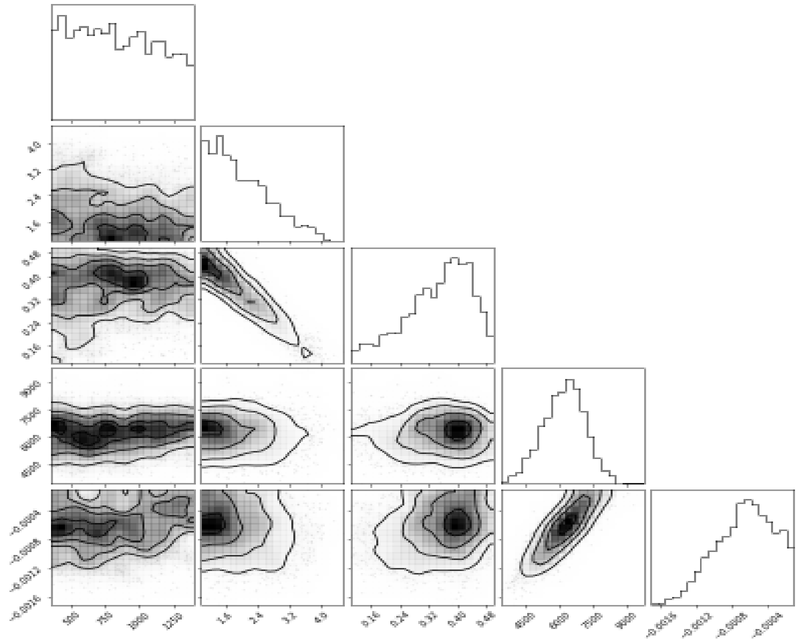
Outline

A. Introduction
B. Model and Kalman Filter
C. Variational Inference filter
D. H_∞ Filter and Augmented H_∞ Filter
E. Results

We address the problem of convergence of sequential variational inference filter using a robust variational objective and H_∞ norm-based correction for a linear Gaussian system.

Introduction





For very high dimensional problems such as:

- Big data, ML problems
- Mapping tropical Pacific sea level
- Monitoring Water Quality

modeled as a linear Gaussian system, traditional Kalman filter is intractable.

Intractable: Matrix storage and computation

Available solutions:

- Reduced rank Kalman Filter
- Direct approximations

Our solution :
Variational formulation aided by the H_∞ filter

Model and Kalman Filter

Linear system with additive Gaussian noise

$$\boldsymbol{y}_t = \boldsymbol{x}_t^T \boldsymbol{\theta}_t + \boldsymbol{\eta}_t$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{A}_t \boldsymbol{\theta}_t + \boldsymbol{w}_t$$

- Index : t
- Output : $\boldsymbol{y}_t \in \mathbb{R}^m$
- Input : $\boldsymbol{x}_t \in \mathbb{R}^n$
- $m << n$
- I.I.D. zero mean Gaussian noises : $\boldsymbol{\eta}_t, \boldsymbol{w}_t$
- Noise covariance matrices : $\boldsymbol{Q}, \boldsymbol{R}$
- Prior on theta : $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0})$

Optimal update: Kalman filter (KF)

$$\boldsymbol{\theta}_{t-1} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t-1}, \boldsymbol{P}_{t-1}\right) \to \boldsymbol{\theta}_t \sim \mathcal{N}\left(\boldsymbol{\mu}_t^{KF}, \boldsymbol{P}_t^{\mathrm{KF}}\right)$$

Intractable: $\boldsymbol{P}_t^{\mathrm{KF}}$ storage and computation

Solution: Estimate a diagonal matrix using Variational Inference

$$\boldsymbol{P}_t \text{ such that } \boldsymbol{P}_t^{KF} \preccurlyeq \boldsymbol{P}_t \preccurlyeq \boldsymbol{P}_{t-1}$$

Issue :

$$\boldsymbol{P}_t^{KF} \preccurlyeq \boldsymbol{P}_t \preccurlyeq \boldsymbol{P}_{t-1} \overset{a.s.}{\Longrightarrow} \boldsymbol{P}_t = \boldsymbol{P}_{t-1}$$

