Robust speech recognition system using bidirectional Kalman filter

Yeh Huann Goh[1] ✉, Paramesran Raveendran[1], Yann Ling Goh[1,2]

[1]*Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Lembah Pantai, 50603 Kuala Lumpur, Malaysia*
[2]*Department of Mathematical and Actuarial Sciences, Lee Kong Chian Faculty of Engineering Science, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, Cheras, 43000 Kajang, Malaysia*
✉ *E-mail: gyhuann@hotmail.com*

Abstract: Kalman filter is normally used to enhance speech quality in a noisy environment, in which the speech signals are usually modelled as autoregressive (AR) process, and represented in the state-space domain. It is a known fact that to identify the changing AR coefficients in every time state requires extensive computation. In this paper, the authors develop a bidirectional Kalman filter and apply it in a speech processing system. The proposed filter uses a system dynamics model that utilises the past and the future measurements to form an estimate of the system's current time state. It provides efficient recursive means to estimate the state of a process that minimises the mean of the squared error. Compared to the conventional Kalman filter, the proposed filter reduces the computation time in two ways: (i) by avoiding the computation of AR parameters in each time state, and (ii) by reducing the dimension of the matrices involved in the difference equations and the measurement equations into constant (1×1) matrices. The speech recognition result shows that the developed speech recognition system becomes more robust after the proposed filtering process, and the proposed filter's low computational expense makes it applicable in the practical hidden Markov model-based speech recognition system.

1 Introduction

Kalman filter is a recursive solution to the discrete-data linear filtering problem. Owing to digital computing, the Kalman filter has been the subject of extensive research and applications. Various approaches based on the Kalman filter have been successfully used for decades at the core of many speech enhancement algorithms [1, 2]. The use of the Kalman filter in denoising the speech signals requires the estimation of the speech model parameters, that is usually modelled as autoregressive (AR) model and represented in the state-space domain [3], it is a known fact that the performance of a Kalman filter system largely depends on the reliability of the estimates of the AR model parameters. Many different algorithms have been proposed to estimate the speech model parameters, such as using expectation–maximisation algorithm [4, 5], subspace non-iterative algorithm based on orthogonal projection in [6], log-spectral amplitude minimum-mean-square-error (MMSE) in [7], power spectral subtraction method in [8, 9], comparison to a masking threshold that computed from both time and frequency-domains simultaneous masking properties of human auditory systems in [10], estimation of the clean-speech short-term predictor parameters from noisy speech using maximum-a-posteriori and MMSE techniques in [11]. Besides, the use of the Kalman filter in the speech recognition system to improve the speech recognition accuracy has been proposed in [12, 13], Mathe *et al.* [14] have used the Kalman filter for speech enhancement purpose.

Speech signal processing using the Kalman filter requires extensive computations and the real-time implementation of this approach is difficult. Different algorithms have been proposed for the fast processing purpose, such as rewriting of the state-space equations to reduce the dimension of the state vector and the amount of computations per iteration in [15]. Decomposing speech signals into subbands to produce low-order AR model that can be processed by low-order Kalman filter, enhanced fullband speech signals are obtained by combining the enhanced subband speech signals [16]. Mai *et al.* [17] proposed a fast adaptive Kalman filter algorithm for

speech enhancement. This algorithm eliminates the matrix operations and only constantly updating the first value of the state vector. Together with the adaptive filtering algorithm to automatically amend the estimation of the environmental noise, the simulation result shows that this algorithm is effective for speech enhancement.

Many noise-robust algorithms have been applied in the front-end feature domain [18–25] and/ or in the back-end model domain [26–28] to reduce the effects of noise on the speech processing system. The study of this paper addresses the problems of noise filtering and the complexity of computation. The proposed bidirectional Kalman filter uses a system dynamics model that utilises the past and the future measurements to form an estimate of the system's current time state. It provides efficient recursive means to estimate the state of a process that minimises the mean of the squared error. This technique is known as smoothing technique that makes the maximum-likelihood estimates of the state variables of linear and non-linear dynamic systems over a finite time interval [29]. Fong *et al.* [30] have applied this smoothing technique in the audio signal enhancement using a Monte Carlo filter. In our design, speech features model parameters are estimated before the recognition process. Unlike the AR model parameters used in the conventional Kalman filter which might change in each time state, the speech features model parameters for the proposed filter are constant throughout.

The organisation of the paper is as follows: Section 2 begins with a brief review of the Kalman filter followed by the derivation of the proposed bidirectional Kalman filter. In Section 3, we show how the model parameters are determined by using a recursion formula and minimum distance measurement. Section 4 compares the proposed bidirectional Kalman filter with the conventional Kalman filter [7] and the fast adaptive Kalman filter [17] according to following criteria: (i) correlation measured, (ii) signal-to-noise ratio (SNR) measured, (iii) weighted spectral slope (WSS) measured and (iv) computation time. We fed the speech signals filtered by the proposed bidirectional Kalman filter into a mel-frequency cepstral coefficient (MFCC)-based speech recognition system and reports our finding in Section 5. Section 6 presents the conclusion and recommendations for future work of this study.

2 Derivation of the discrete bidirectional Kalman filter

The Kalman filter estimates a process by using feedback control. The equations of the Kalman filter fall into two groups: time update equations (predictor equations) and measurement update equations (corrector equations). The time update equations are responsible for projecting forward (in time) the current time state and error covariance estimates to obtain the a priori estimates for the next time state. Our bidirectional Kalman filter addresses the general problem of trying to estimate the state $x \in \mathfrak{R}^n$ of a discrete-time controlled process that is governed by the linear stochastic difference equation

$$x_k = Ax_{k-1} + Bx_{k+1} + w_k \tag{1}$$

with a measurement $z \in \mathfrak{R}^m$ that is

$$z_k = Hx_k + v_k \tag{2}$$

The random variables w_k and v_k represent the process and measurement noise, respectively. They are assumed to be independent of each other, white with normal probability distributions

$$p(w) \sim N(0, Q) \tag{3}$$

$$p(v) \sim N(0, R) \tag{4}$$

The $n \times n$ matrix A and $n \times n$ matrix B in the difference (1) relate the state at the previous time state $k-1$ and the future time state $k+1$, respectively, to the state at the current time state k. Defining $\widehat{x_k^-}$ as an a priori state estimate at state k given knowledge of the process prior to state k, where $\hat{x}_k^- \in \mathfrak{R}^m$. $\hat{x}_k$ is an a posteriori state estimate at state k given measurement z_k, where $\hat{x}_k \in \mathfrak{R}^m$. The following is a list of the equations used in the calculation of each of the important variables

