



HOKKAIDO UNIVERSITY

Title	Wireless OFDM Channel Estimation using a Stochastic Approach
Author(s)	萩原, 淳一郎
Degree Grantor	北海道大学
Degree Name	博士(工学)
Dissertation Number	甲第12485号
Issue Date	2016-12-26
DOI	https://doi.org/10.14943/doctoral.k12485
Doc URL	https://hdl.handle.net/2115/64462
Type	doctoral thesis
File Information	Junichiro_Hagiwara.pdf



Ph.D. Thesis

Wireless OFDM Channel Estimation
using a Stochastic Approach

Junichiro HAGIWARA

Graduate School of Information Science and Technology
Hokkaido University

Abstract

The rapidly increasing demand for wireless communication can be met using orthogonal frequency division multiplexing (OFDM) transmission. OFDM transmission is reliable, is suited to the wideband technology, and has already been implemented in various fields. Unfortunately, multipath fading is an unavoidable phenomenon in mobile communication. Nevertheless, it can be mitigated through the estimation of and compensation for the fading channel. Thus, improving channel estimation accuracy is essential for improving radio capacity. This thesis aims to improve wireless OFDM channel estimation using a powerful stochastic approach that can cope with the ambiguity. In particular, accurate channel estimation requires channel statistics to assess the effects of averaging. This thesis proposes novel methods that appropriately consider time-variable channel statistics in order to significantly improve the channel estimation accuracy. Furthermore, this thesis, for the first time, clarifies the practical difference between batch and sequential estimation processing, which are theoretically equivalent, for stationary processes.

This thesis is organized as follows. Chapter 1 provides an introduction and presents the background, purpose, and structure of the thesis. Chapter 2 provides the preliminary background of the study, including the investigatory assumptions, formulation of OFDM transmission, and fundamentals of stochastic filtering. Chapter 3 proposes a modeling scheme that appropriately reflects the channel statistics. This scheme comprises two methods: one for improving estimation accuracy and the other for reducing computational complexity. The first method is based on a state-space model and considers frequency correlation. The second method is based on the first one and forces the observation matrix into a sparse bidiagonal matrix to decrease the number of mathematical processes. The effectiveness of the proposed scheme is verified using numerical analysis. The coding gain achieved via the first method is up to 3 dB higher than that achievable via a conventional Kalman filter. The second method suppresses increased complexity by up to 2% compared with a conventional Kalman filter.

Chapter 4 proposes a joint channel and its statistics estimation method to track time-varying channel statistics. The channel frequency responses of the pilot subcarrier and its fixed hyperparameters (channel statistics) are estimated using a Liu and West filter (LWF), which is based on the state-space model and the sequential Monte Carlo method. To the best of our knowledge, for the first time, we demonstrate that a conventional LWF biases the hyperparameters because of its poor estimation of the likelihood caused by overfitting in noisy environments. Moreover, this prob-

lem cannot be solved using conventional smoothing techniques. Therefore, we modify the conventional LWF and regularize the likelihood using a Kalman smoother. The effectiveness of the proposed method is confirmed via numerical analysis. When both the Doppler frequency and the delay spread hyperparameters are unknown, the conventional LWF significantly degrades the performance, sometimes below that obtained by the least-squares estimation. By avoiding the failure of hyperparameter estimation, our method outperforms the conventional approach and achieves good performance near the lower bound. The coding gain in our proposed method is up to 10 dB higher than that in the conventional LWF.

Chapter 5 proposes a change-detection method to track abruptly changing channel statistics. Any such change should be immediately followed by channel estimation in order to retain a high estimation accuracy. This immediate follow-up can be achieved by resetting the channel estimation process according to adequate detection of the abrupt change. To this end, we investigate a method for detecting abrupt changes in the channel statistics using innovations obtained through a Kalman filter. We confirm the effectiveness of the method via computer simulations using delay spread as an example of channel statistics. The proposed method relies on the state-space model and enables detection on the basis of statistical theory.

Chapter 6 presents another novel investigation that clarifies characteristics based on estimation processing and practically compares the Wiener and Kalman filters for stationary processes. Under fair conditions, the performance and complexity of each method are numerically investigated. Comparison results show that the performance of the Wiener smoother is slightly better than that of the Kalman one; the former avoids cumulative errors in its sequential processing, while the complexity of the latter is always lower because there is no large matrix operation. Finally, a summary of the discussions in the preceding chapters is given in Chapter 7.

In wireless OFDM channel estimation, channel statistics have not been considered sufficiently until now, having been assumed to be either known or extractable. Our precise and pragmatic proposed methods should considerably improve the quality of mobile communication. Moreover, our proposal is mathematically formulated and is applicable to not only wireless technology but also related areas such as artificial intelligence and econometrics.

Contents

Nomenclature	iv
1 Introduction	1
1.1 Thesis Background	1
1.2 Thesis Purpose	4
1.3 Thesis Structure	5
Bibliography	6
2 Preliminary Background	9
2.1 Assumptions	9
2.2 OFDM Transmission	10
2.3 Stochastic Filtering	12
2.3.1 State-Space Model	12
2.3.2 Linear Filtering	13
2.3.3 Nonlinear Filtering	14
Bibliography	18
3 State-Space Modeling	21
3.1 Preliminary Remarks	21
3.2 Problem Formulation	21
3.3 Method 1: State-Space Model that Appropriately Considers Frequency Correlation	22
3.3.1 Formulation Using the State-Space Model	22
3.3.2 Specification of Hyperparameters	23
3.4 Method 2: Approximation Approach to Reduce Computa- tional Complexity	24
3.4.1 Estimation of Computational Complexity	25
3.4.2 Approximating the Observation Matrix into a Sparse Upper/Lower Bidiagonal Matrix	29
3.5 Numerical Analysis	30
3.5.1 Effects of Appropriately Considering Frequency Cor- relation	30
3.5.2 Effects of Reducing Computational Complexity	40

CONTENTS

3.6	Summary	42
	Bibliography	44
4	Joint Estimation	47
4.1	Preliminary Remarks	47
4.2	Background	48
4.2.1	Problem Formulation	48
4.2.2	Simulation Configuration	50
4.3	Theory	51
4.3.1	Primitive Method Based on Multi-Model	51
4.3.2	Modern Method Based on Sequential Monte Carlo	52
4.4	Methods	56
4.4.1	Why Should LWF be Improved?	56
4.4.2	Conventional Particle Smoothing	57
4.4.3	Novel Proposed Smoothing	58
4.5	Numerical Results and Discussion	62
4.5.1	The Case of a Single Unknown Hyperparameter (Velocity)	62
4.5.2	The Case of Two Unknown Hyperparameters (Velocity and Delay Spread)	67
4.6	Summary	73
	Bibliography	74
5	Detecting Abrupt Changes	79
5.1	Preliminary Remarks	79
5.2	Background	80
5.2.1	Problem Formulation	80
5.3	Theory and Method	81
5.4	Numerical Results and Discussion	81
5.4.1	Simulation Configuration	81
5.4.2	Extreme Change in Delay Spread	84
5.4.3	Moderate Change in Delay Spread	86
5.5	Summary	88
	Bibliography	89
6	Wiener and Kalman Filters: Practical Comparison	91
6.1	Preliminary Remarks	91
6.2	Background	92
6.2.1	Problem Formulation	92
6.3	Conditions for Fair Comparison	93
6.3.1	Type of Estimation	93
6.3.2	Fixed Time Lag L	94
6.4	Numerical Analysis	97
6.4.1	Performance Comparison	97

CONTENTS

6.4.2 Complexity Comparison	100
6.5 Summary	101
Bibliography	102
7 Conclusions	105
Bibliography	109
Appendix A Derivation of Covariance W in State Noise	111
Appendix B Kalman Filter: Another Representation for Complexity Investigation	113
Appendix C How to Derive the Nearest Positive Definite Matrix for a Forced Tridiagonal Correlation Matrix	115
Bibliography	117
Appendix D RB-APF	119
Appendix E RB-APF with the Kernel Smoothing Technique	121
Appendix F Throughput Efficiency	123
Bibliography	126
Acknowledgements	127
Publications	129

Nomenclature

Acronyms

3GPP	3rd Generation Partnership Project
APF	auxiliary particle filter
AR	autoregressive
ARQ	automatic repeat request
BER	bit error rate
CFR	channel frequency response
CIR	channel impulse response
CRSs	cell-specific reference signals
DC	direct current
FFT	fast Fourier transform
GI	guard interval
IMM	interacting multiple model
LMMSE	linear minimum mean squared error
LS	least-squares
LTE	Long Term Evolution
LWF	Liu and West filter
LWS	Liu and West smoother
MAP	maximum a posteriori probability
MCMC	Markov chain Monte Carlo
MIMO	multiple-input multiple-output

NOMENCLATURE

ML	maximum likelihood
MMSE	minimum mean squared error
NMSE	normalized mean squared error
OFDM	orthogonal frequency division multiplexing
PF	particle filter
QPSK	quadrature phase-shift keying
RB	Rao–Blackwellisation
RB-APF	Rao–Blackwellised auxiliary particle filter
RB-LWF	Rao–Blackwellised Liu and West filter
RB-LWS	Rao–Blackwellised Liu and West smoother
RB-PF	Rao–Blackwellised particle filter
RBs	resource blocks
RLS	recursive least-squares
RTS	Rauch–Tung–Striebel
SIR	sequential importance resampling
SIS	sequential importance sampling
SVD	singular value decomposition
WSSUS	wide-sense stationary uncorrelated scattering
ZF	zero forcing

General Notation

$a, \lambda, \dots$	Scalar
$\mathbf{a}, \boldsymbol{\lambda}, \dots$	Column vector
$\mathbf{A}, \boldsymbol{\Lambda}, \dots$	Matrix

Chapter 1

Introduction

1.1 Thesis Background

The use of wireless communication devices such as mobile phones has increased in recent years, demanding greater wireless speed and capacity. These demands can be fulfilled by orthogonal frequency division multiplexing (OFDM) transmission. OFDM efficiently divides wideband (high-rate) signals into orthogonal narrowband (low-rate) signals, as shown in Figure 1.1 [1.1]. OFDM transmission is reliable and well suited to wideband technology, even in mobile environments. Thus, it has already been implemented in 3rd Generation Partnership Project Long Term Evolution (3GPP LTE), wireless local area networks, and terrestrial digital broadcasting.

Unfortunately, multipath fading is unavoidable in mobile communications. One solution to this problem is estimating and compensating for the fading channel. While differential detection does not necessarily re-

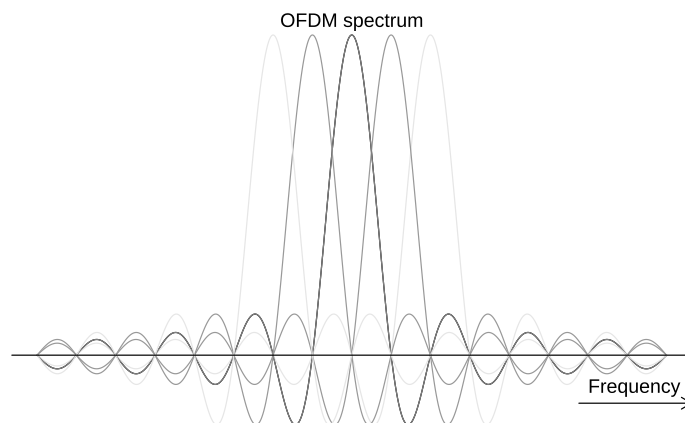


Figure 1.1: Principle of OFDM.

CHAPTER 1. INTRODUCTION

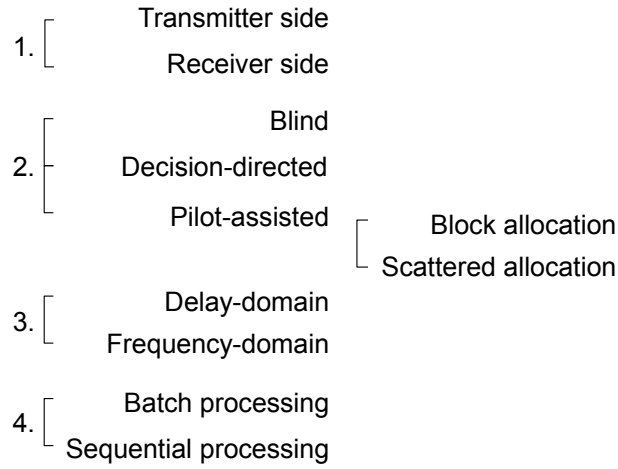


Figure 1.2: Classification of channel estimation methods.

quire channel estimation and compensation [1.2,1.3], coherent detection is currently very popular for accurate communication and requires channel estimation and compensation. Improving the channel estimation accuracy is essential for improving radio capacity. This thesis investigates this improvement using a stochastic approach. Stochastic theory is a universal and powerful tool capable of addressing ambiguity and has been successfully applied in the fields of artificial intelligence and econometrics; however, its potential applications in the communications field are far from fully realized.

Figure 1.2 illustrates the classification of channel estimation methods. Referring to this classification scheme, in viewpoint 1, the transmitter side makes the sending signal pre-distorted according to the channel information fed back from the receiver side [1.4]; rapid time selectivity degrades the estimation accuracy because of the feedback delay. In viewpoint 2, blind estimation must estimate the data symbols as well as channel [1.5, 1.6]. Such a simultaneous estimation is more difficult than other channel estimations that use auxiliary information, and the estimation accuracy is usually insufficient. Decision-directed estimation iteratively improves the provisional channel by using its detection output [1.7, 1.8]. However, this approach actually supposes data coding for accurate estimation and suffers from decoding delays including de-interleaving. Meanwhile, pilot-assisted estimation estimates the channel using a pre-determined specific signal called the pilot. Figure 1.3 shows an example of pilot allocation. Block allocation [1.9] cannot accommodate rapid time selectivity. In viewpoint 3, the delay- and frequency-domain channels correspond to the channel impulse response (CIR) and channel frequency response (CFR), respectively. Figure 1.4 depicts these two essentially equivalent domain channels [1.10]. The principle of OFDM naturally enables frequency-

1.1. THESIS BACKGROUND

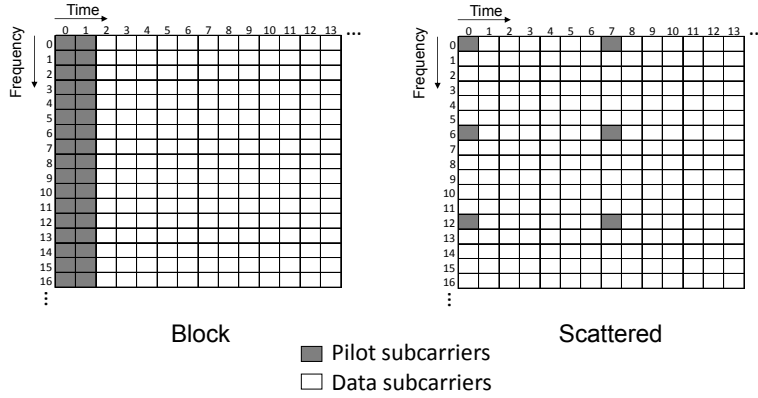


Figure 1.3: Pilot allocation example.

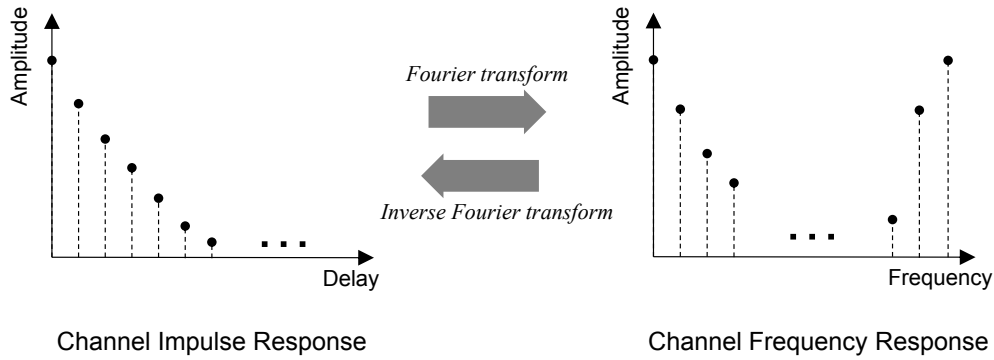


Figure 1.4: Relationship between CIR and CFR.

domain signal processing [1.11]. In viewpoint 4, the batch processing considers multiple received symbols simultaneously [1.12,1.13] and is suitable for off-line analysis — principally for unavoidable process delays. Conversely, sequential processing [1.14–1.17] considers each received symbol individually and is suitable for real-time processing. This thesis supposes the following combination for application to a real OFDM communication device with high estimation performance.

1. Receiver side
2. Pilot-assisted and scattered allocation
3. Frequency-domain
4. Sequential processing

Stochastic estimation for sequentially correlated data such as a channel has a rich history. Kolmogorov and Wiener are widely regarded as the first to solve this problem for linear stationary processes in the late 1930s

CHAPTER 1. INTRODUCTION

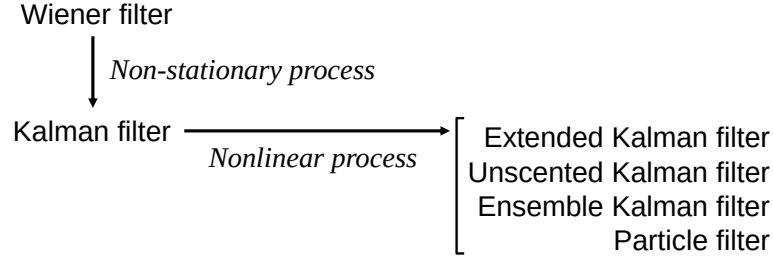


Figure 1.5: History of the stochastic filter.

and early 1940s [1.18]. Kalman then solved this problem for linear non-stationary processes using the state-space model [1.19]. Further efforts to overcome nonlinearity have succeeded the Kalman filter. Figure 1.5 summarizes the historical development of these filters. The Wiener filter [1.20] is typically used for batch processing and can be applied to linear stationary processes. In contrast, the Kalman filter is typically used for sequential processing and can be applied to linear non-stationary processes. The extended Kalman filter [1.21] approximates a nonlinear function in a state-space model as a linear function by Taylor expansion. Meanwhile, the unscented Kalman filter [1.22] approximates the posterior density of state as a normal distribution by a transformation using representative so-called “sigma points.” The ensemble Kalman filter [1.23] approximates the state in the state-space model to its many discrete instances. In contrast, the particle filter [1.24, 1.25] approximates the posterior density of state by its many pairs of realizations and weights. This thesis applies the celebrated Kalman filter for linear processes owing to its applicability for sequential processing and the simple but powerful particle filter for nonlinear processes owing to no principal limitations.

