A comparative analysis of Koopman Kalman Filter and the Extended Kalman Filter

December 6, 2019

1 Abstract

In this we compare a new filter designed based on operator theoretic framework of dynamical systems with a widely used variant of the Klamn Filter. Specifically, we use Koopman operator theory to derive a high-dimensional embedding of the nonlinear dynamics in which the system outputs evolve in a linear fashion. We see that by exploiting this linear structure of the Koopman operator, we can design linear observers like Kalman Filter for nonlinear state estimation. We compare the performance of this Koopman Kalman Filter with Extended Kalman Filter and observe that KKF shows superior performance. This is attributed to the fact that EKF performs a local linearization of the system whie Koopman theory facilitates (nearly) global linearization of nonlinear dynamics.

2 Preliminaries

2.1 Koopman Operator Theory

In this section, we present some preliminaries on the Koopman operator theoretic framework. Consider a discrete controlled nonlinear dynamical system given below:

$$x_k = F(x_{k-1}, u_{k-1}) \tag{1}$$

where $x_k \in \mathcal{X} \subseteq \mathbb{R}^n$ is the state of the system, $u_k \in \mathcal{U} \subseteq \mathbb{R}^m$ is the input, F is the flow map that evolves the system states forward in time, and $k \in \mathbb{Z}$ represents the time step such that $t_k = k\Delta$ where Δ is the sampling time.

Let us define a set of scalar-valued *observables* that are functions of the states and the inputs, $g : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$. Each observable is an element of an infinite-dimensional function space $\mathcal{G}$ which, for example, can be defined by the Lebesgue square-integrable functions, $\mathcal{G} = \ell^2(\mathcal{X} \times \mathcal{U}, \mathbb{R})$, or other appropriate spaces [**?**]. In this infinite-dimensional function space, the flow of the system is governed by the Koopman operator, $\mathcal{K} : \mathcal{G} \to \mathcal{G}$ that defines the dynamics of observables $g \in \mathcal{G}$ along the trajectories of the system as:

$$\mathcal{K}g \triangleq g \circ F \tag{2}$$

By definition, the Koopman operator is linear even though the underlying dynamical system is nonlinear. For all $g_1, g_2 \in \mathcal{G}$ and all $\alpha, \beta \in \mathbb{R}$, it satisfies

$$\begin{aligned} \mathcal{K}(\alpha g_1 + \beta g_2) &= (\alpha g_1 + \beta g_2) \circ F \\ &= (\alpha g_1 \circ F) + (\beta g_2 \circ F) \\ &= \alpha \mathcal{K} g_1 + \beta \mathcal{K} g_2 \end{aligned} \tag{3}$$

The system identification method described below exploits the fact that any finite-dimensional nonlinear system can be equivalently represented using an infinite-dimensional linear system by transforming the traditional state-space to the space of functions (observables) of the system's states and inputs.

2.2 System identification using EDMD

The Koopman operator theory is conceptually developed on the infinite-dimensional function space $\mathcal{G}$. However, it is not practically feasible unless we can determine finite-dimensional approximations to the Koopman operator without a great loss in accuracy. To do this, consider a finite-dimensional subspace $\bar{\mathcal{G}} \subset \mathcal{G}$ spanned by a set of basis functions $\phi(x,u) = [\phi_1(x,u), \phi_2(x,u), \cdots, \phi_{N\phi}(x,u)]^T$. Now, any observable function $\psi \in \bar{\mathcal{G}}$ can be represented as a linear combination of these basis functions as follows:

$$\psi = c_1\phi_1 + c_2\phi_2 + \cdots + c_{N_\phi}\phi_{N_\phi} = c^T\phi \tag{4}$$

For these functions in $\bar{\mathcal{G}}$, we seek to generate the finite-dimensional approximation of the Koopman operator, denoted as $K \in \mathbb{R}^{N_\phi \times N_\phi}$. Because, typically $\bar{\mathcal{G}}$ is not invariant with respect to $\mathcal{K}$, there is a residual term which is minimized in the L_2-norm sense via linear regression [**?**]. Intuitively, if these observables ψ represent the system's outputs (or equivalently, sensor measurements), a cost function to be minimized in an optimal control problem, or nonlinear constraints, then the evolution of these "nonlinear" observables can now be analysed using a linear system via K. This is done by carrying a nonlinear transformation of the system states (outputs) to the so-called "lifted" space using the set of N_ϕ basis functions.

