Data Assimilation lightweight package

Jose Miranda

Florida State University

Department of Scientific Computing

Final project - ISC 4933/5935 – Spring 2024

Overview

- Introduction
- Methods
  - BLUE, KF, EKF, EnKF, 3DVar and 4DVar
- Experiments
- Next steps and challenges

Introduction

- Description: a python library with some DA methods
  - BLUE, KF, EKF, EnKF, 3DVar, 4DVar
- Motivations:
  - Educational/refresher/reference material
  - Reusability
  - Open source - https://github.com/0jrm/DA_101

Methods

Best Linear Unbiased Estimator (BLUE)

- Minimizes the error variance
- Assumes the error distributions are Gaussian and errors are uncorrelated

- x as the prior estimate of the state.
- P as the covariance matrix of the prior estimate.
- y as the new observation.
- R as the covariance matrix of the observation error.
- H as the observation matrix which maps the state space into the observation space.
- W as the weight matrix.

$$\hat{x} = x + W(y - Hx)$$

$$W = PH^T(HPH^T + R)^{-1}$$

$$\hat{P} = (I - WH)P$$

Methods

Kalman Filter (KF)

- Recursive filter that estimates state based on a series of noisy measurements.

- x as the prior state estimate.
- P as the prior covariance matrix.
- y as the observation.
- R as the observation noise covariance.
- H as the observation matrix.
- F as the state transition matrix.
- Q as the process noise covariance.

1. **Predict:**

$$\hat{x}^- = Fx$$

$$P^- = FPF^T + Q$$

2. **Update:**

$$K = P^- H^T (HP^- H^T + R)^{-1}$$

$$\hat{x} = \hat{x}^- + K(y - H\hat{x}^-)$$

$$P = (I - KH)P^-$$

Methods

Extended Kalman Filter (EKF)

- Extension of KF, linearizes the current mean and variance (Taylor series).

- x as the prior state estimate.
- P as the prior covariance matrix.
- y as the observation.
- R as the observation noise covariance.
- Nonlinear state transition function $f(x)$.
- Nonlinear observation function $h(x)$.
- Jacobians F and H of $f(x)$ and $h(x)$ at x, respectively.

1. **Predict:**

$$\hat{x}^- = f(x)$$

$$P^- = FPF^T + Q$$

2. **Update:**

$$K = P^- H^T (HP^- H^T + R)^{-1}$$

$$\hat{x} = \hat{x}^- + K(y - h(\hat{x}^-))$$

$$P = (I - KH)P^-$$

Methods

Ensemble Kalman Filter (EnKF)

- KF variant that uses an ensamble of forecasts (Monte Carlo approach) to handle non-linear DA relationships.

- x_{i-1}^a as the previous assimilated state estimate.
- w_i as the process noise.
- y as the observation.
- R as the observation noise covariance.
- Nonlinear model dynamics function $f(x)$.
- Observation matrix H.
- Covariance of ensemble members P^f.

1. **Predict:**

$$x_i^f = f(x_{i-1}^a) + w_i$$

2. **Update:**

$$K = P^f H^T (H P^f H^T + R)^{-1}$$

$$x_i^a = x_i^f + K(y - H x_i^f)$$

Methods

3D Variational Data Assimilation (3DVar)

- Adjusts a forecast by minimizing a cost function that represents the difference between model state and observations.

- x_a is the analysis state (updated state)
- x_b is the background state (prior estimate).
- B is the background error covariance matrix.
- y is the observation vector.
- H is the observation operator (linear or nonlinear) that maps the state space to the observation space.
- R is the observation error covariance matrix.

$$x_a = \operatorname{argmin}_x J(x)$$

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - Hx)^T R^{-1}(y - Hx)$$

Methods

4D Variational Data Assimilation (4DVar)

- Adjusts a forecast by minimizing a cost function that represents the difference between model state and observations.

