A Spline Framework for Optimal Representation of Semiperiodic Signals

by

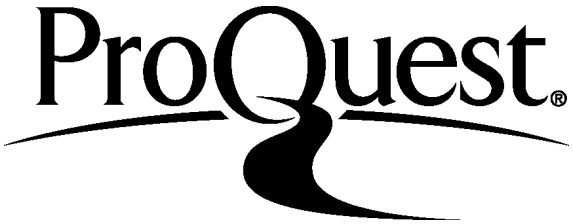
Farzin G. Guilak

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Electrical and Computer Engineering

Dissertation Committee:
James McNames, Chair
Robert Fountain
Dan Hammerstrom
Eric A. Wan

Portland State University
2015



ABSTRACT

Semiperiodic signals possess an underlying periodicity, but their constituent spectral components include stochastic elements which make it impossible to analytically determine locations of the signal's critical points. Mathematically, a signal's critical points are those at which it is not differentiable or where its derivative is zero. In some domains they represent characteristic points, which are locations indicating important changes in the underlying process reflected by the signal.

For many applications in healthcare, knowledge of precise locations of these points provides key insight for analytic, diagnostic, and therapeutic purposes. For example, given an appropriate signal they might indicate the start or end of a breath, numerous electrophysiological states of the heart during the cardiac cycle, or the point in a stride at which the heel impacts the ground. The inherent variability of these signals, the presence of noise, and often, very low signal amplitudes, makes accurate estimation of these points challenging.

There has been much effort in automatically estimating characteristic point locations. Approaches include algorithms operating in the time domain, on various transformations of the data, and using different models of the signal. These methods apply a wide variety of techniques ranging from simple thresholds and search windows to sophisticated signal processing and pattern recognition algorithms. Existing approaches do not explicitly use prior knowledge of characteristic point locations in their estimation.

This dissertation first develops a framework for an efficient parametric representation of semiperiodic signals using splines. It then implements an instance of that framework to optimally estimate locations of characteristic points, incorporating prior knowledge from manual annotations on training data. Splines represent signals in a piecewise manner by applying an interpolant to constraint points on the signal known as knots. The framework allows choice of interpolant, objective function, knot initialization algorithm, and optimization algorithm. After initialization it iteratively modifies knot locations until the objective function is met.

For optimal estimation of characteristic points the framework relies on a Bayesian objective function, the *a posteriori* probability of knot locations given the observed signal. This objective function fuses prior knowledge, the observed signal, and its spline estimate. With a linear interpolant, knot locations after optimization serve as estimates of the signal's characteristic points.

This implementation was used to determine locations of 11 characteristic points on a prospective test set comprising 200 electrocardiograph (ECG) signals from 20 subjects. It achieved a mean error of -0.4 milliseconds, less than one quarter of a sample interval. A low bias is not sufficient, however, and the literature recognizes error variance to be the more important factor in assessing accuracy. Error variances are typically compared to the variance of manual annotations provided by reviewers. The algorithm was within two standard deviations for six of the characteristic points, and within one sample interval of this criterion for another four points.

The spline framework described here provides a complementary option to existing methods for parametric modeling of semiperiodic signals, and can be tailored to represent semiperiodic signals with high fidelity or to optimally estimate locations of their characteristic points.

DEDICATION

For my dear wife, Karen Virginia. For almost thirty years, your love, friendship, selflessness, and unending support has given me the strength, courage, and confidence to take on challenges I wouldn't have accepted otherwise. I am a better person because of you.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. James McNames. Completion of this dissertation spanned many years and included a number of significant obstacles, not all technical in nature. Without his continued support, encouragement, mentorship, advocacy, and great patience for a late-career, part-time student, my effort would not be completed. And without his guidance, depth of knowledge, and high standards for technical rigor and clarity of communication, my effort would not be as strong.

> *"Part of getting this degree is having the perseverance to jump through all of the required hoops, but there aren't usually this many, they're not usually so high, and they're not usually on fire."* [57]

Thank you for everything you did to reduce the number of hoops, lower their height, and extinguish the fires.

My sincere thanks to the members of my committee, Dr. Dan Hammerstrom, Dr. Eric Wan, and Dr. Robert Fountain. I greatly appreciate your participation, time, and efforts in reviewing my dissertation. And additional appreciation to Dr. Hammerstrom for his support, counsel, and mentorship over many years.

I would like to thank Dr. Richard Sloan for graciously providing the ECG data used for the work described in Chapter 4, which was obtained through research funded by NIH Grant R01 HL63872.

And finally, my unending gratitude to my parents, Dr. Nahid T. Guilak and Dr. Hooshang Guilak. Throughout my life they have instilled and encouraged a

love for learning — so much so that it is an integral part of my identity. And they have taught by example, accepting new challenges and embracing new experiences all of their lives. They have both set extremely high standards in their professional and personal lives, standards that I aspire to meet. Thanks for all of your love and support, and for all of the opportunities you worked so hard to provide for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 OVERVIEW

This dissertation addresses the problem of optimally identifying characteristic points in semiperiodic signals. Semiperiodic signals are essentially periodic in nature, but without meeting the strict definition of periodicity due to temporal variations of their harmonics. Like periodic signals, semiperiodic signals demonstrate a repeating pattern creating a distinct and recognizable morphology in each cycle, but semiperiodic morphologies exhibit slow variations (relative to the fundamental frequency of the signal) in the times and amplitudes of their features.

The signal's characteristic points are locations of visual prominence that impart important information regarding the underlying process reflected in the signal, and can be viewed as the morphological features defining each cycle. They are modeled as locations of high curvature or local extrema and are consistent with the notion of mathematical critical points. Identifying such points is important in efficiently representing the signal, delineating it into regions of interest, or in identifying and tracking key morphological markers.

An example used in developing the algorithms, and detailed in subsequent sections, is that of the electrocardiogram signal. The ECG signal is semiperiodic in nature and has numerous clinically-relevant morphological points in each cycle.

These characteristic points directly reflect cardiac activity and are used by clinicians to determine a subject's health state. Unpredictable changes in the subject's physiology, as well as various types of noise that are invariably present in the signal, make optimally locating characteristic points a challenging problem.

The novel algorithmic framework described in this dissertation addresses these challenges using a time-domain representation of semiperiodic signals to determine locations of characteristic points for each cycle. The framework describes semiperiodic signals using splines, which represent signals in a piecewise manner by applying an interpolant to constraint points on the signal known as knots.

To optimally locate the signal's characteristic points they are modeled as knots in a spline representation, and their best locations are determined using Bayesian estimation. A training set annotated by human experts provides the required prior distribution, and is used in conjunction with a likelihood derived from the original signal and its the spline representation to compute a figure of merit, the *a posteriori* probability (or simply, posterior) of knot locations given the observed signal. Finding the knot locations that maximize the figure of merit provides an optimal estimate of the signal's characteristic points.

Adopting a Bayesian approach to locate characteristic points provides a measure of noise tolerance as well as improved performance in the presence of ambiguous or varying characteristic points. Here the term "noise" refers generically to signal distortion as may be caused by the data acquisition system (for example, quantization), interference from man-made signals similar to that of interest (such as other electronic equipment), or inescapable additive noise from natural sources. Ambiguous characteristic points occur when morphological features are not clearly defined or when they vary in the data under analysis. In such cases, the algorithm relies more heavily on the priors for its estimate of knot locations, helping reduce variability of characteristic point estimates and thereby improve their accuracy.

Although any type of learning algorithm requiring training data can be said

to determine priors in some sense, the differentiating factor in this effort is its explicit use of probability density estimates describing characteristic point times and amplitudes. Capturing the priors in this manner enables employing Bayesian estimation in this domain.

Depending on the requirements of a system using this algorithm, the prior information can be determined using annotated data from a single subject or from multiple subjects, or perhaps through other means such as a model of the underlying process. In the case of priors derived using data of a single subject, the priors can be considered to be tuned, or biased, to that particular subject. This gives the optimization algorithm a greater ability to handle noise and ambiguity for that subject's characteristic point locations. However, its ability to generalize to morphologies that are significantly different will be limited. In the case of priors derived from data of multiple subjects, the priors will be more general and capable of representing a wider range of morphologies, but will have reduced ability to tolerate noise or ambiguity.

The algorithm described here can be used in any setting where accurate estimates of characteristic point locations are required. Examples for the ECG signal include offline processing of long-term ECG recordings to assess subjects' health state, in real time transport monitoring, or on wearable devices to determine subjects' stress levels "in-the-moment", and drive therapeutic interventions. It could also be used in diagnostic systems to identify characteristic point locations that are then used to derive specific, established metrics used by clinicians for diagnostic or prognostic purposes.

Comparing the performance of the algorithm described here against existing commercial diagnostic systems is problematic. Although such systems provide metrics derived from characteristic point locations, they do not provide the locations themselves. These algorithms are typically proprietary, and obtaining access to their internal state will require special licensing arrangements with the

manufacturer. In addition, most do not estimate the locations of all eleven characteristic points as described in this effort. Furthermore, interpretive algorithms implemented by commercial diagnostic systems typically operate on multiple leads (usually up to 12, although some systems use 13). This allows multiple "views" of the ECG signal and can aid in reading through noise before performing analysis. In short, performing a consistent, objective comparison is non-trivial as the fundamental approaches and assumptions are very different.

1.2 BACKGROUND INFORMATION

1.2.1 Periodic and semiperiodic signals

Periodic signals

Before introducing semiperiodic signals, a formal definition of periodic signals is required. From [15], a signal $f(t)$ is periodic with period T_0 if for any integer m

$$f(t \pm mT_0) = f(t), \text{ for } -\infty < t < \infty \tag{1.1}$$

That is, function values separated by any integer multiple of the period are identical for the signal. Periodic signals can be expressed in the frequency domain as an infinite summation of sinusoids using the Fourier series expansion. The Fourier series expansion, in exponential form, is defined as

$$f(t) = \sum_{n=-\infty}^{\infty} c(nf_0)e^{j2\pi nf_0 t} \tag{1.2}$$

Where $f_0 = 1/T_0$ and the nth Fourier coefficient, $c(nf_0)$ is given by

$$c(nf_0) = \frac{1}{T_0}\int_{T_0} f(t)e^{-j2\pi nf_0 t}dt \tag{1.3}$$

From the definition of the Fourier expansion in Equation (1.2), it is seen that a periodic signal contains only frequency components that are exact integer multiples of the fundamental frequency, nf_0. These components are called the harmonics of

the signal, and each has an amplitude and phase determined by the corresponding Fourier coefficient $c(nf_0)$.

Letting $c_r(nf_0) = \mathrm{Re}[c(nf_0)]$ and $c_i(nf_0) = \mathrm{Im}[c(nf_0)]$ represent the real and imaginary components of the Fourier coefficient, respectively, the amplitude and phase of the nth harmonic are given by

$$|c(nf_0)| = \sqrt{c_r(nf_0)^2 + c_i(nf_0)^2} \tag{1.4}$$

$$\angle c(nf_0) = \arctan \frac{c_i(nf_0)}{c_r(nf_0)} \tag{1.5}$$

Note that for truly periodic signals, none of the these values is time varying; the frequency, amplitude, and phase are constant for the fundamental and all harmonics.

To illustrate fundamental concepts regarding characteristic points of periodic and semiperiodic signals, a simplified example signal can be useful. One example that is analytically simple, yet powerful enough to roughly approximate certain real-world signals like the ECG, is a rectangular pulse train. Analysis of this waveform will demonstrate tradeoffs between frequency- and time-domain representations of semiperiodic signals.

First define a single rectangular pulse of width τ over the interval of length T_0. With $-T_0/2 \leq t \leq T_0/2$, the pulse is represented by $f(t)$ as

$$f(t) = \begin{cases} A, & |t| < \frac{\tau}{2} \\ 0, & |t| > \frac{\tau}{2} \end{cases} \tag{1.6}$$

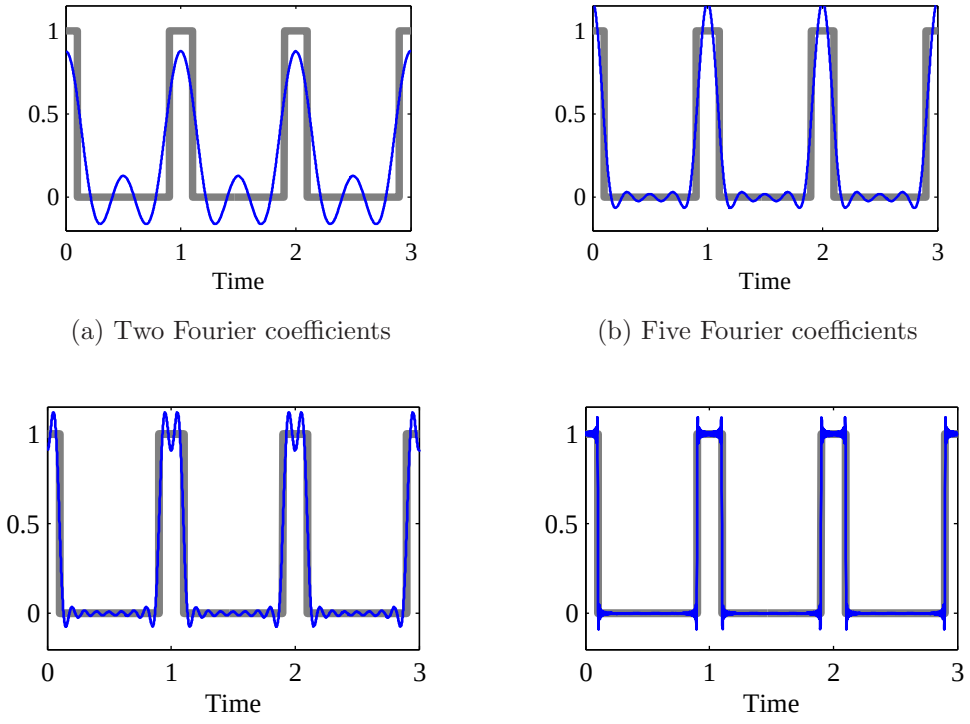
This pulse has "on" value of A for a total duration of τ in each period of duration T_0. Infinitely replicating this pulse every T_0 seconds results in a pulse train with duty cycle of τ/T_0, *i.e.*, the the ratio of time it is "on" to the total period. It can be shown (see, for example [15]) that the Fourier coefficients of this pulse train are given by

$$c(nf_0) = Af_0\tau \,\mathrm{sinc}(nf_0\tau) \tag{1.7}$$

Where

$$\text{sinc}(\lambda) \triangleq \frac{\sin(\pi\lambda)}{\pi\lambda} \tag{1.8}$$

Figure 1.1 shows the Fourier series representation of a rectangular pulse train with $\tau = T_0/5$ (20% duty cycle) for several cases that approximate the signal with an increasing number of harmonics. As the order of the Fourier expansion is increased the resulting synthesized signal more closely resembles the original pulse train. The ability of a frequency-based approach is limited in representing highly time-localized changes such as rapid slopes and sharp peaks. In these cases a large number of coefficients are required for accurate representation.



(a) Two Fourier coefficients

(b) Five Fourier coefficients

(c) Ten Fourier coefficients

(d) 100 Fourier coefficients

Figure 1.1: A comparison of Fourier series syntheses of a rectangular pulse train with a 20% duty cycle. Each subfigure shows the pulse train reconstructed with a different number of coefficients, illustrating that abrupt edges require a greater number of coefficients for accurate reproduction.

For certain semiperiodic signals such as the ECG, accurately representing morphological features of the signal with this approach rapidly becomes intractable due to the number of coefficients required. This is especially true since the "duty cycle" of an ECG signal is even less than this example, typically around 10% for an adult. Additionally, for most domains the Fourier coefficients $c(nf_0)$ have no relevance to experts like clinicians analyzing ECG signals; they will have to be translated into domain-specific annotations to be meaningful to users.

Semiperiodic signals

Qualitatively, semiperiodic signals are similar to periodic signals in that they have a repetitive structure. However, for a semiperiodic signal the fundamental frequency and the amplitudes, frequencies, and phases of the harmonics are all time varying. By definition these variations are bandlimited and change at a rate much lower than the fundamental frequency resulting in morphologies that are similar to each other on a cycle-to-cycle basis, but not identical as would be the case for a truly periodic signal. These differences are manifested as changes in the timing and amplitude of the signal's characteristic points.

There are numerous examples of semiperiodic signals, many of which occur in biological systems. In addition to the electrocardiogram signal which is used in this dissertation, other semiperiodic signals in this domain include hemodynamic waveforms such as those created by blood pressure or pulse sensors; the photoplethysmogram, which can be used to determine blood oxygenation levels by measuring the blood's absorption of different frequencies of light; the electroglottogram, which measures the degree of contact in vocal folds during voiced speech using changes in impedance; respiration signals derived from resistive bands, impedance measurements, or other means; and the neuronal spikes which reflect action potentials of neurons in the brain.

Semiperiodic signals also occur in biomechanics, where electromyographic or

on-body inertial sensors are often used to monitor activity. Their use in quantifying gait, for example, results in semiperiodic signals reflecting step-to-step interval and points of interest corresponding to heel strike, terminal stance, toe off, and foot swing during each step. In other domains, sources of semiperiodic signals include tides, vibration of rotating machines such as motors, and photometric monitoring of stars.

Since semiperiodic signals are not truly periodic, the term *rhythmicity* is sometimes used in referring to the cyclical pattern of the signal. The time-varying components that are summed to synthesize a semiperiodic signal are called *partials* to differentiate them from the harmonics that compose a purely periodic signal.

In [10], a sequence x_k is semiperiodic with rhythm r, if for each $\epsilon > 0$ there exists a positive integer n such that

$$|x_k - x_{k+rn}| < \epsilon, \quad \forall r, k \tag{1.9}$$

Following this example, Equation (1.1) is modified to reflect the changes required for a semiperiodic signal

$$|f(t) - f(t \pm mT_1(t))| < \epsilon, \quad \forall m \tag{1.10}$$

In Equation (1.10), $T_1(t)$ is the time-varying fundamental period of the semiperiodic signal. The effect of variations in the partials' amplitudes, frequencies, and phase values is captured by the inequality; specifically, these variations will result in small changes to the signal's amplitude from cycle to cycle.

To illustrate these concepts, Figure 1.2 shows the periodic rectangular pulse train with 20% duty cycle which was introduced in Figure 1.1. Superimposed on the pulse train are two synthesized waveforms. The first, in red, was generated using the first three Fourier coefficients as defined by Equation (1.7). In this case none of the parameters are time-varying: the fundamental frequency is fixed, harmonic frequencies are exact integer multiples of the fundamental frequency and

the amplitudes and phase values of the harmonics are static. The result is a truly periodic signal.

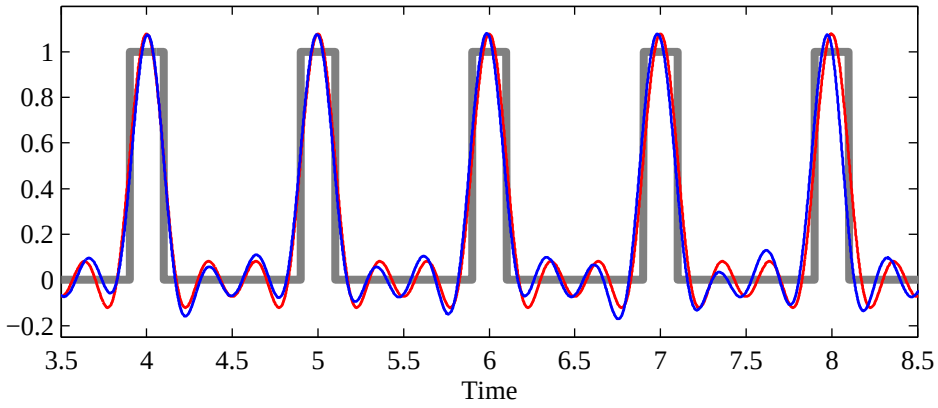


Figure 1.2: Rectangular pulse train with periodic and simulated semiperiodic representations. The red synthesized waveform is purely periodic, and the blue synthesized waveform simulates semiperiodicity with a slowly varying random component in the Fourier coefficients used in its synthesis. Both waveforms are represented using the first three Fourier coefficients.

The second waveform, in blue, was also generated with three components. However, this case uses partials (not harmonics) by introducing a slowly varying random component to the fundamental frequency and the frequencies, amplitudes, and phase values of the Fourier coefficients derived above. The resulting waveform demonstrates key characteristics of a semiperiodic signal. Note how the main peak of the semiperiodic signal shifts off of the red, corresponding to the change in its fundamental frequency. And changes in the amplitudes, frequencies, and phase values of the partials result in changes in the amplitudes and timing of the lobes around each main peak. This sets the stage for the description of a signal's characteristic points.

1.2.2 Characteristic points

Informally, the characteristic points of a semiperiodic signal are points of interest which garner attention due to their prominence. In the literature they are commonly known as fiducial points, and to a lesser extent, as signal singularities. Regardless of the term, it is well understood that these locations "...often carry the most important information" in such waveforms [48].

Characteristic points are typically a local minimum, a local maximum, or a point of high curvature. Their prominence reflects a change that is of interest in the underlying system. Since the characteristic point represents a change of interest, it is reasonable, then, that it can be described with respect to the signal's derivatives. In this work characteristic points are modeled as a subset of the signal's critical points, *i.e.*, locations at which the slope is zero or undefined. All characteristic points are critical points, but the converse is not true: some critical points may not be of interest for a given domain, so they would not be classified as characteristic points.

The first derivative test, described in any introductory calculus text such as [80], is used to analytically determine local minima and maxima of a waveform. First, define the critical points of a signal $f(t)$ as those at which its derivative $f'(t) = 0$ or does not exist, *i.e.*, a discontinuity. Then for all critical points c, if there exists an interval $(c - \delta, c + \delta)$ such that

$$f'(t) < 0, t \in (c - \delta, c) \text{ and} \tag{1.11}$$

$$f'(t) > 0, t \in (c, c + \delta) \tag{1.12}$$

then $f(c)$ is a local minimum. Reversing the inequalities above will determine if the critical point c is a local maximum.

Points of inflection are points at which a waveform exhibits a change of concavity as evidenced by a change in the sign of its second derivative $f''(t)$. The sense of concavity is defined to be "concave up" when $f''(t) > 0$ and "concave down" when

$f''(t) < 0$. Again following [80], a point c is a point of inflection for a function $f(t)$ if its graph is concave in one sense for $t \in (c - \delta, c)$ and concave in the opposite sense for $t \in (c, c + \delta)$. At the point of inflection, the function's second derivative $f''(t) = 0$ or is undefined.

For a purely periodic signal, the locations of local extrema and points of inflection can be determined analytically by solving the equations above; there is no ambiguity in their placement. This can be seen in the red waveform in Figure 1.2, in which the locations of all local extrema and points of inflection are static with respect to each other.

For a semiperiodic signal with a stochastic element in its time-varying partials, the critical points can be determined numerically by calculating estimates of the signal's first and second derivatives from its samples, then using these values with the equations above to locate local extrema and points of inflection. In the noise-free case of the blue waveform in Figure 1.2, the location of critical points is slightly different for each cycle, but given the signal samples, the locations of the critical points can be determined numerically.

The most interesting case — leading to one of the contributions of this dissertation, occurs for a semiperiodic signal with stochastically-varying partials as above, but with the challenge of additive noise. In this case neither of the approaches above will work. Due to the stochastic element of the partials there is no analytic solution describing the critical points. And they cannot be determined numerically as the additive noise will cause many locations not at critical points where the first or second derivatives are zero. As a result, in this case the characteristic points must be determined individually for each cycle.

Once obtained, the characteristic points of semiperiodic signals can be used to delineate the signal and identify important points of interest in the underlying system. The points of interest can be used to represent the signal in an efficient

manner and facilitate analysis, signal classification, or data compression. In addition, they may have domain relevance in and of themselves. For example, for ECG signals, characteristic points reflect changes in electrophysiological activity in the myocardium and are subjects of extensive clinical training in electrophysiology and cardiology. Both absolute and relative changes to these points, as well as metrics derived from their locations, can aid clinicians in analyzing and diagnosing the condition of the heart.

1.2.3 Splines

Splines were originally long, thin, flexible strips of wood or metal that were bent around fixed points in order to create smooth curves for engineering purposes like shipbuilding [69]. They have since been adapted to mathematical curve fitting and are used to approximate signals in a piecewise-continuous manner. They have numerous applications ranging from data smoothing, regression, and curve fitting to computer graphics and geometric modeling.

Splines can provide compact and efficient representations of certain classes of signals. In this dissertation they are used as the basis for algorithms that can optimally represent semiperiodic signals and optimally determine their characteristic points. The spline's flexibility in approximating curves with different degrees of smoothness at different locations [21] is ideal for representing semiperiodic signals as described above. This is especially true when the cyclical components of the signal's partials result in many critical points localized in a short time period during each cycle.

This type of signal could be modeled as a "low duty cycle" semiperiodic signal which — like the ECG — exhibits short durations of low smoothness between long durations of high smoothness. Attempting to fit this pattern with a polynomial will require one of very high order, making it susceptible to severe oscillations of the fitted curve [47] and potential overfitting.

Splines address this problem by providing an alternative that allows representation of a signal piecewise with polynomials of lower degree, including degree one, *i.e.*, linear segments. The segments are bounded on each side by points called knots, selected along the abscissa of the function to be estimated. The knots are used as the endpoints for an interpolant function. The knots themselves may be points of discontinuity or, depending on the implementation, additional constraints such as twice differentiability (for cubic splines) may be required of the interpolation.

In the development of this work it was found that polynomial splines could more accurately represent ECG waveforms, but that linear splines are more effective in determining locations of their characteristic points. This is discussed further in Chapters 3 and 4.

1.2.4 Electrocardiogram

This section presents a very high-level background on the electrocardiogram signal, sufficient to inform its use in developing the spline framework and Bayesian characteristic point optimization algorithms.

A heart beat involves a complex electrochemical process during which an ordered movement of ions into and out of myocardial cells causes them to depolarize and repolarize, resulting in the coordinated contraction of the heart muscle to pump blood.

Electrical depolarization refers to the movement of ions in myocardial cells causing them to contract. The electric field associated with depolarization forms an activation front that is rapidly spread to neighboring cells, thus quickly propagating depolarization across the entire heart muscle. Repolarization consists of the ionic transfers that prepare the cell for its next contraction.

The movement of charged particles, whether for depolarization or repolarization, creates an electrical current between myocardial cells. The current is localized

to myocardial cells, but the corresponding electrical potential induces an electrical field that extends to the body surface. Potential differences of the electrical field can be measured on the skin or sampled and recorded as a function of time with an electrocardiograph. Such a representation is known as the surface electrocardiogram. An ECG signal captured in this manner is a projection of the summation of the electrical activity of all myocardial cells at a given instant onto the surface of the body.

ECG measurement requires at least two electrodes to capture the potential difference caused by the electric field on the body. The location of these electrodes specifies a lead, and the physical placement of a lead on the body determines the "view" of the heart that it provides.

Figure 1.3 illustrates a human heart and provides a high-level view of its electrical system [104]. A normal heart beat starts with the depolarization of the sino-atrial node located at the posterior region of the top of the right atrium. The sino-atrial node is the heart's natural pacemaker; its specialized cells spontaneously depolarize at a rate determined by nervous system activity and hormones in the blood. The propagation of the activation front from the sino-atrial node into the atria (the top chambers of the heart) causes their depolarization resulting in contraction, pumping de-oxygenated blood into the right ventricle and oxygenated blood into the left ventricle.

The activation front continues moving down the heart muscle, where it encounters the atrio-ventricular node near the top of the right ventricle. This node relays electrical activity into the ventricles via the bundle of His and the Purkinje fibers, causing them to depolarize. The resulting contraction of the right ventricle pumps deoxygenated blood to the lungs, and that of the left ventricle pumps oxygenated blood to the body via the aorta.

Figure 1.4 shows an example of a cardiac cycle, comprising two beats. The first beat is annotated with labels indicating the most common constituent waves of the

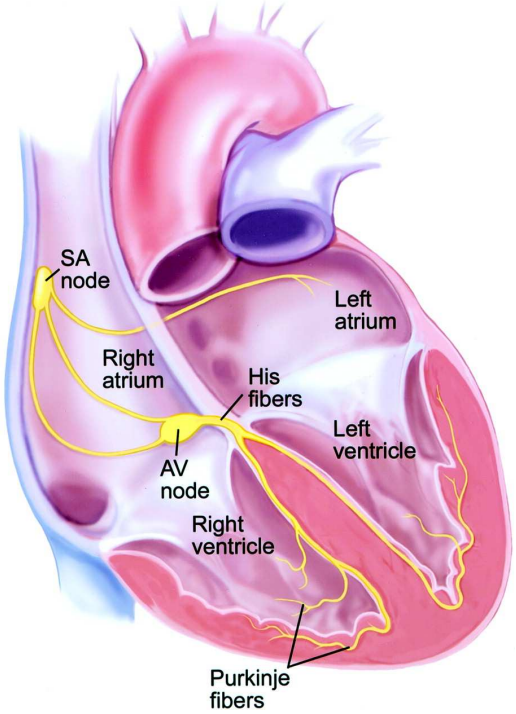


Figure 1.3: The heart and its electrical conduction system. A normally-conducted heart beat starts with an impulse in the SA node, causing the atria to contract and fill the ventricles with blood. The electrical activation then moves to the ventricles through the AV node, causing them to contract and pump oxygenated blood to the body and de-oxygenated blood to the lungs. Figure from P.J. Wang and N.A.M. Estes III, Supraventricular Tachycardia, *Circulation*, 106(25) pages e206–e208, 2002.

ECG signal. The second beat is used to illustrate intra- and inter-beat measures commonly used to analyze and diagnose ECG signals. All of these measures are defined by characteristic points of the signal, *i.e.*, the onsets, peaks, and offsets of component waves.

In Figure 1.4, the P wave corresponds to depolarization of the atria. The large size of the ventricles, combined with their rapid depolarization creates what is generally the most prominent feature of the ECG signal: the QRS complex,

which is composed of the Q, R, S, and sometimes R′ waves as indicated in the figure. The T wave reflects changes caused by repolarization of the ventricles. The repolarization of the atria cannot be seen on the ECG as their amplitude is too small to be detected by surface electrodes [9].

The isoelectric level is the electrical baseline of the heart, and occurs when there are no changes due to depolarization or repolarization. It is used for certain relative measurements such as amplitudes of component waves and deviation of the ST segment.

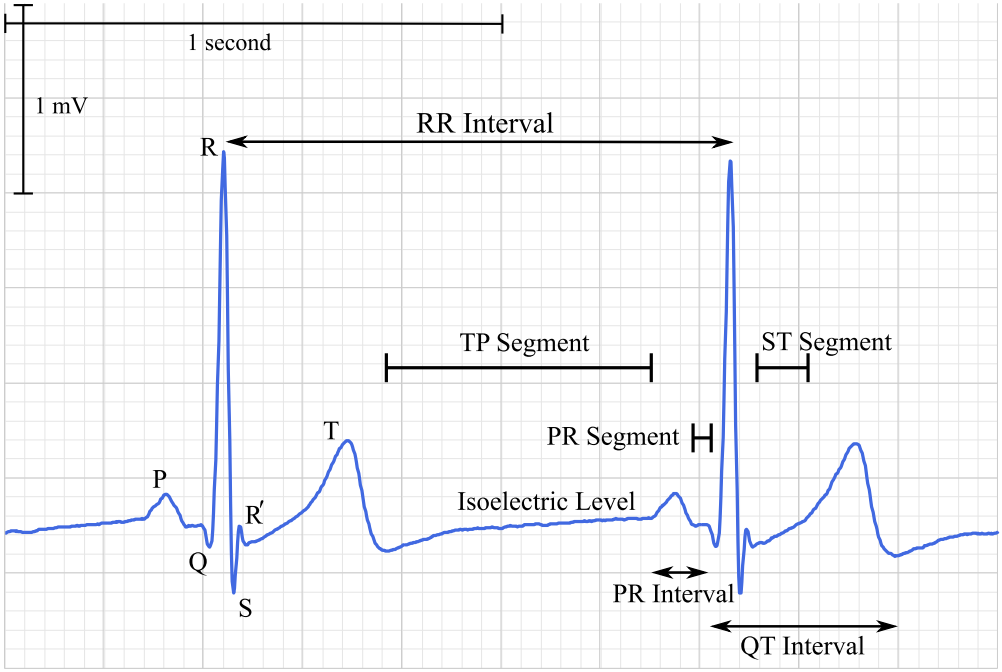


Figure 1.4: An annotated ECG illustrating a cardiac cycle. The first beat is annotated with names and locations of component waves. Characteristic points are generally the onsets, peaks, and offsets of these waves. Other measures, such as the intervals and segments shown above, are used by clinicians to quantify electrophysiological activity of the heart and can indicate disease, stress, and other conditions of interest.

The derived measures called *intervals* refer to time differences between various characteristic points in the signal, while those called *segments* refer to the value of

the signal itself across the specified range.

The time between successive beats is often called the RR interval and designates the time between two depolarizations of the ventricles. The RR interval specifies the fundamental period (rhythm) of the ECG and determines the subject's instantaneous heart rate. For virtually all subjects (the exceptions being a few extremely pathological conditions), the RR interval varies on a beat-to-beat basis. This change by itself is sufficient to make the ECG signal semiperiodic; it is exacerbated by periodic and stochastic changes to partials resulting in measurable changes to all characteristic points from cycle to cycle.

The semiperiodic nature of the ECG waveform manifests not only as beat-to-beat changes in RR interval, but also as changes in amplitude and timing of all constituent characteristic points including the starting and ending points of the P, QRS, and T waves, and the location and magnitude of their peaks. So the morphology of the ECG complex repeats regularly — but with periodic and stochastic variations — and as such cannot be classified as a truly periodic signal.

Locating and quantifying characteristic points in the ECG waveform and obtaining metrics derived from them provides a non-invasive view of cardiac function for analytic, diagnostic, or therapeutic purposes. There are many such metrics; a few commonly-used ones include the QT interval, deviation of the ST segment from the isoelectric level, the PR interval, and variability in RR interval over time. However, the semiperiodic nature of the ECG signal combined with noise from muscle artifact, electrode motion, and other sources makes accurate location of these points challenging. It is for this reason that ECG signal analysis was selected as the first implementation of the spline framework for parametric representation of semiperiodic signals, and for optimal characteristic point determination.

1.3 CONTRIBUTIONS

This section summarizes the main contributions of the dissertation: a novel framework for spline-based representation of semiperiodic signals, an implementation of that framework for optimally locating characteristic points of a semiperiodic signal using Bayesian estimation, the probability density estimate used to incorporate *a priori* information regarding characteristic points for the optimization, and a method for objectively obtaining this prior knowledge.

1.3.1 Spline framework for representing semiperiodic signals

The first contribution of this dissertation is the development of a novel, generic, spline-based framework for parametrically representing semiperiodic signals. As an application, an instance of this framework is developed to address the ECG delineation problem, also known as ECG segmentation. The result of delineation on a cardiac cycle is the representation of the signal in a compact, efficient manner comprising a number of knots whose positions are optimally determined by the algorithm. The output of the algorithm can be used to classify heart beats based on their morphology, compress the ECG signal, or with additional processing, to identify and estimate characteristic point locations.

The framework is shown in Figure 1.5 and comprises a method for knot initialization, a choice of spline interpolant, an objective function (error criterion or figure of merit), and a knot location optimization algorithm. Upon completion of the knot adjustment cycle, indicated by the shaded region in the figure, the objective function is satisfied and knots are placed at their optimal locations.

Choice of the components implemented by the framework is driven by the application. For example, high-fidelity signal representation requires different choices than accurate estimation of characteristic point locations.

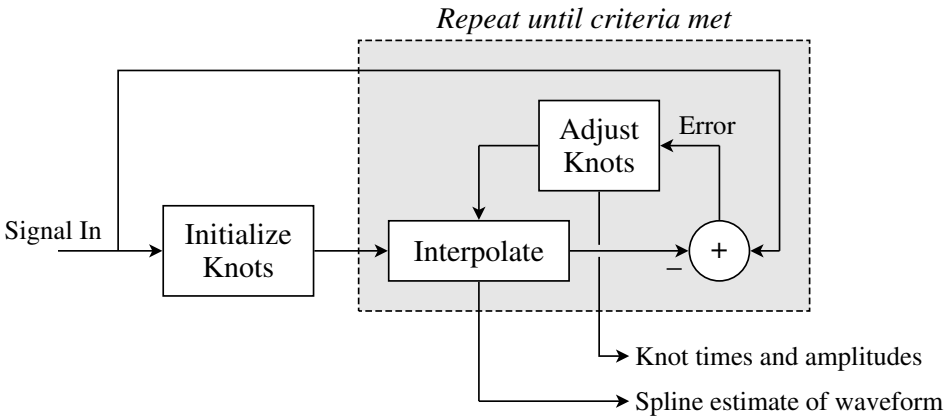


Figure 1.5: The spline framework for representing semiperiodic signals.

For the initial application demonstrating ECG waveform delineation, the framework is used to segment the signal into arbitrary regions for an accurate representation of the signal. Examining the results establishes tradeoffs in the approach that balance computational complexity, reconstructed signal fidelity, and location of resultant knots with respect to true characteristic points.

The spline representation is an efficient one, requiring relatively few knots to represent each cycle of the semiperiodic signal irrespective of the sampling rate. Since the algorithm is run on each cycle of the semiperiodic signal, it is inherently capable of capturing and incorporating small morphological changes that occur due to stochastic or periodic variations in the signal's partials.

An additional benefit with this approach is that the location of the knots can potentially be meaningful, in and of themselves, to domain experts who may not understand Fourier coefficients or other mathematical representations of the signal.

1.3.2 Optimal characteristic point estimation

The second contribution of this dissertation is an implementation of the spline framework described above to optimally locate a predefined set of characteristic points of a semiperiodic signal. The ECG signal is again used as an application

domain for this contribution — in this case by defining a number of clinically-important characteristic points and implementing a Bayesian figure of merit for use by the optimization algorithm to estimate their locations on a beat-by-beat basis.

The algorithm described in Section 1.3.1 initializes knot locations using a recursive partitioning of the waveform and optimizes them using a genetic algorithm to minimize the root mean square error (RMSE) between the observed signal and its interpolated approximation.

Recursive partitioning results in a variable number of knots, which for the ECG example is dependent on QRS complex morphology, making it difficult to maintain correspondence between the knots and specific characteristic points. This limitation necessitates post-processing to map knots back to characteristic points after their locations have been optimized.

Using RMSE as the error criterion optimizes the difference between actual and synthesized signals and results in knot locations providing the highest fidelity signal estimate with the chosen interpolant, not necessarily those providing the most accurate estimates of the signal's characteristic point locations.

Furthermore, the previous approach neglects additional information present in the relative times and amplitudes of knots describing the waveform (*i.e.*, their prior probabilities).

The second contribution uses the same framework, but with different choices for the constituent algorithms, to address these shortcomings. It uses a predetermined number of knots to represent the set of characteristic points whose locations are to be estimated. For each of these knots prior probability distributions are obtained using manual annotation of a training database by human reviewers.

Knot locations are initialized to the means of the priors, and a linear interpolant is used due to its ability to place knots at locations of the desired characteristic points while accurately representing the signal.

In this implementation, a Bayesian figure of merit is used as the objective function. This figure of merit is the *a posteriori* probability (or simply, posterior) of the knot locations given the observed signal, and is calculated using Bayes' Theorem. The benefit of this figure of merit is its ability to fuse information provided by the waveform with prior knowledge derived from the training set.

