Kalman Filtering for Discrete-Time Linear Systems with Infinite-Dimensional Observations

Maxwell M. Varley, Timothy L. Molloy, Girish N. Nair

Abstract— **Estimating the finite-dimensional state of dynamic systems using modern sensors such as cameras, lidar, and radar involves processing increasingly high-dimensional observations. In this paper, we exploit concepts from the theory of infinite-dimensional systems to examine state estimation in the continuum limit of infinite-dimensional observations. Specifically, we investigate state estimation in discrete-time linear systems with finite-dimensional states and infinite-dimensional observations corrupted by additive noise. In contrast to previous derivations of the Kalman filter for infinite-dimensional observations, we are able to derive an explicit solution for the optimal Kalman gain by modeling the infinite-dimensional observation noise as a *stationary* Gaussian Process. We demonstrate the utility of our Kalman filter in a simulation of a linearized system derived from the pinhole camera model.**

I. Introduction

Autonomous navigation and state estimation of physical systems with high-dimensional observations is an area of research with a wide array of applications. These applications include autonomous navigation in indoor environments, outdoor environments, and extraterrestrial environments [1] [2] [3]. This high-dimensional observation space introduces delays to the system, which adversely affects real time state estimation. Due to the high dimensionality of this data, algorithms that process this data in real-time require some method of dimensionality reduction. This takes the form of either feature selection, where there must be some process to omit a large amount of data, or feature extraction, where the data is in some way transformed to a lower dimension. These feature detection methods are generally heuristics-based and involve finding regions of interest via edge detection, corner detection, blob detection, or some other heuristic. For this reason, algorithms that require speedy real-time performance such as those used in autonomous navigation often consist of a front-end stage for feature detection and a back-end stage for optimization of state estimates [3]. State of the art autonomous navigation algorithms such as ORB-SLAM2 [4] and SVO [5] include this front-end/back-end split design. Control techniques for performing visual state estimation, specifically visual SLAM, have recently been developed using noise free observers [6] [7]. There is also recent work on control with high-dimensional states and observations [8].

In contrast to these approaches, this paper proposes a stochastic approach in which we treat high-dimensional observational measurements as infinite-dimensional Gaussian Processes. By approaching the continuum limit of the system, we are able to apply well-known concepts from the theory of stochastic infinite-dimensional systems.

The authors are with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC, 3010, Australia. varleym@student.unimelb.edu.au, tim.molloy @ unimelb.edu.au,gnair@unimelb.edu.au

This work received funding from the Australian Government, via grant AUSMURIB000001 associated with ONR MURI grant N00014-19-1-2571.

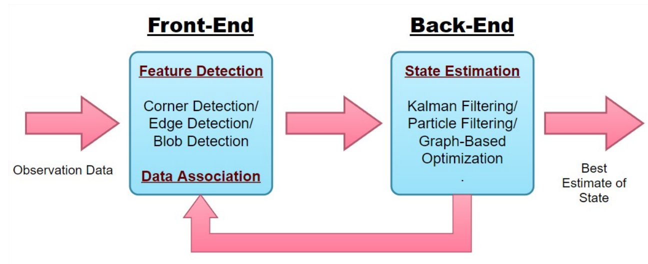


Fig. 1. Example form of vision-based state estimation. Adapted from [3].

In contrast to finite-dimensional systems where the state/measurements are represented by vectors with a discrete i indexing the vector components, an infinite-dimensional system represents the state/measurements with functions that are indexed by a continuous i. These systems can be used to model distributed-parameter systems in areas like material engineering, chemical engineering, and biotechnology [9], as well as systems with high-dimensional measurements such as vision, radar, or lidar. This paper examines performing state estimation in a system with finite-dimensional states and infinite-dimensional measurements by designing a Kalman filter for this system. This filter echoes similar work done in this area. Distributed-parameter Kalman filters for continuous-time stochastic systems have been explored and derived by authors such as Meditch and Tzafestas [10] [11] [12] [13]. Tzafestas has also derived the distributed-parameter Kalman filter for discrete-time stochastic systems using a Bayesian approach [12], and the discrete-time Kalman filter has been derived via the Wiener-Hopf equation by Nagamine et al. [14]. Minimum variance estimators for distributed-parameter systems are discussed in detail by Morris [15] [16], who also discusses the controllability and observability of such systems and optimal sensor location for purposes of estimation. All of these works provide Kalman filtering algorithms for distributed-parameter systems, but rely on the existence of a certain implicitly defined inverse operator, which Tzafestas labels the distributed-parameter matrix inverse.

We are interested in a system with finite-dimensional state but infinite-dimensional measurements on an infinite domain D. These measurements are corrupted by additive zero-mean Gaussian Process noise that is stationary on D, taken to be the real line. This is motivated by a mobile robot that wishes to estimate its low-dimensional pose from high-dimensional observations, the infinite-dimensional measurements representing the continuum limit of the high resolution sensor. These observations could be based on vision, sonar, lidar, or any other high-dimensional sensor modality. We show that applying the previously described distributed-parameter Kalman filter to this scenario may result in an operator inverse that does not exist.

The key contribution of this paper is the formulation of a discrete-time Kalman filter for linear systems with finite-dimensional states and infinite-dimensional observations. By exploiting the stationarity of the infinite-dimensional observation noise we calculate an explicit, well-defined solution for the optimal Kalman gain as the kernel of an integral over the domain of interest D. Many vision-based SLAM techniques rely on heuristic-based front-end feature detection techniques which are separated from back-end state estimation, resulting in the feature extraction procedure taking little to no account of the system's state dynamics [17] [18]. While the simulations in section V are motivated by a simple vision-based system, our solution is applicable to other modalities that are measured via high-dimensional sensors. Such sensors include vision, but also sensors like sonar, radar, and lidar. The current approach is constrained by the assumption of a linear system, but serves as a rigorous foundation for future analysis. In particular, we believe it will lead to an understanding of how to extract state estimates from high-dimensional sensors in a principled manner, and in a way that inherently accounts for the system dynamics.

