2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Asymptotically Optimal Finite-Dimensional Approximations for Linear Filtering with Infinite-Dimensional Measurements.

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

Abstract— **This work proposes a novel approach to approximate optimal linear filters for discrete-time linear Gaussian systems with infinite-dimensional measurements and finite-dimensional states. Assuming scalar-valued states for simplicity, we formulate the problem in terms of optimally selecting N points at which to sample the infinite-dimensional measurement, in order to minimize the mean-squared filtering error. We show that for large N, this problem can be expressed using the notion of an *asymptotic point density function* from the field of high-resolution quantization theory. To the best of the authors' knowledge, this method has not been considered in infinite-dimensional filtering previously. This leads to a characterization in terms of an Urysohn integral equation, which can be solved numerically to yield an asymptotically optimal N-point filter. The mean-squared approximation error is proportional to N^{-4}, which is faster than the typical N^{-2} decay of high-resolution quantization and suggests that this approximation method will be useful even for moderate or small N. These properties are verified by simulations based on a linearized pinhole camera measurement model.**

I. Introduction

State estimation in systems with low-dimensional states and high-dimensional observations is a relevant and ongoing challenge in a number of areas, such as autonomous navigation. In such systems, the state is often 6 degree-of-freedom pose data, while the measurements arise from high-resolution LIDAR scans or cameras [1], [2]. In recent work, an optimal linear filter has been developed for discrete-time Gaussian linear systems with infinite-dimensional measurements but finite-dimensional states [3]. However this filter involves performing an integral over the measurement domain during the online update, which can cause computational delays. For real-time applications it is preferable to deal with a finite number of measurements, raising the question of where to take these measurements.

In this paper, we propose an N-point approximation of our previously derived optimal filter [3] and demonstrate its asymptotic optimality for sufficiently large N, in terms of minimizing the mean squared estimation error. The feasibility of the approach is shown for scalar-valued states, with infinite-dimensional Gaussian measurements on a continuous domain (the generalization of this analysis to vector values is straightforward and will be presented in future work). Despite infinite-dimensional (or distributed-parameter) Kalman filtering having a long history (see [4]–[6] and references therein), the vast majority of past works have focused on situations in which the state and measurement spaces are both infinite-dimensional. Although there has been previous work on performing estimation based on a finite set of observable points, e.g. [7], none appear to explore the approximation of filters for our setting in which the state space is finite-dimensional but the measurement space is infinite-dimensional.

The key element of our approach is an N-point sampling density function defined over the measurement domain (Def. 3.5), a notion borrowed from the field of asymptotic quantization [8]. As N grows large this density function approaches a well-defined density function λ over the measurement domain. We then reformulate the minimization of the N-point filter MSE as a constrained optimization problem in terms of λ (Lemma 3.3). We show that the asymptotically optimal density function is the solution to an Urysohn-like integral equation [9, Section 14.1-2], which we solve numerically via successive approximations (Thm. 3.4, eq. (49)). In the scalar estimation case, this leads to a non-uniform sampling strategy which is asymptotically optimal. These properties are verified by simulations which suggest that the approach works well even for moderate N.

Optimal sensor placement for linear filters has been explored previously for finite and infinite-time horizons, under various cost functions. Although not globally optimal, a number of papers employ greedy algorithms which select the next optimal sensor to place given the previous sensors are optimally placed [10], [11], and analysis has been performed on theoretical performance guarantees on these sequential sensor placing algorithms [12]. The approximation filter proposed in this paper is globally optimal, although only in the asymptotic case. A surprising result of our approach is that the N-point mean-squared approximation error decays as N^{-4}, compared to the decay rate of N^{-2} which is commonly seen in high resolution scalar quantization [13, Section 5.6] and similar work in this area [14]. This suggests that fewer sample points are needed to achieve a feasible state estimation. Due to space limitations, the asymptotic analysis presented here is formal. The technical conditions required for convergence with large N are similar to asymptotic quantization [13], [8] and will be discussed in a longer version.

A property of the optimal asymptotic sampling density function λ is that it depends on the second and higher

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

M. M. Varley and G. N. Nair are with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC, 3010, Australia. (emails: varleym@student.unimelb.edu.au, gnair@unimelb.edu.au)

T. L. Molloy is with the CIICADA Lab, School of Engineering, Australian National University, Canberra, ACT 2601, Australia (e-mail: timothy.molloy@anu.edu.au)

derivatives of the measurement noise covariance and measurement function. This is interesting since, particularly in image processing, second derivatives often arise in heuristic algorithms used for edge detection [15], [16]. Though λ is currently computed off-line based on known system parameters, we envisage that our approach can be extended to form a principled basis for online feature detection.

The rest of this paper is structured as follows. Section II introduces necessary preliminary results and formulates the problem. Section III derives an asymptotically optimal N-point approximation based on the given criteria. Section IV simulates the derived N-point filter based on a linearized pinhole camera measurement model, and compares it with a filter that employs uniform sampling.

II. Preliminaries and Problem Formulation

A. Preliminaries

This article analyzes approximation schemes for a linear filter performing state estimation on a system with finite-dimensional states and infinite-dimensional observations. A derivation and discussion of this filter is given in [3]. A brief description of the system and the corresponding filter is now provided.

