Application of wavelet packet transform in roller bearing fault detection and life estimation

Farzad Hemmati¹, Mohammad Miraskari¹ and Mohamed S. Gadala²,³

¹ Mechanical Engineering Department, University of British Columbia, 2054-6250 Applied Science Lane, Vancouver, BC, V6T 1Z4, Canada.
² Mechanical Engineering Department, University of British Columbia, and Mechanical Engineering Department, Abu Dhabi University, Abu Dhabi, UAE
³ E-mail: gadala@mech.ubc.ca

Abstract. Rolling element bearings are crucial parts of many machines and there has been an increasing demand to find effective and reliable health monitoring technique and advanced signal processing to detect and diagnose the size and location of incipient defects. Condition monitoring of rolling element bearings, comprises four main stages which are, statistical analysis, fault diagnostics, defect size calculation, and prognostics. A novel signal processing algorithm is designed to diagnose localized defects on rolling element bearings components under different operating speeds, loadings, and defect sizes. The algorithm is based on optimizing the ratio of Kurtosis and Shannon entropy to obtain the optimal band pass filter utilizing wavelet packet transform and envelope detection. To experimentally measure the defect size on rolling element bearings using acoustic emission technique, the proposed method along with spectrum of squared Hilbert transform are performed under different rotating speeds, loading conditions, and defect sizes to measure the time difference between the double acoustic emission impulses. Measurement results show the power of the proposed method for experimentally measuring size of different fault shapes using acoustic emission signals. Fatigue life estimation of rolling element bearing has also been investigated utilizing defect size measurements combined with Recursive Least Square Estimation method which is an adaptive algorithm. Experimental results show the effectiveness of recursive least square algorithm for predicting the future defect size on the outer race.

1. Introduction

Condition monitoring of heavy rotating machinery and equipment such as turbines, compressors and generators, is gaining importance in various industries since it keeps the plant at healthy condition for maximum production; helps in detecting faults at early stages; avoid serious accidents and damage; and reduces downtime. Bearings are the common elements used in heavy rotating machinery and equipment because of their high reliability. Bearings start to malfunction due to machine overload, shaft misalignment, rotor unbalance, overheating, etc. Many different techniques based on vibration methods have been developed to extract bearing fault features [1]. However, vibration signals are not sensitive to incipient faults and they are usually masked by background noise caused by mechanical vibration signals from rotating machinery. Hence, it is normally difficult for the vibration techniques to detect bearing faults at an early stage. Acoustic emission (AE) is the phenomenon of transient elastic wave generation due to a rapid release of strain energy caused by relative motion of small particles under mechanical stresses [2]. Interaction of rolling element bearing components and
movement of bearing rollers over defects will produce AE’s. The frequency content of acoustic emission (AE) is typically in the range of 100 kHz to 1 MHz, so AE is not influenced or distorted by imbalance and misalignment which are at low frequency ranges. The high sensitivity of AE technique and AE parameters in detecting the incipient bearing faults has become one of the significant advantages of AE over vibration measurement [3]. A comprehensive review of AE application for bearing fault detection was presented by Mba et al. [4]. A complete overview of bearing and gear health monitoring systems, diagnostics and prognostics, have been provided by Wang et al. [5]. Al-Ghamd and Mba [6] investigated the relationship between AE parameters in time domain and defect size. They concluded that AE burst duration is an effective parameter for identifying defect size on the outer race. However, in recent work performed by researchers there is no sensitivity analysis presented to analyze the most sensitive statistical parameter to incipient faults and defect growth. Acoustic based diagnostic signals in real industrial environments associated with high temperatures, rotating speeds, and pressures are always masked by high levels of noise. Desirable de-noising of these signals is not achievable with conventional techniques. Therefore, to overcome this challenge adaptive signal processing techniques need to be developed to enhance signal to noise ratio of AE signals. Dyer and Stewart [7] first presented the use of kurtosis parameter for bearing fault diagnosis and it was suggested to use kurtosis value in selected frequency bands. Antoni and Randall [8] presented the use of Spectral Kurtosis (SK) to extract transient components from a noisy signal. Sawalhi and Randall [9] proposed minimum entropy deconvolution (MED) technique along with SK to enhance the results of envelope analysis from a vibration bearing fault signal. Discrete wavelet transform (DWT) has been used in signal denoising due to its high resolution in time and frequency domains [10]. For instance, Qiu [11] utilized wavelet filter-based denosing method to enhance weak periodic impulse signature masked by standard Gaussian white noise. In DWT, a digitized signal is decomposed into its low-pass approximation and high-pass detailed signals and further decompositions only apply to the detailed components. Hence, it suffers from insufficient treatment of the high frequency components where the bearing fault impulses exist [12]. Thus, wavelet packet transform (WPT) has been introduced to overcome this issue by treating both low and high frequency components [12]. Lei Y et al. [13] proposed an improved kurtogram method for diagnosing bearing faults. Even though, this method finds the frequency band which has the maximum value of kurtosis using WPT, the optimal wavelet function is not selected along with kurtosis value does not provide any information regarding periodic behavior of bearing fault impulses. Detectable bearing defect area is much smaller than 6.25 mm\(^2\) (0.01 in\(^2\)), which is commonly considered to be a fatal failure size by industry standards [14, 15]. However, to diagnose defect location using characteristic fault frequencies is not sufficient for the purposes of condition-based maintenance. In industry when a defect is diagnosed, the machinery is often forced to shut down which may cause tremendous loss of productivity for changing the faulty bearing. The time or number of cycles to reach the final failure area (6.25 mm\(^2\)) may be greater than its fatigue life [16]. Therefore, it is important to predict the fatigue life and growth rate of defects on rolling element bearings in a prognostic mode in addition to diagnostic mode. A comprehensive review of rolling element bearing life estimation was presented in [17]. The prognostic methods can be classified into two main categories; (1) Data-driven methods which create a model based on the measurement data from condition monitoring [18, 19]. Disadvantage of data-driven approaches is their sensitivity to the measurement data. (2) Physics-based prognostics models typically involve building a comprehensive mathematical model based on a physical model of the system and crack propagation [20-24]. Wang and Tsui provided new statistical methods for bearing life estimation [25, 26]. With reliable prognostic capabilities, bearing maintenance and replacement can be scheduled at optimal times in the interest of overall system productivity and safety. 

