SIGNAL DETECTION AND DISCRIMINATION FOR MEDICAL DEVICES USING WINDOWED STATE SPACE FILTERS

Reto A. Wildhaber
Dept. of Information Technology and
Electrical Engineering ETH Zurich
Bern University of Applied Sciences, Biel
Switzerland
email: wildhaber@isi.ee.ethz.ch

Nour Zalmai[1], Marcel Jacomet[2] and Hans-Andrea Loeliger[1]
[1]Dept. of Information Technology and
Electrical Engineering ETH Zurich
[2]Bern University of Applied Sciences, Biel
Switzerland
email: {zalmai, loeliger}@isi.ee.ethz.ch

ABSTRACT
We introduce a model-based approach for computationally efficient signal detection and discrimination, which is relevant for biological signals. Due to its low computational complexity and low memory need, this approach is well-suited for low power designs, as required for medical devices and implants. We use linear state space models to gain recursive, efficient computation rules and obtain the model parameters by minimizing the squared error on discrete-time observations. Furthermore we combine multiple models of different time-scales to match superpositions of signals of variable length. To give immediate access to our method, we highlight the use in several practical examples on standard and on esophageal ECG signals. This method was adapted and improved as part of a research and development project for medical devices.

KEY WORDS
signal detection, pattern discrimination, low power, biological signals, linear state space models

1 Introduction

For an increasing number of wearable and implantable medical devices, efficient and reliable signal processing is a key element. Since space and energy are a highly limiting design factors in such devices, sampling schemes and efficient signal processing algorithms with low power consumption are required.

Biological signals are essentially of finite duration with a well-defined shape in the time domain and a variety of time scales. Well known examples are ECG (electrocardiogram), EMG (electromyogram), ENG (electroneurogram), EEG (electroencephalogram), and many others. When processing those signals, shape preservation is a fundamental aspect. Additionally, biological signals often contain a baseline due to physical interferences, which needs to be estimated and separated from the signal of interest. Consequently, many signal processing tasks in the field of biology consist of signal detection, separation, and discrimination.

The most often used approaches for those tasks are finite (FIR) or infinite impulse response (IIR) filters. FIR filters are versatile but computationally cumbersome and not flexible enough to deal with different time scales. On the other hand, IIR filters are efficiently implemented but hard to tune for advanced specifications. Furthermore, IIR filters are not shape preserving if not used in a forward-backward processing mode, which might be complicated in real time applications. Another popular method is multi-resolution analysis, including wavelets [1], where the selection of an appropriate wavelet is often hard but crucial. Model based approaches are also very common for those signal processing tasks. The idea is to fit a model to the observed signal by minimizing the squared error (L2 norm). Those methods are flexible for performing signal detection, approximation, and separation. For example spline models [2, 3] are commonly used for signal interpolation and approximation.

Such model based approaches usually preserve the signal shape. Finding a good signal model is often possible, but the error computation and minimization might be computationally intense. However, the use of state space models, which also includes spline models [4, 3], leads to efficient and recursive error computation [5, 6]. Signal estimation with state space models are closely linked to Kalman filters [7, 8] and share the same recursions. Furthermore, the cost is a quadratic form and, thus, a constrained minimization of the cost is either done in closed form or performed with standard iterative algorithms. State space modeling is well established [9, 10, 11], and offers many flexibilities which are often not exploited or only vaguely stated.

In this paper, we use model based LTI (Linear Time-Invariant) state space filters for signal detection and discrimination in practical applications. We locally fit linear state space models to our observations by adding a window of exponentially decaying and/or rectangular shape. This approach leads to introduce a *cost segment* for a specific model in a particular window. Furthermore we combine multiple such models of different time scales to match superpositions of signals of variable length, leading to a *composite cost* consisting of several *cost segment*s. As we use models based on state space systems and choose our windows carefully, we obtain a versatile method with efficient recursive computation rules.

To give an immediate access to our method, we illus-

trate our approach with three practical examples:

1. ***Esophageal ECG Wave Detection and Discrimination*** - Detection and discrimination of atrial (A) and ventricular (V) waves of variable scale in an esophageal ECG signal, ready for online or offline application (Fig. 3).

2. ***Esophageal ECG Wave Detection and Discrimination with a Baseline Signal of Different Time Scale*** - Detection and discrimination of ECG waves of variable scale including baseline modeling of different time scales (Fig. 4).

3. ***P-Wave Discrimination in Multi-Channel ECG*** - Detection and discrimination of P-Wave in a pathologic multi-channel surface ECG signal (Fig. 6).

Our method was adapted and improved as part of a research and engineering project for an implantable esophageal ECG device. Its low memory need and low computational complexity suits for implementation on most hardware platforms, including ASICs (Application Specific Integrated Circuits).

2 Shape Detection and Discrimination

2.1 Signal Models Using Autonomous State Space Representations

We recall that the scalar output s_k of an autonomous (i.e. input free), time-invariant state space system of order N is given by

$$x_k = Ax_{k-1} \tag{1}$$
$$s_k = cx_k \tag{2}$$

with *state-transition matrix* $A \in \mathbb{R}^{N\times N}$, *output vector* $c \in \mathbb{R}^{1\times N}$, *state vector* $x_k \in \mathbb{R}^{N\times 1}$. This sequence can be equivalently written in closed form as

$$s_k(x_0) = cA^k x_0,\ k \in \mathbb{Z}\,, \tag{3}$$

parametrized by an *initial state vector* $x_0 \in \mathbb{R}^{N\times 1}$. Such state space system can generate a wide class of discrete-time signals. Choosing A and c accordingly leads to output sequences consisting of polynomial, sinusoidal, exponential shapes and products of those. For example, for generating a polynomial sequence of degree two, we can choose

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix},\ c = [1\ 0\ 0] \tag{4}$$

and get, given the initial state $x_0 = [\beta_0\ \beta_1\ \beta_2]^\mathsf{T}$, the sequence $s_k(x_0) = cA^k x_0 = \beta_0 + \beta_1 k + \beta_2 k^2$.