Let the discrete-time stochastic linear system (A, Γ) be described by the equations

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

$$z_k(i) = \gamma(i)x_k + v_k(i), \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the state, $w_k \in \mathbb{R}^n$ is the process noise, and $z_k(i) \in \mathbb{R}^m$ and $v_k(i) \in \mathbb{R}^m$ are the measurement and measurement noise, respectively, on the spatial domain $i \in \mathbb{R}^d$. Both the states and measurements are indexed by time step $k \in \mathbb{N}$, and $A \in \mathbb{R}^{n\times n}$ is the state transition matrix. The measurements are infinite-dimensional, and as such the linear observation function $\gamma(i) \in \mathbb{R}^{m\times n}$ for all $i \in \mathbb{R}^d$. The mapping function γ is assumed to be square and absolutely integrable over domain $\mathbb{R}^d$. The process and measurement noise are Gaussian with

$$\begin{aligned} E[w_k] = 0, \qquad & E\left[w_k w_k^\top\right] = Q \in \mathbb{R}^{n\times n} \\ E[v_k(i)] = 0,\ & E\left[v_k(i)v_k(i')^\top\right] = R(i-i')\ \forall i, i' \in \mathbb{R}^d. \end{aligned} \tag{3}$$

Note that the infinite-dimensional measurement noise is represented by a stationary Gaussian field. The definition and properties of Gaussian fields can be found in a number of texts [17], [18].

The form of the linear filter for system (A, Γ) is given by

$$\begin{aligned} \hat{x}_k^\infty &= A\hat{x}_{k-1}^\infty + \int_{\mathbb{R}^d} \kappa_k(i)\left(z_k(i) - \hat{z}_k(i)\right) di \\ \hat{z}_k(i) &= \gamma(i)A\hat{x}_{k-1}^\infty, \end{aligned} \tag{4}$$

where $\kappa_k : \mathbb{R}^m \to \mathbb{R}^n$ and represents the optimal gain, in the sense of minimizing the mean squared estimate error $E\left[\|x_k - \hat{x}_k^\infty\|^2\right]$ at each time step k. The procedure for deriving this optimal gain function is derived in [3] and the resulting Procedure 1 is presented in Appendix A. The optimal state estimate then takes the form

$$\hat{x}_k^\infty = A\hat{x}_{k-1}^\infty + P_k \int_{\mathbb{R}^d} f(i)(z_k(i) - \gamma(i)A\hat{x}_{k-1}^\infty)di, \tag{5}$$

where $f(\cdot) \triangleq \mathcal{F}^{-1}\{\mathcal{F}\{\gamma\}^\top \mathcal{F}\{R\}^{-1}\}$, P_k represents the covariance of the error $(x_k - \hat{x}_k^\infty)$, $\mathcal{F}\{\cdot\}$ denotes the d-dimensional Fourier transform, and $\mathcal{F}^{-1}\{\cdot\}$ denotes the corresponding inverse Fourier transform.

In the rest of this paper, to simplify our analysis we consider only scalar-valued states, measurements, and measurement domains.

B. Problem Formulation

In real-world digital systems, it is not possible for a sensor to sample continuously over the measurement domain in order to implement the integral in (5). Instead the sensor must perform measurements at discrete points within the domain. For this reason, linear filters of the form (4) are not implementable, and some approximation scheme must be employed. Assuming scalar states, let $\hat{x}^N$ be some approximation of $\hat{x}^\infty$, based on sampling the measurement domain at N sample points. Let the cost of a given state estimator $\hat{x}^N$ be given by the steady-state mean-squared error

$$J(\hat{x}^N) \triangleq \limsup_{k\to\infty} E\left[\left(x_k - \hat{x}_k^N\right)^2\right]. \tag{6}$$

In Section III, we will show that for sufficiently large N we may derive the asymptotically optimal N-point approximation for the system described thus far.

III. N-Point Filter Approximation

In this section we will derive an N-point approximation filter that solves (6) as N grows large. In Section III-A we will start with some useful definitions and reformulate the cost function to a form more amenable to analysis. In Section III-B we will derive an update rule for the covariance of the N-point filters error. In Section III-C we will derive approximations of this filter error for sufficiently large N, and introduce an idealized point density function. Finally, in Section III-D we will determine the necessary conditions for this point density function to minimize the cost function (6) and use the result to design a corresponding filter in Section IV.

A. Definitions and Cost Function Reformulation

Let $C = \{C_1, C_2, ..., C_{N+2}\}$ be a partition of the space $\mathbb{R}$ into $N + 2$ intervals, with each bounded interval C_j possessing an assigned length V_j and sample point $i_j \in C_j$. This partition will also have two unbounded (overload) intervals C_{N+1} and C_{N+2} which are assumed to be far enough to the left and right of the real axis to be negligible. We therefore restrict our analysis to the first N bounded (granular) intervals of C. Motivated by (5), we begin by proposing the following form for our N-point approximation.

Definition 3.1 (N-Point State Estimate): We define

$$\hat{x}_k^N \triangleq A\hat{x}_{k-1}^N + P_k \sum_{j=1}^{N} V_j f(i_j) \Big[z_k(i_j) - \gamma(i_j) A\hat{x}_{k-1}^N \Big], \tag{7}$$

as the N-point state estimate at time $k \geq 0$. This is simply approximating the integral operator by a Riemann sum with non-uniform partition intervals. In this work we assume that the partition is time-invariant and chosen off-line.

The error between the optimal filter estimate and the N-point approximation is then defined as follows.

Definition 3.2 (Error Term): The error associated with the N-point state estimate $\hat{x}_k^N$ is

$$e_k \triangleq \hat{x}_k^\infty - \hat{x}_k^N, \tag{8}$$

where $\hat{x}_k^\infty$ is given by (5) (see also line 8 in Procedure 1).

With these definitions, it can be shown that minimizing the cost function (6) is equivalent to reducing the "expected distance" between $\hat{x}_k^\infty$ and $\hat{x}_k^N$, i.e.

$$\arg\min_{\hat{x}^N} J(\hat{x}^N) = \arg\min_{\hat{x}^N} \limsup_{k\to\infty} E\left[e_k^2\right]. \tag{9}$$

B. Error Covariance Update Rule

We now will furnish the following definitions for clarity and ease of exposition.