1.2 Thesis Purpose

Based on the information presented above, this thesis aims to improve the channel estimation accuracy for the CFR of scatter allocated pilot in sequential processing on the receiver side using Kalman and particle filters. Accurate channel estimation requires channel statistics to obtain the averaging effects; thus, this thesis proposes novel methods for appropriately considering the time-variable channel statistics to dramatically improve the channel estimation accuracy. This effort also clarifies the practical difference between batch and sequential estimation processing for stationary processes for the first time. All our proposals are mathematically formulated and applicable to not only wireless technology but also a broad range of related fields such as artificial intelligence and econometrics.

1.3 Thesis Structure

The remainder of this thesis is structured as shown in Figure 1.6. Chapter 2 provides the preliminary thesis background, including the investigation assumptions, formulation in OFDM transmission, and fundamentals of stochastic filtering. Chapters 3–5 present novel investigations aimed at improving the channel estimation accuracy. First, Chapter 3 proposes a modeling for including the channel statistics in a state-space model and its approximation modeling for complexity reduction. Chapter 4 proposes a joint channel and its statistics estimation method to follow time-varying channel statistics. Chapter 5 proposes a change detection method for following abruptly changing channel statistics. The effectiveness of these proposals is verified through a computer simulation based on LTE downlink. Next, Chapter 6 describes a novel investigation intended to clarify the characteristics based on estimation processing, wherein a practical comparison between Wiener and Kalman filters for stationary processes reveals a trade-off between the estimation accuracy and computational complexity. This comparison is also confirmed through computer simulation based on LTE downlink. Finally, Chapter 7 summarizes the findings of this thesis as a whole.

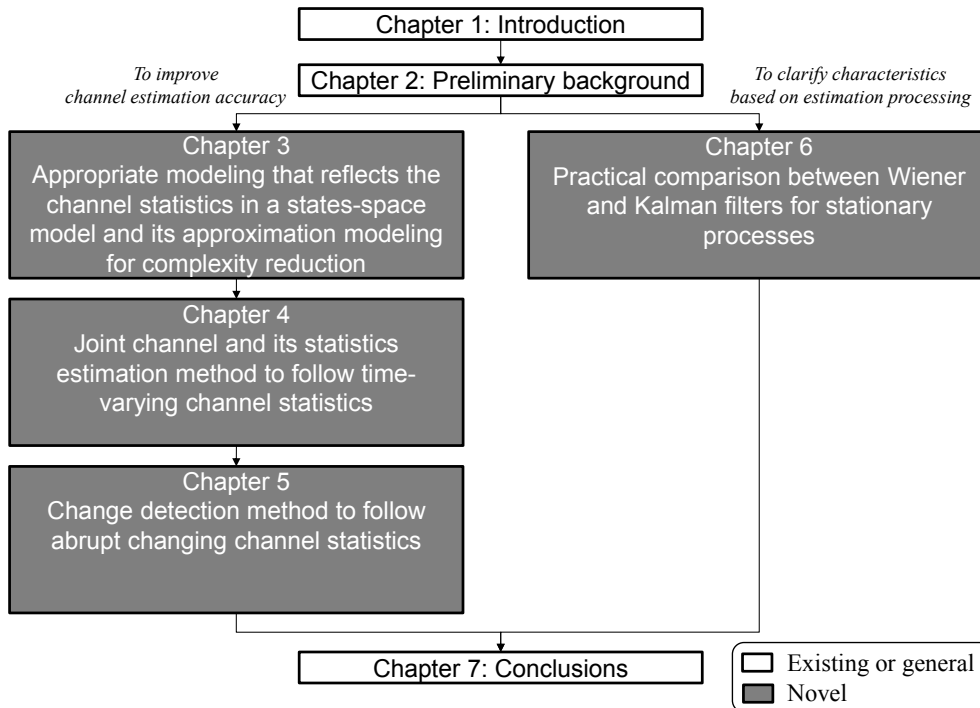


Figure 1.6: Thesis structure.

CHAPTER 1. INTRODUCTION

Bibliography

- [1.1] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. John Wiley & Sons, 2003.
- [1.2] T. Himsoon, W. Su, and K. J. R. Liu, "Single-block differential transmit scheme for broadband wireless MIMO-OFDM systems," *IEEE Trans. Signal Processing*, vol. 54, no. 9, pp. 3305–3314, Sep. 2006.
- [1.3] W. Hashim, F. Said, B. Allen, A. H. Aghvami, and A. F. Ismail, "Performance comparison of differential space-time signalling schemes for OFDM systems," in *Proc. IEEE 13th International Conference on Networks (ICN) jointly held with IEEE 7th Malaysia International Conference on Communications*, vol. 2, Nov. 2005, pp. 1–6.
- [1.4] A. I. El-Arabawy and S. C. Gupta, "Reduced mobile complexity scheme for fast fading channel estimation in OFDM-FDD mobile communication systems," in *Proc. IEEE 6th International Conference on Universal Personal Communications*, vol. 1, Oct. 1997, pp. 274–278.
- [1.5] S. Wu and Y. Bar-Ness, "OFDM channel estimation in the presence of frequency offset and phase noise," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 5, May 2003, pp. 3366–3370.
- [1.6] B. Han, X. Gao, X. You, J. Wang, and E. Costa, "An iterative joint channel estimation and symbol detection algorithm applied in OFDM system with high data to pilot power ratio," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 3, May 2003, pp. 2076–2080.
- [1.7] M. Bossert, A. Donder, and A. Trushkin, "Channel estimation and equalization in orthogonal frequency division multiplexing systems," in *Proc. ITG-Fachbericht, Mobile Kommunikation*, Sep. 1995, pp. 485–492.
- [1.8] V. Mignone, A. Morello, and M. Visintin, "CD3-OFDM: a new channel estimation method to improve the spectrum efficiency in digital terrestrial television systems," in *Proc. International Broadcasting Convention (IBC)*, Sep. 1995, pp. 122–128.
- [1.9] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications — Amendment 4: Enhancements for very high throughput for operation in bands below 6 GHz," IEEE Std. 802.11ac, 2013.

BIBLIOGRAPHY

- [1.10] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 2, pp. 18–48, Apr. 2007.
- [1.11] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [1.12] P. Höher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, Apr. 1997, pp. 1845–1848.
- [1.13] —, "Pilot-symbol-aided channel estimation in time and frequency," in *Multi-carrier Spread-spectrum*, K. Fazel and G. P. Fettweis, Eds. Norwell, MA, USA: Kluwer Academic Publishers, 1997, pp. 169–178.
- [1.14] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Communications*, vol. 3, no. 1, pp. 315–325, Jan. 2004.
- [1.15] W. H. He and Y. Lee, "Low-complexity Kalman channel estimator structures for OFDM systems with and without virtual carriers," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 4, Jun. 2004, pp. 2447–2451.
- [1.16] W. Chen and R. Zhang, "Estimation of time and frequency selective channels in OFDM systems: a Kalman filter structure," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 2, Nov. 2004, pp. 800–803.
- [1.17] M. Liyanage and I. Sasase, "Steady-state Kalman filtering for channel estimation in OFDM systems for Rayleigh fading channels," *IEICE Trans. Communications*, vol. E92-B, no. 7, pp. 2452–2460, 2009.
- [1.18] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall Upper Saddle River, NJ, 2000.
- [1.19] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering, Transactions of the ASME*, vol. 82 (Series D), no. 1, pp. 35–45, Mar. 1960.
- [1.20] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Technology Press and Wiley, NY, 1949.
- [1.21] L. A. McGee and S. F. Schmidt, "Discovery of the Kalman filter as a practical tool for aerospace and industry," National Aeronautics and Space Administration, Tech. Rep., 1985.

CHAPTER 1. INTRODUCTION

- [1.22] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proc. American Control Conference*, vol. 3, Jun. 1995, pp. 1628–1632.
- [1.23] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research: Oceans*, vol. 99, no. C5, pp. 10,143–10,162, 1994.
- [1.24] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2, pp. 107–113, Apr. 1993.
- [1.25] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, Mar. 1996.

Chapter 2

Preliminary Background

2.1 Assumptions

To estimate the sequential OFDM channel, we estimate the CFR (that is narrow-band channel gain) of the pilot subcarriers and interpolate these gains for the non-pilot subcarriers. The interpolation procedure is beyond the scope of this study, as various methods already exist [2.1]. A small-scale fading channel for known pilot signals is sequentially estimated on the receiver side in an OFDM transmission environment. The frequency interval between the OFDM subcarriers is designed sufficiently wide to ignore inter-subcarrier interferences caused by the Doppler effect. The pilot subcarrier is not coded and is scattered in a comb pattern [2.2] in the time and frequency domains. We denote the time and frequency intervals between adjacent pilot subcarriers by dt and df , respectively. Except for the temporal change of channel statistics, multipath wave characteristics are determined by wide-sense stationary uncorrelated scattering (WSSUS) [2.3] and are invariant over the OFDM symbol duration. The maximum multipath wave delay is shorter than the OFDM guard interval (GI). Although modeling the time and frequency selectivity requires many independent measurements, we adopt typical models of land mobile communications [2.4, 2.5] for simplicity. Specifically, we base time selectivity on the Jakes' model [2.6] and frequency selectivity on the exponential delay profile [2.6]. The nonlinearity of the employed analog radio circuit and imperfections in time, phase, and frequency are negligible in coherent detections. We consider only a single-input, single-output environment.

We now define some essential notations. $\mathbf{E}_k$ denotes the $k \times k$ identity matrix. The transpose and complex-conjugate transpose of any matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$ and $\mathbf{A}^H$, respectively. Determinant and inverse of square matrix $\mathbf{A}$ are denoted by $\det(\mathbf{A})$ and $\mathbf{A}^{-1}$, respectively. For Hermitian positive semidefinite matrix $\mathbf{A}$, its square root matrix is denoted by $\mathbf{A}^{1/2}$ which satisfies $\mathbf{A} = \mathbf{A}^{1/2}(\mathbf{A}^{1/2})^H$. The matrix $\text{diag}(A_0, A_1, \dots)$ con-

CHAPTER 2. PRELIMINARY BACKGROUND

tains the elements $A_0, A_1, \dots$ in its sequential diagonal entries. $(\mathbf{A})_{\text{row}, \text{col}}$ denotes the row-th and column-th entry of $\mathbf{A}$. $\otimes$ denotes the Kronecker product. $\mathbf{0}_k$ denotes the $k \times 1$ zero vector. A random vector $\mathbf{r}$ having an independent complex normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Gamma}$ is expressed as $\mathbf{r} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$. $\mathcal{CN}(\mathbf{r}; \boldsymbol{\mu}, \boldsymbol{\Gamma})$ denotes the density of the complex normal distribution at point $\mathbf{r}$. $p(\cdot)$ denotes probability densities or distributions, whereas $P(\cdot)$ denotes probability. $\propto$ denotes proportionality. The expectation, weighted sample mean, covariance and weighted sample covariance of a quantity $\cdot$ are denoted by $E[\cdot]$, $E_\star[\cdot]$, $\text{Var}[\cdot]$, and $\text{Var}_\star[\cdot]$, respectively. The estimator, absolute value, argument, and complex conjugate value of $\cdot$ are expressed as $\hat{\cdot}$, $|\cdot|$, $\arg(\cdot)$, and $\cdot^*$, respectively. $\{\cdot\}$ indicates a set, and $t:t+L$ denotes a discrete time duration $\{t, t+1, \dots, t+(L-1), t+L\}$. j is the imaginary unit $\sqrt{-1}$. $\mathcal{O}(u)$ denotes that complexity is proportional to u .

2.2 OFDM Transmission

Figure 2.1 shows the OFDM transmitter and receiver in an equivalent baseband system [2.7]. On the transmitter side, information bits are parallelized and modulated to the symbols of each subcarrier. An inverse fast Fourier transform is applied to these symbols, which are then serialized. A GI is appended to the output before the OFDM signal is transmitted. On the receiver side, the GI is removed from the received OFDM signal, the output is parallelized, and a fast Fourier transform is applied to the symbols of each subcarrier. These symbols are demodulated and serialized to detect information bits. The relationship between the received and transmitted symbols is expressed as

$${}^F\mathbf{y}(t) = {}^F\mathbf{S}(t){}^F\mathbf{B}\mathbf{h}(t) + {}^F\mathbf{v}(t), \quad (2.1)$$

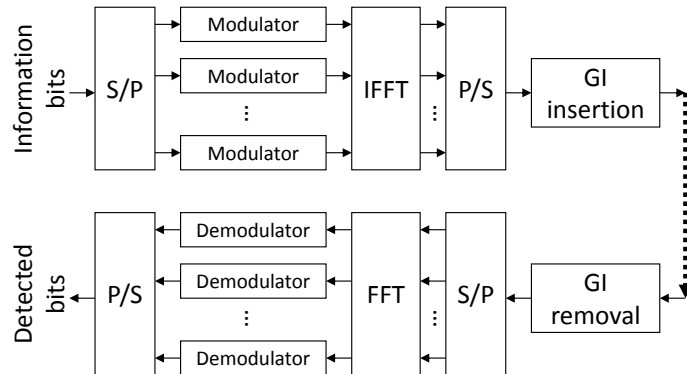


Figure 2.1: Diagram of the OFDM transmitter and receiver in an equivalent baseband system.

2.2. OFDM TRANSMISSION

where left superscript “ F ” means “full subcarrier representation,” ${}^F\mathbf{y}(t) = [y_0(t), \dots, y_n(t), \dots, y_{N-1}(t)]^T$ denotes the N -by-1 received symbol vector at the t -th symbol, and its subscript denotes a subcarrier index in the frequency domain. N corresponds to the maximum number of subcarriers in one symbol duration. ${}^F\mathbf{S}(t) = \text{diag}(s_0(t), \dots, s_n(t), \dots, s_{N-1}(t))$ denotes the N -by- N transmitted symbol matrix at the t -th symbol. ${}^F\mathbf{B} = [\mathbf{b}_0, \dots, \mathbf{b}_n, \dots, \mathbf{b}_{N-1}]^T$ denotes the N -by- D Fourier transform matrix, and $\mathbf{b}_n = [\exp(-j2\pi n0/N), \dots, \exp(-j2\pi nd/N), \dots, \exp(-j2\pi n(D-1)/N)]^T$. $\mathbf{h}(t) = [h_0(t), \dots, h_d(t), \dots, h_{D-1}(t)]^T$ denotes a D -by-1 wideband channel impulse response vector at the t -th symbol, and its subscript denotes the normalized delay of a multipath element wave. $D-1$ corresponds to the maximum delay of the multipath element wave. ${}^F\mathbf{v}(t) \sim \mathcal{CN}(\mathbf{0}_N, {}^F\mathbf{V})$ denotes the N -by-1 additive white Gaussian noise vector at the t -th symbol, and when σ^2 is assumed to be the average noise power of each subcarrier, ${}^F\mathbf{V}$ can be expressed as ${}^F\mathbf{V} = \sigma^2 \mathbf{E}_N$.

Setting ${}^F\tilde{\mathbf{h}}(t) = {}^F\mathbf{B}\mathbf{h}(t)$ leads to the following equation:

$${}^F\mathbf{y}(t) = {}^F\mathbf{S}(t){}^F\tilde{\mathbf{h}}(t) + {}^F\mathbf{v}(t), \quad (2.2)$$

where

$$\begin{aligned} {}^F\tilde{\mathbf{h}}(t) &= {}^F\mathbf{B}\mathbf{h}(t) \\ &= \begin{bmatrix} \sum_{d=0}^{D-1} h_d(t) \exp(-j2\pi 0d/N), \dots, \\ \sum_{d=0}^{D-1} h_d(t) \exp(-j2\pi nd/N), \dots, \\ \sum_{d=0}^{D-1} h_d(t) \exp(-j2\pi (N-1)d/N) \end{bmatrix}^T \\ &= [\tilde{h}_0(t), \dots, \tilde{h}_n(t), \dots, \tilde{h}_{N-1}(t)]^T. \end{aligned} \quad (2.3)$$

$\tilde{h}_n(t)$ in Equation (2.3) denotes the narrow-band channel gain of the n -th subcarrier at the t -th symbol. There are two approaches for deriving ${}^F\tilde{\mathbf{h}}(t)$. One approach (“CIR approach”) [2.2, 2.8] derives ${}^F\tilde{\mathbf{h}}(t)$ via $h_d(t)$, and thus, directly estimates the wideband channel impulse responses. The other approach (“CFR approach”) [2.9, 2.10] derives ${}^F\tilde{\mathbf{h}}(t)$ via $\tilde{h}_n(t)$, and thus, estimates the narrow-band channel gain for multiple subcarriers. These two approaches are essentially equivalent, but each approach has its own features. For CIR approach, a typical accuracy-improving technique is to avoid unnecessary $h_d(t)$ estimation by limiting the maximum delay or path position of a multipath wave [2.11]; CFR approach does not need such consideration. For CFR approach, however, the frequency correlation between elements from Equation (2.3) should be adequately considered in

CHAPTER 2. PRELIMINARY BACKGROUND

order to improve estimation accuracy. Such a consideration of frequency correlation leads to an averaging effect in the frequency domain; conventional methods based on CFR approach do not sufficiently consider this point. This thesis applies the CFR approach based on the OFDM features described in Chapter 1 and further considers the frequency correlation in an appropriate manner.