In this paper, different approaches in developing prognostic models are reviewed and the role of the proposed technique in estimating the defect size using acoustic emission technique in the prognostic model is explained. The prognostic model utilizes an adaptive algorithm to best estimate the rate of defect propagation in a real-time manner. The adaptive algorithm of the model parameters offers the best prediction, in the least square error sense, of the bearing future state for any given diagnostic
system. A run-to-failure experiment was conducted for validating the prognostic model using acoustic emission signals. Then, the prognostic method is applied to the experimentally acquired signals of a faulty bearing where the fault is artificially introduced on the outer race of the rolling element bearing. The predicted bearing defect size as estimated by the prognostic model is compared to the actual measured defect size and the results demonstrate that by using the prognostic method, defect size on the outer race can be accurately predicted at a given number of cycles.

2. Bearing defect size estimation

In this section AE signals are sampled at a rate of 2MHz for maximum accuracy in detecting the entry and exit points. The size of the artificial line shape defect was 1mm (D1) in width and 0.5mm in depth. For measured AE signals the experiments were conducted at speeds of 600 rpm (S2) and 1100 rpm (S3) and under loads of 0 N (L0) and 100 N (L1) to investigate the effects of speed and load on the AE defect size estimation. The time duration of measured AE waveform is selected as 3s (6,000,000 data points). The theoretical defect size on the outer race can be calculated based on shaft rotating speed and relative velocities of outer race and roller elements. The objective is to correlate the theoretical and artificial defect size.