*Theorem 1

Solution :
Introduce cross-correlation information using concepts from the H_∞ filter, bringing robustness

Variational Inference →

Variational Inference

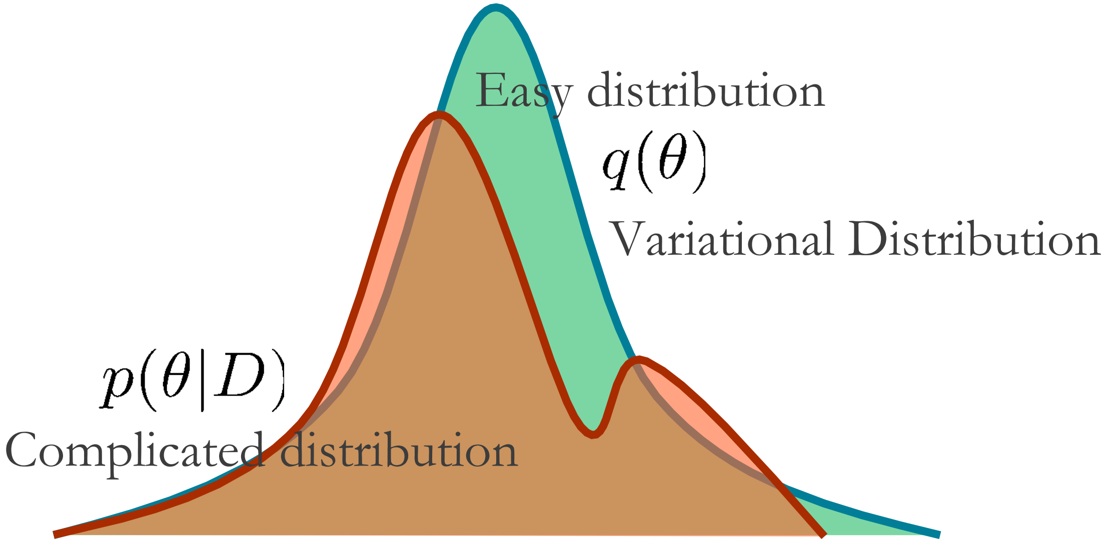
Bayes Rule

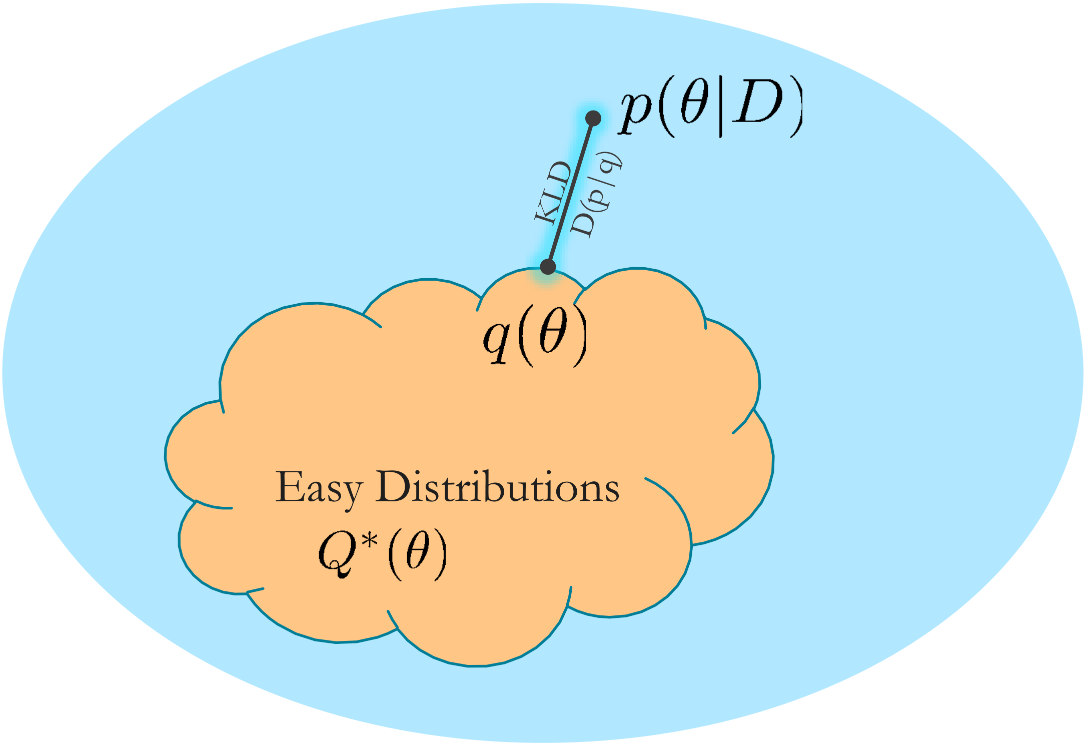
Posterior — Prior — Likelihood

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{\int_\theta p(\theta)p(D|\theta)d\theta}$$

$\mathbb{R}^{100\dots}$

At even moderately high dimensions of $\theta$the amount numerical operations explode.





Variational Inference Filter

$$D_{KL}(q_\phi(\boldsymbol{\theta})|p(\boldsymbol{\theta})) = \int q_\phi(\boldsymbol{\theta}) \log \frac{q_\phi(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta}$$

Posterior pdf: $p(\boldsymbol{\theta})$
Variational pdf: $q_\phi(\boldsymbol{\theta})$