Because the main scope of this thesis is limited to channel estimation for pilot subcarriers, the above expression for full subcarriers should be modified to that for pilot subcarriers only, as in [2.10]. For this purpose, superscript F is removed from vector/matrix variables and the subscript of its element is labeled $i = 0, \dots, I - 1$ for pilot subcarriers, instead of $n = 0, \dots, N - 1$ for full subcarriers. I corresponds to the maximum number of pilot subcarriers in one symbol duration. Thus, Equation (2.2) is modified as

$$\begin{aligned}\mathbf{y}(t) &= \mathbf{S}(t)\tilde{\mathbf{h}}(t) + \mathbf{v}(t) \\ &= [y_0(t), \dots, y_i(t), \dots, y_{I-1}(t)]^T,\end{aligned}\tag{2.4}$$

where

$$\begin{aligned}\mathbf{S}(t) &= \text{diag}(s_0(t), \dots, s_i(t), \dots, s_{I-1}(t)), \\ \tilde{\mathbf{h}}(t) &= [\tilde{h}_0(t), \dots, \tilde{h}_i(t), \dots, \tilde{h}_{I-1}(t)]^T, \\ \mathbf{v}(t) &\sim \mathcal{CN}(\mathbf{0}_I, \mathbf{V}), \quad \mathbf{V} = \sigma^2 \mathbf{E}_I.\end{aligned}$$

Recall that our aim is to estimate $\tilde{\mathbf{h}}(t)$ from the received $\mathbf{y}(t)$ with known $\mathbf{S}(t)$.

2.3 Stochastic Filtering

2.3.1 State-Space Model

A state-space model [2.12] is a flexible framework that enables sequential Bayesian inference for the latent variable (state).

The linear Gaussian state-space model discussed in this thesis consists of the observation equation (Equation (2.5)) and the state equation (Equation (2.6)), and the Kalman filter [2.13] provides the optimal solution for such a state-space model:

$$\mathbf{y}(t) = \mathbf{F}(t) \boldsymbol{\theta}(t) + \mathbf{v}(t),\tag{2.5}$$

$$\boldsymbol{\theta}(t) = \mathbf{G}(t) \boldsymbol{\theta}(t-1) + \mathbf{w}(t),\tag{2.6}$$

where $\mathbf{y}(t)$ denotes the I -by-1 observation vector for pilot subcarriers at the t -th symbol (as in Equation (2.4)), $\mathbf{F}(t)$ denotes the I -by- I observation matrix for pilot subcarriers at the t -th symbol, $\boldsymbol{\theta}(t) = [\theta_0(t), \dots, \theta_i(t),$

2.3. STOCHASTIC FILTERING

$\dots, \theta_{I-1}(t)]^T$ denotes the I -by-1 linear Gaussian state vector for pilot subcarriers at the t -th symbol, and $\mathbf{G}(t)$ denotes the I -by- I transition matrix for pilot subcarriers at the t -th symbol. $\mathbf{v}(t) \sim \mathcal{CN}(\mathbf{0}_I, \mathbf{V})$ denotes the I -by-1 observation noise vector for pilot subcarriers at the t -th symbol; its covariance matrix is set to $\mathbf{V} = \text{diag}(V_0, \dots, V_i, \dots, V_{I-1})$ (as in Equation (2.4)). $\mathbf{w}(t) \sim \mathcal{CN}(\mathbf{0}_I, \mathbf{W})$ denotes the I -by-1 state noise vector for pilot subcarriers at the t -th symbol; its covariance matrix is set to $\mathbf{W} = \text{diag}(W_0, \dots, W_i, \dots, W_{I-1})$.

Furthermore, the general (for example, nonlinear or non-Gaussian) state-space model discussed in this thesis comprises the observation equation (Equation (2.7)) and the state equation (Equation (2.8)), and the particle filter [2.14, 2.15] provides the approximate optimal solution for such a general state-space model:

$$\mathbf{y}(t) \sim p(\mathbf{y}(t) | \mathbf{x}(t)), \quad (2.7)$$

$$\mathbf{x}(t) \sim p(\mathbf{x}(t) | \mathbf{x}(t-1)), \quad (2.8)$$

where $\mathbf{y}(t)$ is the same as that in Equation (2.5), $p()$ in Equation (2.7) denotes the observation probability density for pilot subcarriers at the t -th symbol, $\mathbf{x}(t) = [x_0(t), \dots, x_i(t), \dots, x_{I-1}(t)]^T$ denotes the I -by-1 general state vector for pilot subcarriers at the t -th symbol, and $p()$ in Equation (2.8) denotes the transition probability density for pilot subcarriers at the t -th symbol.

2.3.2 Linear Filtering

2.3.2.1 Kalman Filter

Algorithm 2.1 shows the recursion process of the Kalman filter at the t -th symbol. In Algorithm 2.1, $\mathbf{a}(t) = E[\boldsymbol{\theta}(t) | \mathbf{y}(1:t-1)]$, $\mathbf{R}(t) = \text{Var}[\boldsymbol{\theta}(t) | \mathbf{y}(1:t-1)]$, $\mathbf{f}(t) = E[\mathbf{y}(t) | \mathbf{y}(1:t-1)]$, $\mathbf{Q}(t) = \text{Var}[\mathbf{y}(t) | \mathbf{y}(1:t-1)]$, $\mathbf{m}(t) = E[\boldsymbol{\theta}(t) | \mathbf{y}(1:t)]$, and $\mathbf{C}(t) = \text{Var}[\boldsymbol{\theta}(t) | \mathbf{y}(1:t)]$. Note that the initial values of $\mathbf{m}(t)$ and $\mathbf{C}(t)$ can be set to any finite value through prior knowledge. The filtered state at the t -th symbol (that is $\boldsymbol{\theta}(t) | \mathbf{y}(1:t)$) has density $\mathcal{CN}(\mathbf{m}(t), \mathbf{C}(t))$.

2.3.2.2 Kalman Smoother

The Kalman smoothing process comprises forward filtering and backward smoothing. The forward filtering uses the forward recursion process described in Algorithm 2.1. Algorithm 2.2 shows a backward recursion process at the t -th symbol, which uses the popular RTS algorithm [2.16]. The RTS algorithm implicitly assumes that the forward filtering has been swept in advance, so $\mathbf{m}(t)$, $\mathbf{C}(t)$, $\mathbf{a}(t)$, and $\mathbf{R}(t)$ are assumed to be already calculated and stored via the advanced Kalman filter. In Algorithm 2.2, $\mathbf{u}(t) =$

CHAPTER 2. PRELIMINARY BACKGROUND

Algorithm 2.1 Kalman filter

0. Filtered state at the $t-1$ -th symbol: $\mathbf{m}(t-1), \mathbf{C}(t-1)$
 1. Update process at the t -th symbol
 - *One-step-ahead prediction of state*
 (Mean) $\mathbf{a}(t) \leftarrow \mathbf{G}(t)\mathbf{m}(t-1)$
 (Covariance) $\mathbf{R}(t) \leftarrow \mathbf{G}(t)\mathbf{C}(t-1)\mathbf{G}(t)^H + \mathbf{W}$
 - *One-step-ahead prediction of observation / Likelihood*
 (Mean) $\mathbf{f}(t) \leftarrow \mathbf{F}(t)\mathbf{a}(t)$
 (Covariance) $\mathbf{Q}(t) \leftarrow \mathbf{F}(t)\mathbf{R}(t)\mathbf{F}(t)^H + \mathbf{V}$
 - *Kalman gain*
 $\mathbf{K}(t) \leftarrow \mathbf{R}(t)\mathbf{F}(t)^H\mathbf{Q}(t)^{-1}$
 - *Filtered state*
 (Mean) $\mathbf{m}(t) \leftarrow \mathbf{a}(t) + \mathbf{K}(t) [\mathbf{y}(t) - \mathbf{f}(t)]$
 (Covariance) $\mathbf{C}(t) \leftarrow [\mathbf{E}_I - \mathbf{K}(t)\mathbf{F}(t)] \mathbf{R}(t)$
 2. Filtered state at the t -th symbol: $\mathbf{m}(t), \mathbf{C}(t)$
-

Algorithm 2.2 Kalman smoother

0. Smoothed state at the $t+1$ -th symbol: $\mathbf{u}(t+1), \mathbf{U}(t+1)$
 1. Update process at the t -th symbol
 - *Smoothing gain*
 $\mathbf{A}(t) \leftarrow \mathbf{C}(t)\mathbf{G}(t+1)^H\mathbf{R}(t+1)^{-1}$
 - *Smoothed state*
 (Mean) $\mathbf{u}(t) \leftarrow \mathbf{m}(t) + \mathbf{A}(t) [\mathbf{u}(t+1) - \mathbf{a}(t+1)]$
 (Covariance) $\mathbf{U}(t) \leftarrow \mathbf{C}(t) + \mathbf{A}(t) [\mathbf{U}(t+1) - \mathbf{R}(t+1)] \mathbf{A}(t)^H$
 2. Smoothed state at the t -th symbol: $\mathbf{u}(t), \mathbf{U}(t)$
-

$E[\boldsymbol{\theta}(t) \mid \mathbf{y}(1:t+L)]$ and $\mathbf{U}(t) = \text{Var}[\boldsymbol{\theta}(t) \mid \mathbf{y}(1:t+L)]$, where L is the fixed time lag value. Note that the initial values of $\mathbf{u}(t)$ and $\mathbf{U}(t)$ are set to $\mathbf{m}(t)$ and $\mathbf{C}(t)$, respectively. The smoothed state at the t -th symbol (that is $\boldsymbol{\theta}(t) \mid \mathbf{y}(1:t+L)$) has density $\mathcal{CN}(\mathbf{u}(t), \mathbf{U}(t))$.

2.3.3 Nonlinear Filtering

The particle filter discretely approximates the target distribution by N particles corresponding to combined realizations and weights (Figure 2.2). The notation $\left\{ \cdot^{(n)} \right\}_{n=1}^N$ hereafter defines a set of indexed particles $\cdot^{(n)}$. For ideal performance, particle filters require an infinite number of particles. To enhance the performance with finitely many particles, researchers have developed the auxiliary particle filter (APF) [2.17] and the Rao–Blackwellised particle filter (RB-PF) [2.18, 2.19].

2.3. STOCHASTIC FILTERING

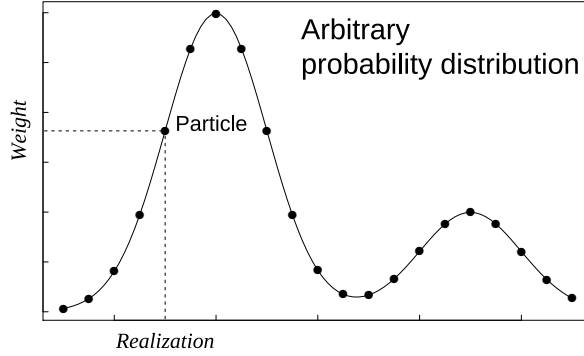


Figure 2.2: Concept of particle filter.

2.3.3.1 Basic Particle Filter

Algorithm 2.3 presents the recursion process for the basic particle filter at the t -th symbol, where the target distribution is the filtered distribution $p(\mathbf{x}(t) \mid \mathbf{y}(1:t))$.

2.3.3.2 Auxiliary Particle Filter

APF first resamples based on the one-step-ahead best guess of the state. The pre-selected state from the prior distribution likely yields data consistent results and can enhance the finite-particle performance of the filter. Algorithm 2.4 presents the recursion process of APF at the t -th symbol, where the target distribution is the filtered distribution $p(\mathbf{x}(t) \mid \mathbf{y}(1:t))$.

2.3.3.3 Rao–Blackwellised Particle Filter

When the general state $\mathbf{x}(t)$ can be divided into the linear and Gaussian part $\boldsymbol{\theta}(t)$ and nonlinear and non-Gaussian part $\boldsymbol{\psi}(t)$, RB-PF hybridizes the Kalman and particle filters and enhances the finite-particle performance by applying an analytical Kalman filter to the linear and Gaussian state $\boldsymbol{\theta}(t)$. The target filtered distribution is decomposed as $p(\mathbf{x}(t) \mid \mathbf{y}(1:t)) = p(\boldsymbol{\psi}(t), \boldsymbol{\theta}(t) \mid \mathbf{y}(1:t)) = p(\boldsymbol{\psi}(t) \mid \mathbf{y}(1:t)) \cdot p(\boldsymbol{\theta}(t) \mid \mathbf{y}(1:t), \boldsymbol{\psi}(t))$, and the particle and Kalman filters are applied to $p(\boldsymbol{\psi}(t) \mid \mathbf{y}(1:t))$ and $p(\boldsymbol{\theta}(t) \mid \mathbf{y}(1:t), \boldsymbol{\psi}(t)) = \mathcal{CN}(\mathbf{m}(t), \mathbf{C}(t))$, respectively. This scheme leads to $\boldsymbol{\theta}(t)^{(n)} = \{\mathbf{m}(t)^{(n)}, \mathbf{C}(t)^{(n)}\}$.

Algorithm 2.3 Basic particle filter

0. Filtered state at the $t-1$ -th symbol:

$$\left\{ \text{Realization } \mathbf{x}(t-1)^{(n)}, \text{Weight } \omega(t-1)^{(n)} \right\}_{n=1}^N$$

1. Update process at the t -th symbol

• **for** $n = 1$ **to** N **do**

a. *Realization*

Draw $\mathbf{x}(t)^{(n)}$ from

proposal distribution $p(\mathbf{x}(t) | \mathbf{x}(t-1)^{(n)}, \mathbf{y}(t))$.

b. *Weight*

$$\omega(t)^{(n)} \leftarrow \omega(t-1)^{(n)} \cdot$$

$$\frac{p(\mathbf{x}(t)^{(n)} | \mathbf{x}(t-1)^{(n)}) p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \mathbf{x}(t)^{(n)})}{p(\mathbf{x}(t)^{(n)} | \mathbf{x}(t-1)^{(n)}, \mathbf{y}(t))}.$$

end for

• *Normalization of the weights:* $\omega(t)^{(n)} \leftarrow \omega(t)^{(n)} / \sum_{n=1}^N \omega(t)^{(n)}$.

• *Resampling*

Draw resampling index k_n from a set $\{1, \dots, N\}$ with $P(k_n = n) \propto \omega(t)^{(n)}$.

Reselect $\mathbf{x}(t)^{(n)}$ along with the k_n and reset $\omega(t)^{(n)}$ to $1/N$.

2. Filtered state at the t -th symbol:

$$\left\{ \text{Realization } \mathbf{x}(t)^{(n)}, \text{Weight } \omega(t)^{(n)} \right\}_{n=1}^N$$

2.3. STOCHASTIC FILTERING

Algorithm 2.4 Auxiliary particle filter

0. Filtered state at the $t-1$ -th symbol:

$$\left\{ \text{Realization } \mathbf{x}(t-1)^{(n)}, \text{Weight } \omega(t-1)^{(n)} \right\}_{n=1}^N$$

1. Update process at the t -th symbol

- *Resampling*

Draw auxiliary index k_n from a set $\{1, \dots, N\}$ with $P(k_n = n) \propto \omega(t-1)^{(n)} p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\mathbf{x}}(t)^{(n)})$, where the one-step-ahead best guess is

$$- \hat{\mathbf{x}}(t)^{(n)} = E[\mathbf{x}(t) | \mathbf{x}(t-1)^{(n)}].$$

- **for** $n = 1$ **to** N **do**

- a. *Realization*

Draw $\mathbf{x}(t)^{(n)}$ from

transition distribution $p(\mathbf{x}(t) | \mathbf{x}(t-1)^{(k_n)})$.

- b. *Weight*

$$\omega(t)^{(n)} \leftarrow \frac{p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \mathbf{x}(t)^{(n)})}{p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\mathbf{x}}(t)^{(k_n)})}.$$

end for

- *Normalization of the weights:* $\omega(t)^{(n)} \leftarrow \omega(t)^{(n)} / \sum_{n=1}^N \omega(t)^{(n)}$.

2. Filtered state at the t -th symbol:

$$\left\{ \text{Realization } \mathbf{x}(t)^{(n)}, \text{Weight } \omega(t)^{(n)} \right\}_{n=1}^N$$

CHAPTER 2. PRELIMINARY BACKGROUND

Bibliography

- [2.1] D. Larsson, "Analysis of channel estimation methods for OFDMA," Master's thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, Dec. 2006.
- [2.2] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Communications*, vol. 3, no. 1, pp. 315–325, Jan. 2004.
- [2.3] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Communications Systems*, vol. 11, no. 4, pp. 360–393, Dec. 1963.
- [2.4] Y. Karasawa, *Radiowave Propagation Fundamentals for Digital Mobile Communications*, 2nd ed. Corona Publishing, 2016, [in Japanese].
- [2.5] H. Iwai, *Radio Propagation in Mobile Communications — Fundamental Knowledge for Simulations Analysis of Wireless Communications —*. Corona Publishing, 2012, [in Japanese].
- [2.6] W. C. Jakes, Jr., *Microwave Mobile Communications*. Wiley, New York, 1974, [reprinted by IEEE Press].
- [2.7] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [2.8] W. H. He and Y. Lee, "Low-complexity Kalman channel estimator structures for OFDM systems with and without virtual carriers," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 4, Jun. 2004, pp. 2447–2451.
- [2.9] W. Chen and R. Zhang, "Estimation of time and frequency selective channels in OFDM systems: a Kalman filter structure," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 2, Nov. 2004, pp. 800–803.
- [2.10] M. Liyanage and I. Sasase, "Steady-state Kalman filtering for channel estimation in OFDM systems for Rayleigh fading channels," *IEICE Trans. Communications*, vol. E92-B, no. 7, pp. 2452–2460, 2009.
- [2.11] M. R. Raghavendra and K. Giridhar, "Improving channel estimation in OFDM systems for sparse multipath channels," *IEEE Signal Processing Letters*, vol. 12, no. 1, pp. 52–55, Jan. 2005.
- [2.12] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Apr. 1991.