Definition 3.3 (S Approximation and Error):

$$S \triangleq \int_{\mathcal{D}} s(i)\, di, \qquad \hat{S} \triangleq \sum_{j=1}^{N} V_j s(i_j) \tag{10}$$

$$s(i) \triangleq f(i)\gamma(i) \qquad \tilde{S} \triangleq S - \hat{S}. \tag{11}$$

Definition 3.4 (G Approximation and Error):

$$G_k \triangleq \int_{\mathcal{D}} f(i) v_k(i) di, \;\; \hat{G}_k \triangleq \sum_{j=1}^{N} V_j f(i_j) v_k(i_j) \tag{12}$$

$$\tilde{G}_k \triangleq G - \hat{G}_k. \tag{13}$$

Observe that with these definitions provided, we may reformulate the N-point state estimate as

$$\begin{aligned}\hat{x}_k^N =& A\hat{x}_{k-1}^N + P_k \sum_{j=1}^{N} V_j f(i_j) \Big[z_k(i_j) - \gamma(i_j) A\hat{x}_{k-1}^N \Big] \\ =& A\hat{x}_{k-1}^N + P_k \hat{S}(x_k - A\hat{x}_{k-1}^N) + \hat{G}_k \\ =& A\hat{x}_{k-1}^N + P_k \hat{S}(x_k - A\hat{x}_{k-1}^\infty + Ae_{k-1}) + \hat{G}_k. \end{aligned} \tag{14}$$

Substituting this formulation, along with (5), into (8), we find that with $M_k \triangleq (I - P_k\hat{S})A$,

$$\begin{aligned} e_k =& \hat{x}_k^\infty - \hat{x}_k^N \\ =& A\hat{x}_{k-1}^\infty + P_k S(x_k - A\hat{x}_{k-1}^\infty) + G_k \\ & - A\hat{x}_{k-1}^N - P_k\hat{S}(x_k - A\hat{x}_{k-1}^\infty + Ae_{k-1}) - \hat{G}_k \\ =& M_k e_{k-1} + P_k\tilde{S}(x_k - A\hat{x}_{k-1}^\infty) + \tilde{G}_k. \end{aligned} \tag{15}$$

Having formulated an update rule for the error, we may now derive an update rule for the error covariance $E[e_k^2]$. Substituting e_k from (15) we have

$$\begin{aligned} E\left[e_k^2\right] =& M_k^2 E\left[e_{k-1}^2\right] + P_k^2\tilde{S}^2 E\left[(x_k - A\hat{x}_{k-1}^\infty)^2\right] \\ & + E\left[\tilde{G}_k^2\right] + \textit{ cross terms}. \end{aligned} \tag{16}$$

The two cross terms involving $\tilde{G}_k$ are zero due to the measurement noise being uncorrelated with the approximation error and filter error terms. The remaining cross term is also zero as

$$\begin{aligned} E\left[(x_k - A\hat{x}_{k-1}^\infty)e_{k-1}\right] =& AE\left[(x_{k-1} - \hat{x}_{k-1}^\infty)e_{k-1}\right] \\ & + E\left[w_{k-1}e_{k-1}\right] \\ =& AE\left[(x_{k-1} - \hat{x}_{k-1}^\infty)e_{k-1}\right], \end{aligned} \tag{17}$$

as it can be shown that $E\left[(x_{k-1} - \hat{x}_{k-1}^\infty)e_{k-1}\right] = 0$. As in Procedure 1, we define the covariance $P_k^- = E\left[(x_k - A\hat{x}_{k-1}^\infty)^2\right]$. Making this final substitution to (16) and removing cross terms ultimately yields an error covariance update rule of

$$E[e_k^2] = M_k^2 E\left[e_{k-1}^2\right] + P_k^2\tilde{S}^2 P_k^- + E[\tilde{G}_k^2]. \tag{18}$$

C. Integral Approximations and Error Terms

We will now discuss how we may approximate the $\tilde{S}$ and $\tilde{G}_k$ terms for sufficiently large N, which will in turn lead us to an approximation of (18). Let $f : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable over $\mathbb{R}$. Then given an interval partition C of $\mathbb{R}$ and sufficiently large N,

$$\int_{\mathbb{R}} f(i)\, di \approx \sum_{j=1}^{N} V_j f(i_j), \quad i_j \in C_j, \tag{19}$$

where i_j is an arbitrary point in C_j. The error terms of this approximation are made explicit in the Lemma 3.1. To derive approximations, we require the following concept of an N-point density function.

Definition 3.5 (N-Point Density Function): Define the N-point density function $\lambda_N : \mathbb{R} \to [0,\infty)$, $\int_{\mathbb{R}} \lambda_N(i)\, di = 1$, by

$$\lambda_N(i) \triangleq \frac{1}{NV_j}, \quad \text{if } i \in C_j, \quad \text{for } j = 1, 2, ..., N, \tag{20}$$

where N is the total points sampled for the non-uniform Riemann sum approximation, V_j is the length of interval C_j.

This notion of a density function is a useful tool often used in vector quantization schemes [8], [13], [19]. When N grows very large, $\lambda_N(i)$ will approximate a density function $\lambda : \mathbb{R} \to [0,\infty)$ with $\int_{\mathbb{R}} \lambda(i)\, di = 1$ over the region of interest. For the analysis undertaken in this work, the sample points i_j^* will be the mid-point of each interval. We will not discuss the optimality of this selection here, but arguments for the mid-point minimizing the mean squared error distortion of a scalar quantizer have been presented [13, Section 6.2].