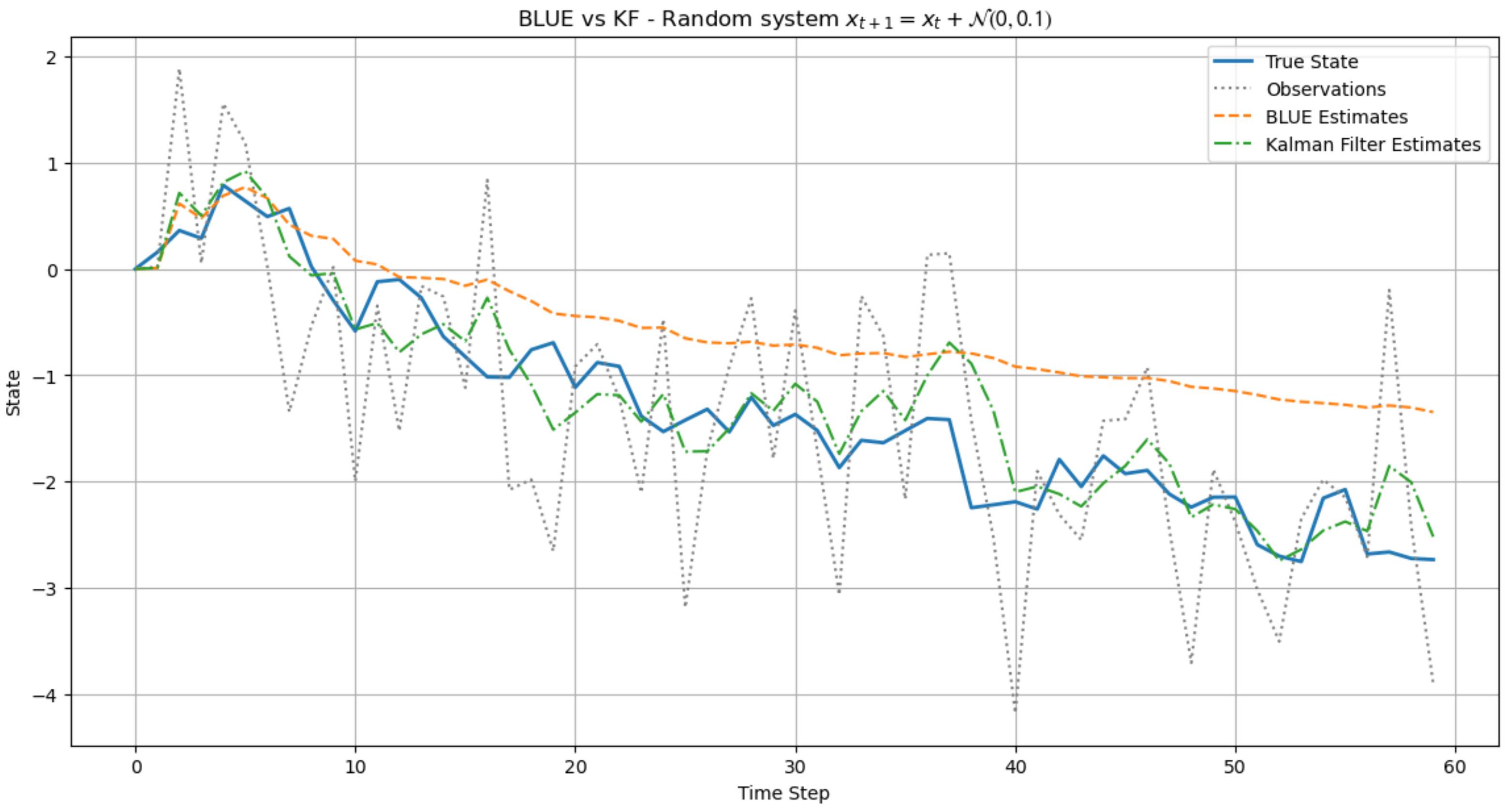
- x_0 is the initial state from which future states x_k are derived.
- x_b is the background state (prior estimate at the initial time).
- B is the background error covariance matrix.
- y_k are the observations at time k.
- H_k are the observation operators at time k.
- R_k are the observation error covariances at time k.
- x_k are states predicted from x_0 using the model dynamics.

$$x_a = \operatorname{argmin}_x J(x)$$

$$J(x_k) = \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b) + \frac{1}{2}\sum_{k=1}^{n}(y_k - H_k x_k)^T R_k^{-1}(y_k - H_k x_k)$$