BIBLIOGRAPHY

- [2.13] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering, Transactions of the ASME*, vol. 82 (Series D), no. 1, pp. 35–45, Mar. 1960.
- [2.14] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2, pp. 107–113, Apr. 1993.
- [2.15] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, Mar. 1996.
- [2.16] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, Aug. 1965.
- [2.17] M. K. Pitt and N. Shephard, "Filtering via simulation: auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, Jun. 1999.
- [2.18] R. Chen and J. S. Liu, "Mixture Kalman filters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 3, pp. 493–508, 2000.
- [2.19] A. Doucet, N. de Freitas, K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks," in *Proc. AUAI 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, Stanford, U.S.A., Jul. 2000, pp. 176–183.

Chapter 3

State-Space Modeling

3.1 Preliminary Remarks

This chapter proposes a novel state-space modeling scheme. The scheme comprises two methods. On the basis of a state-space model, the first method (hereafter “Method 1”) appropriately considers the frequency correlation between pilot signals. The second method (hereafter “Method 2”) is an approximation method based on the first and reduces computational complexity in order to gain pragmatic channel estimation.

The rest of this chapter is structured as follows. Section 3.2 describes the problem formulation. Sections 3.3 and 3.4 describe Methods 1 and 2, respectively. Section 3.5 presents numerical analysis results and verifies the proposed scheme. Section 3.6 summarizes the discussion.

3.2 Problem Formulation

As previously described in Section 2.2, conventional methods based on CFR approach do not sufficiently consider frequency correlation. For example, [3.1] does not consider frequency correlation itself. Although [3.2] considers frequency correlation together with time correlation in the state equation (described later), this approach is just an approximation; the WS-SUS condition means that the time correlation is independent of the frequency correlation, yet the formulation of [3.2] implies that each temporal variation is related with the other along the subcarrier direction, consequently making the time correlation dependent on the frequency correlation. For consistency with the WSSUS condition, instantaneous channel gain must have the frequency correlation. Reference [3.2] also proposes considering frequency correlation during the post-processing phase after the Kalman filter has been applied to each subcarrier, but such an external coupling approach cannot achieve optimal estimation. To overcome these concerns, we propose a novel scheme that adequately considers frequency

CHAPTER 3. STATE-SPACE MODELING

correlation based on a state-space model and reduces computational complexity.

3.3 Method 1: State-Space Model that Appropriately Considers Frequency Correlation

3.3.1 Formulation Using the State-Space Model

The formulation of the state-space model consists of the observation equation, the state equation, and their associated equations.

3.3.1.1 The Observation Equation

The observation equation for the state-space model discussed in this thesis corresponds to Equation (2.4). $\tilde{h}_i(t)$ has a complex Gaussian density with mean 0, according to the central limit theorem (sum of many multipath element waves) and based on the assumption of the Jakes model. It is also correlated to other elements due to the frequency correlation. Considering these characteristics, $\tilde{\mathbf{h}}(t)$ can be expressed in a way that explicitly explains the correlation between each element in the equation, based on a common method for adding the correlation to independent multivariate Gaussian variables [3.3]:

$$\tilde{\mathbf{h}}(t) = \mathbf{\Sigma} \boldsymbol{\theta}(t), \quad (3.1)$$

where $\mathbf{\Sigma}$ is an I -by- I matrix, as shown in Equation (3.2).

$\mathbf{\Sigma} =$

$$\begin{bmatrix} 1 & \rho(\Delta t=0, \Delta f=1df) & \rho(\Delta t=0, \Delta f=2df) & \cdots & \rho(\Delta t=0, \Delta f=(I-2)df) & \rho(\Delta t=0, \Delta f=(I-1)df) \\ \rho^*(\Delta t=0, \Delta f=1df) & 1 & \rho(\Delta t=0, \Delta f=1df) & \cdots & \rho(\Delta t=0, \Delta f=(I-3)df) & \rho(\Delta t=0, \Delta f=(I-2)df) \\ \rho^*(\Delta t=0, \Delta f=2df) & \rho^*(\Delta t=0, \Delta f=1df) & 1 & \cdots & \rho(\Delta t=0, \Delta f=(I-4)df) & \rho(\Delta t=0, \Delta f=(I-3)df) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^*(\Delta t=0, \Delta f=(I-2)df) & \rho^*(\Delta t=0, \Delta f=(I-3)df) & \rho^*(\Delta t=0, \Delta f=(I-4)df) & \cdots & 1 & \rho(\Delta t=0, \Delta f=1df) \\ \rho^*(\Delta t=0, \Delta f=(I-1)df) & \rho^*(\Delta t=0, \Delta f=(I-2)df) & \rho^*(\Delta t=0, \Delta f=(I-3)df) & \cdots & \rho^*(\Delta t=0, \Delta f=1df) & 1 \end{bmatrix}^{1/2}. \quad (3.2)$$

$\mathbf{\Sigma}$ results in the square root matrix of a correlation matrix that expresses the frequency correlation between pilot subcarriers, and $\boldsymbol{\theta}(t)$ results in a complex Gaussian vector whose element is independent of other elements. Thus, Equation (2.4) is accordingly rewritten as

$$\mathbf{y}(t) = \mathbf{S}(t)\mathbf{\Sigma}\boldsymbol{\theta}(t) + \mathbf{v}(t) = \mathbf{F}(t)\boldsymbol{\theta}(t) + \mathbf{v}(t), \quad (3.3)$$

where the observation matrix is set as $\mathbf{F}(t) = \mathbf{S}(t)\mathbf{\Sigma}$. $\rho(\Delta t = 0, \Delta f)$ denotes the frequency correlation coefficient of the narrow-band channel gain, Δt denotes the time difference, and Δf denotes the frequency difference. $\rho(\Delta t = 0, \Delta f)$ should be a complex value in general [3.4]. Its real part denotes either the correlation between in-phase components or

3.3. METHOD 1: STATE-SPACE MODEL THAT APPROPRIATELY CONSIDERS FREQUENCY CORRELATION

that between quadrature-phase components. Its imaginary part denotes the correlation between an in-phase and quadrature-phase component (or between a quadrature-phase and in-phase component). The assumption of an exponential delay profile in this thesis leads us from the frequency correlation coefficient to the following equation [3.5]:

$$\rho(\Delta t = 0, \Delta f) = \frac{1 + j2\pi\Delta f\sigma_\tau}{1 + (2\pi\Delta f\sigma_\tau)^2}, \quad (3.4)$$

where σ_τ denotes the channel delay spread.

3.3.1.2 The State Equation

The temporal dynamics of state vector $\theta(t)$ are often set to the autoregressive (AR) model [3.1, 3.2]. However, [3.6] shows that the optimum dynamics obey the zero-th order polynomial under optimum-weighted recursive least-squares (RLS) channel estimation. The Kalman filter is considered equivalent to the optimum-weighted RLS, so this thesis adopts the random walk model [3.7]. The transition matrix is consequently set to $\mathbf{G}(t) = \mathbf{E}_I$. Accordingly, the state equation is as follows:

$$\theta(t) = \theta(t-1) + \mathbf{w}(t), \quad (3.5)$$

where a covariance matrix of the state noise vector is $\mathbf{W} = 2l^2(1 - \rho(\Delta t = 1dt, \Delta f = 0))\mathbf{E}_I$. l^2 denotes the average power of the narrow-band channel gain for each pilot subcarrier, and $\rho(\Delta t, \Delta f = 0)$ denotes the time correlation coefficient of the narrow-band channel gain. See Appendix A for a detailed derivation of $\mathbf{W}$. $\rho(\Delta t, \Delta f = 0)$ should be a complex value in general. Because the Jakes model is assumed in this thesis, the time correlation coefficient is replaced with the following real Bessel functions of the first kind of order zero [3.5]:

$$\rho(\Delta t, \Delta f = 0) = J_0(2\pi f_D \Delta t), \quad (3.6)$$

where f_D denotes the channel maximum Doppler frequency.

3.3.2 Specification of Hyperparameters

It is difficult to treat hyperparameters $\mathbf{V}$, Σ , and $\mathbf{W}$ as they are known in real environments. These parameters, however, must be specified before channel estimation. In the proposed method outlined in this chapter, such parameter specification is achieved via a short training period before channel estimation, assuming these hyperparameters do not change. In Section 2.1, a frequency/time selectivity model related to Σ or $\mathbf{W}$ is assumed; this assumption can reduce the number of hyperparameters to be specified at the same time. Thus, the maximum likelihood (ML) method [3.7] for the

CHAPTER 3. STATE-SPACE MODELING

state-space model can be practically applied to the training process. This ML method implies that the Kalman filter with different hyperparameters is applied multiple times during the training period. The log-likelihood for the whole training period can be written as

$$\begin{aligned} \ell(\mathbf{V}, \mathbf{\Sigma}, \mathbf{W}) = & -\frac{1}{2} \sum_t \log \det(\mathbf{Q}(t)) \\ & -\frac{1}{2} \sum_t \left(\mathbf{y}(t) - \mathbf{f}(t) \right)^H \mathbf{Q}(t)^{-1} \left(\mathbf{y}(t) - \mathbf{f}(t) \right), \end{aligned} \quad (3.7)$$

where $\mathbf{Q}(t)$ is a covariance matrix for one-step-ahead prediction [3.7] of observation vector $\mathbf{y}(t)$, and $\mathbf{f}(t)$ is the mean vector for one-step-ahead prediction of observation vector $\mathbf{y}(t)$. $\mathbf{Q}(t)$ and $\mathbf{f}(t)$ implicitly depend on $\mathbf{V}$, $\mathbf{\Sigma}$, and $\mathbf{W}$. The explicit definition of $\mathbf{Q}(t)$ and $\mathbf{f}(t)$ in Algorithm 2.1 makes it clear that log-likelihood is described in closed forms of the hyperparameters. Equation (3.7) is numerically maximized to obtain the ML estimator of hyperparameters as $(\hat{\mathbf{V}}, \hat{\mathbf{\Sigma}}, \hat{\mathbf{W}})_{\text{ML}} = \arg\max_{\mathbf{V}, \mathbf{\Sigma}, \mathbf{W}} \ell(\mathbf{V}, \mathbf{\Sigma}, \mathbf{W})$. The feature of each hyperparameter is briefly reconfirmed as follows.

$\mathbf{V}$ is related to E_b/N_0 . This value is formulated through σ^2 (the average noise power of each pilot subcarrier) in this thesis, but may sometimes be treated as known via other possible alternative methods [3.1].

$\mathbf{\Sigma}$ is related to the frequency correlation between pilot subcarriers. This value depends on wideband channel impulse response $\mathbf{h}(t)$, but can be formulated through channel delay spread σ_τ , assuming an exponential delay profile (see also Equation (3.4)).

$\mathbf{W}$ is related to the time correlation between each pilot subcarrier. This value can be formulated through the channel maximum Doppler frequency f_D , assuming the Jakes model as in this thesis (see also Equation (3.6)). The ML method discussed in this chapter specifies whole $2l^2(1 - \rho(\Delta t = 1dt, \Delta f = 0))$ in $\mathbf{W}$ as one hyperparameter. Simple alternative methods that use either forgetting or discount factors [3.8] have also been proposed, but such methods impair estimation accuracy.

3.4 Method 2: Approximation Approach to Reduce Computational Complexity

The scope of this thesis is to enhance OFDM channel estimation accuracy in real time on the receiving side. Method 1, described in Section 3.3, achieves this purpose. Computational complexity, however, is often a critical issue in actual real-time processing; thus, this section discusses an approach that reduces Method 1's computational complexity. For the purposes of this thesis, "computational complexity" refers to the number of multiplication and division processes occurring during sequential processing when a pilot subcarrier is newly received.

3.4. METHOD 2: APPROXIMATION APPROACH TO REDUCE COMPUTATIONAL COMPLEXITY

3.4.1 Estimation of Computational Complexity

The Kalman filter provides optimal estimators for the linear Gaussian state-space model described in Section 3.3, so a method that reduces Method 1's computational complexity will also do the same for the Kalman filter. Various techniques to reduce the Kalman filter's computational complexity — such as the fast Kalman filter [3.9] or the steady-state Kalman filter [3.7] — have been proposed. These methods, however, have their own restrictions: the fast Kalman filter cannot be applied to the formulation in this thesis because its observation matrix must be a Toeplitz matrix, and the steady-state Kalman filter cannot be extended to a situation where a hyperparameter is time-variant. Therefore, a novel method is proposed. Before proceeding further with detailed discussion, the baseline for computational complexity should be clarified. The serial processing of multivariates [3.10] is already known to be quite effective at reducing the computational complexity of the multivariate Kalman filter, which is applied to the proposed formulation (see Section 3.3), so this approach is set as the starting point for Method 2. Tables 3.1 and 3.2 give the computational complexity for pure matrix operations (see Appendix B) and for the serial processing of multivariates, respectively. The tables also describe the vector/matrix size set over the variables as well as each step in the sequential process at t . Compared with Table 3.1, the complexity of Table 3.2 can be reduced when $I \geq 1$, according to $(5I^3 + 2I^2) > (2I^3 + 4I^2)$.

CHAPTER 3. STATE-SPACE MODELING

Table 3.1: Computational complexity for pure matrix operations (complexity of the inverse matrix is assumed to be I^3)

Each step in sequential process at t	Complexity
1. One-step-ahead prediction of state	
(Mean) $\overset{(I,1)}{\mathbf{a}(t)} \leftarrow \overset{(I,1)}{\mathbf{G}(t)}\overset{(I,1)}{\mathbf{m}(t-1)} = \overset{(I,1)}{\mathbf{m}(t-1)}$	0
(Variance) $\overset{(I,I)}{\mathbf{R}(t)} \leftarrow \overset{(I,I)}{\mathbf{G}(t)}\overset{(I,I)}{\mathbf{C}(t-1)}\overset{(I,I)}{\mathbf{G}(t)}^H + \overset{(I,I)}{\mathbf{W}}$ $\quad \quad \quad = \overset{(I,I)}{\mathbf{C}(t-1)} + \overset{(I,I)}{\mathbf{W}}$	0
2. One-step-ahead prediction of observation	
(Variance) $\overset{(I,I)}{\mathbf{q}(t)} \leftarrow \overset{(I,I)}{\mathbf{R}(t)}\overset{(I,I)}{\mathbf{F}(t)}^H$	I^3
$\overset{(I,I)}{\mathbf{Q}(t)} \leftarrow \overset{(I,I)}{\mathbf{F}(t)}\overset{(I,I)}{\mathbf{q}(t)} + \overset{(I,I)}{\mathbf{V}}$	I^3
3. Kalman gain	
$\overset{(I,I)}{\mathbf{K}(t)} \leftarrow \overset{(I,I)}{\mathbf{q}(t)}\overset{(I,I)}{\mathbf{Q}(t)}^{-1}$	$I^3 + I^3$
4. Filtered state	
(Mean) $\overset{(I,1)}{\mathbf{m}(t)} \leftarrow \overset{(I,1)}{\mathbf{a}(t)} + \overset{(I,I)}{\mathbf{K}(t)}[\overset{(I,1)}{\mathbf{y}(t)} - \overset{(I,I)}{\mathbf{F}(t)}\overset{(I,1)}{\mathbf{a}(t)}]$	$I^2 + I^2$
(Variance) $\overset{(I,I)}{\mathbf{C}(t)} \leftarrow \overset{(I,I)}{\mathbf{R}(t)} - \overset{(I,I)}{\mathbf{K}(t)}\overset{(I,I)}{\mathbf{q}(t)}^H$	I^3
5. Go to Step 1 at $t+1$	0
Sum: $5I^3 + 2I^2$	

3.4. METHOD 2: APPROXIMATION APPROACH TO REDUCE COMPUTATIONAL COMPLEXITY

Table 3.2: Computational complexity for the serial processing of multivariates ($\mathbf{F}_i(t)$ denotes the i -th row of $\mathbf{F}(t)$)

Each step in sequential process at t	Complexity
1. One-step-ahead prediction of state	
(Mean) $\mathbf{a}(t) \leftarrow \mathbf{G}(t)\mathbf{m}(t-1) = \mathbf{m}(t-1)$	0
(Variance) $\mathbf{R}(t) \leftarrow \mathbf{G}(t)\mathbf{C}(t-1)\mathbf{G}(t)^H + \mathbf{W}$	0
$\mathbf{R}(t) = \mathbf{C}(t-1) + \mathbf{W}$	0
for $i \leftarrow 0$ to $I-1$ do	
2. One-step-ahead prediction of observation	
(Variance) $\mathbf{q}(t) \leftarrow \mathbf{R}(t)\mathbf{F}_i(t)^H$	I^2
$\mathbf{Q}(t) \leftarrow \mathbf{F}_i(t)\mathbf{q}(t) + V_i$	I
3. Kalman gain	
$\mathbf{K}(t) \leftarrow \mathbf{q}(t)\mathbf{Q}(t)^{-1}$	I
4. Filtered state	
(Mean) $\mathbf{a}(t) \leftarrow \mathbf{a}(t) + \mathbf{K}(t)[y_i(t) - \mathbf{F}_i(t)\mathbf{a}(t)]$	$I+I$
(Variance) $\mathbf{R}(t) \leftarrow \mathbf{R}(t) - \mathbf{K}(t)\mathbf{q}(t)^H$	I^2
end for	
5. Conclusion of the loop for $i \Rightarrow$ Go to Step 1 at $t+1$	
(Mean) $\mathbf{m}(t) \leftarrow \mathbf{a}(t)$	0
(Variance) $\mathbf{C}(t) \leftarrow \mathbf{R}(t)$	0
Sum: $2I^3 + 4I^2$	

Method 2 forces observation matrix $\mathbf{F}(t)$ into a sparse upper/lower bidiagonal matrix in order to decrease the number of multiplication and division processes. Before details regarding Method 2's validity and implementation are described (in the next subsection), its computational complexity is compared with that of the conventional method. Table 3.3 shows the computational complexity of Method 2 based on the serial processing of multivariates. Upper/lower bidiagonalized observation matrix $\mathbf{F}(t)$ with this proposal can reduce the number of non-zero elements in any row to at most two; thus, the computational complexity regarding $\mathbf{F}_i(t)$ decreases. Compared with Table 3.2, the complexity of Table 3.3 can be reduced when $I \geq 2$, according to $(2I^3 + 4I^2) > (I^3 + 4I^2 + 4I)$.