In its most general form, the Koopman linear system is presented as follows:

$$\phi(x_k, u_k) = K(x_{k-1}, u_{k-1})\phi(x_{k-1}, u_{k-1}) \tag{5}$$

where $\phi \in \mathbb{R}^{N_\phi}$ and usually $N_\phi \gg n$. For the augmented state (x,u), the above formulation will admit any nonlinear dynamical system. However, as the dimension of the original system increases, the number of basis functions to be considered can grow to infeasible levels. So, to make the computations tractable and amenable, the simplifications described below are introduced in the structure of the nonlinear transformation [**?**].

Specifically, the objective is to identify a linear dynamical system of the following form (similar to a linear time invariant state-space model) in the space of observables using time-series data generated by the control system Eq. (1).

$$\phi(x_k) = A\phi(x_{k-1}) + Bu_{k-1} \tag{6}$$

For this purpose, we use the recently developed EDMD algorithm. To construct a finite dimensional approximation to the Koopman operator for the controlled system in Eq. (1), the EDMD algorithm requires

1. a time-series data set of N_t snapshot pairs satisfying the dynamical system in Eq. (1) which can be organized in the following matrices.

$$\begin{gathered} X = [x_1, x_2, \cdots, x_{N_t}], \quad Y = [y_1, y_2, \cdots, y_{N_t}], \\ U = [u_1, u_2, \cdots, u_{N_t}] \end{gathered} \tag{7}$$

   Note that we use y instead of x_{k+1} here because the data above need not be temporally ordered as long as it satisfies $y_k = F(x_k, u_k)$.

2. a library of nonlinear basis functions $\{\phi_1, \phi_2, \ldots, \phi_{N_\phi}\}$ whose span is $\bar{\mathcal{G}} \subset \mathcal{G}$.

The EDMD algorithm then seeks to solve a least-squares problem to obtain K which is the transpose of the finite-dimensional approximation to the Koopman operator, $\mathcal{K}$:

$$\min_{K} \quad \sum_{k=1}^{N_t} \|\phi(y_k, u_k) - K\phi(x_k, u_k)\|_2^2 \tag{8}$$

In order to obtain the form described in Eq. (6), the following simplification can be introduced into the formulation. Since we are not interested in prediciting the future values of inputs, without loss of generality we can assume that

$$\phi(x, u) = \begin{bmatrix} \phi(x) \\ u \end{bmatrix}, \qquad \phi(y, u) = \begin{bmatrix} \phi(y) \\ u \end{bmatrix} \tag{9}$$

which is to say that the nonlinear observable functions are applied to the system states alone and not the inputs. Suppose that N nonlinear functions are used to lift the states to the observable space and the m system inputs are added separately to the basis as shown in Eq. (9), we have $N_\phi = N + m$. To obtain the evolution of lifted states, we can then disregard the last m components of each of the terms in Eq. (8), decompose the Koopman matrix as $K = [A, B]$ and use only the first N rows in K, which leads to the following minimization problem

$$\min_{A,B} \quad \sum_{k=1}^{N_t} \|\phi(y_k) - A\phi(x_k) - Bu_k\|_2^2 \tag{10}$$