$$\hat{x}_k = E[x_k] \tag{5}$$

A priori estimate error

$$e_k^- = x_k - \hat{x}_k^- \tag{6}$$

A posteriori estimate error

$$e_k = x_k - \hat{x}_k \tag{7}$$

A posteriori estimate error covariance

$$P_k = E[e_k \ e_k^{\mathrm{T}}] \tag{8}$$

A priori estimate error covariance

$$P_k^- = E[e_k^- \ e_k^{-\mathrm{T}}]$$
$$= AP_{k-1}A^{\mathrm{T}} + BP_{k+1}B^{\mathrm{T}} + AP_{k-1,k+1}B^{\mathrm{T}} + BP_{k+1,k-1}A^{\mathrm{T}} + Q \tag{9}$$

$$P_{k+1,k-1}^- = E[e_{k+1}^- \ e_{k-1}^-]$$
$$= AP_{(k-1)+1,(k-1)-1}A^{\mathrm{T}} + AP_kB^{\mathrm{T}} + BP_{(k+1)+1,(k+1)-1}B^{\mathrm{T}}$$
$$+ Q + BE((x_{k+2} - \hat{x}_{k+2})(x_{k-2} - \hat{x}_{k-2})^{\mathrm{T}})A^{\mathrm{T}} \tag{10}$$

$$P_{k-1,k+1}^- = E[e_{k-1}^- \ e_{k+1}^-]$$
$$= AP_{(k-1)-1,(k-1)+1}A^{\mathrm{T}} + BP_kA^{\mathrm{T}} + BP_{(k+1)-1,(k+1)+1}B^{\mathrm{T}}$$
$$+ Q + AE((x_{k-2} - \hat{x}_{k-2})(x_{k+2} - \hat{x}_{k+2})^{\mathrm{T}})B^{\mathrm{T}} \tag{11}$$

The detail steps of the derivation of (9)–(11) are given in Appendices 1–3. There are two expected values $E((x_{k+2} - \hat{x}_{k+2})(x_{k-2} - \hat{x}_{k-2})^{\mathrm{T}})$ and $E((x_{k-2} - \hat{x}_{k-2})(x_{k+2} - \hat{x}_{k+2})^{\mathrm{T}})$ in (10) and (11), it can be predicted that continue derivation of these two (10) and (11) gives us

new expected values of $E((x_{k+n'} - \hat{x}_{k+n'})(x_{k-n'} - \hat{x}_{k-n'})^{\mathrm{T}})$ and $E((x_{k-n'} - \hat{x}_{k-n'})(x_{k+n'} - \hat{x}_{k+n'})^{\mathrm{T}})$ where n' is an integer that increases by one in every following derivation step. Continue derivation of these two equations may improve the accuracy of the proposed bidirectional Kalman filter but this makes both the equations endless. Therefore, for simplicity, we assume that vector differences $x_{k+2} - \hat{x}_{k+2}$ and $x_{k-2} - \hat{x}_{k-2}$ which are separated by four time states are uncorrelated, this gives us

$$E((x_{k+2} - \hat{x}_{k+2})(x_{k-2} - \hat{x}_{k-2})^{\mathrm{T}}) = 0 \tag{12}$$

$$E((x_{k-2} - \hat{x}_{k-2})(x_{k+2} - \hat{x}_{k+2})^{\mathrm{T}}) = 0 \tag{13}$$

Equations (10) and (11) become

$$P_{k+1,k-1}^- = AP_{(k-1)+1,(k-1)-1}A^{\mathrm{T}} + AP_kB^{\mathrm{T}}$$
$$+ BP_{(k+1)+1,(k+1)-1}B^{\mathrm{T}} + Q \tag{14}$$

$$P_{k-1,k+1}^- = AP_{(k-1)-1,(k-1)+1}A^{\mathrm{T}} + BP_kA^{\mathrm{T}}$$
$$+ BP_{(k+1)-1,(k+1)+1}B^{\mathrm{T}} + Q \tag{15}$$

The measurement update equations are responsible for the feedback, that is, for incorporating a new measurement into the a priori estimate to obtain an improved a posteriori estimate.

The Kalman filter computes an a posteriori state estimate $\hat{x}_k$ as a linear combination of a priori estimate $\hat{x}_k^-$ and a weighted difference between an actual measurement z_k and a measurement prediction $H\hat{x}_k^-$ as shown below

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \tag{16}$$

The difference $z_k - H\hat{x}_k^-$ in (16) is called the measurement innovation, or the residual. The residual reflects the discrepancy between the predicted measurement $H\hat{x}_k^-$ and the actual measurement z_k. A residual of zero means that the two are in complete agreement.

The $n \times n$ matrix K in (16) is chosen to be the gain or blending factor that minimises the a posteriori estimate error covariance (8)

$$K_k = P_k^- H^{\mathrm{T}}(HP_k^- H^{\mathrm{T}} + R)^{-1} \tag{17}$$

Equations (18)–(20) give the error covariance update

$$P_k = (I - KH)P_k^-(I - KH)^{\mathrm{T}} + KRK^{\mathrm{T}} \tag{18}$$

$$P_{k+1,k-1} = (I - K_{k+1}H)P_{k+1,k-1}^-(I - K_{k-1}H)^{\mathrm{T}}$$
$$+ K_{k+1}RK_{k-1}^{\mathrm{T}} \tag{19}$$

$$P_{k-1,k+1} = (I - K_{k-1}H)P_{k-1,k+1}^-(I - K_{k+1}H)^{\mathrm{T}}$$
$$+ K_{k-1}RK_{k+1}^{\mathrm{T}} \tag{20}$$

