Probabilistic Methods for Real-time Unsupervised Anomalous Trajectory Detection

Thinh Hoang Dinh, *Student Member, IEEE,* Vincent Martinez, Pierre Maréchal, Daniel Delahaye

Abstract—Anomaly detection plays a crucial role in numerous Intelligent Transportation Systems (ITS) applications, as unusual or anomalous behaviors often carry significant safety implications. Particularly in road traffic, vehicle trajectories are represented in time-series data, necessitating real-time detection and requiring the immediate identification and alert of anomalous trajectories once enough evidence to substantiate the possibility emerges.

This paper presents a probabilistic framework including likelihood and Bayesian methods to enable real-time hypothesis testing for anomaly detection as observations accumulate. Operating in an unsupervised manner, the framework eliminates the need for explicit anomaly modeling. The time series data undergoes dimensional reduction through Functional Principal Component Analysis (FPCA) and is then fitted to obtain a prior distribution of FPCA scores. After this, the likelihood of time series observations stemming from this prior distribution is calculated in real-time.

We conducted a series of numerical experiments to demonstrate the effectiveness and responsiveness of the proposed methods in detecting common road hazards such as wrong-way driving, overspeeding, and sudden hard-braking. The results showed successful detection of all presented anomalies and a rapid response rate, ranging from 40ms to 160ms delay. The framework produced a significantly lower rate of false alarms compared to the Local Outlier Factor (LOF) method. Additionally, the framework is capable of calibrating to accommodate different sensor types and capabilities. It also demands low computational resources, making it readily implementable across various embedded Advanced Driver Assistance Systems (ADAS) platforms.

Index Terms—anomaly-detection, data-driven, unsupervised, dimensionality-reduction, ADAS.

I. INTRODUCTION

With the roll-outs of radar and V2X equipped vehicles on the market, much traffic information is expected to be captured. It is presumed that this information, in addition to many other already available information sources such as security cameras and traffic crowd-sourcing applications, can deliver more significant benefits not only for traffic authorities but also for road participants thanks to the popularity of ADAS and autonomous driving systems [1] on newer vehicles. However, the availability of data does not necessarily translate to

Thinh Hoang Dinh is with Artificial and Natural Intelligence Toulouse Institute (ANITI), Université de Toulouse, Toulouse, France.

Vincent Martinez is with NXP Semiconductors Toulouse, Toulouse, France.

Pierre Maréchal is with Toulouse Institute of Mathematics, Toulouse, France.

Daniel Delahaye is with École Nationale de l'Aviation Civile, Toulouse, France.

better safety because crucial information is often buried under an overabundance of irrelevant data. It is typically observed that accidents usually follow abnormal traffic behavior. Hence identifying these situations are crucial to the success of any safety system.

Anomaly detection, a matured field of research, boasts a diversity of methodologies suitable for univariate and multivariate time series [2]. These strategies range from simple parametric models [3] to comprehensive deep learning frameworks [4], spanning supervised [5], [6], unsupervised [7], [8], and hybrid learning approaches. Such detection techniques are fundamental to ITS applications, including driver alert systems, collision avoidance, and autonomous driving. Very often, an anomaly detector consist of many distinct detectors for each road hazard type, such as wrong-way driving [9], [10], overspeeding [11], and aberrant braking [12] detectors. Alternatively, unsupervised machine learning techniques have facilitated the creation of a new class of algorithms, which can identify abnormal behaviors based solely on training examples. These techniques have the advantage of being highly adaptable, which may be important for rapid adaptation to different road scenarios.

These methods typically construct a representation of normal behavior trajectories through a dictionary [13], a parametric distribution, or histogram bins [14]. Many such approaches can achieve state-of-the-art online performance; however, they often demand significant computational resources, a requirement typically unmet by onboard embedded platforms. Additionally, deep learning techniques also necessitate large volumes of training data, and their opaque nature inhibits the derivation of analytical insights that might inform performance guarantees, such as the system's adaptability to sensors of differing quality. Furthermore, these techniques often assume the complete availability of trajectories (e.g., from traffic security cameras), yet onboard sensors typically yield fragmented trajectories due to frequent occlusions by other road objects. Additionally, safety-oriented systems must operate under strict time constraints to allow for the identification of escape routes, which may be executed by the driver or an autonomous driving system.

Such considerations have motivated this paper to introduce an anomaly detection framework that is capable of real-time trajectory classification, computationally efficient, adaptable, and simple that enables analyzing of performance guarantees and accommodating various sensor types and capabilities. The primary contributions of this paper are as follows:

1) We introduce an unsupervised anomalous trajectory detection framework, with the trajectories represented in

FPCA scores, and two methods-likelihood-based and Bayesian-based, to perform classification of abnormality of the time series in real-time, as data arrives. The algorithms are recursive, so they are highly efficient with memory consumption and computational complexity.

2) We perform thorough analytical analysis of the mathematical properties of this framework to reveal its characteristics, types of trajectory anomalies suitable for detection.

3) We demonstrate that many common road hazards such as wrong-way driving, overspeed and abarrent braking can be detected using the same framework.

The organization of this paper is as follows: Section II offers a comprehensive review of the existing literature on anomalous trajectory detection. Section III introduces essential background information on FPCA. Section IV outlines our proposed anomaly detection framework, which includes two distinct methodologies: likelihood-based and Bayesian-based. Following this, in Section V, we display a variety of numerical simulations to demonstrate the performance of our algorithm, comparing it to other recognized anomaly detection methods like the Local Outlier Factor (LOF) and Isolation Forest. A deeper interpretation and discussion of these results can be found in Section VI. The paper culminates with a conclusion in Section VII.

II. RELATED WORK

In [15], a Bayesian network was proposed to predict the behavior of the driver based on curvilinear coordinates of the vehicle on a highway road. Trained on the NGSIM US-101 dataset [16], the network predicted the driver would proceed on either keeping or changing lanes. In [17], a measure based on graph centrality function was proposed to measure the deviation or similarity of one vehicle to the neighboring vehicles. In [18], authors used game theory to model the behavior by benefits of taking specific actions against other drivers. Several datasets with labels on human driving behavior are given in [19]. A human evaluated driving style using LASSO for feature selection and prediction by a simple yet efficient linear regression model was constructed in [20].

On the deep learning front, [21] proposed a framework based on generative adversarial network (GAN) to detect anomalies in time series with application to mobility data. In [22], a GAN was proposed to mimic the driving style of the drivers. Although quite powerful, these methods require a very large training dataset, and a powerful hardware to run in real-time.

Reconstruction-based Anomaly Detection is another popular approach that is built on the premise that similar subjects (or trajectories) can be represented by a simpler object—in the context of sparse coding, a very sparse vector—that produces a minimal reconstruction error. In [23], this approach was used to design an online algorithm for detecting anomalies in a video stream. A similar approach was presented in [6], but with the ℓ_0 norm utilized to regularize sparsity. A major drawback of this approach is the necessity of having a complete trajectory before conducting anomaly detection

(or classification). Conversely, [24] proposed a Bayesian non-parametric framework for detecting collisions and nearby passes at intersections, based on the Hidden Markov Model (HMM). Although these models yielded promising results, the HMM is not readily interpretable, making it difficult to derive performance guarantees.

Aside from learned bases, other natural bases for time series such as Fourier and Spline bases have been explored [25]. In [26] and [27], PCA was employed to reduce the dimensionality of the feature vector, thereby improving the robustness of classification algorithms. Particularly with energy functions, PCA in the functional domain (or FPCA) was successfully used to identify anomalous landing trajectories in [28]. Given the appropriate representations, it is possible to apply various machine learning techniques to classify anomalous trajectories. For instance, [29] uses a Self-Organizing Map Neural Network, while [30] utilized K-means clustering on a Fuzzy membership function. Different distances between an observed trajectory and other nominal trajectories, such as the Hausdorff distance [31] or Dynamic Time Warping (DTW) [32], can also be used. However, similar to reconstruction-based algorithms, these approaches also require full trajectory availability prior to classification. Safety assurance systems, on the other hand, often require the time-to-alarm of anomalous driving behavior to be as low as possible.

These discussions underscore the need for extending unsupervised anomaly detection methods in two directions. Firstly, the application of transforms is necessary to mitigate the curse of dimensionality and enhance learning, which can be achieved by fitting distributions within a space of fewer-dimensional representations. Secondly, anomaly detection must still occur in the original time domain, not in the space of representations. This approach is crucial to accommodate partial observations and streaming data, thereby facilitating the development of a real-time online algorithm.

III. BACKGROUND

A. Functional Principal Component Analysis (FPCA)

Consider trajectories in a set $\mathcal{S}$ square-integrable functionals $\mathcal{S} = \{f_k(t)_{k=1}^{N_v}\}$ on Hilbert Space L^2, it is possible to represent them through the dominant modes of variation which are orthonormal basis corresponding to the eigenvectors of the covariance kernel $K(\cdot, \cdot)$. The treatment of this subject follows [33], [34].

