HKF: Hierarchical Kalman Filtering With Online Learned Evolution Priors for Adaptive ECG Denoising

Guy Revach, Senior Member, IEEE, Timur Locher, Nir Shlezinger, Senior Member, IEEE, Ruud J. G. van Sloun, Member, IEEE, and Rik Vullings, Member, IEEE

Abstract—Electrocardiography (ECG) signals play a pivotal role in many healthcare applications, especially in at-home monitoring of vital signs. Wearable technologies, which these applications often depend upon, frequently produce low-quality ECG signals. While several methods exist for ECG denoising to enhance signal quality and aid clinical interpretation, they often underperform with ECG data from wearable technology due to limited noise tolerance or inadequate flexibility in capturing ECG dynamics. This paper introduces HKF, a hierarchical and adaptive Kalman filter, which uses a proprietary state space model to effectively capture both intra- and inter-heartbeat dynamics for ECG signal denoising. HKF learns a patient-specific structured prior for the ECG signal’s intra-heartbeat dynamics in an online manner, resulting in a filter that adapts to the specific ECG signal characteristics of each patient. In an empirical study, HKF demonstrated superior denoising performance (reduced Mean-Squared Error) while preserving the unique properties of the waveform. In a comparative analysis, HKF outperformed previously proposed methods for ECG denoising, such as the model-based Kalman filter and data-driven autoencoders. This makes it a suitable candidate for applications in extramural healthcare settings.

Index Terms—Electrocardiography, expectation-maximization, kalman filters.

I. INTRODUCTION

In the pursuit of halting the increase in healthcare costs and creating a sustainable healthcare system, it is increasingly important to monitor patients outside the hospital environment. In recent years, various technologies have emerged for remote and ambulatory monitoring of vital signs. Notably, wearable Electrocardiography (ECG) monitoring devices have gained prominence [2], [3]. The ECG reflects the electrical activity of the heart and is crucial for monitoring cardiac functions and diagnosing potential pathologies. It plays a key role in the clinical detection of severe heart conditions such as coronary heart diseases, heart attacks, and arrhythmias, which can lead to stroke [4], [5], as well as fetal asphyxia during labor [6]. Specific segments of the ECG waveform are critical diagnostic tools used by cardiologists to assess a patient’s condition or its deterioration. These waveform characteristics may vary depending on the application. For instance, an elongated QT interval is a known risk factor for cardiac arrhythmias [7], [8]. The ST segment often provides vital information for detecting myocardial infarction or fetal hypoxia [6], [9]. Elevated T-wave amplitudes can indicate a hypoxic state [10], and the occurrence of negative T-wave amplitude may indicate compromised cardiac performance [11], [12]. Noisy ECG signals typically obscure such critical features, making accurate extraction challenging, if not impossible. Therefore, it is essential that the ECG recording is as clear as possible, emphasizing the need for robust and effective ECG denoising methods to ensure accurate diagnostic outcomes [13].

In the realm of ECG denoising, recent literature presents a variety of signal-processing approaches. These range from classical techniques such as infinite impulse response (IIR) and Finite Impulse Response (FIR) filters to non-parametric methods like the wavelet transform and Empirical Mode Decomposition (EMD). Additionally, model-based techniques build upon predefined statistical models that capture the intrinsic characteristics of the ECG waveform. At the cutting edge, deep neural network (DNN) architectures harness vast amounts of data to learn an empirically optimal process for denoising an ECG signal [14], [15], [16], [17], [18].

Classical methods such as IIR and FIR filters have been cornerstone techniques for denoising signals, including ECG...
recordings [14]. These filters are favored for their computational efficiency and robust performance under conditions where the characteristics of both the signal and the noise are well-understood and consistent. IIR filters are particularly valued for their efficiency in terms of computational resources, as they require fewer coefficients to achieve a specific filtering effect compared to FIR filters. However, IIR filters are often preferred in scenarios requiring phase linearity to avoid distorting the signal, a critical consideration in ECG analysis.

In clinical and laboratory settings, where external noise sources are minimal, and noise characteristics such as electrical interference or baseline wander are consistent, these classical filters excel. They provide reliable denoising capabilities, enhancing the signal quality necessary for accurate diagnosis and monitoring without the need for adaptive processing. This makes IIR and FIR filters particularly suitable for routine diagnostic tests and basic heart rate monitoring, where environmental conditions and the types of noise interference remain relatively controlled.

However, under more complex and dynamic conditions, where noise characteristics are neither stable nor consistent, classical filtering methods may not suffice. In such environments, intermediate techniques such as EMD and wavelet-based methods offer greater flexibility [19]. EMD [20] adapts to the inherent scales within the data, making it highly suitable for non-linear and non-stationary signals like ECG, where it can effectively isolate and remove noise while preserving crucial signal features [21], [22], [23]. Similarly, wavelet transforms [24], [25], [26] analyze the signal at multiple scales or resolutions, which is particularly useful for identifying and denoising transient noise components while maintaining important aspects of the ECG waveform [27], [28], [29], [30], [31]. While their computational efficiency makes them suitable for real-time clinical applications, their success heavily relies on the correct choice of wavelet and its parameters, and incorrect tuning can lead to ECG distortion. Moreover, in the face of heavy noise, wavelets might not be fully effective.

At-home ECG monitoring often suffers from reduced signal quality compared to in-hospital settings. Wearable garments with embedded electrodes, commonly used for these recordings, generally produce signals with more noise and artifacts than the adhesive electrodes typically employed in hospitals [32]. This issue is compounded in environments where ECG data is subject to varied and unpredictable factors, such as patient movement in ambulatory devices, diverse activities monitored by wearable fitness technology, and complex noise conditions in intensive care units. These scenarios pose significant challenges, particularly during physical activities like exercise stress tests or in veterinary applications where animal movements create unpredictable noise profiles.

While classical methods like IIR and FIR filters have their place, the complexities of modern ECG scenarios often surpass their capabilities. Despite the effectiveness of EMD and wavelet-based methods in handling more complex signal scenarios, our focus shifts towards more advanced techniques capable of addressing specific challenges such as Additive Gaussian Noise (AGN). This type of noise, where the bandwidth overlaps with that of the signal and is not stationary, necessitates the use of methods that can adapt to real-time changes in noise and signal characteristics.

Advanced filtering techniques, particularly adaptive filters such as machine learning approaches and Kalman Filters (KFs), are invaluable in these settings. They not only adjust dynamically to changing conditions but also ensure high fidelity in ECG recordings, which is crucial for accurate health assessments and decision-making across a variety of challenging environments. The evolution of ECG denoising in dynamic settings thus necessitates the adoption of these sophisticated, adaptive denoising techniques, effectively bridging the gap left by classical methods and enhancing diagnostic outcomes.

Deep Learning (DL)-based approaches have revolutionized various fields by training DNNs end-to-end to minimize a loss function, such as the commonly used Mean-Squared Error (MSE), with vast datasets [33]. Specifically in ECG denoising, powerful tools have emerged [34], such as recurrent neural networks [35] and fully convolutional denoising Auto-Encoders (AEs) [36]. Additionally, deep convolutional neural networks (CNNs) have been applied to multi-channel fetal ECG denoising [37], and generative adversarial network architectures have been explored [38]. Another innovative approach involves a DNN-based framework that utilizes stacked cardiac cycle tensors [38]. These advancements underscore DL’s capability to handle complex and varied signal-processing tasks effectively.

Despite these advantages, DNNs present significant challenges. Their performance heavily depends on the quality and diversity of the training data, complicating their tailoring to unique, real-time ECG characteristics of individual, unseen patients. This reliance on datasets that encompass multiple patients often leads to models developing a bias towards a “universal mean waveform” when trained to minimize common loss functions, like MSE. Such bias may obscure vital patient-specific ECG features or introduce artifacts in real-time denoising scenarios, potentially degrading performance. Moreover, medical diagnosis is a conservative field that must adhere to stringent standards and regulations, and black-box DL-based solutions, often not based on known models, lack interpretability and come without guarantees. This makes them challenging to use in contexts where understanding the decision-making process is crucial. Furthermore, these networks have a relatively high computational load that may further limit their application in real-time embedded systems with limited computational resources, such as wearables.

Local approximations of the ECG waveforms have also been explored through windowed state space (SS) models, as in [39], [40], [41], and via autoregressive models [42], [43]. Nevertheless, these approximations can sometimes fall short of capturing the intricate intra-heartbeat dynamics and the inherent quasi-periodicity between consecutive Heartbeats (HBs), posing challenges for optimal denoising.