By substituting (17) into (18), error covariance update (18) can be simplified to

$$P_k = (I - KH)P_k^- \tag{21}$$

3 Determination of model parameters

The matrices A and B in the difference (1) relate the state at the previous time state $k-1$ and the future time state $k+1$ to the current time state k, respectively. To obtain faster processing speed, we reduce all the matrices involved in the Kalman filter computation process into (1×1) matrices. Then, we define x_k in the difference (1) as the speech signal s_k at time k

$$x_k = (s_k) \tag{22}$$

Next, to compute the constant matrices A and B, a linear bidirectional prediction recursion based on (23) is employed

$$\bar{s}_k = \alpha s_{k-1} + (1 - \alpha)s_{k+1} \qquad (23)$$

where the $\bar{s}$ represents the predicted speech signal. Measurement distance (MD) from the predicted speech signals to the original clean speech signals is computed using different values of α as shown in the following equation

$$\text{MD} = \sum_{i=1}^{n} (s_k - \bar{s}_k)^2 / n \qquad (24)$$

For each different value of α, MD is recorded from all the speech signal sentences contained inside the TIDIGIT training subset, where n represents the total number of speech data. The results obtained are plotted in Fig. 1 which shows a minimum distortion at the point $\alpha = 0.5$. From (23), it can be seen that parameter α relates the speech signal of the past time state to the current predicted speech signal, while parameter $(1-\alpha)$ relates the speech signal of the future time state to the current predicted speech signal. As a result, we define

$$A = (\alpha) \qquad (25)$$

$$B = (1 - \alpha) \qquad (26)$$

We set $A = (0.5)$ and $B = (0.5)$ since minimum distortion between the predicted and the original speech signals happens at $\alpha = 0.5$. The constant parameter H in (2) that relates the state x_k to the measurement z_k is fixed as $H = (1)$. In other words, x_k is equal to z_k.

As stated earlier, random variables w_k and v_k represent the process and measurement noise, respectively. In this study, the Kalman filter is used to process speech signals which are collected measurements from a microphone, as a result, we assume that the speech additive white noise is the measurement noise. Besides, we further assume that all the collected speech signals are silence (noise only) at the beginning, and at the ending. The (1×1) matrices Q and R which represents the standard deviation of the normally distributed measurement noise w_k and v_k in each speech sample is found using (27) and (28), respectively

$$Q = \left(\frac{\sqrt{E[s_{\text{intial,ending}}^2]}}{m'} \right) \qquad (27)$$

$$R = \left(\sqrt{E[s_{\text{intial,ending}}^2]} \right) \qquad (28)$$

where $s_{\text{intial, ending}}$ represents the speech data at the beginning, and at the ending of each collected speech sample. Table 1 shows the effects of different m' values to the correlation between the filtered
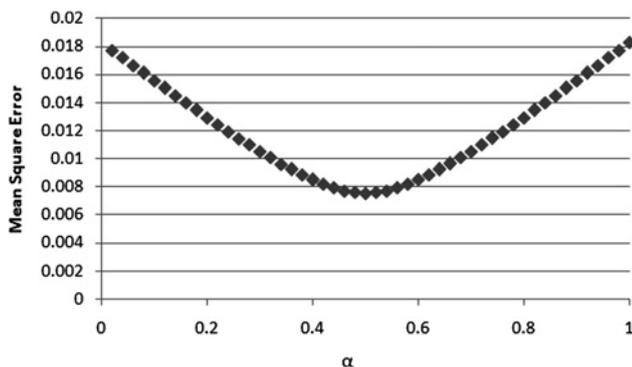

Fig. 1 *MD from predicted speech signals to original clean speech signals using different values of α*

Table 1 Correlation figure between filtered clean speech signals and filtered speech signals at different SNRs using different iteration numbers and different m' values of the proposed bidirectional Kalman filter

Number of iterations	m'					
	5		10		20	
	20 dB	−5 dB	20 dB	−5 dB	20 dB	−5 dB
1	0.969	0.716	0.962	0.738	0.958	0.718
2	0.978	0.722	0.970	0.748	0.962	0.727
3	0.988	0.728	0.979	0.756	0.969	0.730
4	0.994	0.729	0.987	0.763	0.978	0.744
5	0.996	0.730	0.992	0.767	0.984	0.753
6	0.998	0.730	0.995	0.769	0.989	0.760
7	0.998	0.731	0.997	0.769	0.992	0.764
8	0.999	0.731	0.998	0.771	0.994	0.768
9	0.999	0.731	0.998	0.771	0.996	0.769
10	0.999	0.731	0.999	0.771	0.997	0.771
11	0.999	0.731	0.999	0.771	0.997	0.771

clean speech signals and the filtered noisy speech signals (SNR = 20 dB and −5 dB). All speech signals contained inside the TIDIGIT testing subset were used in the test. Only small differences can be seen in the correlation figures at 20 dB region for all the iteration numbers. However, as for the −5 dB region, when it comes to steady state, bidirectional Kalman filter with higher m' shows a higher correlation figure. Besides, for the system using lower m', steady state can be reached at less iteration number. As a result, we pick $m' = 10$ in this study due to its higher correlation and its ability to converge faster.

4 Comparative study

To perform a comparison test, all speech data contained inside the TIDIGIT database testing subset with additive white noise at different SNRs (clean, 20, 15, 10, 5, 0 and −5 dB) were used. Each speech signal was filtered by the conventional Kalman filter [7], by the fast adaptive Kalman filter [17] and by the proposed bidirectional Kalman filter using different iteration numbers. These collected speech data were compared based on the following criteria: (i) correlation measured, (ii) SNR measured, (iii) WSS measured and (iv) computation time in this section. Fig. 2 shows the block diagrams of the speech denoising algorithm of three types of Kalman filters. For the conventional Kalman filter, the state-space model parameters of each frame are obtained directly from the same clean speech signal frame using linear predictive coding (LPC). Obtained parameters are feed into the conventional Kalman filter to process the noisy speech signals of the same speech frame. For both the fast adaptive Kalman filter and the proposed bidirectional Kalman filter, the calculation of the model parameters is omitted since the model parameters are fixed throughout the filtering process. Besides, to avoid overshoot problem happens in the filtered speech signal using the proposed bidirectional Kalman filter, if the absolute value of one particular filtered speech data at the certain time state exceeds 1.2 (original speech signal is scaled to make data peak −1 to 1), the value of that particular speech data will be set to equal to the original unfiltered speech data value.