On L^2 space, the covariance kernel is defined as:

$$K(s,t) = \frac{1}{N_v} \sum_{k=1}^{N_v} f_k(s) f_k(t) \tag{1}$$

The goal of FPCA is to find the orthonormal functional components $\mathcal{V} = \{v_j(t)\}_{j=1}^{\infty}$ such that "projections" of $\mathcal{S}$ on basis $\mathcal{V}$ "spread out" as much as possible. Equivalently saying to maximize the variance of projected functionals, we find $v_j(t)$ such that:

$$\frac{1}{N_v} \sum_{k=1}^{N_v} \langle v_j(t), f_k(t) \rangle^2 \tag{2}$$
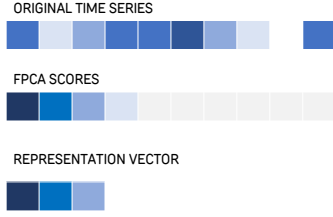
Fig. 1: Dimensionality Reduction Using FPCA: The darker the color shade, the larger the value of the vector component. After applying the FPCA transformation, only a few leading components possess significant values, while the remaining components are nearly zero. We can nullify these components and retain a vector with a significantly reduced number of dimensions.

is maximized under the constraint:

$$\langle v_j(t), v_p(t) \rangle = 0, \forall (j, p) : j \neq p \tag{3}$$

By Mercer's theorem, the covariance kernel admits the following decomposition:

$$K(s, t) = \sum_{i=1}^{\infty} \lambda_j v_j(s) v_j(t) \tag{4}$$

and it was shown in [35] that the solutions $v_j(t)$ to Equation (2) are also the solutions to the eigenvalue problem:

$$\langle K(s, t), v_j(t) \rangle = \lambda_j v_j(s) \tag{5}$$

i.e., the eigenvectors of the covariance kernel. Theoretically, functionals $f_k(t) \in C^{\infty}$ are of infinite dimensions, but we frequently deal with finite, discrete-time series in practice. Two popular methods include (a) approximating $f_k(t)$ with a truncated basis expansion, or (b) opting for a discrete FPCA by approximating the L^2 inner product $\langle \cdot, \cdot \rangle$ by a numerical quadrature form. Either way, Equation (5) becomes a matrix equation and is typically solved by finding the eigenvectors from SVD decomposition (see [34]). The principal component *scores* (sometimes also called *loads*) are then obtained via the following formula:

$$a_{ij} = \langle v_j(t), f_i(t) \rangle \tag{6}$$

and the original functional $f_j(t)$ can be reconstructed, following the Karhunen-Loève theorem:

$$f_i(t) = \sum_{j=0}^{\infty} a_{ij} v_j(t), \quad \text{where } \mathbb{E}[a_{ij}] = 0. \tag{7}$$

Despite the perfect reconstruction requiring an infinite number of bases, in reality, only several leading FPCA *scores* are sufficient to approximate the original functional to a high degree of accuracy (Fig. 1). The significant advantage of FPCA is that by construction, the basis is chosen through a precarious manner that allows representations of long time series with vectors of low dimensions in the representation domain, making machine learning methods on this domain extremely efficient. However, it also requires the availability of full-length time series in order to compute the *scores* per Equation
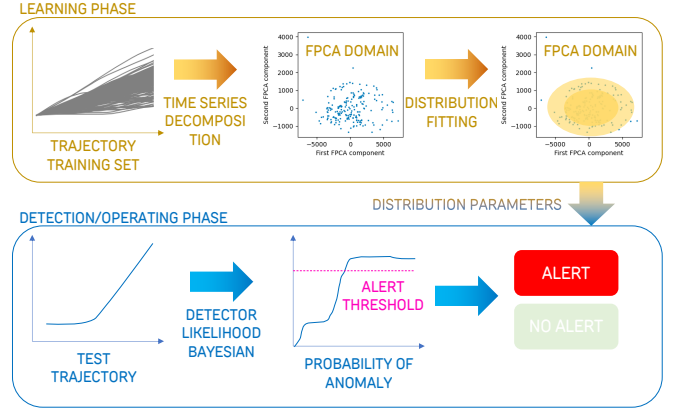

Fig. 2: Outline of the two phases of our detection framework for anomalous time series. The learning phase employs dimensionality reduction for efficient learning of the representation distribution. The detection phase performs real time hypothesis testing of anomalous trajectory.

(6). However, if we assume that a_{ij} follows a multivariate normal distribution, hence $a_{\cdot, j}$ becomes a Gaussian vector that, when lending itself into the linear form of (7) allows us to design a method for online sequential hypothesis testing, as data arrives.

It's important to recognize that FPCA isn't the sole dimensionality reduction method available for time series data. Moreover, the upcoming anomaly detection framework we present can integrate seamlessly with any linear dimensionality reduction technique. Given this versatility, we will use the term "representation" to refer to FPCA scores (or loads).

IV. ANOMALY DETECTION FRAMEWORK

The framework is structured into two distinct stages: the learning phase and the detection phase, as depicted in Fig. 2. The learning phase applies FPCA to reduce the dimensionality of time series trajectory data. This process yields a simplified, low-dimensional representation of the trajectories, facilitating the subsequent fitting of distribution parameters that will be employed during the detection or operation phase.

The detection phase involves real-time hypothesis testing, primarily revolving around the null hypothesis, represented as H_0: observations originate from a "nominal" support of the representation space.

The key objective of this detection framework is to determine whether there is sufficient evidence to reject the null hypothesis.

A. Learning Phase

Consider a specific road patch, denoted as P, which is presumed to be a rectangular section with dimensions $w \times l$. It's important to note that within this patch, a single traffic mode exists at any given time, implying that the traffic pattern remains homogeneous. An illustrative example of such a patch could be a segment of a carriageway, where vehicles are mandated to maintain the same direction and adhere to a uniform speed limit (refer to Fig. 3). A patch containing two
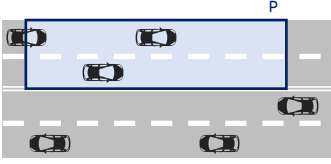
Fig. 3: Illustration of the road patch of interest P. All vehicles should be moving in the same direction within the road patch.

lanes moving in opposite directions would be deemed invalid.

Trajectories of vehicles traversing the designated patch P could be extracted from a security camera like the NGSIM-101 dataset [36] or V2X Cooperative Awareness Message (CAM) traces [37]. This data forms the *training set* $\mathcal{S} = (x_k^i, y_k^i), i \in 0, 1, \ldots, n, k \in 1, 2, \ldots, N_v$. It's worth noting that vehicles may pass through the patch P at various speeds, resulting in a significant variance in the lengths of the trajectories. To address this, we analyze the distribution of these lengths, selecting its mode value as n. We then adjust the trajectories accordingly: those shorter than n are extended using cubic splines, while longer ones are truncated. The moment a vehicle enters the patch is defined as $i = 0$.

Consider a *training set* of vehicle trajectories with a consistent temporal length n denoted as $\mathcal{S}_x = \{\boldsymbol{x}_k \in \mathbb{R}^n, 1 \leq k \leq N_v\}$ for the lateral (lane changing) component, and $\mathcal{S}_y = \{\boldsymbol{y}_k \in \mathbb{R}^n, 1 \leq k \leq N_v\}$. We can utilize Discrete FPCA (as detailed in Section III) to determine the FPCA scores (also referred to as *representations*) for each individual trajectory [34], [35]. In essence, FPCA converts a time-domain trajectory vector $\mathbb{R}^n$ into a representation vector also of dimension $\mathbb{R}^n$. However, only the first few elements carry significant weight, with the rest approaching zero. Consequently, we only consider the first $m << n$ coefficients to derive the sets of representations $\mathcal{P}_x = \{\boldsymbol{z}_{x,k} \in \mathbb{R}^m, 1 \leq k \leq N_v\}$ and $\mathcal{P}_y = \{\boldsymbol{z}_{y,k} \in \mathbb{R}^m, 1 \leq k \leq N_v\}$ for the lateral and longitudinal components of the trajectory, respectively.

The final step involves fitting a Gaussian mixture, denoted as $p_{\mathfrak{M}} \triangleq p(a|\mathfrak{M})$, onto these representation sets $\mathcal{P}_x, \mathcal{P}_y$ using the EM algorithm. Each Gaussian component, represented as p_{a_i} within the mixture, signifies a distinct nominal traffic mode. Consider a situation where a road patch includes a highway ramp, vehicles may decide to proceed straight or make a right turn to follow the ramp off the highway. In such instances, two distinguishable clusters might emerge in the representation space, leading the mixture to comprise two components.

I think this section is already sound.

B. Detection Phase

We formalize our problem as follows. An ego-vehicle E traveling near a potential outlier vehicle O and records a series of state observations from O denoted as $\boldsymbol{x}_{1:n} = [x_1, x_2, \ldots, x_n]^\top$, $n < t_{max}$. The random variable x_k denotes the state observation at time $T_{obs} + (k-1)T_s$, given that observation starts at T_{obs} and the sampling time is T_s. For convenience, suppose $T_{obs} = kT_s, k \in \mathbb{N}$. We also denote x^i

as the random variable of observation at iT_s. In the following, two detection methods will be presented, one based on *likelihood computation* and the other on posterior distribution computation, a *Bayesian method*.

1) Likelihood Method: Given a latent representation a like a vector of FPCA scores, assume that each measurement of the trajectory is independent, the *likelihood distribution* can be written as:

$$p(x^{1:n}|a) = \prod_{i=1}^n p(x^i|a) \tag{8}$$

The learning phase has given us the prior distribution of latent representation $p(a)$ hence it is possible to compute the *marginal likelihood*:

$$\mathcal{L}_{\mathfrak{M}} \triangleq p(x^{1:n}) = \int_a p(x^{1:n}|a)p(a)da = \int_a \prod_{i=1}^n p(x^i|a)p(a)da \tag{9}$$

The marginal likelihood can be used to test the hypothesis, H_0. As will be evident later, the marginal likelihood converges to a value proportional to the distance from the origin to the true latent representation, a^*, which generates the observations.