Finally, to determine the optimal knot locations, a constrained coordinate optimization algorithm is used to find the best value of the posterior. This algorithm is gradient free, and searches for the best location of a knot at every location between its immediate neighbors.

1.3.3 Prior probability density for optimization

To incorporate *a priori* information regarding characteristic points into the figure of merit (and therefore into the optimization), the algorithm uses an estimate of the joint probability density of the time of each knot and a measure of its curvature. As with those used for knot initialization, this prior density estimate is also derived from the reviewers' manual annotations on the training set. In this case, however, additional information is captured by the priors to aid the optimization algorithm.

In the ECG signal, the curvatures of sharp peaks, rounded peaks, and waveform onsets and offsets are markedly different — respectively possessing high, moderate, and mild curvatures. Incorporating a measure of curvature into the priors gives the figure of merit greater ability to assess the goodness of each knot's proposed location during the optimization process. For example, using knowledge of a knot's curvature makes it less likely for the optimization algorithm to place a knot corresponding to a waveform onset at the waveform's peak.

The curvature of each knot is defined by the knot's time and amplitude, and the times and amplitudes of its immediate neighbors to the left and right. Although priors could be constructed from the times and amplitudes of all three points, the corresponding increase in dimensionality of the joint probability density would be

problematic and require exponentially more training data.

Using a curvature metric calculated from relative times and amplitudes of a three-tuple of knots condenses the six-dimensional space to only two dimensions, comprising the knot's time and its curvature. This requires far less training data and makes the approach viable.

1.3.4 Estimation of priors for Bayesian optimization

This regards the means of determining prior distributions required for Bayesian optimization of characteristic point locations. It is a necessary element for the contributions described in Sections 1.3.2 and 1.3.3 above.

One historical criticism of Bayesian methods is their reliance on prior probabilities which are often unknown. Practitioners have used a number of methods to obtain priors, ranging from use of a uniform distribution to subjective approaches which allow human experience and judgment to influence their value. The contribution from this effort is the means to objectively obtain priors from a training data set.

From this data set, a number of events are selected and manually annotated for locations of interest by multiple reviewers. The statistics of annotation locations, and measures derived from these locations, define the priors used by the Bayesian optimization algorithm.

For the ECG application described in this dissertation, several hundred beats were randomly selected from 40 subjects and predefined characteristic point locations were annotated manually by two reviewers. Statistics of each point's location and a measure of curvature (called the knot's relevance value) were then computed to estimate the prior probabilities for each point.

The flexibility of the approach extends to the scope of the priors. They can be calculated across an entire subject population, or "tuned" to an individual, focusing the priors on only one subject's data. This is especially useful for biological

systems whose characteristics can be highly dependent on the individual, and can increase estimation accuracy by eliminating inter-subject variability.

1.4 DISSERTATION OVERVIEW

The remainder of this dissertation includes the literature review and chapters providing motivation for, and descriptions of, the basic spline framework and its implementation for optimal estimation of characteristic point locations.

Chapter 2 is the literature review. It first describes a taxonomy used to frame the review process and defines the focus, goals, perspective, coverage, organization, and audience for the review. The structure provided by this taxonomy helps to scope the review and to present the results in a cohesive manner. The review itself is organized across the conceptual classes discovered in the process of applying the taxonomy, comprising time domain algorithms, transform-based algorithms, pattern recognition methods, model-based algorithms, and spline-based algorithms.

Chapter 3 describes the generic spline framework for parametrically representing semiperiodic signals. An instance of the framework is developed to illustrate its capability for segmentation of the ECG signal, and to help understand tradeoffs in signal fidelity and knot location for different interpolants. Results of the segmentation are presented for a small set of signals exhibiting challenging morphologies.

Chapter 4 presents the main contribution of this dissertation, in which an instance of the spline framework is developed for optimally estimating the characteristic points of a semiperiodic signal using a Bayesian approach. The implementation is tested on a 200 signals from 20 subjects (not in the training set) and the results are compared against manual annotations from human reviewers and an accepted standard.

Chapter 5 provides a summary and conclusion for the dissertation and includes other domains which may benefit from this approach, as well as areas for future research.

Appendix A shows the complete set of priors used in the optimization process of Chapter 4. These are estimates of the joint probability density of the time of each knot and its relevance, which is a measure of the signal's curvature at that point.

Appendix B describes the ECG data that were used in the effort of Chapter 4 for optimal estimation of characteristic points.

Chapter 2

LITERATURE REVIEW

2.1 APPROACH

In order to provide focus and structure to the review process, this section adopts Cooper's taxonomy for literature reviews as described in [18,19,76]. This taxonomy was originally intended for meta-analytic studies in education and psychology, but it provides a good framework for defining and structuring any literature review.

In his taxonomy, Cooper classifies literature reviews across six characteristics: focus, goal, perspective, coverage, organization and audience. The sections below provide a very brief description of these characteristics as well as choices for each.

2.1.1 Focus

The taxonomy defines several potential focus areas, including research outcomes, methods, and practices or applications. For this literature review, the focus is on theories in order to establish the current state-of-the-art in waveform delineation and annotation algorithms and how they relate to each other. This helps to identify higher-level themes in the literature and set the stage for the dissertation's contributions which advance new algorithms in this domain. The other focus areas are intended for social science research and are not directly applicable to this domain.

2.1.2 Goal

Goals for the literature review in the taxonomy include integration, criticism, and identification of central issues. This literature review has two goals: to integrate and synthesize previous research and to identify central issues. Focusing on these goals exposes existing themes in the research and provides rationale motivating this approach.

2.1.3 Perspective

This literature review does not advance a particular perspective; rather it provides a neutral, objective representation of the literature. As in any engineering effort all approaches have benefits and shortcomings and it is important to understand both aspects — not only to inform use of various algorithms but also to help guide future efforts.

2.1.4 Audience

This characteristic primarily manifests itself in the writing style of the review, specifically use of jargon and detail as balanced against focus on implications of the work [18]. The taxonomy identifies audiences of specialized scholars, generalized scholars, practitioners or policy makers, or the general public.

For this literature review (and in fact, the entire dissertation) the audience of this dissertation is assumed to be specialized scholars with deep domain experience in signal processing and pattern recognition, but not necessarily in ECG analysis.

2.1.5 Coverage

This characteristic regards the extent of the literature review. The body of work in waveform segmentation and annotation is very large, precluding any exhaustive review that covers all the literature. Other options identified by the taxonomy

include central or pivotal, and representative.

This review does not provide a review of pivotal literature, as this requires establishing proof that the only publications reviewed are pivotal — and, more problematically, that no literature fitting this criterion are excluded. Instead it provides reviews of a representative sample of algorithms which includes most interesting and important efforts in this space as identified by the search strategy described below. This approach better informs reviewers of efforts historically applied to solve the segmentation problem.

The literature search methodology to obtain representative coverage comprised several stages. The initial stage was "search driven" and intended to identify a starting set of literature in waveform segmentation and annotation of characteristic points. Search terms were chosen to include segmentation, delineation, and annotation of any type of semiperiodic waveform, but virtually all results were related to biomedical signals. And of these results, most pertained to the ECG.

Detailed review of the initial literature identified new search terms and cited publications of interest, which were the subject of subsequent reviews. This process was repeated several times, until the rate of discovery of new publications was significantly diminished and all citations of interest had been previously discovered. Although not exhaustive, this approach was very fruitful and provided significant coverage of algorithm development in this area dating back several decades.

2.1.6 Organization

This characteristic describes how the literature review is arranged: topics can be presented historically, methodologically, or conceptually. A methodological organization groups efforts employing the same research methodologies (usually in the natural sciences, social sciences, or education) and does not lend itself to a literature review focused on theoretical elements like algorithms.

A purely historical organization presents all of the literature in a strictly chronological order. This may be suitable for reviews emphasizing the progression of research efforts or change in practices over time [76], but for this effort a historical organization would result in a highly "fragmented" view of the literature as there is simultaneous effort on numerous approaches and virtually all approaches are revisited over time. A strictly historical organization would result in a great deal of interleaving of concepts in the literature review and make it difficult to integrate and synthesize findings per the goal defined in Section 2.1.2.

To address these concerns, this review is primarily organized in a conceptual manner, grouping works that share abstract ideas [18] — in this case an empirically-determined classification of algorithm types. Within each conceptual class items are organized historically. This approach illustrates the evolution of algorithms over time in the different classes, aiding the goal of integration and synthesis. The following section outlines the conceptual classes used in organizing the literature review, and subsequent sections provide the review itself.

2.2 CONCEPTUAL CLASSES

The problem of automatic segmentation and annotation of semiperiodic signals — specifically those from the ECG — has a very long history. Researchers have applied a vast array of signal processing and pattern recognition techniques to detect QRS complexes and estimate the precise locations of onsets, offsets, and peaks of their component waves reliably and accurately under challenging conditions.

Numerous solutions are described in the literature, and most employ a combination of algorithms to achieve best results under different constraints and assumptions. For example, reducing algorithm complexity at the cost of degradation to accuracy, or improving noise immunity at the cost of increased computational burden. Careful examination of the literature did not identify any broadly-accepted taxonomy of algorithms. However the conceptual classes derived from this review

are consistent with, and expand, those described in other works such as [89], [17], [48], [96], [51], [31], [78], [2].

Since each approach is often comprised of multiple algorithms, it can be difficult to establish its membership unambiguously to a single conceptual class. In these cases the publication is grouped into a conceptual class reflecting the most significant innovation it describes.

This section provides a brief description of the conceptual classes identified by the effort described above; subsequent sections are devoted to literature reviews grouped by these conceptual classes.

Time-domain analysis. Literature in this conceptual class describe algorithms operating in the time domain. Many examine the structural properties of the signal and use empirically-derived thresholds and search windows to determine characteristic point locations.

Transform-based algorithms. Numerous transformations have been applied to semiperiodic signals to help enhance the signal, reduce noise, and improve discrimination of characteristic point locations.

Pattern recognition methods. Publications in this class comprise either "soft compute" or classic pattern recognition methods. Soft compute is defined as a class of algorithms that exploit tolerance for imprecision, ambiguity, partial truth, and approximation and include artificial neural networks (ANN), fuzzy logic, evolutionary computing, and probabilistic reasoning.

Model-based algorithms. Broadly defined, models are mathematical representations of real signals that efficiently capture their essential characteristics [53]. Use of models for waveform delineation allows application of a set of powerful algorithmic tools to improve accuracy and noise tolerance.

Spline-based algorithms. Spline-based methods really belong in the time-domain analysis conceptual class — and more specifically, in the segmentation-based algorithms class that will be described in Section 2.3.2. But because there is no literature describing their use for waveform delineation, Section 2.7 reviews efforts applying splines to ECG processing more generally, *i.e.*, for waveform compression, noise elimination, and certain types of analysis. Presenting this class last also helps set the stage for subsequent chapters providing detailed description of the algorithms that are the subjects of this dissertation.

For virtually all methods, the first step comprises preprocessing to reduce noise in the signal. Typically three types of filtering are performed: high-pass for low-frequency noise sources, low-pass for high-frequency noise sources, and notch filtering for band-pass noise.

In the ECG signal higher frequency noise sources include muscle artifact, typically from movement of skeletal muscles close to the surface electrode, electrode motion artifact caused by changes in electric potential due to relative motion of the electrode against the subject's skin (frequently exacerbated by dry electrodes), and in surgical settings, interference from electrocautery knives. Baseline drift is the most common low-frequency noise, typically caused by slow variations in pressure on electrodes. And the most typical bandpass noise source is electrical power line interference.

The reviews preceding Section 2.7 on spline-based methods below focus on algorithmic innovations for waveform delineation; specifically determining onsets, peaks, and offsets of some subset of the P, QRS, and T waves. Preprocessing methods such as the filtering methods described above are extensively covered in the ECG signal processing literature. So, assuming that the signal has been conditioned prior to the delineation algorithm, the literature review does not provide details on preprocessing unless it is a particularly novel approach that directly

impacts the estimation of characteristic point locations.

Section 2.7 covers the use of splines in ECG processing more generally since there is a dearth of literature describing their use in detection or estimation of characteristic points, even though splines were recognized for this application in 1978 by [61].

2.3 TIME-DOMAIN ANALYSIS

Methods in this conceptual class operate on signals in the time domain, and it contains the majority of algorithms. Due to limitations in computational resources, earlier efforts focused on reducing algorithm complexity to make analysis tractable on available devices. However, despite the advance of Moore's Law which has resulted in incredible gains in microprocessor capability, even more recent efforts (such as [78], [79], [81], [63]) strive for low computational overhead to reduce processing time and power consumption for applications in relatively constrained systems like embedded platforms, mobile devices, and wearables.

These methods can have issues with noise immunity despite filtering because there is significant spectral overlap between the QRS complex and noise caused by muscle artifact and electrode motion. Using conventional filtering to completely remove in-band noise can cause significant losses in the signal of interest and impact the measures used to find characteristic points.

In addition, most methods described in this section are not optimal in any sense and rely heavily on empirically-determined thresholds and search windows for each component wave of the ECG signal. The high degree of variability in electrophysiological signals, both inter-subject and intra-subject, can adversely impact these approaches.

2.3.1 Geometrical

A subclass of time domain algorithms described here as geometrical uses slope-based measures of the signal to determine characteristic points. After performing preprocessing as described above to reduce noise, these methods derive their fundamental features from the signal's first and second differences. All of these methods use heuristically-determined thresholds and search windows, looking for minima, maxima, and zero crossings in various metrics to determine characteristic point locations.

In [33] Hsiao-Shu *et al.* propose an algorithm that first identifies significant points of inflection in the signal then thresholds the angles of each point to determine the location of QRS onsets and offsets. Their empirically-determined thresholds indicate an R wave if the inflection angle $\theta \geq 115°$, and a Q or S wave if $\theta \geq 23°$ and the amplitude differences between Q and R, and between R and S waves exceed the average amplitude variation for the signal segment under analysis.

Laguna's slope-based algorithm described in [46] follows noise filtering with a non-linear operation (squaring) to help discriminate between tall T waves and the QRS complex. The signal is then differentiated to accentuate higher frequencies corresponding to critical points. The resultant signal is first used as the basis for a set of rules and thresholds to identify the QRS complex.

Following detection of the QRS complex, further low-pass filtering attenuates the QRS complex to allow more accurate processing of the lower frequency P and T waves in the signal. A different set of thresholds and rules are then employed to detect onsets and offsets of the P and T waves in windows defined relative to the QRS complex. Although the thresholds and search intervals take into account physiologically significant limits, their values were experimentally determined by the researchers.

Mukhopadhyay *et al.* [63] compute the first derivative of the signal after noise reduction filtering then calculate its Hilbert transform. Using empirically-determined

thresholds and search windows, they identify peaks in the transformed signal to identify an R wave region. They then search the original time-domain signal for a slope reversal in that region to find the precise location of the R wave. Q wave (S wave) locations are determined by searching for slope reversals in the original time domain signal to the left (right) of the R wave peak. QRS onset (offset) points are obtained by searching to the left of the Q wave (right of the S wave) for points of zero slope in the differentiated signal. A similar search in a window to the right of the QRS offset is performed to locate the T wave.

The use of the Hilbert transform in this algorithm is only to accentuate areas of higher frequency concentration in the derivative signal due to the bandpass property of discrete time implementations of the Hilbert transform, not its more customary application in signal processing to obtain an analytical representation of the signal.

Mazomenos [55] describes a time-domain morphology and gradient algorithm which uses a combination of extrema detection, slope information, and adaptive thresholding — and estimates the 11 points indicating the onset, peaks, and offsets of the P, Q, R, S, and T waves.

After noise filtering, a "feature signal" is calculated using a linear combination of the first and second derivatives of the input ECG, which they call "gradients". They experimentally determined coefficients for the linear combination to attenuate the P and T waves while enhancing the QRS complex. After calculating the slopes for each local extrema in the entire complex, they designate the extreme point with largest slope value within an adaptively determined window of the feature signal as the R wave. QRS onset and offset locations are determined by searching the feature signal for values smaller than pre-defined thresholds in regions to the left and right of the R wave.

In [103], Wallace *et al.* use a five-point difference equation to implement their

slope calculation, effectively realizing a bandpass filter that provides some measure of noise attenuation. Following this filter they find local maxima to indicate R wave locations, and search in neighborhoods around the R wave for extrema corresponding to Q and S waves.

In the only non-ECG publication reviewed here, Naidu *et al.* [66] apply a time-domain search to detect characteristic points of the impedance cardiogram (ICG). ICG is a non-invasive procedure used to estimate stroke volume and several other measures of cardiac activity by computing the impedance of the thorax using a low-current signal of 20–100 kHz. In their algorithm the R wave of a simultaneously-captured ECG signal is used as a reference point to determine the peak of the ICG cycle, which then establishes search windows for characteristic points. These points are defined simply by finding maxima and minima in the search windows.

Salih *et al.* [81] describe a geometrical approach that uses rising and falling edges of a potential QRS complex to specify the vertices of a triangle composing the entire complex. The base of the triangle is used in a second stage to search for points of inflection comprising the onsets and offsets of constituent waves.

Numerous other researchers use similar approaches with minor adaptations to filtering, search region boundaries, and threshold values. To address baseline wander in the waveform, Gupta *et al.* [31], applies a linearly-interpolated correction term before applying slope thresholds in the regions preceding and following the QRS complex. Other variations include use of the second derivative to further accentuate changes in slope and highlight characteristic points [32], and squaring the second derivative to amplify changes prior to searching [78].

2.3.2 Segmentation

Segmentation-based methods also operate in the time domain but approaches falling in this conceptual class have a primary innovation based on segmenting the signal and using the endpoints of each segment as potential characteristic points.

Spline-based algorithms are properly a subset of this conceptual class (as they segment the signal in the time domain), but in order to provide a more thorough review of splines in ECG signal processing and to set the stage for the remaining chapters of this dissertation, they are discussed separately in Section 2.7.

In [42], Koski proposes a method of approximating a digital signal with a "suitable continuous broken line" described as a series of endpoints drawn from the original signal samples. The manuscript describes experiments to find the best endpoints by comparing the residual error of linear approximations using various segmentation methods and distance metrics.

The authors recognize the high degree of variability of biomedical signals which greatly complicates determining suitable thresholds, thereby necessitating an adaptive approach. To that end the first step of this algorithm calculates the error distribution of a few segments at the beginning of the signal and uses a statistical approach to set an error threshold for segmentation, *i.e.*, they sort the residual values during the initial training period and take the value that lies above $P\%$ (in practice, they use $P = 90$) of collected residuals as the threshold.

For their distance function, they use a modified Euclidean metric that mitigates errors of a simple amplitude distance in areas with very high signal slope (such as the QRS complex). Their new metric uses the conventional Euclidean distance if the point in question is outside of the normals of the two points used for linear interpolation; otherwise it calculates the distance of the normal between the point and interpolated line.

For segmentation, the algorithm starts by selecting a subset of data points from the signal and initializes the first two endpoints with the first and last points of this segment. As long as the sum of distances between all points in the segment and a linear approximation between the endpoints is less than the threshold (calculated using the modified Euclidean distance described above), they add more points to the subset and recalculate the maximum distance.

When the error threshold is exceeded, they introduce a new endpoint at the data point with greatest distance to the linear approximation. Distance metrics are computed between points and all resultant lines, and new endpoint introduction is repeated until the overall error is once again below the threshold.

This approach yields endpoints at all "significant" points of the signal, as determined by the choice of P and the distance metric. Features represented by the endpoints should certainly include the onset, peak, and offset of the QRS complex due to their prominence. But it is possible that lower-amplitude P waves may be missed in the segmentation process, necessitating a lower error threshold and many more endpoints.

In [41], Keogh *et al.* identify and evaluate several methods for segmenting time series data and propose a hybrid approach that outperforms others. Keogh identifies three major approaches to this segmentation problem: sliding window, top down, and bottom up.

Sliding window algorithms are defined as those that anchor the left point of the potential segment at the first data point of a time series then increasingly lengthen the segment to the right until the approximation exceeds an error bound. The class of top down algorithms considers every possible partitioning of the time series segment under consideration and introduces a new endpoint at the best location (*i.e.*, the one with the greatest error). The resultant subsections are then evaluated, and if the stopping criterion is not met the algorithm recursively continues. Conversely, bottom up algorithms start with the finest possible approximation and merge segments until a stopping criterion is met.

On the test data sets reported in [41], top down and bottom up performed with lower error than sliding window. Sliding Window algorithms, however, are the only ones that support on-line data processing; top down and bottom up require all available data prior to processing and must be run in batch mode. To get the benefit of online processing with the accuracy afforded by the other methods,

Keogh proposes a new approach which combines the sliding window and bottom up algorithms.

In this approach, the working buffer is large enough to support approximately five segments. They apply a bottom up algorithm to the buffer and report the leftmost segment. Subsequently they remove the segmented part of the signal and read more data in using a sliding window algorithm. They repeat this process indefinitely, making it an online algorithm. On a data set comprising ECG, financial, manufacturing, and simulated signals, this hybrid approach outperformed the three existing classes of segmentation algorithms as measured by squared residual error.

Cast in this framework, Koski's algorithm [42] is an example of a sliding window followed by bottom up consolidation, using an adaptive threshold based on signal statistics.

Another example of top down segmentation was described in [27] and [28] for ECG waveform compression, and more recently as part of the spline framework in [29]. In both of these applications, the fidelity of the representation is not a primary concern: for the compression algorithm any signal subtleties not captured by the linear segments become part of the residual which is compressed and transmitted along with the signal endpoints identified by partitioning.

And for the demonstrated applications of the spline framework to determine waveform characteristic points, the accurate estimation of the segment endpoints is the most important result, not fidelity of reconstruction. For future applications of the framework where signal fidelity is important, it may be beneficial to use a hybrid segmentation algorithm as well as a non-linear spline interpolant. Details of this approach are discussed in detail in Section 2.7.

Boucheham *et al.* tout the advantages of a similar top down recursive approach in [11], using the characteristic points so identified for R wave detection and making a case for their potential use in detecting other characteristic points.

In [102], Vullings *et al.* use dynamic time warping (DTW) to locate waves of interest in the ECG signal. DTW provides a mathematical framework for comparing and identifying the best match between two patterns exhibiting temporal variation in their features.

Vullings applies DTW to the slopes of a linear segmentation of the ECG signal generated using the hybrid sliding window and bottom up algorithm described by Koski in [42] and summarized above. Following segmentation, the signal is decomposed into three overlapping regions: one preceding the R wave, a narrow region including the QRS complex, and another region following the QRS complex. Comparing the regions from each beat under consideration against a set of reference beats provides the best match and corresponding characteristic point annotations.

Obtaining the reference beat is obviously of great importance. To handle the inevitable and significant variability in physiological signals, the researchers manually annotated several hundred beats to indicate correct locations of eleven key characteristic points. They then used the DTW algorithm to recursively compare each segment of the beat under consideration against corresponding segments of all reference beats and determine the best match using a slope difference metric.

The effect of DTW is application of segment annotations of the best-matching reference to the beat under consideration. This essentially filters out all extraneous endpoints leaving only the desired characteristic points.

2.3.3 Template based

In [108], Zoghlami *et al.* describe a "fitting" approach based on a template derived from the subject's own ECG signal and intended for use in wireless body networks with limited computational resources. During the calibration stage, a representative beat for a subject is manually segmented into regions containing the QRS complex, P, and T waves. They create templates for each region as tables containing the time and amplitude values of the signal in that region. For each template,

they calculate four parameters specifying the height, width, baseline level, and location of the region in the representative beat.

During the learning stage, a heuristic-based approach first determines the location of the R wave, then searches windows around that location to provide initial estimates of the locations of the P and T waves and their amplitudes. Gradient searching in those regions provides estimates of the remaining two parameters for the beat: width and baseline.

For each region (QRS, P, or T) the appropriate template is fitted to the data points using the Levenberg-Marquardt algorithm (LMA) to obtain the optimal set of four parameters describing the beat. LMA provides a numerical solution minimizing the squared error between an empirical function and a parameterized model curve [108].

Finally, they calculate component wave locations and amplitudes using the parameters that optimally represent the beat with respect to the template.

2.3.4 Filter based

This section describes a few time-domain approaches whose primary contribution is the use of specialized filters to delineate the signal.

In [91] Soria *et al.* describe a system based on an adaptive filter to identify the offset of a T wave for a specific subject. Their algorithm does not appear to be truly (continually) adapting; rather it "adapts to the criterion followed by each cardiologist." After having a cardiologist annotate a characteristic point in a series of beats, they determine an adaptation constant as a function of the distance between the characteristic point and the corresponding peak (which is found by a simple amplitude criterion).

The constant is calculated to have the error signal from the adaptive filter to generate a peak at the desired location representing the offset of the T wave. After applying the adaptive filter to the signal, the first peak that occurs in the output

of the adaptive filter following the T wave corresponds to its offset time.

In [94], Sun *et al.* describe an algorithm using morphological filters to address the shortcomings of linear filters — specifically their inability to address the significant spectral overlap between the ECG signal and various noise sources. Morphological filters are nonlinear filters designed to highlight characteristic shapes of a signal under analysis as indicated by a structuring element: a simple constant shape which the filtering operation probes for fits in the signal.

In this effort, they define a multi-scale morphological derivative operator and use it instead of a conventional differentiation to identify critical points corresponding to the signal's characteristic points.

Morphological derivatives are defined in a manner similar to conventional derivatives, except they use the erosion and dilation operators as the basis for differences in the numerators of the derivatives from the left and right, respectively. They further extend these derivatives by introducing a scaling factor in the denominator of the difference between the left and right derivatives.

Using a flat structuring element as the basis for the erosion and dilation operators results in a computationally simple algorithm to calculate the multi-scale morphological derivative. Examining the output of this operation for local extrema identifies onsets, peaks, and offsets of component waves in the ECG signal.

To discriminate between different component waves, they determined threshold values for the local extrema. The thresholds were based on analysis of histograms of data after filtering by the morphological derivative operation.

2.4 TRANSFORM BASED

Although transformation to the frequency domain has been applied to the ECG signal, the majority of literature describes its use in detecting arrhythmias ([4], [13], [62], [26]) or for compressing the signal ([77], [3], [44]).

There appears to be only one effort using the Fourier transform of the signal

for delineation: in [64], Murthy *et al.* describe an approach using filtering in the frequency domain to delineate component waves of the ECG. Their method is predicated on the observation that a component wave in the time-domain ECG signal (P, QRS, or T waves) maps to a group of complex sinusoids in particular bands in the frequency domain.

For delineation of each component wave, the frequency-domain representation is first shifted so a low-pass filter (in frequency) can be applied to eliminate the other components. The result is then shifted back to its original frequency and subject to an inverse DFT. This recovers a time-domain signal comprising only the component wave of interest. This signal is then smoothed and scanned to find the location of the peak. Windows on each side of the peak are then searched to find the onset and offset points using a heuristically-determined amplitude threshold.

To determine the amount of frequency shift required and the cutoff frequency for the low-pass filter, this method requires knowledge of the locations of each component wave in the original time domain signal. Given this requirement, the authors do not make clear why the delineation could not simply be accomplished in the time domain.

2.4.1 Empirical mode decomposition

Another transform-based method uses the empirical mode decomposition (EMD) algorithm. EMD facilitates analysis of non-linear or non-stationary data by first determining a set of intrinsic mode functions (IMF) which are essentially empirically-determined bases for the signal under analysis. It then decomposes the signal into a sum of IMF components, each representing different "oscillatory modes" present in the signal.

In [6], Arafat describes use of the EMD to detect wave boundaries in the ECG signal. They note that characteristic points of the ECG are best represented by the first three IMFs which correspond to the highest frequency oscillatory modes

present in the signal. To detect the R wave, they create a metric by summing the values of these three IMFs, thresholding, and finding the maxima. QRS onset and offset are determined by finding zero-crossing points of the metric on each side of the R wave.

To find P waves and T waves they first isolate a signal segment by locating the peak in a window to the left (or right) of the QRS complex in the original signal. They then generate an effectively low-pass signal on each side of the QRS complex by subtracting sums of the first two or three IMFs from the original time domain signal. The resultant low-pass signal is searched for local minima or maxima to find the onset and offset points.

2.4.2 Wavelet transform

By far the most widely researched transform-based method for signal delineation in the literature uses the wavelet transform. By applying translated and scaled versions of a basis function (the basic, or "mother" wavelet) to a signal, the wavelet transform provides a multi-scale representation that addresses well-known difficulties of more traditional methods in representing signals requiring simultaneous localization in both time and frequency domains. For example, using a Fourier transform with a long time domain signal provides good resolution in frequency but poor resolution in time. Conversely, a signal with short duration provides good time resolution but poor frequency resolution.

The utility of having simultaneously high resolution in both domains is apparent given the definition of characteristic points presented in Section 1.2.2: they are typified by highly time-localized changes in the signal's frequency components; any method providing accuracy in both domains can help identify such points and allow discrimination of various component waves and certain types of noise.

The first significant work in applying wavelets to ECG delineation is that of Li *et al.* described in [48]. In this effort, Li first determines the scales for the

wavelet decomposition by considering the bandwidths of equivalent filters to the wavelet decomposition at several different scales. Mapping these to the frequency spectrum of the QRS complex suggests the use of four scales spanning the range of 4 Hz to 125 Hz.

Following wavelet decomposition of the signal using the scales identified above, they build a very comprehensive set of rules using search windows, numerous threshold values, and comparisons across different scales to find the "modulus maxima" (zero-crossing points in wavelet transforms) corresponding to component wave peaks of interest. Other characteristic points comprise onset and offset points for the component waves; these are determined by examining output of the wavelet transform at a lower scale (higher frequency) by finding the "beginning" of the modulus maximum occurring before the peak, and "ending" of the modulus maximum following the peak.

Di Virgilio [101] describes a similar (and contemporary) approach to Li's, except they use a digital filter bank — not a direct implementation of the wavelet transform — to perform the multiscale decomposition. They also expand the number of scales to cover six bands ranging from 5 Hz to 320 Hz, but only detect the peaks of component waves (P, Q, R, S, and T). The higher-frequency components (Q, R, and S) are detected by thresholding the output of the third component. The lower-frequency components (P and T) are detected using the fifth component after subtracting the previously-identified QRS complex from the signal.

Andreão [5] and Krimi [43] use wavelets to obtain features which are subsequently analyzed with a hidden Markov model (HMM), reviewed in Section 2.6). The powerful multiscale localization capability of the wavelet approach is used to generate better signal segments as features for the HMM. [5] also provides comparative results in P, QRS, and T detection and delineation between several different wavelet functions, concluding the Mexican Hat Wavelet having the best overall performance.

The wavelet transform provides very good time-localization of frequency changes in the signal, but its output is a superset of the characteristic points. This necessitates a potentially large and complex set of heuristically-derived rules and thresholds, or other post-processing to accurately determine the desired locations.

2.5 PATTERN RECOGNITION

This section reviews publications whose primary innovation is in the class of pattern recognition algorithms such as neural networks, syntactic pattern recognition, fuzzy logic, and statistical methods.

In [59], Mehta *et al.* use Fuzzy rules to identify P and T wave peaks detected in the time domain. The algorithm starts by finding a large number candidate peaks in search windows before (for the P wave) and after (for the T wave) the R wave. Candidates are identified by calculating a conservative threshold from peak-to-peak amplitudes observed in the training data, and Fuzzy rules are used to select from the two best candidate peaks in a given search segment (if only one candidate is identified in a segment, it is taken as the real peak).

Their rules are based on three fuzzy membership functions: total energy, fractional of total energy to incremental energy, and average peak-to-peak amplitudes. The membership functions simply calculate the ratio of each of these measure for a given peak to the sum of the measures for both peaks.

After computing individual membership values they are combined by averaging and defuzzified to make the decision: the candidate peak whose combined membership function exceeds 0.5 is chosen as the real peak.

2.5.1 Neural networks

Szildgyi *et al.* [96] compare a neural network approach against one using wavelets to detect QRS complexes, P, and T waves. They use seven scales for their wavelet

approach, spanning bands from approximately 1 Hz to 100 Hz. Unfortunately they do not provide any detail on how they analyze the wavelet decomposition to determine presence of component waves. For their neural network approach, they use a three-layer network with least mean square training. Their neural network implementation achieved slightly better sensitivity in detecting QRS complexes than their wavelet approach.

In [95], Suzuki *et al.* apply an adaptive resonance theory network (ART2) to detect characteristic points. They adopt a geometric approach to derive features from the underlying signal by fitting a triangle to the region between the Q and R waves of the ECG, and another between the R and S waves. They train their network with 100 such patterns for each side, each with a different base length.

During analysis, they feed a 100 ms signal segment on each side of the R wave into the appropriate ART2 network (from the left side for Q wave, from the right side for S wave). The network provides the best matching triangle from its long-term memory. This result is used as an approximate location for the Q or S wave. Their exact locations are determined by comparing the signal slopes against a threshold value in a heuristically-defined search region.

After every detection, the ART2 network is trained with the actual signal segment determined by the analysis above. As a result, the network self-organizes with new data and the template patterns eventually change from right angles to reflect the true ECG signal.

2.5.2 Clustering

Saini *et al.* present a delineation algorithm for multi-lead ECG data using the k-nearest neighbor (kNN) algorithm in [79]. kNN is a statistical pattern recognition algorithm that learns representative exemplars from a training set then provides the best match based on a similarity measure. An incoming feature vector is determined to belong to the category to which the majority of its k nearest neighbors

belong.

In this effort, squared slope values for each of eight ECG leads are used to form a feature vector. The classifier is trained for QRS detection by setting the label to +1 for each eight-dimensional feature vector occurring during the QRS complex in the training data, and to −1 otherwise. During classification any train of +1 outputs whose duration exceeds the average length of +1 outputs is determined be a true QRS complex.

The onset and offset times of the complex are those times at which the classifier output toggles to +1 and −1, respectively. The location and value of the R wave peak is determined by searching the +1 region for its maximum value in the original time series.

The T wave is detected next. First the QRS complexes are removed by replacing them with the baseline. The kNN classifier is trained as described above with +1 labels for each sample interval during the T wave, and −1 labels otherwise. T wave peak value, and onset and offset times are obtained as described for the R wave above.

Finally, the P wave is detected by replacing the T wave regions with baseline so the signal does not include QRS complexes or T waves. Training and detection of P waves are as described for the T wave above.

2.5.3 Syntactic pattern recognition

A number of segmentation-based algorithms employ syntactic pattern recognition principles. Skordalakis provides a good overview of this approach in [89]. Fundamentally, the syntactic approach parses the signal using a grammar that describes patterns of interest in terms of a set of primitives, which can be considered to be a basis set for representing the signal. Due to ease of computation and encoding, the grammars reviewed in [89] all use primitives obtained from linear segmentation of the signal. These primitives are either based on the slope or derivative energy

computed from the linear segments.

Slope can be encoded simply as positive, negative, or zero (subject to margin of ϵ), or it could be further qualified to various degrees such as small positive, intermediate positive, large positive, etc. Some of the grammars also encode the segment start and end times as part of the primitive. Grammars using primitives derived from slope are used to recognize peaks in waveforms and to describe context-sensitive and time-varying aspects of the ECG.

Only one of the grammars uses an energy metric as a primitive. The metric is computed by squaring the first derivative of the signal and encoding the magnitudes and durations of of the resultant peaks. The goal of this grammar was to detect QRS complexes in the ECG waveform.

In [89] and [42], context-free grammars are noted to be inadequate for representing ECG waveforms due to importance of context in this domain, use of grammars with increased semantic description capability, such as attribute grammars, are recommended as a "proper tool for the description of ECG patterns." Attribute grammars associate numeric, non-structural constraints on the primitives describing the signal. The attributes are determined during primitive extraction and used during recognition to qualify the parser's actions (controlling the parsing process) and quantify parameters.

Although attributes improve the accuracy and noise immunity of syntactic approaches, their presence increases the number of thresholds and heuristic rules that must be determined to benefit from their use.

Following this direction, Trahanias and Skordalakis [99] implement a syntactic approach based on an attribute grammar and expand the set of primitives to include waveform peaks and parabolic segments in addition to the linear segments described above. They assign seven attributes calculated during the extraction phase to each primitive pattern. These include the time and amplitude for a peak location and its boundaries, and an energy metric of the peak. Parabolic and

straight line segments are assigned four attributes: the time and amplitude of the start and end points.

In [99], Trahanias *et al.* extract primitives using empirical criteria. After detecting peaks in the raw signal they model the region around the peak using a cubic spline function. Precise peak locations are indicated by points of maximum curvature in the resulting interpolant. If it is determined that a segment exists between two consecutive peaks, a least-squares fit decides if the segment is linear or parabolic.

The attribute grammar implemented in [99] requires a large number of syntactic rules and numerous empirically-determined thresholds. These thresholds are applied for various width, amplitude, angle, and energy measures in the course of parsing the signal using the grammar and using it to identify points of interest.

In [42], Koski *et al.* describe an segmentation-based approach similar to syntactic methods but using an attributed finite automata instead of a grammar and parser. This algorithm also requires decomposition of the signal into primitives. As with the previous approaches, they first segment the signal into linear regions using a sliding window approach followed by bottom up consolidation. Their primitive comprises a label and two associated attributes. The label is determined from the slope of the segment. Statistical analysis of ECG segment slopes, balanced by the tradeoff between accuracy and computational complexity, led them to map the slope into one of five labels (regions) between $-90°$ and $90°$. The attributes associated with the label to form the primitive are the duration of the segment and the span of its amplitude.

Instead of building a grammar, they create a set of ten *automata* (finite state machines) to represent the signal. Each automaton is responsible for recognition of one component of the ECG signal (*i.e.*, segment, wave, or peak). Automatons invoke each other based on the initial finite state machine in what can be described

as a backtracking approach: if a "called" automaton succeeds in parsing the pattern, the "calling" automaton resumes at the end of the parsed pattern. If the called automaton does not succeed, the calling automaton backtracks and tries another automaton until a match is found or the segment is determined to not comprise a pattern of interest.

Syntactic approaches were pursued in the literature in the 1980s and 1990s, but there is a dearth of publications investigating these approaches after the mid-1990s. Limitations in syntactic analysis are partially responsible — a significant issue identified in [89] is the difficulty in representing time-varying patterns using syntactic methods.

The closest modern analog to these algorithms are those based on Markov Models, which are the natural successor to syntactic parsing and finite state automata in the presence of noise and ambiguity. Such approaches are reviewed in Section 2.6.

2.6 MODEL BASED

Various models have been investigated in the literature to facilitate ECG analysis in general, and in particular, waveform delineation. The ability of models to efficiently represent the signal allows application of mathematical tools to accurately determine characteristic point location.

A pole-zero model of the ECG is proposed by Murthy in [65]. Although this effort does not use the model to estimate characteristic point locations, it does delineate the signal into its component waves and could be expanded to find wave onsets, peaks, and offsets by examining the model parameters or its output. At a high level, the goal of the algorithm is to establish a system of poles and zeros in the z-domain whose amplitude response is highly correlated with the component wave patterns in the time domain signal.

The first step in achieving this goal is to compute the discrete cosine transform

(DCT) of the entire QRS complex. The DCT essentially provides the impulse response of the underlying system being modeled (in this case, the subject's heart). The algorithm next uses the Steiglitz-McBride method to determine an IIR filter with that impulse response, thereby modeling the QRS complex as a set of poles and zeros. The rest of the algorithm comprises expanding the IIR filter into a set of partial fractions, regrouping them, and computing the inverse DCT of the impulse responses of the regrouped functions to obtain the time-domain component waves.

2.6.1 Hidden Markov models

As noted in Section 2.5.3, approaches based on Hidden Markov Models (HMM) have largely replaced those using syntactic pattern recognition. Markov models are probabilistic memoryless state machines. The state of the model at any time is based on only its previous state and a transition probability.