This paper is structured as follows. Section II discusses Gaussian Processes and Gateaux Derivatives. Section III presents the form of a system with finite-dimensional states and infinite-dimensional observations, as well as the noise properties of our system. Section IV derives an explicit solution for the Kalman gain associated with our system and an algorithm for applying the Kalman filter to this system. Section V applies this algorithm to a simulated system and presents the resulting behavior and performance of the filter. Section VI concludes the paper.

II. PRELIMINARIES

A. Gaussian Processes

A Gaussian Process is an $\mathbb{R}^m$-valued random process such that any finite set of samples of that process forms a multivariate Gaussian distribution. A zero-mean Gaussian Process $(v(i))_{i\in D}$ on domain D is completely described by its covariance function $R(i,i') = E[v(i)v(i')^T]$. The Gaussian process is called stationary if $R(i,i') = R(i-i'), \forall i,i' \in D$. All covariance functions are positive semi-definite, i.e. for all $n \in \mathbb{N}$, $i_1,\dots,i_n \in D$, and $u_1,\dots,u_n \in \mathbb{R}^m$,

$$\sum_{j=1}^{n}\sum_{k=1}^{n} u_j^T R(i_j - i_k) u_k \geq 0.$$

B. Gateaux Derivatives

The Gateaux derivative generalizes the directional derivative and can be extended to infinite-dimensional spaces.

Definition 2.1: Let X be a vector space, Y a normed space, and $F : X_1 \to Y_1$ a transformation with $X_1 \subset X, Y_1 \subset Y$. If it exists for all $\phi \in X_1$, the Gateaux derivative is given by

$$dF(x,\phi) = \lim_{\epsilon\to 0} \frac{F(x+\epsilon\phi) - F(x)}{\epsilon}. \tag{1}$$

where $x \in X_1$, ϕ is arbitrary and ϵ is scalar-valued. A necessary condition at any extrema of an unconstrained functional is that the Gateaux derivative must be equal to 0 for any arbitrary test function $\phi(i)$ [19].

III. PROBLEM FORMULATION

Consider a system model with state dynamics

$$x_{k+1} = Ax_k + w_k, \tag{2}$$

where $x_k \in \mathbb{R}^n$, $w_k \in \mathbb{R}^n, A \in \mathbb{R}^{n\times n}$, and $k \in \mathbb{N}$ denote the state vector, process noise, state matrix, and time index respectively. For ease of exposition we assume that the control input is zero; the results here would still hold with minor changes if it were non-zero and available to the estimator. The state is not measured directly but is observed via an infinite-dimensional observation waveform defined on domain D, that is,

$$z_k(i) = (\Gamma x_k)(i) + v_k(i) \in \mathbb{R}^m, \quad i \in D, \tag{3}$$

where $v_k = (v_k(i))_{i\in D}$ is a stationary Gaussian Process with covariance R.

The linear operator $\Gamma : \mathbb{R}^n \to L^2(D,\mathbb{R}^m)$ is defined by

$$(\Gamma x_k)(i) \triangleq \gamma(i)x_k \in \mathbb{R}^m, \tag{4}$$

where the function $\gamma \in L^2(D,\mathbb{R}^{m\times n})$. The initial state, process noise, and observation noise are jointly Gaussian such that $\forall j,k \in \mathbb{N}$ and $i,i' \in D$,

$$E[w_k] = 0 \tag{5}$$

$$E[v_k(i)] = 0 \tag{6}$$

$$E[v_k(i)w_j^T] = 0 \tag{7}$$

$$E[w_k w_j^T] = Q\delta_{j-k}, \qquad Q \in \mathbb{R}^{n\times n} \tag{8}$$

$$E[v_k(i)v_j^T(i')] = R(i-i')\delta_{j-k}, \quad R(i-i') \in \mathbb{R}^{m\times m}, \tag{9}$$

where Q and R are positive-definite and δ_{j-k} is the discrete impulse. We wish to examine the performance of a one-step predictor which takes the following form

$$\begin{cases} \hat{x}_{k+1} &= A\hat{x}_k + K_k[z_k - \hat{z}_k] \\ \hat{z}_k &= \Gamma\hat{x}_k. \end{cases} \tag{10}$$

Here K_k is a linear mapping from the space of stationary Gaussian processes on D to the space of Gaussian random vectors in $\mathbb{R}^n$. We assume the integral form

$$K_k\tilde{z} \triangleq \int_D \kappa_k(i)\tilde{z}(i)di,$$

where $\kappa_k(i) \in L^2(D,\mathbb{R}^{n\times m})$. This stochastic integral is defined in the Ito sense. Alternatively, from an engineering perspective it may be thought of as giving the unique random vector u jointly Gaussian with $\tilde{z}$ such that

$$\begin{aligned} cov[u] &= \int_D \int_D \kappa_k(i) G(i,i') \kappa_k^T(i') di di' \\ E[u\tilde{z}(i)^T] &= \int_D \kappa_k(i') G(i,i') di', \end{aligned}$$

where G is the covariance of the Gaussian process $\tilde{z}$ and the integrals, now deterministic, are defined in the usual Lebesgue sense. For present purposes, we shall assume that G and κ_k are suitably "nice" so that expectations and integrals can be swapped in order.

