Computationally-efficient realtime interpolation algorithm for non-uniform sampled biosignals

Onur Guven¹, Amir Eftekhar¹, Wilko Kindt², and Timothy G. Constandinou¹

¹Department of Electrical & Electronic Engineering, Imperial College, South Kensington Campus, London, SW7 2AZ, UK
²Texas Instruments Corporation

Published in Healthcare Technology Letters; Received on 11/08/2015; Revised on 01/02/2016; Accepted 29/02/2016.

This letter presents a novel, computationally-efficient interpolation method that has been optimised for use in ECG baseline drift removal. In our previous work 3 isoelectric baseline points per heart beat are detected, and here utilised as interpolation points. As an extension from linear interpolation, our algorithm segments the interpolation interval and utilises different piecewise linear equations. Thus the algorithm produces a linear curvature that is computationally efficient while avoiding overshoots on nonuniform samples. The proposed algorithm is tested using sinusoids with different fundamental frequencies from 0.05 Hz to 0.7 Hz and also validated with real baseline wander data acquired from the MIT-BIH Noise Stress Database. The synthetic data results show an RMS error of 0.9 µV (mean), 0.63 µV (median) and 0.6 µV (std. dev.) per heart beat on a 1 mVp-p, 0.1 Hz sinusoid. On real data we obtain an RMS error of 10.9 µV (mean), 8.5 µV (median) and 9.0 µV (std. dev.) per heart beat. Cubic spline interpolation and linear interpolation on the other hand shows 10.7 µV, 11.6 µV (mean), 7.8 µV, 8.9 µV(median) and 9.8 µV, 9.3 µV (std. dev.) per heart beat respectively.

1. Introduction: Interpolation is a method of constructing new data points within the range of a discrete dataset. It is a problem that dates back to ancient civilisations, who were known to use interpolation methods for analysing astronomical data [1]. The mathematical basis of this method was not defined till later, as in the work of Edward Waring [2], which is today attributed to Lagrange.

Lagrange polynomials define the least degree of polynomial curves that pass through a given set of coordinates \( x_i, y_i \). However, as the order of Lagrange polynomials increase, any small perturbations in coordinates results in large overshoots at the end points as known in literature as the Runge Phenomenon [3]. These oscillations may have no relation to the true nature of the overall function itself and without rigorous error monitoring higher order polynomial interpolations degrade accuracy as well as increase complexity of the algorithm.

Later, cubic spline \( (3^{rd} \text{ order}) \) functions were defined [4]. These polynomials are smoothly connected to each other at the coordinates \( x_i, y_i \) and since their continuous first and second derivative exists everywhere, the overall generated curve is smooth. However, spline interpolation algorithms rely on matrix inversion techniques for computing coefficients. Therefore, they are computationally demanding, and although efficient, they are not adaptable to real-time systems without windowing techniques. More adaptable are Rifman’s [5] and Keys’ [6] cubic convolution interpolation methods which involve fitting piecewise cubic polynomials (kernels) within intervals. Similar to cubic splines, these methods are computationally complex and not suitable for certain real-time system designs.

There are of course several algorithms and methods throughout the literature, aiming to approximate smoother curves and better fits. However for real-time systems challenges still exist to balance complexity vs. accuracy, and to allow adaptability to changing signal dynamics. The latter especially the case in biological signal applications.

One example of such a biological signal is the electrocardiogram (ECG). Being prone to interference from physiological and environmental sources has made ambulatory ECG, with a clinical accuracy, a challenge. Techniques exist to remove each of these noise sources, however on occasions where signal integrity is crucial these methods do not meet clinical standards. As discussed in our previous work [7], baseline wander can be removed by detecting fiducial points and estimating the baseline wander by interpolating through those points with a piecewise cubic hermite interpolation (PCHIP). However, PCHIP is still somewhat complex, therefore, a new method is investigated where baseline wander estimation is accurately achieved with less computational hardware resources required.