HMMs are Markov models in which the observed data sequences are also probabilistic, representing "hidden", true states which are unobservable. HMMs provide a powerful framework for estimating the state of a hidden stochastic process based on a set of observations. For the arrhythmia analysis problem, the states correspond to the underlying electrical activity of the heart and the observed sequence is the ECG signal.

In [17], Coast *et al.* apply HMMs to identify component waves and perform a "complete" arrhythmia analysis. Although the manuscript is focused on arrhythmia analysis, to identify supraventricular arrhythmias their algorithm is capable of detecting P waves. As such, it provides the earliest instance of HMMs used to delineate (at least part of) the ECG signal.

An HMM is represented by a four-tuple comprising the set of states in the model, the initial state probabilities, a matrix of transition probabilities between states, and a vector of output probabilities modeling the unobservable ("hidden")

underlying stochastic process. Using the HMM requires estimating the model parameters — such as transition and output probabilities — from training data and calculating the posterior probability that a given observation sequence was generated by the model. Maximizing this posterior probability indicates the optimal state sequence and provides insight into the underlying cardiac state of the subject. Coast's algorithm is patient-dependent: models have to be generated for each individual "before and during analysis."

Coast's algorithm uses a separate model for each class of beat to be detected (normal, ventricular, or supraventricular). In the models a state is assigned to each component wave and to the intervals between them. For training, they manually segmented three examples of each beat class and used the mean values of the segments as features for the model.

During analysis, they use the Viterbi algorithm to determine the best state representing the observed ECG signal on a sample-by-sample basis. By simultaneously evaluating multiple models they achieve continuous arrhythmia analysis. The presence or absence of a P wave is indicated by the best sequence picked by the algorithm: only a normal beat has a P wave which is lacking in supraventricular and ventricular beats.

In [16], Clavier *et al.* present another HMM-based algorithm to improve accuracy and address the patient-dependence of Coast's approach [17]. In this effort, they subdivide each component wave into multiple linear segments and use the segment slope in addition to its mean amplitude as features.

Clavier's model comprises 12 states based on a highly-stylized prototype beat. The states include segments and parts of component waves of a typical QRS complex, *e.g.*, rising (left) and falling (right) parts of the P and T waves, rising and falling parts of the R and S waves, and several isoelectric segments connecting these components.

This preliminary effort was trained with 50 beats from 10 different patients

and demonstrated qualitative success in segmenting several beats from different subjects.

In [43], Krimi *et al.* combine the multiscale benefits of the wavelet transform in localizing edges with the statistical power of HMMs for determining the structural properties of a semiperiodic signal. In this approach Krimi first expands on the wavelet method described by Li in [48]. By combining the output of the wavelet transform at two different scales they determine onset and offset points ("edges") more robustly. These edges are used to define signal segments which are the basis for the Markov model.

An effort by Andreão *et al.* [5] describes a similar approach in combining the wavelet transform and HMM to delineate the ECG waveform. The overall approach is very similar to that of [43], with minor differences in the structure of the HMM.

2.6.2 Dynamical ECG model

In [58], McSharry *et al.* describe a dynamical model to realistically simulate an ECG signal. The model is based on three coupled differential equations which generate a three-dimensional trajectory around a limit cycle in state space. The limit cycle defines a circular orbit of unit radius in two of the dimensions, with each revolution corresponding to one interbeat interval. At distinct points on this radius the trajectory is deflected upward or downward by "attractors" or "repellers" which are Gaussian functions with a specific mean representing the angular location of the wave in the circular trajectory, an amplitude to provide the amount of deflection, and a variance specifying the width of the deflection.

This model provides a very powerful and general representation of the ECG signal, effectively as a sum of Gaussians. The angular velocity and parameters of the Gaussians can be modified to accurately simulate numerous physiological conditions.

Although in the original manuscript the goal is primarily to "provide a benchmark for testing numerous biomedical signal processing techniques" and to establish their properties at "different noise levels and sampling frequencies", the model formed the basis of several promising state space approaches not just for simulation, but applied directly for signal analysis and compression. The following publications use this state space model for delineation of the ECG waveform.

In [83], Sayadi *et al.* describe a model-based method for fiducial point extraction in an ECG with baseline wander (low frequency noise). To remove baseline wander they decompose the signal into several bands using a wavelet transform. Outputs from the transform at scales reflecting the lower-frequency baseline drift components are thresholded. Performing an inverse transform results in a baseline corrected version of the original signal.

To determine fiducial points, they fit the clean signal to the McSharry's dynamical model [58]. This fit is achieved by minimizing the mean squared error between the signal and the model over the three parameters specifying the Gaussian components of the model. These parameters are the mean, which provides the angular location; the amplitude, which is the height of the wave; and the variance, which represents the spread of the wave. To facilitate the optimization process, they first determine approximate location of the desired fiducial points by applying heuristic thresholds to the first and second derivatives of the signal in the vicinity of an R wave, then solve the optimization equation to determine the exact model parameters.

Once the model parameters have been established, they use the 99% confidence limits of each Gaussian (*i.e.*, the 3σ point) as the onset and offset time of the corresponding component waves. Peak locations and their values are obtained directly from the Gaussian mean and amplitudes, respectively, determined from the optimization.

Sayadi followed this effort with another similar approach described in [82] and

[84] which uses an extended Kalman filter (EKF) to estimate the system's state and subsequently all fiducial point locations. The EKF structure incorporates the dynamical state space model of McSharry [58], modified to use autoregressive dynamics: since each Gaussian parameter will have very little variation from one beat to the next, its value can be estimated using a first order autoregressive model.

Their EKF incorporates 17 state variables: the system phase representing the current angular location of the model around the limit cycle, the estimate of the ECG signal (summation of all Gaussians), and 15 state variables describing the mean (location), amplitude (height), and variance (spread) of each of five Gaussians representing the P, Q, R, S, and T component waves. The EKF adapts to changes in the underlying ECG signal on a sample-by-sample basis and eliminates the need for a non-linear optimization at each cycle of the ECG as used by [83].

Fiducial point locations are determined using the 3σ point of each Gaussian (in a manner similar to [83]), but with the addition of limits bounding the location. The limits are derived from zero crossings of slope changes in the estimated signal derived by the EKF, and estimate the amount of fluctuation in each fiducial point's location over time.

This EKF approach is extended by Akhbari *et al.* in [2] to use 25 state variables (*vs.* 17 state variables in [82] and [84]). Their states are: the system phase, three states modeling the P, QRS, and T components (in their entirety), and 21 states modeling the mean, amplitude, and variance of seven Gaussians representing component waves. In their model they use two Gaussians for the P and T waves, in order to allow modeling asymmetries in these components.

To address stability concerns with the EKF approach, in [49] Li *et al.* describe another approach based on McSharry's model but using particle filters to estimate model parameters. Particle filters are a sequential Monte Carlo (SMC) algorithm that approximate the optimal Bayesian state estimate for arbitrary joint probability distributions of a system's states and a set of observations [12]. They are

particularly applicable for estimation of non-linear dynamic systems for which the hidden states and observations are not jointly Gaussian (in which case the optimal estimator would be the Kalman Filter).

In this publication, Li proposes a particle filter whose sample values are adjusted with a General Regression Neural Network to mitigate sample degeneracy issues. In order to better represent asymmetrical component waves (a symptom of several disease conditions), they model each wave (P, QRS, and T) with two Gaussians: one for its left part and another for its right. As in approaches using McSharry's model above, the state variables specify the mean, amplitude, and variance of the Gaussians representing the component waves, and they are evolved by the particle filtering using a random walk process. As before, fiducial point locations representing the onsets and offsets of characteristic waves are obtained using the 3σ point of each estimated Gaussian.

Another Bayesian model-based approach is described by Lin *et al.* in [51]. In that effort Lin describes use of a partially-collapsed Gibbs sampler (PCGS) to delineate the P and T waves of an ECG signal. Gibbs samplers are a Markov-chain Monte Carlo (MCMC) method that facilitate the calculation of a Bayesian marginal posterior distribution of a state given a set of observations.

In this effort, Lin restricts his approach to P and T wave search blocks that are created by extracting and concatenating several successive left (for P) and right (for T) neighborhoods of previously detected QRS complexes using offsets determined by the current R-to-R interval (heart rate).

Lin then models the signal in these search blocks as the convolution of an unknown impulse response with an unknown input pulse sequence, with additive Gaussian noise. By further decomposing the model, he obtains a parameter vector fully describing the model. The model equation is also used to derive a likelihood term representing the conditional probability of the observed data given the parameter vector. Priors are assumed as a Bernoulli-Gaussian sequence.

The complexity of computing the resulting posterior distribution precludes use of the more commonly used estimators such as mean squared error and the maximum *a posteriori* estimator, so the PCGS is used to create the Bayesian estimate.

In [50], Lin *et al.* use the same model as [51] but employ a particle filter approach. Specifically, to address concerns with computational complexity with the increasing number of particles, they propose a marginalized particle filter to reduce the number of parameters (and particles). Linear state parameters are marginalized out and estimated using one Kalman filter per particle. The nonlinear state variables are then estimated using a particle filter.

2.7 SPLINE BASED

The use of splines for ECG processing has been primarily focused in the areas of waveform compression ([35, 39, 45, 52, 86]) baseline wander elimination ([8, 23, 56, 60]), and waveform analysis ([34], [38], [87]).

Generally, compression efforts use the spline's knots as a compact representation of the signal and use the interpolant to accurately recreate the waveform itself. Reconstructed signal fidelity is very important in compression applications, and as noted by [39], "The segmented nature allows splines to adjust very efficiently to local characteristics of the data and represent it better (with smaller deviations) than other classes of functions."

For compression applications, various methods are used to determine knot locations. [35] uses extrema or points of large curvature from the the second difference of the signal. [86] compresses only the segment between consecutive beats, using a syntactic approach with an attribute grammar to determine the best knot location. In a manner similar to bottom up segmenting described in Section 2.3.2, [39] starts with a very large number of knots, and then removes knots until the error (as measured by a least mean-squared metric) in the reconstructed signal exceeds a tolerance. [52] uses an amplitude-threshold based method to determine

knot locations: if the difference between a sample and the one following it by two intervals is greater than a threshold value, it is kept as a knot for cubic spline interpolation — in one variant this technique is applied only to the segment between consecutive beats.

Splines are useful for baseline wander elimination as they provide a good representation of this low-frequency noise with a few consistently-identified regions from each beat. And since the noise filtering is accomplished by subtraction of the interpolated spline representation from the original signal, important low-frequency components of the ECG signal itself, like the ST segment, are not impacted. In one of the first efforts to use splines in baseline wander elimination, [60], Meyer *et al.* use cubic splines generated solely from samples of the PR segment of each beat.

There are only a few efforts describing the use of splines for analysis of the ECG signal. In [8], 100 milliseconds intervals prior to each R wave are searched for a fiducial point indicated by minimizing a figure of merit based on the derivative of the signal. These fiducial points serve as the knots for a cubic spline approximation that provides an accurate isoelectric reference for computing ST segment deviations.

In [34], Huang *et al.* use a spline-based approach to detect QRS complexes in the presence of noise. First, they compute a moving average of the original signal for every N points. They then fit a cubic spline to the averages, replacing each of the N points with the interpolated spline approximation. This interpolation provides an estimate of baseline wander which they subtract from the original signal. They then fit another cubic spline to the local maxima of the resultant signal, and apply a threshold to the approximation to detect QRS complexes.

In [38], Kalovrektis *et al.* segment the QRS complex and calculates a cubic spline representation. The coefficients defining the cubic spline for that signal segment serve as features for a classifier trained to discriminate between normal

sinus rhythm and several arrhythmias.

Another use of splines for analysis is described in [87], for calculating an ECG-derived respiration signal. In this effort Shayei *et al.* detect R wave amplitudes and uses a cubic spline interpolant to generate a waveform approximating the respiration signal. Respiration is present in the R wave amplitude due to fluctuations in the cardiac vector as the heart is moved by the diaphragm during breathing. Fitting a spline to the R peaks also provides a signal that is sampled uniformly in time facilitating analysis of the respiration signal.

Summarizing the survey of spline-based methods, for applications in compression the goal is to reduce the number of knots while maintaining high fidelity in the reconstructed signal, so the knot locations themselves are not of primary importance.

For applications in baseline wander elimination, knot locations are predetermined at particular locations of the waveform that correspond to its isoelectric level to allow the spline approximation to approximate low frequency drift of the signal. In these approaches the isoelectric level is determined first, then the knots are placed correspondingly.

And the spline-based ECG analysis efforts use either interpolant coefficients or the interpolated estimate itself. Knot locations, while important to create a good representation of the signal, are not used in and of themselves for analysis or estimation.

There is a long history of splines in ECG signal processing, but there is no evidence of research into use of knot locations to determine the characteristic points of a semiperiodic signal. Even though this concept was recognized and articulated in 1978 by Mier-Muth and Willsky [61], it was not developed in that effort and there is no subsequent research linking the knots used in a spline representation to a waveform's characteristic points until 2011 [29].

Unlike the spline-based efforts for waveform compression, analysis, and filtering,

the algorithm for optimal ECG delineation described in this dissertation strives to accurately place knots on a signal's critical points, and use those points as estimates of the characteristic points defining onset, peak, and offsets of the signal's component waves.

To do so it applies splines in a novel manner. The approach starts with a parametric representation of the signal using splines defined by a predetermined set of knots, each corresponding to a characteristic point of interest. Using a Bayesian figure of merit, the algorithm then fuses information from the observed signal and its parametric representation with prior knowledge of characteristic point locations (obtained from manual annotations of training data), to determine the optimal knot locations estimating the signal's characteristic points.

Chapter 3

A SPLINE FRAMEWORK FOR REPRESENTING SEMIPERIODIC SIGNALS

Before describing a method to optimally determine the locations of a semiperiodic signal's characteristic points in Chapter 4, this chapter presents a framework that can be used to represent such signals with splines. It also provides the results of an effort to assess one implementation of the spline framework on a set of ECG signals exhibiting a wide variety of morphologies.

The framework provides a general capability for an optimal parametric representation of semiperiodic signals, not limited to characteristic point estimation. By proper choice of constituent algorithms the framework allows customization for various needs including signal compression, noise elimination, and analysis.

3.1 ALGORITHM DESCRIPTION

The spline framework is illustrated in Figure 3.1. For a given input signal, it first determines initial knot locations, then computes a representation of the signal using those knots and the chosen spline interpolant. The optimization algorithm modifies knot locations until an objective function, which can be a fitness function or error criterion, is satisfied. The error value itself may be used in the optimization process, for example, if a gradient-based method is employed.

Upon completion of the optimization process, the framework provides the optimal knot locations using their constituent times and amplitudes, the interpolated estimate of the signal computed using the optimally-placed knots, and the value of the objective function.

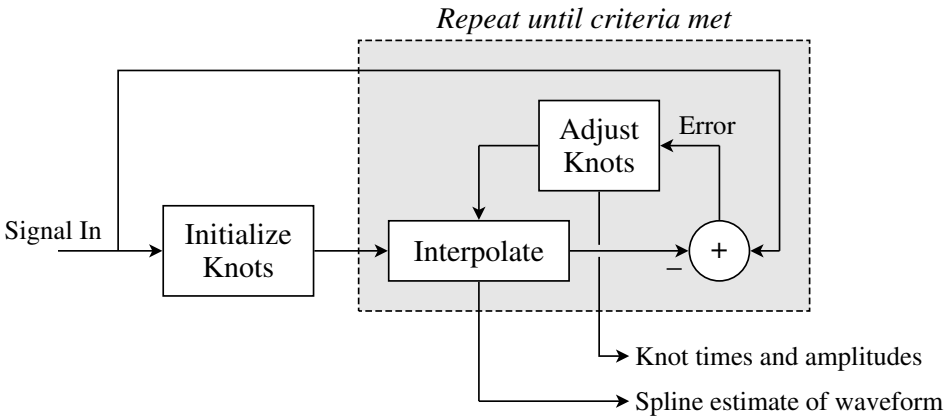


Figure 3.1: The spline framework for representing semiperiodic signals.

By allowing choice of knot initialization algorithm, type of spline interpolant, objective function, and optimization algorithm, the framework provides great flexibility to meet an application's specific requirements. Choices of these components can impact computational resources, speed, fidelity of the interpolated signal, and correspondence of knot locations to a signal's characteristic points.

Although Figure 3.1 shows knot adjustment only, in a more general implementation the optimization algorithm could consider any other parameters influencing the model. For example, the choice of spline interpolant could potentially be considered in the optimization process for each segment of the signal under analysis.

3.1.1 Knot initialization

The number of knots required to represent a particular signal is highly dependent on its morphology, so this step of the framework requires an adaptable, efficient method to determine how many knots are required, and to make the initial knot location assignments.

Using too few knots will adversely impact ability of the spline interpolant to estimate the underlying signal, and will limit its representation of the underlying characteristic points.

On the other hand, using too many knots will decrease efficiency of the spline representation and increase the computational effort required to implement the framework. Additionally, for applications estimating characteristic point locations, too many knots can significantly complicate determination of the correspondence of knots to characteristic points.

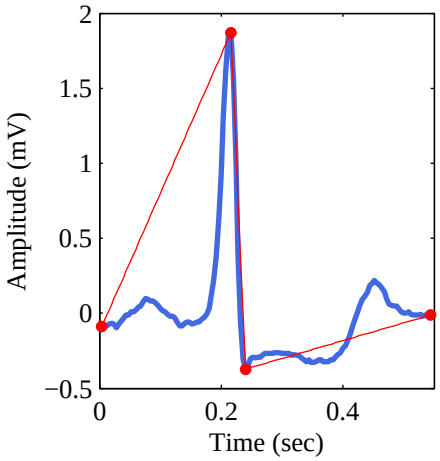
For knot initialization, this effort implements a recursive partitioning algorithm (RPA), described in [27] and [28] for a waveform compression application. RPA belongs to the class of top down segmentation algorithms described in Section 2.3.2.

The recursive partitioning algorithm linearly interpolates between the endpoints of the signal segment and finds the point on the waveform with the greatest error between the signal and its interpolated estimate. It then recursively applies itself to the new line segments generated by each existing endpoint and the point of greatest error. The recursion terminates when the maximum error between the interpolated and original waveforms is less than an empirically determined, application-dependent threshold selected to maintain key features of the signal under analysis.

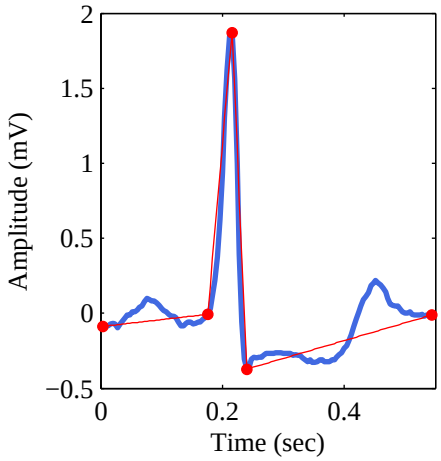
Figure 3.2 illustrates operation of the RPA at different stages on a sample ECG complex extracted from the European ST-T database (EDB) [25]. Figure 3.2a follows the second partition. The next largest difference between the interpolated and original waveforms is the onset of the R wave, which is picked for the third partition in Figure 3.2b. The final result is shown for a threshold value of 0.1 millivolt in Figure 3.2c. Lowering the threshold further would have resulted in partitions at approximately 0.3 and 0.5 seconds.

The apparently-unnecessary knots identified by the RPA on the steep ascending and descending segments of the R wave are due to the morphology of the signal in those regions, where it is a very steep sigmoid with large amplitude. As a result differences between the signal and its linearly-interpolated estimate are relatively great at the points of highest curvature on the sigmoid, which necessitates

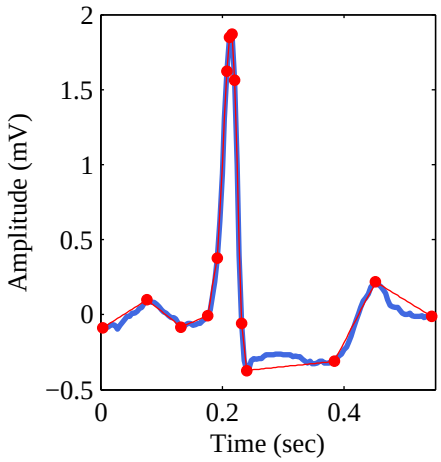
additional knots to meet the desired amplitude threshold.



(a) After second partition. There are four knots with RMSE of 572 μV.



(b) After third partition. There are five knots, with RMSE of 133 μV.



(c) After the final partition achieving the 0.1 mV threshold. There are 14 knots, with RMSE of 48 μV.

Figure 3.2: Example illustrating the Recursive Partitioning Algorithm. This figure shows RPA applied to a beat from EDB record `e0406`, with results after the second, third, and final partitioning operations. The thicker line is the original waveform; the thinner one the linearly-interpolated estimate based on the knots, which are indicated with circular markers. Knots determined in this manner are used as initial locations for the optimization algorithm.

With proper choice of threshold, RPA results in initial knot locations that serve as a good starting point for subsequent optimization.

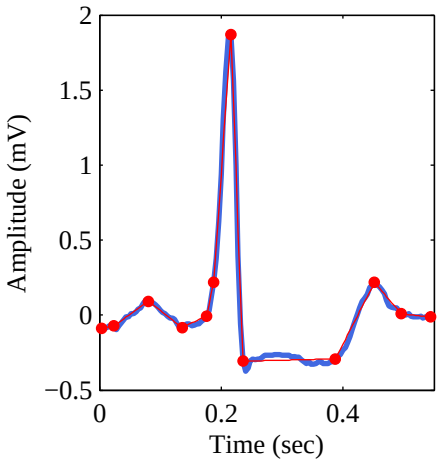
3.1.2 Spline interpolant

The interpolant choice is very important to ensure the goals of a given application of the spline framework are met. For accurate signal reproduction, for example, in applications like waveform compression, a reasonable, but not perfect, criterion is the root mean square error between the interpolated approximation and the original waveform. Due to the large magnitude and steep slope of the QRS complex, small errors in representing the signal in this region can result in values of RMSE that are disproportional to the clinical importance of this region of the ECG signal. In contrast, errors in the smaller component waves (P and T) and important signal segments (such as the ST segment) produce a much smaller RMSE but are clinically more important.

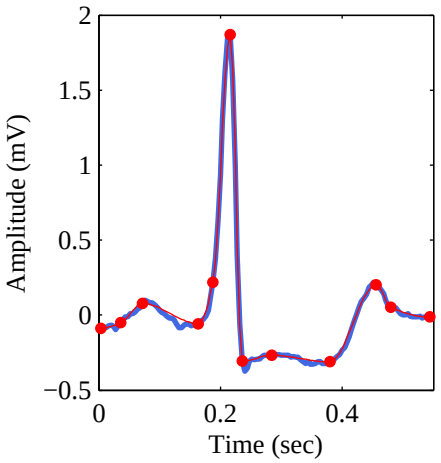
Also, knots optimized using spline interpolants with the RMSE criterion may not fulfill the requirements for waveform delineation, which require knots to fall on the waveform's characteristic points.

To illustrate the effect of interpolant choice on the estimated signal's fidelity and on optimized knot locations, Figure 3.3 shows results of the framework on the same beat used in Figure 3.2. The original beat is superimposed with the results of three different interpolants, which were chosen to explore the effects of differentiability constraints: a linear interpolant with no constraints, a cubic Hermite interpolant that requires the first derivative to be continuous at each knot location, and a cubic spline interpolant that requires the second derivative to be continuous at each knot location.

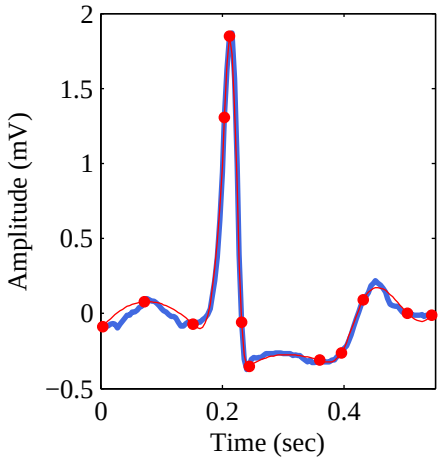
For this exercise each interpolant was run with 12 randomly initialized knots and optimized using a genetic algorithm as described in Section 3.1.3 below.



(a) Linear interpolant, RMSE 44.4 μV



(b) Cubic Hermite interpolant, RMSE 23.0 μV



(c) Cubic spline interpolant, RMSE 43.0 μV

Figure 3.3: A comparison of linear, cubic Hermite, and cubic spline interpolants. Each plot shows optimized knot locations, interpolated signal estimates, and the RMSE for the indicated interpolant on a beat from EDB record `e0406`.

From Figure 3.3 it is clear that with a sufficient number of knots the linear, cubic Hermite, and cubic spline interpolants all can represent the signal effectively. As measured by RMSE, the spline interpolant's estimate of the underlying signal is worst for the linear interpolant with an RMSE of 44.4 μV, followed by the cubic

spline and cubic Hermite interpolants, with RMSE values of 43.0 μV and 23.0 μV, respectively.

The large error in the linear interpolant shown in Figure 3.3a is due primarily to amplitude differences between the spline representation and the original in areas of high slope around the QRS complex, especially on the ascending branch of the wave. The signal in these areas is sigmoidal with large amplitude, resulting in greater differences at regions of high curvature in the sigmoid.

The ability to inherently represent curvature that is provided by the cubic spline interpolant of Figure 3.3b reduces its error in the QRS complex. However, the limited number of knots, coupled with its differentiability constraints, cause this interpolant to not accurately reflect the shape of the signal for the P and T waves.

Figure 3.3 also shows that if the framework is to be used for estimating the signal's characteristic points, RMSE is not the only criterion to consider: knot locations following optimization should ideally reflect locations of the signal's characteristic points. With a linear interpolant the optimization process moved knots to locations providing good estimates of the characteristic points for component waves of the complex. As seen in Figure 3.3a a subset of the knots are positioned very well with respect to these points and correspond to the onsets, peaks, and offsets of the P, QRS, and T waves.

The second-order differentiability constraint imposed by the cubic spline, and the first-order differentiability constraint for the cubic Hermite interpolant resulted in the optimization algorithm to move some of the knots away from the desired characteristic point locations. Specifically, the cubic Hermite interpolant in Figure 3.3b does not have a knot on the P wave offset, and its T wave peak is slightly off. To a lesser degree, the knot representing the onset of the QRS complex is too far to the left.

The errors in knot locations for the cubic spline interpolant shown in Figure 3.3c

are more numerous. There is no knot corresponding to the P wave onset, and although there is a knot between the end of the P wave and start of the QRS complex, it does not fall on the P wave offset or QRS onset so both of these points are impacted. In addition, there is no knot on the T wave peak.

3.1.3 Optimization algorithm

In the space defined by all possible knot locations, the optimization algorithm must determine the location that results in the best fitness function. The effort described in this chapter assumed no constraints on knot locations, and performed the optimization using a genetic algorithm.

Genetic algorithms (GA) are a search technique inspired by natural selection that iteratively evaluate randomly modified permutations of potential solutions to discover the best one, as determined by a fitness function. Benefits of genetic algorithms include that they are relatively easy to implement, and that they do not use the gradient of the fitness function which for this application can be difficult to determine. However, there is no guarantee that they will find a global optimum and they are computationally expensive.

At a high level, the GA approach represents the inputs subject to optimization as a population of individuals encoded as chromosomes, which are then modified and evaluated in iterations called generations. In this application an individual is a set of knots parameterizing the spline to be evaluated for fitness, and is described by a chromosome comprising the knot locations.

The modification process creates new individuals by mutating existing individuals and randomly combining elements from different members of the population. Copies of the most fit individuals are propagated based on the fitness criteria [22]. The following steps outline the GA functionality used by the optimization algorithm:

1. The initialization algorithm creates the initial population by determining the

starting knot locations using RPA, and augmenting the RPA output with a set of randomly-perturbed versions. It provides a pool of 25 individuals defining the population and capturing potential knot locations to be evaluated.

2. Next, the algorithm determines the fitness of each individual in the entire population by computing the RMSE between the original signal and its estimate obtained with the desired spline interpolant parameterized by each set of knots created in step (1).

3. The selection process then chooses two parents from the current population using a fitness-proportional selection criterion, also known as roulette wheel selection. In this method the probability of choosing a parent is directly proportional to its fitness, much like a roulette wheel with non-uniformly sized sections. The fitness-proportional process provides *all* individuals a chance to be selected for propagation, but gives a greater probability to those that are more fit.

4. The algorithm next performs a single-point crossover operation on the selected parents in an attempt to develop an offspring that is more fit than either of the parents. To do so it selects a random point, and switches the contents of the two parent's chromosomes following that point.

5. To maintain diversity in the genetic pool of potential solutions between generations, the algorithm next performs a mutation on the chromosomes. It randomly selects a knot, then randomly perturbs its location in a constrained region around its current location, *i.e.*, bounded by its neighboring knots.

6. The next step augments the population with the best offspring. Replacing the individual in the population having the worst fitness with the best offspring from step (5) results in an improvement to the overall fitness of the population.

7. The algorithm repeats the process starting with step (2) for 500 generations to evaluate a sufficient number of possibilities. This value was determined empirically by observing, over numerous tests, how many generations were required before little or no benefit was seen in the fitness function.

At the end of the process the individual with the best fitness function represents the best knot locations describing the signal with the specified interpolant.

3.2 ALGORITHM ASSESSMENT

To help further understand the limitations and advantages of this implementation of the framework it was tested on a larger number of ECG waveforms extracted from the EDB. Selected waveforms were chosen to demonstrate the algorithm's ability to adapt to a wide range of pathological morphologies including fusion beats, ST segment deviation, T wave deviation and inversion, and premature ventricular contraction. The signals produced by these pathologies are generally more challenging for automated techniques than those from healthy subjects, often having more peaks and greater variability in their features.

The EDB provides ECG data collected using two leads (*i.e.*, electrode placements) on each subject. It also provides a set of "truth" annotators indicating the location of each beat in the signal, using the precise location of its R wave peak. Using these annotators, ECG signals were extracted from both available channels for several subjects prior to exercising the framework. To ensure that the ability of the algorithm to represent P and T waves was tested, the length of the extracted patterns was long enough to include these component waves.

Recursive partitioning, used with beat-dependent thresholds of approximately $100\,\mu$V, determined the number of knots required for each beat and their initial locations. The genetic algorithm described above then optimized their location in different runs using linear, cubic Hermite, and cubic spline interpolants. Each run,

independent of interpolant used, was started with the knots in the initial locations determined by RPA for each beat.

RPA provides a good initial estimate of knot locations prior to optimization, but the choice of threshold is very important: smaller values will result in a more sensitive algorithm which will be likely to identify undesired artifact as initial knot locations. And larger values may miss required features, adversely impacting the algorithm's ability to represent the signal accurately.

Figure 3.4b illustrates a good compromise for a waveform with a large amount of line noise which is seen as oscillations on the baseline of the signal. Smaller threshold values for this example resulted in all peaks and valleys being being selected as knot locations. With proper threshold and RMSE criterion, the interpolated result effectively reduces the line noise while still maintaining the ability to represent characteristic points by the knots and the signal using the interpolated estimate.

3.3 RESULTS AND DISCUSSION

Table 3.1 shows the aggregate error statistics for each beat across for all interpolants used in the evaluation. For each beat it shows the number of knots used in the spline representation and the RMSE values between the original waveform and its spline estimate for several interpolants. These include RMSE for the original non-optimized knot locations determined by RPA, and knots optimized using linear, cubic Hermite, and cubic spline interpolants. The final row reports the mean and standard deviation of the RMSE values for each method.

RMSE values for the cubic Hermite interpolant are superior to those using linear and cubic spline interpolants for all but two cases in which the linear interpolant is better. This experience with the cubic Hermite interpolant is consistent with [1], in which Hermite basis functions are used to model the ECG signal.

The mean RMSE across all beats for the non-optimized initial knots produced

Table 3.1: Spline framework results for linear, cubic Hermite, and cubic spline interpolants on beats from both leads of several subjects in the EDB. n_k is the number of knots determined by the RPA using a threshold appropriate to each beat ($\approx 100\,\mu$V). The RMSE for the initial (pre-optimized) knots determined by RPA is given by e_1. The RMSE for optimized knots using a linear interpolant is given by e_2, using the cubic Hermite interpolant by e_3, and using the cubic spline interpolant by e_4. All RMSE values are in μV.

Record	Lead	n_k	e_1	e_2	e_3	e_4
e0114	MLIII	16	20.4	14.4	14.4	18.0
	V4	18	19.0	14.2	12.1	17.6
e0116	V4	11	57.2	44.1	42.2	139.4
	MLIII	9	30.8	21.7	22.4	39.1
e0123	V4	23	20.7	16.7	9.7	14.8
	MLIII	14	13.3	11.1	13.1	22.7
e0161	V4	22	17.0	12.8	7.9	9.1
	MLIII	15	18.5	11.7	9.0	12.0
e0206	V5	19	29.9	25.8	23.7	40.1
	MLI	14	23.7	16.7	14.6	25.9
e0413	V2	13	36.3	32.7	17.7	144.5
	V5	11	41.0	37.0	27.9	150.4
$\mu \pm \sigma$ (μV)			27.3 ± 12.6	21.6 ± 11.0	17.9 ± 9.9	52.8 ± 56.3

by RPA, using a linear interpolant, is 27.3 μV. Optimizing these locations with a linear interpolant improves the RMSE to 21.6 μV. As with the beat shown in Figure 3.3, optimization using the cubic Hermite interpolant results in the highest fidelity representation, with a mean RMSE of 17.9 μV. Unlike that example, however, optimizing using the cubic spline interpolant resulted in a significantly greater mean RMSE than the linear interpolant, at 52.8 μV.

Comparing figures illustrating the results of these interpolants provides an explanation of this effect. By using the RMSE as the optimization criterion, the algorithm will usually favor moving knots to areas of greatest curvature and slope because these locations generally have the biggest impact on this measure of error. For the linear and cubic Hermite interpolants, limited differentiability constraints allow the interpolant to represent the remaining signal accurately, especially when fewer knots are available to cover areas of lower curvature in the waveform.

With the cubic spline interpolant, however, the second order differentiability constraint precludes it from accurately representing relatively linear regions if there are not a sufficient number of knots remaining after addressing the areas of high slope or curvature in the signal.

The RPA initialization algorithm required relatively few knots for the waveforms of EDB record `e0116`. Interpolated estimates after optimization with the linear and cubic Hermite interpolants, as illustrated in Figures 3.4b and 3.5b, show a sufficient number of them located to accurately represent areas of high curvature and slope, in particular the QRS complexes and the severely elevated ST segment of the bottom complex. Fewer knots are required to represent the flatter areas of these waveforms, and these interpolants provide an estimate that is very close to the original signal in these regions.

There are approximately the same number of knots per signal segment for this record in the optimization using the cubic spline interpolant shown in Figure 3.6b. In this case, however, the interpolant's ability to accurately represent areas of low

slope with only two or three knots is limited. This is especially evident on the top waveform, which has a large sinusoidal interpolated estimate for a linear segment. The long, flat, elevated ST segment of the bottom waveform required three knots to limit the error in that region, which left only three to represent the QRS complex. As a result the interpolant has large error in that region as well.

The waveforms for EDB record `e0413` demonstrate a very similar result. Again, RPA required relatively few knots and the allocation of knots per segment after optimization is similar for all three interpolants. However, while the linear and cubic Hermite interpolants shown in Figures 3.4f and 3.5f represent the remaining parts of the signal quite accurately, the cubic spline in Figure 3.6f again results in a large sinusoidal estimate for the flat parts of both waveforms. In addition, its ability to represent the T wave of the bottom waveform is significantly impacted as well.

3.4 SUMMARY

This chapter presents a novel spline-based framework for parametrically representing semiperiodic waveforms in a highly flexible manner. Like any framework, this one provides a high-level construct that separates out key functionality into various components, each of which can be selected or modified to achieve specific goals. A key advantage of the framework is its incorporation of an optimization component, building in the capability to obtain an optimal representation of some feature of the signal under analysis.

In contrast, previous approaches using splines are much more specific than the spline framework presented in this chapter. Each published method described in Section 2.7 targets a specific usage such as waveform compression, baseline noise elimination, or signal analysis. The choices of knot initialization and type of interpolant are selected to satisfy the problem being addressed.

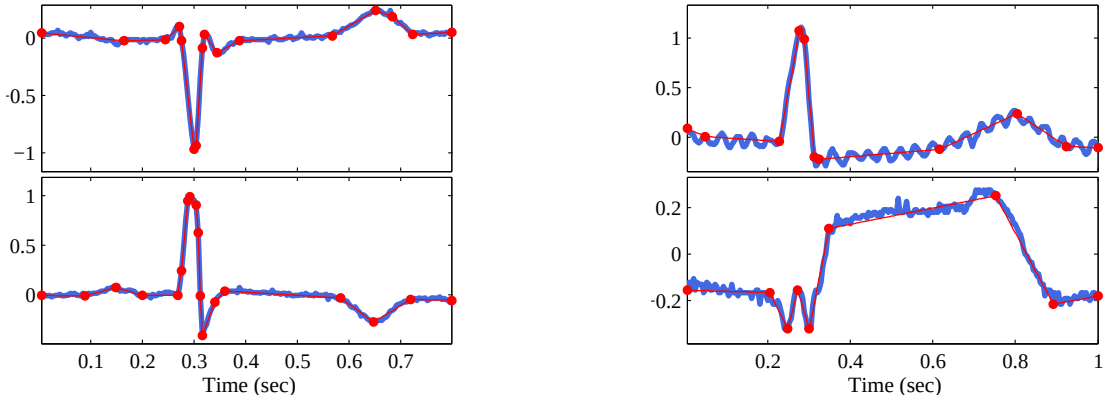
The implementation of the framework described in this chapter was intended

to explore its application to ECG signals and to understand the tradeoffs inherent in the framework for various choices of its constituent algorithms. These include a knot initialization algorithm, error criterion, interpolant, and optimization algorithm; choice of each allows tailoring the framework for a given application.

In applications requiring high-fidelity representation of the signal itself (*e.g.*, for data compression applications), the cubic Hermite interpolant is preferable since it represents the signal more accurately as measured by RMSE. Although it is possible to decrease the RMSE of the cubic spline interpolant for the problematic waveforms described above, those cases will require more knots, reducing the efficiency of the spline representation.

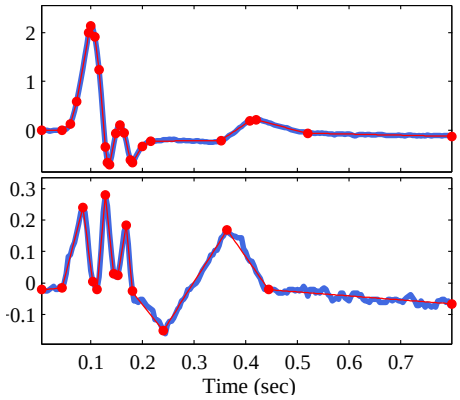
For waveform delineation applications in which the goal is to precisely determine the locations of the signal's characteristic points, the interpolant should produce the best RMSE when optimization locates the knots at the signal's characteristic points. As observed on the few beats evaluated in this effort, the linear and cubic Hermite interpolants behave very similarly. However there are several cases where the linear interpolant performed better in this regard, and it is computationally much simpler. For this reason, the linear interpolant is used for the optimal characteristic point estimation algorithm that will be described in Chapter 4.