Lemma 3.1: For sufficiently large N, the error between the integral, if it exists, of a sufficiently smooth integrable function $f : \mathbb{R} \to \mathbb{R}$ and its mid-point non-uniformly partitioned Riemann sum is

$$\int_{\mathbb{R}} f(i) di - \sum_{j=1}^{N} V_j f(i_j^*) = \frac{1}{4!}\frac{1}{N^2}\int_{\mathbb{R}} \frac{d^2 f}{di^2}(i)\lambda(i)^{-2}\, di + O(N^{-4}) \tag{21}$$

where V_j is the length of the interval C_j, and unbounded intervals are considered negligible. The function $\lambda(i)$ denotes the density in the sense of Definition 3.5.

Proof: Observe that for an integrable, sufficiently smooth function $f : \mathbb{R} \to \mathbb{R}$ with an N-partition C and mid-points $i_j^* \in C_j$,

$$\int_{\mathbb{R}} f(i)di - \sum_{j=1}^{N} V(C_j)f(i_j^*) = \sum_{j=1}^{N} \int_{C_j} f(i) - f(i_j^*)di. \tag{22}$$

The Taylor polynomial of $f(i)$ around i_j^* is given by

$$f(i) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{d^n f}{di^n}\bigg|_{i_j^*} (i - i_j^*)^n. \tag{23}$$

It follows then that

$$\begin{aligned}\int_{C_j} f(i) - f(i_j^*)di &= \int_{C_j} \sum_{n=1}^{\infty} \frac{1}{n!} \frac{d^n f}{di^n}\bigg|_{i_j^*} (i - i_j^*)^n \, di \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} \frac{d^n f}{di^n}\bigg|_{i_j^*} \int_{C_j} (i - i_j^*)^n \, di. \end{aligned} \tag{24}$$

As i_j^* is the mid-point of C_j, $F_n(i) = \int (i - i_j^*)^n \, di$ is an even function around i_j^* when n is odd and an odd function around i_j^* when n is even, leading to cancellation of every odd term and doubling of every even term. This yields

$$\int_{C_j} f(i) - f(i_j^*)di = \sum_{n=1}^{\infty} \frac{1}{(2n)!} \frac{d^{2n} f}{di^{2n}}\bigg|_{i_j^*} \int_{C_j} (i - i_j^*)^{2n} \, di. \tag{25}$$

Let us now evaluate the integral $\int_{C_j} (i - i_j^*)^{2n} \, di$,

$$\begin{aligned}\int_{C_j} (i - i_j^*)^{2n} \, di =& (i - i_j^*)^{2n+1}\big|_{C_j} (2n+1)^{-1} \\ =& 2(2n+1)^{-1} \left(\frac{V_j}{2}\right)^{2n+1} \\ =& 2^{-2n}(2n+1)^{-1} V_j^{2n+1}. \end{aligned} \tag{26}$$

Combining this with (25) and (22), and substituting V_j according to Definition 3.5, for sufficiently large N,

$$\begin{aligned}\int_{\mathbb{R}} f(i)di - \sum_{j=1}^{N} V(C_j)f(i_j^*) &= \sum_{j=1}^{N} \sum_{n=1}^{\infty} \frac{2^{-2n}}{(2n+1)!} \frac{d^{2n} f}{di^{2n}}\bigg|_{i_j^*} V_j^{2n+1} \\ &\approx \sum_{n=1}^{\infty} \frac{2^{-2n}}{(2n+1)!} \int_{\mathbb{R}} \frac{d^{2n} f}{di^{2n}} \frac{1}{(\lambda(i)N)^{2n}} \, di \\ &= \frac{1}{4!} \frac{1}{N^2} \int_{\mathbb{R}} \frac{d^2 f}{di^2} \lambda(i)^{-2} \, di + O(N^{-4}). \end{aligned} \tag{27}$$

■

A result of Lemma 3.1 is that if s, as given by (11), is sufficiently smooth and N is sufficiently large,

$$\tilde{S} = \frac{1}{24} \frac{1}{N^2} \int_{\mathbb{R}} \frac{d^2 s}{di^2} \frac{1}{\lambda^2(i)} \, di + O(N^{-4}). \tag{28}$$

Having found an approximation for $\tilde{S}$, we analyze $\tilde{G}_k$.

Lemma 3.2: For sufficiently large N and sufficiently smooth function $g(i, m) = f(i)R(i - m)f(m)$,

$$E[\tilde{G}_k^2] \approx \frac{1}{576N^4} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\partial^4 g}{\partial i^2 \partial m^2} \frac{1}{\lambda^2(i)\lambda^2(m)} \, di \, dm. \tag{29}$$

Proof: Let $g(i, m) = f(i)R(i - m)f(m)$ with covariance $R(i, m) = E[v(i)v(m)]$, and observe that

$$\begin{aligned} E[\tilde{G}_k^2] &= E\left[\left(\sum_{j=1}^{N} \int_{C_j} (f(i)v_k(i) - f(i_j^*)v_k(i_j^*))di\right)^2\right] \\ &= \sum_{j=1}^{N} \sum_{n=1}^{N} \int_{C_j} \int_{C_n} \Big(f(i)E[v_k(i)v_k(m)]f(m) \\ &\quad - f(i_j^*)E[v_k(i_j^*)v_k(m)]f(m) - f(i)E[v_k(i)v_k(m_n^*)]f(m_n^*) \\ &\quad + f(i_j^*)E[v_k(i_j^*)v_k(m_n^*)]f(m_n^*)\Big) dmdi \\ &= \sum_{j=1}^{N} \sum_{n=1}^{N} \int_{C_j} \int_{C_n} [g(i, m) - g(i_j^*, m) \\ &\quad - g(i, m^*) + g(i_j^*, m_n^*)] \, dm \, di. \end{aligned} \tag{30}$$