Model-based techniques that utilize SS models and variations of Bayesian filters [44], such as the KF [45], have become notable alternatives for ECG denoising. The work in [46], [47], [48] proposes a denoising method for single-channel ECG recordings, while [49] aims to denoise multi-channel signals.
A key aspect of these methodologies is their use of models that describe the signal’s dynamics, which are usually complex and quasi-periodic and are used as evolution priors within the Bayesian framework. This enhances both interpretability and reliability. Unlike DNNs, these models not only provide deep insights but also offer theoretical assurances, building trust among medical practitioners. For instance, [50] introduces non-linear cyclic priors using Partial-Differential-Equation (PDE) models and sums of Gaussians. The study in [46] employs the dynamic model introduced in [50] within a nonlinear Bayesian filtering framework, utilizing 2 state variables, while later research in [47] expanded this model to include 17 state variables, focusing on the quasi-periodicity of heartbeats. The implementation of PDE models and sums of Gaussians requires meticulous tuning to effectively capture the dynamic nature of ECG signals. This complexity poses significant challenges in practical applications and limits flexibility. Therefore, despite their innovative approaches, these models often struggle to accurately capture patient-specific variations. To address this, [51] attempts to automatically fit model parameters using a Least-Squares (LS) optimizer based on several pre-recorded heartbeats. Moreover, the work in [48] seeks to refine the morphological model of cardiac signals, as described in [50], [52], by moving away from the quasi-periodicity assumption and continuously adapting and updating the model’s parameters. Utilizing a nonlinear sum of Gaussians to describe the ECG waveform requires the use of nonlinear Bayesian methods, such as the Extended (EKF) and Unscented (UKF). These methods introduce additional computational complexities and may degrade performance due to the inherent sub-optimality of nonlinear models. Moreover, the real-time implementation of these complex Bayesian filters poses significant challenges on resource-constrained platforms, exacerbated by high computational demands and sensitivity to initial parameter settings and noise characteristics. The research in [53] incorporates the Expectation-Maximization (EM) algorithm to determine the evolution function. When combined with a bank of KFs, this method efficiently targets both high and low-frequency noise. However, its application is constrained by its reliance on a linear prior function, its restriction to filtering signals of fixed length, and its omission of the ECG’s periodic information. The work in [49] omits the within heartbeat evolution and only represents the evolution of consecutive heartbeats with an identity function, based on the premise that, in the absence of arrhythmia, two successive heartbeats centered around the R-peak closely resemble each other. While this strategy enhances the Signal to Noise Ratio (SNR), it also runs the risk of obscuring vital physiological dynamics.

In this work, we propose a Hierarchical Kalman Filter (HKF) for ECG denoising, which enhances signal quality without obscuring dynamic changes potentially linked to pertinent pathophysiology. Our HKF is designed based on our innovative hierarchical SS model, which describes the ECG signal dynamics both within individual heartbeats and across consecutive heartbeats. Specifically, HKF consists of an online learned structured evolution prior for a single heartbeat; a Rauch-Tung-Striebel (RTS) intra-heartbeat smoother [54] that harnesses this prior; and an inter-heartbeat KF [45] for denoising spanning multiple heartbeats. The online warm-up phase is meticulously designed to tackle challenges such as the highly patient-specific heartbeat shape and substantial noise variation resulting from myriad factors, ranging from equipment intricacies to room temperature variations. While one might conceptualize a typical ECG signal shape, the reality is that there’s considerable inter-variability among patients. Not only can the signal shape vary significantly between patients, but there’s also intra-patient variability: the placement and orientation of electrodes can introduce alterations in the observed waveform. This renders the task of crafting a universal prior quite challenging. Crucially, HKF doesn’t require supervised pre-training and is inherently patient-adaptive due to its online covariance estimation and its learned structured evolution prior. Yet, it preserves the transparent and interpretable nature of the KF.

Our experimental study shows that the proposed HKF effectively denoises ECG signals, even in challenging setups, while retaining the subtle, clinically valuable structures within the signals. These attributes make it especially suited for medical and healthcare applications where a high degree of confidence and reliability is essential.

The remainder of this paper is organized as follows: Section II formulates the task and introduces the hierarchical SS model. Section III delves into the details of the proposed HKF. Section IV elaborates on parameter estimation. Finally, Section V presents our empirical study, demonstrating that the HKF surpasses both model-based and data-driven benchmarks.

II. SYSTEM MODEL AND TASK FORMULATION

In this section, we lay the groundwork for the subsequent derivation of the HKF. We begin with a review of essential information on the ECG signal. Following that, we delve into the task of multi-channel ECG signal denoising. We conclude by presenting our unique hierarchical SS model, which serves as the foundation for the HKF design.

A. The ECG Signal

The heart is composed of two main types of muscle: the atrial muscle and the ventricular muscle, as illustrated in Fig. 1(a). During a HB, these different muscles contract and relax in response to electrical impulses, which depolarize and repolarize the heart. These impulses propagate as an electrical field through the body and can be detected by electrodes on the skin. The voltage variation recorded over time is what we refer to as the ECG signal. The typical ECG signal comprises three segments: the P-wave, the QRS complex, and the T-wave, as shown in Fig. 1(b). Each of these segments represents a different stage of heart contraction. The P-wave corresponds to the contraction of the atrial muscle; the QRS complex indicates ventricular depolarization, and the T-wave reflects ventricular repolarization. As its name suggests, the QRS complex consists of three smaller waves (Q-wave, R-wave, and S-wave) associated with the depolarization of the ventricular muscle. The entire QRS complex cycle takes about 100 msec [56]. While a typical
The ECG denoising task is defined as the reconstruction of the underlying clean signal \( x_t \) across \( m \)-channels. This reconstruction is carried out by a denoiser \( \Psi \), which utilizes all available information, including past and current noisy observations, to output \( \hat{x}_t \) as an estimate for \( x_t \). The denoiser is defined by the mapping:

\[
\Psi : \{y_i\}_{i=1}^f \mapsto \hat{x}_t.
\]

An optimal denoiser \( \Psi^* \) from a class of denoisers \( \Psi \), is the one that minimizes a specific cost function \( C(\hat{x}_t, x_t) \) between the estimated denoised value \( \hat{x}_t \) and the clean signal \( x_t \). A natural choice for this function is the commonly used MSE:

\[
\Psi^* = \arg \min_{\Psi \in \Psi} \{ \mathbb{E}[\|\hat{x}_t - x_t\|^2] \}.
\]

Describing a functional \( \Psi^* \) that can optimally denoise the signal for every discrete-time \( t \) and is also computationally efficient is not trivial. Consequently, we assume that the signal can be described using a hierarchical SS model, as elaborated in Subsection II-D. Building on this, we aim to achieve an approximately MSE optimal denoising by employing a combination of the Kalman Smoother (KS) and the KF, as detailed in Section III.

D. Hierarchical System Model

The relationship between the observed ECG signal \( y_t \) to its corresponding noiseless instance \( x_t \) can be modeled as a dynamical system. A canonical way to model dynamical systems in discrete-time is by using SS models [58]. SS models blend well with numerous filtering and smoothing techniques and can accommodate the integration of both physical and data-driven learned models. To exploit the periodicity, which plays a key role in the ECG signal, we model the signal dynamics using a hierarchical SS model.

In the following, we make a reasonable assumption that individual HBs can be accurately segmented. Therefore, we divide the signal into periodic segments of length \( T \), where each such segment represents a single HB. Our model draws inspiration from the decimated components decomposition method, which is used to represent periodic systems and cyclostationary signals [59] as multivariate (tensor) stationary ones. The dynamics within a single HB are described using an intra-HB (internal) SS model, while the dynamics between two consecutive Heartbeats are modeled using an inter-HB (external) SS model, as detailed next.

E. Intra Heartbeat Modeling

The intra-HB SS model defines the dynamics within a single HB with index \( \tau \). For each time step \( t \) within the period \( T \triangleq \{1, \ldots, T\} \) of length \( T \), the model is given by:

\[
\begin{align*}
x_{\tau,t} &= f_{\tau} (x_{\tau,t-1}) + e_{\tau,t}, \quad e_{\tau,t} \sim \mathcal{N}(0, Q_{\tau}), \\
y_{\tau,t} &= x_{\tau,t} + v_{\tau,t}, \quad v_{\tau,t} \sim \mathcal{N}(0, R_{\tau}).
\end{align*}
\]

The matrices \( Q_{\tau} \) and \( R_{\tau} \) represent the time-varying intra-evolution (process) and observation covariance of a Gaussian distribution, respectively. They capture the correlation across multiple channels. Specifically, \( Q_{\tau} \) accounts for the correlation arising from the fact that these channels originate from the same foundational heart activity. In contrast, \( R_{\tau} \) captures the correlation attributed to shared measuring effects: patient-related factors such as breathing and movement, measuring-device-related factors such as quality and age, and environmental factors, including electromagnetic interference and ambient room temperature. As a result, these matrices are assumed to be unknown and are not restricted to be diagonal.
We denote the $\tau$-th HB and its noisy observation as the matrices $X_\tau$ and $Y_\tau$, respectively, namely,
\begin{align}
X_\tau &= [x_{\tau,1}, \ldots, x_{\tau,T}] \in \mathbb{R}^{m \times T}, \\
Y_\tau &= [y_{\tau,1, \ldots, y_{\tau,T}] \in \mathbb{R}^{m \times T}.
\end{align}

F. Inter Heartbeat Modeling

The inter-HB SS model defines the evolution between two consecutive HBs, labeled with indices $\tau - 1$ and $\tau$, namely:
\begin{align}
x_{\tau,t} &= F(x_{\tau-1,t}) + \epsilon_{\tau,t}, \quad \epsilon_{\tau,t} \sim N(0, Q_{\tau,t}), \\
\tilde{y}_{\tau,t} &= x_{\tau,t} + \nu_{\tau,t}, \quad \nu_{\tau,t} \sim N(0, R_{\tau,t}).
\end{align}

Based on the periodicity of the cardiac vector, here, we assume that two consecutive HBs closely resemble each other. Therefore, the inter-state evolution is defined as the identity mapping, i.e., $F(x) = x$. The matrices $Q_{\tau,t}$ and $R_{\tau,t}$ represent the inter-evolution (process) and observation covariance, respectively. These matrices capture the multi-channel correlation, and they adhere to the same assumptions outlined for the inter-HB model.

\begin{align}
Y_\tau &= [\tilde{y}_{\tau,1, \ldots, \tilde{y}_{\tau,T}] \in \mathbb{R}^{m \times T}.
\end{align}

An illustrative overview of the system model can be found in Fig. 2.