The practical solution to the above equation is obtained by using regularization via a truncated singular value decomposition and the value of K that minimizes Eq. (10) is given by

$$K_N := [A, B] = \phi_{XY}\phi_{XX}^{\dagger} \tag{11}$$

where † denotes the pseudoinverse, and the matrices are computed as

$$\begin{gathered}\phi_{XX} = \begin{bmatrix}\phi_X \\ U\end{bmatrix}\begin{bmatrix}\phi_X \\ U\end{bmatrix}^T, \quad \phi_{XY} = \phi_Y \begin{bmatrix}\phi_X \\ U\end{bmatrix}^T \\ \phi_X = [\phi(x_1), \cdots, \phi(x_{N_t})], \quad \phi_Y = [\phi(y_1), \cdots, \phi(y_{N_t})]\end{gathered} \tag{12}$$

with

$$\phi(x_k) = \begin{bmatrix}\phi_1(x_k) \\ \vdots \\ \phi_N(x_k)\end{bmatrix} \tag{13}$$

Please note that any solution to Eq. (11) is a solution to Eq. (10), and the formulation of Eq. (11) has an advantage of being independent of the number of data samples N_t. Therefore, the Koopman linear system obtained using the above algorithm will be in the form of the controlled linear dynamical system given by Eq. (6), where $\phi \in \mathbb{R}^N$ is the lifted state in the observable space, and $A \in \mathbb{R}^{N\times N}$ and $B \in \mathbb{R}^{N\times m}$ are the matrices that describe the system dynamics in the lifted space.

The Koopman linear system Eq. (6), derived using the above algorithm, governs the evolution of the basis ϕ in the lifted space. With the use of these basis functions, all observables of interest $\psi \in \mathbb{R}^{n_\psi}$ can be determined (refer to Eq. (4)) by simply using a matrix of coefficients as $\psi(x_k) = C\phi(x_k)$, where $C \in \mathbb{R}^{n_\psi \times N}$. In reality, not all observables can be contained in the span of the chosen subspace (recall, we are only using a finite-dimensional truncated function space). Thus, the matrix C must be chosen such that the projection of ψ onto the span$\{\phi_1, \cdots, \phi_N\}$ is minimized in the L_2-norm sense as below:

$$\min_C \quad \sum_{k=1}^{N_t} \|\psi(x_k) - C\phi(x_k)\|_2^2 \tag{14}$$

When dealing with dynamical systems, since the goal is stability analysis or controller synthesis, it is advantageous to reproduce the state dynamics. Therefore, the observable of interest here is the state itself, $\psi = x$. In such cases, we can assume that the basis also contains the state observable, i.e., $[\phi_1, \cdots, \phi_n]^T = x$. In particular, such a set of basis functions is said to be *state-inclusive*, and typically the solution to matrix C is trivial and can be obtained by $C = [I_n, 0_{N-n}]$. If the system outputs are described as functions of the state, i.e., $z_k = h(x_k)$, then they can be obtained in a similar fashion with $\psi = z$.

In summary, the evolution of full state observable x_k and the system outputs z_k can be expressed via a linear time invariant system as shown below:

$$\begin{aligned} \phi_k &= A\phi_{k-1} + Bu_{k-1} \\ z_k &= h(x_k) = C^h\phi_k \\ x_k &= C^x\phi_k \end{aligned} \tag{15}$$

2.3 Koopman Kalman Filter

Given the linear form of the Koopman invariant system as shown in Eq.(15), one can then design a standard Kalman Filter as proposed in [cite]. As a part of the preliminaries, discrete-time Kalman filter will be reviewed in this section. Assume a linear system with noise

$$\begin{aligned} x_k &= Fx_{k-1} + Gw_{k-1} \\ z_k &= Hx_k + v_k \end{aligned} \tag{16}$$

where $E[w] = E[v] = 0, E[ww^T] = Q, E[vv^T] = R$ and $E[wv^T] = 0$. Let us suppose it is given that $E[x_{k-1}] = \hat{x}_{k-1}$ and $E[(x_{k-1} - \hat{x}_{k-1})(x_{k-1} - \hat{x}_{k-1})^T] = P_{k-1}$. Then, Kalman Filter gives the optimal estimate of the state $\hat{x}_k$ given the measurement z_k.