Traditionally, solved using ELBO formulation. (ELBO-VI)

But, here we prefer solving for $D_{KL}(p(\boldsymbol{\theta})|q_\phi(\boldsymbol{\theta}))$ which is the Expected Propagation formulation. (EP-VI)

ELBO-VI underestimates uncertainty.

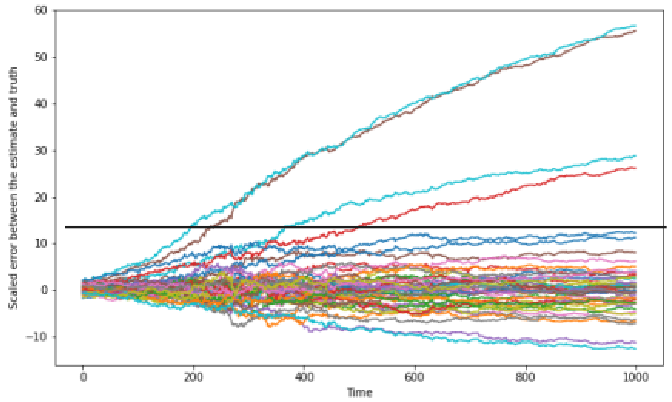
Optimal $P_{EP\text{-}VI}$: solved analytically with a diagonal VI covariance matrix

$$q(\boldsymbol{\theta} \mid \boldsymbol{\phi}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{P}(\boldsymbol{d}))$$

*Theorem 4

where $\boldsymbol{d}$ is the diagonal.

Issue: convergence problem for large dimension
Reason: cannot capture correlation with a diagonal matrix. Since this is not based on a real distance metric, there is no cancellation to hide uncertainty



Plot showing the scaled estimation error response of the EP-VI filter with a dimension n=50.