III. HKF: HIERARCHICAL KALMAN FILTER

Our custom-designed HKF exploits the proposed hierarchical SS model to perform patient-dependent denoising.

A. Overview and Design Rationale

There are two main phases to the algorithm:

- P1 A short online warm-up phase.
- P2 A signal-denoising processing phase.

In P1, the online warm-up phase, we learn the internal parameters of the inter-HB SS model (3), specifically: the prior signal evolution model and the noise covariance matrices. The specific HB shape varies considerably among patients and can be influenced by numerous factors, ranging from physiological to equipment-related and ambient factors. Consequently, pre-training an algorithm on a dataset from multiple patients results in learning an average heartbeat waveform. To avoid this, our algorithm is designed to adapt to individual patients, learning the signal model in an unsupervised manner. In particular, we employ Taylor LS, detailed in Subsection IV-A, for the evolution function approximation. Meanwhile, the unknown noise covariance matrices are estimated using a variant of the EM algorithm, as elaborated upon in Subsection IV-C.

In the signal-denoising processing, denoted as P2, we utilize the mapping $\text{Ψ}$ to transform the current observed HB, $Y_\tau$, into a denoised HB, $\hat{X}_\tau$. This denoising is accomplished by fusing $Y_\tau$ with all previously observed information, $\{Y_{\tau'}\}_{\tau' = 1}^{T}$, namely:
\begin{align}
\text{Ψ}: Y_{\tau'} \rightarrow \hat{X}_\tau.
\end{align}

However, in practice, $\text{Ψ}$ operates using the previous posterior $X_{\tau-1}$ as a sufficient statistic, which efficiently captures all accumulated information, as follows:
\begin{align}
\text{Ψ}: \hat{X}_{\tau-1}, Y_\tau \rightarrow \hat{X}_\tau.
\end{align}

This Markovian assumption is fundamental for Kalman filters, thus establishing a recursive and computationally efficient framework that bypasses the need to directly reprocess observed data. This strategy is underpinned by the SS model outlined earlier and emphasizes the model’s adaptability and our methodological rigor in developing a scalable solution. Here, for enhanced clarity, we denote $X_\tau$ as $\hat{X}_\tau$, which represents the mean of the posterior distribution, given $\tau$ HBs.

In practice, the denoising is implemented in two steps as in the following:

- D1 Intra-HB Kalman smoothing.
- D2 Inter-HB Kalman filtering.

In D1, the RTS smoother [44], [54] is used to compute $\hat{X}_\tau$, an intermediate denoised version, from $Y_\tau$, in a stand-alone manner, based on the inter-HB SS model (3), given the learned prior parameters. More specifically:
\begin{align}
\text{Ψ}: Y_\tau \rightarrow \hat{X}_\tau = [\hat{x}_{\tau,1:T}, \ldots, \hat{x}_{\tau,T:T}] \in \mathbb{R}^{m \times T}.
\end{align}

Here, the notation $\hat{x}_{\tau,t|t}$ denotes the smoothed version of sample $t$ in HB $\tau$ given all $T$ samples from the entire HB. This stage is detailed in Subsection III-B.

In D2, the KF is used to compute $\hat{X}_\tau$, the fully denoised posterior HB, by fusing $Y_\tau = \hat{X}_\tau$, i.e., the stand-alone smoothed version, with $\hat{X}_{\tau-1}$, the fully denoised posterior of the previous heartbeat. This stage is based on the inter-HB SS model (5), with an adaptive covariance estimation. More specifically:
\begin{align}
\text{Ψ}: \hat{X}_{\tau-1}, \hat{X}_\tau \rightarrow \hat{X}_\tau = [\hat{x}_{\tau|\tau-1,T}, \ldots, \hat{x}_{\tau|T,T}] \in \mathbb{R}^{m \times T}.
\end{align}

This operation represents a batch of $T$ parallel KFs. In this context, the additional notation $\tau | \tau$ denotes the update of the smoothed version, incorporating information from the current
filtering fuses the noisy observations from the current heartbeat with all available information from previous heartbeats.

Based on the intra-HB SS model parameters (3):

\[ \text{update} \]

and all previous HBs. Further details are provided in Subsection III-C.

Fig. 3 visualizes the high-level structure of the processing phase, while Fig. 4 is a high level description of our HKF algorithm.

B. First Stage Smoothing

In the processing phase, the first step, D1, involves implementing a KS using the RTS algorithm [54]. This KS operates based on the intra-HB SS model parameters (3): \( \hat{f}_t \), \( Q_t \), and \( \hat{R}_t \), learned during the warm-up phase. Given \( y \), the newly observed HB with index \( \tau \), we employ the KS, denoted as \( \hat{\Psi} \), to obtain an instantaneous estimate of the denoised HB, denoted as \( \hat{X}_\tau \).

The RTS smoother encompasses two-steps:

KS1 A forward pass step, i.e., a KF.

KS2 A backward pass step.

The forward pass, KS1, is defined by two phases: prediction and update\(^1\)

The prediction phase, is given by state prediction:

\[ \hat{x}_{t|t-1} = \hat{f}_t \left( \hat{x}_{t-1|t-1} \right), \quad P_{t|t-1} = P_{t-1|t-1} + Q_t, \quad (11) \]

\(^1\)For brevity we ignore the inter-HB index \( \tau \).

where \( P_{t|t-1} \), the prior covariance, is computed using an identity matrix, as defined in (31), and by innovation prediction:

\[ \Delta y_t = y_t - \hat{y}_{t|t-1} = y_t - \hat{x}_{t|t-1}, \quad S_t = P_{t|t-1} + R_t. \quad (12) \]

The update phase is given by:

\[ \hat{x}_{t|t} = \hat{x}_{t|t-1} + \hat{K}_t \cdot \Delta y_t, \quad P_{t|t} = P_{t|t-1} - \hat{K}_t \cdot S_t \cdot \hat{K}_t^T. \quad (13) \]

Here, \( \hat{K}_t \) is the Kalman gain, given by:

\[ \hat{K}_t = P_{t|t-1} \cdot S_t^{-1}. \quad (14) \]

The backward pass, KS2, of the RTS smoother is given by:

\[ \Delta x_t = \hat{x}_{t+1|T} - \hat{x}_{t+1|t}, \quad \Delta P_{t+1} = P_{t+1|T} - P_{t+1|t}. \quad (15) \]

Here, \( G_t \), the smoothing gain, is given by:

\[ G_t = P_{t|t} \cdot P_{t+1|t}^{-1}. \quad (17) \]

The output of the smoothing step is

\[ \hat{X}_\tau = \hat{\Psi} (\hat{q}_\tau) = [\hat{x}_{\tau, T|T}, \ldots, \hat{x}_{\tau, T|T}]. \quad (18) \]

We assume that the estimated distribution for the instantaneously denoised HB is given by

\[ x_{\tau, T|T} \sim N(\hat{x}_{\tau, T|T}, P_{\tau, T|T}). \quad (19) \]

This distribution is significant for the subsequent filtering state detailed in Subsection III-C. We will use \( \hat{x}_{\tau, \tau|T_T} \), the posterior mean, and \( P_{\tau, T|T} \), the posterior error covariance estimate, from the initial stage as observations and observation noise for the second stage, respectively. This approach not only eliminates the need for an additional estimation step but also offers the optimal estimate of the observation noise matrices under the assumption that all previously estimated values are optimal.

C. Second Stage Filtering

In the second step, D2, of the processing phase, we employ a KF based on the inter-HB SS model (5). Given the immediate estimate of the current HB, denoted as \( \hat{X}_\tau \), and the posterior estimate of the preceding HB accounting for its entire past \( \hat{X}_{\tau-1} \), we produce a posterior estimate for the current HB \( \hat{X}_\tau \). This filter effectively unfolds to TKFs operating simultaneously along the \( \tau \)-axis:

\[ \hat{\Psi}: \hat{x}_{\tau, T|T}, \hat{x}_{\tau-1, \tau-1|T_T} \rightarrow \hat{x}_{\tau, T|T}. \quad (20) \]

In the end, we leverage the temporal correlation between two successive HBs, \( \tau - 1 \) and \( \tau \), as outlined in (5). For every \( t \in T \) and given the estimated covariance matrices, we implement T independent (and parallel) KFs. This procedure is articulated in the ensuing steps; note that for simplicity, we occasionally omit the intra-HB index \( t \).