We consider a particular class of observations as functionals that admit a unique linear decomposition:

$$\boldsymbol{x}|a = \sum_{k=0}^m a_k \phi_k(x), \quad m < +\infty \tag{10}$$

and that $\{a_k\} \in \mathbb{R}^m$ is a Gaussian random representation vector that follows $\mathcal{N}(a; \mu_a; \Sigma_a)$. Hence, in the discrete domain, a segment of vehicle trajectory could be written as:

$$x^i = \mu^{T_{obs}+i} + \sum_{k=1}^m a_k \phi_k^{T_{obs}+i} \tag{11}$$

where $\epsilon^i \sim \mathcal{N}(0; \sigma_\epsilon{}^2)$ (i.i.d) is related to the sensor measurement error.

Note that because T_{obs} is unknown, [38] suggested to replace T_{obs} as an unknown and find T_{obs} that maximizes $\mathcal{L}_{\mathfrak{M}}$:

$$\mathcal{L}_{\mathfrak{M}} = \max_{T_{obs}} \int_a \prod_{i=1}^n p(x^i|a)p(a)da \tag{12}$$

where x^i follows (11).

Recursive analytical solutions. Analytical solutions for problem (12) exist when $p(a)$ is a simple Gaussian distribution (or even a Gaussian mixture, due to the linearity of integral). For the moment, suppose the measurement noise $\epsilon^i = 0$. Consider the term:

$$\mathcal{L}_{\mathfrak{M}} = \int_a \prod_{k=1}^n p(x^k|a)p(a)da = \frac{1}{(\sigma_\epsilon \sqrt{2\pi})^n} \frac{1}{(2\pi)^{m/2}\sqrt{det(\Sigma_a)}}$$
$$\int_a \exp\left[-\frac{1}{2}(a-\mu_a)^\top \Sigma_a^{-1}(a-\mu_a) \right.$$
$$\left. -\frac{1}{2\sigma_\epsilon^2} \sum_{k=1}^n \left(x^k - \mu^{T_{obs}+k} - \sum_{j=1}^m a_j \phi_j^{T_{obs}+k} \right)^2 \right] da \tag{13}$$

Denote $\Phi^{T_{obs}+k} = [\phi_1^{T_{obs}+k}, \phi_2^{T_{obs}+k}, \ldots]^\top$, $\mathcal{L}_{\mathfrak{M}}$ becomes:

$$
\mathcal{L}_{\mathfrak{M}} \propto \int_a \exp\left[-\frac{1}{2}\left(a^\top \Sigma_a^{-1} a - 2 a^\top \Sigma_a^{-1} \mu_a + \mu_a^\top \Sigma_a^{-1} \mu_a \right) \right.
$$
$$
-\frac{1}{2\sigma_\epsilon^2} \sum_{k=0}^{n-1} \left((x^k)^2 + (\mu^{T_{obs}+k})^2 + \left(\Phi^{T_{obs}+k \top} a \right)^2 \right.
$$
$$
\left. \left. -2 x^k \mu^{T_{obs}+k} - 2\mu^{T_{obs}+k} \Phi^{T_{obs}+k \top} a - 2 x^k \Phi^{T_{obs}+k \top} a \right) \right] da
\tag{14}
$$

We further let $C_\times^{T_{obs}+k} = \frac{1}{\sigma_\epsilon^2}(\Phi^{T_{obs}+k} \otimes \Phi^{T_{obs}+k})$ where $\otimes$ being the outer product, and $C_\dagger^{T_{obs}+k} = \frac{1}{\sigma_\epsilon^2}(-\mu^{T_{obs}+k} + x^{T_{obs}+k})\Phi^{T_{obs}+k \top}$, then equation (14) can be rewritten as:

$$
\mathcal{L}_{\mathfrak{M}} \propto \int_a \exp\left[-\frac{1}{2}\left[a^\top \left(\Sigma_a^{-1} + \sum_{k=0}^{n-1} C_\times^{T_{obs}+k} \right) a \right. \right.
$$
$$
\left. \left. -2\left(\Sigma_a^{-1} \mu_a + C_\dagger^{T_{obs}+k} \right)^\top a + \left(\mu_a^\top \Sigma_a^{-1} \mu_a + \sum_{k=0}^{n-1} \iota^{T_{obs}+k} \right) \right] \right] da
$$

where we have replaced:

$$
\iota^{i+k} = \frac{1}{\sigma_\epsilon^2}(x^k - \mu^{T_{obs}+k})^2
\tag{15}
$$

The exponent, after completion of squares, bares the form of a Gaussian distribution's exponent, with the new covariance and mean:

$$
\Sigma_a^{*-1} = \Sigma_a^{-1} + \sum_{k=0}^{n-1} C_\times^{T_{obs}+k}
\tag{16}
$$

and

$$
\mu_a^* = \left(\Sigma_a^{-1} + \sum_{k=0}^{n-1} C_\times^{T_{obs}+k} \right)^{-1} \left(\Sigma_a^{-1}\mu_a + \sum_{k=0}^{n-1} C_\dagger^{T_{obs}+k} \right)
\tag{17}
$$

Equation (16) reflects the gain of information from a bare hypothesis, sequentially with time as new measurements arrive. The mean shift is shown in (17). Finally, to perform the marginalization integral in (12) we use the Euler-Poisson integral to yield:

$$
\mathcal{L}_{\mathfrak{M}} = \frac{1}{(\sigma_\epsilon \sqrt{2\pi})^n} \sqrt{\frac{det(\Sigma_a^*)}{det(\Sigma_a)}}
$$
$$
\exp\left(-\frac{1}{2}\left(-\mu_a^{*\top}\Sigma_a^{*-1}\mu_a^* + \mu_a^\top \Sigma_a^{-1}\mu_a + \sum_{k=0}^{n-1}\iota^{T_{obs}+k} \right) \right)
\tag{18}
$$

It is noteworthy that (16), (17) can be used for the recursive calculation of $\mathcal{L}_{\mathfrak{M}}$. It will become evident later that these expressions also allow for the calculation of the posterior mean and covariance in the Bayesian method. The details are shown in Algorithm 1.

Algorithm 1 Recursive Computation of Posterior Distribution Parameters and Marginal Likelihood.

1: initialize $\Sigma_a^{*-1} \leftarrow \Sigma_a^{-1}$, $\mu_a^{*'} \leftarrow \Sigma_a^{-1}$.
2: **for** $T_{obs} \in \mathbb{N}$ **do**
3:     **if** data arrives **then**
4:         $\Sigma_a^{*-1} \leftarrow \Sigma_a^{*-1} + \frac{1}{\sigma_\epsilon^2}(\Phi^{T_{obs}+k} \otimes \Phi^{T_{obs}+k})$
5:         $\mu_a^{*'} \leftarrow \mu_a^{*'} + \frac{1}{\sigma_\epsilon^2}(-\mu^{T_{obs}+k} + x^{T_{obs}+k})\Phi^{T_{obs}+k \top}$
6:         $\mu_a^* \leftarrow \Sigma_a^* \mu_a^{*'}$
7:         $\mathcal{L}_{\mathfrak{M}}(T_{obs}) \leftarrow$ Equation (18)
8:     **end if**
9: **end for**
10: **return** $\max_{T_{obs}} L_{\mathfrak{M}}$

---

2) Bayesian method: From Bayes Theorem:

$$
p(a|x_{1:n}) = \frac{p(x_{1:n}|a)p(a)}{\int_a p(x_{1:n}|a)p(a)da} \propto_a p(x_{1:n}|a)p(a),
\tag{19}
$$

which is nothing but the integrand of $\mathcal{L}_{\mathfrak{M}}$ in (13). Because a Gaussian form is exhibited in the integrand of equation (IV-B1), it can be deduced that (16) and (17) are the covariance and mean of the posterior distribution, respectively.

The likelihood method yields a single real number, which decreases over time as more observations are incorporated. In contrast, the Bayesian method generates a distribution, wherein the eigenvalues of the covariance matrix decrease with time, suggesting a more reliable posterior estimate. In the subsequent section, the results are presented, enabling the use of these values for hypothesis testing.

C. Real-time Hypothesis Testing with the Marginal Likelihood

Suppose that the trajectories learned in the learning phase came from a ranging sensor that can be modeled with the following measurement model:

$$
x^i = \hat{x}^i + \epsilon^i,
\tag{20}
$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$. For this marginal likelihood method, we also assume $\hat{x}^i$ satisfies (11). The usual way to define anomaly in an unsupervised manner is as follows.