This is done in two recurrent steps, prediction and update. In the prediction step, the KF will propagate the optimal estimate at t_{k-1} to t_k which is called the *prior* denoted as $\hat{x}_k^-$. After measurement z_k comes in, this additional information is utilized to update it to the *posterior*, that is $\hat{x}_k$. The steps are given below:
Prediction:

$$\begin{aligned} \hat{x}_k^- &= F\hat{x}_k \\ P_k^- &= FP_{k-1}F^T + GQG^T \end{aligned} \tag{17}$$

Update:

$$\begin{aligned} \hat{x}_k &= \hat{x}_k^- + P_kH^TR^{-1}(z_k - H\hat{x}_k^-) \\ P_k^{-1} &= (P_k^-)^{-1} + H^TR^{-1}H \end{aligned} \tag{18}$$

This is the optimal solution to the estimation problem in the case of linear system with Gaussian probability density.

Now, in the case of the Koopman Kalman Filter (KKF) the above equations are used to estimate the states in the lifted Koopman space, i.e, ϕ_k where the dynamics are given by Eq. (15). An additional step is required in case of KKF which initializes the lifted states ϕ_0 based on initial state specifications in the original coordinate x_0. Also, the initial distribution of ϕ_0 is obtained using the matrices in Eq. (15) as give below.

- *Initial state*: $\phi_0 = \phi(x_0)$
- *Initial distribution*: Suppose a distribution is specified for the initial state x_0 such as $x_0 \sim \mathcal{N}(\bar{x}_0, P_0^x)$, the authors in [] propose that a numerical procedure can be used to obtain the corresponding distribution on ϕ_0. Specifically, since the Kalman Filter assumes a normal distribution for

the state variables and noise terms the following approximation is used to obtain the corresponding normal distribution in the Koopman space.

$$\bar{\phi}_0 = (C^x)^\dagger \bar{x}_0, \qquad P_0^\phi = (C^x)^\dagger P_0^x ((C^x)^\dagger)^T \tag{19}$$

- *Updated distribution in x*: Once the filter estimates in the Koopman coordinates $\{\hat{\phi}_k\}_{k=1}^t$ have been determined fro a given output sequence $\{z_k\}_{k=1}^t$, the corresponding state estimates in the original coordinates $\{\hat{x}_k\}_{k=1}^t$ and the covariance matrix can be obtained by following equations

$$\hat{x}_k = (C^x)\hat{\phi}_k, \qquad P_k^x = C^x P_k^\phi (C^x)^T \tag{20}$$

Remark 1 *Please note that even when the original state is normally distributed as shown above, the ϕ_0 is not normally distributed because of the nonlinear transformation involved in lifting the state to higher dimension. Due to non-normally distributed initial state, the distribution on ϕ will evolve as a non-normal distribution and Kalman Filter is only an approximation even in the case when the Koopman transformation is an exact linearization and valid globally. This will be seen in the numerical example considered in the next section.*

2.4 Extended Kalman Filter

In order to study the performance of the KKF described above, we compare it with the widely used Extended Kalman Filter (EKF). Here we briefly describe he EKF for the sake of completion.

Assume a nonlinear system,

$$\begin{aligned} x_k &= f(x_{k-1}) + G w_{k-1} \\ z_k &= h(x_k) + v_k \end{aligned} \tag{21}$$

The implementation of EKF is very similar to the traditional KF, except the nonlinear functions f, h above are (locally) linearized using the Jacobian.
Prediction:

$$\begin{aligned} \hat{x}_k^- &= f(\hat{x}_k) \\ P_k^- &= F P_{k-1} F^T + G Q G^T \end{aligned} \tag{22}$$