The measured AE signals are processed using the proposed method for de-noising along with spectral analysis of the squared Hilbert transform. A window size of 2200 point is selected to measure the time duration and frequency between two spikes of each doubles impulses. After the AE frequencies ($F$) of all double spikes are successfully found, the arithmetic mean of all AE frequencies is calculated and reported as the experimentally measured AE frequency ($F_{\text{avg}}$). Defect sizes estimation method is explained on figure 1 [27]. Measured defect sizes obtained from AE signals under different conditions are tabulated in table 1, for comparison purposes.

| Speed (rpm) | Defect size (mm) | Estimated defect size (mm) and optimal wavelet function | Error % |
|-------------|------------------|--------------------------------------------------------|---------|
| (s2)600     | 1                | 1.016 (db44) 1.023 (db22) | L0 L1   | 1.6 2.3 |
| (s3)1100    | 1                | 1.062 (db22) 1.084 (db44) | L0 L1   | 6.2 8.4 |
| (s2)600     | 2                | 2.044 (sym22) 2.104 (db44) | L0 L1   | 2.2 5.2 |
| (s3)1100    | 2                | 2.064(db39) 2.132 (sym24) | L0 L1   | 3.2 6.6 |

Figure 1. Flow chart of defect size estimation.

3. Damage model based on adaptive and stochastic prognostics method

3.1. Numerical simulation using Paris’s law

The primary difficulty for effective implementation of bearing prognostics is the highly stochastic nature of defect growth. For instance, the crack growth on a rolling element bearing beyond its initial defect size is a highly variable process [14]. Fatigue cracks initiate and grow when stresses vary through time and their growth is affected by many factors, e.g. material properties, temperature, geometry and shape factors, stress concentration, and lubrication. Assume the stresses to be fluctuating between the limits of $\sigma_{\text{min}}$ and $\sigma_{\text{max}}$. Stress intensity range per cycle can be calculated using [15]:
where $\beta$ is the stress intensity modification factor and $a$ is the crack length. Cracks usually initiate at free surface or large discontinuity. Assuming an initial crack of length $a_0$, rate of crack growth ($\frac{da}{dN}$) can be calculated as a function of $\Delta K_I$ [15]. Defect size growth rate as a function of stress intensity can be divided into three stages/regions: (I) crack initiation, (II) crack propagation which is a linear function on log-log coordinates, within the domain of linear elastic fracture mechanics validity, and (III) unstable crack [15]. Assuming a crack is discovered early in stage II, the crack growth in region II can be estimated by the Paris equation [28]:

$$\frac{da}{dN} = C(\Delta K)^m$$  \hspace{1cm} (2)

where $a$ is the instantaneous size of a defect or crack and $N$ represents running cycles (running time). $C$ and $m$ are empirical material constants and $\Delta K_I$ can be derived using Equation (1). By substituting $\Delta K_I$ from Equation (1) into Equation (2), the above equation can be simplified into a compact form in a manner similar to Paris's equation which is presented as:

$$\frac{da}{dN} = C(\beta \sigma \sqrt{\pi a})^m = C_0 a^n$$  \hspace{1cm} (3)

where $C_0$ and $n$ are material constants. Equation (3) indicates that the rate of defect size growth is related to the instantaneous defect size $a$ under constant operating conditions. Through integration of Equation (3), based on the assumption that $C_0$ and $n$ are constants and not function of the crack length, the solution can be written as:

$$\frac{da}{a^n} = C_0 dN \rightarrow \int_{a_0}^a \frac{da}{a^n} = \int_{N_0}^N C_0 dN$$  \hspace{1cm} (4)

$$a = (a_0^{1-n} + C_0 (1 - n)(N - N_0))^{\frac{1}{n-1}} = (a_0^{1-n} + C_0 (1 - n)\Delta N)^{\frac{1}{n-1}}$$  \hspace{1cm} (5)

where $a_0$ is the initial defect size and $\Delta N$ is the number of cycles to reach to the defect size $a$. Under the assumption $C_0$ and $n$ are constants and using Equation (5), numerical simulations are performed to analyze the sensitivity of defect size using different typical values of $C_0$, $a_0$, and $n$ (for various materials, geometry factors, and crack sizes). As it can be seen in figure 2 and figure 3, by increasing initial defect size and $C_0$, estimated defect length increases. However, by increasing the parameter $n$, estimated defect size decreases which is illustrated in figure 4.