CHAPTER 3. STATE-SPACE MODELING

Table 3.3: Computational complexity of Method 2

Each step in sequential process at t	Complexity
1. One-step-ahead prediction of state	
(Mean) $\overset{(I,1)}{\mathbf{a}(t)} \leftarrow \overset{(I,1)}{\mathbf{G}(t)}\overset{(I,1)}{\mathbf{m}(t-1)} = \overset{(I,1)}{\mathbf{m}(t-1)}$	0
(Variance) $\overset{(I,I)}{\mathbf{R}(t)} \leftarrow \overset{(I,I)}{\mathbf{G}(t)}\overset{(I,I)}{\mathbf{C}(t-1)}\overset{(I,I)}{\mathbf{G}(t)}^H + \overset{(I,I)}{\mathbf{W}}$ $\quad \quad \quad = \overset{(I,I)}{\mathbf{C}(t-1)} + \overset{(I,I)}{\mathbf{W}}$	0
for $i \leftarrow 0$ to $I-1$ do	
2. One-step-ahead prediction of observation	
(Variance) $\overset{(I,1)}{\mathbf{q}(t)} \leftarrow \overset{(I,I)}{\mathbf{R}(t)}\overset{(I,1)}{\mathbf{F}_i(t)}^H$	$2I$
$\overset{(1,1)}{Q(t)} \leftarrow \overset{(1,I)}{\mathbf{F}_i(t)}\overset{(I,1)}{\mathbf{q}(t)} + \overset{(1,1)}{V_i}$	2
3. Kalman gain	
$\overset{(I,1)}{\mathbf{K}(t)} \leftarrow \overset{(I,1)}{\mathbf{q}(t)}\overset{(1,1)}{Q(t)}^{-1}$	I
4. Filtered state	
(Mean) $\overset{(I,1)}{\mathbf{a}(t)} \leftarrow \overset{(I,1)}{\mathbf{a}(t)} + \overset{(I,1)}{\mathbf{K}(t)}[\overset{(1,1)}{y_i(t)} - \overset{(1,I)}{\mathbf{F}_i(t)}\overset{(I,1)}{\mathbf{a}(t)}]$	$I+2$
(Variance) $\overset{(I,I)}{\mathbf{R}(t)} \leftarrow \overset{(I,I)}{\mathbf{R}(t)} - \overset{(I,1)}{\mathbf{K}(t)}\overset{(1,I)}{\mathbf{q}(t)}^H$	I^2
end for	
5. Conclusion of the loop for $i \Rightarrow$ Go to Step 1 at $t+1$	
(Mean) $\overset{(I,1)}{\mathbf{m}(t)} \leftarrow \overset{(I,1)}{\mathbf{a}(t)}$	0
(Variance) $\overset{(I,I)}{\mathbf{C}(t)} \leftarrow \overset{(I,I)}{\mathbf{R}(t)}$	0
Sum: I^3+4I^2+4I	

3.4. METHOD 2: APPROXIMATION APPROACH TO REDUCE COMPUTATIONAL COMPLEXITY

3.4.2 Approximating the Observation Matrix into a Sparse Upper/Lower Bidiagonal Matrix

Approximating observation matrix $\mathbf{F}(t) = \mathbf{S}(t)\mathbf{\Sigma}$ into a bidiagonal matrix corresponds to approximating $\mathbf{\Sigma}$ into a bidiagonal matrix because the transmitted symbol matrix $\mathbf{S}(t)$ is diagonal. $\mathbf{\Sigma}$ is a square root matrix of the pilot subcarrier frequency correlation matrix. Forcing the original frequency correlation matrix into a sparse matrix [3.11] requires some discussion.

When a pilot subcarrier is scattered in a comb pattern in the time and frequency domains, the pilot interval in the frequency domain should be, as long as possible, taking transmission efficiency into consideration; however, to make interpolation easier, the pilot interval in the frequency domain should not last so long that frequency correlation significantly decreases. Therefore, it is generally supposed that the frequency correlation between adjacent pilot subcarriers is more important than that between beyond-adjacent pilot subcarriers. This consideration implies that the approximation leaves only a frequency correlation between adjacent pilot subcarriers and neglects the correlation between beyond-adjacent pilot subcarriers. Consequently, the frequency correlation matrix can be forced into a tridiagonal matrix, which means that Equation (3.2) can be approximated as

$$\mathbf{\Sigma} \approx \left[\begin{array}{cccc} 1 & g & & \\ g^* & 1 & g & \\ & g^* & 1 & \ddots \\ & & \ddots & \ddots & g \\ & & & g^* & 1 \end{array} \right]^{1/2}, \quad (3.8)$$

where g is a frequency correlation coefficient between adjacent pilot subcarriers.

Deriving $\mathbf{\Sigma}$, the sparseness of such a forced frequency correlation matrix is not always maintained, depending on the process of the square root matrix. Here, Cholesky factorization [3.3] is applied to the matrix square root process in order to retain sparseness. Note that Cholesky factorization may not be applicable to such a forced matrix when it is not positive definite; if the forced correlation matrix is not positive definite, it is further restricted to the nearest positive definite matrix. Several approaches have been proposed for deriving the nearest positive definite matrix [3.12,3.13]. This thesis proposes an analytical approach, making use of the tridiagonal feature; see Appendix C for more details.

In short, forcing a frequency correlation matrix into a tridiagonal matrix is followed by Cholesky factorization in order to achieve a bidiagonal $\mathbf{\Sigma}$ (if the forced correlation matrix is not positive definite, it is first restricted to the nearest positive definite matrix). The observation matrix

CHAPTER 3. STATE-SPACE MODELING

finally becomes upper/lower bidiagonal.

3.5 Numerical Analysis

3.5.1 Effects of Appropriately Considering Frequency Correlation

To verify the benefits of appropriately considering frequency correlation, a computer simulation is performed using 3GPP LTE with a 5 MHz bandwidth. The frequency interval between each subcarrier is 15 kHz, and the symbol length (except for the GI) corresponds to 66.7 μ s. The fast Fourier transform (FFT) size is 512, wherein 300, 1, and 211 are the sizes for the active, direct current (DC), and guard subcarriers, respectively. The guard band is $5 - (300 + 1) \times 0.015 = 0.485$ MHz. Cell-specific reference signals (CRSs) [3.14] are regarded as pilot subcarriers. Figure 3.1 shows the CRS allocation. In the figure, CRSs at timeslots 4, 11, $\dots$ are not used in the simulation, $dt = 0.5$ ms, $df = 90$ kHz, and a specific pseudorandom sequence for the CRSs is modulated by QPSK. There are 250 data subcarriers and 50 pilot subcarriers for one symbol duration. An isolated cell and single-user environment without inter-cell interference and multiple access are assumed. Table 3.4 shows the other assumptions used for the computer simulation.

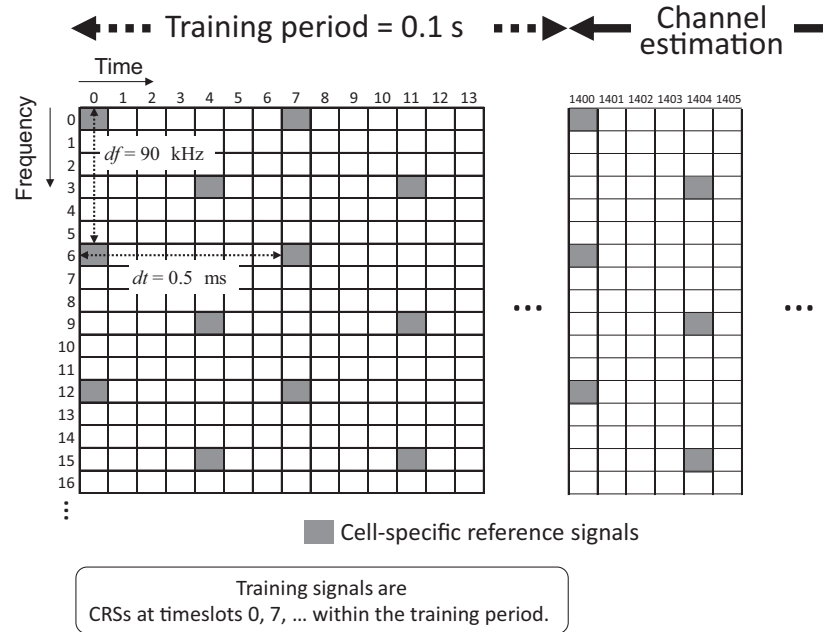


Figure 3.1: Scattered pilot subcarrier allocation.

3.5. NUMERICAL ANALYSIS

Table 3.4: Computer simulation assumptions

Carrier frequency	2 GHz
Maximum Doppler frequency: f_D	14, 140 Hz (i.e., 7.6, 76 km/h)
Delay profile	• 24-path exponential – Delay spread: 1 μs – Maximum delay: 4.7 μs (within GI)
ML specification of hyperparameters	Training period is 0.1 s just before the estimated symbol. The Σ in Method 2 is specified through the truncation for Σ in Method 1.
Numerical computation of Kalman filter	A square root Kalman filter based on singular value decomposition [3.15] is applied to Method 1, Method 2, and conventional Kalman filter for the avoidance of degradation in numerical accuracy.

3.5.1.1 Channel Estimation Example at $E_b/N_0 = 8$ dB and $f_D = 14$ Hz

Table 3.5 shares example results from the hyperparameter specification (f_D is derived with the true average power, l^2 , of the narrow-band channel gain for each pilot subcarrier). Using the ML method for the state-space model, the hyperparameters in Table 3.5 are specified close to each true value, but a certain error is recognized — especially for f_D . Although these results imply that the pure ML method is limited and should ultimately be improved, the specifications are sufficiently accurate and useful for channel estimation [3.16]. Such hyperparameter specifications are used in the following analysis. Figures 3.2 and 3.3 show example channel estimation results for pilot subcarrier index $i = 24$. The X-axes denote the time elapsed since beginning channel estimation (one slot corresponds to one dt), and the Y-axes denote the amplitude of the narrow-band channel gain.

Table 3.5: Example results of hyperparameter specification

Hyperparameter	True value	Specified value
E_b/N_0 [dB]	8	8.5
σ_τ [μ s]	1	1.0
f_D [Hz]	14	42.3

CHAPTER 3. STATE-SPACE MODELING

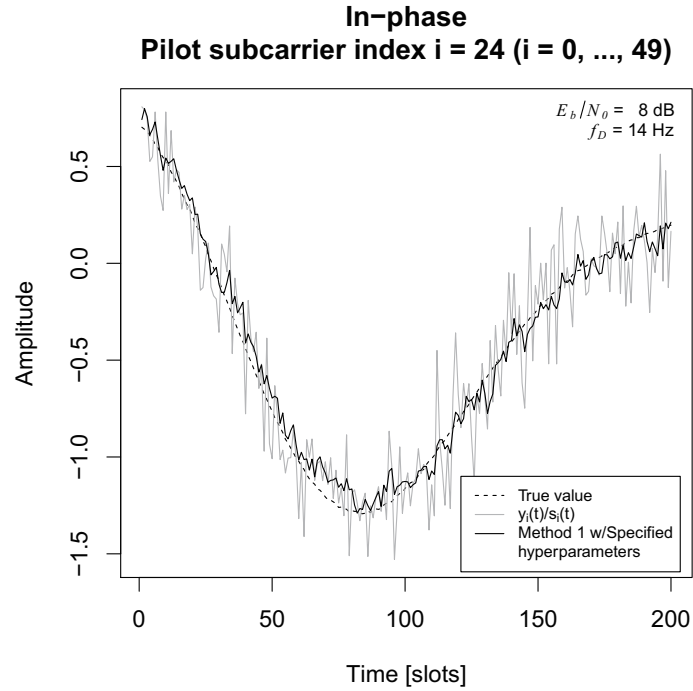


Figure 3.2: Example results of channel estimation (in-phase).

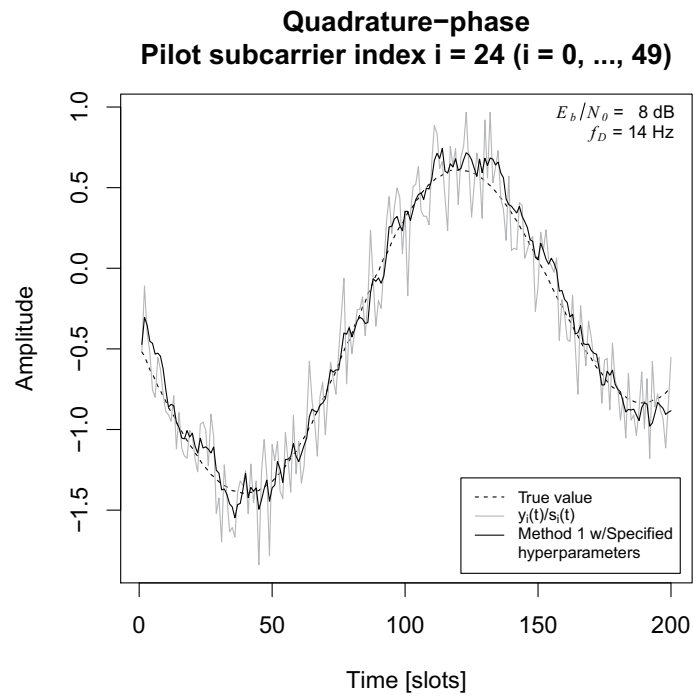


Figure 3.3: Example results of channel estimation (quadrature-phase).

3.5. NUMERICAL ANALYSIS

As shown in Figures 3.2 and 3.3, Method 1 with specified hyperparameters can estimate true values better than the least-squares (LS) estimator $y_i(t)/s_i(t)$.

3.5.1.2 NMSE Performance

Figures 3.4 and 3.5 compare various channel estimation results. The X-axes denote E_b/N_0 and the Y-axes denote the normalized mean squared error (NMSE) = $\sum_{i=0}^{L-1} E[|\tilde{h}_i(t) - \hat{\tilde{h}}_i(t)|^2] / \sum_{i=0}^{L-1} E[|\tilde{h}_i(t)|^2]$. The NMSE performances with both specified and true hyperparameters are plotted in Figures 3.4 and 3.5.

CHAPTER 3. STATE-SPACE MODELING

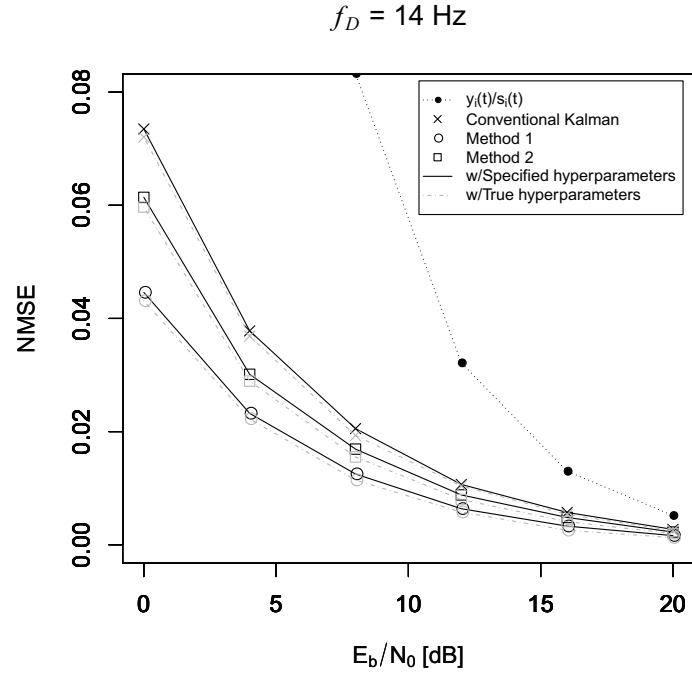


Figure 3.4: Comparison of various channel estimation results ($f_D = 14 \text{ Hz}$).

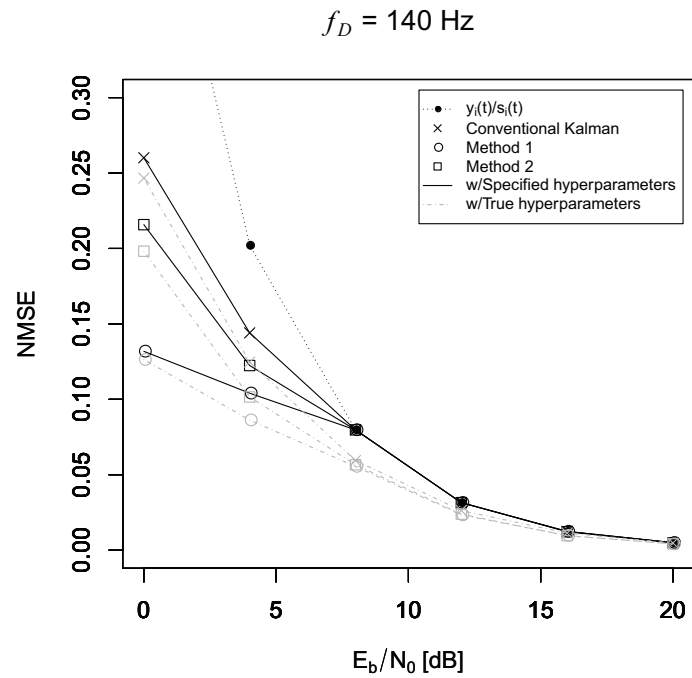


Figure 3.5: Comparison of various channel estimation results ($f_D = 140 \text{ Hz}$).