Definition IV.1. *A trajectory is defined as abnormal if the Euclidean distance from its FPCA representation a^* in (7) to the centroid of the training set's FPCA representations is larger than some threshold $\hat{d}$:*

$$
||a^* - \mathbb{E}[a]||_2 \geq \hat{d},
\tag{21}
$$

and from the Karhunen-Loeve theorem, it is known that $\mathbb{E}[a] = 0$, thus the definition is equivalent to:

$$
||a^*||_2 \geq \hat{d}.
\tag{22}
$$

Given $p\left(x^i|a\right) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon}e^{-\frac{\left(x^i-\mu^i-\Phi^{i\top}a\right)^2}{2\sigma_\epsilon^2}}$ and $x^i = \Phi^{i\top}a^* + \mu^i + \epsilon^i$. Then:

$$p\left(x^i|a\right) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon}e^{-\frac{\left(\Phi^{i\top}\left(a^*-a\right)\ +\epsilon^i\right)^2}{2\sigma_\epsilon^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_\epsilon}e^{-\frac{\left(\Phi^{i\top}\left(a^*-a\right)\ \right)^2}{2\sigma_\epsilon^2}}e^{-\frac{\left(\epsilon^{i2}+2\Phi^{i\top}\left(a^*-a\right)\epsilon^i\right)}{2\sigma_\epsilon^2}}$$

$$(23)$$

Substituting into (9) we have:

$$\mathcal{L}_{\mathfrak{M}} = \int_a \frac{1}{(2\pi)^{\frac{n+m}{2}}\sigma_\epsilon^n\sqrt{\det(\Sigma_a)}}$$

$$e^{-\frac{\sum_{i=1}^n\left(\Phi^{i\top}\left(a^*-a\right)\ \right)^2}{2\sigma_\epsilon^2}-\frac{1}{2}(a-\mu_a)^\top\Sigma_a^{-1}(a-\mu_a)}$$

$$e^{-\frac{\sum_{i=1}^n\left(\epsilon^{i2}+2\Phi^{i\top}\left(a^*-a\right)\epsilon^i\right)}{2\sigma_\epsilon^2}}da$$

Defining the variables $b = a - \mu_a$, $b^* = a^* - \mu_a \Rightarrow a^* - a = b^* - b$ and $da = db$, we have:

$$\mathcal{L}_{\mathfrak{M}} = \int_a \frac{1}{(2\pi)^{\frac{n+m}{2}}\sigma_\epsilon^n\sqrt{\det(\Sigma_a)}}$$

$$e^{-\frac{\sum_{i=1}^n\left(\Phi^{i\top}\left(b^*-b\right)\right)^2}{2\sigma_\epsilon^2}-\frac{1}{2}b^\top\Sigma_a^{-1}b}$$

$$e^{-\frac{\sum_{i=1}^n\left(\epsilon^{i2}+2\Phi^{i\top}\left(b^*-b\right)\epsilon^i\right)}{2\sigma_\epsilon^2}}db$$

The term $1/[(2\pi)^{\frac{n+m}{2}}\sigma_\epsilon^n\sqrt{\det(\Sigma_a)}]$ can be recognized as simply a normalization constant. To condense and streamline the calculation steps, it can be denoted as $C(n)$. Now, by completing the squares:

$$\mathcal{L}_{\mathfrak{M}} = C(n)\int_a \exp\left[-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^n\left(\Phi^{i\top}b^*\right)^2 + \right.$$

$$\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^n\left(\Phi^{i\top}b^*\Phi^{i\top}b\right) - \frac{1}{2}b^\top\left(\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^n\Phi^i\Phi^{i\top} + \Sigma_a^{-1}\right)b\right]$$

$$\exp\left[-\frac{\sum_{i=1}^n\left(\epsilon^{i2} + 2\Phi^{i\top}\left(b^* - b\right)\epsilon^i\right)}{2\sigma_\epsilon^2}\right]db$$

Taking the Poisson-Euler integral and get the log of both sides:

$$\log\mathcal{L}_{\mathfrak{M}} = \log C(n) + \frac{m}{2}\log(2\pi) + \frac{1}{2}\log\det\Sigma^*$$

$$+\frac{1}{2\sigma_\epsilon^2}\left[\left(\sum_i\Phi^{i\top}b^*\Phi^{i\top} + \sum_i\Phi^{i\top}\epsilon^i\right)\right.$$

$$\frac{\Sigma^*}{\sigma_\epsilon^2}\left(\sum_i\Phi^ib^{*\top}\Phi^i + \sum_i\Phi^i\epsilon^i\right) - \sum_i\left(\Phi^{i\top}b^* + \epsilon^i\right)^2\right]$$

$$(24)$$

where

$$\Sigma^* = \left(\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^n\Phi^i\Phi^{i\top} + \Sigma_a^{-1}\right)^{-1}$$

For convenience, we let $\log D(n) = \log C(n) + \frac{m}{2}\log(2\pi) + \frac{1}{2}\log\det\Sigma^*$ and notice that this term is independent of b. As a result, computation of $D(n)$ can be bypassed if we let $b^* = 0$
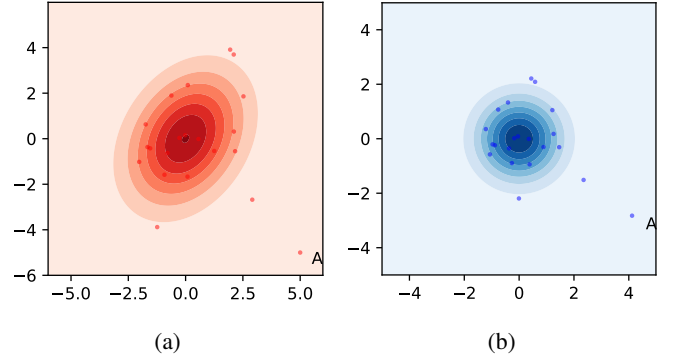
Fig. 4: Density function of two 2D Gaussian latent distributions $p(b^*)$ and $p(R^{-1/2}b^*)$. The latent representation A is identified as an anomaly due to its substantial distance from the cluster's centroid.

to obtain the marginal likelihood curve $\log D(n)$, then subtract $\log D(n)$ from $\mathcal{L}_{\mathfrak{M}}$.

1) Perfect Measurement Case: We solve the simple case of perfect measurements (no noise, or $\epsilon^i = 0$). Then (24) becomes:

$$\log\mathcal{L}_{\mathfrak{M}} = \log D(n)\ + \frac{1}{2\sigma_\epsilon^2}\left[\left(\sum_i\Phi^{i\top}b^*\Phi^{i\top}\right)\frac{\Sigma^*}{\sigma_\epsilon^2}\right.$$

$$\left(\sum_i\Phi^ib^{*\top}\Phi^i\right) - \sum_i\left(\Phi^{i\top}b^*\right)^2\right]$$

The objective of this framework is to determine whether the latent representation of the observed trajectory significantly deviates from the cluster's centroid, which is also the origin point. However, using Euclidean distance as a direct measure of this deviation poses challenges, as the interpretation of distance may vary across different directions. This variation is attributable to a greater spread of data points along the eigenvectors of the covariance matrix, meaning that a data point A at a given distance d in this direction may still have more neighboring points than another data point B, at the same distance d but in a different direction (Fig. 4).

Mathematically, we would like to find a matrix R such that $b^* = Rc^*$ where $c^* \sim \mathcal{N}(0; \boldsymbol{I}_{m\times m})$. Then the distance from b^* to $\boldsymbol{0}$ can be straightforwardly defined with $||c^*||_2$. One straightforward choice is:

$$R = \Sigma_a^{1/2} \qquad (25)$$

Proposition IV.2. *Without measurement noise ($\epsilon^i = 0$), the marginal likelihood $\mathcal{L}_{\mathfrak{M}}$ defined in (9) is bounded below by:*

$$T^i = \lambda_1\left(-U^\top\frac{\Sigma^*}{\sigma_\epsilon^2}U + V\right)d, \qquad (26)$$

if the Euclidean distance from the transformed latent representation $c^* = \Sigma_a^{-1/2}b^*$ to the mean of the transformed prior distribution $p(\Sigma_a^{-1/2}a)$ is d, where U, V are defined in (27) and $\lambda_1(M)$ denotes the largest eigenvalue of the matrix M.

Proof. The log likelihood expression can be rewritten as:

$$\log \mathcal{L}_{\mathfrak{M}} = \log D(n) \ + \frac{1}{2\sigma_\epsilon^2}\left[\left(\sum_i \Phi^{i\top} Rc^* \Phi^{i\top}\right)\frac{\Sigma^*}{\sigma_\epsilon^2}\right.$$

$$\left.\left(\sum_i \Phi^i (Rc^*)^\top \Phi^i\right) - \sum_i \left(\Phi^{i\top} Rc^*\right)^2\right]$$

Using the dot product property, with some transformations:

$$\sum_i \Phi^{i\top} Rc^* \Phi^{i\top} = \sum_i \left(R^\top \Phi^i\right)^\top c^* \Phi^{i\top} = \sum_i c^{*\top}\left(R^\top \Phi^i \Phi^{i\top}\right)$$

$$\sum_i \left(\Phi^{i\top} Rc^*\right)^2 = \sum_i c^{*\top} R^\top \Phi^i \Phi^{i\top} Rc^*$$

Hence:

$$\log \mathcal{L}_{\mathfrak{M}} = \log D(n) \ + \frac{1}{2\sigma_\epsilon^2}\left[\left(c^{*\top}\left(\sum_i \left(\Phi^i \Phi^{i\top} R\right)^\top\right)\right)\frac{\Sigma^*}{\sigma_\epsilon^2}\right.$$

$$\left.\left(\sum_i \left(\Phi^i \Phi^{i\top} R\right)\right)c^* - c^{*\top}\left(\sum_i R^\top \Phi^i \Phi^{i\top} R\right)c^*\right]$$

Let:

$$U = \sum_i \Phi^i \Phi^{i\top} R = \Phi R$$

$$V = \sum_i R^\top \Phi^i \Phi^{i\top} R = R^\top \Phi R \tag{27}$$

We have:

$$\log \mathcal{L}_{\mathfrak{M}} = \log D(n) \ - \frac{1}{2\sigma_\epsilon^2}\left[c^{*\top}\left(-U^\top \frac{\Sigma^*}{\sigma_\epsilon^2}U + V\right)c^*\right]$$

which is a quadratic form. Note that $-U^\top \frac{\Sigma^*}{\sigma_\epsilon^2}U + V$ is positive definite, hence:

$$\log \mathcal{L}_{\mathfrak{M}} \geq \log D(n) \ - \frac{1}{2\sigma_\epsilon^2}\left[\lambda_1\left(-U^\top \frac{\Sigma^*}{\sigma_\epsilon^2}U + V\right)||c^*||^2\right] \tag{28}$$

where λ_1 indicates the maximum eigenvalue of $-U^\top \frac{\Sigma^*}{\sigma_\epsilon^2}U + V$. $\square$