General family of f-divergence

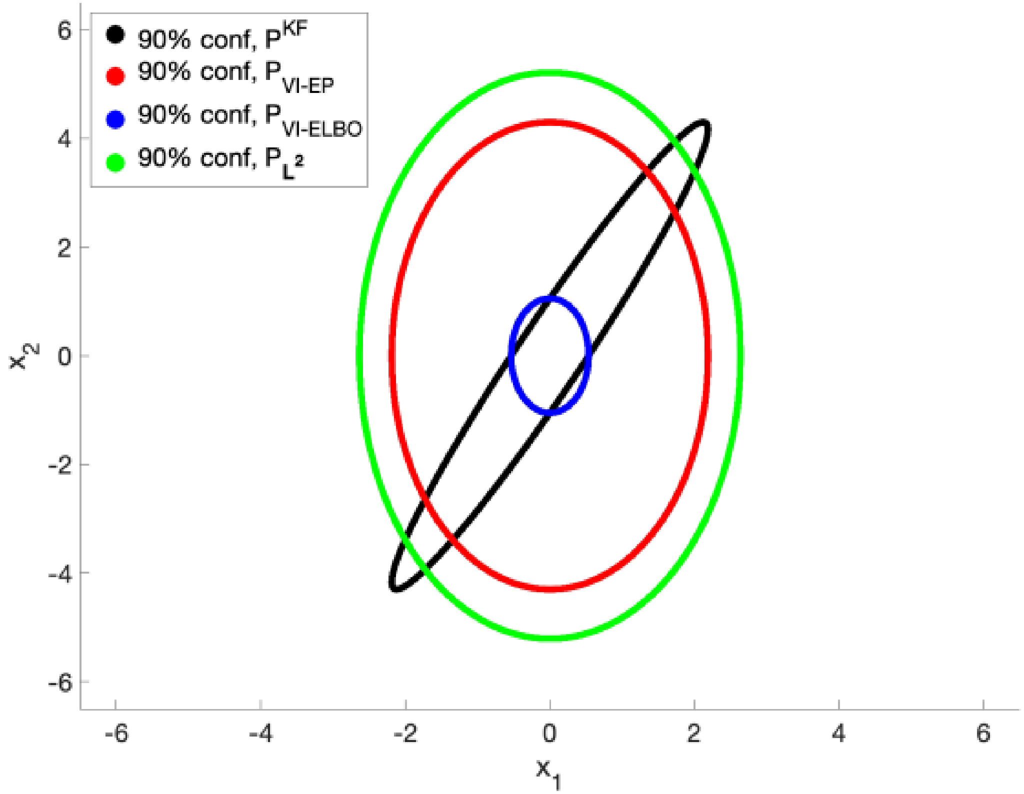
$\mathbb{L}^r$ Filter

$\mathbb{L}^r$ Information pseudometric*:

$$\mathbb{L}^r = \left(\int p(\boldsymbol{\theta}) \left| \log \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \right|^r d\boldsymbol{\theta} \right)^{1/r}; \quad r \geq 1$$

$\mathbb{L}^r$ Filter Update Step:

$$\underset{\boldsymbol{d}_i}{\operatorname{argmin}} \int p(\boldsymbol{\theta}) \left| \log \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta} \mid \boldsymbol{d})} \right|^r d\boldsymbol{\theta} = \boldsymbol{d}^* \rightarrow \boldsymbol{P}_t(\boldsymbol{d}^*)$$



Comparing the 90% confidence level of true posterior (in black), traditional VI update based on KL divergence with EP (in red) and with the ELBO (in blue). The pseudometric optimum avoids suppressing probability in domains that have significant posterior density.

Better, but cross-correlation is still not captured.

Motivation from H_∞ filter

*Duersch, Jed A., and Thomas A. Catanach. "Generalizing information to the evolution of rational belief." *Entropy* 22.1 (2020): 108

H_∞ Filter and augmentation with VI

Goal of H_∞ filter is to estimate $\boldsymbol{\theta}_t$ such that,

$$J = \frac{\sum_{t=0}^{N-1} ||\boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_t||^2_{\boldsymbol{S}_t}}{||\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_0||^2_{\boldsymbol{P}_0^{-1}} + \sum_{t=0}^{N-1} \left(||\boldsymbol{w}_t||^2_{\boldsymbol{Q}^{-1}} + ||\boldsymbol{\eta}_t||^2_{\boldsymbol{R}^{-1}} \right)}$$

can be made less than $1/\gamma$.

The estimation steps are:

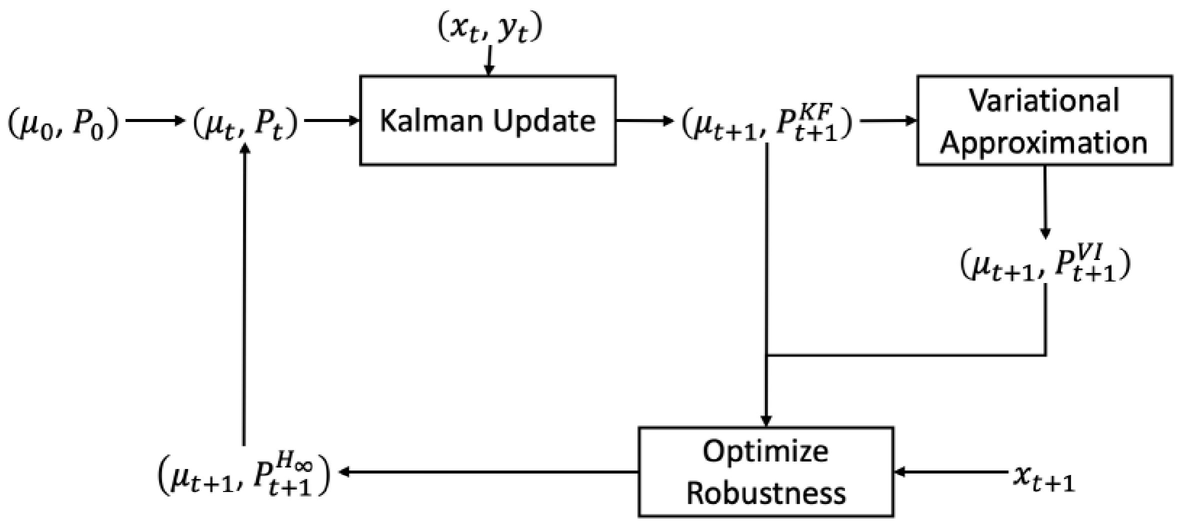
$$\boldsymbol{P}_t = \tilde{\boldsymbol{P}}_t[\boldsymbol{I} - \gamma \boldsymbol{S}_t \tilde{\boldsymbol{P}}_t + \boldsymbol{x}_t \boldsymbol{R}^{-1} \boldsymbol{x}_t^T \tilde{\boldsymbol{P}}_t]^{-1}$$

$$\boldsymbol{K}_t = \boldsymbol{P}_t \boldsymbol{x}_t \boldsymbol{R}^{-1}$$

$$\hat{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{x}_t^T \tilde{\boldsymbol{\theta}}_t)$$

$\tilde{\boldsymbol{P}}_t$: prior covariance

$\hat{\boldsymbol{\theta}}_t$: estimates



Three steps: First, the Kalman update to compute the true update. Then Variational Inference is used to find the best low memory approximation. Finally, comparing the true update and variational approximation a final update which is optimized to robustly assimilate the next observation.

Algorithm

Algorithm 1 Algorithm of Augmented H_∞ Filter Update

$P_{t-1}^{H_\infty}$: Diagonal posterior matrix at step (t-1)
$P_{t}^{H_\infty}$: Diagonal posterior matrix at step t
Calculate:

$$K_t = (A_{t-1} P_{t-1}^{H_\infty} A_{t-1}^T + Q) x_t \times \left[x_t^T (A_{t-1} P_{t-1}^{H_\infty} A_{t-1}^T + Q) x_t + R \right]^{-1}$$

$$P_t^{KF} = (I - K_t x_t^T)(A_{t-1} P_{t-1}^{H_\infty} A_{t-1}^T + Q) \quad (11)$$

Optimization 1: (using Newton Conjugate Gradient)

$$d^{\mathrm{Lr}} = \underset{d_t}{\operatorname{argmin}} \int p(\theta \mid y_t) \log \left[\frac{p(\theta \mid y)}{q(\theta \mid d_t)} \right]^r d\theta \approx P_t^{\mathrm{Lr}}(d) \quad (12)$$

Propagate $P_t^{\mathrm{Lr}}(d)$: $\tilde{P}_{t+1}^{\mathrm{Lr}} = (A_{t+1} P_t^{\mathrm{Lr}} A_{t+1}^T + Q)$

Compute the Kalman update for the next step using P_t^{KF}:

$$K_{t+1}^{KF} = (A_t P_t^{KF} A_t^T + Q) x_{t+1} \times \left[x_{t+1}^T (A_t P_t^{KF} A_t^T + Q) x_{t+1} + R \right]^{-1} \quad (13)$$