2.2 Error Computation Using Cost Functions

We want to compute the squared error between an input signal $y_i \in \mathbb{R}$ for $i \in \mathbb{Z}$ and a signal model $s_{i-k}(x)$ from (3), shifted to index k, with unknown *initial state vector* x. Thereof, we introduce the *cost segment* on the interval $\{k+a, \ldots, k+b\}$ with $a, b \in \mathbb{Z} \cup \{\pm\infty\}$, $a \le b$ as

$$J_a^b(k, x, \theta) = \sum_{i=k+a}^{k+b} \gamma^{i-k} \left(y_i - s_{i-k}(x)\right)^2 \tag{5}$$

with *initial state vector* x and parameter set $\theta = (c, \gamma)$ consisting of the *output vector* c and the decaying factor $\gamma \in \mathbb{R}_+$.

The selection of a, b, and γ define the windowing of the *cost segment*'s sum and its localization around k. In the special case where $\gamma = 1$ and a and b are finite, the *cost segment* is computed on a rectangular window of bounds $k+a$ and $k+b$. In the limit case where $a \to -\infty$ and $\gamma > 1$ (respectively, $b \to +\infty$ and $0 < \gamma < 1$), the *cost segment* is computed on a left-sided (respectively, right-sided) exponentially decaying window (cf. Table 1, last column).

By the virtue of the linear state space model, the cost in (5) can be computed recursively. Indeed, parameterizing the cost as

$$J_a^b(k, x, \theta) = x^\mathsf{T} W_k x - 2x^\mathsf{T}\xi_k + \kappa_k \tag{6}$$

with, using (3),

$$W_k = \sum_{i=k+a}^{k+b} \gamma^{i-k} \left(A^{i-k}\right)^\mathsf{T} c^\mathsf{T} c \left(A^{i-k}\right) \in \mathbb{R}^{N\times N} \tag{7}$$

$$\xi_k = \sum_{i=k+a}^{k+b} \gamma^{i-k} y_i \left(A^{i-k}\right)^\mathsf{T} c^\mathsf{T} \in \mathbb{R}^{N\times 1} \tag{8}$$

$$\kappa_k = \sum_{i=k+a}^{k+b} \gamma^{i-k} y_i^2 \in \mathbb{R}\,, \tag{9}$$

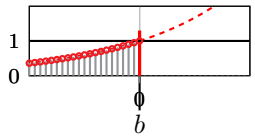
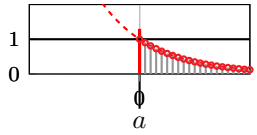
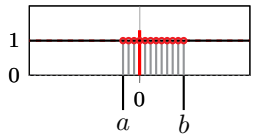
all the sums (7), (8) and (9) can be recursively computed, either as forward recursions (from k to $k+1$) or as a backward recursions (from k to $k-1$).

In order to compute the cost for all $k \in \{1, \ldots, K\}$, $K \in \mathbb{N}$, the forward recursion needs to start at $k = k' = \min(-b, 0)$, or the backward recursion at $k = k' = \max(K-a, K) + 1$, both with initial values $W_{k'} = 0$, $\xi_{k'} = 0$, $\kappa_{k'} = 0$. The full recursions for either rectangular or exponentially decaying windows are listed in Table 1.

2.3 Model Fit by Error Minimization

For many pattern detection tasks, we are interested in estimating the *initial state vector* x at a fixed index k, by minimizing the *cost segment* $J_a^b(k, x, \theta)$, which is a quadratic

Table 1: Selection of Cost Segment Windows

Segment	*Recursions*	*Boundaries* (a, b)
Cost Segment $J_a^b(k,x,\theta) = \sum_{i=k+a}^{k+b} \gamma^{i-k} \left(y_i - s_{i-k}(x)\right)^2$, $k \in \{1,\dots,K\}$		
Left sided $(a{=}-\infty,\, b{=}0,\, \gamma{>}1)$	$W_{k+1} = \gamma^{-1} A^{-\mathsf{T}} W_k A^{-1} + c^{\mathsf{T}} c$ (10) $\xi_{k+1} = \gamma^{-1} A^{-\mathsf{T}} \xi_k + c^{\mathsf{T}} y_{k+1}$ (11) $\kappa_{k+1} = \gamma^{-1} \kappa_k + y_{k+1}^2$ (12)	
Right sided $(a{=}0,\, b{=}\infty,\, \gamma{<}1)$	$W_{k-1} = \gamma A^{\mathsf{T}} W_k A + c^{\mathsf{T}} c$ (13) $\xi_{k-1} = \gamma A^{\mathsf{T}} \xi_k + c^{\mathsf{T}} y_{k-1}$ (14) $\kappa_{k-1} = \gamma \kappa_k + y_{k-1}^2$ (15)	
Finite support $(-\infty{<}a{\le}b{<}\infty)$	$W_{k+1} = \gamma^{-1} A^{-\mathsf{T}} W_k A^{-1} - \gamma^{a-1} w_{k+a} \left(A^{a-1}\right)^{\mathsf{T}} c^{\mathsf{T}} c \left(A^{a-1}\right)$ $+\gamma^b w_{k+b+1} \left(A^b\right)^{\mathsf{T}} c^{\mathsf{T}} c A^b$ (16) $\xi_{k+1} = \gamma^{-1} A^{-\mathsf{T}} \xi_k - \gamma^{a-1} w_{k+a} \left(A^{a-1}\right)^{\mathsf{T}} c^{\mathsf{T}} y_{k+a}$ $+\gamma^b w_{k+b+1} \left(A^b\right)^{\mathsf{T}} c^{\mathsf{T}} y_{k+b+1}$ (17) $\kappa_{k+1} = \gamma^{-1} \kappa_k - \gamma^{a-1} w_{k+a} y_{k+a}^2 + \gamma^b w_{k+b+1} y_{k+b+1}^2$ (18) --- $W_{k-1} = \gamma A^{\mathsf{T}} W_k A + \gamma^a w_{k+a-1} \left(A^a\right)^{\mathsf{T}} c^{\mathsf{T}} c A^a$ $-\gamma^{b+1} w_{k+b} \left(A^{b+1}\right)^{\mathsf{T}} c^{\mathsf{T}} c A^{b+1}$ (19) $\xi_{k-1} = \gamma A^{\mathsf{T}} \xi_k + \gamma^a w_{k+a-1} \left(A^a\right)^{\mathsf{T}} c^{\mathsf{T}} y_{k+a-1}$ $-\gamma^{b+1} w_{k+b} \left(A^{b+1}\right)^{\mathsf{T}} c^{\mathsf{T}} y_{k+b}$ (20) $\kappa_{k-1} = \gamma \kappa_k + \gamma^a w_{k+a-1} y_{k+a-1}^2 - \gamma^{b+1} w_{k+b} y_{k+b}^2$ (21)	
Forward Recursions $(k \to k{+}1)$, $k' = \min(-b, 0)$ | *Backward Recursions* $(k \to k{-}1)$, $k' = \max(K-a, K)+1$. *Initialization* $W_{k'} = 0$, $\xi_{k'} = 0$, $\kappa_{k'} = 0$. $w_k \in \{0,1\}$. Set $w_k = 1$ for $k \in \{1,\dots,K\}$ and $w_k = 0$ otherwise.		