Update:

$$\begin{aligned} \hat{x}_k &= \hat{x}_k^- + P_k H^T R^{-1} (z_k - h)\hat{x}_k^-)) \\ P_k^{-1} &= (P_k^-)^{-1} + H^T R^{-1} H \end{aligned} \tag{23}$$

where

$$F = \frac{\partial f}{\partial x}\Big|_{\hat{x}_{k-1}}, \qquad H = \frac{\partial h}{\partial x}\Big|_{\hat{x}_{k-1}} \tag{24}$$

3 Unicycle robot model

In this section, we demonstrate the comparison of Koopman Kalman Filter with an Extended Kalman Filter for the case of a unicycle robot. The equations of motion of a unicycle robot are presented below

$$\begin{aligned} \dot{x} &= v cos\theta \\ \dot{y} &= v sin\theta \\ \dot{\theta} &= \omega \end{aligned} \tag{25}$$

where x, y denote the position of the unicycle robot with respect to the xy plane, θ denote the angle and v and ω denote the action variables i.e., linear velocity and angular velocity of the unicycle orientation in the xy plane.

We may discretize the above EOM using Δt discretization time as below:

$$\begin{aligned} x' &= x + v cos\theta \Delta t \\ x' &= y + v sin\theta \Delta t \\ \theta' &= \theta + \omega \Delta t \end{aligned} \tag{26}$$

We assume that the "nominal" commanded values of linear and angular velocities $(\bar{v}, \bar{\omega})$ are corrupted by continuous white noise $(\delta v, \delta\omega)$ and their distributions are given by $\delta v \sim N(0, \sigma_v^2)$, $\delta\omega \sim N(0, \sigma_\omega^2)$.

We assume that the GPS measurements are available to us i.e.,

$$z' = [x' \ y']^T + r' \tag{27}$$

where $s \sim \mathcal{N}(0, R), R = diag([\sigma_{s_1}^2 \ \sigma_{s_2}^2])$ denotes the sensor (measurement) noise.

The values considered in the following simulations are given below:

Parameter	Value
$\bar{v}$	1
$\bar{w}$	0.1
σ_v	1
σ_w	$\pi/6$
σ_{s_1}	1
σ_{s_2}	1

3.1 Deriving Koopman Linear System

In the specific system considered above, the dictionary of observable functions considered are shown below:

$$\phi = [x \ y \ \theta \ sin\theta \ cos\theta]^T \tag{28}$$

Using the above dicitonary, we can obtain a Koopman linear form of the original system in Eq.(26). Specifically, if we take sin on both sides of the equation for

θ in Eq.(26) and apply $sin(a+b) = sinacosb + cosasinb$ and similarly taking cos we obtain the follwing linear form of the original system.

$$\begin{bmatrix} x' \\ y' \\ \theta' \\ sin\theta' \\ cos\theta' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & v\Delta t \\ 0 & 1 & 0 & v\Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & cos\omega\Delta t & sin\omega\Delta t \\ 0 & 0 & 0 & -sin\omega\Delta t & cos\omega\Delta t \end{bmatrix} \begin{bmatrix} x \\ y \\ \theta \\ sin\theta \\ cos\theta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta t \\ 0 \\ 0 \end{bmatrix} \omega \tag{29}$$

Note that the above linear form is an exact linearization of the original system and is globally valid in this case.

3.2 Numerical Experiments

Now that the Koopman linear form is determined for the unicylce dynamics, in this section we applied the KKF and EKF to compare the performance of both the filters in estimating the states given a set of GPS measurements. Also, we assumed that the initial condition is uncertain with a normal distribution $x_0 \sim \mathcal{N}(\bar{x}, P_0^x)$ where $\bar{x}_0 = [0\ 0\ 0]^T, P_0^x = diag([5\ 5\ 5])$. In the simulations, the discretization time step is considered to be $\Delta t = 0.1$ and the measurements are assumed to be available at each time step, i.e., at the rate of 10Hz. We considered two case studies, with different linear to angular velocity ratios. For example, a larger value of v/ω implies that the robot traverses a longer arc compared to the case with a smaller v/ω in which the rotation is more compared to the translation. In both cases, the performance of KKF and EKF are compared with respect to the accuracy of the state estimates. To quantify the estimation accuracy, we considered an averaged root mean square error as a function of time for a number of Monte Carlo runs as shown below:

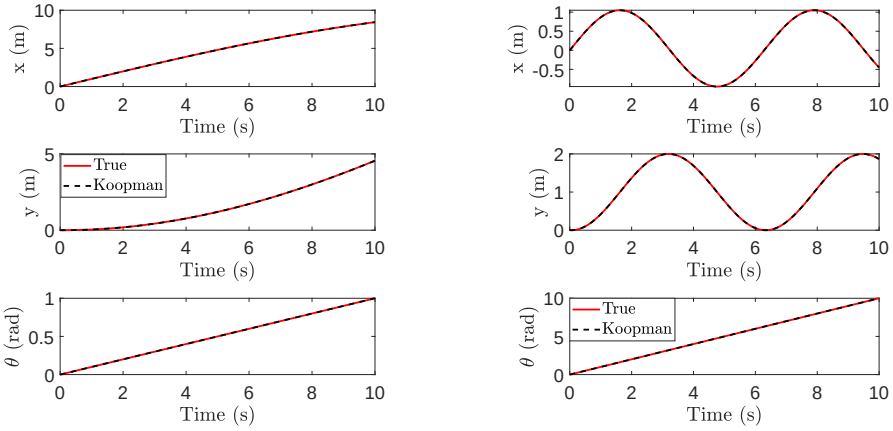
$$\mathcal{R}_k = \sqrt{\frac{1}{N_r}\sum_{i=1}^{N_r} \|x_k^i - \hat{x}_k^i\|_2^2}, \quad k = 1, \cdots, t \tag{30}$$

where x is the actual state, $\hat{x}$ is the estimated state, N_r is the number of Monte Carlo runs with random initial states sampled from the initial distribution described above and $\|\cdot\|_2$ denotes the vector L_2-norm.

3.3 Results

3.3.1 Koopman model validation

Figure 1 shows the output predictions of the above determined Koopman linear model for two different inputs (randomly) chosen for the linear and angular velocities. From the figure it can be seen that the true model and the identified linear model is overlapping in both the cases resulting in a very good accuracy. This is expected because the derived linearization is exact (not approximate) and is therefore valid globally.



(a) (b)

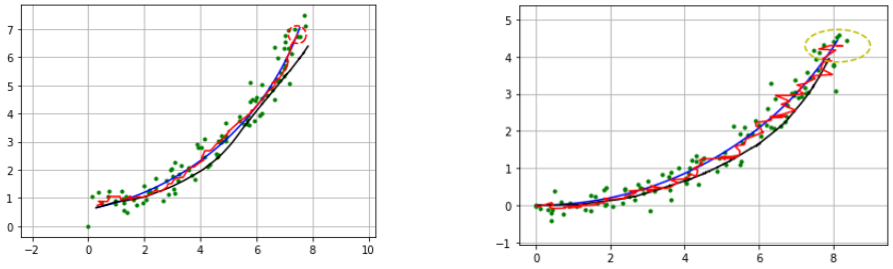
Figure (1) Prediction comparison of the unicycle response for two different inputs. (a) $v/\omega = 10$ and (b) $v/\omega = 1$

3.3.2 State Estimation

Now that the Koopman model has been validated, we used the linear model described in Eq. (29) to perform state estimation using a Kalman Filter as described in the previous section. In order to evaluate the performance of the Koopman based Filter KKF, we compared its estimation performance with the commonly used EKF. We considered two cases: (a) $v/\omega = 10$ and (b)$v/\omega = 100$. Figure shows the estimation results of KKF and EKF respectively for Case (a) and Figure shows the estimation results of KKF and EKF respectively for Case (b). From the figures, we can see that in Case (a) the KKF performs slightly better than EKF, however, in Case (b) the performance of both the filters is comparative. Figures show the averaged root mean squared error for both the cases for EKF and KKF respectively.