Compute the expression for $K_{t+1}^{H_\infty}$:

$$K_{t+1}^{H_\infty} = \tilde{P}_{t+1}^{\mathrm{Lr}} [I - \gamma \tilde{P}_{t+1}^{\mathrm{Lr}} + x_t R^{-1} x_t^T \tilde{P}_{t+1}^{\mathrm{Lr}}]^{-1} \times x_t R^{-1} \quad (14)$$

Optimization 2: (Sequential Least Squares Programming)

$$\gamma_t^* = \underset{\gamma}{\operatorname{argmin}} \| K_{t+1}^{H_\infty} - K_{t+1}^{KF} \|_2$$

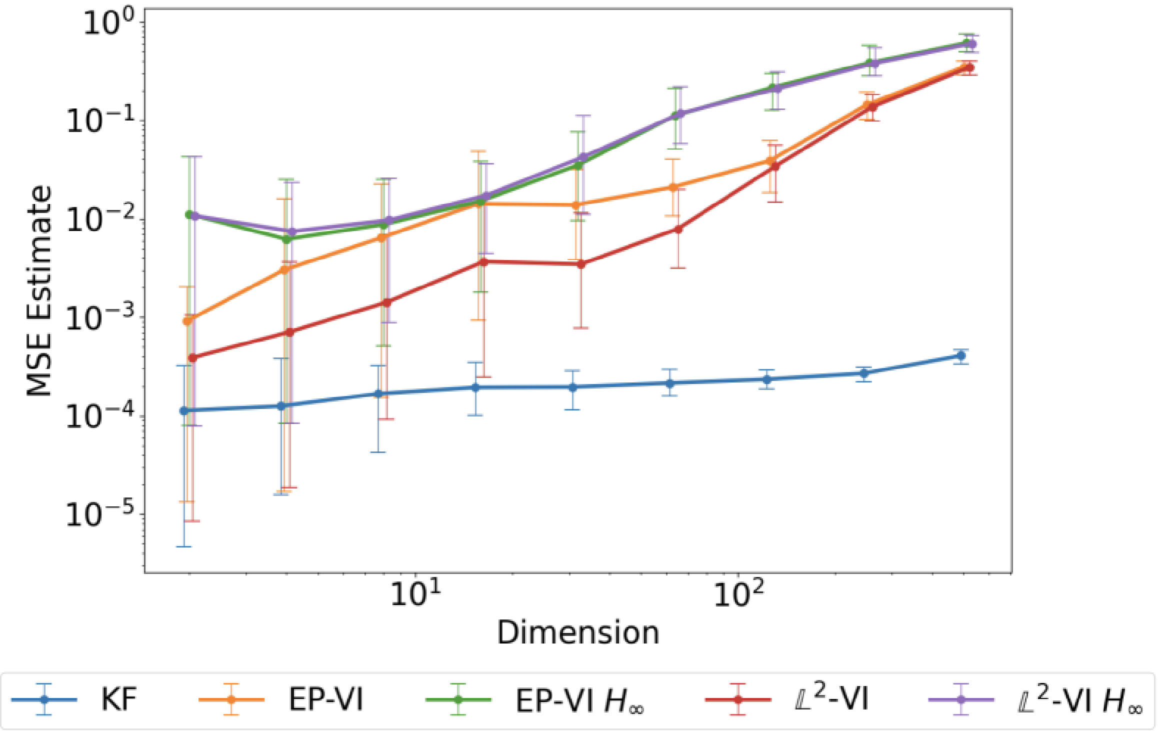
$$P_t^{H_\infty} = P_t^{\mathrm{Lr}}(d) [I - \gamma_t^* P_t^{\mathrm{Lr}}(d) + x_t R^{-1} x_t^T P_t^{\mathrm{Lr}}(d)]^{-1} \quad (15)$$

Outline of the algorithms:

1. Calculate the posterior covariance using the Kalman update.
2. Solve the EP-VI for the diagonal covariance matrix.
3. Use the H_∞ feedback to minimize the norm between the Kalman gain and EP -VI Kalman gain.

