Extended Kalman Filtering for Decoding a Corrupted UART Signal

Casey van der Leek[1]
Email: caseyvdleek@aol.com

I. SYNOPSIS

The pulse sequences of a corrupted UART signal are extracted using an Extended Kalman Filter, a non-linear stochastic filtering algorithm, which is considered the standard in the state estimation of non-linear models. Using the sampled signal data, baud rate of the device and the fact that the noise levels overwhelm the carrier signal only in the latter stages of the sampled interval, the structure of the binary character frames that were transmitted can be deduced. The filter was then used to estimate the voltages and sample numbers of the data frames only. Several tunable parameters exist that serve to optimise the solutions such as, the signal slew rate estimate, the positions of the data bit samplers and the diagonal elements of the covariance (spectral density) matrix that correspond to a maximum measurement error. A non-linear least squares optimiser and a minimisation algorithm were deployed to obtain estimates of the measurement variance. Since the noise effects increase over the sampled interval a dynamic model was ultimately used for the measurement variance. Finally, autocorrelation, autocovariance and power spectral density information of the isolated noise on the carrier signal are provided.

II. INTRODUCTION

The first quarter of the signal data can accurately be decoded by simple auto-threshholding and this sequence is then used to extract the basic character frame structure ; it is also useful to parameterise and test the estimator's accuracy. An analysis of the first few pulses give insight into the possible structure of the full character frames and offer early testing opportunities for the operation of the EKF source code such as the process model effectiveness and sampling strategies.

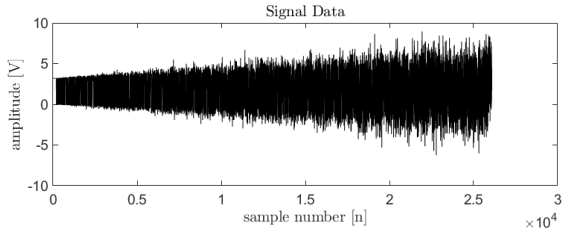

Fig. 1. Sampled UART signal data.

[1]C. van der Leek, MSc(Eng) Mechatronics

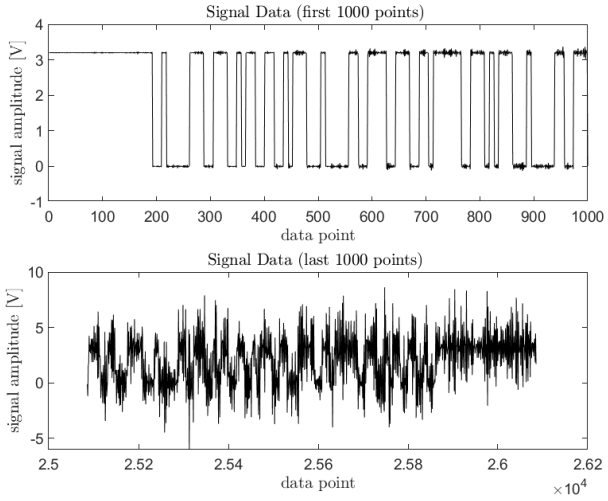

Fig. 2. The pulses in the first quarter of the signal data can be isolated by auto-threshholding code. After sampling number $n = 6589$ the noise levels begin to overwhelm the carrier signal pulses.

III. DATA FRAME STRUCTURE

The early phases of the pulses indicate that the pulse voltages alternate between 0.0 and 3.2V only and the bit rate is therefore equal to the baud rate of 115200 which equates to a bit period of $8.68\mu s$ for the UART device. Since the sample rate is 1MHz the total sample time is $0.026085s$ and approximately 3005 bits were transmitted, including those in the idle state. In order to determine the full character frame structure, code was developed that scanned these early signal sequences. Prior knowledge of the specifications of the UART device enabled the early elimination of candidate frame structures. Initial valid full frames were determined to be 9, 10 or 11 bits in length and with a generic structure that included start(STT), data, parity(PB) and two stop (STP1,STP2) bits, there were a total of 24 permutations considering that there are two parity types, even and odd. The locations of the most significant (MSB) and least significant (LSB) data bits also had to be verified. The code search ultimately revealed a frame size of 11 bits and |STT||DATA||STP1||STP2| structure, with the data bits having a |MSB|.....|LSB| structure (little-endian architecture) and decoded according to the ASCII designations.