KF1 Prior error and innovation covariance:

\[ P_{\tau|\tau-1} = P_{\tau-1} + \hat{Q}_\tau, \quad S_\tau = P_{\tau|\tau-1} + \hat{R}_\tau. \quad (21) \]

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
The update equations:

\[ \hat{X}_{\tau} = \hat{X}_{\tau-1} + K_{\tau} \Delta, \quad P_{\tau} = P_{\tau|\tau-1} - K_{\tau} S_{\tau} K_{\tau}^\top. \] (22)

Here, \( \Delta \) is the innovation, and \( K_{\tau} \) is the Kalman gain:

\[ \Delta = \hat{X}_{\tau} - \hat{X}_{\tau-1} = (\hat{X}_{\tau,t|T} - \hat{X}_{\tau-1,t|T})_{\forall t}, \] (23)

\[ K_{\tau} = P_{\tau|\tau-1} S_{\tau}^{-1}. \] (24)

To successfully deploy the KF the set of inter-HB covariance matrices, i.e.,

\[ \{Q_{\tau,t}, R_{\tau,t}\}_{t=1}, \] (25)

should be estimated on the fly before the filtering step. The estimation of the observation noise is provided in Subsection IV-D, while the estimation of the process noise is provided in Subsection IV-E.

\[ D. \text{ Signal Pre-Processing} \]

Our HKF algorithm operates under the assumption that each HB, represented as \( X_{\tau} \), can be isolated during the preprocessing phase and subsequently processed independently. The core of this process relies on on-the-fly segmentation of each HB, facilitated by R-peak detection. Consequently, the processing of each HB is centered around the R-peak within a set of samples. The size of this processing window is proportional to the sampling frequency of the ECG recording device; for instance, a Heart Rate (HR) of 60 [bpm] with a sampling frequency of 360 [Hz] translates to 360 data points per HB.

Accurate detection of various peaks and intervals within an ECG waveform, such as R-peaks and the QT interval, often provides vital clinical information. Over the years, a multitude of methods have been developed to address this complex task [60]. Classical approaches typically utilize the time derivatives of the recorded signal, applying thresholding on parameters such as slew rate [61] or directly to the derivative [62]. Some methods attempt to fit local SS models to a given signal, detecting peaks by observing sudden changes in the model parameters using a log-cost ratio [63]. Additionally, wavelet transform-based methods [64], [65], and neural network (NN)-based techniques [66] have also been proposed.

Since the focus of our paper is not on the specifics of ECG segmentation, and because our HKF approach does not depend on a particular method for segmentation, we opted for an off-the-shelf solution. Nevertheless, certain properties are critical for real-time applications. The segmentation algorithm must be resilient against typical ECG artifacts, such as electrical noise and baseline drift, to ensure reliability across diverse settings. Moreover, to facilitate deployment and real-time processing on portable and wearable devices, which typically have limited processing power, we require modest computational demands and minimal computational latency—both are essential for systems that require instant feedback.

For these reasons, in our empirical study, we employed the Engelse and Zeelenberg’s algorithm [67], from the BioSPPy library. This segmenter is considered a strong candidate for a segmentation solution due to its streamlined, single-scan approach. However, users of the HKF algorithm in real-world applications can select their preferred segmentation algorithm for the preprocessing step. An example of a pre-processed observation is illustrated in Fig. 5.

\[ E. \text{ Signal Post-Processing} \]

After successfully filtering an individual HB, it can be seamlessly integrated with previously filtered HBs to form a continuous signal. During pre-processing, a fixed window length \( T \) is applied, which can result in two potential scenarios based on the HR: Overlaps (denoted as PP1) and Gaps (denoted as PP2).

PP1 Overlaps: In this scenario, if the first sample index of HB \( \tau \) is equal to or less than the last sample index of HB \( \tau - 1 \), indicating an overlap, a weighted average of the overlapping samples from the two HBs is employed as the estimated continuous signal. This ensures a smooth transition between consecutive HBs.

PP2 Gaps: For gaps between heartbeats, if there is no overlap—meaning the first sample of HB \( \tau \) follows after the last sample of HB \( \tau - 1 \)—linear interpolation is used to bridge the data gap. This approach is justified as there’s typically no significant heart activity between these periods, and the region usually consists predominantly of white noise.

A visual representation of these two methodologies can be referenced in Fig. 6. While this method allows for accurate continuity in the reconstructed acecg signal, more sophisticated methods could be employed for further post-processing of the segmented HBs.
IV. ONLINE PRIOR LEARNING AND PARAMETER ESTIMATION

In this section, we delve into the process of parameter learning. Specifically, we discuss how to determine the parameters for the intra-HB Kalman smoothing (D1) during the online warm-up phase (P1). We will also cover the learning of parameters for the inter-HB Kalman filtering (D2).

A. Online Learned Taylor Priors

The primary objective of the online warm-up phase (P1) is to tailor the intra-HB SS model parameters (3) to each patient, particularly focusing on modeling the evolution of the ECG waveform signal. Given the complexities of the ECG signal, identifying a closed-form, time-invariant function that can accurately depict its evolution proves challenging. Alternatively, one can model the evolution of the state-difference, in a discrete-time interval $\Delta t$, namely:

$$x_{t,t} = f_t(x_{t,t-1}) = x_{t,t-1} + \Delta x_{t,t}, \quad t \in T. \quad (26)$$

Here, $\Delta x_{t,t}$ is a time-dependent point-wise increment that provides insight into how the waveform evolves at each discrete-time point. Owing to the periodic nature of the ECG waveform, characterizing such a prior for one period, $\tau$, ensures its applicability across all periods. For a sufficiently small time-quantization interval $\Delta t$, we can adopt a finite Taylor series expansion for point-wise modeling and approximation of the incremental evolution:

$$\Delta x_{t,t} = x_{t,t} - x_{t,t-1} \simeq \sum_{k=1}^{K} \frac{d^k}{dt^k} x_{t,t-1} \cdot \Delta t^k \quad (27)$$

Here, $\Delta t^k$ and $\frac{d^k}{dt^k} x_{t,t-1}$ are the $k$-th basis function and coefficient, respectively. While the former is data-independent and solely depends on the time-quantization interval $\Delta t$, the latter is data-dependent and will be learned.

The hyperparameter $K$ is the order of the Taylor approximation that determines the number of basis functions and coefficients learned. It requires tuning to optimize model performance, as it inherently affects the bias-variance tradeoff. While $K$ should be sufficiently large to capture the intricate behavior within an HB and to prevent underfitting, it shouldn’t be too large as to cause excessive sensitivity to noise and overfitting. In our numerical evaluations, we found that setting $K$ to 5 provided a good balance between complexity and performance.

To allow the integration of our prior model into a standard formulation of an EKF and to facilitate its learning, we represent the incremental evolution prior in a canonical form as:

$$\Delta x_{t,t} = x_{t,t} - x_{t,t-1} \simeq F_t \cdot \phi(\Delta t), \quad (28)$$

where $F_t \in \mathbb{R}^{m \times K}$ is the increment evolution matrix that contains the learned coefficients, formulated as:

$$F_t = \left( \frac{d}{dt} x_{t,t-1}, \cdots, \frac{d^K}{dt^K} x_{t,t-1} \right), \quad (29)$$

and $\phi \in \mathbb{R}^{K \times 1}$ is a vector of fixed basis functions that solely depends on the quantization interval, namely:

$$\phi(\Delta t) = \left( \frac{\Delta t}{\tau}, \cdots, \frac{\Delta t^K}{\tau^K} \right)^\top. \quad (30)$$

To further integrate the evolution prior into the EKF’s setting, we also need to compute its Jacobian matrix:

$$J_F = \frac{dx_{t,t}}{dx_{t,t-1}} = \frac{df_t(x_{t,t-1})}{dx_{t,t-1}} = I_{m \times m} \quad (31)$$

Since $J_F$ simplifies to an identity matrix, it greatly eases the computation of the error covariance $P_{t|t-1}$ in (11) by eliminating the need to compute and multiply the Jacobian for each sample in every HB.