minimization problem, i.e.

$$\hat{x}_k = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} J_a^b(k,x,\theta) \tag{22}$$

$$= \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \left(x^{\mathsf{T}} W_k x - 2x^{\mathsf{T}} \xi_k + \kappa_k\right) \tag{23}$$

$$= W_k^{-1} \xi_k \,. \tag{24}$$

If we further want to restrict the allowed signal shapes, constraints on x can be added. For instance, linear constraints with offset

$$x = Hv + h \tag{25}$$

with given $H \in \mathbb{R}^{N \times M}$, $h \in \mathbb{R}^{N \times 1}$, $M \in \mathbb{N}, M \le N$ and unknown $v \in \mathbb{R}^{M \times 1}$, leads to the closed form solution

$$\hat{v}_k = (H^{\mathsf{T}} W_k H)^{-1} H^{\mathsf{T}} (\xi_k - W_k h) \,. \tag{26}$$

This and many other alternative constrained quadratic optimization problems are summarized in [12].

2.4 Pattern Detection With Likelihood Ratio

Given the interval $\{k+a,\dots,k+b\}$, we want to discriminate between two classes of patterns. The first class is characterized by a fixed *output vector* c_1 and a linearly constrained *initial state vector* $x_1 = H_1 v_1 + h_1$. The second class is characterized correspondingly by c_2, H_2, and h_2. The discrimination is based on the error ratio

$$q_k = \frac{\min\limits_{v_2} J_a^b(k, H_2 v_2 + h_2, \theta_2)}{\min\limits_{v_1} J_a^b(k, H_1 v_1 + h_1, \theta_1)} \in \mathbb{R}_+ \,. \tag{27}$$

In statistical theory, this energy ratio, up to irrelevant constants, is actually interpreted as a likelihood ratio between two hypotheses [12]. The log-likelihood ratio (LLR) between two models is, using (27),

$$\mathrm{LLR}_k = -\frac{1}{2} \log(q_k) \,. \tag{28}$$

We note that the local minima of the error ratio coincides with the local maxima of the LLR. Since the error ratios are

better displayed in a log scale, we subsequently use LLR's. Note that if the patterns to compare only differ by their *initial state vectors* x_1 and x_2, but share a common parameter set $\theta = \theta_1 = \theta_2$, then the recursions coincide.

3 The Composite Cost

As we will see in the first example in Section 4.1, *cost segments* are useful to detect known classes of patterns buried in noise. Nevertheless, biological signals are commonly superimposed on some interferences. Luckily, interferences, such as a wandering baseline, often have a wider time-scale than the patterns of interest. To improve the robustness of pattern detection, we include the interference with its own time-scale into our model.

For this purpose, we define $P \in \mathbb{N}$ *cost segments* of respective parameters $\theta_p = (c_p, \gamma_p)$ and boundaries $a_p \in \mathbb{Z} \cup \{-\infty\}$, $b_p \in \mathbb{Z} \cup \{+\infty\}$, $a_p \leq b_p$, $p \in \{1, \dots, P\}$ and define $\mathbf{\Theta} = (\theta_1, \theta_2, \dots, \theta_P)$, leading to the general form of a *composite cost*

$$\tilde{J}(k, x, \mathbf{\Theta}) = \sum_{p=1}^{P} J_{a_p}^{b_p}(k, x, \theta_p) \; . \tag{29}$$

Note that the observation vectors c_p and decaying factor γ_p, contained in the parameter set θ_p, are specific to each *cost segment*, but the *state-transition matrix* A and the *state vector* x are in common to all *cost segments*. This is actually not a restriction since c_p can select models when using model superposition (cf. Section 3.2).

Since each *cost segment* $J_{a_p}^{b_p}(k, x, \theta_p)$ has its recursive computation with quantities $W_k^{(p)}, \xi_k^{(p)}$ and $\kappa_k^{(p)}$ (see Section 2.2), the *composite cost* $\tilde{J}(k, x, \mathbf{\Theta})$ is thereof obtained according to (6), using the substitutes

$$W_k = \sum_{p=1}^{P} W_k^{(p)}, \; \xi_k = \sum_{p=1}^{P} \xi_k^{(p)}, \; \kappa_k = \sum_{p=1}^{P} \kappa_k^{(p)} \; . \tag{30}$$

Furthermore, the minimization of *cost segment* (Section 2.3) also applies for *composite costs*. Thus, the *composite cost* (29) is still efficiently computed. In the following, we highlight different uses of such *composite costs* with direct applications. Note that a *composite cost* of *composite costs* is also a *composite cost*, which allows recursive application for more flexibility.

3.1 Segment Chaining

Segment chaining refers to *composite costs* with non-overlapping segments, i.e. such that in (29) $b_p < a_{p+1}$ for $p \in \{1, \dots, P-1\}$.