We highlight that the state estimation error previously defined as $e_k \triangleq x_k - \hat{x}_k$ satisfies[1]

$$e_{k+1} = M_k e_k + n_k,$$

where $M_k = A - K_k\Gamma \in \mathbb{R}^{n\times n}$, and $n_k = -K_k v_k + w_k \in \mathbb{R}^n$. Hence, the state estimation error dynamics and the best (component) gains $\kappa_k(i)$ to apply to points $i \in D$ in the observation domain depend on the dynamics of the system, A. In particular, as $k \to \infty$, the steady-state gain operator K must be chosen so that $A - K\Gamma$ is stable to yield bounded mean-square estimation errors. In other words, the gain must take account of the underlying dynamics. This observation is important since standard methods of processing high-dimensional observations (e.g., the front-end/back-end based approaches in vision-based state estimation illustrated in Fig. 1) typically perform the detection and weighting of observation features in isolation to the system dynamics. In this paper, we shall explore this point further by deriving the optimal Kalman gain K_k with kernel $\kappa_k(i)$, with optimality defined as minimizing the trace of the error covariance at every step.

IV. Derivation of Optimal Kalman Gain

In this section, we derive optimality conditions for the Kalman gain and our novel explicit form for it.

A. Optimal Kalman Gain Conditions

We now derive the form of the Kalman filter for this system by finding the gain kernel $\kappa_k(i)$ that reduces the mean square error via minimization of the covariance matrix at each time step. We begin by defining

$$P_k^- = AP_{k-1}A^T + Q \tag{11}$$

$$\hat{x}_k^- = A\hat{x}_{k-1}. \tag{12}$$

Here $P_k^- \in \mathbb{R}^{n\times n}$ defines the *a priori* estimation error covariance at time index k, i.e. the expected covariance of the error at time k, given measurements up to time $k-1$. $\hat{x}_k$ defines the estimated state vector at time k.

The application of (3) and (10) now gives us a formulation of the error at each time step, namely

$$\begin{aligned} e_k &= x_k - \hat{x}_k \\ &= x_k - \hat{x}_k^- - K_k(\Gamma x_k + v_k - \Gamma\hat{x}_k^-) \\ &= (I - K_k\Gamma)(x_k - \hat{x}_k^-) - K_k v_k. \end{aligned}$$

[1]See section VII-A for derivation.

Lemma 4.1: The gain function $\kappa_k(i)$ that minimizes the trace of the error covariance P_k satisfies the equation

$$\int_D \kappa(i') \left[R(i,i') + \gamma(i')P_k^-\gamma(i)^T\right] di = P_k^-\gamma(i)^T$$

Proof: First note that P_k may be expressed as[2]

$$\begin{aligned} P_k =& E\left[e_k e_k^T\right] \\ =& P_k^- - \int_D \kappa_k(i')\gamma(i') di' P_k^- - P_k^- \int_D \gamma^T(i)\kappa_k^T(i) di \\ &+ \int_D \kappa_k(i')\gamma(i') di' P_k^- \int_D \gamma^T(i)\kappa_k^T(i) di \\ &+ \int_D\int_D \kappa_k(i) R(i,i') \kappa_k(i') di di' \end{aligned} \tag{13}$$

We now minimize the trace of P_k by finding the functional extrema. We do this by calculating when the Gateaux derivative is equal to zero. The Gateaux derivative is given by [3]

$$\begin{aligned} \frac{dtr(P_k)}{dK} =& \int_D \phi(i)\, 2[-\gamma(i)P_k^- + \int_D \gamma(i)P_k^-\gamma^T(i')\kappa_k(i')^T di' \\ &+ \int_D R(i,i')\kappa_k(i')^T di'] di. \end{aligned} \tag{14}$$

We now set (14) to zero. Note that $\phi(i)$ is an arbitrary function and thus for the integral to be zero, the term in square brackets must be zero.

$$\begin{aligned} &\int_D \phi(i)\Big[-2\gamma(i)P_k^- + 2\int_D \gamma(i)P_k^-\gamma^T(i')\kappa_k(i')^T di' \\ &\quad + 2\int_D R(i,i')\kappa_k(i')^T di'\Big] di = 0 \\ \implies& -2\gamma(i)P_k^- + 2\int_D \gamma(i)P_k^-\gamma^T(i')\kappa_k(i')^T di' \\ &\quad + 2\int_D R(i,i')\kappa_k(i')^T di' = 0 \\ \implies& \int_D \kappa_k(i')\Big(R(i,i') + \gamma(i')P_k^-\gamma(i)^T\Big) di' = P_k^-\gamma(i)^T. \end{aligned} \tag{15}$$

■

Note that the combination of (15) and (13) yields[4]

$$P_k = (I - K_k\Gamma)P_k^-. \tag{16}$$

It is useful to check if this reduces to the standard Kalman filter if measurements are finite-dimensional. In this the measurement operator Γ is a matrix, and the Kalman gain acts like the matrix. The integral representation (15) can then be inverted to yield

$$K_k = \left(R + \Gamma P_k^-\Gamma^T\right)^{-1} P_k^-\Gamma^T.$$

This is the well-known Kalman gain for standard Kalman filters. Similar expressions to (15) can be found in previous works [10] [11] [12] [13] [14]. In these papers, (15) is solved by defining an operator inverse $(\cdot)^\dagger$ via the integral relationship

$$\int_D G(i,i')G^\dagger(i',i_1)di' = I\delta(i-i_1), \tag{17}$$

[2]See section VII-B for derivation.
[3]See section VII-D for derivation.
[4]See section VII-C for derivation.

where $\delta(\cdot)$ is the Dirac delta function. When this inverse operation is employed, the Kalman gain operator can be formulated as [10] [13] [14]

$$K_k = P_k^- \Gamma^T (R + \Gamma P_k^- \Gamma^T)^\dagger.$$

Alternative forms can also be found where the operator inverse is applied to the measurement noise covariance kernel $R(i,i')$, however this operator inverse is not guaranteed to exist and its existence is assumed. This operator inverse can also lead to results that are not easily implementable. As an example we examine the Ornstein-Uhlenbeck process, which is used to model many physical processes.