B. Learning a Taylor Approximation

The matrix $F_t$, yet to be determined, can be learned in an unsupervised manner. For brevity, we focus on a single channel and omit its index; this process should be repeated for each channel. Let $\theta_i$ represent a single row of matrix $F_t$ as a column vector in $\mathbb{R}^K$. To optimize it, our aim is to minimize a loss function $L(\theta_i, \lambda^2)$ that depends on the data:

$$\theta_i^* = \arg\min_{\theta_i \in \mathbb{R}^K} L(\theta_i, \lambda^2). \quad (32)$$

We adopt a LS loss function with an $\ell_2$ (Tikhonov) regularization, also known as Ridge regression:

$$L(\theta_i, \lambda^2) = \sum_{i=1}^{N} (\Delta y_{t,i} - \phi(\Delta t) \cdot \theta_i)^2 + \lambda^2 \cdot \| \theta_i \|^2 \quad (33)$$

The scalar $\Delta y_{t,i}$ is the noisy observed difference taken from the designated channel and from HB $\tau_i$ out of $N$ HBs, where the set of noisy observations is defined as:

$$\{ \Delta y_{t,i} = y_{t,i} - y_{t-1,i} \}_{i=1}^{N}. \quad (34)$$

This optimization problem admits a clear, closed-form solution:

$$\hat{\theta}_i = \Phi^{-1}(\Delta t, \lambda^2) \cdot \phi(\Delta t) \cdot \Delta \bar{y}_t. \quad (35)$$

Here, $\Phi(\Delta t, \lambda^2)$ on the time-quantization interval and the regularization parameter $\lambda^2$, and is given by:

$$\Phi(\Delta t, \lambda^2) = \phi(\Delta t) \cdot \phi(\Delta t)^\top + \frac{1}{N} \cdot \lambda^2 \cdot I_{K \times K}. \quad (36)$$

$\Delta \bar{y}_t$ is a data-dependent term, calculated as the average of the observed data:

$$\Delta \bar{y}_t = \frac{1}{N} \cdot \sum_{i=1}^{N} \Delta y_{t,i}. \quad (37)$$

Given the nature of the problem, the term in (35) can be further simplified and efficiently computed:

$$\theta_i = \frac{1}{N^\lambda} + \| \phi(\Delta t) \|^2 \cdot \phi(\Delta t) \cdot \Delta \bar{y}_t. \quad (38)$$

Therefore, we derive a closed-form formula for the evolution prior:

$$\Delta \hat{x}_{t,t} = \frac{\| \phi(\Delta t) \|^2}{N \lambda^2} \cdot \Delta \bar{y}_t. \quad (39)$$

Here, $\lambda^2$ is the regularization parameter that accounts for uncertainty and should be fine-tuned to optimize performance in every channel.
The reliability of our estimator is influenced by the number of observations used in its computation, which in turn depends on the number of HBs utilized to gather data. To ensure a smooth evolution of our ECG waveform, we aggregate multiple observations close to the time-step and adapt our loss function accordingly, replacing $\Delta \mathbf{y}_t$ in (39) with $\Delta \tilde{\mathbf{y}}_t$, a weighted average of the noisy data within a window spanning $2M + 1$ decaying weights $\alpha_j$ that sum to one:

$$\Delta \tilde{\mathbf{y}}_t = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=-M}^{M} \alpha_j \cdot \Delta \mathbf{y}_{\tau,t+j}, \quad \sum_{j=-M}^{M} \alpha_j = 1. \quad (40)$$

This introduces an additional degree of freedom, the window type (e.g., rectangular, triangular, Gaussian), and the window’s half-length $M$, which can be tuned for performance. See Fig. 7 for illustration.

### C. Online Intra-Heartbeat Covariance Learning

In the second stage of our proposed warm-up phase P1, the primary objective is to estimate the missing covariance matrices for the intra-HB SS model (3). Specifically, we aim to determine the intra-evolution covariance, $\mathbf{Q}_t$, and the intra-observation covariance, $\mathbf{R}_t$. To achieve this, we employ the EM algorithm. The EM algorithm, an iterative method, has been frequently adapted for the task of parameter estimation in SS models [68], [69], [70]. The EM algorithm operates by alternating between the E-step and the M-step. During the E-step, we compute the conditional expectation. For our application, this involves calculating the posterior moments using the RTS smoother for each time-step $t \in T$. Specifically, from the RTS smoother, we derive the first-order moment $\mathbf{x}_{t|T}$, the covariance matrix $\mathbf{P}_{t|T}$, and the backward smoothing gain $G_{t}^T$. With these values, we can then determine the posterior second-order moments, namely:

$$\mathbf{X}_t^{||} = \mathbf{x}_{t|T} \cdot \mathbf{x}_{t|T}^\top + \mathbf{P}_{t|T}, \quad \mathbf{Y}_t^{||} = \mathbf{y}_t \cdot \mathbf{y}_t^\top, \quad (41)$$

and the posterior correlations:

$$\mathbf{C}_{t,t}^{\text{xy}} = \mathbf{x}_{t|T} \cdot \mathbf{y}_t^\top, \quad (42a)$$

$$\mathbf{C}_{t,t-1}^{\text{xx}} = \mathbf{x}_{t|T} \cdot \mathbf{x}_{t-1|T} + \mathbf{P}_{t|T} \cdot G_{t}^T. \quad (42b)$$

In the M-step, we employ maximum likelihood estimation, drawing upon the results of the previous step, to derive an instantaneous estimate for the unknown covariance matrices, as outlined next:

$$\hat{\mathbf{Q}}_t = \mathbf{X}_t^{||} - 2 \cdot \mathbf{C}_{t,t-1}^{\text{xx}} + \mathbf{X}_{t-1}, \quad (43a)$$

$$\hat{\mathbf{R}}_t = \mathbf{Y}_t^{||} - 2 \cdot \mathbf{C}_{t,t}^{\text{xy}} + \mathbf{X}_t^{||}. \quad (43b)$$

Given the distinct shape of the HB, we operate under the assumption that the observation noise statistic remains invariant within a single HB; that is, it stays consistent over the timescale of a HB. Consequently, by averaging the instantaneous estimators across the entire HB, we obtain a single estimate that holds valid for all time-steps, as illustrated next:

$$\hat{\mathbf{R}}_t = \mathbf{R} = \frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{R}}_t. \quad (44)$$

While our previous assumption applies to the observation noise statistic, it does not hold true for the process noise statistic. We assume that this statistic is not invariant within a single HB and exhibits temporal correlations over brief periods. Consequently, we take an average of the instantaneous estimators within a short time window of length $L = L_1 + L_2 + 1$, obtaining:

$$\hat{\mathbf{Q}}_t = \frac{1}{L_1 + L_2 + 1} \sum_{\ell=-L_1}^{L_2} \hat{\mathbf{Q}}_{t+\ell}. \quad (45)$$

To get a more robust and accurate estimate, we can exploit inter-HB correlations and unfold the EM algorithm over multiple HBs. Given a sequence of HBs, the most straightforward way is to apply the EM algorithm on each HB independently, and then average the single HB based estimators to a multi-HB based estimator. The second alternative, which is more aligned with online learning, is to run a single EM iteration on the data from HB with index $\tau$, and then plug-in the results into HB with index $\tau + 1$ and perform another EM iteration. This process should continue until some convergence criteria are met, namely the change in the estimated parameters is small enough. In addition to the fact that the second alternative has lower computational complexity, it often yields much more stable results, as we observe in the empirical study (Section V).

### D. Inter-Heartbeat Observation Noise Estimation

To estimate $\mathbf{R}_{\tau,t}$, the inter-HB observation noise covariance, we observe that $\mathbf{P}_{\tau,t|T}$ is the error covariance of the first stage estimation, namely

$$\mathbf{x}_{\tau,t|T} - \mathbf{x}_{\tau,t} \sim \mathcal{N}(0, \mathbf{P}_{\tau,t|T}). \quad (46)$$

Since $\tilde{\mathbf{y}}_{\tau,t} = \mathbf{x}_{\tau,t|T}$, the input for the second stage filtering is equal to the output of the first stage smoothing, then we can set $\mathbf{P}_{\tau,t|T}$ as $\hat{\mathbf{R}}_{\tau,t}$, an instantaneous estimate for the second-stage observation noise covariance.

To further smooth our estimate and to improve the algorithm’s stability, we exploit inter-HB correlations by averaging the instantaneous estimates in a window of size $L = L_1 + L_2 + 1$ and get

$$\bar{\hat{\mathbf{R}}}_{\tau,t} = \frac{1}{L_1 + L_2 + 1} \sum_{\ell=-L_1}^{L_2} \hat{\mathbf{R}}_{\tau,t+\ell}. \quad (47)$$
E. Inter-Heartbeat Process Noise Estimation - Diagonal Case

To estimate $Q_\tau$, the process covariance of HB $\tau$, we establish its relationship with the innovation covariance $\hat{S}_\tau$ based on the SS model in equation (5) and the KF equation (21), namely:

$$\hat{S}_{\tau,t} = R_{\tau,t} + Q_{\tau,t} + P_{\tau-1,t}, \quad \forall t$$  \hspace{1cm} (48)

The maximum likelihood estimate for $S_\tau$ [71] is given by:

$$\hat{S}_{\tau,t} = \Delta \cdot \Delta^\top$$ \hspace{1cm} (49)

From that, we can extract an estimate for the process covariance that is only guaranteed to be Positive-Semidefinite (PSD) when we omit the correlation between different channels during the inter-filtering process, and presume a diagonal structure for the covariance matrix. Namely:

$$\hat{Q}_{\tau,t} = \hat{q}^2_i \otimes I_{m \times m},$$ \hspace{1cm} (50)

where all diagonal entries are positive

$$\tilde{q}_i^2 = \max \{ \hat{Q}_{\tau}(i, i), 0 \}, \quad \hat{Q}_\tau = \hat{S}_{\tau,t} - \tilde{R}_\tau - P_{\tau-1}. \hspace{1cm} (51)$$

To get a smoother and more robust estimator, we can exploit intra-HB and inter-HB correlations. We first average our estimator in a local time window

$$Q^*_{\tau,t} = \frac{1}{L_1 + L_2 + 1} \sum_{\ell = -L_1}^{L_2} \hat{Q}_{\tau,t+\ell} \hspace{1cm} (52)$$

and then apply an exponential smoothing, i.e., a simple IIR filter, namely:

$$\hat{Q}_\tau = \alpha \cdot Q^*_\tau + (1 - \alpha) \cdot \hat{Q}_{\tau-1}. \hspace{1cm} (53)$$

Here, $0 < \alpha < 1$ is the forgetting factor.