**Figure 2.** Effects of initial defect size on defect size propagation for $n = 2.25$ and $C = 1.5 \times 10^{-4}$. Initial crack size $a_{01} = 0.006$ mm, $a_{02} = 0.008$ mm, $a_{03} = 0.01$ mm, $a_{04} = 0.012$ mm, $a_{05} = 0.014$ mm.

**Figure 3.** Effects of parameter $C$ on defect size propagation for the conditions where $a_0 = 0.006$ mm and $n = 2.45$. $C_1 = 1.5 \times 10^{-4}$, $C_2 = 1.4 \times 10^{-4}$, $C_3 = 1.3 \times 10^{-4}$, $C_4 = 1.2 \times 10^{-4}$, $C_5 = 1.1 \times 10^{-4}$.
Figure 4. Effects of parameter $n$ on defect size propagation for the condition where $a_0 = 0.006$ mm and $C = 1.5 \times 10^{-4}$. $n1 = 2.45$, $n2 = 2.35$, $n3 = 2.25$, $n4 = 2.15$, $n5 = 2.05$.

3.2. Defect size prognostics utilizing recursive least square estimation

As shown in figure 2, figure 3, and figure 4, instantaneous defect size $a$ is highly sensitive to the parameters $C_0$ and $n$. Experiments show that stochastic nature of defect growth cannot be accurately described by constant parameters $C_0$ and $n$, hence, $C_0$ and $n$ must be accurately determined and updated during the integration process (running time) to accurately obtain the remaining life [20]. Thus, direct integration of Equation (3) is not easy to perform without an explicit dependency of the parameters on the crack length (or time). To determine and update the parameters given in Equation (3), recursive least squares (RLS) approach is introduced for estimating and predicting the defect size. Equation (3) can be linearized by taking logarithm and written in a form of,

$$\ln \left( \frac{da}{dN} \right) = \ln(C_0) + n\ln(a) + e(N)$$

Where $e(N)$ is the random measurement error. Thus, we are working with a linear system which can be defined as:

$$\gamma(N) = H(N) x(N) + e(N)$$

Where $\gamma(N) = \ln \left( \frac{da}{dN} \right)$, $H = [1 \quad \ln(a)]$, and $x = [\ln(C_0) \quad n]^T$. From online condition monitoring using acoustic emission technique along with the proposed method introduced in by F. Hemmati et al. [27], measured noisy AE signals can be de-noised and defect size on different locations of rolling element bearing can be measured utilizing the spectrum of squared Hilbert transform. The objective is to update parameters, $C_0$ and $n$, at a given cycle recursively by taking defect size measurements and predict the next defect size through direct integration of Equation (3).

In Equation (6), $a$ can be obtained through the proposed method explained in [27]. Defect size propagation rate $\frac{da}{dN}$ is determined through locally fit the best power function to five preceding measured data points in a least square sense. For doing that it is assumed defect size can be written as a power function of cycle $N$.

$$a = C_1 N^{n1}$$

For online condition monitoring, it is useful to estimate the parameters in a recursive manner, hence, the estimated parameters can be updated as new data become available. Since $C_0$ and $n$ change at each cycle, they needed to be updated when new measured defect size become available. For this purpose, recursive least squares (RLS) algorithm is used to adaptively update the values of $C_0$ and $n$ [29, 30]. A detailed description of the RLS algorithm is presented in Appendix A. The RLS algorithm for the given model, Equation (7), can be written as:

$$K(N) = P(N-1)H^T(N)(H(N)P(N-1)H^T(N) + R(N))^{-1}$$

$$P(N) = (I - K(N)H(N))P(N-1)$$
\[ \hat{x}(N) = \hat{x}(N-1) + K(N)(y(N) - H(N)\hat{x}(N-1)) \]  

(11)

where \( \hat{x}(N) \) is the estimate value of \( x(N) \). \( K(N) \) is the estimator gain matrix and \( P(N) \) is the covariance matrix of estimation error. Without prior knowledge of the initial conditions, \( x = [\ln(C_0) \ n]^T \), initial covariance matrix is chosen as a unit matrix with a large positive scalar of 10000. \( R(N) \) is the noise covariance matrix which is assumed all measured data points have the same covariance of 0.001. The flow chart of the adaptive algorithm is presented in the figure 5.