The Ornstein-Uhlenbeck process is zero mean, with covariance kernel $R(i,i') = Ie^{-|i-i'|}$. Taking the domain D to consist of the entire real number line $\mathbb{R}$, we use (17) to find

$$\begin{aligned}
&\int_D Ie^{-|i-i'|} R^\dagger(i', i_1) di' = I\delta(i - i_1) \\
&\implies Ie^{-|i|} * R^\dagger(i', i_1) = I\delta(i - i_1) \\
&\mathcal{F}\{Ie^{-|i|} * R^\dagger(i', i_1)\} = \mathcal{F}\{I\delta(i - i_1)\} \\
&\implies I\frac{1}{\omega^2 + 1}\bar{R}^\dagger = Ie^{j\omega i_1} \\
&\implies \bar{R}^\dagger = Ie^{j\omega i_1}\left(\omega^2 + 1\right) \\
&\implies R^\dagger = I[\delta(i - i_1) - \delta''(i - i_1)]
\end{aligned}$$

where $\delta(i)$ and $\delta''(i)$ denote the Dirac delta function and its second derivative respectively.

This introduces an implementation issue. The Dirac delta function is only well-defined under an integral, and must be handled carefully when implemented on real-world hardware. As we introduce the doublet, triplet[5] and higher derivatives it also causes difficulties in implementation. This may also be particularly troublesome for purposes of estimation and discretisation.

We propose an alternative formulation that gives an explicit definition of $\kappa(i)$, as well as potentially avoiding the issues relating to the implementation of generalized functions such as the Dirac delta function and its derivatives.

B. Explicit Solution for Optimal Kalman Gain

We will now show that if we impose the condition of stationarity on the infinite-dimensional observation noise, we can derive an alternative formulation that does not rely on the problematic operator inverse discussed above. This formulation of the optimal Kalman gain avoids some of the stated drawbacks of the inverse and leads to an explicit solution.

Theorem 4.1: The Kalman gain $\kappa(i)$ that satisfies (15) for a system (2)-(9) with stationary Gaussian Process measurement noise $v_k(i)$ on domain $D = \mathbb{R}$ is explicitly given by

$$\kappa_k(i) = P_k^- \left(I + \int_D f(i)\gamma(i)\, di P_k^-\right)^{-1} f(i).$$

Here $f(i) = \mathcal{F}^{-1}\{\bar{\gamma}^T(j\omega)\bar{R}(j\omega)^{-1}\}$,

where $\bar{\gamma}$ and $\bar{R}$ are the respective Fourier transforms of γ and R. $\mathcal{F}$ and $\mathcal{F}^{-1}$ denotes the Fourier and inverse Fourier transform respectively.

Proof: From Lemma 4.1,

$$\int_D \kappa_k(i')\left(R(i,i') + \gamma(i')P_k^-\gamma^T(i)\right) di' = P_k^-\gamma^T(i) \tag{18}$$

$$\implies \int_D \kappa_k(i') R(i,i')\, di' = \left(I - \int_D \kappa_k(i)\gamma(i)\, di\right) P_k^-\gamma^T(i)$$

$$\implies \int_D \kappa_k(i') R(i,i')\, di' = F(\kappa_k)\gamma^T(i)$$

$$\text{where } F(\kappa_k) = \left(I - \int_D \kappa_k(i)\gamma(i)\, di\right) P_k^-. \tag{19}$$

As the covariance kernel is stationary, and taking the domain D to be the entirety of $\mathbb{R}$, we can deduce that

$$\begin{aligned}
\int_D \kappa_k(i') R(i - i')\, di' &= F(\kappa_k)\gamma^T(i) \\
\implies (\kappa_k * R)(i) &= F(\kappa_k)\gamma^T(i) \\
\implies \mathcal{F}\{(\kappa_k * R)(i)\} &= \mathcal{F}\{F(\kappa_k)\gamma^T(i)\} \\
\implies \bar{\kappa}_k(j\omega)\bar{R}(j\omega) &= F(\kappa_k)\bar{\gamma}^T(j\omega) \\
\implies \bar{\kappa}_k(j\omega) &= F(\kappa_k)\bar{\gamma}^T(j\omega)\bar{R}(j\omega)^{-1} \\
\implies \kappa_k(i) &= F(\kappa_k)f(i).
\end{aligned} \tag{20}$$

Where $f(i) = \mathcal{F}^{-1}\{\bar{\gamma}^T(j\omega)\bar{R}(j\omega)^{-1}\}$ and $j = \sqrt{-1}$.

By substituting (20) into (19) we find

$$\begin{aligned}
F(\kappa_k) &= \left(I - F(\kappa_k)\int_D f(i)\gamma(i)\, di\right) P_k^- \\
\implies F(\kappa_k)&\left(I + \int_D f(i)\gamma(i)\, di P_k^-\right) = P_k^- \\
\implies F(\kappa_k) &= P_k^-\left(I + \int_D f(i)\gamma(i)\, di P_k^-\right)^{-1} \\
\implies \kappa_k(i) &= P_k^-\left(I + \int_D f(i)\gamma(i)\, di P_k^-\right)^{-1} f(i),
\end{aligned} \tag{21}$$

where the last equality follows from (20) again. ■

Combination of (10), (11), (12), (16), and (21) allows us to state our Kalman filter as follows:

Algorithm 1: Kalman Filter for Infinite-Dimensional Measurements

Input: $P_{k-1}, \hat{x}_{k-1}, z_k, S \triangleq \int_D f(i)\gamma(i)di$
$\hat{x}_k^- = A\hat{x}_{k-1}$
$P_k^- = AP_{k-1}A^T + Q$
$P_k = (I + P_k^- S)^{-1} P_k^-$
$\hat{x}_k =$
$\hat{x}_k^- + P_k^-\left(I + SP_k^-\right)^{-1}\int_D f(i)\left(z_k(i) - \gamma(i)\hat{x}_k^-\right) di$
Output: $P_k, \hat{x}_k$

