**Algorithm name: Approximation-based least squares method (Nonlinear version)**

Developers: Dae Wook Kim, Caleb Mayer, Minki P. Lee, and Daniel B. Forger

Overview: This algorithm was developed to extract an underlying circadian rhythm in heart rate from wearable heart rate, steps, and sleep data.

**Attached file lists**

Computer codes

• main.m: This is the main execution file for analyzing wearable measurements and extracting circadian parameters.

• ALSM\_nonlinear\_version.m: This function applies the approximation-based least squares method to wearable measurements.

• constant\_steps\_filter2.m: This function was used in the data preprocessing part of the ALSM\_nonlinear\_version.m to exclude wearable hear rate data collected when steps count is consecutively zero during more than 2 hours.

• remove\_sleep.m: This function was used in the main.m to exclude wearable heart rate data collected when a subject sleep.

• angleCalc.m: This function was used in the ALSM\_nonlinear\_version.m to compute the angle from and .

• circ\_std.m: This function was used in the ALSM\_nonlinear\_version.m to compute the circular standard deviation of the circadian phase estimate.

• circ\_r.m: This function was used in the circ\_std.m to compute mean resultant vector length for circular data.

Wearable data

• subject.txt: a basic subject information including deidentified subject ID and time-zone.

• heart\_rate.csv: A two column array, the first column lists the epoch time date in days (e.g., 738157 is 01-Jan-2021), and the second colum lists the heart rate value at that time.

• steps.csv: A two column array like the heart rate data, but the second column is the steps value at the respective time.

• sleep.csv: A two column array like the heart rate data, but the second column describes whether the subject sleep. Zero value denotes wakefulness, and positive values denote sleep. Even if we have no sleep data, we can use the algorithm only with heart rate and steps data.

**Remarks**

1. In the analysis, we do not use wearable data collected during sleep episode because cardiac rhythmicity is regulated differently during sleep as previous work (Bowman et al., 2021). Specifically, the model fitted to 2-day intervals centered at the sleep episode between.

2. Because we fitted the model to 2-day intervals centered at the sleep episode between, the step count in the output file (i.e., result.csv) is computed from 2-day of data. Thus, it is not a daily step count.

3. We assumed that the epoch time date in days in the first column of heart\_rate.csv, steps.csv, and sleep.csv are recorded in UTC. Thus, we personalize the time date using the time zone information in subject.txt. If there is no time zone information in subject.txt, and thus we cannot correct it, we just use the epoch time date in the first column with correction.

4. In the data preprocessing, when steps data are missing, we fill in the gaps in steps data with zeros following previous work (Bowman et al., 2021). This might lead to several problems in the estimation. For instance, if heart rate data exist, but steps and sleep data do not exist during sleep episode, the heart rate data collected during sleep episode are not excluded because there are no steps data, and thus we cannot use constant\_steps\_filter2.m to determine whether a subject sleep or not. As a result, the heart rate data collected during sleep episode might be used for the estimation with steps data interpolated by zero.

**Brief description of the algorithm**

**A harmonic-regression model of the human heart rate rhythm used in the algorithm**

The human HR shows circadian variation (Massin et al., 2000). This heart rate (HR) circadian signal has also been successfully analyzed with a harmonic-regression-plus-first-order-autoregressive model (Bowman et al., 2021) described below.

(1)

where

,

,

and it is assumed that the circadian period hr, , the values are distributed as Gaussian random variables with mean zero and variance , and the is the activity level (i.e., step count) at time and is the increase in HR per step (HRpS). The autoregressive noise process describes the ongoing effects of external factors on the HR. The Gaussian noise represents new external influences, for example, from stress, hormones, and measurement error. The autocorrelation factor represents the ongoing contribution of external factors.

**Approximation-based nonlinear LSM to estimate the HR model parameters**

We next describe the nonlinear version of the ALSM to efficiently estimate the HR model parameters. We reformulate and approximate Eq. (1) as follows:

(2)

Due to the reformulation and approximation (Eq. (2)), the noise process is converted from autoregressive noise to independent Gaussian noise. Thus, we can now calculate the mean estimate of , denoted by , and its uncertainty (i.e., the covariance matrix ) by using any LSMs for nonlinear squares curve-fitting problems. In this study, we used the Levenberg-Marquardt algorithm (Gavin 2020: https://people.duke.edu/~hpgavin/ExperimentalSystems/lm.pdf). See below for how to compute the probability density of phase and amplitude from the estimates.

**Computing the probability density of phase and amplitude with Monte Carlo methods**

The time of the HR minimum, and the time of the minimum of 24hr rhythm, denoted by the circadian phase, cannot always be simply related to the model parameters (i.e., the harmonic coefficients). For example, if an oscillatory signal is modelled using a single-harmonic function, the circadian phase can be expressed in closed form:

where , , and are the first-order cosine and sine coefficients, respectively. However, even this simple case with the single harmonic does not allow direct calculation of the probability density of the phase estimate from that of the parameter estimate that can be computed as illustrated in Eq. (5) in Materials and Methods. Thus, we derive the probability density of from by means of Monte Carlo methods (Brown et al., 1992). Then, we calculate the mean and variance (i.e., uncertainty) of the phase estimates, denoted by and , respectively. The steps in the Monte Carlo algorithm are as follows:

*Step 1.* Draw at random from the probability density . Note that when analyzing BT data and when analyzing HR data.

*Step 2.* Compute the estimated harmonic coefficients, and , from using the equations, and .

*Step 3.* Compute a sample phase of interest, , by simulating the harmonic function with and and identifying the time of minimum value or the time of minimum of 24hr rhythm.

*Step 4.* Repeat *Step* *1* to *3* a large number of times, , and make a histogram of its respective replicates, which is the probability density of . Here, is set to be 100,000.

*Step5.* Compute and from the probability density of .

Similarly, we can derive the probability density of the estimate of half the range of the oscillatory signal, denoted by the amplitude, and that of the estimate of the average difference between maximum and minimum of 24hr rhythms, denoted by the circadian amplitude. This allows for the calculation of the mean and variance of the amplitude estimates.