Expressing $g(i, m)$ and the $g(i_j^*, m), g(i, m_n^*)$ terms as their Taylor series around (i_j^*, m_m^*) we have

$$\begin{aligned} & g(i, m) - g(i_j^*, m) - g(i, m_n^*) + g(i_j^*, m_n^*) \\ =& \sum_{\beta=0}^{\infty} \sum_{\alpha=0}^{\infty} \frac{\partial^{\alpha+\beta} g}{\partial i^\alpha \partial m^\beta}\bigg|_{(i_j^*, m_n^*)} (i - i_j^*)^\alpha (m - m_n^*)^\beta \frac{1}{\alpha!\beta!} \\ & - \sum_{\alpha=0}^{\infty} \frac{\partial^\alpha g}{\partial i^\alpha}\bigg|_{(i_j^*, m_n^*)} (i - i_j^*)^\alpha \frac{1}{\alpha!} - \sum_{\beta=0}^{\infty} \frac{\partial^\beta g}{\partial m^\beta}\bigg|_{(i_j^*, m_n^*)} (m - m_n^*)^\beta \frac{1}{\beta!} \\ & + g(i_j^*, m_n^*) \\ =& \sum_{\beta=1}^{\infty} \sum_{\alpha=1}^{\infty} \frac{\partial^{\alpha+\beta} g}{\partial i^\alpha \partial m^\beta}\bigg|_{(i_j^*, m_n^*)} (i - i_j^*)^\alpha (m - m_n^*)^\beta \frac{1}{\alpha!\beta!}. \end{aligned} \tag{31}$$

If i_j^*, m_n^* are both the mid-point of C_j, C_n respectively, then the integral $\int_{C_j} (i - i_j^*)^\alpha di$ operates on an even function around i_j^* when α is even and an odd function when α is odd. The same argument may be made for $\int_{C_j} (m - m_n^*)^\beta$, leading to cancellation when either α or β is odd, and doubling when α and β are , so the double integral of (31) is

$$\begin{aligned} & \sum_{\beta=1}^{\infty} \sum_{\alpha=1}^{\infty} \frac{1}{(2\alpha)!(2\beta)!} \frac{\partial^{2\alpha+2\beta} g}{\partial i^{2\alpha} \partial m^{2\beta}}\bigg|_{(i_j^*, m_n^*)} \\ & \cdot \int_{C_j} (i - i_j^*)^{2\alpha} di \int_{C_n} (m - m_n^*)^{2\beta} dm. \end{aligned} \tag{32}$$

Taking the fourth order Taylor series and truncating yields

$$\begin{aligned} & \int_{C_j} \int_{C_n} g(i, m) - g(i_j^*, m) - g(i, m_n^*) + g(i_j^*, m_n^*) didm \\ \approx& \frac{1}{4} \frac{\partial^4 g}{\partial i^2 \partial m^2}\bigg|_{(i_j^*, m_n^*)} \int_{C_j} (i - i_j^*)^2 di \int_{C_n} (m - m_n^*)^2 dm \\ =& \frac{1}{4} \frac{1}{9} \frac{\partial^4 g}{\partial i^2 \partial m^2}\bigg|_{(i_j^*, m_n^*)} (i - i_j^*)^3\bigg|_{C_j} (m - m_n^*)^3\bigg|_{C_n} \\ =& \frac{1}{36} \frac{\partial^4 g}{\partial i^2 \partial m^2}\bigg|_{(i_j^*, m_n^*)} 2\left(\frac{V_j}{2}\right)^3 2\left(\frac{V_n}{2}\right)^3 \\ =& \frac{1}{576} \frac{\partial^4 g}{\partial i^2 \partial m^2}\bigg|_{(i_j^*, m_n^*)} V_j^3 V_n^3. \end{aligned} \tag{33}$$

Substituting the density function from Definition 3.5 into (33) leads to the following approximation for large N

$$E[\tilde{G}_k^2] \approx \frac{1}{576}\sum_{j=1}^{N}\sum_{m=1}^{N}\left.\frac{\partial^4 g}{\partial i^2 \partial m^2}\right|_{(i_j^*, m_n^*)} V_j^2 V_j V_n^2 V_n$$
$$\approx \frac{1}{576}\frac{1}{N^4}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\partial^4 g}{\partial i^2 \partial m^2}\frac{1}{\lambda^2(i)\lambda^2(m)}\,di\,dm. \quad (34)$$

■

D. Solution via Density Function

Having derived appropriate approximations for $\tilde{S}, E[\tilde{G}_k^2]$ as N grows large, we approximate the asymptotic error $E[e_k^2]$ as N grows large. This is achieved via the following lemma.

Lemma 3.3: Let $\psi_k = N^4 E[e_k^2]$. As $N \to \infty$,

$$\psi_\infty \to \frac{1}{1 - M_\infty^2}\frac{1}{576}\left(\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\eta(i,m)}{\lambda(i)^2\lambda(m)^2}\,di\,dm\right), \quad (35)$$

where $\eta(i,m) = P_\infty^2 P_\infty^- \frac{d^2 s(i)}{di^2}\frac{d^2 s(m)}{dm^2} + \frac{\partial^4 g(i,m)}{\partial i^2 \partial m^2}$. (36)