Here, f is defined after (20). We remark that this explicit characterization arises from the stationarity of R and the assumption that $D = \mathbb{R}$. This yields a convolution that converts to a product in frequency domain, leading to (20). As in the standard Kalman filter, the covariance matrices are independent of the observations and can be pre-computed ahead of time. The matrix $S = \int_D f(i)\gamma(i)di$ may also be pre-computed. We have not solved all of the issues originally imposed by the operator inverse $(\cdot)^\dagger$. For example,

[5] First and second derivative of the Dirac delta function.

the inverse Fourier transform yielding f after (20) may not be well-defined if the measurement noise power spectrum $\bar{R}(j\omega)$ decays to zero faster than $\bar{\gamma}(j\omega)$ as $\omega \to \pm\infty$. In rough terms, this suggests that for a well-defined solution to exist, the effective bandwidth of the measurement kernel should be smaller than that of the noise covariance kernel.

The filter just derived is the optimal filter in the minimum mean square error sense. In the context of localization in mobile robotics, it operates on the pixel level i. In a real-world system it is not physically possible to perform computations on an infinite number of pixels, and so some approximations must be made. These approximations will apply to the integral $\int_D f(i)(z_k(i) - \gamma(i)\hat{x}_k^-)di$ in Algorithm 1. Indeed, any approximations of this integral can be viewed as a form of feature detection. By examining the continuum limit, where a high-dimensional and high-resolution noisy image or lidar scan is represented as a Gaussian process on a continuous domain, we have potentially gained the ability to perform feature detection in a principled manner via exploitation of the structure of $f(i)$.

V. Simulation Results

In this section, we present simulation results for the Kalman filter with infinite-dimensional measurements, as described in Algorithm 1. The system consists of a state vector $\mathbf{x}_k = [x_k, \dot{x}_k]^T$ representing the position and velocity of an agent. The motion model consists of a state transition model where Δ_t is the change in time from step to step and η is a velocity damping constant. A patterned line is fixed perpendicular to the observer's line of motion. The pattern of this line is represented by some function $C(y)$, which is then related to the observer's sensor via $\Gamma(i, x_k)$. Both the motion model and the observation model are affected by zero-mean additive Gaussian noise such that $E[w_k w_k^T] = Q$, $E[v_k(i)v_k(i')^T] = R(i, i')$.

The system equations are given by

$$\mathbf{x}_{k+1} = \begin{bmatrix} 1 & \Delta_t \\ 0 & 1-\eta \end{bmatrix} \begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix} + w_k$$
$$z_k(i) = -x_k \frac{i e^{-(\frac{i\bar{x}}{5F})^2}(25F\sin(\frac{i\bar{x}}{F}) + 2i\bar{x}\cos(\frac{i\bar{x}}{F}))}{25F^2} + v_k(i). \tag{22}$$

The observation equation is motivated by a simple pinhole camera model, as demonstrated in Figure 3, where a gray-scale intensity pattern

$$C(y) = e^{-(\frac{y}{5})^2}\cos(y) + 1. \tag{23}$$

is observed on a wall that is situated along the y-axis. In terms of the image domain i, we are then able to derive the non-linear observation function

$$\begin{aligned} z_k(i, x_k) &= C(\frac{i x_k}{F}) + v_k(i) \\ &= e^{-(\frac{i x_k}{5F})^2}\cos(\frac{i x_k}{F}) + 1 + v_k(i). \end{aligned}$$

We then linearize this function around some point $\bar{x}$ via a first order Taylor expansion. Neglecting the constant terms of this function, we arrive at (22).

TABLE I
SIMULATION PARAMETERS

System Variable	Notation	Value
State Matrix	A	$\begin{bmatrix} 1 & \Delta_t \\ 0 & 1-\eta \end{bmatrix}$
Process Noise Covariance	Q	$\begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_{\dot{x}}^2 \end{bmatrix}$
Observation Kernel	$\gamma(i)$	See (22)
State Vector	$\mathbf{x}_k$	$[x_k, \dot{x}_k]^T$
Initial State	$\mathbf{x}_0$	$[5, 1]^T$
Integral Domain	D	[-1,1]
Wall Pattern	$C(y)$	$e^{-\frac{y}{5}^2}\cos(y) + 1$
Measurement Covariance Kernel	$R(i, i')$	$\frac{\zeta}{\sqrt{2\pi l}} e^{-(\frac{i-i'}{\sqrt{2}\ell})^2}$
Initial Error Covariance	P_0	Q
Initial State Estimate	$\hat{\mathbf{x}}_0$	$\mathbf{x}_0$
Linearization Point	$\bar{x}$	5
Time Interval	Δ_t	1
Position Variance	σ_x^2	0.01
Velocity Variance	$\sigma_{\dot{x}}^2$	0.01
Measurement Noise Coefficient	ζ	0.01
Friction Coefficient	η	0.1
Focal Length	F	0.2
Length Scale	ℓ	0.02
Sample Spacing	Δ_s	0.002

Comparing (22) to (3)-(4) our gamma function is given by

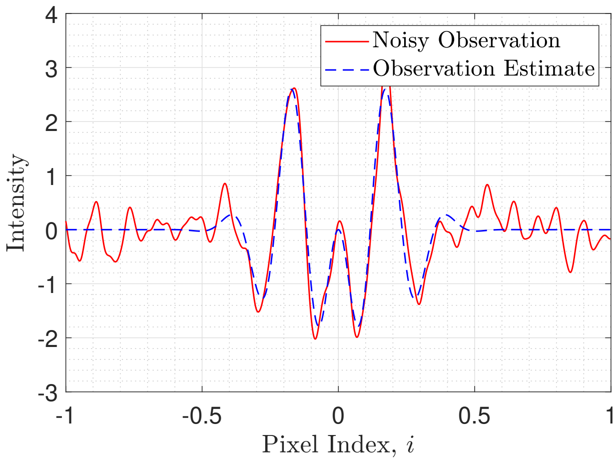
$$\gamma(i) = -\frac{i e^{-(\frac{i\bar{x}}{5F})^2}(25F\sin(\frac{i\bar{x}}{F}) + 2i\bar{x}\cos(\frac{i\bar{x}}{F}))}{25F^2}.$$

In this simulation the parameters given in Table I were used.