4.1 Correlation figures

Table 2 shows the Pearson correlation figures between the filtered noisy speech signals at different SNRs (20, 15, 10, 5, 0 and −5 dB) and the filtered clean speech signals. Correlation figures between the unfiltered noisy speech signals and the clean speech signals are used as the reference figures. The higher the correlation figure, the more similar both the filtered clean and the filtered noisy speech signals are. Higher correlation figures between the clean and noisy speech signals filtered by the conventional Kalman filter can be achieved by using a larger number of
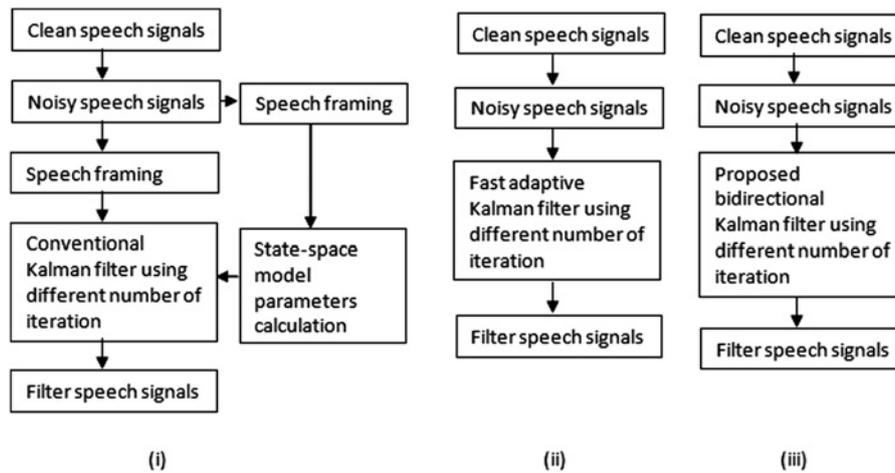
Fig. 2 *Block diagram of (i) the conventional Kalman filter (ii) fast adaptive Kalman filter and (iii) the proposed bidirectional Kalman filter in the speech denoising process*

iterations. As stated earlier, state-space model parameters for the conventional Kalman filter are obtained directly from the clean speech frame, this makes the correlation figures for the conventional Kalman filter converge to their steady state at the second iteration. Further iteration process (third iteration and above) cannot improves the system correlation figure. Recorded correlation figures at the third iteration are the same as the correlation figures measured at the second iteration.

Compared to the reference correlation figures, better correlation figures between the filtered clean and the filtered noisy speech signals can be achieved at the first iteration using the fast adaptive Kalman filter algorithm. When it goes to the second iteration, the only slight difference can be observed at SNRs higher than 5 dB. At −5 dB, the correlation figure is undefined. This implies that some of the filtered −5 dB speech signals have zero variance, these speech signals have become fixed DC signals after being filtered by the fast adaptive Kalman filter. In this case, the optimum iteration number for the fast adaptive Kalman filter in this paper has been set to one.

As for the proposed bidirectional Kalman filter, the system can only achieves a narrow improvement from the 10th iteration to the 11th iteration, this clearly implies that the proposed system reaches its steady state at the 11th iteration. Compared to the required

iteration number for the conventional Kalman filter to reach its steady state, this iteration number is higher. This is mainly caused by the fixed model parameters A and B that are used in the proposed Kalman filter are relatively lower in dimension compared to the state-space model parameters used in the conventional Kalman filter and thus larger number of iterations is needed to reach its steady state. At high SNR regions (20, 15 and 10 dB), compared to the unfiltered speech signals, the correlation figures between the filtered clean and noisy speech signals processed by the proposed bidirectional Kalman filter drop in certain values at the 1st iteration. Although these correlation figures drop, however, these values are still higher than 0.95, this shows that a small portion of the original speech information is lost during the filtering process. This problem can be solved by replacing the (1×1) matrices A and B with higher dimension matrices that relate more speech data in the future time state and the past time state to the current time state so that the differences between the predicted and the actual speech signals can be further reduced. At low SNR regions (5, 0 and −5 dB), correlation figures get improved follow by the increases in the number of iterations. Besides, for all SNR regions, correlation figures achieved by the conventional and the proposed Kalman filters are nearly the same and all these figures are slightly higher than the correlation figures of the fast adaptive Kalman filter at the steady state. Significant improvements can be observed for all three types of Kalman filters at the steady state at low SNR regions (5, 0 and −5 dB). Optimum number of iterations for the proposed bidirectional Kalman filter has been found to be 11.

4.2 SNR figures

The SNR of the filtered signal is obtained using the following equation

$$\text{SNR} = 10\log \frac{\text{KF}_n(s_{\text{clean}})}{\text{KF}_n(s_{\text{noise}}) - \text{KF}_n(s_{\text{clean}})} \quad (29)$$

where s represents speech signals at the clean or noisy environment and KF_n represents the energy level of the nth iteration Kalman filter processed speech signals. Table 3 shows the SNR values for three types of Kalman filters, these SNR figures show a similar trend as the earlier results of the correlation figures. Because of the relatively simple (1×1) matrices A and B used in the bidirectional Kalman filter compared to the (8×8) state-space model parameters obtained from the clean speech frame in the conventional Kalman filter, again the proposed Kalman filter requires more iterations to achieve its steady state. The number of iterations needed to reach the steady state for three types of Kalman filters is the same as the

Table 2 Correlation figure between filtered clean speech signals and filtered speech signals at different SNRs using different iteration numbers of the conventional Kalman filter, fast adaptive Kalman filter and the bidirectional Kalman filter

Number of iterations	20 dB	15 dB	10 dB	5 dB	0 dB	−5 dB
Unfiltered speech signal						
	0.997	0.989	0.968	0.907	0.772	0.564
Conventional Kalman filter, dB						
1	0.997	0.990	0.968	0.908	0.774	0.567
2	0.999	0.996	0.988	0.966	0.906	0.772
3	0.999	0.996	0.988	0.966	0.906	0.772
Fast adaptive Kalman filter, dB						
1	0.998	0.995	0.986	0.959	0.891	0.748
2	0.999	0.995	0.985	0.954	0.856	undefined
Bidirectional Kalman filter, dB						
1	0.962	0.958	0.953	0.935	0.876	0.738
2	0.970	0.965	0.959	0.940	0.883	0.748
3	0.979	0.974	0.966	0.947	0.889	0.756
4	0.987	0.983	0.976	0.956	0.897	0.763
5	0.992	0.989	0.982	0.961	0.901	0.767
6	0.995	0.992	0.985	0.964	0.904	0.769
7	0.997	0.994	0.987	0.965	0.905	0.770
8	0.998	0.995	0.988	0.966	0.906	0.771
9	0.998	0.996	0.989	0.967	0.906	0.771
10	0.999	0.996	0.989	0.967	0.906	0.771
11	0.999	0.996	0.989	0.967	0.907	0.771