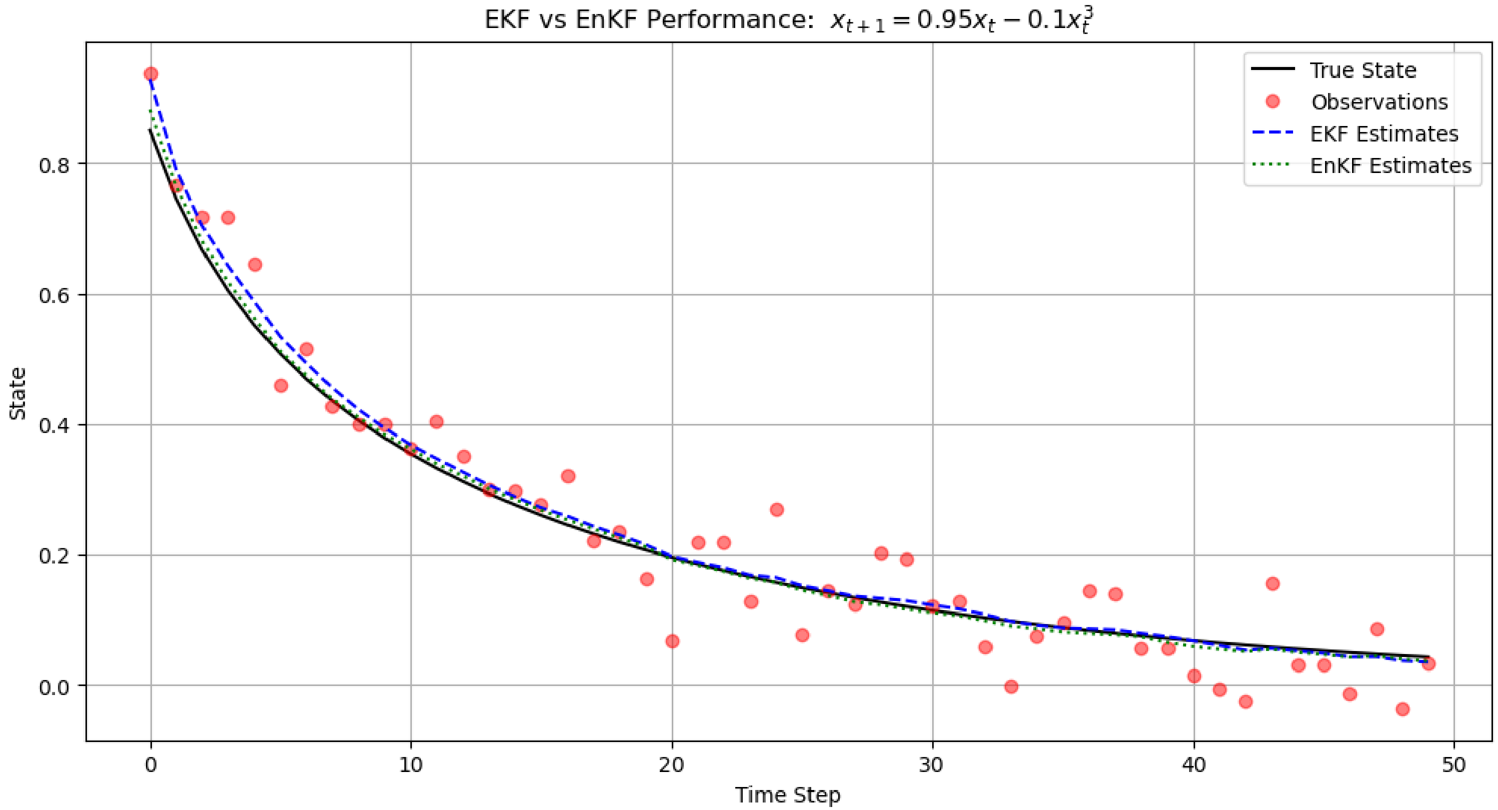
Experiment 1

- Random walk $x_{t+1} = x_t + \mathcal{N}(0, 0.1)$
- 60 time steps.
- State transition and observation matrices = 1
- Process noise covariance = 0.1
- Observation noise covariance = 1
- Initial state= 0
- Initial covariance = 10
- BLUE vs KF



Experiment 2

- Damped pendulum $x_{t+1} = 0.95x_t - 0.1x_t^3$
- Jacobians for EKF linearization
- 50 time steps
- Initial state = 1.0
- Process noise standard deviation 0.02
- Observation noise standard deviation = 0.05
- Observation noise covariance = 0.05
- Ensemble size for EnKF = 100
- EKF vs EnKF estimates



Next steps and Challenges

- Test localization, inflation
- Finish/troubleshoot 3DVar and 4DVar
- Make it more flexible for larger and multivariate problems
- Implement more complex/realistic examples
- WIP: Jupyter notebook with concepts and examples.

Thank you!

- Questions?