Results from a small but diverse set of beats indicate that the spline framework is a viable, complementary option to existing methods for parametric modeling of the ECG waveform. And with appropriate choices of knot initialization, interpolant, optimization algorithm, and objective function, can be used to efficiently represent signals with high fidelity or to estimate locations of the waveform's characteristic points.

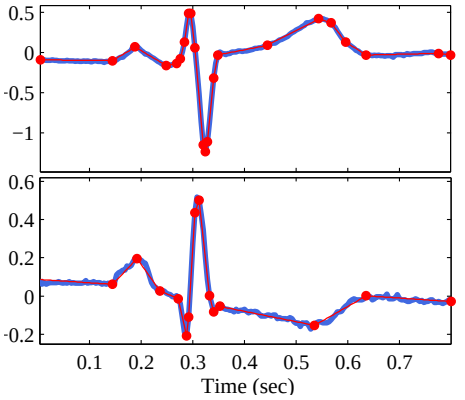


(a) e0114

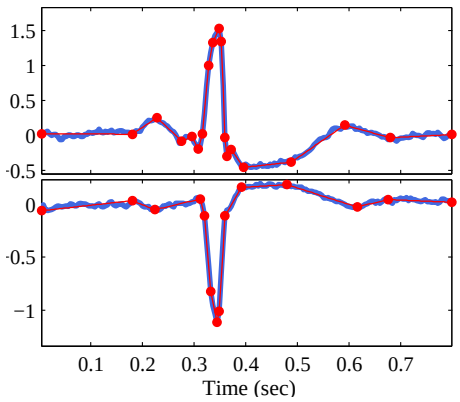
(b) e0116



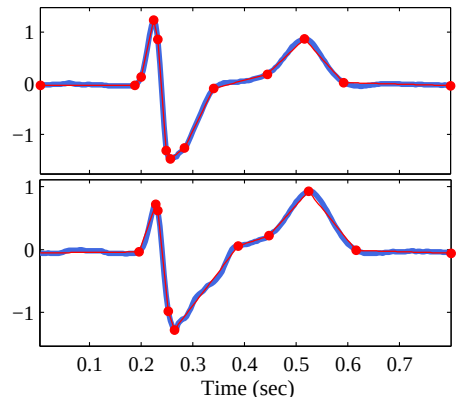
(c) e0123



(d) e0161

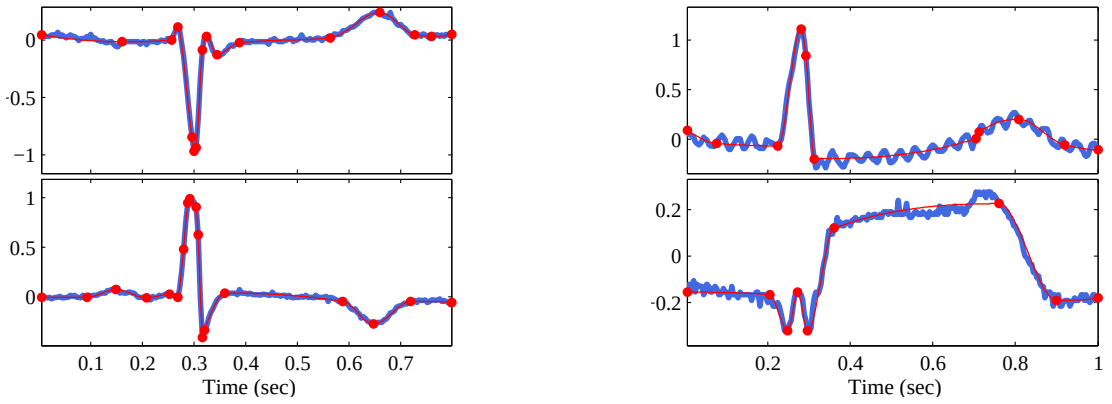


(e) e0206



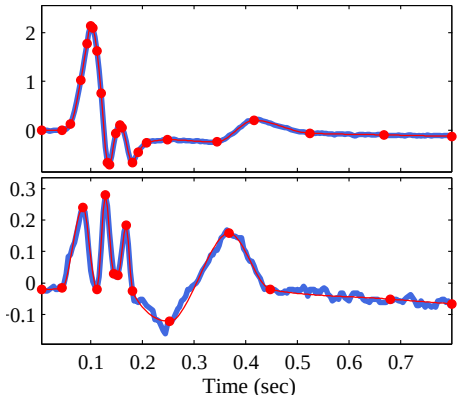
(f) e0413

Figure 3.4: Results of optimization with a linear interpolant. All amplitudes are in mV. Two leads are shown for each beat, details in Table 3.1. The thicker line is the signal and the thinner one is the interpolated estimate. Circular markers indicate optimized knot locations.

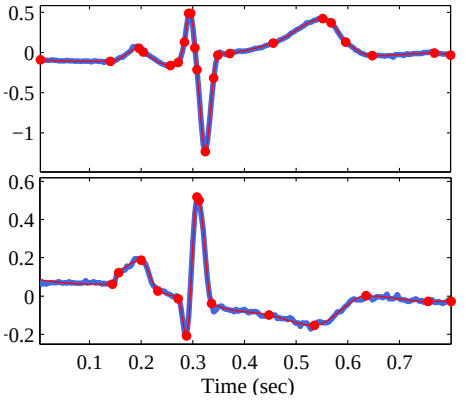


(a) e0114

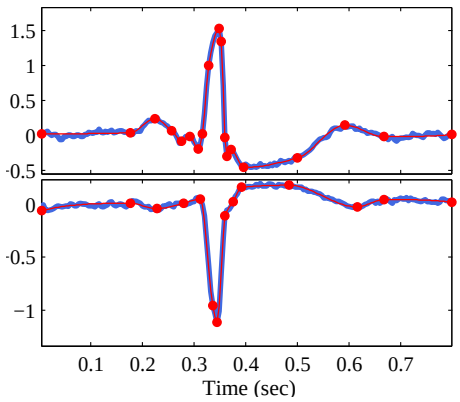
(b) e0116



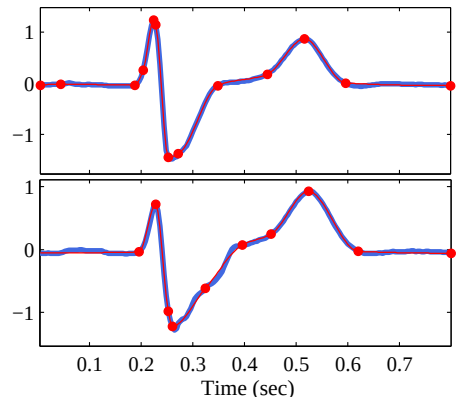
(c) e0123



(d) e0161

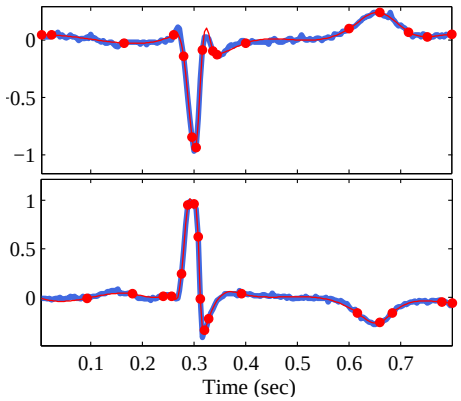


(e) e0206

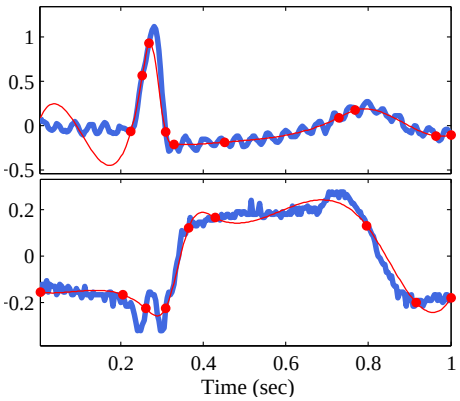


(f) e0413

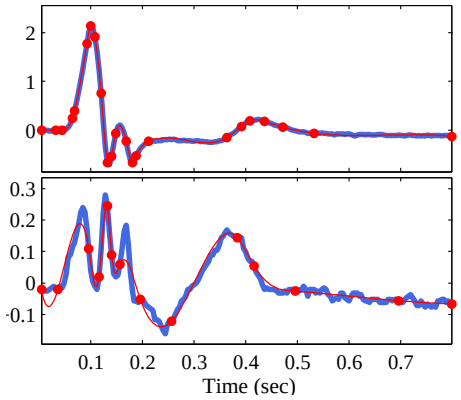
Figure 3.5: Results of optimization with a cubic Hermite interpolant. All amplitudes are in mV. Two leads are shown for each beat, details in Table 3.1. The thicker line is the signal and the thinner one is the interpolated estimate. Circular markers indicate optimized knot locations.



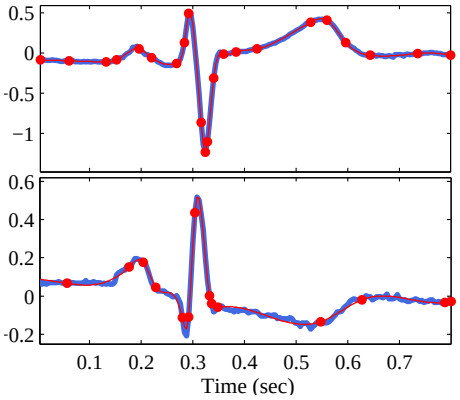
(a) e0114



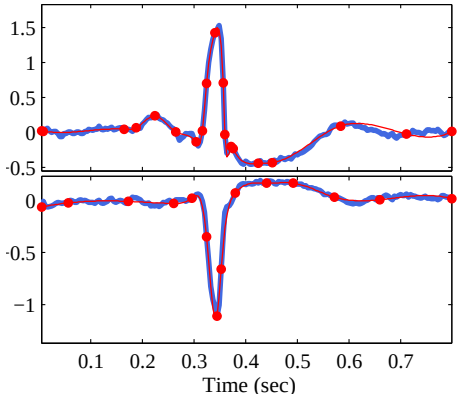
(b) e0116



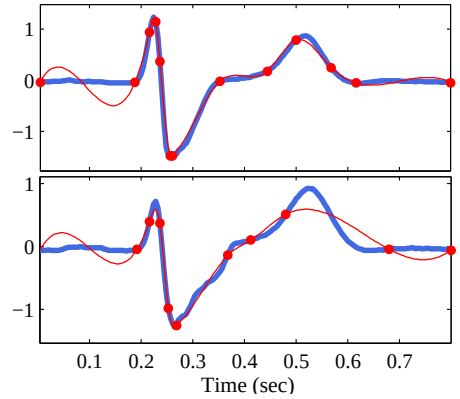
(c) e0123



(d) e0161



(e) e0206



(f) e0413

Figure 3.6: Results of optimization with a cubic spline interpolant. All amplitudes are in mV. Two leads are shown for each beat, details in Table 3.1. The thicker line is the signal and the thinner one is the interpolated estimate. Circular markers indicate optimized knot locations.

Chapter 4

OPTIMIZED CHARACTERISTIC POINT ESTIMATION FOR SEMIPERIODIC SIGNALS

This chapter builds on the work described in Chapter 3 by creating an instance of the generic spline framework to determine the location of a pre-defined set of characteristic points using Bayesian optimization. The algorithm described here is a refined and improved version of the one described in [30]. As before, the example signal used here is the electrocardiogram, although the approach can be applied to any semiperiodic signal requiring precise location of characteristic points on a cycle-by-cycle basis.

Previously it was shown that the spline framework illustrated in Figure 3.1 provides a general algorithmic approach that is highly customizable to represent semiperiodic signals based on knot locations and an interpolant. The knots serve as constraint points to the interpolant which generates an approximation to the underlying signal.

Use of the framework requires specification of a knot initialization algorithm, an interpolant function, an objective function (which could be a figure of merit or error criterion), and an optimization algorithm. The initial knot locations are iteratively modified by the optimization algorithm to determine the locations resulting in the best figure of merit for a given interpolant. The choices that define an implementation of the framework are very important and depend on the goals of the application. For example, it was shown in Chapter 3 that knot locations of a linear interpolant can effectively identify a waveform's characteristic points after optimization, even though that interpolant's ability to accurately represent the

signal as measured by RMSE is generally inferior to a cubic Hermite interpolant.

The previous effort initializes knots using a recursive partitioning of the waveform and optimizes their positions using the RMSE between the observed signal and its interpolated approximation as its objective function. Recursive partitioning is highly dependent on QRS complex morphology and results in a variable number of knots. This makes it very difficult to maintain correspondence between the knots and specific ECG characteristic points, a limitation that necessitates post-processing to map knots back to characteristic points after their locations have been determined.

Furthermore, using RMSE as the sole error criterion provides an overall measure of error between the signal and its interpolated estimate, while neglecting information present in the times and amplitudes of knots describing the waveform.

The implementation of the framework described in this chapter addresses some of the shortcomings of the previous approach as they pertain to accurately determining a signal's characteristic points. It does not address applications requiring accurate, compact representation of the signal itself.

The algorithm described in this chapter uses a Bayesian figure of merit to fuse information provided by each cycle of the semiperiodic waveform with prior knowledge derived from a separate training data set, in order to determine the best knot locations corresponding to a fixed number of the waveform's characteristic points.

The intent of this work is to demonstrate an effective algorithm for analyzing ECG signals in order to better understand and characterize changes to the waveform in long recordings. This approach enables a more thorough exploration of ECG signal morphology, especially on extremely large data sets where manual annotation of multiple points for all beats is not feasible. Researchers can specify points of interest, possibly differing from those described here, and quantify prior knowledge for those points by manually annotating a representative sample of their

data. The algorithm can then automatically analyze a very large number of beats and determine the best locations for the desired points.

4.1 SPLINE FRAMEWORK CONFIGURATION

To maintain correspondence of knots to the characteristic points of the underlying ECG waveform, this effort uses a fixed number of knots with constrained relative locations determined by the desired characteristic points. This facilitates finding and tracking specific points defining waveform onset and offset times, peaks, intervals, segments, and other locations that may be of interest in tracking subtle changes to the waveform over time. A number of these measures are illustrated in Figure 1.4.

The algorithm implements the spline framework described in Chapter 3 with the following customizations:

Interpolant: the algorithm uses a linear interpolant to estimate the signal, due to its computational simplicity and desirable properties in locating the waveform's characteristic points at boundaries of signal segments in the spline representation.

Knot initialization: knots are initially placed on the waveform at the mean time values of the *a priori* probabilities of the corresponding characteristic points (frequently shortened to simply "priors"), which are determined from manual annotations on the training data.

Figure of merit: to find the best knot location during optimization iterations, the algorithm implements the maximum *a posteriori* probability (MAP) of a Bayesian estimator. The MAP fuses goodness of fit of the interpolated signal estimate with *a priori* probabilities of characteristic points determined from a manually-annotated training data set.

Optimization algorithm: knot locations are optimized using the cyclic coordinate method. Each knot representing the waveform is successively swept through all locations between its immediate neighbors, which are kept fixed. The location with the best figure of merit is kept as the optimal location. This process is repeated several times in succession for all knots to help ensure convergence to the best figure of merit.

4.2 CHARACTERISTIC POINTS AND SUPPORT POINTS

One of the key differences between this effort and the one described in Chapter 3 is use of a fixed number of knots in the spline representation, each of which corresponds to one of the characteristic points modeled with prior knowledge of waveform morphology.

4.2.1 Characteristic points

The algorithm estimates locations of characteristic points chosen to describe the waves comprising a QRS complex. Knowledge of the location of the onsets, positive peaks (maxima), negative peaks (minima), and offsets of component waveforms allows calculation of the clinically-important intervals, durations, and segments shown in Figure 1.4, among numerous other possibilities.

An independent beat detection algorithm run on the ECG signal identifies the precise location of R wave peak, subsequently designated as R_p, prior to optimization. As such it is excluded from the following list of characteristic points whose locations are estimated by the algorithm:

P_o — onset of the P wave as the signal increases from its isoelectric level.

P_p — the peak value of the P wave.

P_f — offset of the P wave as the signal returns to its isoelectric level.

Q_o — onset of the QRS complex.

Q_p — the negative peak of the Q wave, which is not always present.

S_p — the negative peak of the S wave, which is not always present.

R'_p — the typically small peak of the R′ wave, which is not always present.

S_f — offset of the QRS complex.

T_o — onset of the T wave, which is often not discernible.

T_p — the peak value of the T wave.

T_f — offset of the T wave as the signal returns to its isoelectric level.

The entire time-ordered set of characteristic points is defined by

$$\mathcal{C} = \{P_o, P_p, P_f, Q_o, Q_p, S_p, R'_p, S_f, T_o, T_p, T_f\} \tag{4.1}$$

All knots k are completely defined by their times and amplitudes on the observed waveform. The knots used to represent the characteristic points and to estimate their locations are specified as

$$k_i = (t_i, a_i), \quad \text{where} \quad i \in \mathcal{C} \tag{4.2}$$

Figure 4.1 shows a set of manually-annotated points on a representative waveform from the training set. It includes all of the points defined in $\mathcal{C}$, from left to right in the order presented above.

There are a number of ways possible to specify knot times and amplitudes. In the most general sense they can be specified in an absolute manner with respect to a global standard, or relative to a local reference value.

Because the optimization process is performed on a beat-by-beat basis the values t_i need only to be capable of representing knot times within the beat undergoing

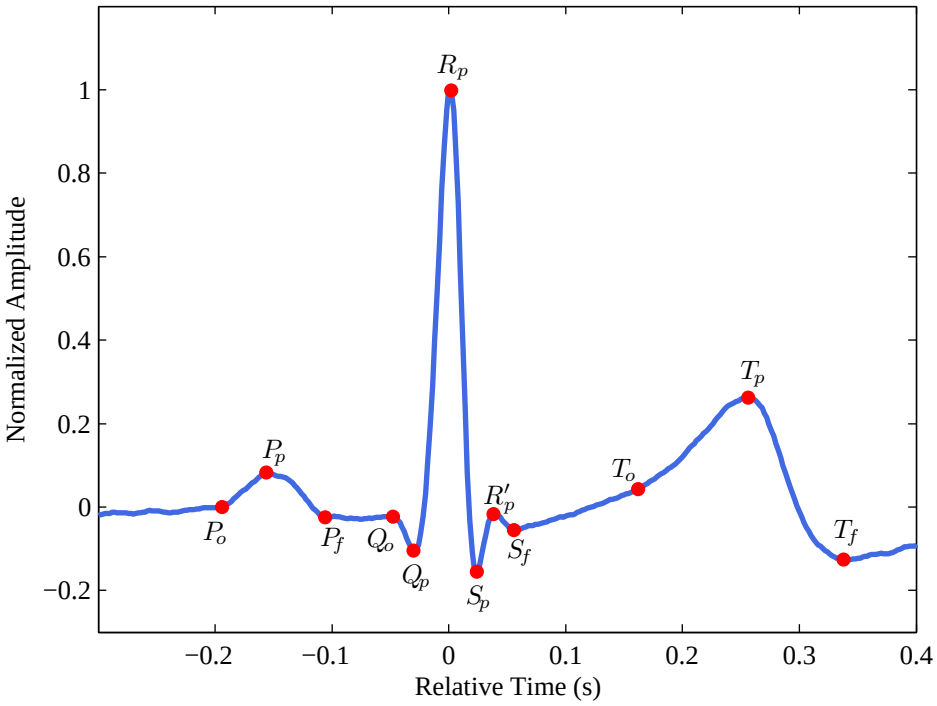


Figure 4.1: A representative beat from the training data set annotated with locations of R_p and manual annotations for all of the characteristic points in $\mathcal{C}$. Times and amplitudes of all points are represented using Equation 4.3.

analysis; there is no need for them to represent times in a more absolute manner that can span multiple beats.

The ECG monitor used to collect the signal provides amplitude values in absolute terms, typically reported in millivolts. The absolute amplitudes of ECG waveform features can be affected by a number of factors even though the waveform morphologies are similar. These factors include inter-subject variability, slight differences in electrode position, intra-subject variability due to changes in health state, and stress level. There is even possibility of beat-to-beat changes due to breathing, as motion of the diaphragm during inspiration and expiration can cause the heart to tilt, changing the cardiac axis and modulating the amplitude of the

QRS complex as seen on the surface ECG. Using the absolute amplitude of the waveform's characteristic points can therefore result in higher variance in estimates of the *a priori* probability density.

These arguments imply that both time and amplitude values should be represented relative to a reference point local to each beat. This reference is the R wave peak because it is generally the most prominent feature of the ECG waveform and defines the location of each complex within the entire signal. R wave locations are found by the beat detection algorithm which is run prior to optimization.

Letting R_p represent the characteristic point for the R wave peak of the beat under analysis, y the observed signal, and t' the time in the absolute reference frame,[1] relative times for that beat are defined as an offset from the time of the R wave peak, and amplitudes are normalized relative to the R wave peak's amplitude. So the t_i and a_i used to define each knot k_i in Equation 4.2 are defined as

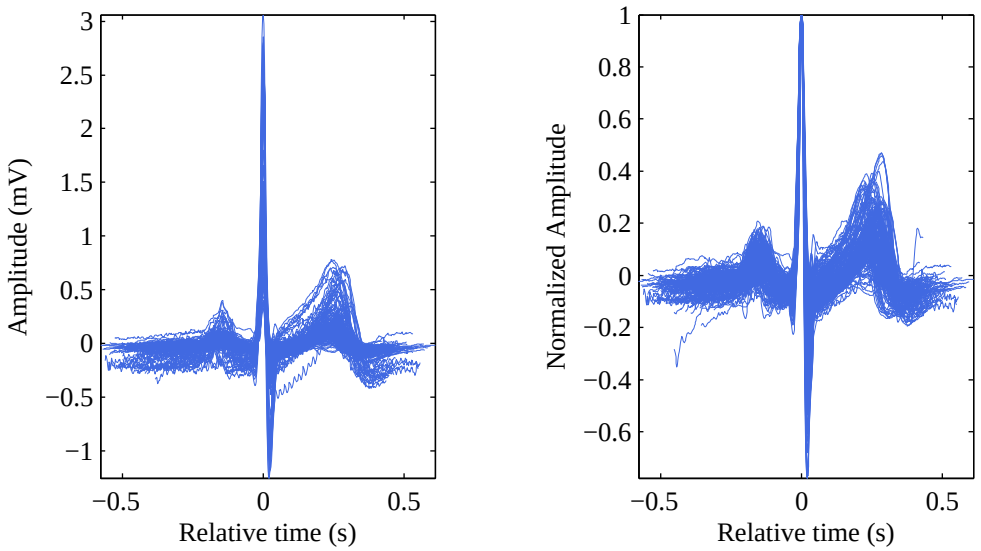
$$t_i = t'_i - t'_{R_p} \tag{4.3a}$$

$$a_i = \frac{y(t'_i)}{y(t'_{R_p})} \tag{4.3b}$$

To depict the variability possible in healthy subjects, Figure 4.2 shows a superimposed display of 200 beats from the training set, aligned in time on the location of the R wave peak. In Figure 4.2a, the amplitude scale is in millivolts, while in Figure 4.2b, values are normalized using the amplitude of the R wave peak as defined by Equation 4.3b.

Table 4.1 shows amplitude statistics for several waves in each beat, computed across 200 beats of the training data. Statistics with subscript y use unnormalized amplitudes, and were calculated using the values of $y(t_i)$ for $i \in \{P_p, Q_p, S_p, T_p\}$. Statistics with subscript a used normalized amplitudes as defined by Equation 4.3.

[1] In this case, the absolute reference frame is with respect to the data collection process. So the amplitudes of y are in millivolts, and times are the sequential sample number as the data are digitized. Absolute times could be with respect to the start of a buffer previously extracted for offline processing, or with respect to the start of data collection in a scenario performing real-time analysis.



(a) (b)

Figure 4.2: An illustration of variability in ECG waveforms. Overlaid plots of 200 beats from the training set, aligned on time of each beat's R wave peak. Amplitude values of the plots in (a) are in millivolts as captured by the ECG monitor. Those of the plots in (b) are normalized by the amplitude of each beat's R wave as defined in Equation 4.3b.

Although there is little difference between mean amplitude values μ_y and μ_a, the standard deviations of the S and T peaks, S_p and T_p are smaller for the normalized data. This reflects a correlation between the strength of cardiac depolarization captured by the R and S wave peaks, and the corresponding repolarization captured by the T wave peak. The reduction in variance from the normalization allows the priors to provide a more effective contribution to the figure of merit.

4.2.2 Support points

In estimating the ECG signal with a linear interpolant, as is necessary for the knot location optimization algorithm, areas of high curvature between characteristic points in $\mathcal{C}$ will have a large error which can adversely affect the optimization algorithm. Adding a number of "support" points to complement the characteristic

Table 4.1: Means and standard deviations of ECG waveform amplitudes for several waves, obtained using 200 beats from the training set. μ_y and σ_y are the statistics for the unnormalized data, expressed in millivolts. μ_a and σ_a are for data normalized using the R wave amplitude of each beat.

$\mathcal{C}$	μ_y	σ_y	μ_a	σ_a
P_p	0.05	0.07	0.05	0.06
Q_p	-0.06	0.07	-0.05	0.05
S_p	-0.27	0.26	-0.24	0.17
T_p	0.17	0.16	0.14	0.09

points around the peaks of the P, R, and T waves in $\mathcal{C}$ allows the linear estimate to more accurately represent the signal in these regions. These points are:

P_1 — support between P wave onset and its peak.

P_2 — support between P wave peak and its offset.

Q_1^R, Q_2^R — two supports between Q wave peak and R wave peak.

R_1^S, R_2^S — two supports between R wave peak and S wave peak.

T_1 — support between T wave onset and its peak.

T_2 — support between T wave peak and its offset.

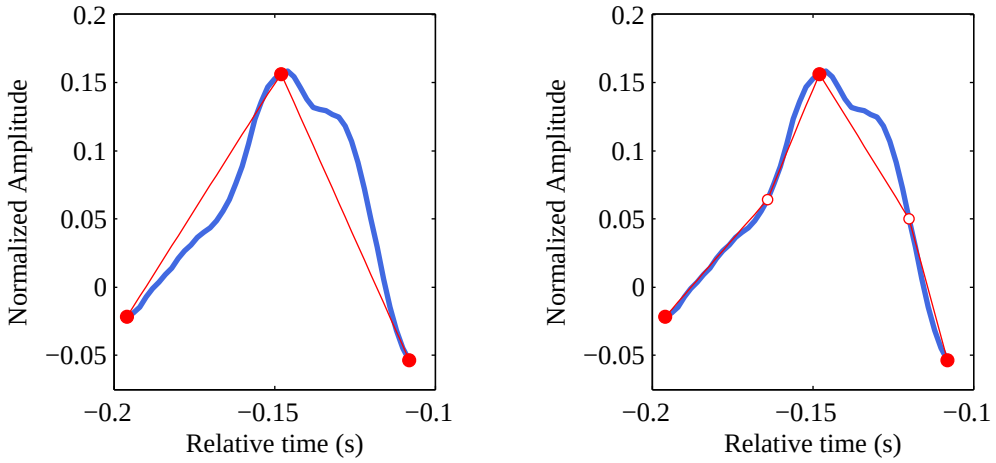
The entire time-ordered set of support knots is defined by

$$\mathcal{S} = \{P_1, P_2, Q_1^R, Q_2^R, R_1^S, R_2^S, T_1, T_2\} \tag{4.4}$$

As with knots in $\mathcal{C}$, those in $\mathcal{S}$ are also completely specified by their times and amplitudes. For the case of support knots, however, the values of t_i and a_i are

calculated using the normalized times and amplitudes of surrounding knots from $\mathcal{C}$, as will be described below.

To help explain the motivation for support knots, Figure 4.3 compares the original signal of a P wave extracted from a beat in the test set against linear estimates created with, and without, the use of support knots. The interpolation in Figure 4.3a relies only on the manually-annotated onset, peak, and offset characteristic points drawn from $\mathcal{C}$, *i.e.*, $\{P_o, P_P, P_f\}$. This linear interpolant clearly has a limited capability to represent the signal in areas of high curvature around the peak.



(a) P wave and its linear estimate without support knots, RMSE 62 μV.

(b) P wave and its linear estimate with support knots, RMSE 30 μV.

Figure 4.3: A comparison of linearly-interpolated estimates of a P wave with and without support knots. Figure (a) illustrates the estimate using only the characteristic point knots P_o, P_p, and P_f drawn from $\mathcal{C}$, all of which are indicated with filled circles. Figure (b) adds the P_1 and P_2 support knots from $\mathcal{S}$ which are indicated with smaller unfilled circles. Use of support knots results in a linear estimate more closely approximating the underlying signal.

Figure 4.3b shows the same P wave and manually-annotated characteristic points from $\mathcal{C}$, but also introduces two additional knots: The P_1 support between

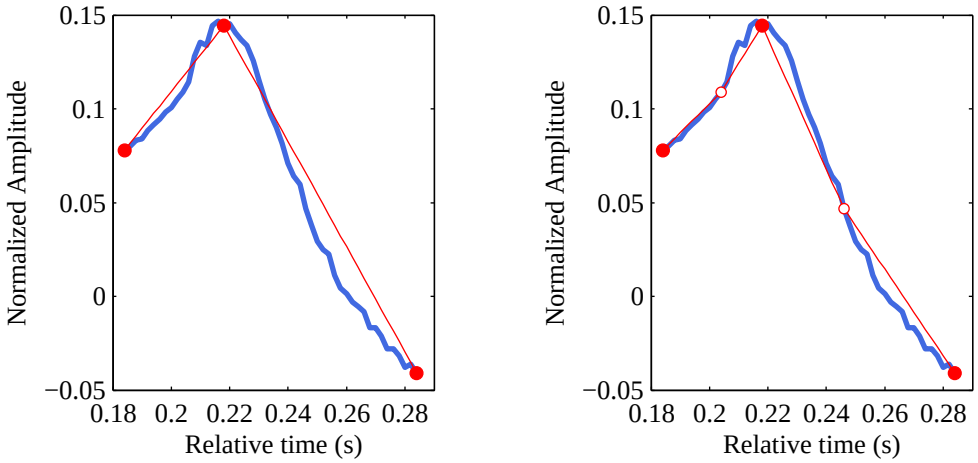
the onset and peak, and the P_2 support between the peak and offset. The linear interpolant using these additional knots represents the underlying signal more closely, with a much smaller RMSE. As previously noted, sole use of RMSE will not determine the best characteristic point locations. However, it does play an important role in the figure of merit that will be described in Section 4.3, which incorporates error in the spline representation as well as prior knowledge to determine optimal location estimates. If the error in the linear spline estimate is too large, the optimization algorithm will favor this term disproportionally to the priors, resulting in poor knot placement.

Figure 4.4 illustrates the same concept for a T wave extracted from another beat in the test set, in this case with the T_1 and T_2 support knots. As with the P wave, incorporating the support knots improves the linear estimate of the signal and results in a smaller RMSE.

Because the P and T waves are not very large in amplitude, one support knot on each side of the peak is sufficient for the linear interpolant to represent the underlying waveform for the optimization process. P and T waves exhibit a wide variety of morphologies, including peaks that are symmetric, asymmetric, sharp, rounded, and "notched". To minimally bias the location of the support knots and allow them to support a variety of morphologies, they are defined as the amplitude midpoints between onsets, peaks, and offsets.

Specifically the amplitude midpoints between the onsets and peaks of the P and T waves define the P_1 and T_1 support knots. And similarly, the amplitude midpoints between the peaks and offsets define the P_2 and T_2 support knots. These are the locations illustrated with the open circles indicating support knots in Figures 4.3 and 4.4.

Figure 4.5 shows a QRS complex extracted from the test set, and its corresponding linear estimates with and without support knots. Note the amplitude scale for the QRS complex is much larger than that of the P and T waves. The



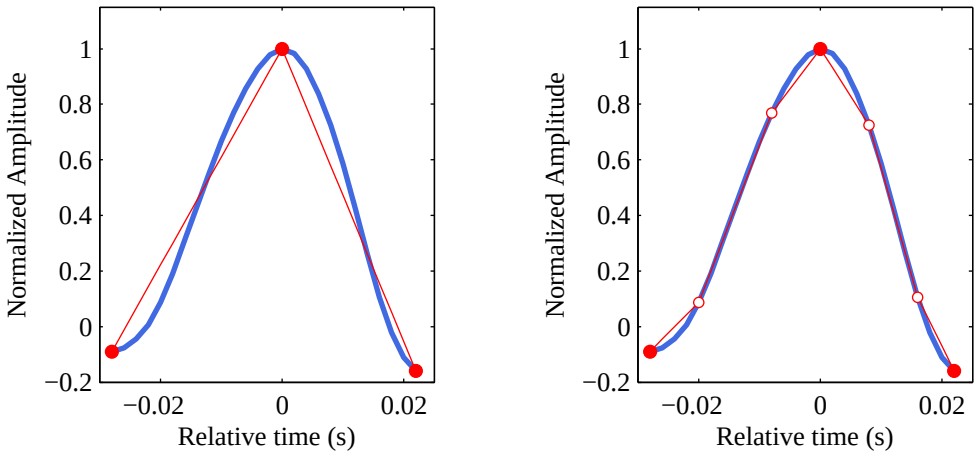
(a) T wave and its linear estimate without support knots, RMSE 9 μV.

(b) T wave and its linear estimate with support knots, RMSE 5 μV.

Figure 4.4: A comparison of linearly-interpolated estimates of a T wave with and without support knots. Figure (a) illustrates the estimate using only the characteristic point knots T_o, T_p, and T_f drawn from $\mathcal{C}$, all of which are shown using filled circles. Figure (b) adds the T_1 and T_2 support knots from $\mathcal{S}$ which are shown using smaller unfilled circles. Use of support knots results in a linear estimate more closely approximating the underlying signal.

ascending (QR slope) and descending (RS slope) segments of the QRS complex are very steep and tall sigmoid shapes with points of inflection approximately midway between the R wave peak and the Q and S peaks.

As with the P and T wave peaks described above, a linear interpolation using only knots from the characteristic point set $\mathcal{C}$ will have a large error, as seen in Figure 4.5a. In this case, however, since the QRS complex is much larger than the P or T waves, two support knots are required on each side of the peak. Use of only one knot at the midpoint of amplitudes as was done with the P and T waves would result in the support knots falling close to the points of inflection of the signal. The points of inflection are very close to where the linear estimate crosses the waveform in Figure 4.5a, so adding support knots at these locations would not significantly improve the linear estimate (*i.e.*, the resulting estimate would be very



(a) QRS complex and its linear estimate without support knots, RMSE 168 μV.

(b) QRS complex and its linear estimate with support knots, RMSE 55 μV.

Figure 4.5: A comparison of linearly-interpolated estimates of a QRS complex with and without support knots. Figure (a) illustrates the estimate using only R_p and the characteristic point knots Q_p and S_p drawn from $\mathcal{C}$, all of which are shown using filled circles. Figure (b) adds support knots from $\mathcal{S}$ to improve the linear estimate. These include two knots, Q_1^R and Q_2^R on the ascending limb of the complex, and two knots, R_1^S and R_2^S on the descending limb. Support knots from $\mathcal{S}$ are shown using smaller unfilled circles.

close to that obtained without supports shown in Figure 4.5a).

To improve the linear estimate, the locations for these four knots should be at the points of maximum curvature surrounding the points of inflection of the sigmoids. A computationally simple, yet effective, approximation can be obtained by calculating two points symmetrically distant from the amplitude midpoint of the QR slope and RS slope.

The optimization algorithm's overall accuracy in estimating characteristic point locations over the entire training set was used to determine the best locations for these support knots. A comparison of accuracies with the support knots placed at one-third and two-thirds of the amplitude range against knots placed at one-fifth and four-fifths of the range resulted in slightly higher accuracies for the latter. So

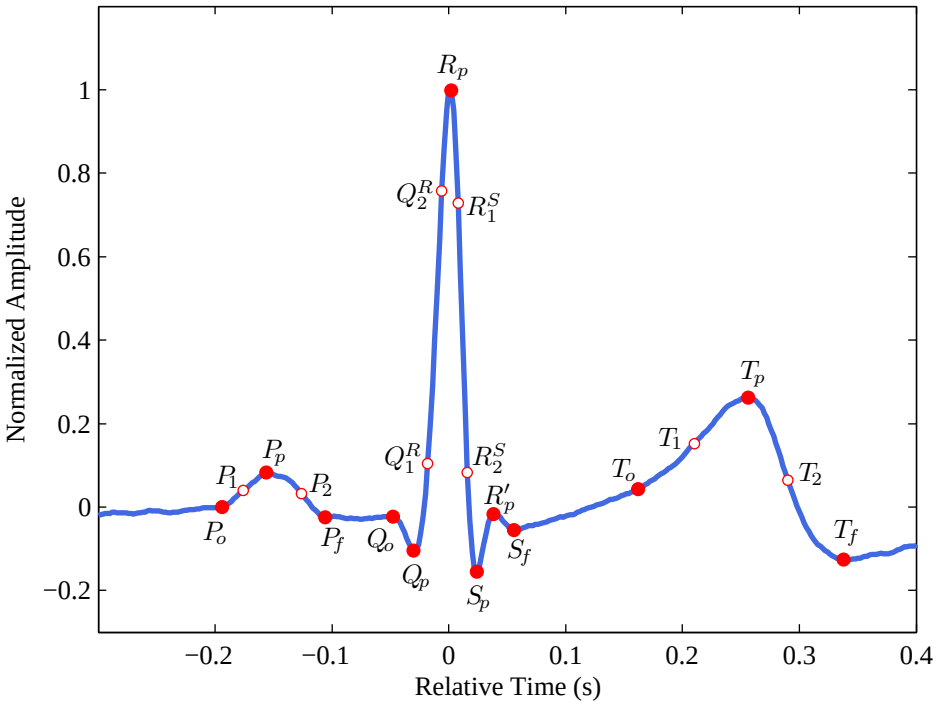


Figure 4.6: A representative beat from the training data set annotated with locations of all knots defined by $\mathcal{K}$. This set includes R_p and all characteristic points in $\mathcal{C}$ indicated with filled circles, and support knots in $\mathcal{S}$ indicated with open circles.

the results on the test set that will be presented in Section 4.8, as well as the support knots shown in Figure 4.5b are at the one-fifth and four-fifth locations.

The entire set of knots, $\mathcal{K}$, is defined as the time-ordered union of all characteristic point knots, support knots, and the R wave peak (which is represented separately since it is determined by the QRS detector prior to optimization). This complete set provides the basis for spline interpolation by the framework:

$$\mathcal{K} = \mathcal{C} \cup \mathcal{S} \cup R_p \tag{4.5}$$

Figure 4.6 shows the same waveform as Figure 4.1, but also includes the support knots calculated as described above.

4.2.3 Motivation for support knots

This section provides a qualitative view of the importance of support knots in the optimization, and gives a representative example to illustrate their benefit. The underlying concepts will be fully described in Section 4.3, which formalizes the terminology and derives the Bayesian figure of merit used by the optimization algorithm.

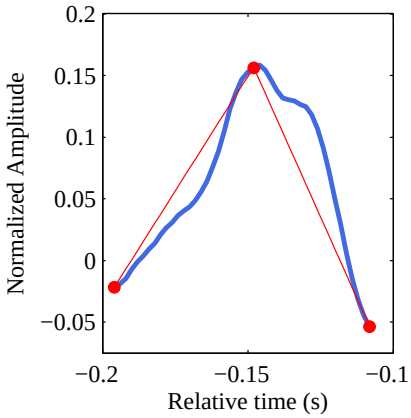
Fundamentally, the Bayesian approach fuses prior knowledge with an observation, in this case to create a figure of merit enabling optimal parameter estimation. The contribution of the observed signal, known as the likelihood, is determined by how well the model governed by the parameters represents the signal. In this case, the parameters are the knots in set $\mathcal{K}$ and the model is the linear interpolant.