IV. STATE ESTIMATION

An Extended Kalman Filter (EKF) was used to estimate the sample number and the corresponding voltage states over the entire transmitted data only. The initial voltage state for

the EKF was set at 1.600V each time it started an estimation of a data frame. Once the voltage states were computed they were sampled according to a pre-defined indexing and

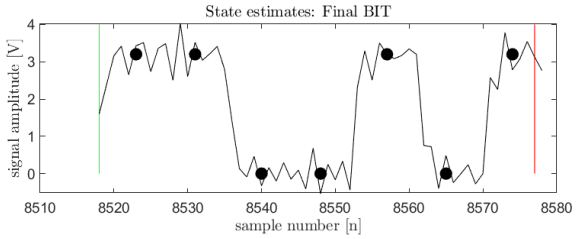

Fig. 3. State estimation of a data bit in the highly corrupted region beyond $n = 8510$. The estimated data voltages between the demarcation data bit boundaries for the interval, start(green) and end(red), with the sampling positions for the bits (dots).

converted to logic values high and low and thus the data frame binary sequence is then converted into its decimal equivalent and finally its ASCII encoding. For a UART pulse the slew rate is given by

$$SR = 2\pi f_{baud} V_{peak}, \tag{1}$$

and with $f_{baud} = 115200$ and $V_{peak} = 3.2$, gives a slew rate of $2.31V/\mu s$ which translates to 30% of bit period ($8.68\mu s$) or $2.597\mu s$ for maximum transition times. This fact, together with the apparent delayed convergence of the estimator at the start of each data sequence, was used when considering the bit sample indexing values.

A. TUNING AND OPTIMISATION

If the assumption is made that the ground truth voltage state is available for a specific iteration by setting it to 0 or 3.2 depending on the current state then the minimisation of a cost function

$$Err(\boldsymbol{R}) = \sum_{i=0}^{N} \|x_{i,ekf}(\boldsymbol{R}) - x_{i,gt}\|^2, \tag{2}$$

can be computed, where $x_{i,ekf}$ is the estimated voltage state by the filter at time step i, $x_{i,gt}$ is the ground truth state at the same time step and N is the total number of iterations. To calibrate the measurement covariance used by the filter processing real sensor data then $\boldsymbol{R}$ can be chosen according to

$$\boldsymbol{R}_{cal} = \arg\min_{\boldsymbol{R}} Err(\boldsymbol{R}). \tag{3}$$

The variance of $\boldsymbol{R}$ can also be obtained by parameterising the final update state equations in the EKF algorithm and applying a non-linear least squares solver to find $\boldsymbol{R}$.

The residual (error) covariance matrix of the state estimates and its trace is an indication of the performance of the filter. In addition, the closer the Kalman gain and the pre-fitted measurement error is to a white noise, the closer the EKF is to performing optimally.

The lower the root mean square error of the measurement predictions and the corresponding observations imply a higher optimality too. The 'noise' covariance matrix $\boldsymbol{Q}$ of the dynamic system process and the measurement uncertainty covariance matrix $\boldsymbol{R}$ values must also be set so that they minimise the covariances of states matrix $\boldsymbol{P}$. The initial values of $\boldsymbol{P}$ at the start of the data frame estimation are also important considerations. Although the additive noise signal on the carrier pulses has close to zero mean and is uncorrelated with itself at any lag, it does not have a constant autocovariance with time so it cannot be assumed to be a white noise across the entire sampling period. It could be approximated as white noise across each of the estimated data frames which have sample lengths of about 70. For a stationary time series sequence in the wide sense, its mean and variance are constant over time.

The mid-point positions for sampling the data voltages are critical in identifying the optimal bit kernel positions. By noting the basic mechanics of the generation of pulses and their estimation, as well as knowledge of the positions of the $STP1$ bit assists in the selection of these positions.

V. NON-LINEAR LEAST SQUARES (LSQ) OPTIMISATION

A non-linear least squares optimisation was also performed using the signal data as model data for the parameterisation of the single voltage variable. The implicitly defined model was a simple unknown DC voltage variable multiplied by 3.2, which had to be solved for, for each of the eight bits data frame. Depending on whether the value of the estimated voltage was above or below the mid-level threshold, a bit was assigned logic high or low. With only the expected value of the estimate having to be specified by the user, there were no other variables or tunable settings.

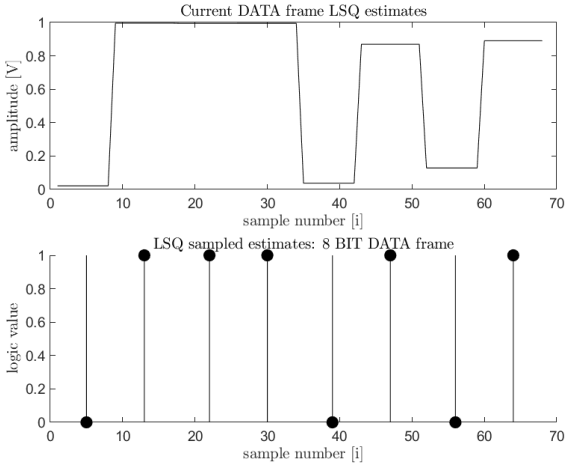

Fig. 4. An 8 bit data frame that has been estimated using a non-linear squares approximation. The estimated voltages for each of the data bits with the resulting logic values after sampling.