Table 3 Speech signals SNR level after being processed by the conventional Kalman filter, fast adaptive Kalman filter and the bidirectional Kalman filter using different iteration numbers

Number of iterations	20 dB	15 dB	10 dB	5 dB	0 dB	−5 dB
Conventional Kalman filter, dB						
1	26.45	21.48	16.48	11.46	6.42	1.36
2	30.27	25.60	20.78	15.74	10.55	5.22
3	30.36	25.61	20.96	15.75	10.76	5.32
Fast adaptive Kalman filter, dB						
1	29.17	24.23	19.24	14.16	9.10	4.15
2	29.26	24.16	18.84	13.15	7.29	1.99
Bidirectional Kalman filter, dB						
1	16.54	15.91	14.92	12.67	8.92	4.19
2	17.33	16.55	15.37	12.97	9.12	4.32
3	18.91	17.82	16.21	13.41	9.33	4.43
4	21.12	19.63	17.48	14.08	9.60	4.53
5	23.09	21.19	18.45	14.53	9.77	4.59
6	25.04	22.57	19.21	14.84	9.87	4.62
7	26.68	23.61	19.70	15.01	9.92	4.62
8	28.03	24.39	20.04	15.13	9.96	4.63
9	28.99	24.87	20.21	15.18	9.97	4.63
10	29.65	25.19	20.34	15.22	9.98	4.63
11	30.06	25.35	20.39	15.23	9.98	4.63

Table 4 Speech signals WSS measures after being processed by the conventional Kalman filter, fast adaptive Kalman filter and the bidirectional Kalman filter using different iteration numbers

Number of iterations	Clean	20 dB	15 dB	10 dB	5 dB	0 dB	−5 dB
Unfiltered speech signal							
	0.00	22.06	26.30	30.83	35.54	40.27	44.79
Conventional Kalman filter, dB							
1	1.51	21.25	25.66	30.50	35.60	40.80	45.80
2	11.41	22.38	24.71	27.35	30.34	33.62	36.99
Fast adaptive Kalman filter, dB							
1	23.95	34.54	36.55	38.86	41.31	44.12	47.24
Bidirectional Kalman filter, dB							
1	12.99	29.07	31.57	34.12	36.94	40.09	43.50
2	12.79	27.37	29.68	32.00	34.63	37.71	41.14
3	10.74	27.56	29.95	32.35	34.98	38.01	41.32
4	10.37	25.87	28.33	30.85	33.65	36.88	40.36
5	8.36	25.29	27.89	30.56	33.51	36.82	40.31
6	9.19	24.19	26.90	29.73	32.89	36.40	40.03
7	6.77	23.62	26.48	29.49	32.78	36.36	40.00
8	6.72	22.94	25.90	29.04	32.49	36.21	39.93
9	5.79	22.55	25.62	28.90	32.44	36.19	39.92
10	5.65	22.15	25.29	28.67	32.32	36.13	39.90
11	5.22	21.93	25.15	28.61	32.30	36.13	39.89

results obtained in the correlation test before, 2 iterations are needed for the conventional Kalman filter and 11 iterations are needed for the bidirectional Kalman filter. As for the fast adaptive Kalman filter, at the second iteration, SNRs for 15 to −5 dB drop compared to the SNRs at the first iteration, this again shows that the optimum iteration number for the fast adaptive Kalman filter is one. Except for the first iteration at 20 dB region for the proposed bidirectional Kalman filter, when the number of iterations increases, the SNR level increases for the conventional and the proposed Kalman filters. This exception again shows that only a small portion of speech signals has lost during the proposed bidirectional Kalman filter filtering process. At the steady state, the conventional Kalman filter shows slightly higher SNR levels than the proposed bidirectional Kalman filter and the fast adaptive Kalman filter.

4.3 WSS distance measured

The WSS distance measured is a direct spectral distance measured. Recently, WSS has been studied extensively for objective speech quality measures [31]. It is based on the comparison of smoothed spectra from the clean and distorted speech samples. In this paper, the smoothed spectra are obtained from the MFCC-based cepstrum liftering. The implementation of WSS can be defined as follows

$$d_{\text{WSS}} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} W(j, m)(S_c(j, m) - S_d(j, m))^2}{\sum_{j=1}^{K} W(j, m)} \quad (30)$$

where K is the number of bands, M is the total number of frames, $S_c(j, m)$ and $S_d(j, m)$ are the spectral slopes of the jth band in the mth frame for clean and distorted speeches, respectively. Table 4 shows the WSS measures for the speech signals at different SNRs compared to the unfiltered clean speech signals. For the unfiltered speech signals, WSS measure for the clean speech is 0 since the test speech signals and the compared speech signals are exactly the same. When SNR goes higher, the WSS measure becomes higher since the distortion between the noisy speech and the clean speech gets larger. These WSS measures for unfiltered speech will act as the reference measures in this section.

For the conventional Kalman filter, the differences between the WSS measures for unfiltered speech signals and the first iteration speech signals are small at all different SNRs. At high SNR regions (clean and 20 dB), WSS measures increase when the iteration number goes from 1 to 2. This result is probably caused by the LPC parameters of the speech signals are directly obtained from the clean speech signals, over filtering happens. At other lower SNR regions, WSS measures at the second iteration (steady

state) show lower values than the WSS measures at the first iteration. For the fast adaptive Kalman filter algorithm, compared to the reference WSS measures, filtered speech signals show larger WSS measures, especially at the clean and high SNR regions, significant raise can be observed. This implies that lots of speech information lost during the filtering process.