While this option is simpler to compute, it is not always optimal as can be seen in the empirical results in Section V. Next, we relive that assumption to exploit inter-channel correlation.

F. Inter-Heartbeat Process Noise With Channels’ Correlation

To enhance the performance of our denoising algorithm, we aimed to relax the constraint that mandates the inter-HB process covariance matrix be diagonal, thereby leveraging inter-channel correlations. A primary constraint on a covariance matrix is its PSD [72]. However, when estimating a covariance matrix as an outcome of gradient descent, adherence to the PSD condition is not guaranteed, rendering it unsuitable for accurately defining the covariance of a Gaussian distribution.

Unlike diagonal matrices, which naturally satisfy PSD conditions with positive real entries, there isn’t a straightforward criterion to ensure a matrix’s PSD property without significant alterations that may lead to filter instabilities.

To address this challenge, we employed a variant of the Riemann Manifold Gradient Descent (RMGD) algorithm [73], an advanced optimization technique that adapts traditional gradient descent to operate on curved spaces called manifolds [74]. PSD matrices of size $m \times m$ span a manifold $S_+$ within $R_{m \times (m+1) \times \frac{m(m+1)}{2}}$. The key ingredient of the RMGD algorithm is the exponential mapping, which is a key operation in differential geometry used to transition points from the tangent space of a manifold back to the manifold itself. The manifold is a curved surface, and the tangent space is a flat plane that touches the manifold at a point without intersecting it. The exponential map takes a vector in this flat tangent space and projects it back onto the curved manifold in a way that respects the manifold’s intrinsic geometry [75]. See an illustration in Fig. 8.

The update rule that computes $Q_{\tau,t}$, the covariance matrix in HB $\tau$ is based on the gradient of the log-likelihood with respect to $Q_{\tau,t}$, namely:

$$\hat{Q}_{\tau,t} = \exp_{\tau-1,t} (-\alpha \cdot \nabla q_{\tau,t} \ell (\Delta)) \hspace{1cm} (54)$$

where $\alpha > 0$ is the step size, and the gradient of the log-likelihood is given by

$$\nabla q_{\tau,t} \ell (\Delta) = -\frac{1}{2} \cdot S_{\tau,t}^{-1} + \frac{1}{2} \cdot S_{\tau,t}^{-1} \cdot \Delta_{\tau,t} \cdot S_{\tau,t}^{-1} \cdot \Delta_{\tau,t}^\top \cdot S_{\tau,t}^{-1}. \hspace{1cm} (55)$$

Here, we use the relationship in (48) to estimate $\hat{S}_{\tau,t}$ based on $\hat{Q}_{\tau-1,t}$ as an anchor value.

Among the available metrics [76], we chose the log-Cholesky metric [77] due to its efficient closed-form solution for the exponential map, a crucial component of RMGD. The technical details that define this specific exponential mapping are beyond the scope of this paper and are detailed in [77] and in our source code3.

G. Discussion

Our proposed HKF is designed as a patient-adaptive, online, multi-channel ECG denoising algorithm. It utilizes a hierarchical statistical SS model that integrates both intra- and inter-HB relationships. This model, represented as a tensor, capitalizes on the quasi-periodicity inherent in ECG waveforms to optimize the exploitation of spatial relationships within and between heartbeats, and across multiple channels. For data processing, we employ the KS to denoise each single heartbeat independently based on the intra-HB model, and the KF to

3The source code used in our empirical study along with hyperparameters is at https://github.com/KalmanNet/HKF_Thesis
integrate information across consecutive heartbeats using the inter-HB model. These renowned algorithms extract essential information for robust performance while maintaining computational efficiency and tractability. Addressing the unique waveform characteristics of individual patients, our approach does not rely on offline training, as is typical with NNs. Instead, we learn a patient-specific intra-HB evolution prior that enhances the tailoring of our filtering algorithm to individual patient needs. The fundamental filtering operators within our Kalman algorithms are the Kalman filtering and smoothing gains. These gains control the update process from the prior to the posterior, making their optimal computation a critical aspect of the model. To account for the non-stationary nature of the specific data being processed. For example, the inter-HB phase operates under the assumption of an online, infinite time horizon, where a continuous stream of heartbeats is processed. In this setting, only forward processing using a KF is practical, as retrospective adjustments to denoised heartbeats would not influence the current posterior or subsequent priors. This forward-only approach is consistent with the real-time nature of ECG signal monitoring, where each new HB is processed sequentially without revisiting prior heartbeats. It is crucial to note that while our method primarily utilizes forward filtering in the inter-HB phase, it retains flexibility to adapt to different operational conditions. In scenarios where all data is pre-recorded, implementing backward processing for smoothing is not only feasible but could also enhance the accuracy of our estimates by considering all available observations simultaneously. Moreover, in real-time settings that can tolerate slight delays—such as a few HBs—it is feasible to implement fixed-lag smoothing. This method allows for forward filtering while also permitting a brief look ahead, followed by smoothing back to enhance the analysis of the current heartbeat. Such adjustments, while practical, do not fundamentally alter the core principles of our algorithm. Instead, they refine its application to meet specific operational demands and constraints effectively. These enhancements and the inherent flexibility of our HKF framework ensure it is not only robust but also adaptable to a wide range of clinical and technical environments, underscoring its suitability for advanced ECG signal processing.

While Section V demonstrates that our online, patient-adaptive, model-based algorithm surpasses a black-box NN trained offline, namely the AE approach, we recognize the potential for integrating NN approaches into our framework. A natural extension of HKF could involve replacing the intra- and inter-HB smoothing and filtering stages with NN-based methods such as RTSNet and KalmanNet [78], [79]. Specifically, the Adaptive KalmanNet [80], designed to adapt to non-stationary noise, could be employed. These architectures focus on learning the forward and backward Kalman gains, which are pivotal in balancing the information from prior distributions and new observations to compute the posterior. We envision that such modifications would yield a hybrid model that leverages both model-based and data-driven approaches for ECG denoising. This hybrid approach not only promises to enhance the noise tolerance of HKF but also to boost overall performance. Explorations into these extensions are reserved for future work, where we anticipate significant advancements in the efficacy of ECG signal denoising.

In clinical settings, precise characteristics of the ECG waveform, such as the QT interval and T-wave amplitude, are critical for diagnosing and monitoring various cardiac conditions. An elongated QT interval, for example, is a known risk factor for predicting cardiac arrhythmias [7], [8], while elevated T-wave amplitudes can indicate a hypoxic state [10]. The noisy signals that served as input for our method typically obscure such critical features, making accurate extraction challenging if not impossible. Our empirical results demonstrate that our method effectively suppresses noise and restores the typical ECG waveform, making characteristic peaks and intervals clearly visible. This enhanced waveform clarity permits precise measurement and analysis of vital clinical features and simplifies the use of methods like those proposed in [10] for delineating the ECG and accurately determining features such as QT intervals and T-wave amplitudes.
as the QT interval. Improved fidelity in waveform representation further supports more reliable clinical assessments and informed decision-making.

To support more practical clinical settings and facilitate broader use by physicians, it is feasible to extend HKF into a versatile algorithm capable of filtering ECG signals under any conditions, including the presence of arrhythmias. Arrhythmias manifest in diverse ways—from accelerated heart rates (atrial fibrillation) to reduced rates (bradycardia) and sporadic episodes of unusually rapid heart rates at rest (supraventricular tachycardia). These irregularities can be precursors to more severe conditions, including strokes [5]. In cases where arrhythmia is evident, the assumption that sequential heartbeats are identical is invalid. Specifically, the inter-HB evolution function, \( F (\cdot) \) from (5), does not remain an identity function. Consequently, enhancing performance through the merging of consecutive heartbeats becomes impractical. Thus, we recommend using only the initial smoothing stage, bypassing the subsequent filtering stage. To achieve this dual functionality, an arrhythmia detection test should be employed, based either on an external off-the-shelf algorithm, such as [81], or on an internal mechanism. Given that HKF is anchored in a Bayesian approach, we can exploit our statistical SS model and devise a likelihood-ratio test, for example, to effectively contrast the intra- and inter-HB filters. However, exploring this application extends beyond the scope of this paper and is reserved for future work.