Although the interplay of the priors and likelihood is critical to Bayesian optimization, if the error between the model's representation and the observation is sufficiently large, the optimization will be increasingly biased against the priors and favor parameters that improve the likelihood instead of the optimal characteristic point locations. This can cause the knot estimates to move off of their desired locations during optimization.

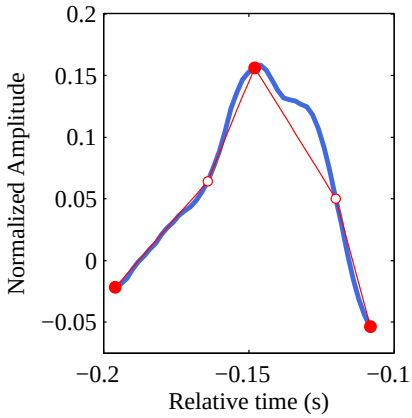
All of the support knots identified in $\mathcal{S}$ serve to reduce this bias. By improving the model's ability to represent the signal in key areas of the waveform that exhibit high curvature, the effect described above is mitigated.

To illustrate how the support knots serve this purpose, Figure 4.7 uses the same P wave signal used in Figure 4.3, which is repeated in the top two figures. The lower two figures, however, show an alternate peak location for the P wave which is indicated by the square marker.

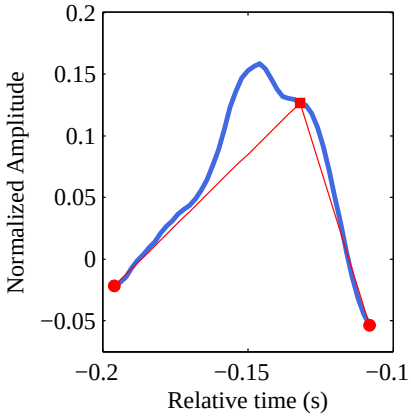
Because the optimization process (to be described in Section 4.7) evaluates the figure of merit for every point in the region between the knots immediately neighboring the knot whose location is being optimized, the location indicated by the square marker will be evaluated as a potential location for the P wave peak.



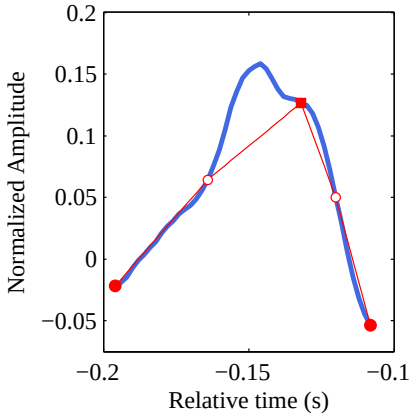
(a) P wave with a knot on the correct peak, and its linear estimate without support knots, RMSE 62 μV.



(b) P wave with a knot on the correct peak,and its linear estimate with support knots, RMSE 30 μV.



(c) P wave with a knot on an incorrect peak, and its linear estimate without support knots, RMSE 60 μV.



(d) P wave with a knot on an incorrect peak and its linear estimate with support knots, RMSE 51 μV.

Figure 4.7: A comparison of linearly-interpolated estimates of a P wave with knots at correct and incorrect peak locations. Circles in figures (a) and (b) show knots at the correct peak location. Square markers in figures (c) and (d) show knots at an incorrect peak location that would be evaluated during optimization. Linear estimates in (a) and (c) were created without support knots. Those in (b) and (d) add the P_1 and P_2 support knots, as indicated with smaller unfilled circles. The best RMSE occurs at the correct peak location when support knots are used.

In the case without support knots, a comparison of the error between the linear estimate at the true peak in Figure 4.7a with that of the incorrect peak in Figure 4.7c shows that the linear estimate is better for the *incorrect* peak with an RMSE of 60 μV, compared to an RMSE of 62 μV for the correct peak.

Adding the support knots has the desired effect. Comparing the error from the estimate using the incorrect peak location shown in Figure 4.7d to that of the estimate using the correct peak locations in Figure 4.7b shows that the RMSE of the estimate using the incorrect peak is greater, at 51 μV, than that of the estimate with the correct peak, which has an RMSE of 30 μV.

By augmenting the set of knots in $\mathcal{C}$ with those in $\mathcal{S}$, the optimization is more likely to report the best figure of merit when knots are at their correct locations, especially when they are in proximity of areas of high curvature in the signal. Without these supports the likelihood component of the Bayesian optimization can dominate the contribution of the priors, and for certain waveform segments, cause the knots to move away from their correct locations.

4.3 FIGURE OF MERIT

The figure of merit is the criterion used by the spline framework described in Chapter 3 to evaluate potential solutions during the optimization process. It is a key component of the framework and its choice must be informed by the purpose of each instance of the framework. In this chapter, the goal is optimal estimation of characteristic point locations of a semiperiodic waveform, while incorporating prior knowledge to improve accuracy. A systematic way to include the use of prior knowledge in the estimation is an important contribution, as existing solutions do not explicitly incorporate priors. This effort represents priors as the *a priori* probability density estimate of the parameters of interest, and they will be described in Section 4.4. It will be shown here that adopting a Bayesian approach for parameter estimation yields an elegant and effective figure of merit that incorporates prior

knowledge, the observed signal, and the effect of model parameters in representing the signal's characteristic points.

The general technique of Bayesian estimation is used for classification in the pattern recognition literature ([20], [85]), and for parametric estimation in the statistical signal processing literature ([40]). It is predicated on modeling the parameter of interest as a random variable with a distribution that is known *a priori*. The goal of the estimation process is then to determine the realization of that random variable resulting in the greatest posterior probability given the observed data.

In this application the model is the linear spline representation of the signal, the model parameters are locations of the knots representing characteristic points $\mathcal{C}$, supports $\mathcal{S}$, and the R wave peak R_p. The signal under analysis is the observed data, and priors are estimated from manual annotations on training data.

The figure of merit is the *a posteriori*, or posterior, probability calculated using Bayes' Theorem. The algorithm determines optimal knot locations by finding the parameters that produce the maximum *a posteriori* probability, or MAP, value. Details on the calculation used to obtain the MAP during the optimization process will be provided in Section 4.6.

Letting k represent the knot locations for all characteristic points in $\mathcal{C}$ as defined in Equation 4.2, and y represent the observed ECG signal, the posterior is given by the conditional probability $p(k|y)$. Using Bayes' Theorem it is defined as

$$p(k|y) = \frac{p(y|k)p(k)}{\sum_k p(y|k)p(k)} = \frac{p(y|k)p(k)}{p(y)} \tag{4.6}$$

Here, $p(k)$ is the *a priori* probability density of the parameters. In the general case priors quantify uncertainty in model parameters before taking the observation into account. For this application they fulfill this role by representing knowledge of knot locations known before the signal is analyzed. They may be obtained from the underlying physiology or estimated from other data sets, as was done in this

effort.

$p(y|k)$, also known as the likelihood, reflects the degree of belief in the observed signal conditioned on knowledge of the model parameters. It is effectively a measure of how well the model, as defined by its current parameters, represents the signal. The likelihood will be derived in Section 4.5.

The term in the denominator of Equation (4.6) is the probability of the observation $p(y)$. It is independent of the parameters k being estimated and serves only as a scaling factor to normalize the posterior probability density. As such it does not affect the comparisons used in maximizing the figure of merit and can be eliminated from the equation. Doing so produces a figure of merit represented simply by the product of the likelihood and the priors

$$p(k|y) \propto p(y|k)p(k) \tag{4.7}$$

Maximizing this figure of merit over the vector of knots k for a given signal y provides the best knot locations corresponding to the optimal estimates of characteristic points $\mathcal{C}$ of the ECG complex.

Application of Bayes' Theorem to define the *a posteriori* probability, and computation of the MAP to determine the best knot locations, provide a systematic way to balance the tradeoff between prior knowledge $p(k)$, the observation y, and parameter estimate k.

4.4 PRIOR PROBABILITIES

The algorithm uses prior knowledge captured as estimates of the *a priori* probability density of the characteristic points of interest in two ways. First, it uses statistics of characteristic point locations in time to initialize knot locations prior to the optimization. These *location priors* are described in Section 4.4.1.

It also uses priors as part of the optimization process. Incorporating the *a priori* probabilities in the optimization biases the solution towards the prior knowledge,

helping to reduce variability of the estimates and improve accuracy. This is particularly helpful in waveforms in which one or more points are ambiguous, there is noise, or a particular feature is absent. In this case a more complete set of priors serve to capture additional information regarding the waveform's characteristic points, and are specified by a joint probability density estimate of each point's time and curvature. These *time-relevance* priors are described in Section 4.4.4.

The *a priori* probability density estimates were calculated using manual annotations of all characteristic points in $\mathcal{C}$ on a subset of the data that was used for training, as described in Appendix B.

The training set comprises 400 randomly-selected beats from 40 randomly-selected subjects. The first set of 200 beats was manually annotated by one reviewer, and the second set of 200 beats was manually annotated by two reviewers. In total this provides up to 600 potential manual annotations of points in $\mathcal{C}$, although due to noise, ambiguity, and missing features in the training data there are fewer than 600 annotations for all points. The annotated locations were processed as described in the following sections to obtain an estimate of the *a priori* probability densities for each characteristic point.

4.4.1 Location priors

This section describes statistics computed from manual annotations which are used for initializing knot locations before invoking the optimization algorithm. These are named *location priors* to distinguish them from the priors used for optimization that are described in Section 4.4.4.

The probabilistic representation of the time and amplitude components of each characteristic point in $\mathcal{C}$ models them as being normally distributed based on the central limit theorem. This allows fully specifying the location prior of each characteristic point by its mean and variance as obtained from manual annotations of the training data.

Table 4.2 shows the mean and standard deviation for the time and amplitude components of all of the characteristic points in $\mathcal{C}$, along with the number of annotated points used in their calculation. Counts n with values of less than 600 reflect characteristic points that were not labeled by one or both reviewers.

Figure 4.8 illustrates the location priors graphically against a representative waveform from the training data set. The rectangles are centered on the mean time and amplitude values. Their width and height are two standard deviations in time and amplitude, respectively, effectively bounding each prior in time with $\mu_t \pm \sigma_t$, and in amplitude with $\mu_a \pm \sigma_a$.

Since the knot locations used to calculate the priors are specified with respect to the R wave peak as defined by Equation (4.3), its mean time and amplitude are exactly zero and one, respectively, and variances of both time and amplitude are zero. For this reason its bounding box is not shown in the figure.

The characteristic points that are part of the QRS complex, Q_o through S_f, exhibit relatively small standard deviations in time, indicating less temporal variability in the vicinity of the R wave during ventricular depolarization. In addition, the sharp curvatures defining these features makes it easier for reviewers to place manual annotations that are consistent with each other. However, several of these points show large amplitude variance reflecting a wide range of amplitudes for the corresponding characteristic point in the training set. For example, depending on subject and lead the S wave peak may be very small, or be present with great amplitude in what is commonly called a biphasic waveform. The Q wave and R′ peaks exhibit a similarly large amplitude variances for the same reason, although their maximum amplitude is generally far less than that of the S wave.

Conversely, the P and T wave characteristic points have much larger variances in the time dimension than those of the QRS complex. Because these features are more bandlimited, it is more difficult for different reviewers to precisely and consistently place the annotations since they can lack sharp demarcations. Furthermore,

Table 4.2: Means and standard deviations of times (μ_t, σ_t) and amplitudes (μ_a, σ_a) of characteristic points in $\mathcal{C}$ for the entire training set. As defined by Equation (4.3), times are in milliseconds offset from the R wave peak and amplitudes are relative to the R wave peak. n is the number of beats in the training set which were manually annotated for each characteristic point. There were a total of 400 beats in the training set; each beat was annotated by at least one of the reviewers, for a maximum possible count of 600. Some of the characteristic points may not have had manual annotations due to noise or lack of the corresponding characteristic point in the ECG signal.

$\mathcal{C}$	μ_t	σ_t	μ_a	σ_a	n
P_o	-190.9	25.4	-0.01	0.03	590
P_p	-141.9	21.9	0.09	0.05	596
P_f	-94.8	21.1	-0.04	0.04	593
Q_o	-39.1	7.2	-0.04	0.03	597
Q_p	-30.0	3.0	-0.10	0.14	393
S_p	24.4	6.1	-0.30	0.21	585
R'_p	39.2	7.4	-0.03	0.05	363
S_f	51.2	9.0	-0.06	0.05	595
T_o	162.8	37.9	0.03	0.07	225
T_p	254.5	27.3	0.20	0.10	563
T_f	329.2	29.1	-0.08	0.06	536

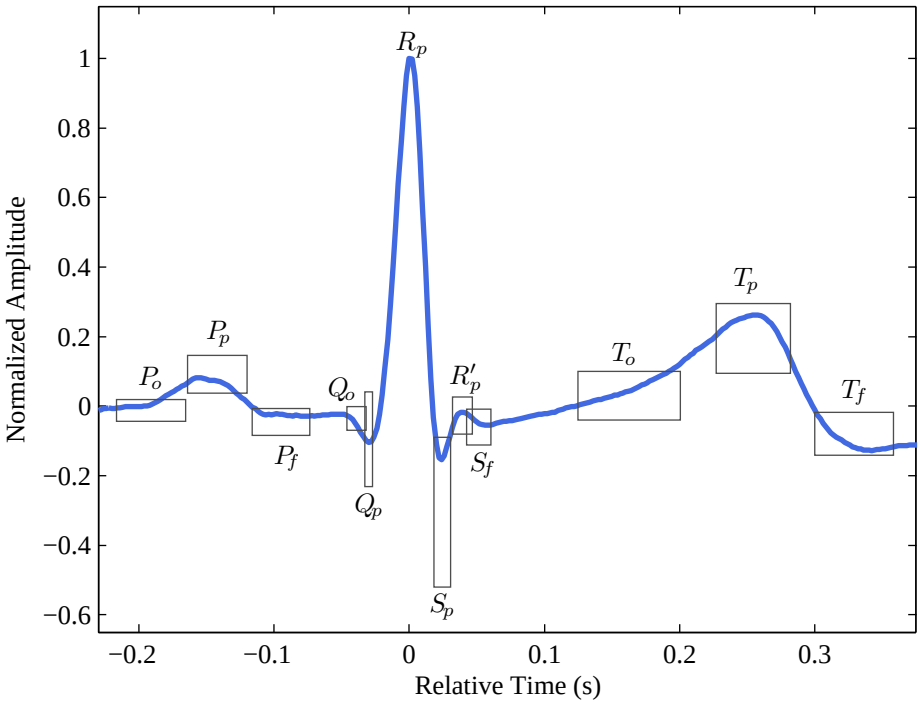


Figure 4.8: A representative beat from the training data set annotated with rectangles representing the location priors for characteristic points in $\mathcal{C}$. These correspond to values shown in Table 4.2. Rectangles are centered at the time and amplitude means of the joint time-amplitude density estimates and are two standard deviations in each dimension; each rectangle is bounded by $\mu \pm \sigma$ in time and amplitude.

the smaller amplitudes of the P and T waves makes them more susceptible to noise, also increasing the variance of manual annotations.

4.4.2 Augmenting manual annotations

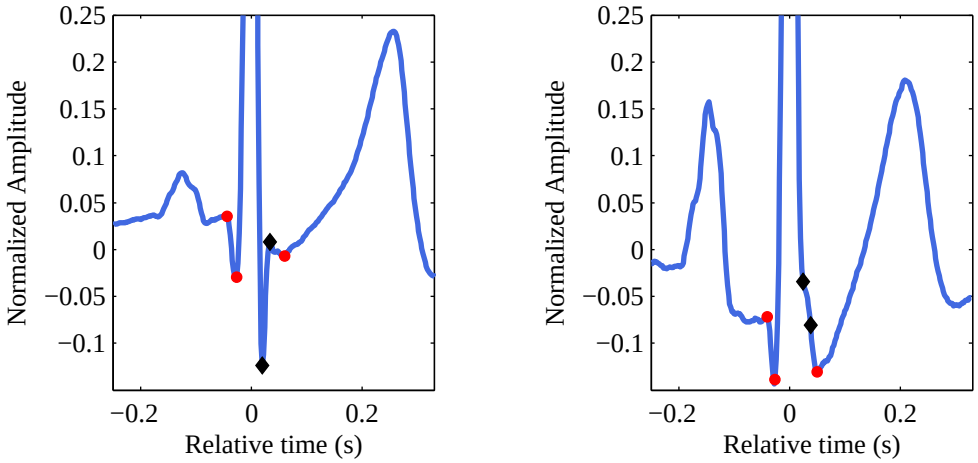
Certain characteristic points may not have been annotated by the reviewers for some QRS complexes in the training set, as evidenced by counts n less than 600 in Table 4.2.

These missing manual annotations fall into two classes. The first class includes

characteristic points that are present in all waveforms but which may not be observable due to low signal amplitude for a particular subject and lead, or because noise on the ECG signal obscures the point's location. While the ability of a reviewer to annotate virtually any characteristic point can be adversely impacted by the presence of noise and low amplitude, the annotations most affected in this manner are the onset, peak, and offset of the P wave, and the peak and offset of the T wave. Despite their relatively low amplitude and lack of sharp features, all of these points have over 500 manual annotations each, and are well represented for use as priors.

The second class of missing manual annotations is more challenging, comprising other characteristic points that may simply not be present in the given lead for a subject. These include the Q wave peak, R′ wave peak, and the T wave onset. The missing peaks in this class are evidenced by lack of a pronounced local maximum or minimum at the expected location of the waveform. A missing T wave onset is the most common missing characteristic point with only 225 manual annotations. It occurs when there is a highly linear transition from the end of the QRS complex leading into the T wave peak. In this case there is no discernible change in curvature corresponding to the start of the T wave that would serve to designate an onset point.

Since the algorithm described in this chapter estimates the location of *all* characteristic points in $\mathcal{C}$, the knots corresponding to the second class of potentially-missing characteristic points will be present and used in the interpolation and optimization even if the underlying features they represent are not in the waveform being analyzed. To illustrate how missing characteristic points in the underlying waveform affect knot location, Figure 4.9 shows two QRS complexes from the test set on an expanded time and amplitude scale. Markers on these figures indicate the onset and offset times of the QRS complex, and peaks of the Q wave, S wave, and R′ wave.



(a) QRS complex exhibiting a Q wave, S wave, and R′ wave. From left to right, the markers indicate the QRS onset, Q wave peak, S wave peak, R′ wave peak, and QRS offset.

(b) QRS complex without S or R′ wave peaks. In this case the optimization process has moved the corresponding knots, indicated by diamond-shaped markers, to an appropriate location on the waveform.

Figure 4.9: An illustration of knot locations determined by the optimization algorithm for features that may not be present in an ECG waveform. The diamond-shaped markers in (b) correspond to the S and R′ peaks, which do not exist in that waveform and, if in the training set, would not have been manually annotated.

The example in Figure 4.9a has distinct characteristic points corresponding the S and R′ wave peaks indicated by the third and fourth markers. The complex shown in Figure 4.9b, however, does not have peaks for the S or R′ waves. In this example the corresponding knots, indicated with filled diamond-shaped markers, have moved to an appropriate location on the waveform as determined by the figure of merit during the optimization process.

Since the figure of merit relies on accurate priors, the locations of these characteristic points must be represented in the training set even for beats that don't exhibit these features, and as a result, have no corresponding manual annotations.

To address this contingency, prior to computing the prior probability density estimates used for optimization, the manual annotations need to be augmented. Knots corresponding to missing missing Q wave and R′ peaks must be placed on the ascending QR slope or ascending S slope respectively. And knots corresponding to a missing T wave onset must be placed between the QRS offset and the T wave peak.

To more closely match the the figure of merit during optimization, the point in the segment resulting in the best likelihood value determines the locations for these augmented knots. Performing this important step before calculating the priors used for the optimization ensures that they reflect both the presence and absence of the characteristic point, and as will be seen in Section 4.4.4, result in a bimodal *a priori* probability density estimate for these, and in some cases their immediately neighboring, points.

4.4.3 Knot relevance

The location priors described above reflect previous knowledge of knot location as given by the times and corresponding amplitudes of the manually-annotated characteristic points. However, additional information exists in the signal, specifically regarding the relevance, or importance, of each knot in representing the signal with a linear interpolant.

Some knots, such as the R wave peak, are critical to the spline representation of the waveform because they indicate a point of great curvature reflecting an abrupt change in the underlying physiology. Other points, such as the T wave onset, often cannot be distinctly identified because they occur when the waveform is nearly linear without a clear change in slope or point of high curvature.

Mathematically, the curvature of a smooth plane curve at an arbitrary point is defined as the rate of change of the tangent to the curve at that point ([80], [37]). Since in this implementation of the spline framework the signal is represented by

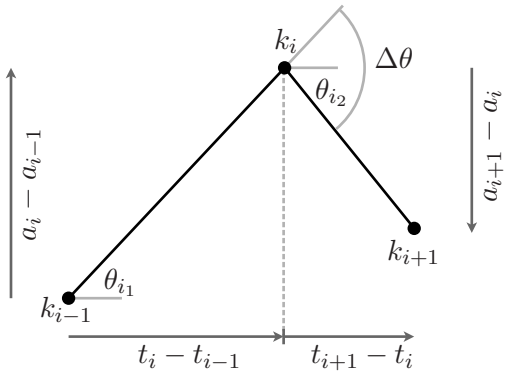


Figure 4.10: Relevance calculation for knot k_i. With typical values for the R wave peak, R_p, Equation (4.8) results in $\theta_{i_1} = \tan^{-1}(0.33/0.01) = 1.54$ and $\theta_{i_2} = \tan^{-1}(-0.23/0.006) = -1.55$. Using these in Equation (4.9) provides the relevance $\rho = (1.54 + 1.55)/\pi = 0.98$, indicating a very sharp concave down peak, as expected for R_p.

linear segments bounded by the knots in $\mathcal{K}$, the tangent to the interpolated signal estimate is not defined at any knot joining non-collinear segments.

Noting that the slopes of the linear segments used to estimate the signal can only change at knot locations suggests a straightforward measure of curvature. It is a special case of the more general definition provided above and is simply stated as the normalized angular change of the line segments surrounding the knot under consideration. This *knot relevance* parameter is designated by ρ, and for each knot $k_i \in \mathcal{C}$ is calculated using the times and amplitudes of the immediately preceding and following knots, k_{i-1} and k_{i+1}. The first and last knots in $\mathcal{C}$ use the signal segment's start and end points as their preceding and following knots, respectively. Relevance values for the start and end points are not defined — nor are they needed — as the segment end points are not represented in $\mathcal{C}$.

Figure 4.10 illustrates the relevance calculation for a typical R wave peak knot drawn from a beat in the training set. First, it is necessary to calculate the angles

of the line segments preceding and following the knot under consideration

$$\theta_{i_1} = \tan^{-1} \frac{a_i - a_{i-1}}{t_i - t_{i-1}} \tag{4.8a}$$

$$\theta_{i_2} = \tan^{-1} \frac{a_{i+1} - a_i}{t_{i+1} - t_i} \tag{4.8b}$$

Then ρ_i is obtained by computing the normalized difference of the angles

$$\rho_i = \frac{\theta_{i_1} - \theta_{i_2}}{\pi} = \frac{\Delta\theta}{\pi} \tag{4.9}$$

The relevance value computed in this manner is bounded by $-1.0 \leq \rho \leq 1.0$. A relevance of 0.0 corresponds to a knot on an exactly-linear segment of the waveform where $\theta_{i_1} = \theta_{i_2}$; a relevance approaching +1.0 indicates a knot on an extremely abrupt, rapid transition that corresponds to a concave down, or positive, peak such as that of an R or R′ wave. A relevance approaching −1.0 indicates a similarly abrupt knot on a concave up, or negative, peak such as a Q or S wave. This metric quantifies not only the shapes of peaks in the waveform, but can also provide valuable information regarding curvatures of waveform onset and offsets.

Information about the curvature of the waveform can indicate the importance of any knot to the waveform's spline representation. Since the algorithm described in this chapter optimizes the location of all knots in $\mathcal{C}$, even if the underlying characteristic point does not exist for a particular waveform, after optimization the relevance value can be used determine whether or not the characteristic point corresponding to a particular knot is actually present. For example, the Q_p knot for a waveform that does not have a Q wave peak will have very low relevance, whereas for a waveform that exhibits the peak it will have a high relevance.

The more important application of the knot relevance, however, is its use as a component of the *a priori* probability density estimate to help identify the best knot locations during Bayesian optimization.

By augmenting the description of each characteristic point estimated by the spline framework with prior knowledge of its relevance, the figure of merit operates on more information, resulting in improved knot location estimates. Table 4.3

shows statistics of the relevance values calculated for all manually-annotated characteristic points in the training set, and can be compared against Figure 4.1 which shows manual annotations of all characteristic points $\mathcal{C}$ on a representative beat.

The relevance values of the knots representing peaks of the P and T waves have mean values of 0.69 and 0.78 respectively, indicating a moderate curvature. This is consistent with expectations for these peaks since they are usually rounded. Sharp positive peaks such as those of the R and R′ waves have greater mean relevance values that approach +1.0, and sharp negative peaks such as those of the Q and S waves have mean values near −1.0. The mean relevance for onsets and offsets of all waves are much lower than those of the peaks, indicating mild curvature.

The relevance value provides an additional benefit to characteristic points that may not be present in a given waveform as described in Section 4.4.2. When the peaks of the Q and R′ waves are not present, the relevance values are close to zero since the corresponding knot locations tend to fall on signal segments that are relatively linear, as illustrated in Figure 4.9b. For these cases, the probability density estimate of relevance values can can be modeled with multiple modes and provide a more complete representation of prior knowledge for the optimization. One mode is near a relevance value of zero corresponding to missing characteristic points, and the other mode is at the relevance value indicated by the manual annotations.

4.4.4 Time-relevance priors

The approach described in Section 4.4.1 calculates location priors for characteristic points using manual annotations on the training set. However, the location priors in isolation do not incorporate the knot relevance value ρ of each characteristic point which provides valuable *a priori* information about the curvature of the points being estimated. Ideally, the prior used for optimization should incorporate time, amplitude, and relevance information.

Table 4.3: Means and standard deviations of knot relevance values ρ calculated for R_p and characteristic points in $\mathcal{C}$ using manual annotations on the training set. n is the number of beats in the training set for which ρ could be calculated. The characteristic point itself, as well as required surrounding points, must all be present.

$\mathcal{C} \cup R_p$	μ_ρ	σ_ρ	n
P_o	-0.34	0.10	587
P_p	0.69	0.16	583
P_f	-0.43	0.14	588
Q_o	0.43	0.13	387
Q_p	-0.88	0.05	391
R_p	0.98	0.00	378
S_p	-0.95	0.03	363
R'_p	0.86	0.09	358
S_f	-0.56	0.12	101
T_o	-0.18	0.09	224
T_p	0.78	0.07	219
T_f	-0.47	0.13	530

An effective estimate that captures the time, amplitude, and relevance of each knot is the joint probability density of the knot's time with respect to its corresponding R wave peak and its relevance value ρ. Incorporating the relevance directly into a joint density avoids the more problematic approach of building the prior in a higher dimension that would have to include the times and amplitudes of each three-tuple of knots, since for every knot the time and amplitude of the knots immediately preceding and following it are required to compute relevance.

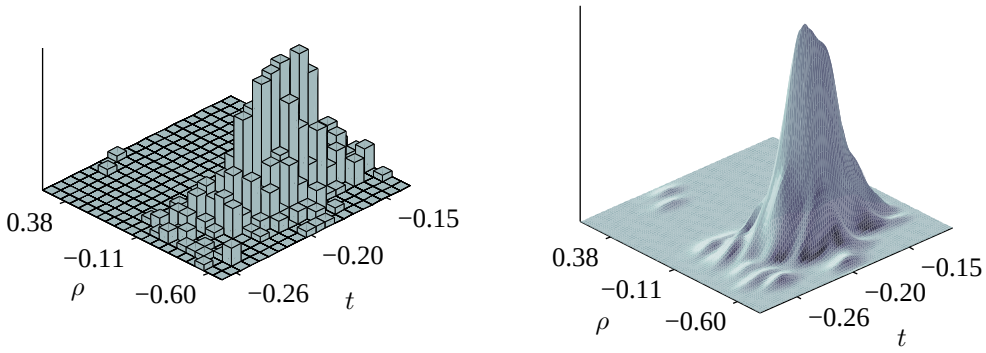
The joint probability density of the prior information is therefore represented simply as $p(t, \rho)$ and estimated from directly from the training set using the time and relevance values for every manually-annotated characteristic point in $\mathcal{C}$.

Although a histogram provides a reasonable estimate of the underlying probability density, it lacks continuity which is desirable when employing the joint probability density estimate in the optimization process. Specifically, a small change in time or relevance should be reflected as a commensurate change in the prior probability, otherwise the priors may have an incorrect effect on the optimization.

One possible solution is to increase the granularity of the histogram by reducing its bin sizes. Although this will provide increased smoothness the resulting growth in number of bins will require much more training data. Otherwise, as the number of bins increases the count per bin will decrease, limiting the histogram's ability to create a meaningful estimate of the underlying probability density.

Using a kernel density estimation (KDE) addresses this shortcoming. KDE is a non-parametric technique that provides a smooth density estimate by summing continuous kernel functions according to the underlying data [107]. The width of the kernel function, called the bandwidth, determines the degree of smoothing provided by the estimator. There are a number of openly-available implementations; this effort used the `gkde2` Bivariate Kernel Density Estimator [14], which incorporates an effective automatic bandwidth selection algorithm. The resultant density estimate provides prior values that can easily be indexed with a given knot's time

and relevance to calculate the figure of merit during optimization.

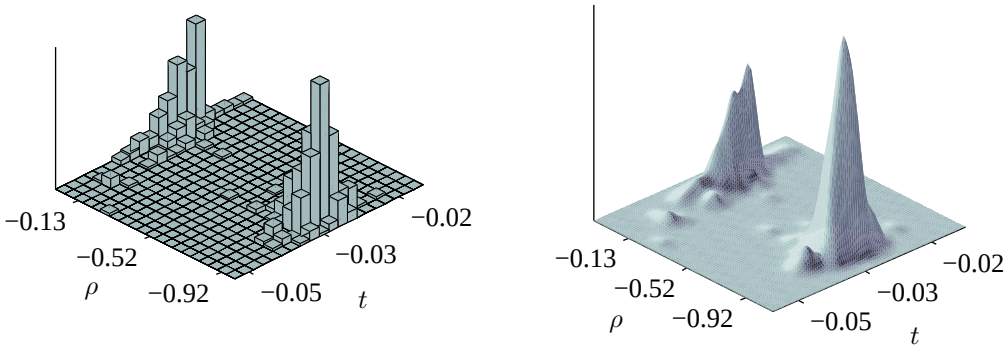


(a) P_o histogram. Limited granularity of the histogram due to bin size impacts its use for optimization.

(b) P_o KDE, reflecting the histogram but with much greater granularity to provide a more continuous basis for estimating values at any given time and relevance.

Figure 4.11: Histogram and KDE estimating the joint time-relevance prior probability density of the P wave onset characteristic point P_o, obtained using using manual annotations on the training set. t is the time of the characteristic point in seconds with respect to the R wave peak, and ρ is its relevance.

Figure 4.11 illustrates the histogram and corresponding KDE of the P wave onset characteristic point. The KDE for the P wave onset characteristic point in Figure 4.11b reflects a smoothed version of the histogram in Figure 4.11a. This characteristic point is present in all of the training data, even though it may not be annotated for every waveform due to noise or ambiguity. The density estimate is unimodal with peaks in both dimensions reflecting values for the P_o shown in Table 4.2 for time, and Table 4.3 for relevance: the mode in the time dimension is approximately 200 milliseconds prior to the R wave, and the mode in the relevance dimension is at $\rho \approx -0.3$, indicating a mild curvature corresponding to the beginning of the P wave.



(a) Q_p histogram.

(b) Q_p KDE.

Figure 4.12: Histogram and KDE estimating the joint time-relevance prior probability density of the Q wave peak characteristic point Q_p, obtained using using manual annotations on the training set. The bimodal nature reflects the fact that this characteristic point may not be present in some waveforms: the peak at $\rho \approx -0.9$ corresponds to sharp concave up curvature of a Q wave, while the peak at $\rho \approx 0.05$ reflects lack of this feature, with the corresponding knot on a linear part of the QR slope.

Figure 4.12 shows the histogram and KDE of the Q wave peak. Since the Q wave peak is one of the characteristic points that may not be present in some waveforms, there are cases where no manual annotation can exist. For signals in the training set where this the case the preprocessing algorithm described in Section 4.4 adds a knot at an appropriate location following the Q wave onset, which typically falls on a relatively linear part of the waveform.

As a result this density estimate has one mode peaking at $\rho \approx -0.9$ and indicating a very sharp concave up waveform corresponding to presence of a Q wave peak, and another mode at $\rho \approx 0.05$ corresponding to complexes in the training set that did not have a Q wave peak and whose corresponding knot was placed on the ascending QR slope of the complex. The relevance value of the first mode is consistent with the mean relevance for Q_p in Table 4.3; that of the second mode

is not reflected in the table, as the statistics presented there do not incorporate augmented points.

In time, both modes of this estimate occur at approximately 30 milliseconds prior to the R wave peak, again consistent with the corresponding location prior in Table 4.2. Complementing the training set of characteristic points allows the prior to contribute properly when optimizing knot locations for potentially missing features.

Figures showing the histograms and KDE estimates of the *a priori* time-relevance probability densities for all characteristic points defined by $\mathcal{C}$ are shown in Appendix A.

4.5 LIKELIHOOD

The likelihood was introduced as a component of the Bayesian figure of merit in Section 4.3, and was described qualitatively as the degree of belief in the observed signal given the models parameters, represented as $p(y|k)$. This value provides an effective measure of how well the model represents the observation with its current parameters.

To derive the likelihood, the observed signal is modeled as the sum of its spline estimate and a noise term. The noise term represents the error between the linear interpolation and the signal, and is assumed to be an uncorrelated stochastic process described by a zero-mean normal distribution.

With y representing the original signal, $\hat{y} = f(k)$ the interpolated spline estimate, and ϵ the error, the model is given by

$$y = \hat{y} + \epsilon \tag{4.10}$$

Solving for the error term yields

$$\epsilon = y - \hat{y} \tag{4.11}$$

which is assumed to be normally distributed with zero mean and variance of σ_ϵ^2, represented as $\mathcal{N}(0, \sigma_\epsilon^2)$. Now the likelihood can be expressed as

$$p(y|k) = p_\epsilon(y - \hat{y}|k) \tag{4.12}$$

The vector of likelihoods expressed in this manner is obtained by calculating values of the normal probability density $\mathcal{N}(0, \sigma_\epsilon^2)$ for each element comprising the error vector computed at every point of the observed signal estimated by the spline interpolant.

The optimization process described in Section 4.7 operates in succession on the characteristic point locations specified in $\mathcal{C}$, determining each one's optimal location between the knots immediately preceding and following it in the full set $\mathcal{K}$. As uncorrelated, normally-distributed processes are independent, the overall likelihood value corresponding to a set of knots is computed as the product of m components, one for each sample of the waveform being estimated using spline interpolation on the current three-tuple of knots. Starting with Equation (4.12), and applying the assumption of independence,

$$\begin{aligned} p(y|k) &= p_\epsilon(y - \hat{y}|k) && \text{(4.13a)} \\ &= p_\epsilon(\epsilon_1, \ldots, \epsilon_m) && \text{(4.13b)} \\ &= \prod_{j=1}^{m} p_\epsilon(\epsilon_j) && \text{(4.13c)} \end{aligned}$$

Where j is the index indicating the location at which the error between the observed signal and its interpolated estimate is calculated, and m indicates the number of samples spanned by the three-tuple of knots, the center knot of which is being optimized. So the likelihood is given by the product of individual values of the normal density function $\mathcal{N}(0, \sigma_\epsilon^2)$ at each error value ϵ_j.

This value reflects how well the model, a linear interpolant governed by its parameters k_i, represents the observed signal y. The smaller the error across the entire signal, the closer the ϵ is to zero. Since the mode of a zero-mean normal

distribution is at zero, the cumulative probability defining the likelihood will have its maximum value when the signal is perfectly estimated by the model.

The noise variance σ_ϵ^2 specifies the spread of the error distribution, which in turn governs the "responsiveness" of the likelihood term to errors in the estimate. Smaller values of σ_ϵ^2 result in a distribution with lower variance, providing a sharper response; *i.e.*, a change in the underlying error will cause a larger degradation of the likelihood. Larger values of σ_ϵ^2 model a distribution with greater variance, producing a smaller response in likelihood for a given change in error.

A systematic comparison of optimization accuracies on the training data determined the best value of noise variance to be $\sigma_\epsilon^2 = 0.001$. The optimization to assess algorithm performance on the prospective test set used this value, with results reported in Section 4.8.

4.6 CALCULATING THE POSTERIOR

As described in Section 4.3, the figure of merit used in the optimization is the posterior probability of the parameters conditioned on the observation, defined using Equation (4.7) as $p(k|y) = p(y|k)p(k)$.

The optimization process determines best placement by finding the knot locations that produce the best figure of merit. Knots resulting in the maximum *a posteriori* probability are expressed mathematically as

$$k_{\text{MAP}} = \underset{k}{\operatorname{argmax}}\, p(k|y) \tag{4.14a}$$

$$= \underset{k}{\operatorname{argmax}}\, p(y|k)p(k) \tag{4.14b}$$

Substituting in the likelihood defined in Equation (4.13c) and the joint *a priori* density estimate $p(k) = p(t, \rho)$ allows expressing the MAP as the cumulative product

$$k_{\text{MAP}} = \underset{k}{\operatorname{argmax}}\; p(t, \rho) \prod_{j=1}^{m} p_\epsilon(\epsilon_j) \tag{4.15}$$

In practice, to prevent issues with numerical computation on values that can span a large range, the logarithm of the constituent likelihood and priors is used, converting the product above to a summation. So the optimal knot locations are given by

$$k_{\text{MAP}} = \underset{k}{\text{argmax}} \log \left(p(t, \rho) \prod_{j=1}^{m} p_{\epsilon}(\epsilon_j) \right) \tag{4.16a}$$

$$= \underset{k}{\text{argmax}} \left(\log p(t, \rho) + \sum_{j=1}^{m} \log p_{\epsilon}(e_j) \right) \tag{4.16b}$$

The optimization algorithm described in Section 4.7 searches for the best knot location for the characteristic point under consideration by finding the maximum *a posteriori* estimate using Equation (4.16b).

4.7 OPTIMIZATION

The optimization algorithm described here finds the best knot locations for a given beat using the MAP calculation outlined above. It assumes that for each beat under analysis the beat detector has located the R wave peak, and that the ECG signal comprising the beat under consideration has been extracted. In addition, a preprocessing step converts the times and amplitudes of the waveform samples to be relative to the time and amplitude of the R wave peak as defined by Equation (4.3).

The first step of the optimization is to place all knots $\mathcal{K}$ representing both characteristic points and support locations on their initial locations. The optimization algorithm will modify their locations based on the figure of merit in its search for the best knot locations representing the signal's characteristic points.

The algorithm determines initial knot locations, prior to optimization, using the R wave location for each beat obtained from the beat detector, and the location prior probabilities derived from the training data as described in Section 4.4.1. Referring to the set of knots representing the characteristic points $\mathcal{C}$

described in Equation (4.1), the subset of knots corresponding to the QRS complex $\{Q_o, Q_p, R_p, S_p, R'_p, S_f\}$ are initialized to the means of the time priors shown in Table 4.2. Initial knot amplitudes are set to the waveform amplitudes at those times.