3.2 Model Superposition

We consider the superposition of $M \in \mathbb{N}$ models with parameters c_m, A_m and $x^{(m)}$, and common γ and common bounds a and b, i.e.

$$\sum_{i=k+a}^{k+b} \gamma^{i-k} \left(y_i - \sum_{m=1}^{M} c_m A_m^{i-k} x^{(m)} \right)^2 \; , \tag{31}$$

which is equivalent to a *composite cost* consisting of a single *cost segment* $J_a^b(k, x, \theta)$ with parameters

$$A = \operatorname{diag}(A_1, \dots, A_M), \tag{32}$$
$$x^{\mathsf{T}} = [(x^{(1)})^{\mathsf{T}}, \cdots, (x^{(M)})^{\mathsf{T}}], \tag{33}$$
$$\theta = ([c_1, \dots, c_M], \gamma) \; . \tag{34}$$

3.3 Multi Model, Multi Time-Scale

We here combine the ideas from Sections 3.1 & 3.2. Assume we have $M \in \mathbb{N}$ signal models. Each model is described by its parameters $(A_m, c_m, x^{(m)})$, $m \in \{1, \dots, M\}$ with its own time-scale, i.e. each model is active in some specific intervals. In order to deal with those multiple time-scales, we rearrange the time axis into $P \in \mathbb{N}$ non-overlapping segments, and define the assignment matrix $\Lambda \in \{0, 1\}^{P \times M}$ such that $\Lambda_{p,m} = 1$ indicates that model "m" is active in segment "p".

We denote $C = \operatorname{diag}(c_1, \dots, c_M)$, $A = \operatorname{diag}(A_1, \dots, A_M)$, and $x^{\mathsf{T}} = \left[(x^{(1)})^{\mathsf{T}} \;\; \cdots \;\; (x^{(M)})^{\mathsf{T}}\right]$. Each *cost segment* is characterized by its lower and upper bounds $a_p \in \mathbb{Z}$ and $b_p \in \mathbb{Z}$, $a_p \leq b_p$, $b_p < a_{p+1}$, and with the associated factor γ_p. We define $\theta_p = (\Lambda_p C, \gamma_p)$ with Λ_p the p^{th} row of Λ, such that all models active in segment "p" are selected. Hence, we get a *composite cost* as in (29). Note that the very left and the very right segment might be of infinite length ($a_1 = -\infty$ or $b_P = +\infty$).

3.4 Multiple Input Channels

We have $S \in \mathbb{N}$ input channels $y = [y^{(1)}, \dots, y^{(S)}]$. Each channel s is associated with its signal model of order $N_s \in \mathbb{N}$ and parameters $c_s \in \mathbb{R}^{1 \times N_s}$, $A_s \in \mathbb{R}^{N_s \times N_s}$ and $x^{(s)} \in \mathbb{R}^{N_s \times 1}$, and common γ, and common bounds a and b, leading to the *cost segment* $J_a^b(k, x, \theta_s)$ computed on the observation channel s with parameters

$$A = \operatorname{diag}(A_1, \dots, A_S), \tag{35}$$
$$x^{\mathsf{T}} = [(x^{(1)})^{\mathsf{T}}, \cdots, (x^{(M)})^{\mathsf{T}}], \tag{36}$$
$$\theta_s = ([\underbrace{0, \dots, 0}_{\sum_{\sigma=1}^{s-1} N_\sigma}, c_s, \underbrace{0, \dots, 0}_{\sum_{\sigma=s+1}^{S} N_\sigma}], \gamma) \; . \tag{37}$$

Then, the total cost, equivalent to the *composite cost* $\tilde{J}(k, x, \mathbf{\Theta})$, is

$$\sum_{s=1}^{S} \sum_{i=k+a}^{k+b} \gamma^{i-k} \left(y_i^{(s)} - c_s A_s^{i-k} x^{(s)} \right)^2 \; . \tag{38}$$

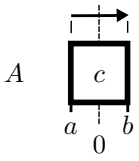


Figure 1: Graphical symbol of the *cost segment* $J(k, x, \theta)$ with *system matrix* A, *output vector* c and boundaries a and b. The top arrow represents the window (rectangular) and the computation direction of the recursions (forward).

4 Examples

4.1 Example: ECG Wave Discrimination

This first example illustrates the ideas of pattern detection and discrimination using a single *cost segment*. We focus on the detection of known and repetitive signal shapes. While in standard ECG signals atrial waves (known as P waves) are of weak amplitude, in esophageal electrocardiography atrial waves are of similar amplitude as the ventricular waves. This results from the anatomical proximity of the esophagus to the atria. After having identified a single atrial or ventricular wave in the signal, it is of high interest to efficiently locate all its repetitions within the signal. As common for ECG signals, the exact amplitudes are of minor diagnostic value and can vary due to external factors (e.g. breathing). Thus, any amplitude scale of the shape shall be allowed.

Problem Setup

Let $y \in \mathbb{R}^{K\times 1}, K \in \mathbb{N}$ be a single-channel ECG signal and k_A, $k_V \in \{1, \ldots, K\}$ the indices of the identified reference atrial (A) and ventricular (V) shapes. We want to localize and discriminate all repetitions of these atrial and ventricular shapes, up to an amplitude scale.

Solution

We model both shapes with a fourth order polynomial of finite length with parameters

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad c = [1\,0\,0\,0\,0]\ . \tag{39}$$

With this choice, for any given *initial state vector* $x = [\beta_0\ \beta_1 \ldots \beta_4]^\mathsf{T}$, we have the equality $s_k(x) = cA^k x = \beta_0 + \beta_1 k + \beta_2 k^2 + \ldots + \beta_4 k^4$. We use a single *cost segment* with boundaries $-a = b = 50$ and compute W_k, ξ_k and κ_k for the *cost segment* using forward recursions (16), (17) and (18) for all $k \in \{1, \ldots, K\}$. A graphical representation of this *cost segment* with all its relevant parameters is given in Fig. 1.