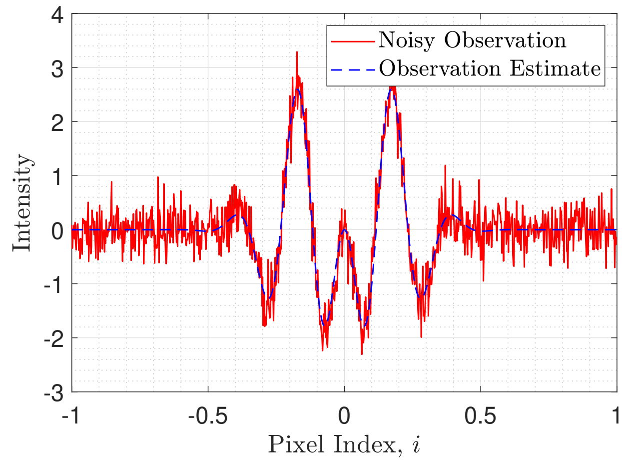
The process noise of the position and velocity are considered uncorrelated. The measurement noise is represented via a squared exponential Gaussian Process. Interestingly, given an exponential squared kernel $R(i, i') = e^{-(\frac{(i-i')^2}{2\ell^2})}$ we can calculate the kernel inverse defined in (17) as $\bar{R}^\dagger = I\frac{1}{\ell}e^{\frac{\ell^2\omega^2}{2}}$. This function does not admit an inverse Fourier transform and hence $R^\dagger$ does not exist in this case. However, if γ is such that $(\bar{\gamma}^T\bar{R})(j\omega)$ does admit a Fourier inverse, then we may calculate $f(i)$ and apply algorithm 1.

With these parameters, algorithm 1 was simulated using Matlab. It was necessary during implementation to discretise the system for numerical computation. Numerical functions were evaluated with spacings equal to Δ_s, or a frequency of 500 samples per unit of i. Algorithm 1 requires numerical integration for the posterior state update step, and this was implemented using the built-in trapz function. To allow numerical integration, a finite pixel domain $D = [-1, 1]$ was used. This is larger than the effective pixel-width of γ and much larger than the effective width l of the noise covariance R, and thus we expect little impact on the results obtained. Other methods of numerical integration may exist that exploit the structure of $f(i)$ and hence lead to more efficient computation. Analysis of these techniques and their levels of efficiency will be an important question for future work in this area.

As the filter is unbiased, the mean square error (MSE) can be analytically computed via the main diagonals of the error covariance matrix in steady state. As the simulated



(a) An observation with smooth noise ($\ell = 0.02$).



(b) An observation with coarse noise ($\ell = 0.001$).

Fig. 2. The length scale ℓ determines the extent to which the correlation between each pixel and its neighbors will decay. A large value of ℓ indicates that each pixel is highly correlated with the surrounding pixels, leading to smoother noise functions. A small value of ℓ indicates that the correlation between a pixel and its neighbors quickly decays with distance, leading to coarser noise functions.

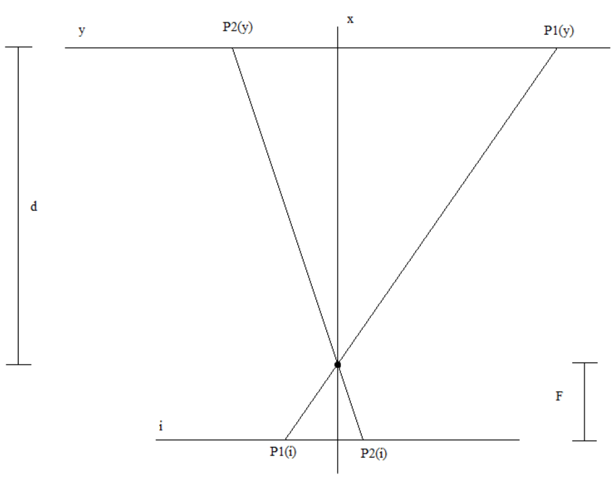


Fig. 3. A simple pinhole camera model, it is apparent from similar triangles that $\frac{x}{y} = -\frac{F}{i}$ and so $y = -\frac{xi}{F}$. F is equivalent to the focal length of the camera.

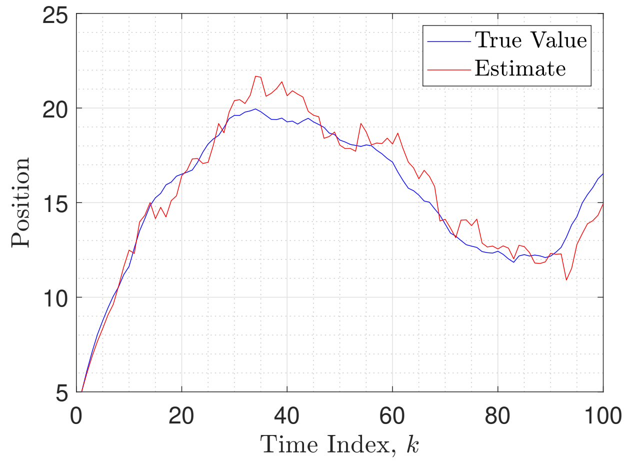
error covariance is updated, it will approach this steady state value. The first main diagonal corresponds to the MSE of the position while the second main diagonal corresponds to the MSE of the velocity. Figure 5 plots the mean square error of the system's state and compares this with the expected state MSE. This simulation was allowed to run for 5000 time steps to allow sufficient decay of transient errors. The asymptotic error covariance matrix for Figure 5 is