This letter presents a new interpolation algorithm that allows a better tradeoff between computational efficiency and signal distortion than prior methods. ECG signals are used as our test application, wherein we measure the distortion of the ST segment (an indicator of heart malfunction) while estimating the baseline wander. These baseline wander signals are low frequency noise artefacts that can be modelled as sinusoids with amplitudes up to 300 µV. The letter is organised as follows: Section 2 describes the overall system concept and methods; Section 3 describes the artificial and real test datasets; Section 4 presents and discusses results with complex algorithms; Section 5 concludes the letter.

2. Methodology: The overall methodology to the proposed algorithm is illustrated in Fig. 1. The main purpose of the algorithm is to estimate curvatures (turning points) with a better approximation than linear interpolation and in other cases simply use linear interpolation to reduce computational complexity. The algorithm is divided into two stages: (1) Turning point detection and (2) Weighted Piecewise Linear Interpolation (WPL)

2.1. Turning Point Detection: First we define the slopes, \( M_i \), between adjacent interpolation points. These slopes are then used to determine if a turning point exists. Here we utilise two criteria. The first condition checks if the slopes change sign such that either a local/absolute minima or maxima exists. However, this condition on its own is not enough to capture all turning points such that on occasions, when adjacent slopes do not change sign, there might be possible curvatures like during \( M_i \) instant as shown in Fig. 1. Therefore, a second condition is required such that even when the slopes do not change sign, these turning points are detected accurately. We have found that when the magnitude of adjacent slopes satisfies Condition 2 in Eq. 1, the algorithm accuracy improves even though no local/absolute minima or maxima is
2.2. Interpolation Methods:

2.2.1. Linear Interpolation: This method only requires current slope, \( M_i \), and a fraction of this slope is added for every interpolation point in between \( y_i \) and \( y_{i+1} \). Therefore, the number of operations required is minimal and the algorithm can interpolate both uniformly and non-uniformly sampled data since interpolation is based on addition operation and the only condition is to meet \( x_{i+1}, y_{i+1} \) coordinates. In Fig. 1 linear interpolation occurs at intervals \( M_{1,2,4,5,6} \).

2.2.2. Weighted Piecewise Linear Interpolation: An improvement to linear interpolation, is achieved when a turning point is detected as shown in Fig. 1. Following this detection, the distance between \( x_i \) and \( x_{i+1} \) is calculated and this interval is divided into 3 equal smaller segments. A counter checks this segment distribution and on events where the distance can not be divided accurately, a compensation factor is added to the final sample such that \( x_{i+1}, y_{i+1} \) coordinates are met. As mentioned in linear interpolation, this characteristic shows that both uniformly and non-uniformly sampled data can be interpolated and in each of these segments, weighted piecewise linear interpolation is achieved where every clock cycle, the corresponding segment slope \( H_{i-1} \), \( H_i \), \( H_{i+1} \) is added to the previous sample such as in linear interpolation. The first segment’s slope, \( H_{i-1} \), is the average of \( M_{i-1} \) and \( M_i \), which estimates the concavity/convexity with its past knowledge. The second slope, \( H_i \), is defined as \( M_i \), and the last \( H_{i+1} \) is shown as defined in Eq. 2. The error function of the WPL interpolation in this case are bounded by \( M_i \) and \( M_{i+1} \) slopes and although limited to three segments, the algorithm could be segmented further with increased complexity.

\[
\begin{align*}
\text{Condition 1} & \rightarrow M_{i-1} > 0 & M_i < 0 & M_{i-1} < 0 & M_i > 0 \\
\text{Condition 2} & \rightarrow \frac{3}{4} \left| M_{i-1} \right| > \left| M_i \right| & \frac{3}{4} \left| M_i \right| > \left| M_{i-1} \right|
\end{align*}
\]

\[H_{i+1} = \frac{M_{i-1} + M_i}{2} \]  

3. Test Data: To test our algorithm we use two sets of data: synthetic and real data. The former models ECG baseline wander while the latter is real data that we shall describe. We first generate interpolation points that are realistic isoelectric fiducial points that define the baseline wander [7]. These fiducial points are generated over 2243 heartbeats of the MIT-BIH Arrhythmia Database signal 100m.mat. These points are therefore realistic representations of non-uniformly distributed interpolation points for baseline wander estimation and are therefore used on both synthetic and real data.