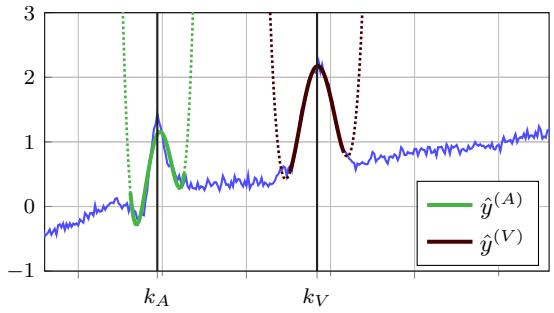


Figure 2: Squared error minimized fit of a 4th order polynomial to the reference atrial (A) and ventricular (V) ECG waves by estimating the initial states x_{k_A} and x_{k_V} accordingly. The atrial signal estimate is $\hat{y}_k^{(A)} = cA^{k-k_A}\hat{x}_{k_A}$ and the ventricular estimate $\hat{y}_k^{(V)} = cA^{k-k_V}\hat{x}_{k_V}$. The solid intervals of the polynomial lines $\hat{y}^{(A)}$ and $\hat{y}^{(V)}$ correspond to the intervals $\{k_A+a, \ldots, k_A+b\}$ and $\{k_V+a, \ldots, k_V+b\}$ considered in the cost function with $-a = b = 50$.

Let the *initial state vectors* $\hat{x}_{k_A}$ and $\hat{x}_{k_V}$ be the unconstrained estimates (24) that minimize the *cost segment* at reference index k_A and k_V, respectively. In our example, we get

$$\hat{x}_{k_A}^\mathsf{T} = [1.11\ 0.045\ -0.013\ -1.2{\times}10^{-4}\ 4.0{\times}10^{-5}]$$
$$\hat{x}_{k_V}^\mathsf{T} = [2.15\ 0.008\ -0.009\ 1.3{\times}10^{-6}\ 1.3{\times}10^{-5}].$$

For the pattern detection and discrimination, we want to allow an offset and an amplitude scale (to be estimated) of the (A) and (V) shapes. This can be encoded by constraining the *initial state vector* $x_k = Hv_k$ as in (25) with $v_k \in \mathbb{R}^2$ and where the first component of v_k is the amplitude scale and the second component the offset. For each shape, we have a specific H, denoted as H_A and H_V, and defined as

$$H_A^\mathsf{T} = \begin{bmatrix} 0 & \beta_1^{(A)} & \beta_2^{(A)} & \beta_3^{(A)} & \beta_4^{(A)} \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{40}$$

where $\beta_i^{(A)}$ is the i^{th} component of $\hat{x}_{k_A}$. H_V is defined correspondingly. Note that the offset components $\beta_0^{(A)}$ and $\beta_0^{(V)}$ are not used because the shapes are superimposed with an unknown offset. The fit to the atrial and ventricular signal using the constraints H_A and H_V is displayed in Fig. 2. For a robust detection of each shape, we additionally need a model when there is no shape. For this case, we use an offset model, parametrized by

$$H_0^\mathsf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We define the log-likelihood ratio for the atria as

$$\mathrm{LLR}_k^{(A)} = -\frac{1}{2}\log\frac{\min\limits_{v_A} J(k, H_A v_A, \theta)}{\min\limits_{v_0} J(k, H_0 v_0, \theta)} \in \mathbb{R}_+\ , \tag{41}$$

and, correspondingly, $\mathrm{LLR}_k^{(V)}$ for the ventricle.

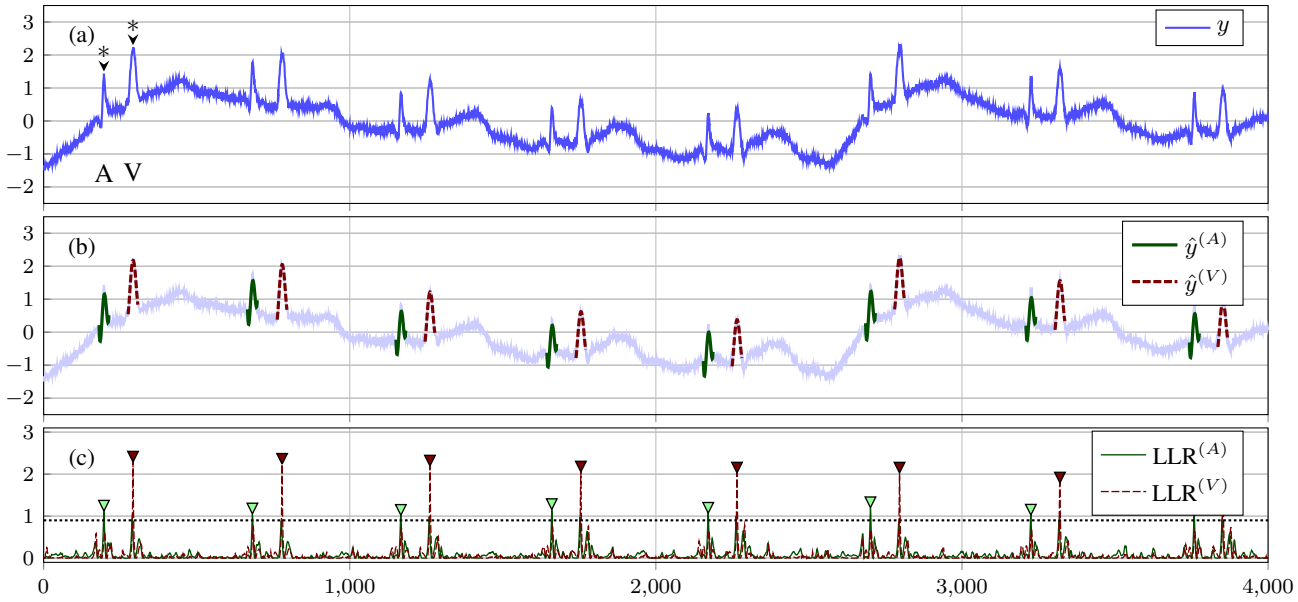


Figure 3: Atrial (A) and ventricular (V) wave pattern detection and discrimination in a esophageal ECG signal y. (Unlike in standard ECG signals, in esophageal ECGs, atrial and ventricular signals are of similar shapes and amplitudes.) (a) Recorded esophageal ECG signal y. The polynomial coefficients used in the pattern-defining constraints H_A and H_V are taken from the index waves (*). (b) ECG signal y, superimposed by detected shapes. (c) Likelihood ratios for atrial ($\mathrm{LLR}^{(A)}$) and ventricular ($\mathrm{LLR}^{(V)}$) patterns with detected peaks (triangles) and chosen detection threshold (dotted line).