3.5. NUMERICAL ANALYSIS

Specified hyperparameter NMSE performance When specified hyperparameters are used, as shown in Figures 3.4 and 3.5, the NMSE improves in order of $y_i(t)/s_i(t)$, the conventional Kalman filter (which does not consider frequency correlation), Method 2, and Method 1. Method 2 does not improve the NMSE as much as Method 1 because Method 2 approximates the frequency correlation matrix in order to reduce computational complexity. Consideration of frequency correlation in Methods 1 and 2 effectively improves the NMSE. In particular, the lower E_b/N_0 becomes, the more effective Methods 1 and 2 become at improving the NMSE. For example, at $E_b/N_0 = 0$ dB, Method 1 improves the NMSE by 39% for $f_D = 14$ Hz and 49% for $f_D = 140$ Hz, compared with using the conventional Kalman filter. This improvement occurs because the lower E_b/N_0 becomes, the harder it is to obtain accurate channel estimators, thus making the averaging effect in the frequency domain (which is achieved by considering frequency correlation) more desirable. However, the NMSEs for all methods are almost the same when $f_D = 140$ Hz and $E_b/N_0 \geq 8$ dB. In such a situation, the averaging effect in the time and frequency domains, which is yielded by the Kalman filter, is unattractive for two reasons: a) the higher f_D becomes, the more frequent channel temporal variations become, which means that the wider the relative dt becomes, the fewer the equivalent averaging samples become; and b) the higher E_b/N_0 becomes, the easier it becomes to obtain an accurate channel estimator.

True hyperparameter NMSE performance True hyperparameters (as compared with specified hyperparameters) improve NMSE performance to some extent for all methods — Methods 1 and 2 and the conventional Kalman filter — because the specification error is eliminated. By comparing Figures 3.4 and 3.5, we see that the higher f_D becomes, the greater the improvement in performance becomes. This improvement occurs because hyperparameter specification errors increase with higher f_D values. As previously mentioned, the averaging effect in the time domain, which is yielded by the Kalman filter, becomes less attractive for higher f_D values. Thus, the higher f_D becomes, the harder it is to perform ML specifications of hyperparameters based on the Kalman filter.

3.5.1.3 Bit Error Rate Performance

Figure 3.6 shows example results of the bit error rate (BER) performance for $f_D = 14$ Hz using the data subcarrier assumptions given in Table 3.6. The BER and E_b/N_0 are denoted on the Y-axis and X-axis, respectively. The BER performances with both specified and true hyperparameters are plotted in Figure 3.6. Regarding the lower bound, two types of BER performance are plotted: Lower bounds 0 and 1. These two bounds are differentiated depending on whether the channel value is interpolated for data

CHAPTER 3. STATE-SPACE MODELING

Table 3.6: Data subcarrier assumptions

Modulation	QPSK
Interpolation of channel estimator	• Frequency domain: Linear insertion • Time domain: The nearest neighbor
Equalization	• Zero forcing (ZF) [3.17] for LS channel estimator $y_i(t)/s_i(t)$ • Minimum Mean Squared Error (MMSE) [3.17] for the other channel estimators
Channel coding	• Convolutional codes – Rate: 1/2 – Constraint length: 7
Channel decoding	Soft-decision Viterbi algorithm
Interleaving	• Random interleaver – Unit: Symbol-by-symbol – Depth: 5 Resource blocks (RBs) [3.14], which corresponds to 400 symbols

subcarriers or not. The Lower bound 1 is hereafter used to refer to the comparison with other methods in order to match the the interpolation condition.

3.5. NUMERICAL ANALYSIS

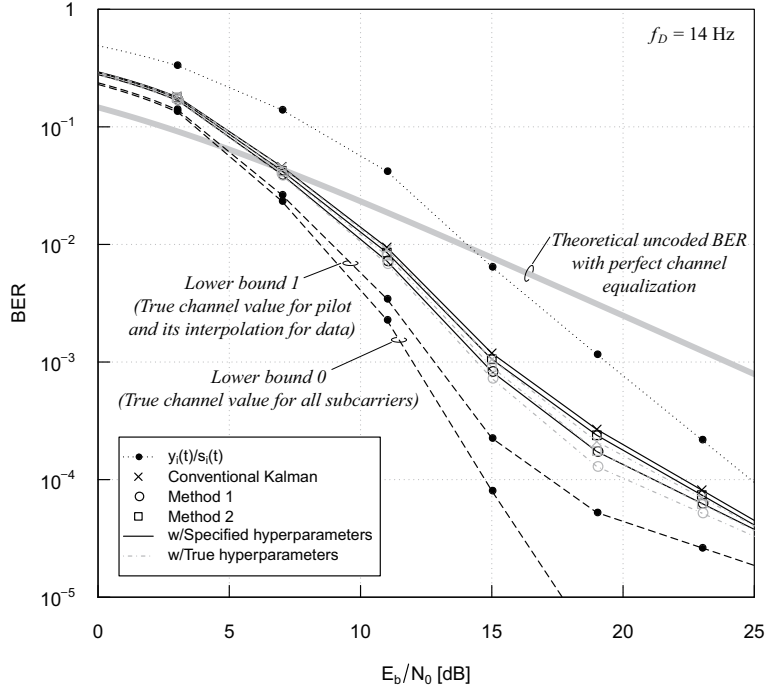


Figure 3.6: Example results of BER performance.

Specified hyperparameter BER performance According to Figure 3.6, the BER performance improves in order of $y_i(t)/s_i(t)$, the conventional Kalman filter, Method 2, and Method 1 — similarly to the NMSE results in Figures 3.4 and 3.5. Table 3.7 gives the E_b/N_0 values that achieve the required BER (10^{-3} and 10^{-4}) for each method with specified hyperparameters. Based on these results, the following conclusions regarding coding gain with specified hyperparameters are drawn.

- Coding gain improvements using Methods 1 and 2 are larger than those for $y_i(t)/s_i(t)$.
- Coding gain improvements using Method 2 are slighter than those for the conventional Kalman filter.
- Coding gain improvements using Method 1 are acceptable compared with those for the conventional Kalman filter.
- The coding gain difference between Method 1 and Lower bound 1 is wide.

CHAPTER 3. STATE-SPACE MODELING

Table 3.7: E_b/N_0 values that achieve required BER (with specified hyperparameters)

Required BER	E_b/N_0				
	$y_i(t)/s_i(t)$	Conventional Kalman filter	Method 2	Method 1	Lower bound 1
10^{-3}	19.36	15.46	15.15	14.65	13.00
10^{-4}	24.99	22.31	21.96	21.16	17.23

True hyperparameter BER performance True hyperparameters (rather than specified ones) improve the BER performance to some extent for all methods — Methods 1 and 2 and the conventional Kalman filter — because the specification error is eliminated. Table 3.8 gives the E_b/N_0 values that achieve the required BER (10^{-3} and 10^{-4}) for each method with true hyperparameters. Views on coding gain with true hyperparameters basically do not change from those with specified hyperparameters.

BER performance comparison with Lower bound 1 We now consider the difference in BER performance between the proposed methods and the Lower bound 1. This difference in BER performance is caused by estimation errors for pilot subcarriers, which are related to hyperparameter specification errors and the averaging effect yielded by the Kalman filter. When specification errors are eliminated, BER performance improves to some extent (Table 3.8). The averaging effect is generally improved by increasing averaging samples; thus, the averaging effect in this thesis is improved by decreasing time interval dt (which is equivalent to maximum Doppler frequency f_D) or frequency interval df (which is equivalent to channel delay spread σ_τ). Such improvements in the averaging effect in the time domain are confirmed through the fact that NMSE performance with true hyperparameters in Figure 3.4 is lower than that in Figure 3.5. Note that these improvements in the time domain are applicable to both

Table 3.8: E_b/N_0 values that achieve required BER (with true hyperparameters)

Required BER	E_b/N_0				
	$y_i(t)/s_i(t)$	Conventional Kalman filter	Method 2	Method 1	Lower bound 1
10^{-3}	19.36	15.12	14.82	14.43	13.00
10^{-4}	24.99	21.71	21.15	20.12	17.23

3.5. NUMERICAL ANALYSIS

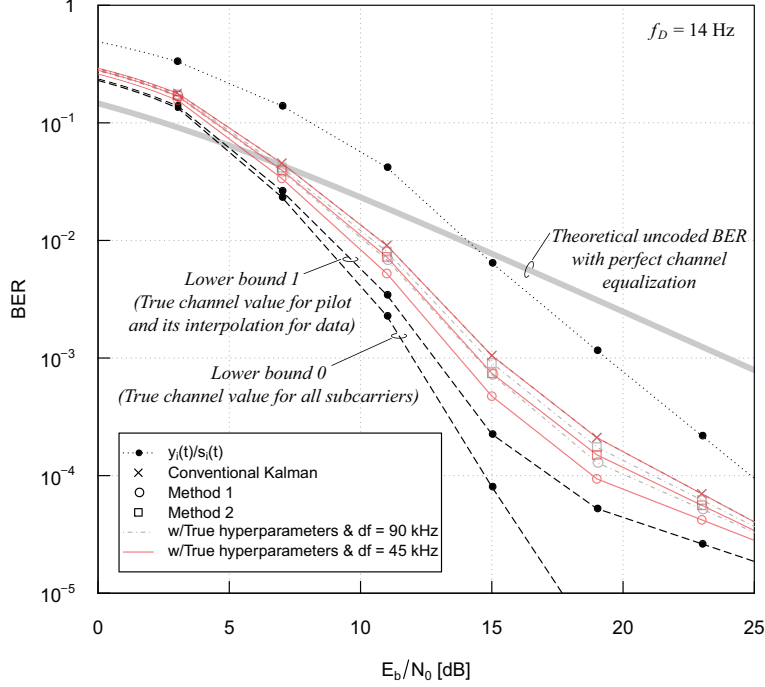


Figure 3.7: Example results of BER performance (with different df values).

conventional Kalman filter and proposed methods; on the other hand, improvements in the frequency domain are not applicable to the conventional Kalman filter (which does not consider frequency correlation) but are applicable to the proposed methods. Therefore, the condition where the proposed methods are effective compared with the conventional Kalman filter is when there is a small df (equivalent to a small σ_τ). To confirm this, Figure 3.7 shows example results of BER performance for $f_D = 14$ Hz with $df = 90$ kHz and $df = 90/2 = 45$ kHz. For comparison, true hyperparameters are applied to both conditions. According to Figure 3.7, the BER performance is not improved by using the conventional Kalman filter (overlapped lines) but is improved by using Methods 1 and 2 with decreasing df values. Table 3.9 gives the E_b/N_0 values that achieve the required BER (10^{-3} and 10^{-4}) for each method with true hyperparameters and $df = 45$ kHz.

CHAPTER 3. STATE-SPACE MODELING

Table 3.9: E_b/N_0 values that achieve required BER (with true hyperparameters and $df = 45$ kHz)

Required BER	E_b/N_0				
	$y_i(t)/s_i(t)$	Conventional Kalman filter	Method 2	Method 1	Lower bound 1
10^{-3}	19.36	15.12	14.48	13.76	13.00
10^{-4}	24.99	21.71	20.65	18.84	17.23

Based on these results, the following conclusions regarding coding gain with true hyperparameters and $df = 45$ kHz are obtained.

- Coding gain improvements using Methods 1 and 2 are very large compared with those for $y_i(t)/s_i(t)$.
- Coding gain improvements using Method 2 are acceptable compared with those for the conventional Kalman filter.
- Coding gain improvements using Method 1 are large compared with those for the conventional Kalman filter.
- The coding gain difference between Method 1 and Lower bound 1 is narrow.

Based on these results, the proposed scheme is shown to improve channel estimation accuracy.

3.5.2 Effects of Reducing Computational Complexity

Here, we compare the computational complexity for Methods 1 and 2 and for the conventional Kalman filter. The computational complexity for Method 1 corresponds to $2I^3 + 4I^2$ in Table 3.2, and that for Method 2 corresponds to $I^3 + 4I^2 + 4I$ in Table 3.3. The computational complexity for the conventional Kalman filter (which does not consider frequency correlation) corresponds to $I^3 + 3I^2 + 2I$ in Table 3.10, because $\mathbf{F}(t) = \mathbf{S}(t)\mathbf{\Sigma} = \mathbf{S}(t)\mathbf{E}_I^{1/2} = \mathbf{S}(t)$.

3.5. NUMERICAL ANALYSIS

Table 3.10: Computational complexity for the conventional Kalman filter

Each step in sequential process at t	Complexity
1. One-step-ahead prediction of state	
(Mean) $\overset{(I,1)}{\mathbf{a}}(t) \leftarrow \mathbf{G}(t)\overset{(I,1)}{\mathbf{m}}(t-1) = \overset{(I,1)}{\mathbf{m}}(t-1)$	0
(Variance) $\overset{(I,I)}{\mathbf{R}}(t) \leftarrow \mathbf{G}(t)\overset{(I,I)}{\mathbf{C}}(t-1)\overset{(I,I)}{\mathbf{G}}(t)^H + \mathbf{W}$ $= \overset{(I,I)}{\mathbf{C}}(t-1) + \mathbf{W}$	0
for $i \leftarrow 0$ to $I-1$ do	
2. One-step-ahead prediction of observation	
(Variance) $\overset{(I,1)}{\mathbf{q}}(t) \leftarrow \overset{(I,I)}{\mathbf{R}}(t)\overset{(I,1)}{\mathbf{F}_i}(t)^H$	I
$\overset{(1,1)}{Q}(t) \leftarrow \overset{(1,I)}{\mathbf{F}_i}(t)\overset{(I,1)}{\mathbf{q}}(t) + \overset{(1,1)}{V_i}$	1
3. Kalman gain	
$\overset{(I,1)}{\mathbf{K}}(t) \leftarrow \overset{(I,1)}{\mathbf{q}}(t)\overset{(1,1)}{Q}(t)^{-1}$	I
4. Filtered state	
(Mean) $\overset{(I,1)}{\mathbf{a}}(t) \leftarrow \overset{(I,1)}{\mathbf{a}}(t) + \overset{(I,1)}{\mathbf{K}}(t)[\overset{(1,1)}{y_i}(t) - \overset{(1,I)}{\mathbf{F}_i}(t)\overset{(I,1)}{\mathbf{a}}(t)]$	$I+1$
(Variance) $\overset{(I,I)}{\mathbf{R}}(t) \leftarrow \overset{(I,I)}{\mathbf{R}}(t) - \overset{(I,1)}{\mathbf{K}}(t)\overset{(1,I)}{\mathbf{q}}(t)^H$	I^2
end for	
5. Conclusion of the loop for $i \Rightarrow$ Go to Step 1 at $t+1$	
(Mean) $\overset{(I,1)}{\mathbf{m}}(t) \leftarrow \overset{(I,1)}{\mathbf{a}}(t)$	0
(Variance) $\overset{(I,I)}{\mathbf{C}}(t) \leftarrow \overset{(I,I)}{\mathbf{R}}(t)$	0
Sum: $I^3 + 3I^2 + 2I$	

Figure 3.8 shows the complexity ratio of Methods 1 and 2 to the conventional Kalman filter. As shown in Figure 3.8, the computational complexity for Methods 1 and 2 increases compared with that for the conventional Kalman filter. This result shows the basic trade-off relationship between performance (described in Section 3.5.1) and computational complexity. However, Method 2 can reduce computational complexity compared with Method 1 and suppress the increased complexity compared with the conventional Kalman filter. In particular, Methods 1 and 2 having computational complexity ratio values of 1.96 and 1.02, respectively, at $I = 50$ corresponding to the computer simulation environment discussed in Section 3.5.1. In such a case, Method 2 reduces computational complexity by 48% compared with Method 1 and suppresses the increased complexity up to 2% compared with the conventional Kalman filter.

CHAPTER 3. STATE-SPACE MODELING

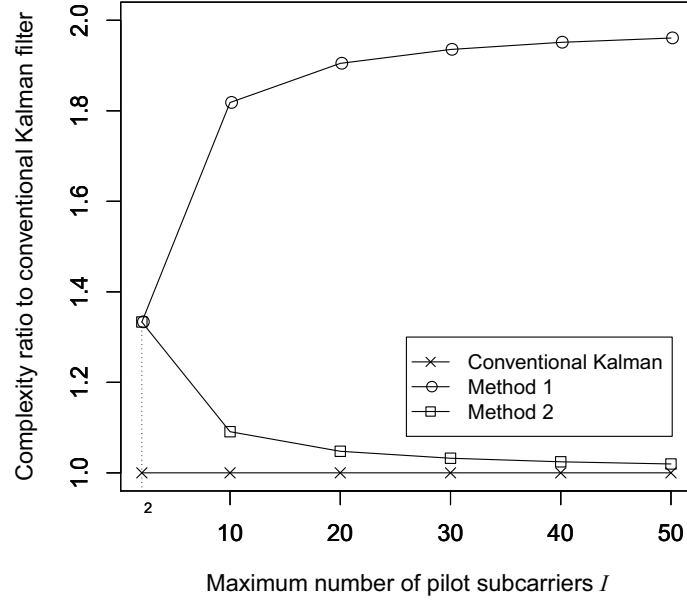


Figure 3.8: Computational complexity comparison.

3.6 Summary

This chapter proposes a novel scheme for sequential OFDM channel estimation on the receiver side; it comprises two methods: Method 1, which improves estimation accuracy, and Method 2, which reduces computational complexity. Method 1 appropriately considers frequency correlation based on a state-space model. Hyperparameters in the state-space model are specified by the ML method, assuming these hyperparameters do not change during a short training period before channel estimation; the ML specification error does not have a dominant influence on performance. Method 2 is based on Method 1 and forces the observation matrix into a sparse bidiagonal matrix. The effect of the proposed scheme is confirmed on the basis of numerical analysis that assumes CRSs in 3GPP LTE as pilot subcarriers. The results of that analysis lead to the following conclusions regarding performance:

- NMSE
 - The NMSE is improved in order of the LS estimator $y_i(t)/s_i(t)$, the conventional Kalman filter (which does not consider frequency correlation), Method 2, and Method 1.
 - NMSE improvements reach, at most, about 50% compared with the conventional Kalman filter.
 - The lower the E_b/N_0 value becomes, the more effective the proposed methods are at improving the NMSE.