Proposition IV.3. *The bound of (28) is tight. Assume that Q is full-rank. In case of infinite observations, the lower bound converges to:*

$$T^\infty = -\frac{1}{2}||c^*||_2^2 \tag{29}$$

Proof. See Appendix A. $\square$

2) Noisy Measurement Case: In this section, we derive the likelihood of anomaly when measurements are contaminated with Gaussian white noise, that is $\epsilon^i \neq 0$ generally. Rewriting (24) as follows:

$$\log \mathcal{L}_{\mathfrak{M}} = \log D(n) + \frac{1}{2\sigma_\epsilon^2}\left[-c^{*\top}\left(-U^\top \frac{\Sigma^*}{\sigma_\epsilon^2}U + V\right)c^*\right.$$

$$+2\left(\sum_i \epsilon_i \Phi^{i\top}\right)\frac{\Sigma^* \Phi R}{\sigma_\epsilon^2}c^* + \left(\sum_i \epsilon^i \Phi^{i\top}\right)\frac{\Sigma^*}{\sigma_\epsilon^2}\left(\sum_i \Phi^i \epsilon^i\right)$$

$$\left.-\sum_i \left(\epsilon^{i2} + 2\Phi^{i\top} Rc^* \epsilon^i\right)\right]$$

which can be rewritten as:

$$\log \mathcal{L}_{\mathfrak{M}} = \log D(n) + \frac{1}{2\sigma_\epsilon^2}\left[-c^{*\top}\left(-U^\top \frac{\Sigma^*}{\sigma_\epsilon^2}U + V\right)c^*\right.$$

$$+2\left(\sum_i \epsilon^i \Phi^{i\top}\right)\left(\frac{\Sigma^* \Phi}{\sigma_\epsilon^2} - I\right)Rc^*$$

$$+ \left(\sum_i \epsilon^i \Phi^{i\top}\right)\frac{\Sigma^*}{\sigma_\epsilon^2}\left(\sum_i \Phi^i \epsilon^i\right) - \sum_i \epsilon^{i2}\right]$$

$$= \log C(n) \ + \frac{1}{2\sigma_\epsilon^2}[A + B + C - D]$$

Term A is the same as in the perfect measurement case. For term B, let us define $\gamma = \sum_i \phi^i \epsilon^i$ (note $\mathbb{E}\gamma = 0$). Define G and use the matrix inversion lemma:

$$G = \frac{\left(\frac{1}{\sigma_\epsilon^2}\Phi + \Sigma_a^{-1}\right)^{-1}\Phi}{\sigma_\epsilon^2} - I = \left(I + \left(\frac{\Sigma_a \Phi}{\sigma_\epsilon^2}\right)^{-1}\right)^{-1} - I$$

$$= -\left(\frac{\Sigma_a \Phi}{\sigma_\epsilon^2} + I\right)^{-1}$$

Then:

$$B = 2\gamma^\top GRc^*$$

Since $\frac{\Sigma_a \Phi}{\sigma_\epsilon^2}$ is positive definite, there exists an eigen-decomposition $Q\Lambda Q^T$:

$$G = -\left(Q\Lambda Q^T + QQ^T\right)^{-1} = -Q(\Lambda + I)^{-1}Q^T$$

Hence, the eigenvectors of G are also the eigenvectors of $\frac{\Sigma_a \Phi}{\sigma_\epsilon^2}$. In a general FPCA decomposition, the eigenvalues of G can potentially be very large, thus the measurement noise is amplified. However, it is worth noting that this only causes a false alarm if the anomalous vehicle is whose latent representation of the trajectory c^* is parallel to v_1 where v_1 is the eigenvector corresponding to the largest eigenvalue of G, and γ hits the same direction as well. Given the randomness of γ, $B \approx 0$.

For C, as $\left(\frac{1}{\sigma_\epsilon^2}\Phi + \Sigma_a^{-1}\right)^{-1} \to 0$ and at the beginning of the detection, $||\gamma||$ is small, C is usually small throughout the whole detection process. That leaves us with:

$$\log \mathcal{L}_{\mathfrak{M}} \approx \log D(n)$$

$$+ \frac{1}{2\sigma_\epsilon^2}\left[-c^{*\top}\left(-U^T \frac{\Sigma^*}{\sigma_\epsilon^2}U + V\right)c^* - \sum_i \epsilon^{i2}\right] \tag{30}$$

It should be noted that, according to Proposition IV.3, the first term converges to the distance between the latent representation c^* and the origin. This effectively represents the degree of anomaly apparent in the current stream of observations. Therefore, this term can be interpreted as the "signal" component of the marginal likelihood. On the other hand, the second term, $\sum_i \epsilon^{i2}$, introduces noise that disrupts the signal, thereby adding complexity to the detection process.

Observe that σ_ϵ^2 is a parameter of the detector. Since the first term converges to $||c^*||^2$ irrespective of σ_ϵ^2, we can employ this variable to optimize the signal-to-noise ratio. Let's consider the case where $Var[\epsilon^i] = \sigma_{\epsilon,m}^2$:

$$\log \mathcal{L}_{\mathfrak{M}} - \log D(n) \to -\frac{1}{2}||c^*||^2 - \frac{\sigma_{\epsilon,m}^2}{2\sigma_\epsilon^2}\sum_i \frac{\epsilon^{i2}}{\sigma_{\epsilon,m}^2} \tag{31}$$

Then it is obvious that for best performance, one should keep the ratio $\sigma_{\epsilon,m}^2/\sigma_\epsilon^2 \approx 0$, or σ_ϵ^2 sufficiently large.

Because $\sum_i \epsilon^{i2}$ follows a χ^2 distribution, (30) suggests a way to test the hypothesis H_0 with the marginal likelihood $\mathcal{L}_{\mathfrak{M}}$. Let P_n be the *nominal probability*, i.e., probability that observations are coming from the transformed latent representation that is less than a distance of d from the origin:

$$
\begin{aligned}
P_n = F_{\chi^2}\Bigg[&- 2\frac{\sigma_\epsilon^2}{\sigma_{\epsilon,m}^2}(\log \mathcal{L}_{\mathfrak{M}} - \log D(n)) \\
&+ \frac{1}{\sigma_{\epsilon,m}^2}\left(-c^{*\top}\left(-U^T\frac{\Sigma^*}{\sigma_\epsilon^2}U + V\right)c^*\right)\Bigg]
\end{aligned}
\tag{32}
$$

where F_{χ^2} denotes the cumulative distribution function (CDF) of a chi-squared random variable.

D. Bayesian Anomaly Detection

The noise term $\sum_{i=1}^n \epsilon^{i2}$, as derived from equation (30), may dominate the signal term under certain conditions, including: (1) when n is large, (2) when the generative noise ϵ^i values do not represent independent samples from a Gaussian distribution, and (3) when the approximation provided by equation (10) is unsatisfactory. Consequently, these factors can significantly diminish the performance of an anomaly detector relying on marginal likelihood.

A potential solution is to rely on the Bayesian view to design an alternative anomaly detector. Let $Q = R^{-1}$. Conditioned on past observations $x_{1:n}$, we have the posterior estimate of $a^* \sim \mathcal{N}(\mu_a^*, \Sigma_a^*)$ where μ_a^* and Σ_a^* are computed recursively from Algorithm 1. Like before, we introduce a transformed latent representation $c^* = R^{-1}a^*$ whose prior is the multivariate normal distribution $\mathcal{N}(0, \boldsymbol{I})$:

$$
c^* \sim \mathcal{N}(Q\mu_a^*, Q\Sigma_a^* Q^\top)
$$

We further define:

$$
Q^* = (Q\Sigma_a^* Q^\top)^{1/2}
$$

then the random variable:

$$
\nu = Q^{*-1}(c^* - Q\mu_a^*)
$$

is a normal random vector. Equivalently:

$$
c^* = Q^*\nu + Q\mu_a^*,
$$

which leads to:

$$
\begin{aligned}
c^{*\top}c^* &= (Q^*\nu + Q\mu_a^*)^\top(Q^*\nu + Q\mu_a^*) \\
&= \nu^\top Q^{*\top}Q^*\nu + 2\mu_a^{*\top}Q^\top Q^*\nu \\
&\quad + \mu_a^{*\top}Q^\top Q\mu_a^*
\end{aligned}
\tag{33}
$$