The final decision on atrial shape (H_A) or ventricular shape (H_V) at any index k is taken by setting an appropriate threshold on the LLR's and taking the one with the highest LLR value. This threshold is the only tuning parameter. Therefore its proper selection is often easily done manually, as in this example (see dotted line in Fig. 3c). If artifacts create false-positive detections, the allowed range of shape's amplitudes (first component of v_k) can additionally be restricted in order to increase the detection robustness. In our example we would optionally restrict the shape amplitude to values around 1, which corresponds to the reference pulse amplitude. Fig. 3 shows a typical recording with results.

4.2 Example: ECG Wave Discrimination Using Multi-time-scale Baseline Estimates

This second example demonstrates the use of *composite costs* by additionally estimating the baseline signal. Including the baseline into the model increases the detection robustness on signals with large baseline intereferences or additive noise. A baseline signal usually has a wider time-scale than the ECG waves. Therefore, our algorithm needs to concurrently deal with signals of different time-scales.

Problem Setup

The problem setup is equivalent to the one in Example 4.1 but with increased additive Gaussian noise on the input signal y.

Solution

We model both (A) and (V) shapes with a fourth order polynomial of finite length with parameters (A_1, c_1) as in (39). Additionally, we model the baseline with a third order polynomial of parameters (A_2, c_2). To deal with a baseline with a wide time-scale, we use a *composite cost* consisting of 2 models and 4 *cost segments* as in Section 3.3 with the segments (cf. Fig. 5):

segm.	a	$b+1$	γ	$c \in \mathbb{R}^{1\times 9}$
1	$-\infty$	$-\Delta$	1.01	$[0_{1\times 5}\ c_2]$
2	$-\Delta$	0	1.005	$[c_1\ c_2]$
3	0	$+\Delta$	0.995	$[c_1\ c_2]$
4	$+\Delta$	$+\infty$	0.99	$[0_{1\times 5}\ c_2]$

Note that the baseline model is active in all segments, while the shape model is only active in segments 2 and 3. $W_k^{(p)}$, $\xi_k^{(p)}$ and $\kappa_k^{(p)}$ are computed in forward direction for segment 1 and 2, and computed in backward direction for segment 3 and 4.

Let the *initial state vectors* $\hat{x}_{k_A}$ and $\hat{x}_{k_V}$ be the unconstrained estimates (24) that minimize the *composite cost* at reference index k_A and k_V, respectively. Since we have a baseline model, we only allow an amplitude scale of the shape. This can be encoded by constraining the *initial state vector* $x_k = Hv_k$ as in (25) with $v_k \in \mathbb{R}^5$ and where the first component of v_k is the amplitude scale of the shape and the remaining four components the coefficients of the polynomial baseline. For each shape we have a specific H,

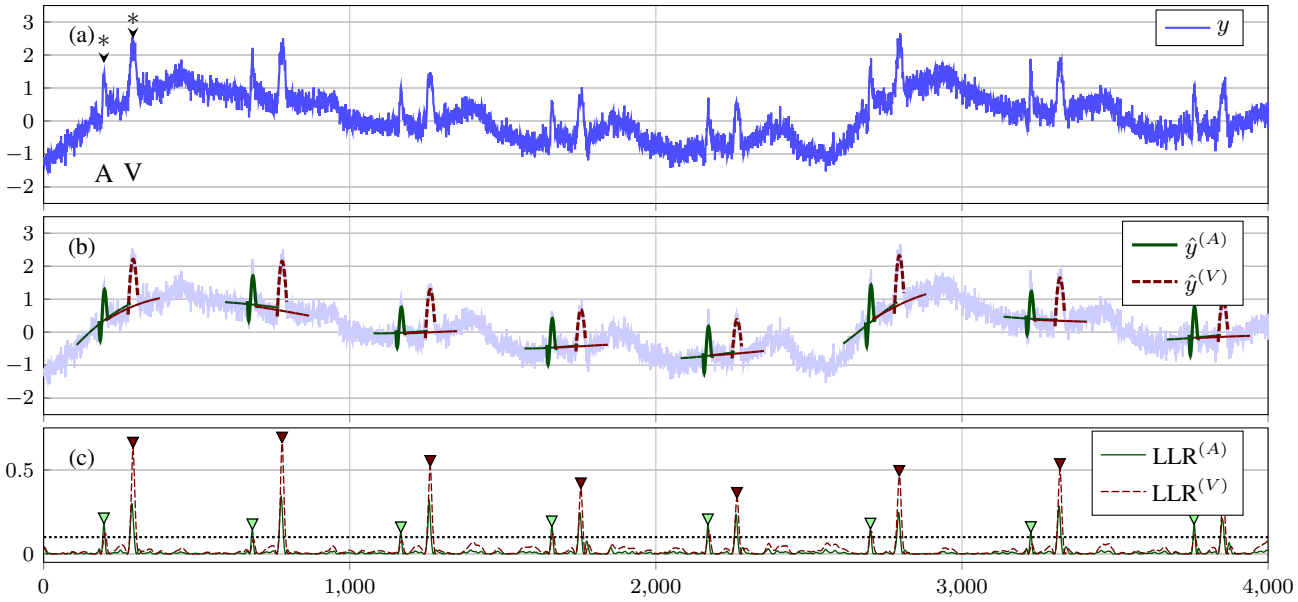


Figure 4: Atrial (A) and ventricular (V) wave pattern detection and discrimination with baseline estimation in a esophageal ECG signal with increased additive Gaussian noise. (a) Recorded esophageal ECG signal y. The polynomial coefficients used in the pattern-defining constraints H_A and H_V are taken from the index waves (*). (b) ECG signal y with superimposed detected shapes and local baseline estimate (horizontal lines). (c) Likelihood ratios for atrial ($\mathrm{LLR}^{(A)}$) and ventricular ($\mathrm{LLR}^{(V)}$) patterns with detected peaks (triangles) and chosen detection threshold (dotted line).