Figure 5. Flow chart of defect size estimation using recursive least squares (RLS).

4. Experimental procedure and validation

Figure 6. Measured defect size with respect to number of cycles using the proposed method by measuring the time difference between two spikes and the effect of initial conditions on the defect size prediction. Estimation I (E1) \( C_0 = 1.5 \times 10^{-5}, n_i = 2 \). Estimation II (E2) \( C_0 = 1.5 \times 10^{-9}, n_i = 3.2 \).

Figure 7. Deviation of error with respect to number of cycles. Estimation I (E1) \( C_0 = 1.5 \times 10^{-5}, n_i = 2 \). Estimation II (E2) \( C_0 = 1.5 \times 10^{-9}, n_i = 3.2 \).

Figure 8. Deviation of parameter \( n \) with respect to number of cycles. Estimation I (E1) \( C_0 = 1.5 \times 10^{-5}, n_i = 2 \). Estimation II (E2) \( C_0 = 1.5 \times 10^{-9}, n_i = 3.2 \).

Figure 9. Deviation of parameter \( C_0 \) with respect to number of cycles. Estimation I (E1) \( C_0 = 1.5 \times 10^{-5}, n_i = 2 \). Estimation II (E2) \( C_0 = 1.5 \times 10^{-9}, n_i = 3.2 \).
The experimental setup shown in [27] is employed to evaluate and demonstrate the prognostic algorithms described in this section. In order to accelerate a defect propagation process, an initial radial shape defect of 1 mm diameter is artificially introduced on the outer race of a rolling element bearing by an engraving tool. Radial preload of 100 N is also applied in the middle of the shaft as shown in the [27]. The experiment was conducted for 17.7 hours (1 million cycles) at a constant rotating speed of 600 rpm. To measure the defect size using the proposed method, AE signals were measured every 34 minutes, total number of 54 measurements, at sampling rate of 2MHz for time duration of 3 seconds. Figure 6 illustrates measured defect size and the predicted defect sizes using the adaptive prognostic model. Defect size is predicted using two different initial conditions $C_0$ and $n$ parameters. The results in figure 6 imply that the adaptive prognostic system can effectively predict the bearing defect propagation process without a prior knowledge of $C_0$ and $n$. Figure 7 shows the errors between the predicted defect size and the measured defect and it implies the fact that the prediction accuracy is not strongly affected by the choice of initial $C_0$ and $n$ values. Figure 8 and figure 9 show the variation of $n$ and $C_0$ parameters with respect to the number of cycles. As illustrated in figure 8 and figure 9, and for the specific case of roller bearing cracks/defects, it can be concluded that after $8 \times 10^7$ number of cycles, $n$ and $C_0$ parameters converge to constant values of 2.78 and $4.5 \times 10^{-7}$ respectively. In the adaptive algorithm the major prediction error sources are attributed to the noisy measurements of AE signals and diagnostic model for measuring the defect size using the spectrum of the squared Hilbert transform.

5. Conclusions
To estimate the defect size, the proposed method was utilized to de-noise the noisy measurement data. Then spectrum of squared Hilbert transform was performed to measure the time travel between double impulses for calculating the defect sizes at different locations. Experimental measured defect sizes were compared with their actual values which shown a maximum error of 10%.

Proposed prognostics methodology incorporates the time-variant nature of defect growth while providing the best prediction possible, in the least square error sense, for any given diagnostic system. The defect size as predicted by the adaptive algorithm using RLS along with fatigue crack propagation model was compared to the measured actual defect size utilizing de-noising algorithm. The adaptive prognostics effectively predict the bearing defect propagation and the error did not exceed 10% for the outer race notch defect shape. It should be noted that by increasing the number of measured AE signals, estimated error will be decreased.