For the proposed Kalman filter, WSS measures for the filtered speech signals at SNRs higher than 5 dB raise up at the first iteration filtering process. These WSS measures are higher than the WSS measures of the conventional Kalman filter, but lower than the WSS measures of the fast adaptive Kalman filter. This result indicates that the proposed bidirectional Kalman filter causes moderate speech information lost at the first iteration loop. For the following iterations, WSS measures are slowly reduced for all SNRs. At the steady state, except for the clean speech signal, WSS measures for all other SNRs are lower than the reference WSS measures. This result shows that although speech information lost during the proposed first iteration filtering process, however, the speech quality gets better at the steady state.

4.4 Computation time

The fast adaptive Kalman filter algorithm eliminates the matrix operations and only constantly updating the first value of the state vector. This algorithm shows the fastest processing speed among these three types of Kalman filters. In the conventional Kalman filter test, we first measured the state-space model parameters of each clean speech signal block using LPC. Then, using the obtained model parameters, the conventional Kalman filter process was run to filter the same speech signal block but contaminated with Gaussian white noise. As for the proposed bidirectional Kalman filter, since the model parameters A, B and H for the difference equation and the measurement equation constant throughout, the process of speech framing and model parameters computation are not necessary. Speech signals contaminated with the same noise as before was directly processed by the proposed bidirectional Kalman filter. Besides, these model parameters are also smaller in matrix dimension. Although the conventional Kalman filter only needs 2 iterations to reach its steady state while the proposed bidirectional Kalman filter requires 11 iterations, experimental result shows that the processing speed of the proposed bidirectional Kalman filter to reach its steady state is faster than the conventional Kalman filter to reach its steady state. Using a computer with Pentium Dual-Core 3 GHz processor, running Matlab R2009b under Windows XP operating system, to filter all the speech signals contained inside the TIDIGIT test subset, and to reach the steady state, respectively, the fast adaptive Kalman filter is 5.7 times faster than the proposed

Kalman filter. However, the proposed bidirectional Kalman filter is 7.9 times faster than the conventional Kalman filter.

5 Experimental study

The sample utterances used in this study were selected from the TIDIGIT speech database. The European Telecommunications Standards Institute (ETSI) front-end feature processing algorithm was applied to the speech signals and 12 MFCC-features were extracted. Hidden Markov model (HMM)-based speech recognition systems were built based on 12 MFCC-features, 1 energy level, 12 delta MFCC-features and 1 delta energy level. The HMM models consist of eleven 16-state continuous words (except silence and pause, that have 3 and 1 states, respectively), with 4 Gaussians per state, respectively. These HMM-based speech recognition systems were trained using 8598 clean sentences contained inside the TIDIGIT training subset. They were trained using (i) the original clean sentences, (ii) the filtered clean sentences using spectral subtraction method, (iii) the filtered clean sentences using the conventional Kalman filter and (iv) the filtered clean sentences using different number of iterations of the proposed bidirectional Kalman filter. All these HMM-based speech recognition systems were tested using 8700 sentences with no noise and with noise having SNR of 20, 15, 10, 5, 0 and −5 dB. The HTK tool kit was used in building HMM models.

Table 5 shows the WAcc of the MFCC-based speech recognition system using seven different types of speech signal as follows: (i) unfiltered speech signals, (ii) filtered speech signals processed by two iterations conventional Kalman filter (CKF2), (iii) filtered speech signals processed by spectral subtraction method (SS), (iv) filtered speech signals processed by one iteration fast adaptive Kalman filter (FAKF1) and (v) filtered speech signals processed by 6, 9 and 11 iterations bidirectional Kalman filter (BKF6, BKF9 and BKF11, respectively). For comparison purpose, another HMM-based speech recognition system based on 12 RASTA PLP-features (R-PLP), 1 energy level, 12 delta RASTA PLP-features and 1 delta energy level has been built. All types of speech recognition system results show that a decrease in SNR value causes the recognition rate to drop. We first look at the speech recognition rates using the proposed Kalman filter using different iteration numbers. At high SNR regions (clean, 20 and 15 dB), recognition rate drops slightly follows the increment in the iteration number, this is mainly caused by more speech information are lost due to the more iteration filtering process. At low SNR regions (5, 0 and −5 dB), the recognition rate improves significantly follows the increment in the iteration number. Comparing the proposed method to the unfiltered speech, all the recognition rates are similar at the high SNR regions, significant improvements are achieved at low SNR regions. Then, we compare the recognition rates achieved by the proposed Kalman filter at the steady state (11th iteration) to the conventional Kalman filter at the steady state (2nd iteration). Overall, the conventional Kalman filter shows better recognition rate then the proposed Kalman filter, especially at the high SNR regions. This is mainly caused by the larger dimension model parameters used in the conventional Kalman filter provides a more accurate state prediction than the smaller dimension fixed model parameters used in the proposed Kalman filter. Finally, we compare the proposed Kalman filter at the steady

state to the speech signals filtered by the spectral subtraction method and the RASTA PLP-based speech recognition systems. Both spectral subtraction and R-PLP-based methods perform better than the proposed Kalman filter at high SNR regions, but once the SNR is below or equal to 5 dB, the proposed Kalman filter shows significant better recognition rate than both the spectral subtraction and R-PLP-based methods. The speech recognition result on speech processed by the fast adaptive Kalman filter is not good. This is mainly caused by speech information lost during the filtering process. This data analysis shows that the proposed bidirectional Kalman filter does not improve the speech recognition rate at high SNR regions. However, as the SNR gets lower, the proposed bidirectional Kalman filter improves the robustness of the speech recognition system.

6 Conclusion

We have proposed a bidirectional Kalman filter that relates the future time state and the past time state to the current time state. Comparison results show that the correlation figures and the SNR figures for both the conventional and the proposed Kalman filters at the steady state are similar. Although the proposed bidirectional Kalman filter requires eleven iterations to reach its steady state while the conventional Kalman filter only requires two iterations, however, due to the smaller dimension and constant model parameters for the difference equations and measurement equations used in the bidirectional Kalman filter, experimental results show that the proposed Kalman filter is 7.9 times faster than the conventional Kalman filter. Besides, although the fast adaptive Kalman filter algorithm shows the best processing speed, however, due to the information lost, this algorithm cannot gives a promising result in speech recognition test. Overall, comparative study in speech recognition test shows that the proposed Kalman filter improves the robustness of the speech recognition system. By considering the system processing speed and the recognition rate, the proposed Kalman filter is therefore more suitable to be used in a practical speech processing system than the other two types of Kalman filter algorithms. Model parameters with larger dimension that relate more speech data from the future time state and the past time state to the current time state should be used in the future work so that the accuracy in the state prediction can be improved and the speech information lost can be reduced.