Proof: Set $\psi_k = N^4 E[e_k^2]$, substitute into (18), and apply approximations (28), (33), to find that as $N \to \infty$,

$$\psi_k = M_k^2 \psi_{k-1} + N^4 P_k^2 P_k^- \tilde{S}^2 + N^4 E[\tilde{G}_k^2]$$
$$\to M_k^2 \psi_{k-1}^2 + P_k^2 P_k^- \frac{1}{576}\left(\int_{\mathbb{R}}\frac{d^2 s}{di^2}\lambda(i)^{-2} di\right)^2$$
$$+ \frac{1}{576}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\partial^4 g}{\partial i^2 \partial m^2}\lambda(i)^{-2}\lambda(m)^{-2}\,di\,dm, \quad (37)$$

where $M_k \to (1 - P_k S)A$ as $N \to \infty$, from (10) and the definition below (14). By rewriting the squared integral on the RHS as a double integral, we derive the asymptotic value

$$\psi_\infty = M_\infty^2 \psi_\infty + \frac{P_\infty^2 P_\infty^-}{576}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{d^2 s(i)}{di^2}\frac{1}{\lambda(i)^2}\frac{d^2 s(m)}{dm^2}\frac{1}{\lambda(m)^2}di\,dm + \frac{1}{576}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\partial^4 g}{\partial i^2 \partial m^2}\lambda(i)^{-2}\lambda(m)^{-2}\,di\,dm$$
$$\psi_\infty = \frac{1}{576}\frac{1}{1 - M_\infty^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\eta(i,m)}{\lambda(i)^2\lambda(m)^2}\,di\,dm, \quad (38)$$

where the kernel function $\eta(i,m)$ is given by

$$\eta(i,m) = \left(P_\infty^2 P_\infty^- \frac{d^2 s(i)}{di^2}\cdot\frac{d^2 s(m)}{dm^2} + \frac{\partial^4 g(i,m)}{\partial i^2 \partial m^2}\right) \quad (39)$$

and $M_\infty = (1 - P_\infty S)A$, which has magnitude < 1 under mean-squared stability of the infinite-dimensional filter (5).

■

Having derived the value ψ_∞ in terms of a density function, we now find the minimizing point density function.

Theorem 3.4 (Optimal Density Function Conditions): The point density function λ that minimizes the asymptotic MSE of the N-point filter must satisfy the following necessary condition

$$\lambda(i)^3 = \frac{4}{\omega}\int_{\mathbb{R}}\eta(i,m)\lambda(m)^{-2}\,dm, \quad (40)$$

where ω is a normalization constant ensuring $\int_{\mathbb{R}}\lambda(i)\,di = 1$, and the kernel $\eta(i,m)$ is given by Lemma 3.3.

Proof:

$$\arg\min_\lambda E[e_\infty^2] = \arg\min_\lambda \psi_\infty \quad (41)$$
$$= \arg\min_\lambda \int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\eta(i,m)}{\lambda(i)^2\lambda(m)^2}\,di\,dm. \quad (42)$$

Note that λ is a density function and so we impose non-negativity and normalization constraints $\lambda(i) \geq 0\ \forall i$ and $\int_{\mathbb{R}}\lambda(i)\,di = 1$. This constrained minimization problem yields the Lagrangian [20, Section 9.4]

$$L(\lambda,\omega) = \int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\eta(i,m)}{\lambda(i)^2\lambda(m)^2}\,di\,dm + \omega\left(\int_{\mathbb{R}}\lambda(i)\,di - 1\right) - \int_{\mathbb{R}}\mu(i)\lambda(i)\,di. \quad (43)$$

A necessary condition for minimization is then given by the condition $\frac{\partial}{\partial\lambda}L = 0$, this can be shown to be equivalent to the condition

$$\lambda(i)^3 = \frac{4}{\omega}\int_{\mathbb{R}}\eta(i,m)\lambda(m)^{-2}\,dm. \quad (44)$$

■

Equation (44) is an Urysohn equation of the second kind. Sufficient conditions for existence and uniqueness of solutions to such equations have been studied in [9, Section 14.1-2], and will be explored further in our future work. For our current purposes, a solution can be found numerically and we illustrate the application of successive approximation[9, Section 14.3-4] to solve (44) in Section IV. A normalized solution to (44) yields our desired point density function. Once an asymptotically optimal point density function λ is found, an N-point approximation filter may be designed in two steps. First, the cumulative sum Λ of this point density function is calculated. This sum represents the expected proportion of points observed as we sweep through the image domain, analogous to a cumulative distribution function. Then, for an N-point filter, we simply acquire the values $\{\Lambda^{-1}(\frac{1}{N}), \Lambda^{-1}(\frac{2}{N}), .., \Lambda^{-1}(1)\}$, corresponding to our partition's boundaries. This is the companding model of quantization [13, Section 5.5]. Having determined the appropriate partition scheme for our system, appropriate substitutions are made to (7) and the filter may be implemented.

IV. Simulation Results

In this section, Procedure 1 is applied in simulations, with line 8 (cf. (5)) modified to the N-point form given in (7).

We consider a system governed by the equations,

$$x_{k+1} = a x_k + w_k, \quad (45)$$
$$z_k(i) = -e^{-(\nu L_f^{-1}\bar{x}i)^2}\left(2(\nu L_f^{-1}i)^2\bar{x}\cos(\xi\bar{x}L_f^{-1}i) + \xi L_f^{-1}i\sin(\xi\bar{x}L_f^{-1}i)\right)x_k + v_k(i). \quad (46)$$

The observation equation is motivated by a linearization of the pinhole camera model, a diagram of this model is presented in [3, Fig. 3]. The linearization is of the pattern

$$C(p) = e^{-\nu^2 p^2}\cos(\xi p) + 1. \quad (47)$$

This pattern is observed via the pinhole camera model of vision, resulting in a non-linear observation function

$$\begin{aligned}\bar{z}_k(i, x_k) =& C(ix_k L_f^{-1}) \\ =& e^{-\nu^2 (ix_k L_f^{-1})^2} \cos\left(\xi(ix_k L_f^{-1})\right) + 1. \end{aligned} \quad (48)$$