References
[1] McFadden P and Smith J 1984 Model for the vibration produced by a single point defect in a rolling element bearing Journal of Sound and Vibration 96(1) 69-82
[2] Pao YH, Gajewski R and Ceranoglu AN 1979 Acoustic emission and transient waves in an elastic plate The Journal of the Acoustical Society of America 65(96)
[3] Tandon T and Choudhury A 1999 A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings Tribology International 32(8) 469-480
[4] Mba D and Rao RN 2006 Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines; Bearings, Pumps, Gearboxes, Engines and Rotating Structures
[5] Wang D, Tsui K, and Miao Q 2018 Prognostics and Health Management: A Review of Vibration based Bearing and Gear Health Indicators IEEE Access 6 665-676
[6] Al-Ghamd AM and Mba D 2006 A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size Mechanical Systems and Signal Processing 20(7) 1537-1571
[7] Dyer D and Stewart R 1977 Detection of rolling element bearing damage by statistical vibration analysis
[8] Antoni J and Randall R 2006 The spectral kurtosis: application to the vibratory surveillance and
diagnostics of rotating machines Mechanical Systems and Signal Processing 20(2) 308-331
[9] Sawalhi N, Randall R, and Endo H 2007 The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis Mechanical Systems and Signal Processing 21(6) 2616-2633
[10] Yiakopoulos C and Antoniadis I 2002 Wavelet based demodulation of vibration signals generated by defects in rolling element bearings Shock and Vibration 9(6) 293-306
[11] Qiu H, Lee J, Lin J, and Yu G 2006 Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics Journal of Sound and Vibration 289(4) 1066-1090
[12] Zarei J and Poshtan J 2007 Bearing fault detection using wavelet packet transform of induction motor stator current Tribology International 40(5) 763-769
[13] Lei Y, Lin J, He Z, and Zi Y 2011 Application of an improved kurtogram method for fault diagnosis of rolling element bearings Mechanical Systems and Signal Processing
[14] Hoeprich M 1992 Rolling element bearing fatigue damage propagation Journal of Tribology 114 328
[15] Shigley JE, Mischke CR, Budynas RG, Liu X, and Gao Z 1989 Mechanical engineering design. McGraw-Hill New York
[16] Harris TA 1991 Rolling bearing analysis New York, USA
[17] Heng A, Zhang S, Tan AC and Mathew J 2009 Rotating machinery prognostics: State of the art, challenges and opportunities Mechanical Systems and Signal Processing 23(3) 724-739
[18] Altmann J and Mathew J 2001 Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis Mechanical Systems and Signal Processing 15(5) 963-977
[19] Vachtsevanos G and Wang P 2001 Fault prognosis using dynamic wavelet neural networks 2001 pp. 857-870: IEEE
[20] Li Y, Billington S, Zhang C, Kurfess T, Danyluk S and Liang S 1999 Adaptive prognostics for rolling element bearing condition Mechanical Systems and Signal Processing 13(1) 103-113
[21] Li Y, Kurfess T and Liang S 2000 Stochastic prognostics for rolling element bearings Mechanical Systems and Signal Processing 14(5) 747-762
[22] Qiu J, Seth BB, Liang SY and Zhang C 2002 Damage mechanics approach for bearing lifetime prognostics Mechanical Systems and Signal Processing 16(5) 817-829
[23] Orsagh RF, Sheldon J and Klenke CK 2003 Prognostics/diagnostics for gas turbine engine bearings
[24] Jantunen E 2006 Prognosis of rolling bearing failure based on regression analysis and fuzzy logic 2006 pp 390-399
[25] Wang D and Tsui KL 2017 Statistical Modeling of Bearing Degradation Signals IEEE Transactions on Reliability 66(4) 1331-1344
[26] Wang D and Tsui KL 2018 Two novel mixed effects models for prognostics of rolling element bearings Mechanical Systems and Signal Processing 99 1-13
[27] Hemmati F, Orfali W, and Gadala MS 2016 Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation Applied Acoustics 104 101-118
[28] Cempel C, Natke H and Tabaszewski M 1997 A passive diagnostic experiment with ergodic properties Mechanical systems and signal processing 11(1) 107-117
[29] Diniz PS 2008 Adaptive filtering: algorithms and practical implementation Springer Verlag
[30] Simon D 2006 Optimal state estimation: Kalman, H [infinity] and nonlinear approaches John Wiley and Sons
Appendix A