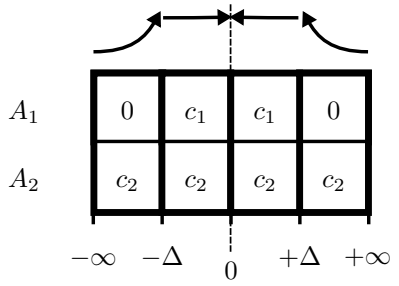


Figure 5: Graphical representation of the *composite cost* $\tilde{J}(k, x, \boldsymbol{\Theta})$ with two models (rows) and four segments (columns). Upper row corresponds to shape model (1), the lower row, to the baseline model (2). The top arrows represent the window shape (decaying or rectangular) and the computation direction (forward or backward).

denoted as H_A and H_V, and defined as

$$H_A^{\mathsf{T}} = \left[\begin{array}{c|c} \hat{x}_{k_A}^{\mathsf{T}} & 0_{1\times 4} \\ \hline 0_{4\times 5} & I_4 \end{array}\right] \tag{42}$$

with the identity matrix $I_4 \in \mathbb{R}^{4\times 4}$. H_V is defined correspondingly. When there is no shape, only the baseline is present, which can be expressed with

$$H_0^{\mathsf{T}} = \left[\begin{array}{c|c} 0_{4\times 5} & I_4 \end{array}\right] . \tag{43}$$

We define the log-likelihood ratios $\mathrm{LLR}_k^{(A)}$ and $\mathrm{LLR}_k^{(V)}$ as in (41). Finally, A and V waves are localized and discriminated using appropriate thresholding on the LLR's. The threshold on the LLR is again the only tuning parameter and is selected manually. Fig. 4 shows the results. Note that, despite the additive Gaussian noise on y, the LLR peaks in this example are sharpened and the detection robustness is increased in contrast to Example 4.1. This is the benefit of including the baseline into the model.

4.3 Example: P Wave Discrimination in Multi-Channel ECG

This example demonstrates the use of multi-channel input signals to enhance the robustness of detection in low signal-to-noise regimes. The task is to detect repetitions of similar shapes of P waves in a standard 3 lead surface ECG signal. By contrast to many other proposed approaches for P wave detection, our method will not need any prior QRS wave masking. Therefore our approach is also appropriate for any arrhythmia disorder of repetitive signal shapes (e.g. atrioventricular dissociation, atrial flutter, etc.).

Problem Setup

Let $y = [y^{(1)}, y^{(2)}, y^{(3)}] \in \mathbb{R}^{K\times 3}, K \in \mathbb{N}$ be a (pathologic) multi-channel ECG signal (Lead I, II, III) with two distinct atrial triggers: sinus node (n) and an additional atrial focus (f). Let k_n, $k_f \in \{1, \ldots, K\}$ be the indices of the two identified reference P waves. We want to localize all repetitions of these distinct atrial triggers by shape comparison in the multi-channel signal y.