There are numerous potential applications for HKF, such as its incorporation into smartwatches or its use in monitoring oxygen levels during labor. Each potential application presents its own set of challenges that warrant exploration. For instance, it would be vital to assess the interplay between the algorithms separating fetal from maternal signals in relation to HKF or to address the subtleties of lower amplitude signals typical in these scenarios.

V. EMPIRICAL STUDY

Next, we empirically evaluate the HKF algorithm’s performance. We begin by comparing HKF with multiple benchmark algorithms, using two different ECG recording datasets. Initially, we discuss the performance of our algorithm for patients without arrhythmia, representing our algorithm’s primary use case. We then showcase its efficacy for patients with arrhythmia and detail how the algorithm should be adapted for such scenarios. In the appendix, we provide an analysis illuminating the implications of certain design choices.

For the numerical evaluation, two distinct ECG recording datasets were utilized. The first is the widely recognized MIT-BIH dataset, which is publicly accessible [82], while the second dataset is proprietary. This selection ensures the algorithm’s compatibility across various recording devices, sampling frequencies, and channel configurations.

Given that the datasets contain clean ECG recordings, we had to artificially introduce AGN to simulate an ECG signal in a noisy environment.

The MSE in relation to the clean signal served as our primary performance metric. However, it’s important to acknowledge that in clinical evaluations, other factors, such as wave shape and clarity, are also significant.

A. Benchmark Algorithms

For our evaluation, the complete HKF was employed. The intermediary output of the first stage KS operation carries the label (KS-intra), and the final independent HKF output is denoted as (HKF). We’ve also presented the multichannel estimate under the label (HKF RMGD). Furthermore, results from applying only the inter-HB filtering, as highlighted in [49], are labeled as (KF-inter).

In addition to these, we included a filtered estimate generated by a data-driven convolutional AE following the methodologies in [36], [37]. To ensure optimal AE performance, training encompassed the entire dataset, excluding the tested subject. The full details are provided in our code.

B. Patients Without Arrhythmia - Proprietary Dataset

We next turn our attention to the evaluation of our algorithm on patients without arrhythmia, using a proprietary dataset. This dataset offers clean ECG recordings of adults, captured at a sampling rate of 500 [Hz]. To emulate fetal ECG signals, two modifications were applied to this data:

1) Given that the fetal HR is typically double that of an adult, the ECG signals were subsampled by a factor of two. This creates signals that appear to have a HR twice as high, with a resultant sampling rate of 250 [Hz].

2) An attenuation of the signal amplitude was performed to simulate the often weaker amplitude seen in fetal ECGs. This diminished amplitude can be attributed to the smaller size of the fetal heart and the typical recording conditions where the ECG is taken from the mother’s abdomen rather than directly from the fetal thorax.

Our algorithm’s robust performance, especially when compared with alternative methods on ECG signals corrupted to 0 [dB] SNR, is detailed in Table I. Given that the ECG signal was captured using \( m = 12 \) observation channels, our HKF with RMGD effectively leverages correlations between these multiple channels, consistently outpacing other methods in performance. Visual representations showcasing our algorithm’s efficacy can be viewed in the subsequent figures: 10(a)–10(c), 11, 12, and 13.
Fig. 10. Consecutive HBs - amplitude vs. time with noisy observations, ground truth, intra-HB smoother, HKF, HKF RMGD.
C. Patients without Arrhythmia - MIT BIH Dataset

Next, we assess our algorithm using patients without arrhythmia from the MIT-BIH (Arrhythmia) dataset. This dataset comprises 48 two-channel ambulatory ECG signals recorded by the BIH Arrhythmia Laboratory between 1975 and 1979. These signals were recorded at a rate of 360 samples per second with an 11-bit resolution over a 10mV range. This dataset is widely recognized in the ECG denoising community and serves as a reliable benchmark for evaluating algorithm performance. Examples of its application can be found in references [21], [27], [83].

The results outlined in Table II, showcase HKF’s strong performance, especially when compared with alternative methods, on ECG signals corrupted to 3 [dB] SNR. Given that the ECG signal was acquired using m = 2 observation channels, our HKF, which employs a diagonal process covariance matrix, typically outperforms the RMGD version—though performance may vary depending on the specific patient. Visual representations of our algorithm’s efficacy can be found in Figs.: 10(d), 10(e), 14, and 15.

The algorithm demonstrates significant capability in denoising ECG signals. Particularly in cases free of arrhythmia, noticeable SNR enhancements are evident. It’s also worth highlighting that the inter-channel filter doesn’t merely force a prior onto a signal. This is apparent in Fig. 10(f), where the signal’s

4In the tables, the colors green , blue , and yellow represent rankings of 1, 2, and 3, respectively.
shape shifts between the 6th and 8th seconds, a change that the HKF successfully captures. This adaptability positions HKF as an excellent tool for monitoring variations in signal shape, potentially useful in contexts like detecting heart attacks or acute hypoxia during labor.

D. Patients With Arrhythmia

Next, we delve deeper into the ECG recordings from the MIT-BIH dataset, focusing on patients who exhibit various types of arrhythmia, such as patients 102 [Fig. 10(f)], 107 [Fig. 10(g)], and 108. Given the unpredictable nature of arrhythmias in terms of shape, frequency, and amplitude, these recordings provide an excellent benchmark for testing the limits of our filter. The evaluation was conducted similarly to previous tests: a clean ECG recording was artificially corrupted by AGN to achieve a SNR, and the denoised output was compared with the original.

In cases where arrhythmia is evident, the assumption that sequential heartbeats are identical is no longer valid. Consequently, as shown in our numerical MSE findings in Table III, the full HKF configuration is outperformed by the simplified intra-HB smoother, also denoted as KS-intra. This outcome underscores that performance enhancement through the merging of consecutive heartbeats is impractical in arrhythmia cases. Thus, we recommend using only the initial stage of smoothing for patients with arrhythmia, bypassing the subsequent filtering stage. See the discussion in Subsection IV-G.

E. Heart-Rate Reconstruction

Our algorithm’s objective is to reconstruct a continuous ECG signal from the separately filtered HBs while maintaining the integrity of the HR. In Fig. 9, it’s evident that the reconstructed HR closely matches the true HR, as indicated by the dataset labels.

VI. CONCLUSION

In this paper, we introduced HKF, a novel strategy for filtering noisy ECG signals. Our approach entailed modeling the ECG system as a hierarchical SS system, upon which HKF was then constructed. This design encompassed both intra- and inter-HB dynamics, aiming to harness the maximum amount of information inherent in an ECG. Integrating both the KF and RTS smoothing, the method provides a clear and computationally efficient framework suitable for online filtering. Notably, the proposed method operates without the need for supervised training, relying instead on a highly patient-specific prior to adjust to variations in standard ECG waveforms. Our results have shown that HKF consistently surpasses other similar methods in performance. Importantly, in scenarios with significant arrhythmia in the ECG recordings, the intermediary output of HKF demonstrates adaptability to alterations in the inter-HB model.

ACKNOWLEDGMENT

The authors thank Hans-Andrea Loeliger for helpful discussions and Mehdi Bakka for helping with the empirical evaluation.

REFERENCES

[1] T. Locher, G. Revach, N. Shlezinger, R. J. G. van Sloun, and R. Vullings, “Hierarchical filtering with online learned priors for ECG denoising,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2023, pp. 1–5.
[2] Y. Bai, C. Tompkins, N. Gell, D. Dione, T. Zhang, and W. Byun, “Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting,” PloS One, vol. 16, no. 5, 2021, Art. no. e0251975.
[3] F. Sana, E. M. Isselbacher, J. P. Singh, E. K. Heist, B. Pathik, and A. A. Armoundas, “Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review,” J. Amer. College Cardiol., vol. 75, no. 13, pp. 1582–1592, 2020.
[4] “Electrocardiogram.” N. H. Service. Accessed: Nov. 16, 2023. [Online]. Available: https://www.nhs.uk/conditions/electrocardiogram/
[5] “Arrhythmia,” N. H. Service. Accessed: Nov. 16, 2023. [Online]. Available: https://www.nhs.uk/conditions/arrhythmia/
[6] I. Amer-Wählin et al., “Cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram for intrapartum fetal monitoring: A Swedish randomised controlled trial,” Lancet, vol. 358, no. 9281, pp. 534–538, 2001.
[7] P. J. Schwartz, M. Periti, and A. Malliani, “The long QT syndrome,” Amer Heart J., vol. 89, no. 3, pp. 378–390, 1975.
[8] P. Schwartz and S. Wolf, “QT interval prolongation as predictor of sudden death in patients with myocardial infarction,” Circulation, vol. 57, no. 6, pp. 1074–1077, 1978.
[9] “Heart attack.” N. H. Service. Accessed: Nov. 16, 2023. [Online]. Available: https://www.nhs.uk/conditions/heart-attack/
[10] K. R. Greene, G. S. Dawes, H. Lilja, and K.-G. Rosén, “Changes in the ST waveform of the fetal lamb electrocardiogram with hypoxemia,” Amer. J. Obstetrics Gynecol., vol. 144, no. 8, pp. 950–958, 1982.
[11] N. A. MacIachlan, J. A. Spencer, K. Harding, and S. Arulkumaran, “Fetal acidemia, the cardiotocograph and the T/QRS ratio of the fetal ECG in labour,” BJOG Int. J. Obstetrics Gynaecol., vol. 99, no. 1, pp. 26–31, 1992.