Recursive Least Square Estimation

A.1. Weighted Least Square Estimation

Suppose \( x = (x_1, x_2, ..., x_k)^T \) is an unknown vector which is going to be estimated, and \( y = (y_1, y_2, ..., y_k)^T \) is the measured data of size \( k \). Every measured data \( y_i \) is known to be a linear function of \( x_i \).

\[
y = Hx + \omega \quad (A.1)
\]

where \( \omega_i, 1 \leq i \leq k \) is the noise of length \( n \). Given an estimate \( \hat{x} \), the difference between the noisy measurements and the \( H\hat{x} \), estimated values, can be written as:

\[
e = y - H\hat{x} \quad (A.2)
\]

It can be assumed that each measured data may be taken under a different condition; hence, the estimated value of the measurement noise can be defined as,

\[
E(\omega_i^2) = \sigma_i^2 \quad (A.3)
\]

where \( \sigma_i, 1 \leq i \leq k \) is the variance of the measurement noise. Assume that the noise for each measurement has zero mean and is independent. The covariance matrix for all measurement noise is,

\[
R = E(\omega\omega^T) = \begin{bmatrix}
\sigma_1^2 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sigma_k^2
\end{bmatrix} \quad (A.4)
\]

Assume the difference \( y - H\hat{x} \) as \( e = (e_1, e_2, ..., e_k)^T \). The summation of squared differences weighted over the variations of the measurements can be minimized by,

\[
J(\hat{x}) = e^TR^{-1}e = \frac{e_1^2}{\sigma_1^2} + \frac{e_2^2}{\sigma_2^2} + \cdots + \frac{e_k^2}{\sigma_k^2} \quad (A.5)
\]

The cost function \( J \) can be expanded as follows,

\[
J(\hat{x}) = (y - H\hat{x})^TR^{-1}(y - H\hat{x}) = y^TR^{-1}y - \hat{x}^TH^TR^{-1}y - y^TR^{-1}H\hat{x} + \hat{x}^TH^TR^{-1}H\hat{x} \quad (A.6)
\]

To obtain the optimal solution of \( \hat{x} \), the partial derivative of \( J \) must be zero.

\[
\frac{\partial J}{\partial \hat{x}} = -2y^TR^{-1} + 2\hat{x}^TH^TR^{-1}H = 0 \quad (A.7)
\]

By solving the above equation, the best estimate of \( x \) can be written as,

\[
\hat{x} = (H^TH^{-1}H)^{-1}H^TR^{-1}y \quad (A.8)
\]

It should be noted that under the condition of non-singularity of matrix \( R \), the above equation has a solution.

A.2. Recursive Least Square Estimation

Equation (A.8) is adequate when all measurements exist. More often, we obtain measurements sequentially and want to update our estimate with each new measurement. In this case, the matrix \( H \) is needed to be augmented to have an updated estimate of \( \hat{x} \) according to equation (A.8) for every new measurement. Suppose we have an estimate \( \hat{x}_{k-1} \) after \( k-1 \) measurements and obtain a new measurement \( y_k \). The algorithm below shows how to update a new estimate of \( \hat{x}_k \) without solving equation (A.8). A linear recursive estimator can be written in the following form,

\[
y_k = H_kx + \omega_k \quad (A.9)
\]

\[
\hat{x}_k = \hat{x}_{k-1} + K_k(y_k - H_k\hat{x}_{k-1}) \quad (A.10)
\]

where \( K_k \) is a \( n \times k \) matrix referred to as the estimator gain matrix. Namely, the new estimate \( \hat{x}_k \) is updated from the previous estimate \( \hat{x}_{k-1} \) with \( y_k - H_k\hat{x}_{k-1} \), correlation term, via the gain matrix. The current estimate error is \( e_k = x - \hat{x}_k \). The mean value of this error can be computed as follows,