This expression bears the form of a Generalized Chi-Squared random variable $x^\top Q_2 x + \boldsymbol{q_1}^\top x + q_0$ and a closed form analytical solution generally does not exist. However, note that as $\Sigma_a^* \to 0$ and Q is full rank, the second order term converges to zero faster than the first order term. Hence, as Σ_a^* is sufficiently small:

$$
c^{*\top}c^* \approx 2\mu_a^{*\top}Q^\top Q^*\nu + \mu_a^{*\top}Q^\top Q\mu_a^*
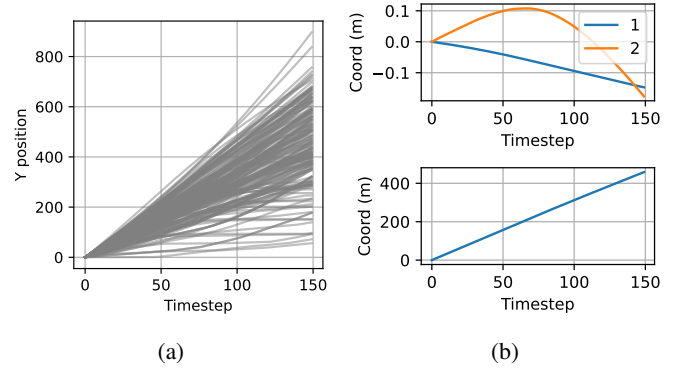\tag{34}
$$

Fig. 5: All longitudinal trajectories of vehicles from the NGSIM dataset (a), and (b) the FPCA components f_1, f_2 of (6) (top) the mean component (bottom) used to dimensionally reduce the trajectories to 2D representations in Fig. 6.

which is simply a Gaussian distribution with parameters:

$$
\begin{aligned}
\mu_p &= \mu_a^{*\top}Q^\top Q\mu_a^* \tag{35} \\
\sigma_p^2 &= 4\mu_a^{*\top}Q^\top Q^* Q^{*\top}Q\mu_a^* \tag{36}
\end{aligned}
$$

and the nominal p-value is:

$$
p_n = F_{\mathcal{N}}\left(\frac{d^2 - \mu_p}{\sigma_p}\right)
\tag{37}
$$

where $F_{\mathcal{N}}$ is the CDF of a normal distribution.

V. NUMERICAL SIMULATIONS

A. Data Preprocessing

The NGSIM US-101 dataset [16] comprises genuine vehicle trajectories collected from traffic cameras installed along a section of the US-101 highway in California. Each trajectory boasts a temporal resolution of up to 0.1 seconds, which is sufficiently precise for macroscopic traffic modeling. The dataset includes labels that extend beyond mere vehicle positions and timestamps to encompass vehicle dimensions (length and width), the lane the vehicle occupies, and measurements such as headway, headway distance, and time with respect to leading vehicles.

We defined the road patch P whose width spans all the lanes and length is 900ft (≈ 274m) and randomly collected 200 vehicle trajectories passing through P. Let t_M be the mode of the distribution of trajectories length in $\mathcal{S}$. Because the Discrete FPCA required a time series of equal length, the outliers significantly longer or shorter than t_M were removed. Those that fell within the acceptable range of ± 10 time steps were either trimmed or extrapolated with a second-order cubic spline to meet the requirement (see Fig. 5a).

B. Learning the Nominal Distribution

The discrete FPCA process, described in section III-A revealed that the first two principal components (Fig. 5b) explained up to 99.67% of variance in the dataset. This allowed us to assume that the first two FPCA components were sufficient to reconstruct $\mathcal{S}$ as well as to discriminate
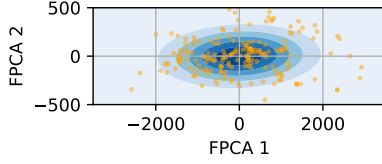
Fig. 6: Decomposition of time series in the training set to representation forms. Data points are yellow scattered dots. The fitted Gaussian distribution contours are also shown.

TABLE I: Latent Representation c^* and the Responsiveness of the Detection Algorithm with Detection Threshold $d = 1$.

Curve	Latent Represent. $c^* = Rb^*$	Alarm Delay
0	$(0, 0)$	∞
1	$(0.5, 0)$	∞
2	$(0, 1)$	∞
3	$(0, 0.99)$	∞
4	$(0.99, 0)$	∞
5	$(1, 1.21)$	14

against anomalies. However, it is noteworthy that although theoretically on the functional domain C^∞, FPCA will yield an "almost exact" reconstruction of any functional $f \in C^\infty$ but realistically, the discrete implementation leads to increasing error towards the end of the trajectories. For this reason, we proposed to only perform detection on the first 90% segment of the trajectory, leaving the remaining 10% as a safety margin.

Each time series decomposed with the two principal components above will become a 2D vector in the representation domain. Since the movement is unimodal in this road patch, we opted for fitting a Gaussian distribution $p(a)$ over the set of representation vectors $\mathcal{P}$ (Fig. 6).

C. Noiseless Detection using the Marginal Likelihood

In the absence of measurement noise, the anomaly threshold boundary is defined by Proposition IV.3. Fig. 7 presents a total of six trajectories, numbered from 0 to 5. Their corresponding representation form, c^*, as described in Proposition IV.2, is provided in the second column of Table I. We categorize trajectories with latent representations located within the unit disc as nominal trajectories, that is, those corresponding to $d = 1$. Under this classification, approximately 68% (equivalent to 1 standard deviation) of trajectories are considered nominal.

As anticipated, curves with $||c^*|| \leq 1$ did not trigger any alarms. Alarm was only raised for curve 5, where $||c^*|| \approx 1.57$, after only 14 of 150 observations. This highlights a significant distinction between our methods and other comparison techniques like sparse representation or sliding window-based techniques. The latter requires the window to be fully populated before further analysis can occur, resulting in additional delays.

D. Detection of Anomaly in The Case of Noise-Contaminated Observations Using the Marginal Likelihood

The marginal likelihood detector detailed in (32) is showcased as an unsupervised tool to detect aberrant driving
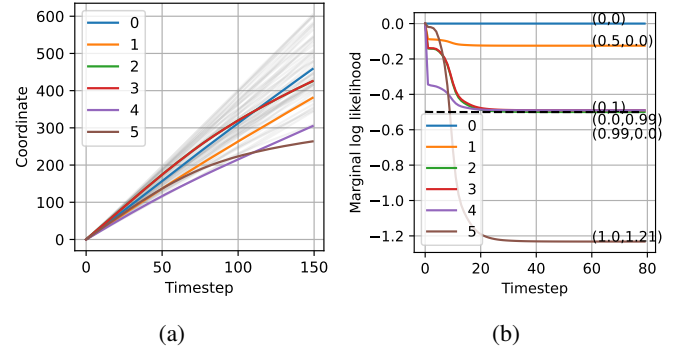

Fig. 7: On the left-hand side, six sample trajectories are depicted, with the "nominal trajectories" - those with transformed latent representations residing within the unit disc - presented in gray. On the right-hand side, the real-time computation of the marginal likelihood curves that correspond to these six sample trajectories. The threshold corresponding to $d = 1$ is the disconnected black line. Marginal likelihood curves exceeding this line will trigger the alarm of anomaly, which mark significant deviation from "nominal trajectories".
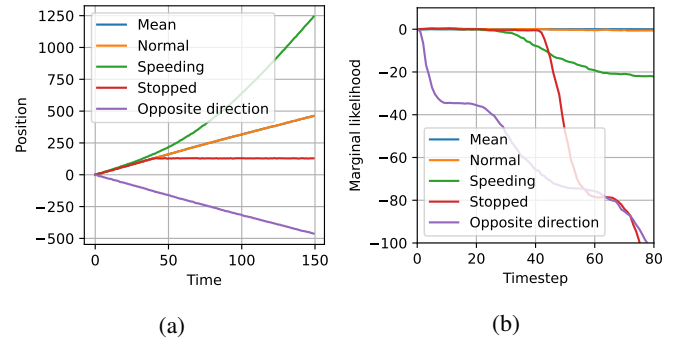

Fig. 8: (a) Common anomalous trajectories. (b) The marginal likelihood computed by Algorithm 1. The mean trajectory was added for visualization purpose only.

behavior through the demonstration of four test trajectories depicted in Fig. 8(a). These trajectories are representative of: (1) a typical driving trajectory from the dataset, (2) a vehicle exceeding the average traffic speed by up to two-fold, notably from timestep 25, (3) a vehicle coming to an unexpected halt at the 40th timestep, and (4) a vehicle proceeding in an incorrect direction. The simulation is further refined by the addition of zero-mean Gaussian White Noise (GWN) with a variance of $1/\sigma_{\epsilon,m}^2 = 1$, mimicking sensor noise. The marginal likelihoods associated with these trajectories are illustrated in Fig. 8(b).

The anomaly p-value for each trajectory, calculated using equation (32), is depicted in Fig. 9. In this configuration, we have set the variance $1/\sigma_\epsilon^2$ to 0.08 and the detection threshold d to 1. The black dotted line signifies the alarm threshold set at $p = 0.05$. Within these examples, all abnormal behaviors were accurately classified. The alarm for the speeding vehicle was triggered at $t = 35$, while the alarm for the halted vehicle was activated at $t = 46$, representing a delay of just 6 timesteps.

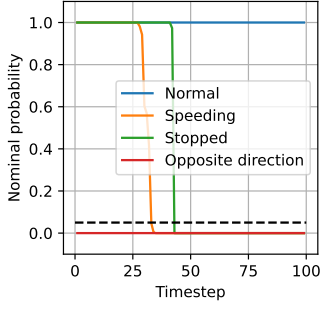To underscore the findings, we compare the results with those obtained using the well-established Local Outlier Fac-

Fig. 9: Marginal likelihood method: probability that the trajectory is nominal. The dotted black line is the alarm threshold.