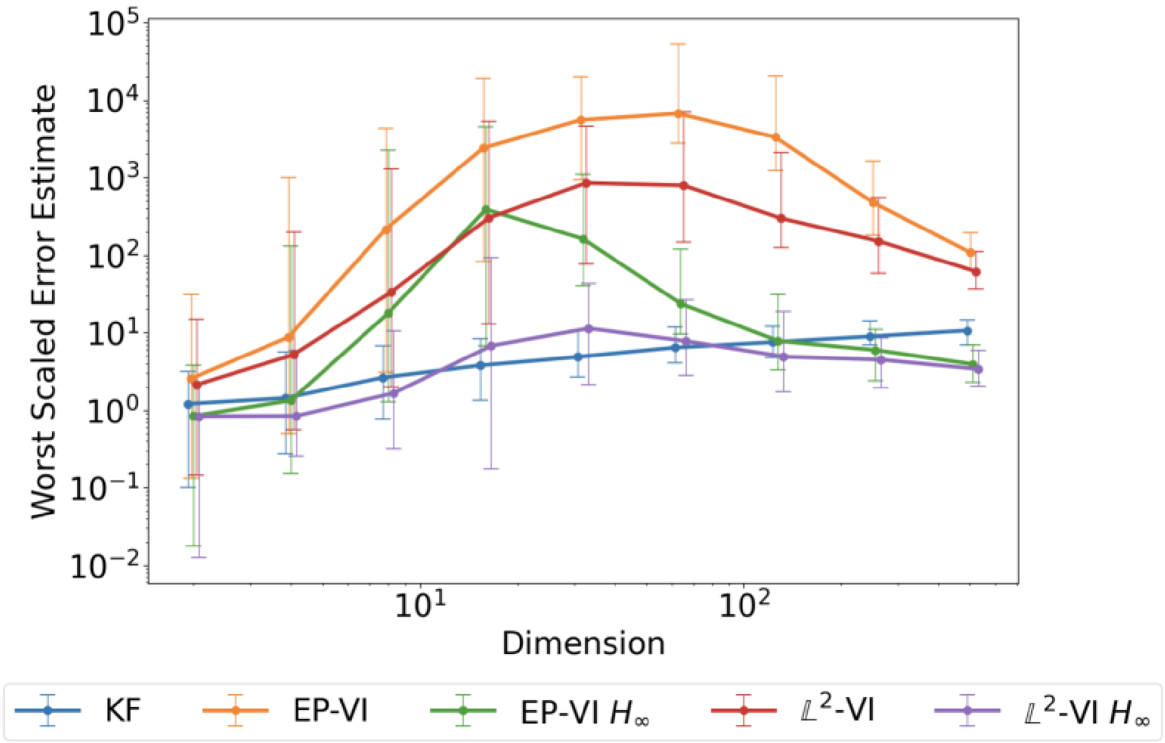
Note: The outcome is a low-rank covariance matrix, that can robustly represent the uncertainty in the estimates, without leading to divergence with growing problem size.

Result - I



Mean Square Error for the five different filtering algorithms with varying problem dimensions after 1000 sequentially observed data points. The error bars correspond to 93% confidence intervals. Two main observations are that 1) the L2 formulation of the VI problem outperforms standard EP VI and 2) the two H∞ filters under-perform, but as we will see are not overconfident so are more trustworthy. A small perturbation exists on the different filter's dimension coordinates to improve readability.

Result - II



Worst Case Scaled Error (e.g. the largest absolute error divided by estimated standard deviation) for the five filters. The error bars are 93% confidence intervals. We see that 1) the variational filters without added robustness can be significantly overconfident and 2) the two H∞ filters, particularly the L2-H∞ better capture the worst-case estimate uncertainty. These methods trade robustness and lower bias for slower convergence. A small perturbation exists on the different filter's dimension coordinates to improve readability.

Takeaways

- The standard KF becomes exceedingly memory intensive as the dimension of the underlying state space increases.
- We propose a reduced memory filter based upon variational inference with an information pseudometric and $H\infty$ filter to resolve the storage and computational issue related to the error covariance matrix while retaining robustness.
- Our filters, specifically the L2-$H\infty$ exhibits slower learning, but enable information update along all the directions of the state space, keeping the worst-case performance better than other filters.
- We are looking forward to improving the performance of our proposed algorithm

Thank You