Fig. 15. Single HB - patient 101 - MIT-BIH - amplitude vs. time.

TABLE III
MSE [dB] for the MIT-BIH Testset - 3 [dB] SNR

| Patient | Noise Floor [dB] | Alt. KF-Intra | KS-Intra | HKF | HKF (RMGD) |
|---------|-----------------|--------------|----------|-----|------------|
| 102     | -4.74           | -4.44        | -9.78    | -14.64 | -12.32 | -12.99 |
| 107     | -5.96           | -6.06        | -10.63   | -14.98 | -8.24    | -12.98 |
| 108     | -14.27          | -15.05       | -9.8     | -22.38 | -12.77   | -21.87 |
[38] O. Sayadi and M. B. Shamsollahi, “ECG denoising with adaptive bionic

[27] I. Daubechies,  

[24] D. Zhang et al., “An efficient ECG denoising method based on empirical

[23] J. Gilles, “Empirical wavelet transform,” IEEE Trans. Signal Process.

[30] G. Georgieva-Tsaneva and K. Tcheshmedjiev, “Denoising of electrocar-

[29] M. Ayat, M. B. Shamsollahi, B. Mozaffari, and S. Kharabian, “ECG
denoising using modulus maxima of wavelet transform,” in Proc. Annu.
Conf. IEEE Eng. Med. Biol. Soc., 2009, pp. 416–419.

[37] E. Fotiadou and R. Vullings, “Multi-channel fetal ECG denoising with deep CNNS,” Frontiers Pediatrics, vol. 8, p. 508, Aug. 2020.

[36] A. Rasti-Meymandi and A. Ghaffari, “A deep learning-based framework
for ECG signal denoising based on stacked cyclic cycle tensor,” IEEE
Trans. Signal Process. Control, vol. 71, 2022, Art. no. 103275.

[35] R. A. Wildhaber, N. Zalmail, M. Jacomet, and H.-A. Loeliger, “Windowed
state-space filters for signal detection and separation,” IEEE Trans. Signal
Process., vol. 66, no. 14, pp. 3768–3783, Jul. 2018.

[34] R. A. Wildhaber, “Localized state space and polynomial filters
with applications in ECG,” Ph.D. dissertation, ETH Zürich, Zürich,
Switzerland, 2019.

[33] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature
Commun. Nonlin. Sci. Numer. Simul., vol. 57, pp. 415–421, 2018.

[32] A. Gruetzmann, S. Hansen, and J. Müller, “Novel dry electrodes for
applications in ECG,” Ph.D. dissertation, ETH Zürich, Zürich,
Switzerland, 2019.

[31] P. Singh, G. Pradhan, and S. Shahnawazuddin, “Denoising of ECG
signal by non-local estimation of approximation coefficients in DWT,”
Biocybernetics Biomed. Eng., vol. 37, no. 3, pp. 599–610, 2017.

[30] A. Gruetzmann, S. Hansen, and J. Müller, “Novel dry electrodes for
ECG monitoring,” Physiol. Meas., vol. 28, no. 11, pp. 1375–1390,
Nov. 2007.

[29] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521,
pp. 436–444, 2015.

[28] C. T. Arsene, R. Hankins, and H. Yin, “Deep learning models for
denoising ECG signals,” in Proc. 27th Eur. Signal Process. Conf.
(EUSIPCO), 2019, pp. 1–5.

[27] K. Antezak, “Deep recurrent neural networks for ECG signal denoising,”
2018, arXiv:1807.11551.

[26] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y.
Chien, “Noise reduction in ECG signals using fully convolutional
denoising autoencoders,” IEEE Access, vol. 7, pp. 60806–60813, 2019.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
models,” Current Directions Biomed. Eng., vol. 8, no. 2, pp. 101–104, 2022.

[64] P. J. M. Fard, M. Moradi, and M. Tajvidi, “A novel approach in R peak detection using hybrid complex wavelet (HCW),” Int. J. Cardiol., vol. 124, no. 2, pp. 250–253, 2008.

[65] Y. Zou et al., “An energy-efficient design for ECG recording and R-peak detection based on wavelet transform,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 62, no. 2, pp. 119–123, Feb. 2015.

[66] J. Lautala et al., “Robust ECG R-peak detection using LSTM,” in Proc. 35th Annu. ACM Symp. Appl. Comput., 2020, pp. 1104–1111.

[67] W. A. Engelse and C. Zeeukenberg, “A single scan algorithm for QRS-detection and feature extraction,” Comput. Cardiol., vol. 6, no. 1979, pp. 37–42, 1979.

[68] Z. Ghahramani and G. E. Hinton, “Parameter estimation for linear dynamical systems,” Univ. Toronto, Dept. CS, Tech. Rep. CRG-TR-96-2, 1996.

[69] J. Dauwels, A. Eckford, S. Kori, and H.-A. Loeliger, “Expectation maximization as message passing—Part I: Principles and Gaussian messages,” 2009, arXiv:0910.2832.

[70] Sophocles J. Orfanidis, Applied Optimum Signal Processing, Rutgers University, Jan. 2007.

[71] P. De Jong, “The likelihood for a state space model,” Biometrika, vol. 75, no. 1, pp. 165–169, 1988.

[72] R. Bhattacharya, Positive Definite Matrices (Princeton Ser. Appl. Math.), Princeton, NJ, USA: Princeton Univ. Press, 2007.

[73] S. Bonnabel, “Stochastic gradient descent on Riemannian manifolds,” IEEE Trans. Autom. Control, vol. 58, no. 9, pp. 2217–2229, Sep. 2013.

[74] J. M. Lee, Introduction to Riemannian Manifolds, vol. 176. New York, NY, USA: Springer-Verlag, 2018.

[75] S. Sa and R. Hosseini, “Conic geometric optimization on the manifold of positive definite matrices,” SIAM J. Optim., vol. 25, no. 1, pp. 713–739, 2015.

[76] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, “Geometric means in Riemannian manifolds,” SIAM J. Matrix Anal. Appl., vol. 29, no. 1, pp. 328–347, 2007.

[77] Z. Lin, “Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition,” SIAM J. Matrix Anal. Appl., vol. 40, no. 4, pp. 1353–1370, 2019.

[78] G. Revach, N. Shlezinger, Z. Ni, A. L. Escoriza, R. J. G. van Sloun, and Y. C. Eldar, “KalmanNet: Neural network aided Kalman filtering for partially known dynamics,” IEEE Trans. Signal Process., vol. 70, pp. 1532–1547, 2022.

[79] G. Revach, X. Ni, N. Shlezinger, R. J. G. van Sloun, and Y. C. Eldar, “RTSNet: Learning to smooth in partially known state-space models,” IEEE Trans. Signal Process., vol. 71, pp. 4441–4456, 2023.

[80] X. Ni, G. Revach, and N. Shlezinger, “Adaptive KalmanNet: Data-driven Kalman filter with fast adaptation,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2024, pp. 5970–5974.

[81] S. Dhyan, A. Kumar, and S. Choudhury, “Analysis of ECG-based rhythmia detection system using machine learning,” Circulation, vol. 101, no. 23, pp. e215–e220, 2000.

[82] P. Singh and G. Pradhan, “A new ECG denoising framework using generative adversarial network,” IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 18, no. 2, pp. 759–764, Mar./Apr. 2021.

Guy Revach (Senior Member, IEEE) received the B.Sc. (cum laude) and M.Sc. degrees from the Department of Electrical & Computer Engineering, Technion – Israel Institute of Technology, Haifa, in 2008 and 2017, respectively. His master’s thesis, under the supervision of Prof. Nahum Shimkin, explored planning for cooperative multigagents. Since 2019, he has been a Ph.D. Candidate with the ETH Zürich, under the guidance of Prof. Dr. Hans-Andrea Loeliger. He is a Statistical Signal Processing and Machine Learning Researcher with the Institute for Signal and Information Processing (ISI), ETH Zürich, Switzerland.

Rik Vullings (Member, IEEE) received the M.Sc. degree in applied physics and the Ph.D. degree in electrical engineering from Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands, in 2005 and 2010, respectively. Currently, he is an Associate Professor with the Biomedical Diagnostics Laboratory, Signal Processing Systems Group, Department of Electrical Engineering, TU/e. His research interests include probabilistic model-driven and data-driven signal-processing techniques for biomedical applications.