(a) $1/\sigma_\epsilon^2 = 0.08$　　　　　(b) $1/\sigma_\epsilon^2 = 8$

Fig. 11: Bayesian method: probability that the trajectory is nominal. The dotted black line is the alarm threshold set corresponding to $p = 0.05$.
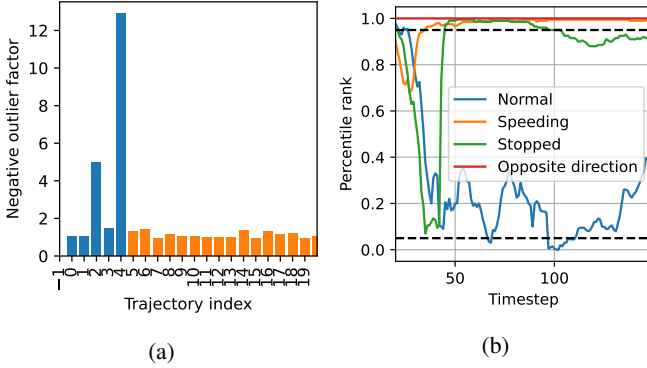

(a)　　　　　　　(b)

Fig. 10: Calculating Negative LOF with $N_{neighbors} = 5$: (a) The LOF (Local Outlier Factor) for all trajectories, with the first five representing test trajectories and the remaining ones constituting the training set. (b) The LOF's progression over time corresponding to partial observations. The dotted lines represent the thresholds for the 5% and 95% percentiles.

Fig. 12: Anomaly scores computed in real-time with Isolation Forest

tor (LOF) and Isolation Forest (IF) methods for anomaly detection. Our results expose significant limitations within LOF specifically, but these shortcomings are speculated to be pervasive within the family of local density anomaly detection techniques as well. Despite having access to all measurements, the LOF method only accurately identified the second (speeding) and fourth (wrong-way driving) trajectories (Fig. 10), failing to detect the anomaly of a vehicle that suddenly stopped. Moreover, the progression of LOF curves as measurements became sequentially available indicated a susceptibility to false alarms with this approach. For instance, even during normal trajectory, false alarms were triggered at timesteps 65 and 96 when the percentile rank curve crossed the 5% threshold line.

E. Anomaly Detection with Bayesian Method

To illustrate the efficacy of the proposed Bayesian anomaly detector (referenced in Eq. (37)), we employ the same test trajectories outlined in the preceding section. Fig. 11(a) depicts the time-varying p-value. It is evident that, upon receiving complete observations, all anomalies were accurately classified. Contrastingly, a comparison with Fig. 9 indicates a relatively reduced reactivity of the Bayesian method, and
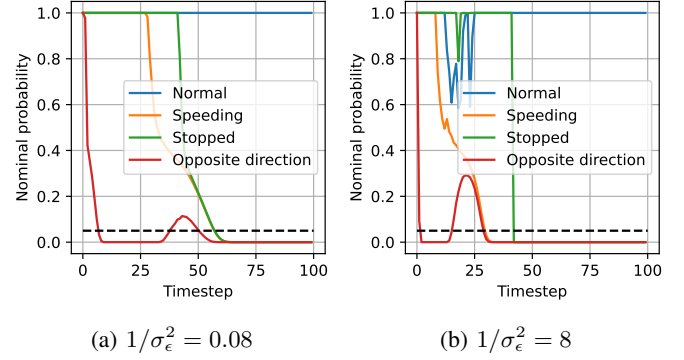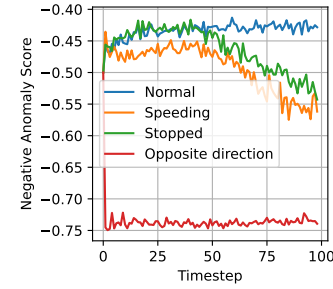
increase in susceptibility to false alarms. Nonetheless, the performance significantly surpasses the baseline LOF, marking a substantial improvement.

For example, the vehicle made an abrupt stop at timestep 40. While the confidence level started to decline approximately at the same time, the system didn't trigger an alarm until timestep 56, resulting in a delay of 16 timesteps. This delay is notably longer than the 6 timestep delay demonstrated by the marginal likelihood method. The identification of a vehicle driving in the wrong direction also experienced a 4 timestep delay, along with a false negative period spanning from timestep 36 to timestep 50.

Performance trends become apparent when comparing various values of σ_ϵ^2. Decreasing σ_ϵ^2 is entailed by faster convergence of Σ_a^* to zero. This, in turn, allows the distance between the transformed latent representation and the origin to be estimated more accurately, resulting in a more reactive detector. However, this sensitivity escalates the likelihood of false alarms, as can be observed from Fig. 11(b), where there are notable dips in confidence early in the time series, and the false negative period of the wrong-way driving trajectory lasts longer and hits higher confidence peak.

The Isolation Forest algorithm (Fig. 12) generally exhibits good detection performance alongside a reduced false alarm rate. However, the initial overlapping anomaly score values suggest that the initial dozens of timesteps can be disregarded. Interestingly, the system accurately identified instances of opposite direction driving from the onset. Furthermore, it

TABLE II: Comparisons between different anomaly detection methods. AT: alarm time, FD: maximum false negative time period.

Method	Normal		Speeding		Stopped		Opposite	
	AT	FD	AT	FD	AT	FD	AT	FD
Marginal Likelihood	∞	0	35	0	46	0	1	0
Bayesian	∞	0	56	0	56	0	10	14
LOF	65; 96	4	25	0	48	50	1	0
Isolation Forest	∞	0	10	0	65	0	1	0

correctly classified a suddenly stopped vehicle with approximately a 25 timestep delay, at around timestep 65. Regarding the detection of a speeding vehicle, the anomaly was recognized as early as timestep 10. This performance significantly outpaces the marginal likelihood and Bayesian detectors, which only activated at timesteps 35 and 56 respectively.

Table II provides a comprehensive comparison of the selected methods. Generally, both the marginal likelihood and Bayesian methods exhibit excellent detection rates. However, the Isolation Forest method surpasses them in terms of responsiveness. It's essential to remember that the FPCA-based detector incorporates measurement noise into its computation, necessitating a greater number of observations for a reliable testing of hypothesis H_0.

VI. Discussion

This paper introduces a comprehensive unsupervised framework designed for identifying abnormal road vehicle trajectories. The framework draws on two probabilistic methods: the marginal likelihood and the Bayesian posterior estimate. Detailed formulas pertaining to the Gaussian prior latent representation case and the Gaussian white noise measurement model, were presented. The entailed in-depth mathematical analysis provided grounds for a general anomaly detection framework for PCA and other representational methods, such as [28]. This augments existing literature by offering a framework that can carry out real-time, online anomaly classification, in contrary to using a sliding window. Numerical simulations underscored the efficacy of the system in promptly detecting common road anomalies like speeding or wrong-way driving with a mere 4-16 time steps delay.

The marginal likelihood method and the Bayesian method operate within two distinct domains. The former approximates nominal behaviors and reconstructs them within the time domain, performing anomaly classification there. Given that the FPCA transformation is linear, this can be equated to the marginal likelihood method detecting anomalies throughout the entire $\mathbb{R}^n$ space. Conversely, the Bayesian method seeks to classify anomalies within the representation domain. This is akin to projecting observations onto a subspace $\mathbb{R}^m, m < n$, and calculating the anomaly score subsequently. Through this interpretation, it becomes clear that the marginal likelihood method offers quicker responsiveness and is potentially less susceptible to false alarms, provided that the approximation or reconstruction (10) remains valid.

The framework presented in this paper is not confined solely to the trajectories of road vehicles. Instead, it can be adapted to a variety of time series data, including aircraft trajectories in aviation or the detection of malicious attacks in computer networking.

Nevertheless, the framework has several identifiable shortcomings. One such limitation is that the formulas provided are only applicable in cases where there's a Gaussian prior over the latent representations. While this is a reasonable assumption for situations with a singular traffic mode, like highways, more complex scenarios such as road merges or intersections may necessitate modeling with a Gaussian mixture. The unsupervised nature of the system may also be perceived as a potential limitation. While the detector can discern if an observed trajectory deviates from what it learned from the training set, it lacks the capacity to explicitly categorize anomalies into specific groups such as wrong-way driving or speeding. It is projected that a machine learning classifier could be instrumental in addressing this scenario. Moreover, similar to the Fourier Transform, FPCA experiences the "window effect" when applied to finite time series, which means the approximation power diminishes near the trajectory's endpoint. This factor should be duly considered when constructing the road patch P.

VII. Conclusion

In this paper, we have introduced a novel unsupervised anomaly detection framework and provided an in-depth exploration of its mathematical properties. Our numerical simulations have convincingly demonstrated that this method exhibits a robust capability for detecting a wide array of anomalous driving behaviors, without the need for explicit teaching. Despite its current limitation to unimodal data, the framework offers significant potential in identifying anomalies within time series data.

Recognizing this limitation to unimodal data, our future work will be devoted to addressing this issue to extend the applicability of the framework. The proposed anomaly detection framework is sufficiently versatile to be relevant to a diverse range of other disciplines, such as aviation and computer networking. It is particularly pertinent to fields associated with human safety, where the requirement for a rapid, responsive detection system is paramount.

References

[1] C. Englund, E. E. Aksoy, F. Alonso-Fernandez, M. D. Cooney, S. Pashami, and B. Åstrand, "AI perspectives in Smart Cities and Communities to enable road vehicle automation and smart traffic control," *Smart Cities*, vol. 4, no. 2, pp. 783–802, 2021.

[2] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–33, 2021.

[3] Y. Li, W. Liu, and Q. Huang, "Traffic anomaly detection based on image descriptor in videos," *Multimedia tools and applications*, vol. 75, no. 5, pp. 2487–2505, 2016.