7 Acknowledgment

8 References

1 Paliwal, K., Basu, A.: 'A speech enhancement method based on Kalman filtering'. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 1987, vol. 12, pp. 177–180

2 Goh, Z., Tan, K., Tan, B.: 'Kalman-filtering speech enhancement method based on a voiced-unvoiced speech model', *IEEE Trans. Speech Audio Process.*, 1999, **7**, (5), pp. 510–524

3 Gabrea, M.: 'Adaptive Kalman filtering-based speech enhancement algorithm'. IEEE Canadian Conf. on Electrical and Computer Engineering, 2001, vol. 1, pp. 521–526

4 Gannot, S., Burshtein, D., Weinstein, E.: 'Iterative and sequential Kalman filter-based speech enhancement algorithms', *IEEE Trans. Speech Audio Process.*, 1998, **6**, (4), pp. 373–385

5 Lee, K., Jung, S.: 'Time-domain approach using multiple Kalman filters and em algorithm to speech enhancement with nonstationary noise', *IEEE Trans. Speech Audio Process.*, 2000, **8**, (3), pp. 282–291

6 Grivel, E., Gabrea, M., Najim, M.: 'Subspace state space model identification for speech enhancement'. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 1999, vol. 2, pp. 781–784

7 You, C., Rahardja, S., Soo Ngee Koh, *et al.*: 'Autoregressive parameter estimation for Kalman filtering speech enhancement'. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2007, vol. 4, pp. 891–913

8 Sorqvist, P., Handel, P., Ottersten, B.: 'Kalman filtering for low distortion speech enhancement in mobile communication'. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 1997, vol. 2, pp. 1219–1222

Table 5 WAcc for continuous digit recognition for speech recognition system with different types of filtering process

Types of filtering process	Clean	20 dB	15 dB	10 dB	5 dB	0 dB	−5 dB
unfiltered	99.1	91.4	86.5	75.5	61.4	21.3	7.29
CKF2	97.4	93.5	91.0	87.0	80.0	54.4	25.2
SS	97.49	95.62	90.41	79.42	62.39	32.16	11.57
FAKF1	79.0	69.2	60.7	55.2	41.4	30.6	11.0
BKF6	98.5	91.8	87.0	81.0	67.4	36.8	14.6
BKF9	98.1	91.7	87.5	81.2	70.5	51.1	18.7
BKF11	97.6	90.4	86.1	80.3	72.3	58.4	26.6
R-PLP	98.5	94.5	92.5	83.7	58.0	24.4	10.5

9   You, C., Koh, S., Rahardja, S.: 'Kalman filtering speech enhancement incorporating masking properties for mobile communication in a car environment'. IEEE Int. Conf. on Multimedia and Expo, 2004, vol. 2, pp. 1343–1346

10 Ma, N., Bouchard, M., Goubran, R.: 'Speech enhancement using a masking threshold constrained Kalman filter and its heuristic implementations', *IEEE Trans. Audio Speech Language Process.*, 2006, **14**, (1), pp. 19–32

11 Kuropatwinski, M., Kleijn, W.: 'Estimation of the short-term predictor parameters of speech under noisy conditions', *IEEE Trans. Audio Speech Language Process.*, 2006, **14**, (5), pp. 1645–1655

12 Jeong, S., Hahn, M.: 'Speech quality and recognition rate improvement in car noise environments', *Electron. Lett.*, 2001, **37**, (12), pp. 800–802

13 Ma, J., Deng, L.: 'Efficient decoding strategies for conversational speech recognition using a constrained nonlinear state-space model', *IEEE Trans. Speech Audio Process.*, 2003, **11**, (6), pp. 590–602

14 Mathe, M., Nandyala, S.P., Kishore Kumar, T.: 'Speech enhancement using Kalman filter for white, random and color noise'. IEEE Int. Conf. on Devices, Circuits and Systems (ICDCS), 2012, pp. 195–198

15 Mustiere, F., Bolic, M., Bouchard, M.: 'Improved colored noise handling in Kalman filter-based speech enhancement algorithms'. Canadian Conf. on Electrical and Computer Engineering, 2008, pp. 497–500

16 Wu, W., Chen, P.: 'Subband Kalman filtering for speech enhancement', *IEEE Trans. Circuits Syst. II, Analog Digital Signal Process.*, 1998, **45**, (8), pp. 1072–1083

17 Mai, Q., He, D., Hou, Y., Huang, Z.: 'A fast adaptive Kalman filtering algorithm for speech enhancement'. IEEE Conf. on Automation Science and Engineering (CASE), 2011, pp. 327–332

18 Shaughnessy, D.: 'Improving speech analysis methods for robust automatic recognition'. IEEE, Canadian Conf. on Electrical and Computer Engineering, 2004, vol. 1, pp. 161–164

19 Boll, S.: 'Suppression of acoustic noise in speech using spectral subtraction', *IEEE Trans. Acoustics Speech Signal Process.*, 1979, **27**, (2), pp. 113–120

20 Atal, B.: 'Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification', *J. Acoust. Soc. Am.*, 1974, **55**, (6), pp. 1304–1312

21 Hermansky, H., Morgan, N.: 'Rasta processing of speech', *IEEE Trans. Speech Audio Process.*, 1994, **2**, (4), pp. 578–589

22 Hermansky, H.: 'Perceptual linear predictive (plp) analysis of speech', *J. Acoust. Soc. Am.*, 1990, **87**, (4), pp. 1738–1752

23 Cui, X., Alwan, A.: 'Noise robust speech recognition using feature compensation based on polynomial regression of utterance snr', *IEEE Trans. Speech Audio Process.*, 2005, **13**, (6), pp. 1161–1172

24 Kovacevic, B., Milosavljevic, M., Veinovic, M.: 'Robust recursive ar speech analysis', *Signal Process.*, 1995, **44**, (2), pp. 125–138

25 Cohen, I., Berdugo, B.: 'Speech enhancement for non-stationary noise environments', *Signal Process.*, 2001, **81**, (11), pp. 2403–2418

26 Gales, M., Young, S.: 'Robust continuous speech recognition using parallel model combination', *IEEE Trans. Speech Audio Process.*, 1996, **4**, pp. 352–359