$$P_\infty = \begin{bmatrix} 1.137 & 0.108 \\ 0.108 & 0.040 \end{bmatrix}$$

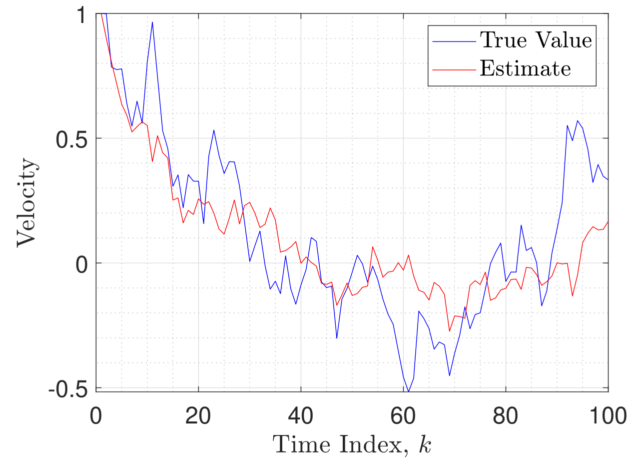
The asymptotic position MSE is then predicted to have a value of 1.137 while the asymptotic velocity MSE is predicted to have a value of 0.040.

VI. Conclusions

In this work we have presented a novel formulation of the Kalman filter for systems with finite-dimensional states and infinite-dimensional observations affected by stationary noise. This formulation avoids some of the implementation issues that previous formulations incur. We believe this filter will provide a rigorous foundation for localization tasks in systems with observations provided by high-dimensional sensors. An algorithm for this filter is derived and implemented in a simulated environment that is motivated by the pinhole camera model of vision. Future work in this area includes modeling other observation modalities such as radar, sonar, and lidar, all of which yield high-dimensional data. This work could also be generalized to non-linear systems by allowing the filter to update its linearized model of the system at each step and hence deriving an extended Kalman filter for systems with infinite-dimensional measurements.



(a) True position and estimated position.



(b) True velocity and estimated velocity.

Fig. 4. The system was simulated for 100 time steps. The true trajectory of the state was recorded alongside the filters trajectory estimate. Figure 4a presents the true and predicted position over time. Figure 4b presents the true and predicted velocity over time.

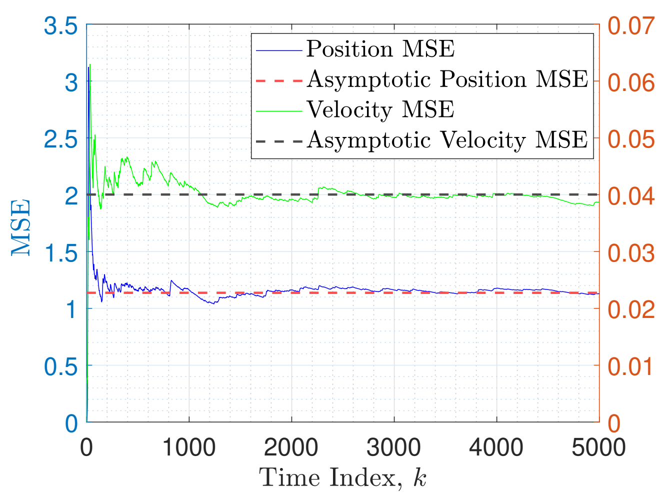


Fig. 5. Empirical and Asymptotic Mean Square Error. Note that the left axis denotes values relating to position, while the right axis denotes values relating to velocity.

VII. Appendix

A. Filter Error Trajectory

$$
\begin{aligned}
e_{k+1} &= x_{k+1} - \hat{x}_{k+1} \\
&= Ax_k + w_k - A\hat{x}_k - K_k(z_k - \hat{z}_k) \\
&= Ae_k - K_k(\Gamma x_k - \Gamma \hat{x}_k + v_k) + w_k \\
&= Ae_k - K_k(\Gamma e_k + v_k) + w_k \\
&= (A - K_k\Gamma)e_k - K_k v_k + w_k.
\end{aligned}
$$

B. Covariance Error Update

Let $e_k^- = x_k - \hat{x}_k^-$ noting that $e_k^- = Ae_{k-1} + w_{k-1}$ and hence $cov(e_k^-, e_k^-) = AP_{k-1}A^T + Q = P_k^-$.

$$
\begin{aligned}
P_k =& E\left[e_k e_k^T\right] \\
=& E\left[\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T\right] \\
=& E\left[\left(\left(I - K_k\Gamma\right)e_k^- - K_k v_k\right)\left(\left(I - K_k\Gamma\right)e_k^- - K_k v_k\right)^T\right] \\
=& \left(I - K_k\Gamma\right)E\left[e_k^-\left(e_k^-\right)^T\right]\left(I - K_k\Gamma\right)^T \\
&+ E\left[K_k v_k\left(i\right)\left(K_k v_k\left(i'\right)\right)^T\right] \\
=& \left(I - K_k\Gamma\right)P_k^-\left(I - K_k\Gamma\right)^T + E\left[K_k v_k\left(i\right)\left(K_k v_k\left(i'\right)\right)^T\right].
\end{aligned}
\tag{24}
$$

C. Simplified Covariance Update

Note that our covariance update is given by

$$P_k = \left(I - K_k\Gamma\right)P_k^-\left(I - K_k\Gamma\right)^T + KRK^T,$$

but our optimal Kalman gain K_k is given by

$$K\left(R + \Gamma P_k^- \Gamma^T\right) = P_k^- \Gamma^T$$

The combination of these allow

$$
\begin{aligned}
P_k &= P_k^- - K_k\Gamma P_k^- - P_k^-\Gamma^T K_k^T + K_k\Gamma P_k^-\Gamma^T K_k^T + K_k R K_k^T \\
&= P_k^- - K_k\Gamma P_k^- - P_k^-\Gamma^T K_k^T + P_k^-\Gamma^T K_k^T \\
&= (I - K_k\Gamma)P_k^-.
\end{aligned}
$$