\[
E(e_k) = E(x - \hat{x}_k) = E(x - \hat{x}_{k-1} - K_k(y_k - H_k\hat{x}_{k-1}))
\]

\[
= E(e_{k-1} - K_k(H_kx + \omega_k - H_k\hat{x}_{k-1}))
\]

\[
= E(e_{k-1} - K_kH_k(x - \hat{x}_{k-1}) - K_k\omega_k)
\]

\[
= (I - K_kH_k)E(e_{k-1}) - K_kE(\omega_k) \quad (A.11)
\]
where \( I \) is the \( n \times n \) identity matrix. If \( E(\omega_k) = 0 = E(\omega_{k-1}) = 0 \), then \( E(e_k) = 0 \). Thus, if the measurement noise \( \omega_k \) has zero mean for all \( k \), and the initial estimate of \( x \) is set equal to its expected value, then \( \hat{x}_k = x_k \) for all \( k \) values. The key is to determine the optimal value of the gain matrix \( K_k \). In order to do that, the aggregated variance of the estimation errors is minimized at time \( k \).

\[
J_k = E(\|x - \hat{x}_k\|^2) = E(e_k^T e_k) = E(\text{tr}(e_k e_k^T)) = \text{tr}(P_k)
\]  
(A.12)

where \( \text{tr} \) is the trace operator, and \( P_k = E(e_k e_k^T) \) is the estimation error covariance. With substitution of equation (A.11) \( P_k \) can be obtained as,

\[
P_k = E\left((I - K_k H_k)E(e_k) - K_k \omega_k\right)((I - K_k H_k)E(e_k) - K_k \omega_k)^T
\]

\[
= (I - K_k H_k) E(e_{k-1} e_{k-1}^T) (I - K_k H_k)^T - K_k E(\omega_k e_{k-1}) (I - K_k H_k)^T
\]

\[
- (I - K_k H_k) E(e_{k-1} \omega_k^T) K_k^T + K_k E(\omega_k \omega_k^T) K_k^T
\]

\[
= (I - K_k H_k) P_{k-1} (I - K_k H_k)^T + K_k R_k K_k^T
\]  
(A.13)

The last step is to use the definition of \( R_k = E(\omega_k \omega_k^T) \) as a covariance of \( \omega_k \), plus a fact that the estimation error \( e_{k-1} \) at time \( k - 1 \) is independent of the measurement noise \( \omega_k \) at time \( k \). The latter implies that,

\[
E(\omega_k e_{k-1}^T) = E(\omega_k e_{k-1}) = 0
\]  
(A.14)

Equation (A.13) is the recurrence for the covariance of the least squares estimation error. It should be noted that \( P_k \) as a covariance matrix is a positive definite matrix. To find the optimal value of \( K_k \) that minimizes the cost function given by equations (A.11) and (A.12), \( J_k \) is differentiated with respect to \( K_k \).

\[
\frac{\partial J_k}{\partial K_k} = 2(I - K_k H_k) P_{k-1} (-H_k^T) + 2K_k R_k
\]  
(A.15)

Setting the partial derivative to zero, \( K_k \) can be written as,

\[
K_k = P_{k-1} H_k^T (H_k P_{k-1} H_k^T + R_k)^{-1}
\]  
(A.16)

Assume \( S_k = H_k P_{k-1} H_k^T + R_k \) so,

\[
K_k = P_{k-1} H_k^T S_k^{-1}
\]  
(A.17)

Substitute the above for \( K_k \) into equation (A.13) to find \( P_k \) which is summarized as bellow,

\[
P_k = (I - P_{k-1} H_k^T S_k^{-1} H_k) P_{k-1} (I - P_{k-1} H_k^T S_k^{-1} H_k)^T + P_{k-1} H_k^T S_k^{-1} R_k S_k^{-1} H_k P_{k-1}
\]

\[
= P_{k-1} - P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} - P_{k-1} H_k^T S_k^{-1} H_k P_{k-1}
\]

\[
+ P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} + P_{k-1} H_k^T S_k^{-1} R_k S_k^{-1} H_k P_{k-1}
\]  
(A.18)