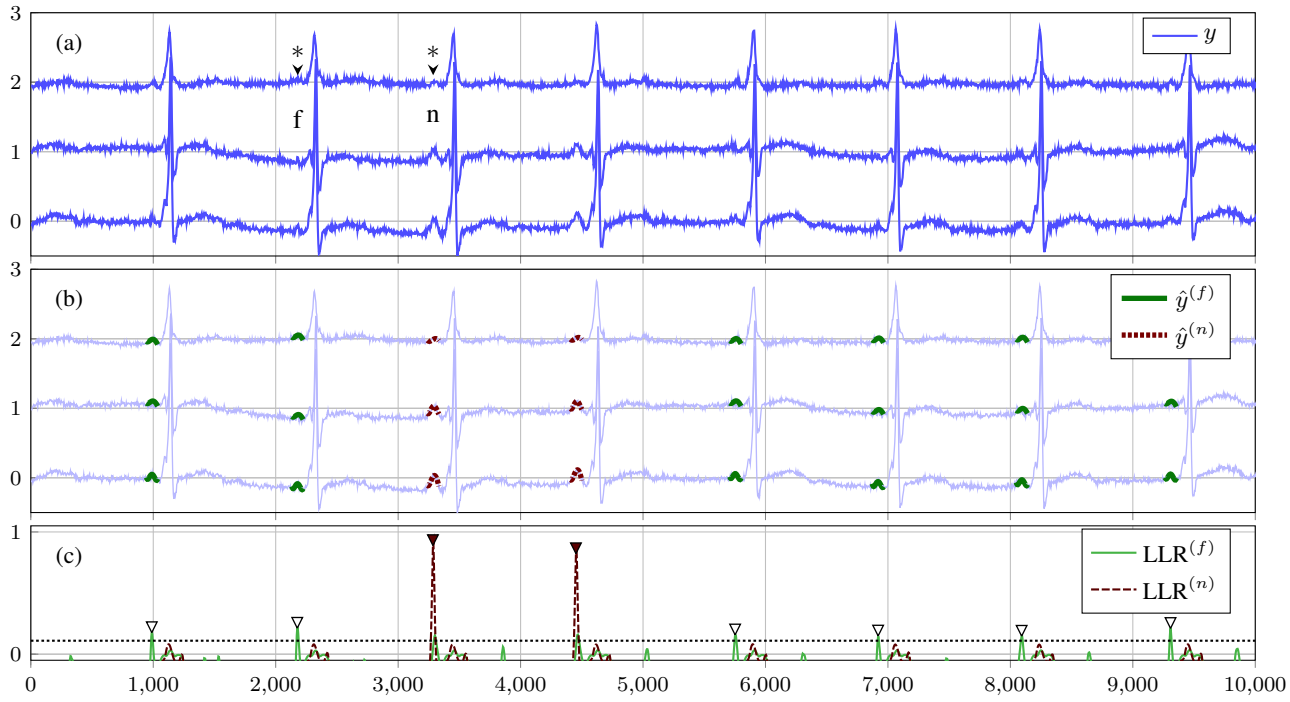


Figure 6: P wave detection and discrimination in a multi-channel surface ECG signal (Lead I, II and III). (a) Recorded surface ECG signals y with reference P waves (*). (b) ECG signal y with superimposed detected shapes $\hat{y}^{(n)}$ and $\hat{y}^{(f)}$. (c) $\mathrm{LLR}^{(n)}$ and $\mathrm{LLR}^{(f)}$ with detected peaks (triangles) and chosen detection threshold (dotted line).

Solution

For each channel we use a *composite cost* consisting of 2 models and 4 segments as in Example Fig. 5. For the P wave, we use a polynomial model of fourth order with parameters (A_1, c_1). For the baseline, we use an offset model with parameters $A_2 = [1]$ and $c_2 = [1]$. The three *composite costs* are combined as in Section 3.4, leading to a new *composite cost* with $A \in \mathbb{R}^{18\times 18}$ and $c \in \mathbb{R}^{1\times 18}$.

Let the *initial state vectors* $\hat{x}_{k_n}^{(s)}$ be the unconstrained estimates (24) that minimizes the *composite cost* of channel s at reference index k_n and similarly for $\hat{x}_{k_f}^{(s)}$. Since we do not allow amplitude scale, we estimate the baseline offset in each channel, leading to a constrained state $H_n v_n + h_n$ when there is a shape (n), and $H_n v_n$ when there is no shape, with

$$H_n = \begin{bmatrix} 0_{5\times 1} \\ \hline 1 \\ \hline 0_{5\times 1} \\ \hline 1 \\ \hline 0_{5\times 1} \\ \hline 1 \end{bmatrix} \in \mathbb{R}^{18\times 1}\,,\; h_n = \begin{bmatrix} x_{k_n}^{(1)} \\ \hline 0 \\ \hline x_{k_n}^{(2)} \\ \hline 0 \\ \hline x_{k_n}^{(3)} \\ \hline 0 \end{bmatrix} \in \mathbb{R}^{18\times 1}\,.$$

Having no amplitude scale reduces the flexibility but increases the overall discrimination robustness. Finally, we use the likelihood ratio

$$\mathrm{LLR}_k^{(n)} = -\frac{1}{2}\log\frac{\min\limits_v \tilde{J}(k, H_n v + h_n, \Theta)}{\min\limits_v \tilde{J}(k, H_n v, \Theta)}\,. \quad (44)$$

We proceed equally for $\mathrm{LLR}_k^{(f)}$ with the second pattern defined by h_f and H_f using $\hat{x}_{k_f}$. The final decision on the P wave locations is taken by appropriate thresholding on the LLR's. Fig. 6 shows the results. We note that the algorithm is not restricted to detect or discriminate only two shapes but can handle several shapes.

5 Conclusion

We have proposed computationally efficient methods for signal detection, separation, and discrimination, based on state space models. We have shown how to deal with patterns superimposed with interference signals of variable time scales, such as baseline signals. The proposed methods were evaluated within a research project on a diagnostic esophageal ECG device, and are internally widely used on data sets obtained from the associated clinical trials in human cardiology. We processed a large amount of ECG and other biological signal recordings. Once getting used to these methods, the development of fast and efficient signal processing solutions is straightforward, even for rapid algorithm prototyping and experimental analysis. These filter methods are complement to the standard IIR and FIR

filters, but are also a substitute since our LTI state space methods are a generalization of IIR and FIR filters.

We will continue to unify our ideas to provide a versatile framework, including interpolation methods and methods to dynamically deal with signals of variable timescales. We also work towards profound benchmarking of our method and comparisons with known methods from literature, rating computational complexity and detection performance. In addition, we will provide validated practical applications implemented on dedicated hardware in the field of cardiology, based on the illustrative examples given in this work.

6 Acknowledgment

We are grateful for the valuable collaboration with the Department of Cardiology, Bern University Hospital. All illustrative examples in this work use esophageal or surface ECG signals obtained during a collaborative clinical trial. The trial protocol as well as the trial equipment were provided by the Bern University of Applied Sciences. This research was supported by the Swiss National Science Found (SNF) grant CR23I2_166030.

References

[1] S. Mallat. *A wavelet tour of signal processing*. Academic press, 2009.

[2] G. Wahba. *Spline models for observational data*, volume 59. Siam, 1990.

[3] G. Wahba. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(3):364–372, 1978.

[4] W. E. Wecker and C. F. Ansley. The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78(381):81–89, 1983.

[5] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang. The factor graph approach to model-based signal processing. *Proceedings of the IEEE*, 95(6):1295–1322, 2007.

[6] N. Zalmai, R. A. Wildhaber, D. Clausen, and H.-A. Loeliger. Inferring depolarization of cells from 3d-electrode measurements using a bank of linear state space models. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.

[7] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

[8] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Journal of Fluids Engineering*, 83(1):95–108, 1961.

[9] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear estimation*, volume 1. Prentice Hall Upper Saddle River, NJ, 2000.

[10] Reller C. *State-Space Methods in Statistical Signal Processing: New Ideas and Applications*. PhD thesis, No. 20584, ETH Zurich, 2012.

[11] J. Durbin and S. J. Koopman. *Time series analysis by state space methods*. Number 38. Oxford University Press, 2012.

[12] N. Zalmai, R. A. Wildhaber, and H.-A. Loeliger. Detection and estimation in multi-channel signals with autonomous linear state space models. In *(submitted)*, 2016.