[4] R. Hinami, T. Mei, and S. Satoh, "Joint detection and recounting of abnormal events by learning deep generic knowledge," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3619–3627.

[5] S. Lee, H. G. Kim, and Y. M. Ro, "STAN: Spatio-temporal adversarial networks for abnormal event detection," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 1323–1327.

[6] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2720–2727.

[7] H. Y. Ngan, N. H. Yung, and A. G. Yeh, "Outlier detection in traffic data based on the Dirichlet process mixture model," *IET Intelligent Transport Systems*, vol. 9, no. 7, pp. 773–781, 2015.

[8] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, "Learning deep representation of appearance and motion for anomalous event detection. arXiv," *arXiv preprint arXiv:1510.01553*, 2015.

[9] Z. Rahman, A. M. Ami, and M. A. Ullah, "A real-time wrong-way vehicle detection based on YOLO and centroid tracking," in *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 2020, pp. 916–920.

[10] S. Simpson, D. Bruggeman *et al.*, "Detection and warning systems for wrong-way driving," Arizona. Dept. of Transportation, Tech. Rep., 2015.

[11] M. A. Khan and S. F. Khan, "IoT based framework for Vehicle Over-speed detection," in *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*. IEEE, 2018, pp. 1–4.

[12] Y. Marumo, H. Tsunashima, T. Kojima, and Y. Hasegawa, "Analysis of braking behavior of train drivers to detect unusual driving," *Journal of Mechanical systems for Transportation and Logistics*, vol. 3, no. 1, pp. 338–348, 2010.

[13] Z. Zhang, X. Mei, and B. Xiao, "Abnormal event detection via compact low-rank sparse learning," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 29–36, 2015.

[14] Y. Zhang, H. Lu, L. Zhang, X. Ruan, and S. Sakai, "Video anomaly detection based on locality sensitive hashing filters," *Pattern Recognition*, vol. 59, pp. 302–311, 2016.

[15] R. Song, "Driver intention prediction using model-added Bayesian network," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 235, no. 5, pp. 1236–1244, 2021.

[16] J. Colyar and J. Halkias, "US highway 101 dataset. Federal Highway Administration Research and Technology Fact Sheet. Publication number: FHWA-HRT-07-030," *Tech. Rep.*, 2007.

[17] R. Chandra, U. Bhattacharya, T. Mittal, A. Bera, and D. Manocha, "Cmetric: A driving behavior measure using centrality functions," *arXiv preprint arXiv:2003.04424*, 2020.

[18] N. Li, D. W. Oyler, M. Zhang, Y. Yildiz, I. Kolmanovsky, and A. R. Girard, "Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems," *IEEE Transactions on control systems technology*, vol. 26, no. 5, pp. 1782–1797, 2017.

[19] V. Ramanishka, Y.-T. Chen, T. Misu, and K. Saenko, "Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7699–7707.

[20] E. Cheung, A. Bera, E. Kubin, K. Gray, and D. Manocha, "Identifying driver behaviors using trajectory features for vehicle navigation," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3445–3452.

[21] A. Geiger, D. Liu, S. Alnegheimish, A. Cuesta-Infante, and K. Veeramachaneni, "TadGAN: Time series anomaly detection using generative adversarial networks," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 33–43.

[22] A. Kuefler, J. Morton, T. Wheeler, and M. Kochenderfer, "Imitating driver behavior with generative adversarial networks," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 204–211.

[23] B. Zhao, L. Fei-Fei, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in *CVPR 2011*. IEEE, 2011, pp. 3313–3320.

[24] V. Nguyen, S. Gupta, S. Rana, C. Li, and S. Venkatesh, "A Bayesian nonparametric approach for multi-label classification," in *Asian conference on machine learning*. PMLR, 2016, pp. 254–269.

[25] C. Li, Z. Han, Q. Ye, and J. Jiao, "Visual abnormal behavior detection based on trajectory sparse reconstruction analysis," *Neurocomputing*, vol. 119, pp. 94–100, 2013.

[26] S. W. Liu, H. Y. Ngan, M. K. Ng, and S. J. Simske, "Accumulated relative density outlier detection for large scale traffic data," *Electronic Imaging*, vol. 2018, no. 9, pp. 239–1, 2018.

[27] X. Jin, Y. Zhang, L. Li, and J. Hu, "Robust PCA-based abnormal traffic flow pattern isolation and loop detector fault detection," *Tsinghua Science & Technology*, vol. 13, no. 6, pp. 829–835, 2008.

[28] G. Jarry, D. Delahaye, F. Nicol, and E. Feron, "Aircraft atypical approach detection using functional principal component analysis," *Journal of Air Transport Management*, vol. 84, p. 101787, 2020.

[29] J. Owens and A. Hunter, "Application of the self-organising map to trajectory classification," in *Proceedings Third IEEE International Workshop on Visual Surveillance*. IEEE, 2000, pp. 77–83.

[30] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 9, pp. 1450–1464, 2006.

[31] X. Wang, K. Tieu, and E. Grimson, "Learning semantic scene models by trajectory analysis," in *European conference on computer vision*. Springer, 2006, pp. 110–123.

[32] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2000, pp. 285–289.

[33] J. O. Ramsay and B. W. Silverman, *Applied functional data analysis: methods and case studies*. Springer, 2002, vol. 77.

[34] Y. Hong, "scikit-fda: Principal Component Analysis for Functional Data," B.S. thesis, 2020.

[35] J. Ramsay and B. Silverman, "Principal components analysis for functional data," *Functional data analysis*, pp. 147–172, 2005.

[36] V. Alexiadis, "Video-Based Vehicle Trajectory Data Collection," in *Transportation Research Board 86th Annual Meeting*. Citeseer, 2006.

[37] CAR 2 CAR Consortium. Survey on ITS-G5 CAM statistics. [Online]. Available: https://www.car-2-car.org/

[38] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

APPENDIX

A. Proof of Proposition IV.3

We rename Φ to $S_n = \Phi = \sum_{i=1}^{n} \Phi^i \Phi^{i\top}$ for emphasis on the evolution of Φ with time. From (27), $U = S_n R, V = R^T S_n R$.

$$\frac{1}{\sigma_\epsilon^2} \left(U^\top \frac{\Sigma^*}{\sigma_\epsilon^2} U - V \right) = \frac{1}{\sigma_\epsilon^2} R^\top \left(S_n^\top \Sigma^* \frac{1}{\sigma_\epsilon^2} - I \right) S_n R$$

$$= \frac{1}{\sigma_\epsilon^2} R^\top \left(S_n^\top \left(\frac{1}{\sigma_\epsilon^2} S_n + \Sigma_a^{-1} \right)^{-1} \frac{1}{\sigma_\epsilon^2} - I \right) S_n R$$

As Σ_a^{-1} is positive definite, we can perform an eigen-decomposition $\Sigma_a^{-1} = Q \Lambda Q^\top$ and use the matrix inversion lemma:

$$\left(\frac{1}{\sigma_\epsilon^2} S_n + \Sigma_a^{-1} \right)^{-1} = \left(\frac{S_n}{\sigma_\epsilon^2} \right)^{-1} - \left(\frac{S_n}{\sigma_\epsilon^2} \right)^{-1}$$

$$Q \left(\Lambda^{-1} + Q^\top \left(\frac{S_n}{\sigma_\epsilon^2} \right)^{-1} Q \right)^{-1} Q^\top \left(\frac{S_n}{\sigma_\epsilon^2} \right)^{-1}$$

Hence:

$$\frac{1}{\sigma_\epsilon^2} \left(U^\top \frac{\Sigma^*}{\sigma_\epsilon^2} U - V \right) = -RQ \left(\Lambda^{-1} + Q^\top \sigma_\epsilon^2 S_n^{-1} Q \right)^{-1} Q^\top R$$

and as $S_n \to \infty$, combined with Q being full rank:

$$\frac{1}{\sigma_\epsilon^2} \lambda_1 \left(U^\top \frac{\Sigma^*}{\sigma_\epsilon^2} U - V \right) \to \lambda_1(-R\Sigma_a^{-1} R) = -1,$$

by definition of $R = \Sigma_a^{1/2}$. Proposition IV.3 follows.

Thinh Hoang Dinh is currently pursuing his PhD at the ENAC Optimization Laboratory in Toulouse, France. His research focuses on the application of artificial intelligence in the field of transportation, aiming to improve efficiency and sustainability within this sector.

Vincent Martinez received the M.Sc. degree in mathematical and statistical modeling from INSA Toulouse, France. He is working as a Senior Principal Engineer at NXP Semiconductors since 2006. He is currently a Wireless Signal Processing and Physical Layer Expert, and is contributing to various V2X standardization bodies.

Pierre Maréchal is Professor of Mathematics at the University of Toulouse III Paul Sabatier. Although his initial training was in engineering, he obtained a PhD in Image processing in 1997, and then turned to mathematics through a postdoc at Simon Fraser University. His areas of interest are mainly, but not exclusively, optimization and inverse problems.

Daniel Delahaye leads the Optimization and Machine Learning Team at ENAC Research Laboratory and chairs the "AI for ATM and Large Scale Urban Mobility" research at ANITI in Toulouse. After earning his engineering degree and master's from ENAC and National Polytechnic Institute of Toulouse, he obtained his Ph.D. in automatic control and later tenure in applied mathematics at ENAC. He has worked at the French Civil Aviation Study Center and MIT. His research, conducted in collaboration with MIT, Georgia Tech, and NASA, centers on AI and mathematical optimization for airspace and aircraft trajectory optimization.