27 Leggetter, C., Woodland, P.: 'Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models', *Comput. Speech Language*, 1995, **9**, (2), p. 171

28 Cui, X., Gong, Y.: 'A study of variable-parameter Gaussian mixture hidden Markov modeling for noisy speech recognition', *IEEE Trans. Audio Speech Language Process.*, 2007, **15**, (4), pp. 1366–1376

29 Bryson, A., Frazier, M.: 'Smoothing for linear and nonlinear dynamic systems'. Proc. of the Optimum System Synthesis Conf., 1962, pp. 353–364

30 Fong, W., Godsill, S.J., Doucet, A., West, M.: 'Monte Carlo smoothing with application to audio signal enhancement, signal processing', *IEEE Trans.*, 2002, **50**, (2), pp. 438–449

31 Kondo, K.: 'Subjective quality measurement of speech' (Springer, 2012)

8   Appendices

8.1   *Appendix 1: Mathematical derivation 1*

Mathematical steps to derive the a priori estimate error covariance (9)

$$
\begin{aligned}
\boldsymbol{P}_k^- &= E[\boldsymbol{e}_k^- \boldsymbol{e}_k^{-\mathrm{T}}]\\
&= E[(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-)^{\mathrm{T}}]\\
&= E[(\boldsymbol{A}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1}) + \boldsymbol{B}(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1}))(\boldsymbol{A}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1}))^{\mathrm{T}}] + E[\boldsymbol{W}_{k-1}\boldsymbol{W}_{k-1}^{\mathrm{T}}]\\
&= E[\boldsymbol{A}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\\
&\quad + \boldsymbol{A}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1})^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1})(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1})(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1})^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}] + \boldsymbol{Q}\\
&= \boldsymbol{A}\boldsymbol{P}_{k-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{P}_{k+1}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{P}_{k-1,k+1}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{P}_{k+1,k-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{Q}
\end{aligned}
$$

8.2   *Appendix 2: Mathematical derivation 2*

Mathematical steps to derive the a priori estimate error covariance (10)

$$
\begin{aligned}
\boldsymbol{P}_{k+1,k-1}^- &= E[\boldsymbol{e}_{k+1}^- \boldsymbol{e}_{k-1}^{-\mathrm{T}}]\\
&= E[(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1}^-)(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1}^-)^{\mathrm{T}}]\\
&= E[(\boldsymbol{A}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k) + \boldsymbol{B}(\boldsymbol{x}_{k+2} - \hat{\boldsymbol{x}}_{k+2}))(\boldsymbol{A}(\boldsymbol{x}_{k-2} - \hat{\boldsymbol{x}}_{k-2})\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k))^{\mathrm{T}}] + \boldsymbol{Q}\\
&= E[\boldsymbol{A}(\boldsymbol{x}_{(k-1)+1} - \hat{\boldsymbol{x}}_{(k-1)+1})(\boldsymbol{x}_{(k-1)-1} - \hat{\boldsymbol{x}}_{(k-1)-1})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\\
&\quad + \boldsymbol{A}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_{k+2} - \hat{\boldsymbol{x}}_{k+2})(\boldsymbol{x}_{k-2} - \hat{\boldsymbol{x}}_{k-2})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_{(k+1)+1} - \hat{\boldsymbol{x}}_{(k+1)+1})(\boldsymbol{x}_{(k+1)-1} - \hat{\boldsymbol{x}}_{(k+1)-1})^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}] + \boldsymbol{Q}\\
&= \boldsymbol{A}\boldsymbol{P}_{(k-1)+1,(k-1)-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{P}_k\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{P}_{(k+1)+1,(k+1)-1}\boldsymbol{B}^{\mathrm{T}}\\
&\quad + \boldsymbol{Q} + \boldsymbol{B}E((\boldsymbol{x}_{k+2} - \hat{\boldsymbol{x}}_{k+2})(\boldsymbol{x}_{k-2} - \hat{\boldsymbol{x}}_{k-2})^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}
\end{aligned}
$$

8.3   *Appendix 3: Mathematical derivation 3*

Mathematical steps to derive the a priori estimate error covariance (11)

$$
\begin{aligned}
\boldsymbol{P}_{k-1,k+1}^- &= E[\boldsymbol{e}_{k-1}^- \boldsymbol{e}_{k+1}^{-\mathrm{T}}]\\
&= E[(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1}^-)(\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1}^-)^{\mathrm{T}}]\\
&= E[(\boldsymbol{A}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k) + \boldsymbol{B}(\boldsymbol{x}_{k-2} - \hat{\boldsymbol{x}}_{k-2}))(\boldsymbol{A}(\boldsymbol{x}_{k+2} - \hat{\boldsymbol{x}}_{k+2})\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k))^{\mathrm{T}}] + \boldsymbol{Q}\\
&= E[\boldsymbol{A}(\boldsymbol{x}_{(k-1)-1} - \hat{\boldsymbol{x}}_{(k-1)-1})(\boldsymbol{x}_{(k-1)+1} - \hat{\boldsymbol{x}}_{(k-1)+1})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\\
&\quad + \boldsymbol{A}(\boldsymbol{x}_{k-2} - \hat{\boldsymbol{x}}_{k-2})(\boldsymbol{x}_{k+2} - \hat{\boldsymbol{x}}_{k+2})^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\\
&\quad + \boldsymbol{B}(\boldsymbol{x}_{(k+1)-1} - \hat{\boldsymbol{x}}_{(k+1)-1})(\boldsymbol{x}_{(k+1)+1} - \hat{\boldsymbol{x}}_{(k+1)+1})^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}] + \boldsymbol{Q}\\
&= \boldsymbol{A}\boldsymbol{P}_{(k-1)-1,(k-1)+1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{P}_k\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{P}_{(k+1)-1,(k+1)+1}\boldsymbol{B}^{\mathrm{T}}\\
&\quad + \boldsymbol{Q} + \boldsymbol{A}E((\boldsymbol{x}_{k-2} - \hat{\boldsymbol{x}}_{k-2})(\boldsymbol{x}_{k+2} - \hat{\boldsymbol{x}}_{k+2})^{\mathrm{T}})\boldsymbol{B}^{\mathrm{T}}
\end{aligned}
$$