We linearise this function with respect to x_k around a linearization point $\bar{x}$ and apply additive Gaussian measurement noise $v_k(i)$ to arrive at (46). The asymptotic optimal density function for this system is calculated numerically via successive approximation with the following update equation

$$h_{t+1}(i) = \left(\int_a^b \eta(i, m) h_t(m)^{\frac{-1}{3}} \, dm \right)^2, \quad (49)$$

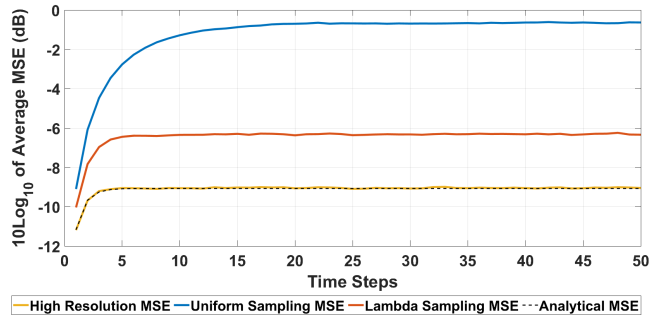
which has a fixed point solution equivalent to condition (44) with $h(i) = \left(-\frac{4}{\omega}\right)^{\frac{6}{5}} \lambda(i)^6$. This form is used to remove explicit dependence on the unknown normalization parameter ω and also to ensure a non-negative solution.

A candidate $h_0(i)$ is chosen and defined over an appropriate domain. The bounds are chosen such that any values outside of $[a, b]$ are negligible. The $\eta(i, m)$ function is derived from the system via (39). For the system considered in this paper this procedure converges on a fixed point. Conditions to ensure convergence, existence and uniqueness of this solution will be an important area of future work. The fixed point solution $h(i)$ can be used to calculate $\omega^{\frac{1}{7}} \lambda(i) = h(i)^{\frac{1}{8}}$, with ω selected to ensure $\int_{\mathbb{R}} \lambda(i) \, di = 1$. As discussed in Section IV, the companding model is used to select an appropriate partition of the observation domain. System parameters are given in Table I.

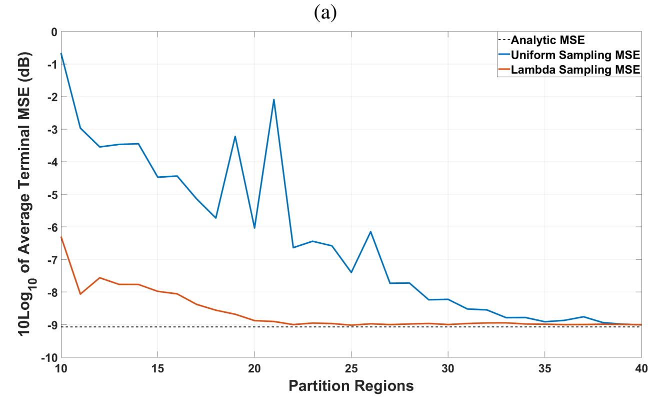
Three filters are implemented and compared. The high-resolution filter samples the observations uniformly with a spacing of Δ_s corresponding to 200 samples. The uniform-sampling filter samples the observations uniformly with 10 samples. The lambda-sampling filter samples the observations according to our companding model with 10 samples. The results are shown in Fig. 1a, alongside the analytically calculated optimal MSE. We also analyze the performance

TABLE I: Simulation Parameters

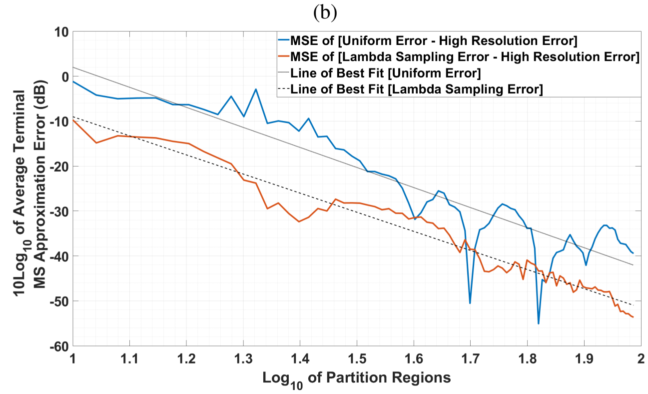
System Variable	Notation	Value
State Transition	A	0.9
Process Noise Covariance	Q	0.1^2
Observation Kernel	$\gamma(i)$	See (46)
Initial State	$\mathbf{x}_0$	100
Integral Domain	D	[-1,1]
Wall Pattern	$C(p)$	$e^{-\nu^2 p^2} \cos(\xi p) + 1$
Measurement Covariance Kernel	$R(i, i')$	$\frac{\rho}{\sqrt{2\pi}\ell} e^{-\left(\frac{i-i'}{\sqrt{2}\ell}\right)^2}$
Initial Error Covariance	P_0	Q
Initial State Estimate	$\hat{\mathbf{x}}_0$	$\mathbf{x}_0$
Linearization Point	$\bar{x}$	1
Focal Length	L_f	0.01
Decay Parameter	ν	0.05
Frequency Parameter	ξ	0.8
Length Scale	ℓ	0.04
Observation Noise Intensity	ρ	1
Sample Spacing	Δ_s	0.01



(a)



(b)



(c)

Fig. 1: Simulation Results: (a) MSE (dB scale) of various filters averaged over 50,000 trials. The uniform-sampling filter and lambda-sampling filter sample 10 points in the measurement domain. (b) Average terminal value of MSE (dB scale) for uniform and lambda sampling schemes as the interval number N increases. (c) Average terminal value of MS of approximation errors (dB) for uniform and lambda sampling schemes. Upper and lower lines of best fit have gradients -44.6 and -42.5 dB/decade respectively.

of these filters as the number of intervals increase. Fig. 1b presents the MSE's of both the lambda-sampling filter and the uniform-sampling filter as the number of intervals increase. As expected, both approach the theoretically optimal MSE as the number of intervals grow large. The MSE of the approximation errors (differences between the high resolution and the uniform/lambda-sampling filters) are plotted in Fig. 1c. We see a noisy but noticeable trend of a -40dB/decade roll-off, which aligns with the theoretical decay rate of N^{-4} predicted by Lemma 3.3.