The higher variability of P and T wave locations with respect to the R wave necessitates more care in their initial placement to reduce the chance of the optimization algorithm finding local optima for the characteristic points belonging to these waves. The knot initialization algorithm makes a preliminary estimate of the locations of these peaks by searching for a local maximum amplitude value in the time range bounded by $\mu_t \pm \sigma_t$ for P_p and T_p, as defined in Table 4.2. It then initializes the onset and offset knots given by the set $\{P_o, P_f, T_o, T_f\}$ by translating their location time priors relative to the preliminary peak locations. Knot amplitudes are again initialized to values of the waveform at these times.

The next step is initialization of the support knots $\mathcal{S}$ in Equation (4.4) defined mathematically in Section 4.2. The knot initialization algorithm places the P wave peak support knot P_1 at the midpoint of amplitudes of $\{P_o, P_p\}$, and the support knot P_2 at the midpoint of amplitudes of $\{P_p, P_f\}$. It similarly places the T wave support knots T_1 and T_2 at amplitude midpoints of $\{T_o, T_P\}$ and $\{T_p, T_f\}$, respectively.

The remaining support knots include two knots on the ascending slope of the QRS complex $\{Q_1^R, Q_2^R\}$, and two on the descending slope $\{R_1^S, R_2^S\}$. As defined previously, the knot initialization algorithm places these knots at one-fifth and four-fifths of the amplitude difference between Q_p and R_p for the QR support knots, and at one-fifth and four-fifths of the amplitude difference between R_p and S_p for the RS support knots.

The knot optimization, corresponding to the loop shown as the shaded rectangle in the spline framework Figure 3.1, can be started once initialization is complete. The choice of optimization algorithm was driven by the need to efficiently find

an optimum without the use of a gradient, due to the difficulty of analytically representing the the gradient of the figure of merit. The optimization algorithm used in this implementation of the spline framework determines the optimal knot locations by finding k_{MAP} as defined by Equation 4.16b using the cyclic coordinate method (CCM) [100].

CCM is a gradient-free algorithm which exhaustively searches the coordinate axes to be optimized in succession, constraining the search to only one direction at a time. It is computationally efficient and suited to problems where the number of parameters is not very large, as is the case for this application, which optimizes only the 11 characteristic point locations defined by $\mathcal{C}$.

The coordinate axes searched by CCM are the set of all possible knot locations for each characteristic point knot in $\mathcal{C}$, as bounded by its immediate neighbors to the left and right in the full set of knots $\mathcal{K}$. The first and last knots in $\mathcal{C}$ use the signal's start and end points as their left and right bounds, respectively.

With this approach the input to each CCM iteration is a three-tuple of neighboring knots, with the location of the center knot being optimized. The locations of the calculated support knots in $\mathcal{S}$ are not optimized, but they are used as bounding points for neighboring knots in $\mathcal{C}$ that are being optimized. The algorithm updates support knot locations after every CCM iteration to reflect changes in the characteristic point knot locations $\mathcal{C}$.

After finding the optimal location of a knot, optimization proceeds to the next one in $\mathcal{C}$. It repeats the process on the entire set of knots until there is no change in the figure of merit, indicating the best knot locations k_{MAP} have been found.

4.8 RESULTS

The prospective test set used to evaluate algorithm performance comprises 200 beats from 20 randomly-selected subjects, as described in Appendix B. Two reviewers manually annotated all 200 beats with locations of characteristic points in $\mathcal{C}$.

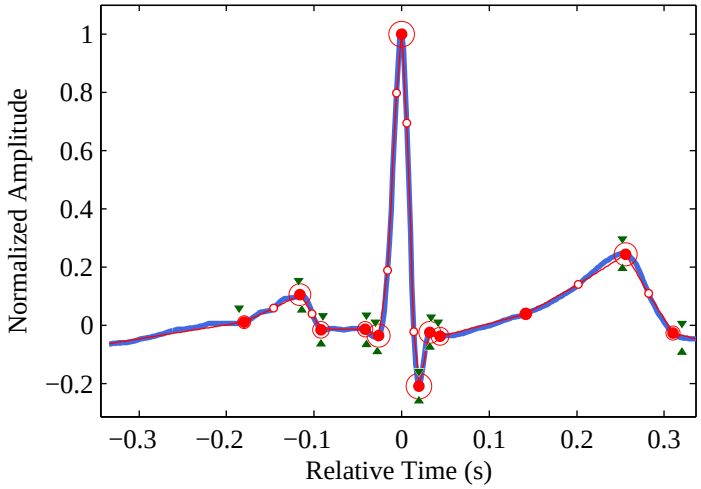
This section first graphically presents results of the Bayesian optimization algorithm on a few beats from the test set to illustrate results of the approach on a number of interesting morphologies. Aggregate accuracies are then presented for all beats in the test set, comparing the algorithm's optimal locations to an established standard when available, and against the reviewers' manually-specified locations for characteristic points that lack an established standard.

Each of the results illustrated in Figures 4.13 through 4.16 shows the underlying ECG signal, its optimized locations and their relevances, and the linearly-interpolated signal estimate. The figures also show the reviewers' manual annotations when available, but because the reviewers were not mandated to annotate points that did not exist or were ambiguous, each characteristic point may have zero, one, or two annotations.

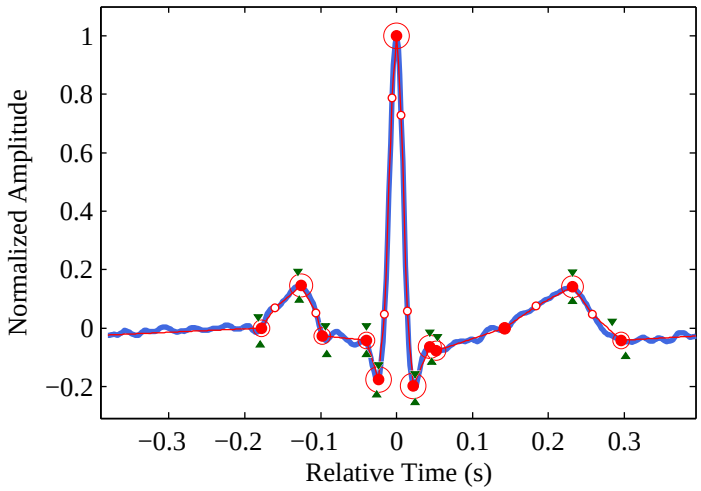
The complexes in Figure 4.13 exhibit all characteristic points in $\mathcal{C}$ with the exception of the T wave onset. Note that Reviewer 1 did not specify a location for the P wave onset P_o in Figure 4.13a, but the algorithm's location closely matches Reviewer 2's annotation. Both of these signals are relatively linear in the region between the end of the QRS complex and the T wave peak, so there is no clear T wave onset T_o. As a result the optimized knot has a low relevance as indicted by the very small outer ring.

The complexes in Figure 4.14 were chosen to illustrate biphasic waveforms characterized by a large S peak. It these cases both are also lacking Q wave and R′ peaks. For both of these waveforms optimization has moved the Q_p knot onto the ascending QR slope between Q_o and the Q_1^R support knot. Similarly, the R'_p knot is on the slope between S_p and the QRS offset S_f.

Figure 4.15 shows two examples with a broad R wave, one missing the Q wave peak and the other missing the R′ peak. In Figure 4.15a the Q_p knot has migrated to a break in the ascending QR slope between Q_o and the Q_1^R support. Similarly, in Figure 4.15b knots for S_p and R'_p are on the descending waveform between the

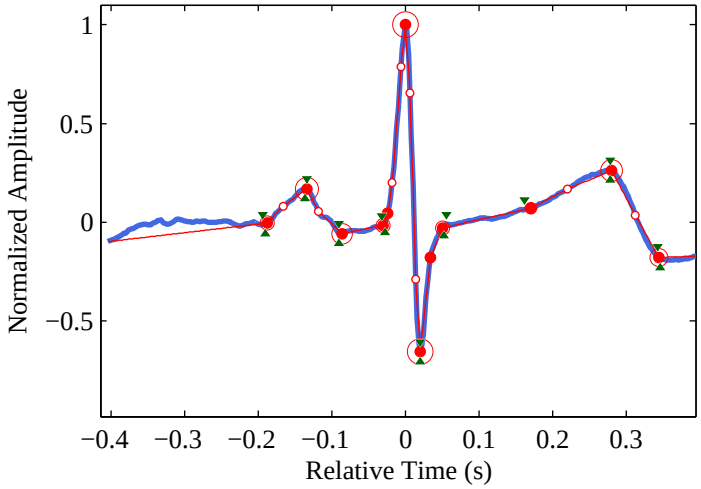


(a) Very small Q_p and R'_p, no T_o.

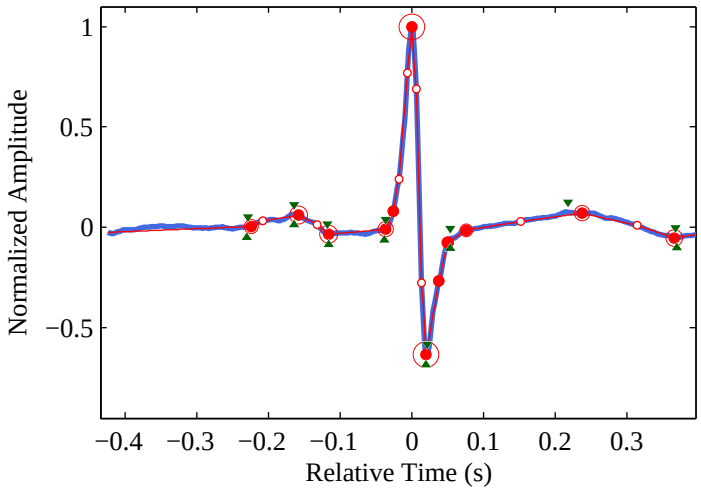


(b) Very small R'_p, no T_o.

Figure 4.13: Results of the optimization on examples from the test set exhibiting most of the characteristic points $\mathcal{C}$. Small open circles indicate the support knots $\mathcal{S}$. Larger red-filled circles are the locations of the optimized knots representing the characteristic points $\mathcal{C}$. The latter are circumscribed with circles of variable size with diameters proportional to $|\rho|$. Triangles indicate locations of reviewers' annotations when available; Reviewer 1's annotations are below the signals, and Reviewer 2's are above.



(a) No Q_p or R'_p.

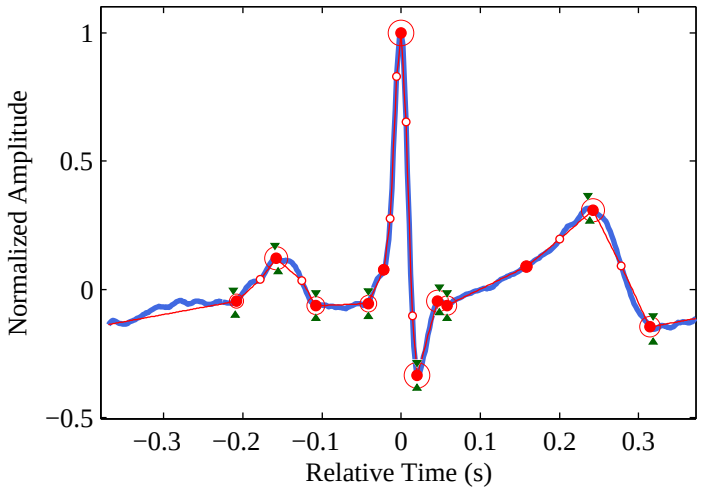


(b) No Q_p or R'_p, very small T_p.

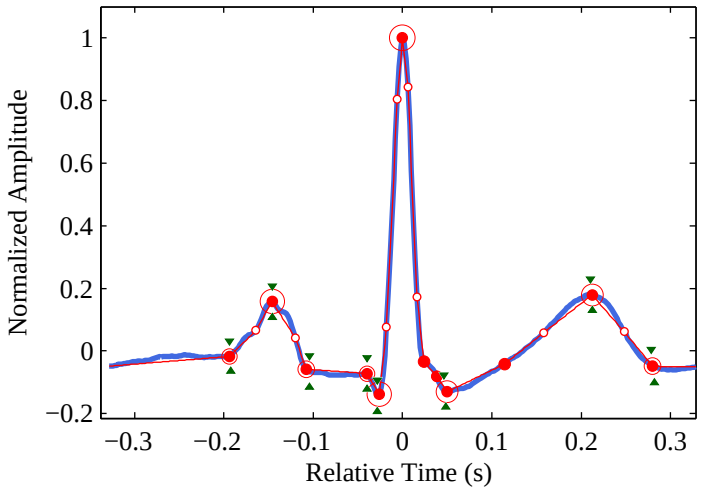
Figure 4.14: Results of the optimization on examples from the test set exhibiting biphasic waveforms, characterized by large S_p. See Figure 4.13 for legend.

R_2^S support and the QRS complex offset S_o.

Figure 4.16 shows waveforms with line noise. The algorithm generally performs well, with the contribution of priors in the optimization helping mitigate the effect of noise on knot placement. However, some locations are still impacted: for



(a) No Q_p and very small R'_p.

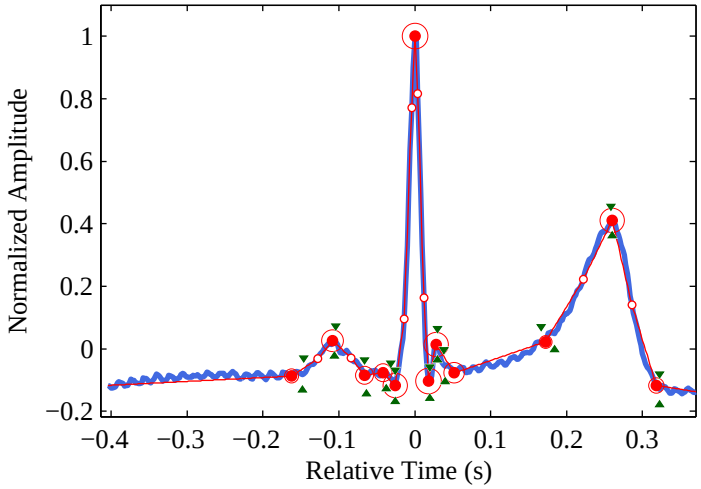


(b) Small Q_p, no R'_p, and no S_p.

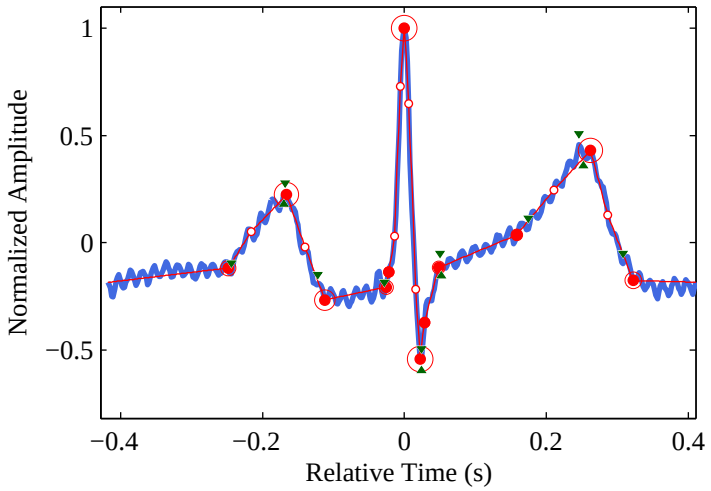
Figure 4.15: Results of the optimization on examples from the test set with wide R wave. See Figure 4.13 for legend.

example, P_o and S_f for Figure 4.16a and T_p for Figure 4.16b.

Table 4.4 compares the reviewers' annotations against each other and against optimal locations determined by the algorithm. For each characteristic point in $\mathcal{C}$, the table shows the mean μ_R and standard deviation σ_R of the difference, in time,



(a) All points in $\mathcal{C}$ present.



(b) Missing Q_p, R'_p, and T_o.

Figure 4.16: Results of the optimization on examples from the test set corrupted with line noise. See Figure 4.13 for legend.

of the two reviewers' annotations for all beats in the training set. It also shows the mean μ and standard deviation σ of the error between the algorithm-optimized knot locations and the average value of the reviewers' manually-specified characteristic point locations. The number of complexes from the 200-beat prospective

test database used in each error statistic calculation are shown in the columns labeled n_R and n. The inter-reviewer comparison count, n_R, is the number of beats which *both* reviewers annotated, and for which an inter-reviewer difference could be computed. The algorithm evaluation count, n, is the number of beats which *either* reviewer annotated, and against which the algorithm could be evaluated. Reviewers were not mandated to annotate any points, and missing manual annotations may be due to lack of that feature in the underlying waveform or to sufficient noise or ambiguity in the morphology to prevent the reviewers from making a manual annotation.

In general, the reviewers' annotations show good agreement. All mean differences μ_{R} are four milliseconds or less, corresponding to two or fewer sample intervals. The standard deviations of the reviewers' differences σ_{R} are also good, with a maximum of 8.5 milliseconds for the T wave onset, which can be very difficult to ascertain due to its lack of curvature. A two-sample t-test on the reviewers' manual annotations for each characteristic point indicated no significant differences between the means of the reviewers' annotations at a 5% level of significance for any of the characteristic points.

It is recognized in the literature that low bias in characteristic point estimates is a necessary but not sufficient criterion to quantify performance of an algorithm estimating their locations. A more important indicator is the variance of the error in relation to established standards. For this comparison, the table includes the standard deviation tolerance values $2\sigma_{\mathrm{CSE}}$, as specified by the Common Standards for Quantitative Electrocardiography (CSE) Working Party in [71]. These values, when available, are *"two standard deviations of the difference between the median of the individual and final referee estimates,"* and estimate *"what can be expected from an expert cardiologist."* This is a demanding standard for automated approaches and subsequent efforts such as [54] define "strict" and "loose" criteria around the CSE tolerance, specifically, an algorithm satisfies the loose criterion if $\sigma < 2\sigma_{\mathrm{CSE}}$,

Table 4.4: Characteristic point estimation errors on the prospective test data set. For each characteristic point in $\mathcal{C}$, this table presents the mean, standard deviation, and sample count for inter-reviewer differences as well as the algorithm's error against the mean of the reviewers' annotations. • indicates the standard deviation meets the loose criterion $\sigma < 2\sigma_{\text{ref}}$, ∘ indicates it is within one sample point of $2\sigma_{\text{ref}}$, and ⋆ indicates that it meets the strict criterion $\sigma < \sigma_{\text{ref}}$.

		Error statistics μ, σ in milliseconds							
		Rev1 vs. Rev2			Mean vs. Algorithm			σ_{ref}	
Char. Point	$\mathcal{C}$	μ_{R}	σ_{R}	n_{R}	μ	σ	n	$2\sigma_{\text{CSE}}$	$2\sigma_{\text{R}}$
P onset	P_o	-0.6	3.7	157	0.6	11.8 ∘	190	10.2	–
P peak	P_p	-1.0	2.5	151	1.4	4.0 •	171	–	5.0
P offset	P_f	-2.8	6.2	168	2.6	12.1 •	192	12.7	–
QRS onset	Q_o	-0.3	3.8	168	-1.0	6.9 ∘	189	6.5	–
Q peak	Q_p	0.0	1.0	92	3.3	3.4 ∘	115	–	2.0
S peak	S_p	0.1	0.7	190	-1.2	1.3 •	190	–	1.4
R′ peak	R'_p	0.4	1.1	83	-5.1	7.6	125	–	2.1
QRS offset	S_f	3.5	5.4	174	0.1	9.2 •	199	11.6	–
T onset	T_o	0.0	8.5	28	-3.8	18.3 ∘	71	–	17.0
T peak	T_p	-1.3	2.1	176	3.7	4.4 ∘	188	–	4.3
T offset	T_f	-4.0	8.3	151	-3.9	8.5 ⋆	174	30.6	–

and the strict criterion if $\sigma < \sigma_{\text{CSE}}$.

Unfortunately, the CSE Working Party does not define tolerance values for most of the characteristic points specified by $\mathcal{C}$ and estimated by the algorithm described here. Since the manual reviewer annotations fulfill the strict criterion $\sigma_{\text{R}} < \sigma_{\text{CSE}}$ for all points specified by CSE except Q_o (and for which $\sigma_{\text{R}} \ll 2\sigma_{\text{CSE}}$),

this suggests a corresponding tolerance value $2\sigma_{\mathrm{R}}$ using the standard deviation of reviewer error as the reference value for characteristic points not available in [71]. So the subsequent comparisons of algorithm annotations use $\sigma_{\mathrm{ref}} = \sigma_{\mathrm{CSE}}$ when σ_{CSE} is available, and $\sigma_{\mathrm{ref}} = \sigma_{\mathrm{R}}$ when it is not.

Symbols are used in Table 4.4 to characterize the algorithm's error variance results as compared to the applicable reference σ_{ref} as determined above. The symbol $\star$ indicates a standard deviation that satisfies the strict criterion $\sigma < \sigma_{\mathrm{ref}}$. The symbol $\bullet$ indicates a standard deviation that satisfies the loose criterion $\sigma < 2\sigma_{\mathrm{ref}}$. And the symbol $\circ$ indicates a standard deviation that is within one sample point (two milliseconds) of the loose criterion, $\sigma < 2\sigma_{\mathrm{ref}} + 2\mathrm{ms}$.

4.9 THE IMPACT OF PRIORS

In order to objectively determine the contribution of *a priori* probability density estimates on the optimization process, the test data were evaluated on an instance of the framework that incorporated only the likelihood term in the figure of merit, ignoring the joint time-relevance prior. No other functionality of the algorithm differed in this instance, including its knot initialization algorithm which used location priors as described in Section 4.7.

Table 4.5 shows means and standard deviations of errors for all characteristic points in $\mathcal{C}$ as compared against the reviewers' manual annotations. These values are computed using the location (in time) of each characteristic point estimate in the test set. Results using the Bayesian figure of merit presented in Table 4.4 are repeated here and augmented with those from the version using the likelihood-only figure of merit.

An aggregate metric is useful in summarizing the performance of both versions of the algorithm across all characteristic points. The mean squared error (MSE) can be used to assess the quality of an estimator, and is defined the average of the squared errors of each location estimate. MSE incorporates both the variance

and bias of the estimator into a single value that can be used to characterize the algorithm's performance for each characteristic point.

Using Equation 4.2, let t_i be the location of knot k_i as determined by manual annotations for all characteristic points $i \in \mathcal{C}$, and let $\hat{t}_i$ be the algorithm's estimated location for that point. Then μ_i and σ_i are the mean and standard deviation of the error in the location estimate $t_i - \hat{t}_i$, computed across all characteristic points i in the test set.

The MSE is defined in terms of μ_i and σ_i (see, for example, [7] or [40]), as the sum of the estimator's variance and squared bias, or

$$\text{MSE} = \sigma_i^2 + \mu_i^2 \tag{4.17}$$

A good estimator is one that is both accurate, having low bias — and precise, with low variance.

The root mean square error is defined as the square root of the mean squared error as defined by Equation 4.17:

$$\text{RMSE} = \sqrt{\sigma_i^2 + \mu_i^2} \tag{4.18}$$

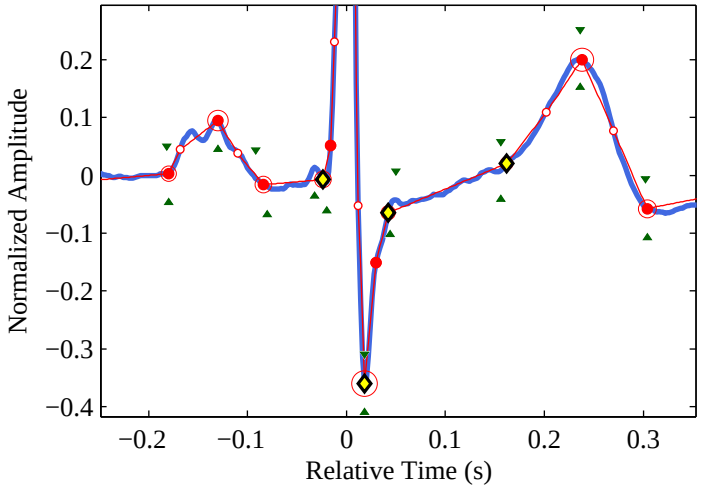
and is shown in the bottom row of Table 4.5 for each approach. RMSE provides a measure of the estimator's performance in units of time, and indicates a significant degradation in accuracy performance when the priors are not used. The RMSE degrades by approximately 50% from 9.6 milliseconds to 14.3 milliseconds when priors are not incorporated by the figure of merit.

There are a few characteristic points that show a very small improvement with the likelihood-only figure of merit. The standard deviations of P_f, Q_p, and T_o show reductions of 2.5, 0.3, and 1.7 milliseconds respectively. These improvements are overwhelmed by the performance of several other points that degraded by multiple sample intervals. The reason for this phenomenon is a subject for future research; it may be due to a combination of the linear interpolant's ability to localize critical points and more consistent knot placement in the absence of the influence of priors.

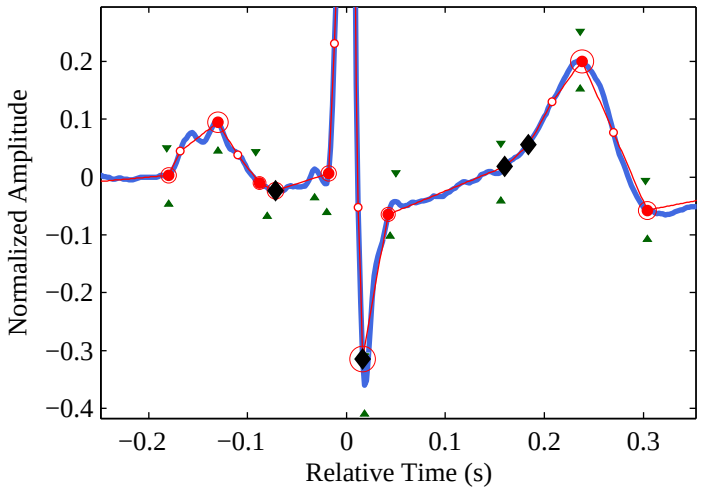
Table 4.5: A comparison of characteristic point estimation errors on the prospective test data, as evaluated with the Bayesian figure of merit and a figure of merit comprised solely of the likelihood. For each characteristic point in $\mathcal{C}$, this table presents the mean and standard deviation of errors against manual annotations from the reviewers. The RMSE reflects the aggregate error of the estimator across all points with each figure of merit, and shows significant degradation when priors are not used. All values are in milliseconds.

		Bayesian		Likelihood		
Char. Point	$\mathcal{C}$	μ	σ	μ_{L}	σ_{L}	n
P onset	P_o	0.6	11.8	-0.9	11.9	190
P peak	P_p	1.5	4.2	1.7	4.6	171
P offset	P_f	2.4	11.7	0.1	9.2	192
QRS onset	Q_o	-0.8	7.0	-10.8	20.7	189
Q peak	Q_p	3.3	3.4	4.0	3.1	115
S peak	S_p	-1.2	1.2	-2.0	2.0	190
R′ peak	R'_p	-5.1	7.6	-6.0	9.8	125
QRS offset	S_f	-0.5	8.8	9.4	24.5	199
T onset	T_o	-4.7	18.8	0.8	17.1	71
T peak	T_p	3.5	4.1	3.2	9.5	188
T offset	T_f	-3.9	8.7	-5.6	12.5	174
RMSE		9.6		14.3		

To illustrate how incorporating priors improves estimates, Figure 4.17 shows expanded views of a beat from the test set analyzed with both versions of the algorithm. Figure 4.17a shows results of the optimization using the Bayesian figure



(a) Knots optimized with the Bayesian figure of merit.



(b) Knots optimized with a figure of merit using only the likelihood.

Figure 4.17: A comparison of optimal knot locations using the Bayesian figure of merit and a figure of merit comprised solely of the likelihood. Highlighted knots indicate, from left to right, the QRS onset, S wave peak, QRS offset, and T wave onset. Eliminating priors from the figure of merit causes these knots to move from the locations indicated by yellow markers in (a) to the locations indicated by black markers in (b). Knots representing the Q and R′ wave peaks are also adversely impacted.

of merit that incorporates the priors. Figure 4.17b shows results using a figure of merit comprised solely of the likelihood.

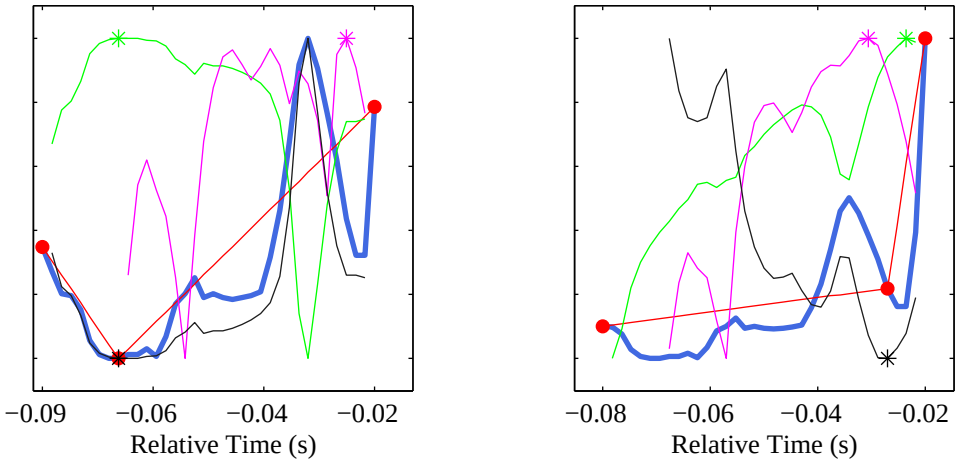
These figures highlight four knot locations: from left to right, they correspond to the locations of QRS onset Q_o, S wave peak S_p, QRS offset S_f, and T wave onset T_o. Locations of the knots representing peaks of the Q wave Q_p and the R′ wave peak are also affected, but not as much as those that are highlighted.

The highlighted markers in Figure 4.17a show locations for these characteristic points that are generally good, matching annotations provided by the reviewers to within a few samples. The corresponding markers in Figure 4.17b are significantly off of their desired positions. The Q_o knot is far to the left of its proper location and immediately follows the P wave offset. The knot for S_p is no longer on the peak, having moved to the left. The QRS offset knot S_f is very far to the right and aligns with the T wave onset marked by both reviewers. This has caused optimization to move the T_o knot too far to the right on the T wave.

To obtain insight into how the Bayesian figure of merit fuses the priors and likelihood to provide improved estimates of characteristic point locations, Figures 4.18 through 4.20 show the final CCM iteration that produced optimal knot locations for three of the points highlighted in Figure 4.17. For each characteristic point, results are shown for both figures of merit.

Each of these figures shows, in blue, the segment of the original signal surrounding the optimized knot. That knot, its immediate neighbors to the left and right (from $\mathcal{K}$), and the resultant linear spline estimate of the segment are displayed in red. The green line indicates the value of the likelihood computed at each location of the center knot as CCM moves it across the domain of the signal segment, *i.e.*, between its neighboring knots. The magenta line similarly represents the value of the time-relevance prior at each point, and the black line is the computed figure of merit.

The best values for the likelihood, prior, and figure of merit are indicated by



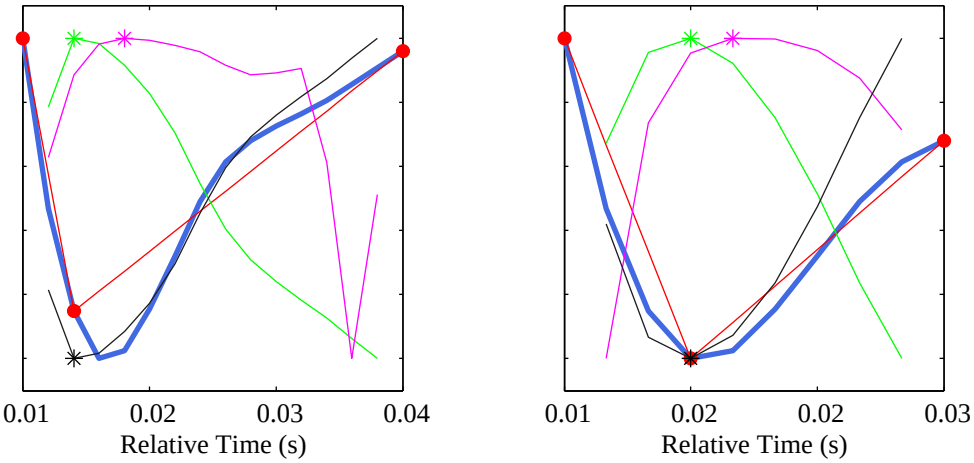
(a) Knot optimized using a figure of merit comprised only of the likelihood. The optimal location is 25 samples (50 ms) to the left of the correct location, and too close to the P wave offset.

(b) Knot optimized with Bayesian figure of merit. Incorporating the priors moves this knot to within one sample of its desired location at -22 ms, at the base of the QRS complex.

Figure 4.18: A comparison of the final optimization search for Q_o using a figure of merit comprised solely of the likelihood against the Bayesian figure of merit. Figures show values of the likelihood in green, the time-relevance prior in magenta, and the computed figure of merit in black, for every location of the center knot between its neighbors. The center knot is the optimal location for each figure of merit.

an asterisk of the appropriate color. In order to increase visibility at the top of the figures, the figure of merit line is inverted, and its best value is at the bottom of the plot.

Displaying such a wide variety of data on a single plot is challenging, so all values were normalized to the range available. As a result the vertical axes do not indicate true magnitudes of the displayed data. They are, however, useful in assessing relative changes in each value as the center knot is scanned across the signal segment by CCM. Their time axes are expressed relative to the R wave peak, and correspond with the larger depictions shown in Figure 4.17.



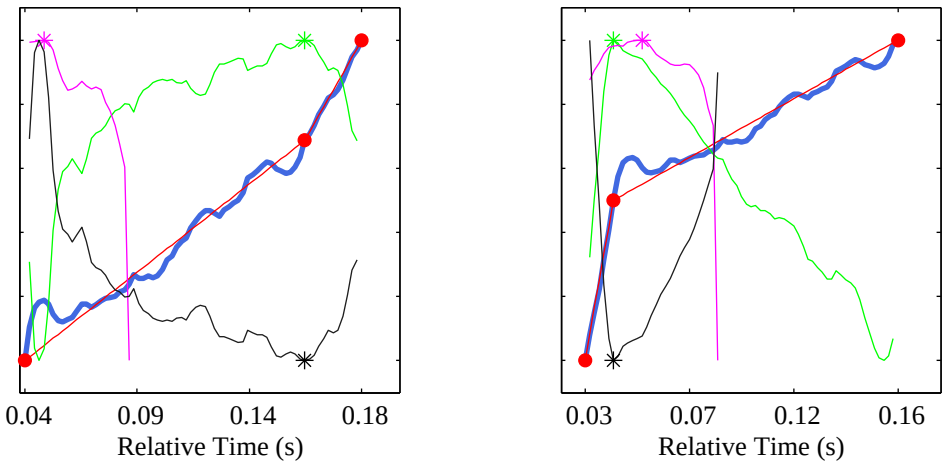
(a) Knot optimized using a figure of merit comprised only of the likelihood. The signal's large slope degrades the likelihood when the knot is correctly placed on the peak; as a result it is placed one sample to the left of the peak.

(b) Knot optimized with Bayesian figure of merit. Incorporating the priors moves this knot to exactly the peak of the S wave at 18 ms.

Figure 4.19: A comparison of the final optimization search for S_p using a figure of merit comprised solely of the likelihood against the Bayesian figure of merit. The figures show values of the likelihood in green, the time-relevance prior in magenta, and the computed figure of merit in black, for every location of the center knot between its neighbors. The center knot is the optimal location for each figure of merit.

The left sub-figure of all of these figures, labeled (a), displays results of the final optimization cycle with a figure of merit comprised solely of the likelihood. Results of the final optimization cycle with the Bayesian figure of merit are shown in the right sub-figure, labeled (b).

In order to give an indication of how the time-relevance priors might have impacted the (a) sub-figures, the values of the prior are shown even though they are not used in the computed figure of merit. So for all of the (a) sub-figures, the figure of merit (black) are an inverted version of the likelihood (green), indicating



(a) Knot optimized using a figure of merit comprised only of the likelihood. In this long segment, the point of maximum likelihood is at 160 ms, 57 samples to the right of its proper location.

(b) Knot optimized with Bayesian figure of merit. Incorporating the priors moves this knot to 42 ms, within two samples of its correct location.

Figure 4.20: A comparison of the final optimization search for S_f using a figure of merit comprised solely of the likelihood against the Bayesian figure of merit. The figures show values of the likelihood in green, the time-relevance prior in magenta, and the computed figure of merit in black, for every location of the center knot between its neighbors. The center knot is the optimal location for each figure of merit.

exact equality.

Because of the iterative nature of CCM, and that it is applied multiple times to each segment, knot locations for the two figures of merit are generally different. For each characteristic point, the time axis can be used to determine the location of the segments with respect to each other, and with respect to the larger depictions shown in Figure 4.17.

Figure 4.18 shows the plots described above for the QRS onset characteristic point, Q_o. The best location shown in Figure 4.18a corresponds to the maximum

value of the likelihood, which occurs at -72 milliseconds. This point misses the correct QRS onset by 50 milliseconds (25 sample intervals). Results with the Bayesian figure of merit are shown in Figure 4.18b. In this case, the priors' influence has moved the knot to one sample of the desired location at -22 milliseconds.

Figure 4.19 shows the plots described above for the S wave peak, S_p. The location of the peak based only on the likelihood in Figure 4.19a is impacted by the large slope of the descending RS wave. With the center knot at the correct position (*i.e.*, at the negative peak of S_p), the difference between the linear estimate and the signal causes a sizable decrease in likelihood, so the knot is placed to the left of the peak. Incorporating the priors causes the knot to move to the exact location of the peak, as shown in Figure 4.19b.

Finally, the CCM optimization for the QRS offset, S_f, is shown in Figure 4.20. In this relatively long signal segment, the likelihood-only figure of merit peaks at 160 milliseconds, far to the right of the desired location at 44 milliseconds, as shown in Figure 4.20a. Incorporating the priors causes the knot to move to within one sample of the correct location as shown in Figure 4.19b.

4.10 REDUCING THE TRAINING SET

The previous section explored the effect of completely eliminating priors from the figure of merit. To help determine the effect of the quantity of training data on the accuracy of characteristic point location estimates at a more granular level, this section reports results of the optimization on the prospective test set using different priors calculated from reduced sets of training data.

In five separate runs, priors were recalculated from the manual annotations of a randomly-selected set of subjects. Each of the five runs used priors calculated from data of one, two, four, eight, 16, and 32 subjects. There were a total of 40 subjects in the training set.

Table 4.6 shows the results of the optimization for all of these runs. Each row

of the table corresponds to a different number of subjects, n_s, from the training set whose data were used to compute priors. Each column, R, indicates the run number for a different randomly-selected set of subjects used in obtaining the priors.

The entries in the table for each run are the RMSE values calculated from the means and standard deviations of the errors in the estimates using Equation 4.18. The rightmost columns, μ_{RMSE} and σ_{RMSE} are the means and standard deviations of the RMSE values for each row.