D. Gateaux Derivative

Recall that the Gateaux derivative is given by (1). We can immediately see that for any function $\zeta(i)$

$$
\begin{aligned}
\frac{d}{dK}\int_D \kappa(i)\zeta(i)di &= \lim_{\epsilon\to 0}\int_D \left[\kappa(i) + \epsilon\phi(i)\right]\zeta(i) - \kappa(i)\zeta(i)di\epsilon^{-1} \\
&= \int_D \phi(i)\zeta(i)di.
\end{aligned}
$$

It is also readily apparent that

$$\begin{aligned}
&\frac{d}{dK}\int_D\int_D \kappa(i)\zeta(i)\kappa(i')^T di di' \\
&= \lim_{\epsilon\to 0}\int_D\int_D (\kappa(i)+\epsilon\phi(i))\zeta(i)(\kappa(i')+\epsilon\phi(i')^T) \\
&\quad - \kappa(i)\zeta(i)\kappa(i')^T di di' \epsilon^{-1} \\
&= \lim_{\epsilon\to 0}\int_D\int_D \phi(i)\zeta(i)\kappa(i')^T + \kappa(i)\zeta(i)\phi(i')^T \\
&\quad + \epsilon\phi(i)\zeta(i)\phi(i')^T di di' \\
&= \int_D\int_D \phi(i)\zeta(i)\kappa(i')^T + \kappa(i)\zeta(i)\phi(i')^T di di'
\end{aligned}$$

we can then see that the Gateaux derivative of the trace of (13) can be calculated as follows

$$\begin{aligned}
tr\left(P_k\right) =& tr\left(P_k^-\right) - 2tr\left(\int_D \kappa_k\left(i'\right)\gamma\left(i'\right) di di' P_k^-\right) \\
&+ tr\left(\int_D \kappa_k\left(i'\right)\gamma\left(i'\right) di' P_k^- \int_D \gamma^T\left(i\right)\kappa_k^T\left(i\right) di\right) \\
&+ tr\left(\int_D \kappa_k\left(i\right) R\left(i,i'\right)\kappa_k\left(i'\right) di di'\right).
\end{aligned}$$

Taking advantage of the cyclic permutation of the trace operator to find

$$\begin{aligned}
\frac{dtr\left(P_k\right)}{dK} =& -2\left[\int_D \phi\left(i\right)\gamma(i) di P_k^-\right] \\
&+ 2\left[\int_D\int_D \kappa_k\left(i'\right)\gamma(i')P_k^-\gamma(i)^T\phi\left(i\right) di di'\right] \\
&+ 2\left[\int_D\int_D \phi\left(i\right) R\left(i,i'\right)\kappa_k\left(i'\right) di di'\right] \\
=& \int_D \phi\left(i\right)\Big[-2\gamma(i)P_k^- + 2\int_D \gamma(i)P_k^-\gamma^T(i')\kappa_k(i')^T di' \\
&+ 2\int_D R(i,i')\kappa_k(i')^T di'\Big] di.
\end{aligned}$$

References

[1] Mark Maimone, Yang Cheng, and Larry Matthies. Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics*, 24(3):169–186, 2007.

[2] P. Newman, D. Cole, and K. Ho. Outdoor slam using visual appearance and laser ranging. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 1180–1187, 2006.

[3] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016.

[4] Raúl Mur-Artal and Juan D. Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

[5] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. Svo: Fast semi-direct monocular visual odometry. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 15–22, 2014.

[6] Robert Mahony and Jochen Trumpf. Equivariant filter design for kinematic systems on lie groups. *IFAC-PapersOnLine*, 54(9):253–260, 2021. 24th International Symposium on Mathematical Theory of Networks and Systems MTNS 2020.

[7] Pieter van Goor, Robert Mahony, Tarek Hamel, and Jochen Trumpf. Constructive observer design for visual simultaneous localisation and mapping. *Automatica*, 132:109803, 2021.

[8] John Baillieul. Perceptual control with large feature and actuator networks. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3819–3826, 2019.

[9] Han-Xiong Li and Chenkun Qi. Modeling of distributed parameter systems for applications—a synthesized review from time–space separation. *Journal of Process Control*, 20(8):891–901, 2010.

[10] J. S. Meditch. Least-squares filtering and smoothing for linear distributed parameter systems. *Automatica*, 7(3):315–322, May 1971.

[11] S.G. Tzafestas and J.M. Nightingale. Optimal filtering, smoothing and prediction in linear distributed-parameter systems. *Proceedings of the Institution of Electrical Engineers*, 115:1207–1212(5), August 1968.

[12] S. G. Tzafestas. Bayesian approach to distributed-parameter filtering and smoothing. *International Journal of Control*, 15(2):273–295, 1972.

[13] S. G. Tzafestas. On the distributed parameter least-squares state estimation theory. *International Journal of Systems Science*, 4(6):833–858, 1973.

[14] H. Nagamine, S. Omatu, and T. Soeda. The optimal filtering problem for a discrete-time distributed parameter system. *International Journal of Systems Science*, 10(7):735–749, 1979.

[15] Kirsten Morris. *Controller Design for Distributed Parameter Systems*. Springer, 2020.

[16] Kirsten A. Morris. Optimal output estimation for infinite-dimensional systems with disturbances. *Systems & Control Letters*, 146:104803, 2020.

[17] Peter Corke. *Robotics, Vision and Control: Fundamental Algorithms in MATLAB*. Springer, 1st edition, 2013.

[18] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[19] David G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., USA, 1st edition, 1997.