VI. ANALYSIS OF NOISE

The additive noise signal was extracted using the original sequence and the state estimates and a qualitative analysis shows that a first order probability distribution average like the variance is dependent on time, implying that the noise is non-stationary, as expected. Second order averages such as

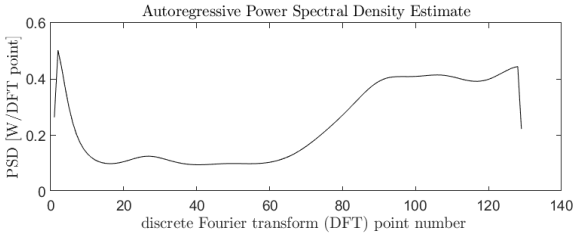

Fig. 5. Power spectral density estimate of the entire isolated noise signal. Notice the uniform values of around 0.1. The variance is 1 for all frequencies for the idealised model for a pseudo-random noise generator.

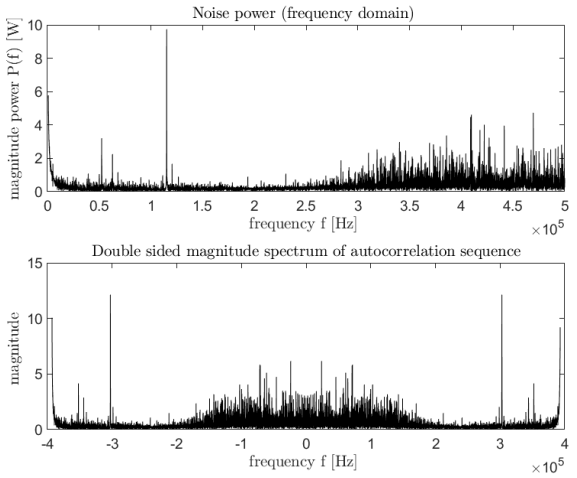

Fig. 6. The noise power spectral density estimates and magnitude spectrum of the autocorrelation estimates of the isolated noise signal.

the aperiodic autocovariance and autocorrelation sequences are dependent on time differences only, and they partially describe the time variation of a random signal. The time dependent Fourier transforms of these non-stationary sequences often exist and provide a useful description of the signal properties with time.

VII. POWER SPECTRAL DENSITY (PSD) ESTIMATES

The spectral representation of second order averages can be realised by taking the Fourier transforms of them resulting in power spectral density (PSD) estimates which typically involves a trade-off between time and frequency resolution.

A. Periodogram Analysis

A good spectral estimate is the smoothed periodogram, which is the squared magnitude of the discrete Fourier transform (DFT) of the windowed (finite segment length) autocorrelation sequence.

B. Fourier transform of the autocorrelation sequence

Another way to estimate the PSD is by taking the time dependent Fourier transform of the estimated autocorrelation sequence. If these transforms are likened to a series of filters, then the frequency response of each filter corresponds to the transform of the windowed sequence, frequency shifted to the DFT frequencies.

VIII. NOTES ON SOURCE CODE

The source code for the program that solves the decoding has been coded in both Matlab(*.mlx) and Python(*.py) versions, with the first few lines of each indicating the line locations of key user defined settings.

REFERENCES

[1] E. Vlasbom, "Nonlinear State Estimation for a Bipedal Robot", MSc Thesis Project, Delft Center for Systems and Control, Delft University of Technology, June 2014.

[2] V. Lertpiriyasuwat, M.C.Berg, K.W. Buffinton, "Extended Kalman Filtering Applied to a Two-Axis Robotic Arm with Flexible Links", *The International Journal of Robotics Research*,Vol.19,No.3,pp.254-270, March 2000.

[3] L. Aguiar, M.R.O.A. Maximo, T. Yoneyama, S. Pinto, "Kalman Filtering for Differential Drive Robots Tracking", SBAI, Porto Alegre, 4 October 2017.

[4] Alan V. Oppenheim, Ronald W. Schafer, "Discrete-Time Signal Processing", *Signal Processing Series, Prentice-Hall International, Inc.*, 1989.

[5] Tibet Mimar, Compression Labs Inc., "Programming and Designing with the 68000 Family", *Prentice-Hall, Inc.*, 1991.

[6] Norman Morrison, "Introduction to Fourier Analysis", *John Wiley Sons, Inc., Wiley Interscience Publication*, 1994.

[7] Ralph J.Smith, Richard C. Dorf, "Circuits, Devices and Systems", Fifth Edition, *John Wiley Sons, Inc.*, 1992.

[8] Paul Horowitz, Winfield Hill, "The Art of Electronics", Second Edition, *Cambridge University Press, UK*, 1995.