3.6. SUMMARY

- BER
 - The BER is improved in the same order as the NMSE.
 - Coding gain improvements reach, at most, about 1 dB compared with the conventional Kalman filter.
 - The condition where the proposed methods are effective, compared with the conventional Kalman filter, is when there is a small frequency interval, df , between adjacent pilot subcarriers (which is equivalent to a small channel delay spread, σ_τ). When df is reduced to half, coding gain improvements with true hyperparameters reach, at most, about 3 dB compared with the conventional Kalman filter and the BER performance is closer to the lower bound.
- Computational complexity
 - The proposed methods' complexity is higher than that of the conventional Kalman filter.
 - Method 2 reduces complexity by about 50% compared with Method 1 and suppresses increased complexity up to a few percents compared with the conventional Kalman filter.

Thus, the proposed scheme improves channel estimation accuracy and reduces computational complexity.

This chapter assumes time-invariant hyperparameters for channel estimation, but this assumption is not always practical in real environments. The online joint estimation method for hyperparameters [3.16] can be applied to time-varying hyperparameters and has been actively studied in recent years. Such an extension is investigated in the next chapter.

CHAPTER 3. STATE-SPACE MODELING

Bibliography

- [3.1] M. Liyanage and I. Sasase, "Steady-state Kalman filtering for channel estimation in OFDM systems for Rayleigh fading channels," *IE-ICE Trans. Communications*, vol. E92-B, no. 7, pp. 2452–2460, 2009.
- [3.2] W. Chen and R. Zhang, "Estimation of time and frequency selective channels in OFDM systems: a Kalman filter structure," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 2, Nov. 2004, pp. 800–803.
- [3.3] J. R. Schott, *Matrix Analysis for Statistics*, 2nd ed. John Wiley & Sons, Inc., 2005.
- [3.4] S. Stein and J. J. Jones, *Modern Communication Principles with Application to Digital Signaling*. McGraw-Hill, 1967.
- [3.5] W. C. Jakes, Jr., *Microwave Mobile Communications*. Wiley, New York, 1974, [reprinted by IEEE Press].
- [3.6] T. K. Akino, "Optimum-weighted RLS channel estimation for rapid fading MIMO channels," *IEEE Trans. Wireless Communications*, vol. 7, no. 11, pp. 4248–4260, Nov. 2008.
- [3.7] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Apr. 1991.
- [3.8] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models*, 2nd ed. Springer-Verlag, 1997.
- [3.9] D. Falconer and L. Ljung, "Application of fast Kalman estimation to adaptive equalization," *IEEE Trans. Communications*, vol. 26, no. 10, pp. 1439–1446, Oct. 1978.
- [3.10] B. Anderson and J. Moore, *Optimal Filtering*. Prentice-Hall Englewood Cliffs, N.J., 1979, vol. 11.
- [3.11] K. Saito, J. Webber, T. Nishimura, T. Ohgane, and Y. Ogawa, "Tri-diagonalizing approach on frequency domain equalization in a doubly-selective channel," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 2009, pp. 1–5.
- [3.12] N. J. Higham, "Computing the nearest correlation matrix — a problem from finance," *IMA Journal of Numerical Analysis*, vol. 22, no. 3, pp. 329–343, 2002.
- [3.13] D. L. Knol and J. M. F. ten Berge, "Least-squares approximation of an improper correlation matrix by a proper one," *Psychometrika*, vol. 54, no. 1, pp. 53–61, 1989.

BIBLIOGRAPHY

- [3.14] “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation,” The 3rd Generation Partnership Project (3GPP), TS 36.211, Dec. 2009.
- [3.15] L. Wang, G. Libert, and P. Manneback, “Kalman filter algorithm based on singular value decomposition,” in *Proc. 31st IEEE Conference on Decision and Control (CDC)*, Tucson, U.S.A., Dec. 1992, pp. 1224–1229.
- [3.16] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic Linear Models with R*. Springer, 2009.
- [3.17] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

Chapter 4

Joint Estimation

4.1 Preliminary Remarks

As previously confirmed in Chapter 3, channel statistics such as velocity and delay spread are needed for accurate channel estimation. These are rarely known, but can be estimated by various methods. For instance, the velocity can be estimated by autocorrelating the received signal [4.1,4.2]. Hagiwara [4.3] proposed that several channel statistics including velocity can be set as hyperparameters in a stochastic model and estimated by marginal likelihood maximization. Such methods, classified as batch methods, rely on sufficient data for accurate estimation. For example, Hagiwara's method [4.3] requires a ~ 0.1 -s training period before the channel estimation. However, real channel statistics in mobile communications temporally vary and may abruptly respond to human/vehicle behaviors and communication environments. Therefore, batch methods might estimate outdated statistics in real-time communication environments, which degrades their performance. To adapt to temporal changes, we prefer real-time methods that estimate while receiving data. For this purpose, we have developed an on-line/sequential joint estimation method.

Several on-line/sequential joint estimation methods have been proposed, some of which model the temporal changes in channel statistics; e.g., Huang et al. [4.4] modeled the velocity dynamics using an auto regressive (AR) model for blind detection under a flat fading channel. However, the AR model is not readily adaptable to abrupt changes. The stochastic model of Nemeth et al. [4.5] handles abrupt changes in object-location tracking, but assumes a known average ratio of the abrupt change, which may not apply to real mobile communication environments. Because the channel statistics in real mobile communications change on a case-by-case basis, they are inherently difficult to model universally. To track the dynamics of channel statistics, this study pragmatically adopts a model-free approach (fast repetitive on-line/sequential estimation) rather than a

CHAPTER 4. JOINT ESTIMATION

model-based approach. Thus, the new method quickly estimates the fixed channel statistics.

Among the existing on-line/sequential methods for estimating joint channels and their fixed statistics, we adopt the modern and effective Liu and West filter (LWF) [4.6] based on the state-space model and sequential Monte Carlo method. For the first time, we demonstrate that the conventional LWF biases the hyperparameter from poor likelihood estimation caused by overfitting [4.7] in noisy environments. To overcome this problem, which cannot be solved by conventional smoothing techniques, we modify the conventional LWF and regularize the likelihood using a Kalman smoother. Using numerical analyses, we confirm the higher estimation accuracy of our proposed method considering the Rao–Blackwellisation (RB) technique over the conventional LWF and least-squares (LS) methods.

The remainder of this chapter is structured as follows. Section 4.2 provides the background of the study (problem formulation and simulation configuration), and Section 4.3 introduces existing theories. Section 4.4 demonstrates the need to improve the existing method and proposes our novel method with enhanced estimation accuracy. Section 4.5 analyses the numerical results, and Section 4.6 summarizes the discussions.

4.2 Background

4.2.1 Problem Formulation

OFDM channel can be estimated by various models [4.3, 4.8–4.11]. This study adopts the state-space model of [4.3], but excludes the extended formulation for simplicity. The observation and state equations in this model are expressed as [4.3]:

$$\mathbf{y}(t) = \mathbf{F}(t) \boldsymbol{\theta}(t) + \mathbf{v}(t), \quad (4.1)$$

$$\boldsymbol{\theta}(t) = \mathbf{G}(t) \boldsymbol{\theta}(t-1) + \mathbf{w}(t), \quad (4.2)$$

where (for the pilot subcarriers at the t -th symbol) $\mathbf{y}(t)$ and $\boldsymbol{\theta}(t) = [\theta_0(t), \dots, \theta_i(t), \dots, \theta_{I-1}(t)]^T$ denote the $I \times 1$ observation vector and the $I \times 1$ state vector, respectively, and $\mathbf{F}(t)$ and $\mathbf{G}(t)$ denote the $I \times I$ observation matrix and $I \times I$ transition matrix, respectively. For the pilot subcarriers at the t -th symbol, the covariance matrix in the $I \times 1$ observation noise vector $\mathbf{v}(t) \sim \mathcal{CN}(\mathbf{0}_I, \mathbf{V})$ is set to $\mathbf{V} = \text{diag}(V_0, \dots, V_i, \dots, V_{I-1})$, while the covariance matrix in the $I \times 1$ state noise vector $\mathbf{w}(t) \sim \mathcal{CN}(\mathbf{0}_I, \mathbf{W})$ is set to $\mathbf{W} = \text{diag}(W_0, \dots, W_i, \dots, W_{I-1})$. $\mathbf{y}(t)$ and $\mathbf{v}(t)$ in Equation (4.1) are identical with them in Equation (2.4).

4.2. BACKGROUND

The terms in the observation equation are specified below:

$$\tilde{\mathbf{h}}(t) = \mathbf{\Sigma} \boldsymbol{\theta}(t), \quad (4.3)$$

$$\mathbf{F}(t) = \mathbf{S}(t)\mathbf{\Sigma}, \quad (4.4)$$

$$\mathbf{\Sigma} = \left(\mathbf{\Omega}_f\right)^{1/2}, \quad (4.5)$$

$$\left(\mathbf{\Omega}_f\right)_{\text{row, col}} = \rho(\Delta t = 0, \Delta f = (\text{col} - \text{row})df), \quad (4.6)$$

$$\rho(\Delta t = 0, \Delta f) = \frac{1 + j2\pi\Delta f\sigma_\tau}{1 + (2\pi\Delta f\sigma_\tau)^2}, \quad (4.7)$$

where $\rho(\Delta t, \Delta f)$ denotes the time versus frequency correlation coefficient of the narrow-band channel gain, Δt and Δf denote the time and frequency differences, respectively, and σ_τ is the channel delay spread.

The terms in the state equation are given as

$$\mathbf{G}(t) = \mathbf{E}_I, \quad (4.8)$$

$$\mathbf{W} = 2l^2(1 - \rho(\Delta t = 1dt, \Delta f = 0))\mathbf{E}_I, \quad (4.9)$$

$$\rho(\Delta t, \Delta f = 0) = J_0(2\pi f_D \Delta t), \quad (4.10)$$

l^2 denotes the average power of the narrow-band channel gain of each pilot subcarrier, J_0 represents a zero-order Bessel function of the first kind [4.12], and f_D is the maximum Doppler frequency of the channel.

In this model, the narrow-band channel gains correspond to the parameter $\boldsymbol{\theta}(t)$ and the channel statistics correspond to hyperparameters σ^2 , σ_τ , l^2 , and f_D (the latter is equivalent to velocity). For simplicity, we assume fundamental hyperparameters for OFDM mobile communication and that σ_τ and f_D can be unknown, while σ^2 and l^2 are known via some other method. Hereafter, the hyperparameters are grouped as $\boldsymbol{\psi} = \{\sigma_\tau, f_D\}$. The lower and upper bounds of the hyperparameters $\boldsymbol{\psi}$ [4.13] are specified in Table 4.1. When the hyperparameters $\boldsymbol{\psi}$ are specified, the above linear Gaussian state-space model can be solved by a Kalman filter [4.14]. The recursion process of the Kalman filter at the t -th symbol is presented as Algorithm 2.1. In the above model, the Kalman filtering process depends only on the hyperparameters, prior distribution whose mean and covariance correspond to $\mathbf{m}(t-1)$ and $\mathbf{C}(t-1)$ in Algorithm 2.1, respectively, and received data and is hereafter represented by $\mathcal{KF}(\text{hyp}, \text{pri}, \text{dat})$.

Table 4.1: Lower and upper bounds of hyperparameters

Hyperparameter	Lower bound	Upper bound
σ_τ [μs]	0.2	5.0
Velocity [km/h]	0	300

CHAPTER 4. JOINT ESTIMATION

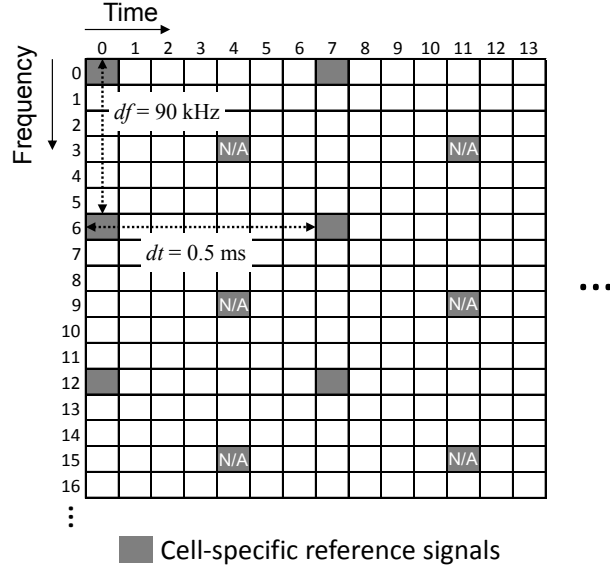


Figure 4.1: Allocation of scattered pilot subcarriers.

4.2.2 Simulation Configuration

In the simulations, we used 3GPP LTE specifications with a 5 MHz bandwidth. The frequency interval between each subcarrier was 15 kHz, and the symbol length (excluding the GI) was $66.7 \mu\text{s}$. The fast Fourier transform (FFT) size was 512, wherein 300, 1, and 211 were the sizes for the active, direct current (DC), and guard subcarriers, respectively. The guard band was $5 - (300+1) \times 0.015 = 0.485 \text{ MHz}$. As the pilot subcarriers, we adopted cell-specific reference signals (CRSs) [4.15], allocated as shown in Figure 4.1. As shown in the figure, the CRSs at timeslots 4, 11, $\dots$ were excluded for simplicity. We set $dt = 0.5 \text{ ms}$, $df = 90 \text{ kHz}$ and modulated the specific pseudorandom sequence for the CRSs by quadrature phase-shift keying (QPSK). There were 250 data subcarriers and 50 pilot subcarriers for one symbol duration. An isolated cell and single-user environment without inter-cell interference and multiple access were assumed. The other assumptions of the computer simulation are summarized in Table 4.2.

Table 4.2: Assumptions of the computer simulation

Carrier frequency	2 GHz
Maximum Doppler frequency: f_D	28, 140 Hz (i.e., 15.3, 76.5 km/h)
Delay profile	• 24-path exponential – Delay spread: 0.5 μs for $f_D = 28$ Hz 1.0 μs for $f_D = 140$ Hz – Maximum delay: 4.7 μs (within GI)
Numerical computation of Kalman filter/smoothen	A square root Kalman filter/smoothen based on singular value decomposition [4.16] is applied to suppress the degradation in numerical accuracy.

4.3 Theory

Numerical examples of two existing methods with velocity as the unknown hyperparameter are presented.

4.3.1 Primitive Method Based on Multi-Model

This method selects the model with the best criteria among multiple models. In this study, the criterion is the instantaneous likelihood¹. When performing parallel Kalman filters for the multiple models, Algorithm 2.1 leads instantaneous likelihood at the t -th symbol as

$$\begin{aligned}
p(\mathbf{y}(t) | \mathbf{y}(1:t-1)) &= \mathcal{CN}(\mathbf{y}(t); \mathbf{f}(t), \mathbf{Q}(t)) \\
&= \mathcal{CN}(\mathbf{y}(t); \mathbf{F}(t)\mathbf{a}(t), \mathbf{F}(t)\mathbf{R}(t)\mathbf{F}(t)^H + \mathbf{V}) \\
&= \mathcal{CN}(\mathbf{y}(t); \mathbf{F}(t)\mathbf{G}(t)\mathbf{m}(t-1), \\
&\quad \mathbf{F}(t)[\mathbf{G}(t)\mathbf{C}(t-1)\mathbf{G}(t)^H + \mathbf{W}]\mathbf{F}(t)^H + \mathbf{V}). \quad (4.11)
\end{aligned}$$

Similar to the Kalman filter, the instantaneous likelihood depends only on the hyperparameters, prior distribution, and received data and is hereafter denoted as $\mathcal{L}(\text{hyp}, \text{pri}, \text{dat})$.

The performance of this method was evaluated in computer simulations with $E_b/N_0 = 20$ dB, $f_D = 140$ Hz, and $\sigma_\tau = 1.0$ μ s. The multi-model velocities were divided into 10 equal segments, i.e., 10 models with velocities $33.3 \times \{0, 1, \dots, 9\}$ km/h. Figure 4.2 shows the simulation results.

¹Maximizing the time integration of this instantaneous log-likelihood is equivalent to batch-type marginal likelihood maximization.

CHAPTER 4. JOINT ESTIMATION

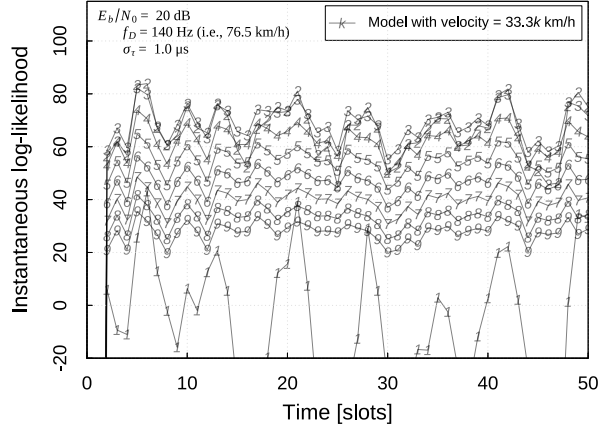


Figure 4.2: Instantaneous log-likelihood in multiple models. The X-axis denotes the time elapsed in the simulation (one slot corresponds to one dt).

According to Figure 4.2, the model with velocity around the true value (75.6 km/h) almost always exhibits the highest instantaneous likelihood, indicating that this velocity is quickly estimated with some fluctuations. These fluctuations can be suppressed by methods such as the interacting multiple model (IMM) [4.17,4.18], which weights multiple models by their transition probabilities. The multi-model approach provides fairly accurate hyperparameter estimation, but requires finer granularity on the hyperparameter settings for higher accuracy. For example, the velocities in the multi-model should be divided into 20, 30, or 40 segments rather than 10 segments. Naively applying this approach to joint estimation with many hyperparameters becomes unrealistically complex because the combination of hyperparameter settings explodes with exponential order; the so-called curse of dimensionality [4.19]. To overcome this drawback, we restrain the complexity by the method described in the next subsection.

4.3.2 Modern Method Based on Sequential Monte Carlo

To enhance the estimation accuracy while constraining the complexity, we must track the likely hyperparameters rather than fix them in a multi-model. Such a mechanism is enabled by the particle filter [4.20,4.21] based on sequential Monte Carlo methods. This simple and powerful method solves nonlinear and non-Gaussian state-space models [4.22] and has remarkably developed in recent years. LWF is popularly used in joint estimates of parameters and fixed hyperparameters based on particle filters. Therefore, LWF forms the basis of the present proposal.