3.1. Synthetic Data: Baseline wander can be modelled as a sinusoid around 0.15 – 0.3 Hz [8] that increases with exercise. Therefore, synthetic datasets are generated with 16 sinusoids with each fundamental frequencies that change from 0.05 Hz up to 0.7 Hz corresponding to a respiration rate of 3 to 42 per minute respectively. These sinusoids are sampled at 360 Hz and last for 30 min (i.e. \( \approx 650k \) samples).

3.2. Real Data: Real data is obtained from the MIT-BIH Noise Stress Database [9]. These datasets (BWM1.mat and BWM2.mat) are baseline wander recordings, sampled at 360 Hz with a gain of

Figure 1. Weighted piecewise interpolation methodology showing: (a) example input signal; (b) algorithm flowchart; (c) illustration of concept.

Figure 2. Synthetic data: Mean & standard deviation of RMS errors per heart beat of different interpolation methods.
Table 1 Real data - RMS and maximum error per heartbeat and ST segment

<table>
<thead>
<tr>
<th>Interpolation method</th>
<th>Signal (Hz)</th>
<th>RMS error (µV) per heartbeat†</th>
<th>Max. error (µV) per ST segment†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>µ</td>
<td>median</td>
</tr>
<tr>
<td>Linear</td>
<td>BWM1</td>
<td>14.8</td>
<td>10.6</td>
</tr>
<tr>
<td></td>
<td>BWM2</td>
<td>8.4</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>BWM1</td>
<td>13.5</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>BWM2</td>
<td>7.9</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>BWM1</td>
<td>13.7</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>BWM2</td>
<td>8.1</td>
<td>6.9</td>
</tr>
</tbody>
</table>

†2243 Heartbeats detected via MIT-BIH Arrhythmia Database (100m.mat)

Figure 3 Comparison of our algorithm with linear interpolation using a 1 V_{p-p} sinusoidal signal. Signals are denoted as sinusoidal (Green), linear interp. (Blue), WPL interp. (Red) Shown are: (a) 0.3 Hz sinusoid response; (b) sample by sample error analysis (linear vs WPL interp.); (c) 0.5 Hz sinusoid response; (d) sample by sample error analysis (linear vs triangular interp.).

200 V/V. Each recording lasts for 30 min (i.e. ≈650k samples) and the FFT of these signals show that the respiration frequency is mostly around 0.1 Hz with white Gaussian noise is present throughout the whole sample set. Due to this white noise, the baseline wander signals have been filtered with a 16-point moving average filter. prior to testing. This filter order was deemed sufficient to reduce the white noise below 5µV in worst conditions. Otherwise the noise floor is defined by the white noise within the recorded signal itself making it impossible to test interpolation methods thoroughly.

4. Results & Discussion:

4.1. Synthetic data: Fig. 2 compares three different interpolation methods: linear, cubic spline and our proposed interpolation method, when applied to sinusoidal baseline wander signals. Among all frequencies, weighted piecewise linear interpolation yields better accuracy when compared to linear interpolation alone. However, as the frequency of the sinusoids increase, all three algorithms’ performance degrades. This is due to there being less interpolation points per period available. When the respiration rate increases, this reflects an increase in pulse rate, maintaining a ratio of approximately 1 breath for every 3-4 heartbeats [10] and upon successful detection of fiducial points as in our previous work [7] at least 9 to 12 interpolation points should be located per period of the baseline wander. Using the 100m.mat signal however, the heart rate of the patient is around 72 bpm, and as the frequency of the synthetic data increases less interpolation points can be used per period therefore we see a degradation in accuracy.