Table 4.6: RMSE values for characteristic point estimates on the prospective test set, obtained using priors calculated from annotations on a reduced set of subjects. Each row corresponds to the number of subjects, n_s, whose data were used to calculate priors. Columns R_1 through R_5 indicate different sets of randomly-selected subjects for each row. The rightmost columns, μ_{RMSE} and σ_{RMSE} are the means and standard deviations of the RMSE values for each row.

n_s	R_1	R_2	R_3	R_4	R_5	μ_{RMSE}	σ_{RMSE}
1	16.69	16.62	13.44	12.75	19.01	15.70	2.58
2	10.84	10.80	11.43	14.62	11.35	11.81	1.60
4	10.85	9.76	11.09	11.10	10.39	10.64	0.57
8	9.17	9.40	9.66	9.41	10.40	9.61	0.48
16	9.84	10.00	9.56	8.71	9.19	9.46	0.52
32	9.82	9.79	9.55	9.60	9.31	9.61	0.21

It is clear from the table that RMSE values are highest when only one subject is used for obtaining priors, and generally decrease as the number of subjects is increased. Neither the mean nor the standard deviation decreases monotonically with increasing number of subjects. However, the mean values for the three bottom

rows (corresponding to eight, 16, and 32 subjects) are within two percent of each other indicating very little change, and the standard deviation obtained using 32 subjects is less than half its nearest value.

To better illustrate the impact of training set size on accuracy, Figure 4.21 shows superimposed plots corresponding to the values in Table 4.6. The plots in gray are the RMSE values for each individual run, R_1 through R_5. The plot in blue is the mean RMSE value across all five runs for each value of n_s. Also in blue is a whisker plot showing one standard deviation above and below each mean value.

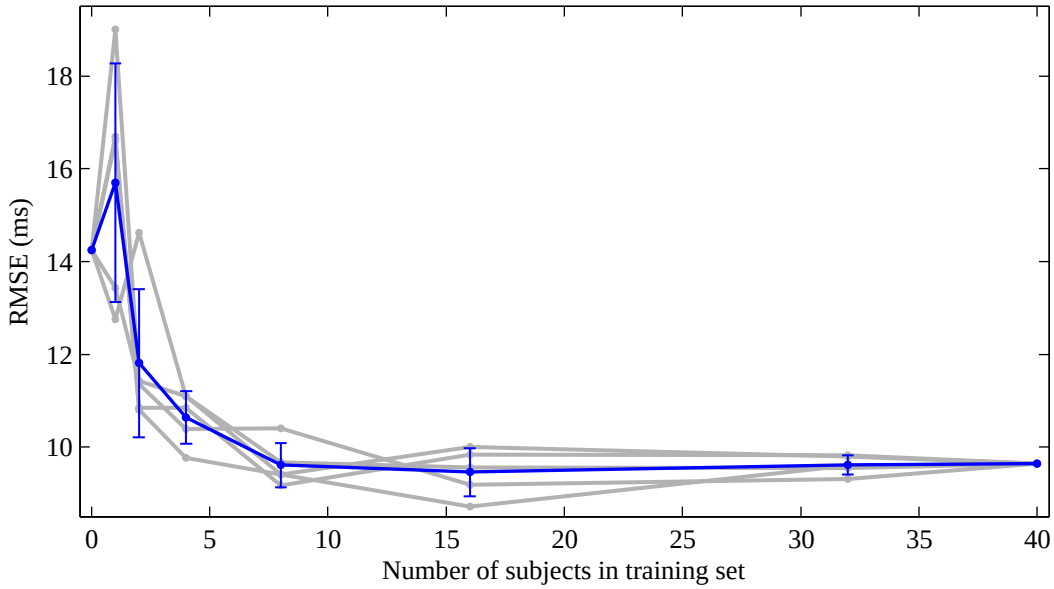


Figure 4.21: Plot of RMSE values for characteristic point estimates on the prospective test set, obtained using priors calculated from annotations on a reduced set of subjects. These correspond to values shown in Table 4.6, with the first and last points being the RMSE values for the likelihood-only and full training previously reported. The plots in gray are the RMSE values for runs R_1 through R_5. The plot in blue is the mean RMSE value across all five runs for each value of n_s. The whisker plot shows one standard deviation above and below each mean value.

The first and last points in the plot are not shown in Table 4.6, and correspond

to the RMSE values for the zero-subject case and for the 40-subject case. The former was presented in Section 4.9, and was obtained using the likelihood-only figure of merit. It has an RMSE value of 14.3 milliseconds. The latter corresponds to the full training set, whose results were reported in Section 4.8 and again in Section 4.9. It has an RMSE value of 9.6 milliseconds. Since there is only one run corresponding to each of these cases, there is no standard deviation.

The likelihood-only case has lower RMSE than several of the the single-subject runs (R_1, R_2, R_5) and one of the two-subject runs (R_4). The limited data used in these cases biases the optimization in favor of the one or two subjects whose data were used in calculating the priors, and away from the average that more accurately represents all the subjects in the test set. As a result the errors are generally larger and have greater variance between runs.

There is a knee in the curve at eight subjects, after which the mean shows very little change. However, the standard deviation has its minimum value with 32 subjects. These data suggest that — for cases where the subjects used in training are not the same as those in the test set (which is the case here) — at least eight subjects should be in the training set to reduce mean error, and that further increasing the number of subjects can reduce RMSE variance.

4.11 SUMMARY

This chapter described an implementation of the spline framework to represent ECG waveforms and optimally estimate the locations of several characteristic points that can be used to describe the signal. The framework integrates prior information about the amplitude and timing of these points obtained from a large training database, a flexible spline representation to identify the features of interest, and Bayesian inference to balance the prior information with the observed data. With suitably defined knots and training data, the framework can readily be extended to include other leads, different characteristic points, or be applied to

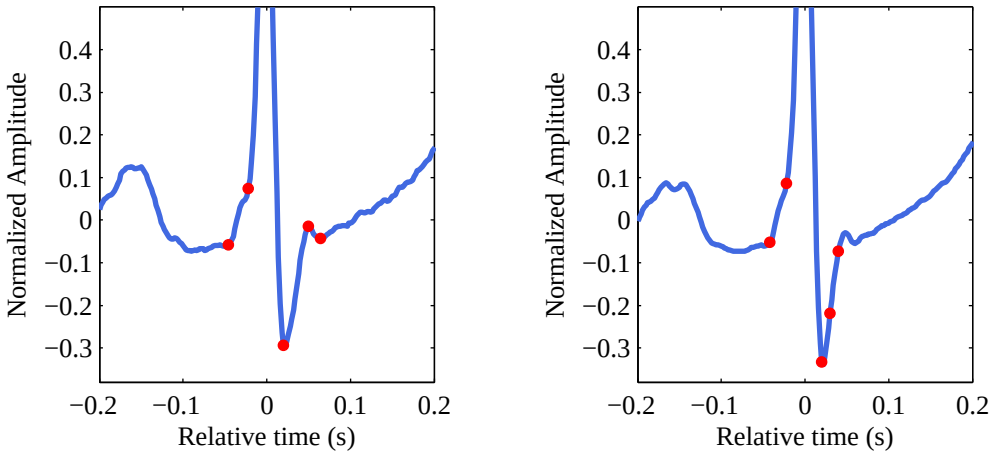
other semiperiodic signals.

The implementation performed well on a prospective test data set, accurately locating knots corresponding to the characteristic points $\mathcal{C}$ across a wide range of normal morphologies comprising 200 beats from 20 subjects. Table 4.4 shows low mean errors for the algorithm when compared to the average of the reviewers' manual annotations. The largest mean errors are approximately five milliseconds, which corresponds to 2.5 signal samples at the 500 Hz sampling rate used for this data.

The variance of the errors is recognized to be more important than their mean. To assess characteristic point location variability the standard deviations of errors were compared against a reference standard deviation. When available, the reference was the $2\sigma_{\text{CSE}}$ value defined by the CSE Working Party in [71]. For those characteristic points not defined by the CSE, the $2\sigma_{\text{R}}$ value derived from manual annotations by the reviewers was used (as described in Section 4.8).

As shown in Table 4.4, of the 11 error standard deviations corresponding to characteristic point locations estimated by the algorithm, one meets the strict criterion, four meet the loose criterion, and five are within one sample (two milliseconds) of the loose criterion. The R′ peak is the only point that is not close to the reference with a difference of a little over five milliseconds.

An analysis of errors in R′ knot placement showed that the problem is due to the low amplitude of the peak in much of the data, which can cause the feature to be missed during the optimization process. Figure 4.22 shows two beats from the same subject illustrating an error representative of those resulting in high error variance for the R′ peak. In Figure 4.22a the R′ wave is detected correctly despite its small amplitude. In Figure 4.22b, however, slight changes in the beat's morphology in addition to the R′ wave's small amplitude cause the knots for the R′ wave peak and QRS offset to move down to a relatively linear region of the S wave, missing their correct locations.



(a) Correctly estimated R′ and S_o.

(b) Incorrectly estimated R′ and S_o.

Figure 4.22: Analysis of the error in R′ location estimates. Shown are two QRS complexes from the same subject in the test set, indicating (from left to right) the characteristic points Q_o, Q_p, S_p, R′ , and S_f as determined by the optimization. In (a), the algorithm properly places all knots. Slight differences in the morphology of (b), compounded by the very small amplitude of the R′ peak, result in incorrect placement of both the R′ wave peak and QRS offset.

Overall the algorithm performs well for beats that are described by the selected characteristic points $\mathcal{C}$ and supported by the *a priori* probability density estimates developed using the training data set. Beats whose morphology is not described by $\mathcal{C}$, such as the test beats used to illustrate the spline framework in Chapter 3, will not be handled well by this instance of the framework. Properly analyzing such beats will require a different set of characteristic points and associated priors.

Chapter 5

CONCLUSION

This dissertation presents a framework for representing semiperiodic signals using splines, and an implementation of that framework to optimally estimate the locations of a signal's characteristic points. The optimization algorithm uses a Bayesian approach that incorporates prior probabilities of the characteristic points of interest derived from a manually-annotated training data set.

It also documents an extensive search exploring algorithms in the literature that address the signal segmentation and delineation problem. There is a significant amount of work in this field broadly categorized into conceptual classes including algorithms that operate in the time domain, on transformed signals, using models, or with pattern recognition techniques. None of the algorithms found in the literature review explicitly use prior information in signal delineation or for optimally estimating characteristic point locations.

Splines, which are inherently a time domain technique, provide an efficient signal representation and have been applied to semiperiodic signals like the electrocardiogram in the literature. However, their existing use only addresses noise elimination, data compression, and waveform analysis (for example, to locate the ST segment of an ECG waveform). The efforts described in Chapters 3 and 4 are novel in developing a generic framework and then applying it to optimally estimate an arbitrary set of a signal's characteristic points.

5.1 CONTRIBUTIONS AND KEY LEARNINGS

The first contribution of this dissertation is the framework for parametric representation of semiperiodic signals as described in Chapter 3. This framework allows representation of semiperiodic waveforms on a cycle-by-cycle basis using splines. It is very flexible, allowing choice of knot initialization method, spline interpolant, figure of merit, and optimization algorithm. These choices facilitate tradeoffs between factors including computational complexity, fidelity of signal representation, and ability to estimate characteristic point locations.

The choice of spline interpolant impacts computational complexity, fidelity of the signal's interpolated estimate, and even the ability to accurately estimate characteristic point locations. Although more sophisticated interpolants can estimate the signal itself with higher fidelity as measured by RMSE, differentiability constraints can adversely impact their use in characteristic point estimation by moving the knots off of true characteristic points during the optimization process. In applications estimating characteristic point locations, a linear interpolant serves well and with low computational burden.

Using a dynamic knot initialization algorithm such as RPA allows representing a cycle of virtually any semiperiodic morphology, independent of assumptions regarding specific characteristic points. However its use for estimating characteristic point locations can be problematic because each dynamically allocated knot created by the recursive partitioning would have to be mapped to a corresponding characteristic point after its optimal location is determined.

The second contribution of this dissertation addresses the knot-to-characteristic point mapping problem by developing an implementation of the spline framework for optimal estimation of a fixed number of characteristic point locations in an ECG signal. The algorithm described in Chapter 4 uses a pre-defined set of knots, each representing one characteristic point of interest. Instead of using RPA as

before, knots are initialized to locations calculated from the prior distributions which are obtained from manual annotations of a training data set.

This implementation also introduces a new figure of merit for optimization: the Bayesian *a posteriori* probability of a set of knot locations, given the observed signal. The figure of merit is computed from an estimate of the *a priori* joint probability density of the times and curvatures (or relevances) of the characteristic points, which is the third major contribution of this dissertation.

As with knot initialization, these priors are estimated from manual annotations of the training data. In this case, however, the priors include additional information regarding the curvature of each knot that is computed from the relative times and amplitudes of each three-tuple of knots; *i.e.*, the knot under consideration and its immediate neighbors to the left and right. This prior allows the figure of merit to incorporate known curvatures of each characteristic point in assessing each knot location.

In this domain, locations of low curvature like waveform onsets and offsets are distinguished from those with moderate or high curvature, such as peaks of P and T waves, or Q, S, and R′ peaks. Incorporating curvature into the optimization process makes it less likely for the figure of merit to result in a high value when the knot undergoing optimization is at an incorrect location. For example, this could occur if a knot representing a waveform's peak is located at the waveform's onset or offset.

A cyclic coordinate method search exhaustively explores the search space of each knot between its immediate neighbors, finding the location resulting in the maximum *a posteriori* probability, or MAP. Several successive applications of the CCM on the entire set of knots help ensure the solution converges to the best figure of merit possible given the limited number of knots and their constrained locations.

The fourth contribution of this dissertation is the objective means to obtain

priors from training data. In this implementation manual annotations of the characteristic provide the times and amplitudes used to model the location priors used for knot initialization, as well as the joint time-relevance priors used for optimization.

Using priors in the optimization process is intuitively appealing as it mimics, in a very limited way, the approach human experts use when analyzing physiological waveforms. To diagnose or interpret a particular subject's ECG, for example, a clinician will review that subject's waveform preceding the time of interest in order to establish a baseline. Changes, as measured against this baseline, help the clinician identify conditions of interest. Of course, a clinician's prior knowledge also includes extensive training and other information about the subject and their state. In a manner loosely analogous to Bayesian optimization, the clinician fuses their prior knowledge with observations in order to make an estimate of the subject's current health state (*i.e.*, a diagnosis).

Many methods in the literature use heuristic thresholds and windows to determine characteristic point locations. Using prior knowledge in the form of estimates of the *a priori* probability densities of characteristic point locations reduces the need for such empirically-determined values.

The Bayesian approach is an effective one, but only to the extent that the priors incorporated in the optimization reflect reality. Human physiology exhibits tremendous variation, between subjects and even for a given subject over time. The power that priors bring to this approach can be obviated, if they are not representative of the signals under analysis. The more accurate the priors are in this regard, the greater their benefit in complementing the likelihood and providing the best estimate.

The figure of merit has the most direct impact in achieving the goal for each instance of the framework. For a high-fidelity representation of the signal an overall measure comparing the interpolated estimate against the observed signal, such as

RMSE, can be useful. To estimate characteristic point locations, other factors, such as the time of the knot and its relevance should also be incorporated in the figure of merit. The maximum *a posteriori* Bayesian figure of merit described in Sections 4.3 and 4.6 incorporates the observed signal, as well as prior knowledge of knot times and relevances. Using this figure of merit, the algorithm obtained low mean errors, and variances close to a standard based on reviewers' annotations, for virtually all characteristic points.

The optimization method is a key component of the framework, finding the model parameters (the knot locations for a spline representation) producing the best figure of merit. Choice of this algorithm is driven by balancing computational complexity against its ability to find the globally-optimal set of characteristic points. An exhaustive search of all possible locations would produce the globally optimal points. However, due to increased computational load this approach can be prohibitive for a large number of signals or if resources are limited. Employing CCM as described in Section 4.7 performs an exhaustive search, but on a constrained subset of the full search space. Although there are a few cases where this approach settles on a local optimum, overall it provides an acceptable tradeoff with good performance.

Since this process occurs in the time domain and the optimized knots map directly to the characteristic points they estimate, there is no translation required to make the algorithm results accessible to clinical domain experts. Another advantage of the time domain representation is the compactness of its representation. The relatively low duty cycles of the QRS complex, in particular, would require a greater number of features if represented in the frequency domain at equal fidelity.

The algorithm described in Chapter 4 complements existing approaches with a flexible alternative to locate characteristic points and delineate semiperiodic signals. This flexibility provides the capability to readily specify, train, and estimate optimal locations for points not currently used in ECG analysis (a superset of

$\mathcal{C}$). Using this approach, researchers can investigate large databases more extensively with reduced manual effort. Automatic, accurate, tracking of characteristic point locations over extended times may lead to new metrics indicative of disease, physiological stress, or other conditions of interest.

5.2 SIGNIFICANCE AND APPLICATIONS

Automatic, accurate estimation of characteristic point locations for semiperiodic signals is especially important in analysis of physiological signals such as the electrocardiogram. In this domain, characteristic points correspond to specific electrophysiological changes in cardiac tissue that can be of clinical importance.

The locations and amplitudes of constituent waves, segments, and intervals in the ECG signal are used to diagnose disease states and assess treatment efficacy. Metrics derived from the characteristic points also reflect the state of the subject's autonomic nervous system providing a real time view of stress.

Temporal trajectories of characteristic points or metrics derived from them have clinical value as well. In some leads there are beat-to-beat changes in the amplitude of the R wave reflecting a subject's respiration rate. Fluctuations in instantaneous heart rate are driven by the autonomic nervous system, causing RR intervals to exhibit reduced variability when the sympathetic nervous system is activated and increased variability when the parasympathetic nervous system is activated. ST segment deviation, in which a small segment of the ECG waveform becomes elevated or depressed with respect to the isoelectric level, is used in cardiac stress tests to identify ischemia in the heart muscle. Other measures whose temporal evolution is important include the QT interval and T wave alternans. Prolongation or shortening of the QT interval has been established to be an independent risk factor for sudden cardiac death [93], [24]. T wave alternans is a very low amplitude beat-to-beat variation in the morphology of the T wave and has similarly been linked to sudden cardiac death [36].

For these reasons clinicians and researchers can benefit from tools that reliably locate points of interest in large data sets and potentially in real time on wearable devices. This capability can enable more thorough exploration of physiological phenomena across large populations. And if deployed on-body, it can be used to drive in-the-moment therapeutic interventions or to generate alarms in the case of life-threatening events.

Although the examples documented in this effort involve the ECG signal, the methods described can be used in other domains involving semiperiodic signals where a spline representation or optimized estimate of the characteristic points are of interest.

Other physiological signals that could benefit include cardiopulmonary signals such as those obtained from a photoplethysmograph, pulse transducer, respiration sensor (using resistive or impedance transducers), or blood pressure transducer (invasive or non-invasive).

In the case of hemodynamic signals, characteristic points of pulsatile waveforms are used to determine cardiovascular parameters such as pulse onset, systolic peak, dichrotic notch, and dichrotic peak as determined by the confluence of the percussion wave, tidal wave, and dichrotic wave ([105], [72]).

For respiratory signals, characteristic points can be used to identify tidal breath cycles. Each cycle is can be described by only two characteristic points: the first represents the beginning of the cycle at the start of the inspiratory phase (which is also the end of the expiratory phase of the previous cycle). The second characteristic point is at the transition of the inspiratory phase to the expiratory phase. In [106], Wang *et al.* describe a method based on signal derivatives and heuristic rules to automatically identify these characteristic points. In addition to use in diagnostic applications, accurate identification of tidal cycles can be used in therapeutic applications such as medication delivery systems. For example, tidal cycle identification is used by [88] to adapt drug delivery to the patient's breathing

pattern, providing inhaled medication at the appropriate time during inspiration to reduce waste and increase effectiveness.

Electroglottography (EGG) is a technique that measures the degree of contact between vocal folds during production of voiced speech by detecting changes in electrical impedance as measured by electrodes placed on a subject's throat. This noninvasive method is used by speech pathologists to assess vocal fold dysfunction. The EGG waveform is semiperiodic and the locations of characteristic points provide important information about the vibration properties of the vocal folds, including the time at which the lower margins of the folds contact, the point of tightest contact, and when the folds start to separate [97]. The methods described in this dissertation can be used to determine the optimal locations of these characteristic points.

Another potential use of the spline framework and Bayesian optimization algorithms is for spike sorting, a technique used in neuroscience to group spikes, the action potentials produced by the firing of neurons, into clusters based on morphological similarity, facilitating assignment spikes to the neurons generating them [74]. In [70], spikes are characterized using features obtained from the minima and maxima of first and second derivatives of the action potential signals. The action potential signals could instead be represented by the spline framework with optimized knot locations corresponding to the characteristic points defining each spike's shape. The optimized knot locations could then be used as features for clustering and classification.

In biomechanics, electromyographic or on-body inertial sensors are often used to monitor and assess activity. When captured for gait analysis, such signals are semiperiodic with a fundamental frequency derived from the step-to-step interval during walking or running. Identifying various points of interest such as heel strike, terminal stance, toe off, and foot swing during each gait cycle can be useful to characterize the gait signal for coaching purposes, rehabilitation, and in tracking

progression of degenerative conditions like Parkinson's Disease. In these applications points of interest in the gait cycle can be represented by characteristic points whose locations are optimally estimated using the spline framework.

Tides, the variation of sea level caused by the gravitational attraction of the moon and sun, are also semiperiodic. The study of tides, and in particular their prediction, is an important aspect of physical oceanography and is used in marine navigation, conservation, fishing, and construction [68]. In addition, tidal datums, the base elevations used as reference values, are used as the basis for establishing land ownership, economic zones, territorial seas, and high seas boundaries [67].

As tides are semidiurnal in most areas, there are generally two high tides and two low tides each day. The corresponding daily water elevation waveform comprises two high water peaks ("higher high water" and "lower high water"), and two low water negative peaks ("higher low water" and "lower low water"). These are the characteristic points of this semiperiodic waveform, and can be used by the spline framework to model the signal. The corresponding optimized knot locations can then be used to derive tidal period, tidal range, and tidal amplitude values. These values, and their statistics taken over many observations, are used in specifying tidal datums [92].

Photometry — the measurement of an object's brightness — has long been used in astronomy to ascertain important information regarding celestial objects. The orbit of extra-solar planets (exoplanets) [73] or other objects such as magnetospheric clouds [98], around distant stars can impart a semiperiodic dimming in the observed flux, or brightness, of the star. Characteristic points of the resultant light curve waveform correspond to various points of the observed object's orbit around the star. These points can be used to indicate locations of initial, maximum, and final occlusion of the star by the orbiting body. These, in turn, can provide information regarding the radii of the star and exoplanet, the mass of the star, and the exoplanet's orbital speed, among other values of interest.

In some cases such as with multiple orbiting planets, orbiting clouds, or in the presence of starspots which may be occluded by the orbiting object, the resultant light curve may have additional features requiring a larger number of knots for proper characterization [75].

5.3 FUTURE WORK

A number of improvements to the approach described in Chapter 4 can increase accuracy, extend applicability of the algorithm, or provide additional information regarding the signal under analysis.

5.3.1 Optimization improvement

Although the CCM optimization algorithm is an effective one, the implementation described in Section 4.7 can find a local optimum resulting in poor knot placement. The incorrectly estimated R′ peak and QRS offset described in Section 4.11, and shown in Figure 4.22, are an example of how constraints imposed on the CCM can cause this to occur.

As described in Sections 4.5 and 4.7, in order to optimize a knot's location, CCM searches all possible locations for the center knot of a three-tuple, bounded by its immediate neighbors on each side. In the case described above, due to initial knot locations and the morphology of the particular waveform, these constraints preclude CCM from even from considering the proper locations for R'_p and S_f.

Figure 5.1 shows the same beat, illustrating the three-tuple surrounding the QRS offset S_f and the corresponding linear estimate of the signal used to calculate the likelihood. Figures 5.1b and 5.1c show other locations considered by CCM during the optimization search. Both of these locations result in a poor estimate of the underlying signal by the linear interpolant. The corresponding error results in a low likelihood and figure of merit, so these locations are not selected. Figure 5.1a

shows the best location determined by CCM. However, as previously discussed, the locations of R'_p and S_f are not as desired.

To reduce the likelihood of identifying a local maximum, the proposed improvement to the CCM optimization method scans two coordinate axes simultaneously, achieving a more thorough exploration of the search space: instead of searching for the best location of the center knot of a three-tuple, it searches for the best locations of the *two* center knots of a four-tuple.

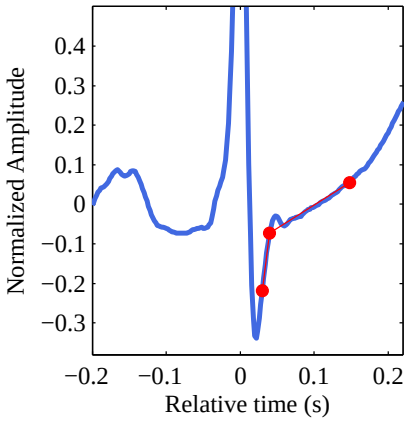
In the example above, this means scanning the region between the S wave peak S_p and T wave onset T_o for the optimal locations of both R'_p and S_f. Figure 5.2 illustrates this approach. Figures 5.2a and 5.2b show two locations encountered by the CCM optimization. Both of these have a large error in their linear estimate and as before, result in a low value for the figure of merit. The locations for R'_p and S_f shown in Figure 5.2c are the desired ones and with this improvement would be identified as the optimal locations.

There is of course a cost associated with this improvement as considering a larger part of the search space requires additional computation and time.

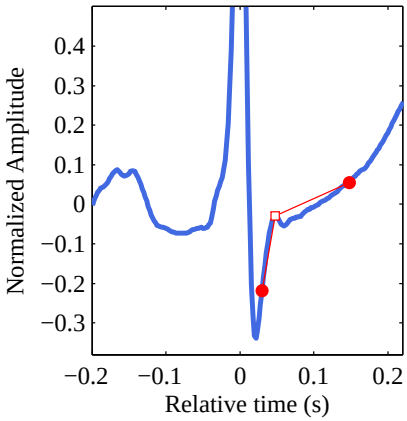
5.3.2 Explicit use of knot amplitude priors

Another potential improvement supports analysis of noisy signals. The effort described in Chapter 4 uses the amplitudes of characteristic points only to compute the knot's relevance value; priors for the amplitude values themselves are not explicitly used. Furthermore, the optimization algorithm determines amplitude values by using the signal's value at the knot time under consideration, effectively constraining it to lie on the signal.

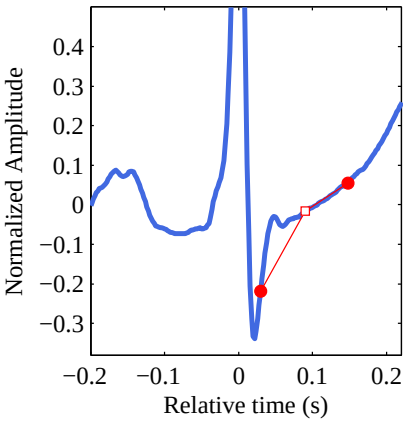
Amplitude values obtained in this manner are susceptible to noise on the signal. If there is a large amount of noise at the current knot time being analyzed by CCM, it will result in a spurious relevance value to the optimization algorithm. This will in turn cause an erroneous figure of merit value and adversely impact knot location.



(a) Local optimum for S_f with existing implementation, showing the best linear estimate that can be obtained between the two constraining knots.

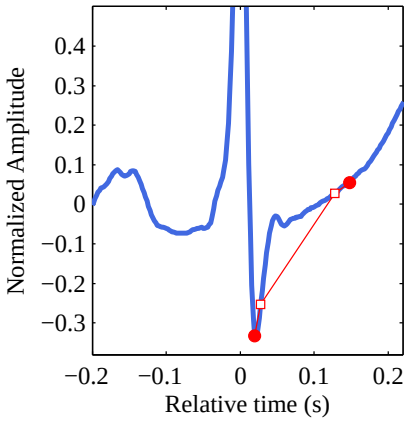


(b) The square marker indicates an S_f location explored by CCM, with poor likelihood.

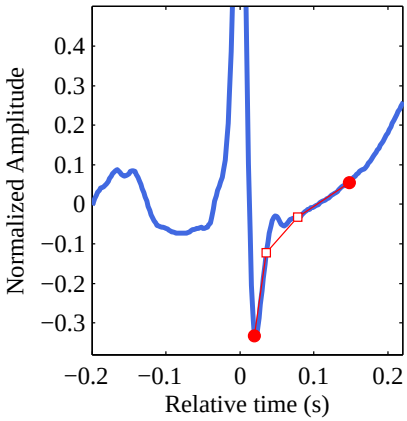


(c) Another potential S_f explored by CCM, with lower likelihood than (b).

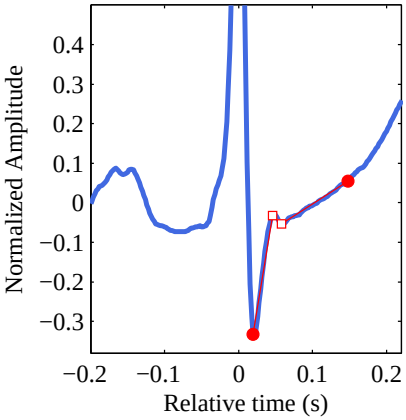
Figure 5.1: An example of CCM finding a local optimum. The QRS complex of Figure 4.22 with markers indicating the R′ peak (R'_p), S wave offset (S_f), and T wave onset (T_o). When evaluated by CCM, the locations indicated by square markers in (b) and (c) had a large error in their linear estimate of the signal resulting in a poor likelihood and low figure of merit. The center knot location shown in (a) is the locally optimal one for S_f, but is not at the desired location.



(a) The square markers indicate locations explored for R'_p and S_f by CCM, with poor likelihood.



(b) Another location explored for R'_p and S_f by CCM, with poor likelihood.



(c) CCM has identified the locations of R'_p and S_f with greatest likelihood.

Figure 5.2: Finding the global optimum by searching two coordinate axes simultaneously. This improvement to the CCM optimization algorithm would search two coordinate axes simultaneously, exploring all possible locations for R'_p and S_f (square markers) between S_p and T_o. Locations in (a) and (b) have large error in their linear estimate of the signal resulting in a poor figure of merit. The center knots in (c) have the best likelihood, are globally optimal, and correspond to the desired locations for R'_p and S_f.

The proposed improvement complements the existing two-dimensional probability density estimate explicitly with knot amplitudes, resulting in a joint time-amplitude-relevance prior. As before, the time dimension is the offset of the knot from the R wave peak, and the relevance dimension captures the curvature at the current point by incorporating the relative times and amplitudes of the three-tuple of knots. With this enhancement, these values are augmented with the amplitude of the knot expressed relative to the R wave peak.

There is a significant increase in computational cost, however, as the optimization algorithm now needs to consider different amplitudes as well as times because the knot under optimization is no longer constrained to lie on the signal. CCM must therefore search a two-dimensional region around a knot's location, and for each of these candidate locations use its time, amplitude, and relevance values to obtain the prior and compute the figure of merit.

Explicitly factoring amplitude priors into the optimization will give the algorithm a greater ability to deal with noise, as it will have a more complete representation of prior knowledge for the characteristic points. In addition to increased computation cost during optimization, more training data will also be required as a result of the increased dimensionality of the joint probability density to be estimated.

5.3.3 Algorithm characterization via synthesized waveforms

Using simulated waveforms in a test platform allows for a systematic, quantitative characterization of numerous aspects of the algorithm. Such a characterization could provide a more detailed understanding of the impact of priors than previously reported, and can determine the effect of precisely-controlled noise levels, of various types, on the algorithm's accuracy in estimating characteristic points.

Although the impact of priors is explored and reported in Sections 4.9 and 4.10, a future effort could increase the depth and breadth of that work. Specifically,

synthetic waveforms could be generated using knots drawn from distributions described by a given set of priors, and the algorithm's performance evaluated. To assess the impact of differences between the priors and actual distributions, subsequent runs could iteratively modify the distributions used to synthesize the waveforms to increasingly vary from the priors.

This effort would provide a objective indicator of characteristic point location accuracy as a function of the distance between the prior densities used in optimization and the densities used to simulate the test waveforms. The results of this investigation are important to understand how the algorithm behaves if a subject's morphology changes during monitoring, as might be the case for a degenerative disease state, *i.e.*, a healthy subjects develops indications of a cardiac condition while being monitored.

Another effort using simulated data could add calibrated amounts of bandlimited noise to signals under analysis, and compare algorithm's accuracy in estimating characteristic point locations with, and without, use of the priors. The intent of this exercise is to quantify the impact of priors in noisy settings.

There is a challenge in synthesizing waveforms in this manner, however. Drawing characteristic point locations from the priors used to initialize knot locations, as described in Section 4.4.1, is not sufficient for accurate signal synthesis.

Fundamentally, the problem arises from the fact that location priors do not preserve much of the information required to synthesize meaningful waveforms, and simulating waveforms using solely these priors often results in unrealistic morphologies. Even though drawn from valid distributions describing each point, the combination of times and amplitudes obtained in this manner often results in invalid curvatures. For example, due to the variances of characteristic point amplitudes, shown Figure 4.8, it is possible for a waveform peak to be less than its corresponding onset and offset value. Or the peak can be higher than the onset, but lower than the offset.

To synthesize a valid waveform represented by the underlying priors, the simulation must incorporate relative times and amplitudes of neighboring knots so that location has a proper curvature. Ideally, the curvatures should be drawn from the time-relevance joint density estimates obtained by KDE as described in Section 4.4.4. This can be achieved using a Monte Carlo algorithm, such as Gibbs sampling, to obtain samples corresponding to the desired distributions.

5.3.4 Nonparametric probability density estimate for likelihood

As described in Section 4.5, the likelihood $p(y|k)$ is a value reflecting how well the model, a linear interpolant defined by the current value of its parameters, the knots, represents an observed signal. In the effort described in Chapter 4 the error values defining the likelihood are modeled as being distributed normally with zero mean and an experimentally determined variance. Likelihood values are drawn from such a distribution for error values corresponding to each point of the interpolated estimate.

The assumption of normality, although convenient, may not be an accurate one. Future research should determine if better results are obtained by using likelihood values drawn from probability density estimates determined completely from the data. Such density estimates could better represent the physical reality of the errors and provide increased estimation accuracy.

This approach can be implemented by performing a kernel density estimation of the error values obtained for each manually specified knot location in the training set. It is completely analogous to the KDE approach used to determine the probability density estimates of the time-relevance priors used by the figure of merit for optimization, as described in Section 4.4.4. In this case, however, the KDE need only estimate a one-dimensional probability density: that corresponding to the errors in the interpolated estimate for each knot.

Using the notation of Section 4.5, the KDE is performed on error values calculated as the difference between the observed signal and its interpolated estimate for each manually specified knot. These are represented as $\epsilon_j = y_j - \hat{y}_j$, with $1 < j < m$, where m is the number of samples spanned by the three-tuple of knots, the center knot of which is being optimized.

During the optimization process, the figure of merit calculation obtains a likelihood by looking up the KDE value of the error in the interpolated estimated at the current knot location.

5.3.5 Explore the bias-variance tradeoff

Section 4.4.4 describes the development of the time-relevance priors from manual annotations of the training set. The resulting estimates of the joint probability densities are illustrated in Appendix A, and are used to provide the *a priori* probabilities for calculating figure of merit during optimization.

The granularity of the histograms, and the bandwidth used in calculating the KDE, determine the tradeoff between bias and variance in these non-parametric estimators. Increasing the bandwidth of the KDE will further smooth the estimate and decrease its variance at the expense of increased bias. Decreasing its bandwidth will reduce the estimator's bias, but increase its variance. The KDEs shown in Appendix A served reasonably well on the test set, but many exhibit very small localized maxima outside of the main modes, indicating potential undersmoothing.

A systematic analysis of the bias-variance tradeoff of these estimates would be beneficial and provide objective evidence regarding the degree of smoothing required for minimum error (both bias and variance) in the characteristic point estimates.

5.3.6 Expand support for other morphologies

There are a variety of morphologies not present in the training and test data used for this effort. Depending on the application of the framework, the subjects to be evaluated, and their health state, there may be need to incorporate priors supporting other ECG morphologies including negative or asymmetric P and T waves, notched T waves, and U waves.

The framework can readily be extended to support such morphological features. In the case of biphasic T waves, for example, it would be necessary to add an additional T peak knot in $\mathcal{C}$ intended to capture the second peak, and then re-derive priors from a training set which includes biphasic T waves. A similar approach could be adopted to identify U wave onset, peak and offset.

The implementation already supports biphasic RS complexes as long as there is a detectable R wave to allow relative knot representations as described in Equation (4.3). The implementation cannot represent inverted QS complexes which lack an R wave because the priors are calculated relative to the R wave peak. An implementation using a different fiducial point which does not rely on the R wave could address this limitation and allow more general complexes to be modeled.

5.3.7 Improving the priors

Using prior density estimates computed on a per-subject basis can improve algorithm performance in the presence of noise. When computed using data from only one individual, the prior probability estimates will have less variance since they are calculated using data exhibiting a much more limited set of morphologies.

As a result of this decreased variance they will have correspondingly more power over the likelihood term in the figure of merit, *i.e.*, the modes of the joint time-relevance density estimates shown in Appendix A will be sharper and result in greater differences in the priors for smaller changes in the time and relevance

values. However, implementing per-subject priors will cause them to be biased to the individual used for training, and as such they will almost certainly not work as well for other subjects.

5.3.8 Additional support knot

For some T wave morphologies an additional support knot will help produce more consistent and accurate estimation of the T wave offset characteristic point. There are several subjects in the training and test sets that could benefit from this improvement.

Specifically, for cases exhibiting a very steep offset wave, the existing support knot T_2 is insufficient. Much like the description provided for the QRS complex in Section 4.2.2, the downward waveform from the T wave peak can be sigmoidally shaped. Positioning the support knot at the midpoint of the peak and offset does help, but it is not enough to capture the curvature of the sigmoid.

As a result the error in the linear estimate adversely impacts the likelihood term, and causes the knot to be placed to the left of the desired location. An additional support knot on the downward slope of the T wave will allow the linear interpolant to represent the signal more accurately, preventing this problem.

5.3.9 Use of relevance following optimization

In the current implementation, relevance values are computed on the manually-annotated characteristic points in the training set to allow estimation of the *a priori* probability density. During optimization, they are calculated on knots only to determine the value of the time-relevance prior at the knot location under consideration as part of the figure of merit calculation.

After optimization is completed, the algorithm can determine presence or absence of characteristic points that may be lacking for a given ECG signal (such as

the peaks of the Q, S, and R′ waves and onset of the T wave) by using the relevance value of the corresponding optimized knot. A comparison of the relevance value against the bimodal density for these characteristic points allows automatic determination of presence of absence of these knots.

REFERENCES

[1] A. Ahmadian, S. Karimifard, H. Sadoughi, and M. Abdoli. An efficient piecewise modeling of ECG signals based on Hermitian basis functions. In *Proceedings of IEEE Engineering in Medicine and Biology Society*, pages 3180–3183. IEEE, 2007.

[2] M. Akhbari, M.B. Shamsollahi, and C. Jutten. ECG fiducial points extraction by extended Kalman filtering. In *2013 36th International Conference on Telecommunications and Signal Processing*, pages 628–632, July 2013.

[3] H. Al-Nashash. A dynamic Fourier series for the compression of ECG using FFT and adaptive coefficient estimation. *Medical engineering & physics*, 17(3):197–203, 1995.

[4] H. D. Ambos, J. Markham, and M.E. Cain. Use of fast Fourier transform analysis to detect patients prone to sustained ventricular arrhythmias. *Computers in Cardiology*, pages 3268–3270, 1984.

[5] R. V. Andreão and J. Boudy. Combining wavelet transform and hidden markov models for ECG segmentation. *EURASIP Journal on Applied Signal Processing*, 2007(1):95–95, 2007.