V. Conclusion

We derive a finite-dimensional approximation for an optimal linear filter with finite-dimensional states and infinite-dimensional measurements. The approximation is converted to an optimization problem in terms of minimizing the mean-squared estimation error. The derived filter samples N discrete measurements across the measurement domain and has a mean-squared approximation error that decays like N^{-4} as opposed to the N^{-2} decay rate commonly seen in high-resolution quantization schemes. The filter is implemented in simulation and these properties are verified and compared with a uniform-sampling filter. This work is constrained to scalar-valued states, measurements and measurement domains, and future work will extend the analysis to the vector case. Future work will also examine conditions for the uniqueness and existence of an asymptotically optimal density function, as well as optimal time-varying and measurement-dependent sampling strategies.

Appendix

A. Optimal Linear Filter Procedure

Procedure 1 Optimal Linear Filter [3]

Inputs: $A, Q, R(i, i'), P_0, \hat{x}_0^\infty, \gamma(i)$
for $k \geq 1$ **do**
    $f(i) = \mathcal{F}^{-1}\{\mathcal{F}\{\gamma(i)\}^\top \mathcal{F}\{R(i)\}^{-1}\}$
    $S = \int_{\mathbb{R}^d} f(i)\gamma(i)di$
    $P_k^- = AP_{k-1}A^\top + Q$
    $P_k = P_k^-(I + SP_k^-)^{-1}$
    Obtain measurement - $z_k(i)$
    $\hat{x}_k^\infty = A\hat{x}_{k-1}^\infty + P_k \int_{\mathbb{R}^d} f(i)(z_k(i) - \gamma(i)A\hat{x}_{k-1})di$
end for
Outputs: $\{P_1, P_2, ..., P_T\}, \{\hat{x}_1^\infty, \hat{x}_2^\infty, ..., \hat{x}_T^\infty\}$

Note that $P_k, P_k^- \in \mathbb{R}^{n\times n}$ represent the covariances of $x_k - \hat{x}_k^\infty$ and $x_k - A\hat{x}_{k-1}^\infty$ respectively. The $\mathcal{F}(\cdot)$ operator denotes the d-dimensional Fourier transform.

References

[1] G. Gallego, J. E. Lund, E. Mueggler, H. Rebecq, T. Delbruck, and D. Scaramuzza, "Event-Based, 6-DOF Camera Tracking from Photometric Depth Maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 10, pp. 2402–2412, 2018.

[2] E. Mueggler, N. Baumli, F. Fontana, and D. Scaramuzza, "Towards evasive maneuvers with quadrotors using dynamic vision sensors," in *2015 European Conference on Mobile Robots (ECMR)*, pp. 1–8, 2015.

[3] M. M. Varley, T. L. Molloy, and G. N. Nair, "Kalman filtering for discrete-time linear systems with infinite-dimensional observations," in *2022 American Control Conference (ACC)*, pp. 296–303, 2022.

[4] P. L. Falb, "Infinite-dimensional filtering: The Kalman-Bucy filter in Hilbert space," *Information and Control*, vol. 11, no. 1, pp. 102–137, 1967.

[5] R. Curtain, "A survey of infinite-dimensional filtering," *SIAM Review*, vol. 17, no. 3, pp. 395–411, 1975.

[6] K. Morris, *Controller Design for Distributed Parameter Systems*. Springer Cham, 01 2020.

[7] M. Todescato, A. Carron, R. Carli, G. Pillonetto, and L. Schenato, "Efficient spatio-temporal Gaussian regression via Kalman filtering," *Automatica*, vol. 118, p. 109032, 2020.

[8] A. Gersho, "Asymptotically optimal block quantization," *IEEE Transactions on Information Theory*, vol. 25, no. 4, pp. 373–380, 1979.

[9] P. Polyanin and A. Manzhirov, *Handbook of Integral Equations: Second Edition*. Handbooks of mathematical equations, CRC Press, 2008.

[10] E. M. Hernandez, "Efficient sensor placement for state estimation in structural dynamics," *Mechanical Systems and Signal Processing*, vol. 85, pp. 789–800, 2017.

[11] C. Zhang and Y. Xu, "Optimal multi-type sensor placement for response and excitation reconstruction," *Journal of Sound and Vibration*, vol. 360, pp. 112–128, 2016.

[12] L. F. O. Chamon, G. J. Pappas, and A. Ribeiro, "Approximate supermodularity of Kalman filter sensor selection," *IEEE Transactions on Automatic Control*, vol. 66, pp. 49–63, 2019.

[13] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Kluwer Academic Publishers, 1992.

[14] A. Aalto, "Spatial discretization error in Kalman filtering for discrete-time infinite dimensional systems," *IMA Journal of Mathematical Control and Information*, vol. 35, pp. i51–i72, 04 2017.

[15] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[16] R. M. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 1, pp. 58–68, 1984.

[17] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Adaptive computation and machine learning, MIT Press, 2006.

[18] R. J. Adler, *The Geometry of Random Fields*. Society for Industrial and Applied Mathematics, 2010.

[19] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[20] D. G. Luenberger, *Optimization by Vector Space Methods*. USA: John Wiley & Sons, Inc., 1st ed., 1997.