4.2. Real data (MIT-BIH): As described, the algorithm is also tested on baseline wander signals acquired from the MIT-BIH Noise Stress Database. Table 1 shows that BWM1 and BWM2 results differ. When we observed these two signals BWM1 signal has a higher standard deviation (93 vs 36) and a higher kurtosis (15.6 vs 4.3) indicating BWM1 signal varies more in amplitude and peakedness which would make it more difficult to interpolate. Possible causes of this variance can be due to gender, stress test conditions, lung capacity since baseline wander occurs due to the impedance change seen by the amplifier as mentioned in the works of Friesen and et.al. [8]. Also, when comparing interpolation methods, we focused on both RMS and maximum errors since the baseline wander can be modelled as a sinusoid, RMS error would carry good measure of its effect, whereas maximum error seen during ST segment carries crucial information. As mentioned in Section 3.2, the fundamental frequency of these signals is mostly around 0.1 Hz. On occasions where the respiration rate increases, the errors become more comparable to the residual Gaussian noise errors present. Fig. 4 shows a 0.12 Hz respiration signal with residual Gaussian noise comparable to the error results at 0.4 Hz respiration rate. This is due to the fact that, since less interpolation points can be used, any high frequency content can not be captured due to the Nyquist sampling theorem. Therefore, not all of the errors reported in Table 1 are due to interpolation errors.

Fig. 5 shows that for almost all curves the algorithm results in smaller errors than linear interpolation except for one case, where an overshoot occurs, whereas Fig. 6 shows that the histogram of errors is more spread for linear interpolation, while WPL interpolation’s spread more closely resembles that of cubic spline interpolation.
interpolation. In addition, these figures comply with American Heart Association (AHA) and International Electrotechnical Commission (IEC) standards which allow a maximum error of 100 µV for clinical ECG systems [11]. Also, Fig. 6 histogram results and Table 1 results show that, RMS errors per heartbeat are in accordance with maximum ST segment errors. Even though all of these methods comply with the standards, in the event of missing fiducial point detections these errors would increase. This is also similar to the Fig. 2 result; as the fiducial point count remained constant, frequency increase of the baseline wander degraded system performance due to decreased sampling rate.

4.3. Complexity: As the linear interpolation takes only two coordinates, the complexity of the WPL interpolation and cubic spline increases due to past knowledge requirements. In the case of WPL interpolation the complexity of the algorithm is an additional slope calculation, interval segmentation and piecewise slope generation. As fiducial points are non-uniformly sampled depending on heart rate and characteristics (P, QRS and T waves) as mentioned in [7], an accurate complexity measure is hard to achieve. But under normal conditions an estimation of interpolation point generation per 100 sample is a reasonable estimate. Table 2 shows the complexity requirements for each interpolation method under this assumption. As can be seen WPL interpolation requires 8 if statements, 8 additions, 2 shift operations and 4 multiplications to generate a piecewise interpolation. The actual computational requirement on the other hand is much lower since the algorithm utilises these resources only when a turning point is detected. When we quantify the complexity measure of cubic spline interpolation, a single sample generation requires 14 floating point multiplications, 8 if statements, 8 additions, 2 shift operations and 4 multiplications to generate a piecewise interpolation. Therefore, the polynomial approach needs much more complexity, however the advantages in return such as the continuity of the interpolation estimation gets disturbed by the quantisation noise and the accuracy results do not show an effective improvement.

5. Conclusion: In this letter, we have described a computationally efficient interpolation algorithm that is suitable for real-time ECG baseline wander estimation. Using both synthetic and real data (from the MIT Noise Stress Database), we have shown that WPL interpolation is more accurate than linear interpolation, more computationally efficient than cubic spline interpolation and in compliance with clinically valid diagnosis.

6. Acknowledgment: This work was in part supported by Texas Instruments Corporation.

References

Table 2 Comparison in computational complexity between different interpolation methods

<table>
<thead>
<tr>
<th>Interpolation method</th>
<th>Number of operations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>additions</td>
</tr>
<tr>
<td>Linear</td>
<td>Per Sample</td>
</tr>
<tr>
<td></td>
<td>Per Fiducial†</td>
</tr>
<tr>
<td>Cubic spline (Win. N=3)</td>
<td>Per Sample</td>
</tr>
<tr>
<td></td>
<td>Per Fiducial†</td>
</tr>
<tr>
<td>WPL</td>
<td>Per Sample</td>
</tr>
<tr>
<td></td>
<td>Per Fiducial†</td>
</tr>
</tbody>
</table>

†Fiducial points are detected every 100 samples under normal conditions