[6] A. Arafat and T. Hasan. Automatic detection of ECG wave boundaries using empirical mode decomposition. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 461–464, April 2009.

[7] S.F. Arnold. *Mathematical Statistics*. Prentice-Hall, 1990.

[8] F. Badilini, A.J. Moss, and E.L. Titlebaum. Cubic spline baseline estimation in ambulatory ECG recordings for the measurement of ST segment displacements. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 584–585, Oct 1991.

[9] D.E. Becker. Fundamentals of electrocardiography interpretation. *Anesthesia Progress*, 53(2):53–64, 2006.

[10] I.D. Berg and A. Wilansky. Periodic, almost-periodic, and semiperiodic sequences. *The Michigan Mathematical Journal*, 9(4):363–368, 1962.

[11] B. Bouchehamand, Y. Ferdi, and M.C. Baatouche. A characteristic points unified approach for ECG analysis and compression. In *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering*, pages 188–189, Oct 2003.

[12] M.G.S Bruno. Sequential Monte Carlo methods for nonlinear discrete-time filtering. *Synthesis Lectures on Signal Processing*, 6(1):1–99, 2013.

[13] M.E. Cain, H.D. Ambos, F.X. Witkowski, and B.E. Sobel. Fast-Fourier transform analysis of signal-averaged electrocardiograms for identification of patients prone to sustained ventricular tachycardia. *Circulation*, 69(4):711–720, 1984.

[14] Y. Cao. Bivariate Kernel Density Estimation, `gkde2 v2.1`. *MATLAB Central File Exchange,* `http://www.mathworks.com`, 2013.

[15] A. B. Carlson. *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*. McGraw Hill, 1975.

[16] L. Clavier and J.M. Boucher. Segmentation of electrocardiograms using a Hidden Markov Model. In *Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 1409–1410, Oct 1996.

[17] D.A. Coast, R.M Stern, G.G. Cano, and S.A. Briller. An approach to cardiac arrhythmia analysis using Hidden Markov Models. *IEEE Transactions on Biomedical Engineering*, 37(9):826–836, 1990.

[18] H.M. Cooper. A taxonomy of literature reviews. *Education Resources Information Center,* `http://eric.ed.gov/?id=ED254541`, 1985.

[19] H.M. Cooper. Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1):104–126, 1988.

[20] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, Inc., 1973.

[21] J. Fan and I. Gijbels. *Local Polynomial Modeling and its Applications*. Chapman and Hall, 1996.

[22] D.B. Fogel. *Evolutionary computation: toward a new philosophy of machine intelligence*, volume 1. John Wiley & Sons, Inc., 2006.

[23] J. Froning, V.F. Froelicher, and M.D. Olson. Application and limitations of continuous baseline estimation and removal using a cubic-spline technique during exercise ECG testing. In *Proceedings of the Computers in Cardiology Conference*, pages 537–540, 1987.

[24] F. Gaita, C. Giustetto, F. Bianchi, C. Wolpert, R. Xchimpf, R. Riccardi, S. Grossi, E. Richiardi, and M. Borggrefe. Short QT syndrome a familial cause of sudden death. *Circulation*, 108(8):965–970, 2003.

[25] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C. Peng, and H.E. Stanley. Physiobank, Physiotoolkit, and Physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[26] H. Gothwal, S. Kedawat, and R. Kumar. Cardiac arrhythmias detection in an ECG beat signal using fast Fourier transform and artificial neural network. *Journal of Biomedical Science and Engineering*, 4(04):289, 2011.

[27] F.G. Guilak. ECG waveform compression by recursive partitioning and decimated residual encoding. In *Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 523–524. IEEE, 1992.

[28] F.G. Guilak. High-quality, low-bit-rate method of compressing waveform data, August 12 1997. US Patent 5,657,398.

[29] F.G. Guilak and J. McNames. A spline framework for ECG analysis. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 957–960. IEEE, 2011.

[30] F.G. Guilak and J. McNames. A Bayesian-optimized spline representation of the electrocardiogram. *Physiological Measurement*, 34(11):1467, 2013.

[31] R. Gupta, M. Mitra, K. Mondal, and S. Bhowmick. A derivative-based approach for QT-segment feature extraction in digitized ECG record. In *Second International Conference on Emerging Applications of Information Technology*, pages 63–66, Feb 2011.

[32] M.A. Haque, M.E. Rahman, C.A. Al Sayeed, and B.M.Z. Uddin. A fast algorithm in detecting ECG characteristic points. In *Proceedings of Second International Conference on Electrical and Computer Engineering*, pages 26–28, 2002.

[33] H. Hsiao-Shu, C.K. Cheng, and G.J. Jan. A real-time QRS detection and delineation algorithm based on the inflection angle of the ECG waveform. In

Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume 1, pages 138–139, Nov 1989.

[34] H. N. Huang, S.Y. Wu, S.T. Chen, and C.Y. Hsu. QRS complex detection using cubic splines. In *Fourth International Conference on Genetic and Evolutionary Computing*, pages 787–790, Dec 2010.

[35] H. Imai, N. Kimura, and Y. Yoshida. An efficient encoding method for electrocardiography using spline functions. *Systems and Computers in Japan*, 16(3):85–94, 1985.

[36] J.M.Pastore, S.D. Girouard, K.R. Laurita, F.G. Akar, and D.S. Rosenbaum. Mechanism linking T-wave alternans to the genesis of cardiac fibrillation. *Circulation*, 99(10):1385–1394, 1999.

[37] G.B. Thomas Jr. *Elements of Calculus and Analytic Geometry.* Addison-Wesley Publishing Company., 1976.

[38] K. Kalovrektis, T. Ganetsos, N. Shammas, I. Taylor, and J. Andonopoulos. Development of a computerized ECG analysis model using the cubic spline interpolation method. In *Proceedings of the 5th WSEAS international conference on Circuits, systems and signals*, pages 186–189. World Scientific and Engineering Academy and Society (WSEAS), 2011.

[39] M. Karczewicz and M. Gabbouj. ECG data compression by spline approximation. *Signal Processing*, 59(1):43–59, May 1997.

[40] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*, volume 1. Prentice-Hall, 1993.

[41] E. Keogh, S. Chu, D. Hart, and M. Pazzani. Segmenting time series: A survey and novel approach. *Data mining in time series databases*, 57:1–22, 2004.

[42] A. Koski, M. Juhola, and M. Meriste. Syntactic recognition of ECG signals by attributed finite automata. *Pattern Recognition*, 28(12):1927–1940, 1995.

[43] S. Krimi, K. Ouni, and N. Ellouze. An approach combining wavelet transform and hidden markov models for ECG segmentation. In *3rd International Conference on Information and Communication Technologies: From Theory to Applications*, pages 1–6. IEEE, 2008.

[44] P.K. Kulkarni, V. Kumar, and H.K. Verma. Diagnostic acceptability of FFT-based ECG data compression. *Journal of medical engineering & technology*, 21(5):185–189, 1997.

[45] G. Lachiver, J.M. Eichner, F. Bessette, and W. Seufert. An algorithm for ECG data compression using spline functions. *Computers in Cardiology*, pages 575–578, 1986.

[46] P. Laguna, D. Vigo, R. Jane, and P. Caminal. Automatic wave onset and offset determination in ECG signals: Validation with the cse database. In *Proceedings of the Computers in Cardiology Conference*, pages 167–170, Oct 1992.

[47] P. Lancaster and K. Šalkauskas. *Curve and Surface Fitting, an Introduction.* Academic Press, 1990.

[48] C. Li, C. Zheng, and C. Tai. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42(1):21–28, Jan 1995.

[49] G.J. Li, X.N. Zhou, S.T. Zhang, and N.Q. Liu. ECG characteristic points detection using general regression neural network-based particle filters. In *International Symposium on Bioelectronics and Bioinformatics*, pages 155–158, Nov 2011.

[50] C. Lin, A. Giremus, C. Mailhes, and J.Y. Tourneret. Beat-to-beat P and T wave delineation in ECG signals using a marginalized particle filter. In *Proceedings of the 20th European Signal Processing Conference*, pages 479–483, Aug 2012.

[51] C. Lin, C. Mailhes, and J.Y. Tourneret. P- and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler. *IEEE Transactions on Biomedical Engineering*, 57(12):2840–2849, Dec 2010.

[52] O. Malgina, E. Plesnik, A. Kosir, M. Zajc, and J.F. Tasic. Methods for ECG signal compression with reconstruction via cubic spline approximation. In *1st Middle East Conference on Biomedical Engineering*, pages 287–291, Feb 2011.

[53] D.G. Manolakis, V.K. Ingle, and S.M Kogon. *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing*, volume 46. Artech House Norwood, 2005.

[54] J.P. Martínez, R. Almeida, S. Olmos, A.P. Rocha, and P. Laguna. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, 51(4):570–581, 2004.

[55] E.B. Mazomenos, T. Chen, A. Acharyya, A. Bhattacharya, J. Rosengarten, and K. Maharatna. A time-domain morphology and gradient based algorithm for ECG feature extraction. In *IEEE International Conference on Industrial Technology*, pages 117–122, March 2012.

[56] C.D. McManus, U. Teppner, D. Neubert, and S.M. Lobodzinski. Estimation and removal of baseline drift in the electrocardiogram. *Computers and Biomedical Research*, 18(1):1–9, February 1985.

[57] J. McNames. Personal communication, February 2008.

[58] P.E. McSharry, G.D. Clifford, L. Tarassenko, and L.A. Smith. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 50(3):289–294, 2003.

[59] S.S. Mehta, S.C. Saxena, and H.K. Verma. Recognition of P and T waves in electrocardiograms using fuzzy theory. In *Proceedings of the First Regional Conference of the IEEE Engineering in Medicine and Biology Society and 14th Conference of the Biomedical Engineering Society of India*, pages 2–54. IEEE, 1995.

[60] C.R. Meyer and H.N. Keiser. Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques. *Computers and Biomedical Research*, 10(5):459–470, October 1977.

[61] A.M. Mier-Muth and A.S. Willsky. A sequential method for spline approximation with variable knots. *International Journal of Systems Science*, 9(9):1055–1067, September 1978.

[62] K. Minami, Y. Ohkuma, H. Nakajima, and T. Toyoshima. Real-time ventricular arrhythmia detection with Fourier analysis and neural network. In *Proceedings of the Computers in Cardiology Conference*, pages 545–548. IEEE, 1996.

[63] S.K. Mukhopadhyay, M. Mitra, and S. Mitra. Time plane ECG feature extraction using hilbert transform, variable threshold and slope reversal approach. In *International Conference on Communication and Industrial Application*, pages 1–4, Dec 2011.

[64] I.S.N. Murthy and U.C. Niranjan. Component wave delineation of ECG by filtering in the Fourier domain. *Medical and Biological Engineering and Computing*, 30(2):169–176, 1992.

[65] I.S.N. Murthy and G.S.S.D. Prasad. Analysis of ECG from pole-zero models. *IEEE Transactions on Biomedical Engineering*, 39(7):741–751, July 1992.

[66] S.M.M. Naidu, P.C. Pandey, and V.K. Pandey. Automatic detection of characteristic points in impedance cardiogram. In *Proceedings of the Computing in Cardiology Conference*, pages 497–500, Sept 2011.

[67] National Oceanographic and Atmospheric Administration. Tidal datums. `http://tidesandcurrents.noaa.gov/datum_options.html`, 2013.

[68] National Oceanographic and Atmospheric Administration. Why do we study tides? `http://oceanservice.noaa.gov/facts/tidestudy.html`, 2014.

[69] P.V. O'Neil. *Advanced Engineering Mathematics*. Wadsworth Publishing Company, 1986.

[70] S.E. Paraskevopoulou, D.Y. Barsakcioglu, M.R. Saberi, A. Eftekhar, and T.G. Constandinou. Feature extraction using first and second derivative extrema (FSDE) for real-time and hardware-efficient spike sorting. *Journal of Neuroscience Methods*, 215(1):29–37, 2013.

[71] CSE Working Party. Recommendations for measurement standards in quantitative electrocardiography. *European Heart Journal*, 6(10):815–825, Oct 1985.

[72] Z. Pei-Yong and W. Hui-Yan. A framework for automatic time-domain characteristic parameters extraction of human pulse signals. *EURASIP Journal on Advances in Signal Processing*, pages 1385–1394, 1999.

[73] M. Perryman. *The Exoplanet Handbook*. Cambridge University Press, 2011.

[74] R.Q. Quiroga. Spike sorting. *Scholarpedia*, 2(12):3583, 2007. revision 137442.

[75] M. Rabus, R. Alonso, J.A. Belmonte, H.J. Deeg, R.L. Gilliland, J.M. Almenara, T.M. Brown, D. Charbonneau, and G. Mandushev. A cool starspot or a second transiting planet in the TrES-1 system? *Astronomy & Astrophysics*, 494(1):391–397, 2009.

[76] J.J. Randolph. A guide to writing the dissertation literature review. *Practical Assessment, Research & Evaluation*, 14(13):2, 2009.

[77] B.R. Reddy and I.S.N. Murthy. ECG data compression using Fourier descriptors. *IEEE Transactions on Biomedical Engineering*, BME-33(4):428–434, 1986.

[78] D. Sadhukhan and M. Mitra. Detection of ECG characteristic features using slope thresholding and relative magnitude comparison. In *Third International Conference on Emerging Applications of Information Technology*, pages 122–126, Nov 2012.

[79] I. Saini, D. Singh, and A. Khosla. Delineation of ECG wave components using k-Nearest Neighbor (KNN) algorithm: ECG wave delineation using KNN. In *Tenth International Conference on Information Technology: New Generations*, pages 712–717, April 2013.

[80] S.L. Salas and E. Hille. *Calculus: One and Several Variables*. John Wiley & Sons, Inc., 1978.

[81] S.K. Salih, S.A. Aljunid, S.M. Aljunid, O. Maskon, and A. Yahya. High speed approach for detecting QRS complex characteristics in single lead electrocardiogram signal. In *IEEE International Conference on Control System, Computing and Engineering*, pages 391–396, Nov 2013.

[82] O. Sayadi and M.B. Shamsollahi. Model-based ECG fiducial points extraction using a modified extended Kalman filter structure. In *First International Symposium on Applied Sciences on Biomedical and Communication Technologies*, pages 1–5, Oct 2008.

[83] O. Sayadi and M.B. Shamsollahi. Model-based fiducial points extraction for baseline wandered electrocardiograms. *IEEE Transactions on Biomedical Engineering*, 55(1):347–351, Jan 2008.

[84] O. Sayadi and M.B. Shamsollahi. A model-based Bayesian framework for ECG beat segmentation. *Physiological Measurement*, 30:335–352, 2009.

[85] R.J. Schalkoff. *Pattern Recognition: Statistical, Structural, and Neural Approaches*. John Wiley & Sons, Inc., 1992.

[86] H.I. Shahein and H.M. Abbas. ECG data compression via cubic-splines and scan-along polygonal approximation. *Signal processing*, 35(3):269–283, 1994.

[87] A. Shayei, S.P. Ehsani, and M. Shabany. Efficient implementation of real-time ECG derived respiration system using cubic spline interpolation. In *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, pages 1083–1086, May 2013.

[88] Philips Healthcare Web Site. Adaptive Aerosol Delivery (AAD) product description. `http://www.healthcare.philips.com`, April 2015.

[89] E. Skordalakis. Syntactic ECG processing - a review. *Pattern Recognition*, 19(4):305–313, 1986.

[90] R.P. Sloan, P.A. Shapiro, E.E. Gorenstein, F.A. Tager, C.E. Monk, P.S. McKinley, M.M. Myers, E. Bagiella, I. Chen, and R. Steinman. Cardiac autonomic control and treatment of hostility: a randomized controlled trial. *Psychosomatic medicine*, 72(1):1–8, 2010.

[91] E. Soria, M. Martinez, J. Calpe, J.F. Guerrero, J.V. Frances, and J. Espf. Adaptive system for the location of characteristic points in an ecg. In *Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 1626–1627, Oct 1996.

[92] R.H. Stewart. *Our Ocean Planet — Oceanography in the 21st Century*. Texas A&M University, 2005.

[93] S.M.J.M. Straus, J.A. Kors, M.L. De Bruin, C.S. van der Hooft, A. Hofman, J. Heeringa, J.W. Deckers, J.H. Kingma, M.C.J.M. Miriam, B.H. Stricker, et al. Prolonged QTc interval and risk of sudden cardiac death in a population of older adults. *Journal of the American College of Cardiology*, 47(2):362–367, 2006.

[94] Y. Sun, K.L. Chan, and S.M. Krishnan. Characteristic wave detection in ECG signal using morphological transform. *BMC cardiovascular disorders*, 5(1):28, 2005.

[95] Y. Suzuki, K. Itakura, S. Saga, and J. Maeda. Signal processing and pattern recognition with soft computing. *Proceedings of the IEEE*, 89(9):1297–1317, Sep 2001.

[96] S.M. Szildgyi, L. Szildgyi, and L. David. Comparison between neural-network-based adaptive filtering and wavelet transform for ECG characteristic points detection. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 272–274, Oct 1997.

[97] S. Tang, C. Zhang, S. Wang, and M. Wan. A preliminary study for a slantwise-placed electroglottography. *Journal of Voice*, 29(1):129.e19–129.e27, 2015.

[98] R.H.D. Townsend, Th. Rivinius, J.F. Rowe, A.F.J. Moffat, J.M Matthews, D. Bohlender, C. Neiner, J.H. Telting, D.B. Guenther, T. Kallinger, et al. MOST observations of σ Ori E: Challenging the centrifugal breakout narrative. *The Astrophysical Journal*, 769(1):33, 2013.

[99] P. Trahanias and E. Skordalakis. Syntactic pattern recognition of the ECG. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):648–657, 1990.

[100] V.S Vassiliadis and R. Conejeros. Cyclic coordinate method. In C.A. Floudas and P.M Pardalos, editors, *Encyclopedia of Optimization*, pages 595–596. Springer US, 2009.

[101] V. Di Virgilio, C. Francaiancia, S. Lino, and S. Cerutti. ECG fiducial points detection through wavelet transform. In *Proceedings of the 17th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 1051–1052, Sep 1995.

[102] H.J.L.M. Vullings, M.H.G. Verhaegen, and H. Verbruggen. Automated ECG segmentation with dynamic time warping. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 163–166. IEEE, 1998.

[103] R.B. Wallace, R.M. Dansereau, and R.A. Goubran. Methods for the detection of ECG characteristic points. In *Medical Measurements and Applications Proceedings (MeMeA), 2012 IEEE International Symposium on*, pages 1–6, May 2012.

[104] P.J. Wang and N.A.M. Estes III. Supraventricular tachycardia. *Circulation*, 106(25):e206–e208, 2002.

[105] Q. Wang and Z.W. Liu. Detection of the pulse waveform characteristic points by wavelet transform using multiscale differential operator. In *Second International Conference on Biomedical Engineering and Informatics*, pages 1–5, Oct 2009.

[106] Z. Wang, Y.Ding, D.F. Parham, and K. Lee. Cycle identification and artifact detection in tidal breathing signals. In *Fifth International Conference on Bioinformatics and Biomedical Engineering*, pages 1–4, May 2011.

[107] W.A. Winner. Introduction to kernel density estimation. *Defense Technical Information Center,* `http://www.dtic.mil`, 1985.

[108] E. Zoghlami, E.P. Ayari, R. Tielert, and N. Wehn. A new fitting approach for online electrocardiogram component waves delineation. In *Proceedings of the Computers in Cardiology Conference*, pages 861–864, Sept 2008.

Appendix A

CHARACTERISTIC POINT PRIORS

This appendix includes figures of the histograms and kernel density estimates (KDE) used to model the joint time-relevance probability densities of all the characteristic points in the set $\mathcal{C}$. The histograms and KDEs were obtained from manual annotations on the training data as described in Section 4.4.4. In all figures, t is the time of the characteristic point in seconds expressed relative to the R wave peak, and ρ is its relevance as defined by Equation 4.9.

$$\mathcal{C} = \{P_o, P_p, P_f, Q_o, Q_p, S_p, R'_p, S_f, T_o, T_p, T_f\}$$

P_o — onset of the P wave as the signal increases from its isoelectric level.

P_p — the peak value of the P wave.

P_f — offset of the P wave as the signal returns to its isoelectric level.

Q_o — onset of the QRS complex.

Q_p — the negative peak of the Q wave, which is not always present.

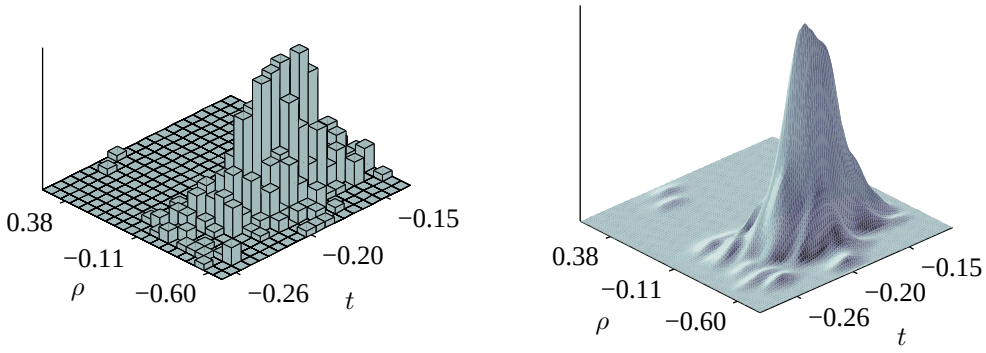
S_p — the negative peak of the S wave, which is not always present.

R'_p — the typically small peak of the R′ wave, which is not always present.

T_o — onset of the T wave, which is often not discernible.

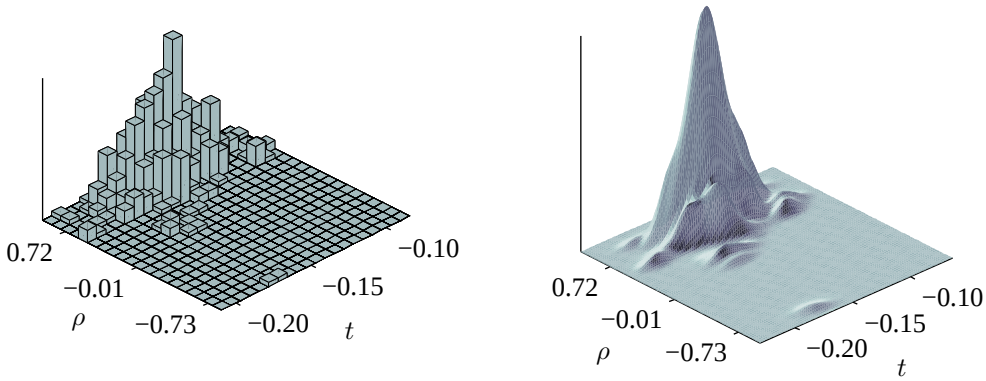
T_p — the peak value of the T wave.

T_f — offset of the T wave as the signal returns to its isoelectric level.



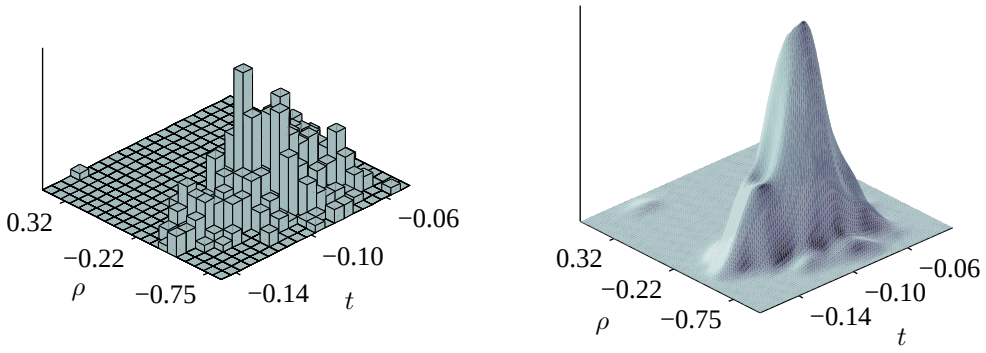
(a) P_o histogram. (b) P_o KDE.

Figure A.1: Histogram and KDE estimating the joint time-relevance prior probability density of the P wave onset characteristic point P_o.



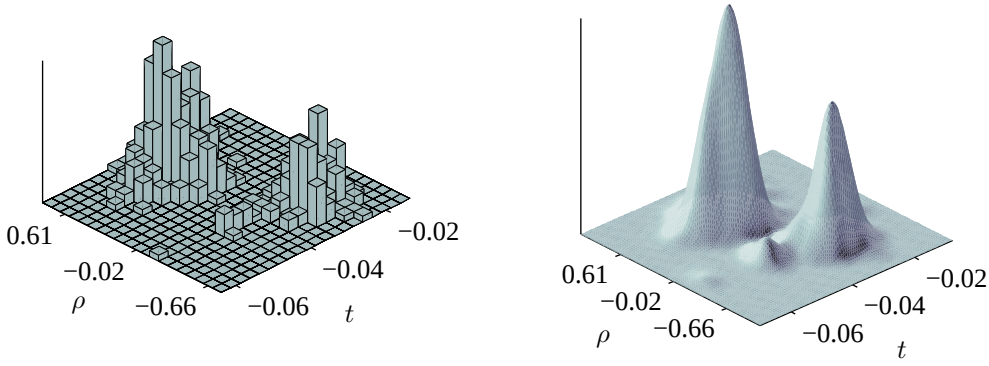
(a) P_p histogram. (b) P_p KDE.

Figure A.2: Histogram and KDE estimating the joint time-relevance prior probability density of the P wave peak characteristic point P_p.



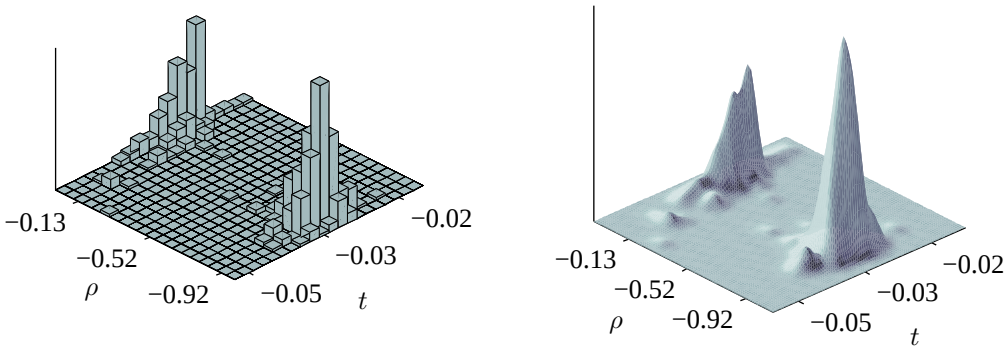
(a) P_f histogram. (b) P_f KDE.

Figure A.3: Histogram and KDE estimating the joint time-relevance prior probability density of the P wave offset characteristic point P_f.



(a) Q_o histogram. (b) Q_o KDE.

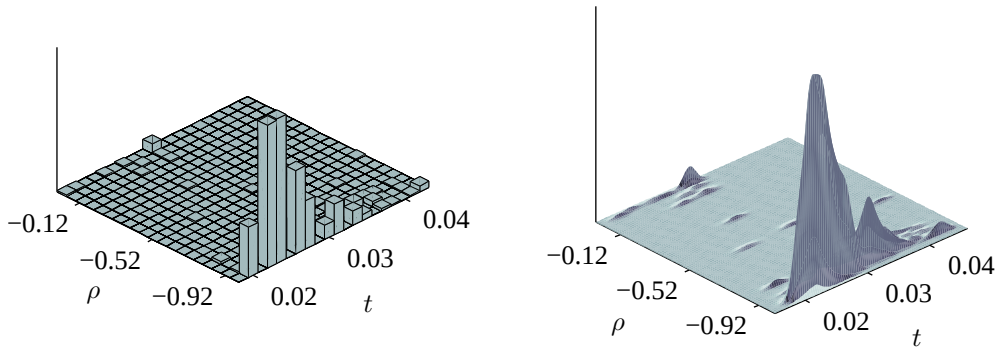
Figure A.4: Histogram and KDE estimating the joint time-relevance prior probability density of the Q wave onset characteristic point Q_o. The mode in the density estimates at $\rho \approx 0.6$ results from signals *with* a Q wave peak Q_p, and indicates a moderate downward curvature of the signal from its isoelectric level toward the negative peak. The mode at $\rho \approx -0.3$ results from signals *without* Q_p, and indicates a mild upward curvature at QRS onset toward the R wave peak.



(a) Q_p histogram.

(b) Q_p KDE.

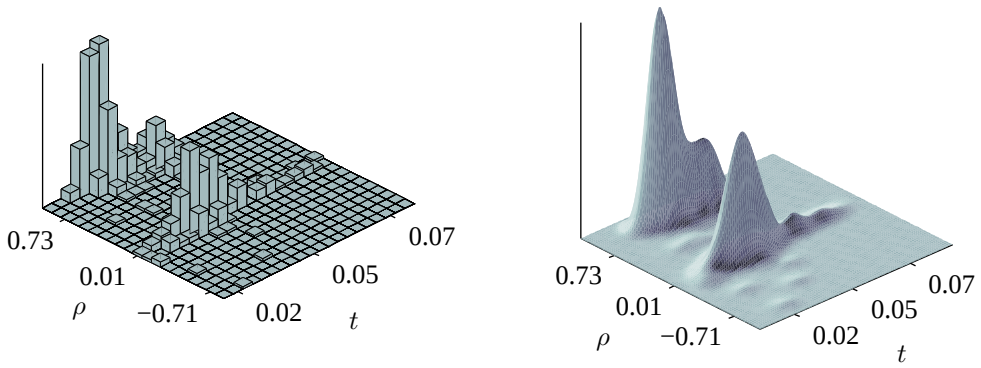
Figure A.5: Histogram and KDE estimating the joint time-relevance prior probability density of the Q wave peak characteristic point Q_p. These estimates are bimodal, reflecting beats in the training data with, and without, the Q wave peak. The mode at $\rho \approx -0.9$ indicates a very sharp concave up waveform corresponding to presence of Q_p. The mode at $\rho \approx 0.05$ corresponds to complexes in the training set that did not have a Q wave peak, and for which the corresponding knot was on the linear part of the ascending QR segment.



(a) S_p histogram.

(b) S_p KDE.

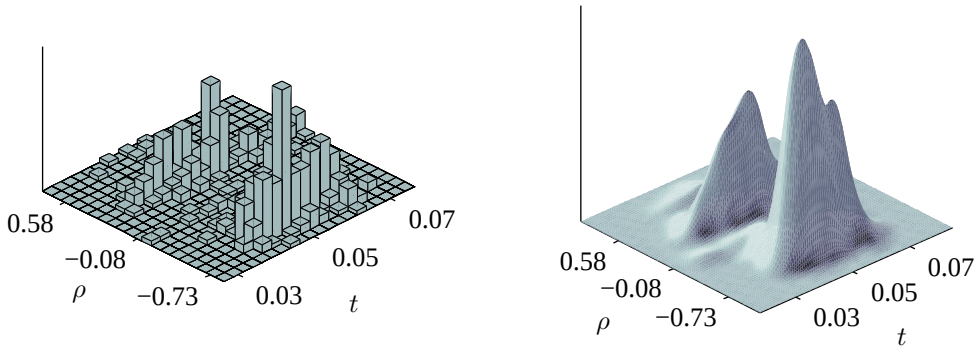
Figure A.6: Histogram and KDE estimating the joint time-relevance prior probability density of the S wave peak characteristic point S_p. Although it is not uncommon for ECG waveforms to be missing an S wave peak, virtually all of the training data used in this effort exhibited S_p. As a result its density estimate is not bimodal like those of Q_p and R'_p.



(a) R'_p histogram.

(b) R'_p KDE.

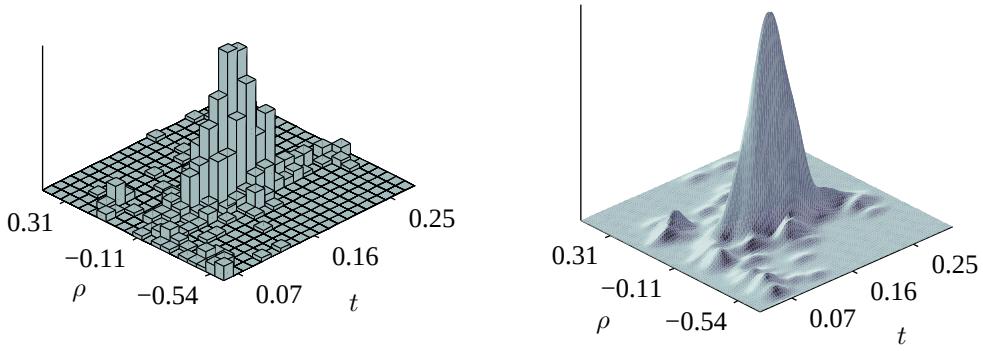
Figure A.7: Histogram and KDE estimating the joint time-relevance prior probability density of the R′ wave peak characteristic point R'_p. The R′ wave may be missing in some waveforms of the training data set, resulting in a bimodal density. The mode at $\rho \approx 0.8$ indicates a sharp concave down peak corresponding to the presence of the positive R′ peak. The mode at $\rho \approx 0.01$ reflects complexes in the training set without this feature, and for which the corresponding knot was on the linear part of the waveform.



(a) S_f histogram.

(b) S_f KDE.

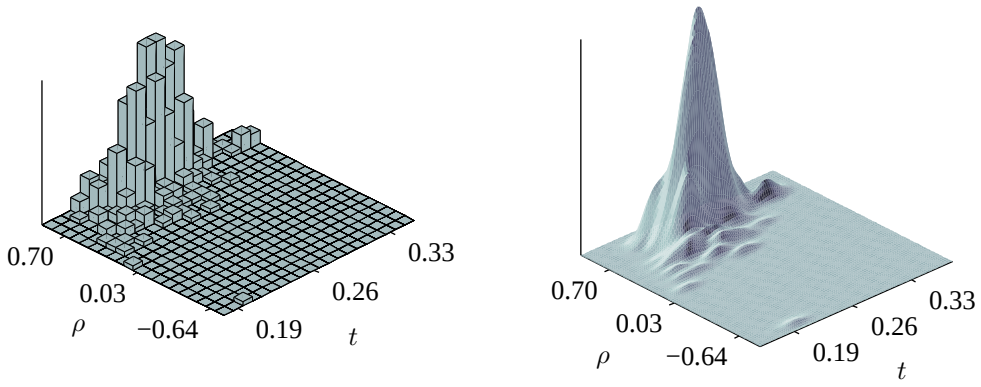
Figure A.8: Histogram and KDE estimating the joint time-relevance prior probability density of the QRS offset characteristic point S_f. The mode in the density estimates at $\rho \approx -0.6$ indicates a moderate curvature down to the isoelectric level from the R′ peak when it is present, or from the R peak (when the signal is lacking both S and R′ peaks). The mode at $\rho \approx 0.4$ indicates a mild curvature to the isoelectric level from the negative S wave peak. It results from waveforms exhibiting the S wave but lacking an R′ wave.



(a) T_o histogram.

(b) T_o KDE.

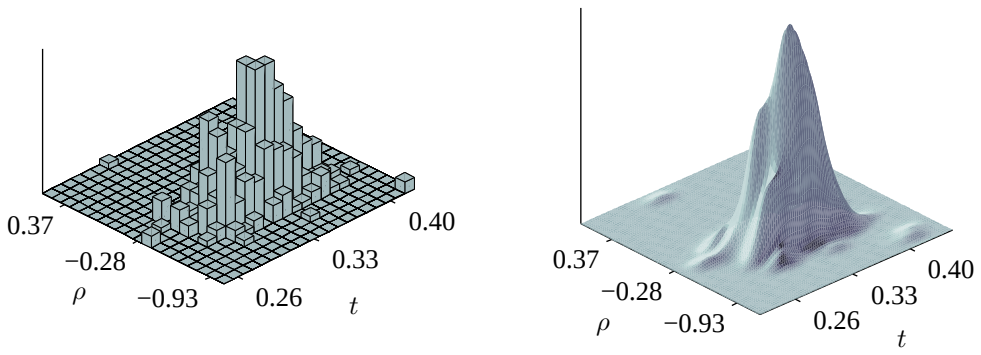
Figure A.9: Histogram and KDE estimating the joint time-relevance prior probability density of the T wave onset characteristic point T_o. The T wave onset does not exist for many waveforms in the training set, leading to an expectation of a bimodal density with modes reflecting presence and absence of this characteristic point. However, the bin width and bandwidth parameters used for the histogram and KDE, respectively, preclude discrimination of very low relevance values associated when the onset exists from the near-zero relevance value when it does not. As a result, the estimate is unimodal.



(a) T_p histogram.

(b) T_p KDE.

Figure A.10: Histogram and KDE estimating the joint time-relevance prior probability density of the T wave peak characteristic point T_p.



(a) T_f histogram.

(b) T_f KDE.

Figure A.11: Histogram and KDE estimating the joint time-relevance prior probability density of the T wave offset characteristic point T_f.

Appendix B

DATA

The data used in this effort comprised lead II ECG signals from 176 healthy 20–45 year-olds as reported in [90]. The signals were captured during a precisely-controlled psychophysiology protocol which included physical, mental, and emotional stressors. The mental stressors were arithmetic and Stroop color-word tests, the emotional stressors comprised preparation for and presentation of a short speech, and the physical stressor was a 70° head-up tilt.

Data for each subject includes approximately one hour of ECG signal, between two and five minutes captured during each phase of the stressor protocol, including baseline and recovery periods. Baseline periods were intended to establish individuals' behavior in periods without active stressors and recovery periods were to allow time for recovery between stressors. The ECG signal was sampled at 500 Hz and digitized with a resolution of 16 bits.

The training set was constructed by first randomly selecting 40 of the 176 subjects. The the `ecgpuwave` QRS detector [25] was run on their full ECG signals to locate beats and establish the precise location of their R wave peaks. This detector was chosen as it is well-established in the literature, and in addition to detecting the R wave also provides onset, peak, and offset locations for the P and T waves. Its estimation of characteristic points for the P and T waves, however, was not as accurate as the manual annotations, so this functionality of the detector was not used.

After running the QRS detector, ten beats were extracted from each of the 40 subjects' data by selecting one beat randomly from each of the ten phases of

the protocol, resulting in a training set comprising 400 beats. The beat extraction algorithm employed a commonly used, simple means to determine the waveform boundaries between successive beats: the midpoint in time between adjacent R wave peaks. So the start and end points for extracting a beat were set to one-half the beat-to-beat interval to the preceding and following beats, respectively.

The first set of 200 beats of the training set was manually annotated by one reviewer, and the second set of 200 beats was manually annotated by two reviewers. In total this provides up to 600 potential manual annotations of points in $\mathcal{C}$, although due to noise, ambiguity, and missing features in the training data there are fewer than 600 annotations for all points. These were normalized to the R wave location and amplitude as defined by Equation (4.3). The location priors and time-relevance priors were computed from the locations of these manually-specified characteristic points as described in Section 4.4.

To prospectively evaluate the algorithm performance another 20 subjects were randomly selected, explicitly different than those chosen for the training set. As with the training data, a single beat was extracted at random from each of the 10 phases of these 20 subjects' data to provide a total of 200 beats. All 200 beats in the test set were manually annotated by two reviewers, resulting in 400 sets of manual annotations of points in $\mathcal{C}$. Algorithm assessment was based on differences in knot locations determined by the algorithm and the two full sets of independent manual annotations, as described in Section 4.8.