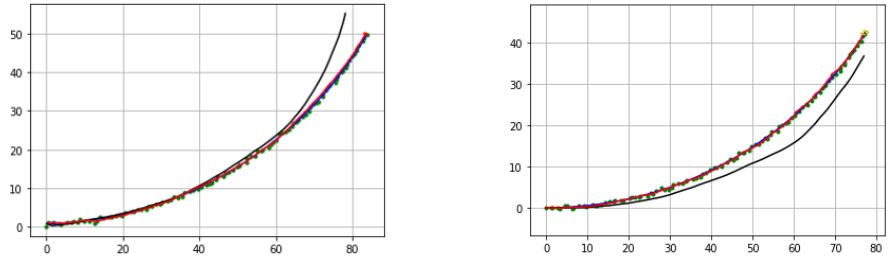
4 Conclusions

Koopman operator theory offers a new direction in state estimation theory. Specifically, it facilitates linear observer design for nonlinear state estimation. The idea is to obtian a Koopman linear model for the underlying nonlinear system and then design any liner filters available to perform the estimation. In the example considered in this project we obtained an exact linearization of the original system which makes the Koopman model valid globally. Because it is a global linearization it is expected to perform better than EKF. However, due to the approximation present in lifitng the initial distribution where we assume Gaussian even for the lifted variable (this is clearly not valid as we



(a) (b)

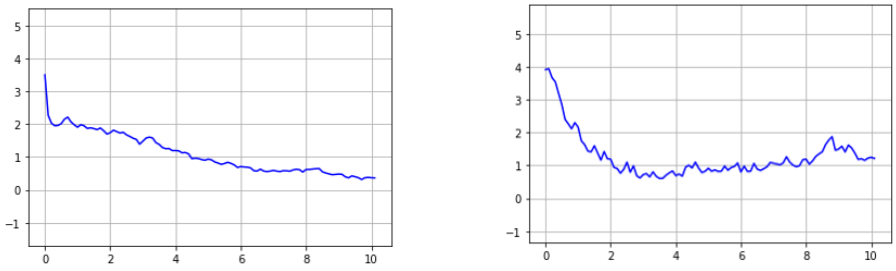
Figure (2) Estimation comparison of the unicycle response for two different Filters for the case $v/\omega = 10$ (a) EKF (b) KKF



(a) (b)

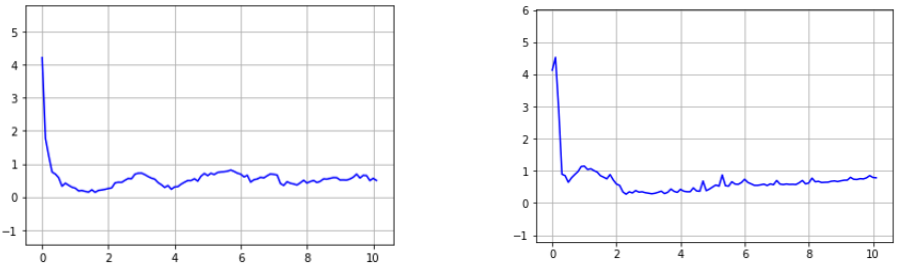
Figure (3) Estimation comparison of the unicycle response for two different Filters for the case $v/\omega = 100$ (a) EKF (b) KKF

perform a nonlinear transformation) the performance is comparable if not better in the case of KKF. However, in future it will be interesting to study KKF for more complex systems where EKF fails. Additionally, it will be of interest to study the performance when the frequency of measurements is lower than what was considered in this project because it is well known that EKF performance degrades with a lower rate of measurements.



(a) (b)

Figure (4) Average Estimation error comparison of the unicycle response for two different Filters for the case $v/\omega = 10$ (a) EKF (b) KKF



(a) (b)

Figure (5) Average Estimation error comparison of the unicycle response for two different Filters for the case $v/\omega = 100$ (a) EKF (b) KKF