4.3.2.1 Algorithm

In one of the state-space-model-based approaches for joint parameter and hyperparameter estimation, the hyperparameter is a parameter in an augmented state. This consistent approach is called the self-organizing state-space model [4.23]. Even though the original state follows a linear and Gaussian state-space model, the augmented state generally obeys a nonlinear and non-Gaussian state-space model, which cannot be solved by the Kalman filter. LWF is based on a self-organizing state-space model, and its nonlinear and non-Gaussian state-space model is solved using a particle filter.

In this study, we discuss the application of the Rao–Blackwellised auxiliary particle filter (RB-APF) as the first step. The RB-APF is a combination of the Rao–Blackwellised particle filter reported in Section 2.3.3.3 and the auxiliary particle filter reported in Section 2.3.3.2. The augmented state in the self-organizing state-space model is denoted by $\mathbf{x}(t) = \{\text{Nonlinear and non-Gaussian part, Linear and Gaussian part}\} = \{\psi, \theta(t)\}$. Hereafter, the hyperparameter ψ is expressed as an explicit function of time, $\psi(t)$, considering time variation during estimation. Algorithm D.1 in Appendix D presents the recursion of the RB-APF at the t -th symbol, where the target distribution is the filtered distribution $p(\psi(t) \mid \mathbf{y}(1:t))$.

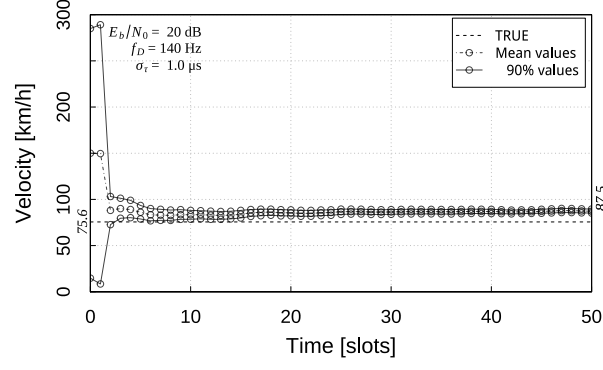
In the purest fixed hyperparameter estimation by the RB-APF, the realizations $\psi(t)^{(n)}$ are selected from only its initial prior distribution. However, this approach will probably achieve insufficient estimation accuracy or fail because the particle diversity is limited. LWF avoids this problem using a kernel smoothing technique [4.24], which retains the particle diversity without increasing the variance. The kernel smoothing technique proceeds in two steps: artificial moving averaging and then a new realization drawn from a continuous distribution with mean and variance set to the moving average and the dispersion decrement by the moving average, respectively. Liu and West [4.6] suggested 0.974–0.995 as the forgetting factor a in the artificial moving average. To ensure stable estimation, we adopt the lower value (0.974). Algorithm E.1 in Appendix E presents the recursion process of RB-APF with the kernel smoothing technique at the t -th symbol. The target distribution is the filtered distribution $p(\psi(t) \mid \mathbf{y}(1:t))$. Lower and upper bounds of the hyperparameters are listed in Table 4.1 and the continuous distribution (a-1 of Algorithm E.1) is the scaled beta distribution, as in [4.25]. The remainder of this study is based on Algorithm E.1. Note that APF with the kernel smoothing technique is usually called LWF [4.26], so Algorithm E.1 is referred to as RB-LWF.

CHAPTER 4. JOINT ESTIMATION

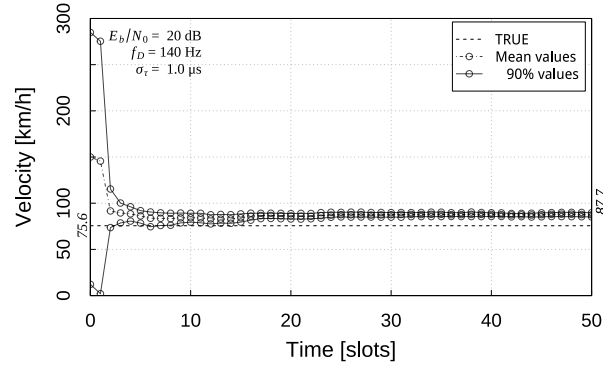
4.3.2.2 Example of Algorithm Performance

To minimize the complexity, we identified the minimum number of particles that did not compromise the accuracy. Figure 4.3 plots the simulation results for $N = 10, 100$, and 1000 . Other conditions are $E_b/N_0 = 20$ dB, $f_D = 140$ Hz, and $\sigma_\tau = 1.0$ μ s. The value at 0 slots corresponds to prior distribution, and the 90% values indicate the upper 95% and lower 5% values. These representations are commonly used in the following descriptions. The prior velocities are equally divided into particle number segments. When one hyperparameter is unknown, 10 particles achieve similar behavior and estimation accuracy to 100 and 1000 particles. Slight difference among converged estimators does not indicate a clear difference of accuracy, because LWF is a stochastic method and essentially gives some fluctuations to the estimation. Therefore, except in special cases, simulations involving one unknown hyperparameter were performed with 10 particles.

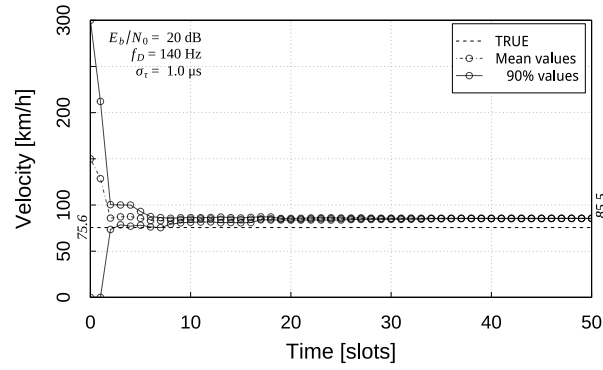
4.3. THEORY



(a)



(b)



(c)

Figure 4.3: Hyperparameter estimation using RB-LWF with 1000 particles (top), 100 particles (center), and 10 particles (bottom).

4.4 Methods

This section reveals the awkward overfitting problem in sequential hyperparameter estimation, which has not been clearly recognized to date. When velocity is the unknown hyperparameter, the conventional LWF biases the hyperparameter from the poor likelihood estimation caused by overfitting in noisy environments. Moreover, this problem cannot be solved by conventional smoothing techniques. Therefore, our new method modifies the conventional LWF and regularizes the likelihood using a Kalman smoother.

4.4.1 Why Should LWF be Improved?

As confirmed in Section 4.3.2.2, LWF estimates sufficiently accurate hyperparameters at high E_b/N_0 , but its performance may degrade at low E_b/N_0 . To show this, Figure 4.4 plots the simulation result for $E_b/N_0 = 0$ dB, $f_D = 140$ Hz, and $\sigma_\tau = 1.0$ μ s. The hyperparameter is clearly overestimated. This result is not a chance occurrence but is caused by the noisy environment. To facilitate understanding, we show another example.

Figure 4.5 plots the simulation results of the Kalman filter and smoother at $E_b/N_0 = 0$ dB, $f_D = 28$ Hz, and $\sigma_\tau = 0.5$ μ s. Here, the filtering hyperparameter f_D is set to true (28 Hz) and intentionally large false (84 Hz) values, while the smoothing hyperparameter is set to its true value. Filtering with the true hyperparameter yields a phase-shifted estimator because the relative future information is not principally considered [4.27]. The lower the E_b/N_0 and the larger the estimation error, the larger the phase shift. On the other hand, the false filtering hyperparameter suppresses the phase shift because high f_D facilitates sensitive adaptation to noise. The marginal log-likelihoods under filtering with true and false hyperparam-

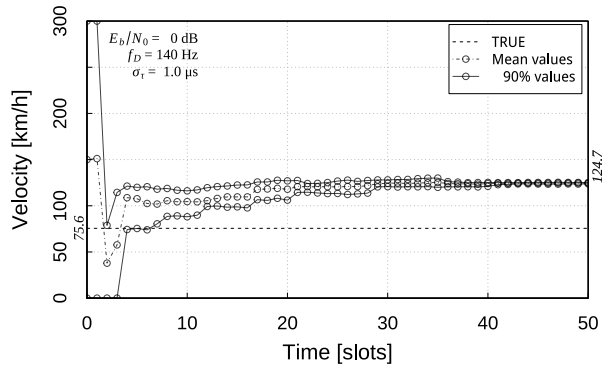


Figure 4.4: Hyperparameter estimation by RB-LWF in a noisy environment.

4.4. METHODS

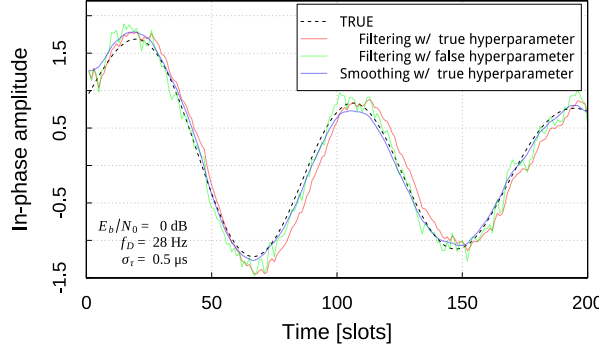


Figure 4.5: Channel estimation using the Kalman filter and smoother. The Y-axis denotes the in-phase amplitude of the narrow-band channel gain at pilot subcarrier index $i = 5$ ($i = 0, \dots, 49$).

eters are 1258.8 and 2431.4, respectively; the false hyperparameter yields the better result. In machine learning theory, this phenomenon is called overfitting. RB-LWF proceeds from the wrong likelihood in noisy environments and thus overestimates the hyperparameter. Smoothing with the true hyperparameter solves the overfitting phenomenon by considering the relative future information (Figure 4.5). This thesis enhances the estimation accuracy by fixed-lag smoothing [4.28], which suppresses the increase in latency. Following [4.29], the tolerable lag is hereafter set to 5 slots ($5 \times 0.5 = 2.5$ ms).

4.4.2 Conventional Particle Smoothing

Among the smoothing methods [4.30, 4.31] applied to particle filters, we apply the most basic technique to RB-LWF; the sequential importance resampling (SIR) smoother [4.21]. The SIR smoother considers the augmented state (including the previous lagged states); consequently, the filtered state reflects the relative future information. The particles in this augmented state are affected by repeated resampling of lag times, which renders them degenerate. To avoid such degeneracy, number of particles in the SIR smoother is increased to 1000. Figure 4.6 plots the simulation results for $E_b/N_0 = 0$ dB, $f_D = 140$ Hz, and $\sigma_\tau = 1.0$ μ s with 1000 particles. The behaviors of RB-LWF are similar in the simulations with 10 particles (c.f. Figure 4.4). Comparing the filtered and smoothed results, we find that the SIR smoother suppresses early converging fluctuations but does not improve the final converged value. Like most existing particle smoothers, the SIR smoother reselects the filtered particles *after filtering*, based on the new criteria established from the relative future information. This simple backward process directly realizes basic Bayesian smoothing [4.28], and

CHAPTER 4. JOINT ESTIMATION

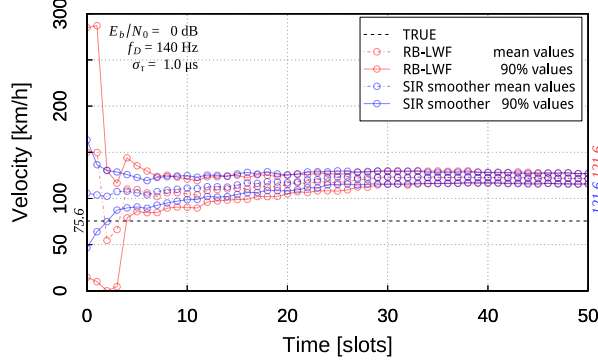


Figure 4.6: Hyperparameter estimation using RB-LWF and SIR smoother with 1000 particles.

the popular Rauch–Tung–Striebel (RTS) algorithm [4.32] is easily applied to the analytical Kalman filter [4.33]. On the other hand, to converge the estimator, RB-LWF repeats the resampling process without increasing the variance. Therefore, the converging particle gradually loses the possibility of reselection. There is a difficulty of applying conventional particle smoothers to the RB-LWF. To improve the hyperparameter estimation accuracy of RB-LWF, we must correct the wrong likelihood *while filtering*.

4.4.3 Novel Proposed Smoothing

To correct or regularize the likelihood during filtering, this study adopts the Kalman smoother for RB-LWF. During filtering with resampling, the fixed lag duration is assumed sufficiently short to validate the Kalman smoother with initial hyperparameter values, and the Kalman filter is replaced by the Kalman smoother in the likelihood calculation. The recursion of the Kalman smoother at the t -th symbol is presented as Algorithm 2.2. According to Algorithms 2.1 and 2.2, Kalman smoothing depends only on hyperparameters, prior distribution, and received data and is hereafter denoted as $\mathcal{KS}(\text{hyp}, \text{pri}, \text{dat})$. The proposed plan basically replaces the filtered state and $\mathcal{KF}(\text{hyp}, \text{pri}, \text{dat} = \mathbf{y}(t))$ in Algorithm E.1 with the smoothed state and $\mathcal{KS}(\text{hyp}, \text{pri}, \text{dat} = \mathbf{y}(t:t+5))$, respectively. However, in this approximation, regular application of the Kalman smoother at every pilot symbol is inadequate. Particularly, the realization can be analytically derived, but the likelihood regularization should be moderately restricted, because it directly affects the sensitive stochastic resampling process. Thus, the frequency of likelihood regularization should be tuned in the algorithm design. At the best regularization frequency, the greedy pre-search under different hyperparameter combinations would achieve the smallest estimation error on average. Accordingly, Algorithm E.1 was

4.4. METHODS

partially modified to Algorithm 4.1, where the modifications are colored in blue. Estimation accuracy is improved by the modifications where the smoothing considering relative future information is adequately applied to the sequential calculations of likelihood and realization.

CHAPTER 4. JOINT ESTIMATION

Algorithm 4.1 Proposed smoothing algorithm

0. Filtered state at the $t-1$ -th symbol:

$$\left\{ \text{Realization } \psi(t-1)^{(n)}, \mathbf{m}(t-1)^{(n)}, \mathbf{C}(t-1)^{(n)}, \text{Weight } \omega(t-1)^{(n)} \right\}_{n=1}^N$$
1. Update process at the t -th symbol
 - *Artificial moving average for hyperparameters*

$$\mu^{(n)} \leftarrow a\psi(t-1)^{(n)} + (1-a)E_{\omega(t-1)^{(n)}}[\psi(t-1)^{(n)}],$$

$$\Gamma \leftarrow (1-a^2)\text{Var}_{\omega(t-1)^{(n)}}[\psi(t-1)^{(n)}].$$
 - *Resampling*
 Draw auxiliary index k_n from a set $\{1, \dots, N\}$ with $P(k_n = n) \propto \omega(t-1)^{(n)} p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(n)})$, where
 - $\hat{\psi}(t)^{(n)} = \mu^{(n)}$,
 - $p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(n)})$ is equal to
 - if regularization timing then**

$$\mathcal{L}(\text{hyp}=\hat{\psi}(t)^{(n)}, \text{pri}=\{\hat{\mathbf{u}}(t-1)^{(n)}, \hat{\mathbf{U}}(t-1)^{(n)}\}, \text{dat}=\mathbf{y}(t)), \text{ where}$$

$$\{\hat{\mathbf{u}}(t-1)^{(n)}, \hat{\mathbf{U}}(t-1)^{(n)}\} \text{ are derived via } \mathcal{KS}(\text{hyp}=\hat{\psi}(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(n)}, \mathbf{C}(t-1)^{(n)}\}, \text{dat}=\mathbf{y}(t:t+5)),$$
 - else**

$$\mathcal{L}(\text{hyp}=\hat{\psi}(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(n)}, \mathbf{C}(t-1)^{(n)}\}, \text{dat}=\mathbf{y}(t)).$$
 - end if**
 - **for $n = 1$ to N do**
 - a-1. *Realization* (for nonlinear and non-Gaussian part)
 Draw $\psi(t)^{(n)}$ from

$$\text{continuous distribution } (\text{Mean}=\mu^{(k_n)}, \text{Variance}=\Gamma).$$
 - a-2. *Realization* (for linear and Gaussian part)
 Derive $\{\mathbf{u}(t)^{(n)}, \mathbf{U}(t)^{(n)}\}$ via $\mathcal{KS}(\text{hyp}=\psi(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(k_n)}, \mathbf{C}(t-1)^{(k_n)}\}, \text{dat}=\mathbf{y}(t:t+5)).$
 - b. *Weight*

$$\omega(t)^{(n)} \leftarrow \frac{p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \psi(t)^{(n)})}{p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(k_n)})}, \text{ where}$$
 - $p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \psi(t)^{(n)})$ is equal to
 - if regularization timing then**

$$\mathcal{L}(\text{hyp}=\psi(t)^{(n)}, \text{pri}=\{\mathbf{u}(t-1)^{(n)}, \mathbf{U}(t-1)^{(n)}\}, \text{dat}=\mathbf{y}(t)), \text{ where}$$

$$\{\mathbf{u}(t-1)^{(n)}, \mathbf{U}(t-1)^{(n)}\} \text{ are derived via previous a-2,}$$
 - else**

$$\mathcal{L}(\text{hyp}=\psi(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(k_n)}, \mathbf{C}(t-1)^{(k_n)}\}, \text{dat}=\mathbf{y}(t)).$$
 - end if**
 - end for**
 - *Normalization of the weights:* $\omega(t)^{(n)} \leftarrow \omega(t)^{(n)} / \sum_{n=1}^N \omega(t)^{(n)}.$
2. Filtered state at the t -th symbol:

$$\left\{ \text{Realization } \psi(t)^{(n)}, \mathbf{m}(t)^{(n)} \leftarrow \mathbf{u}(t)^{(n)}, \mathbf{C}(t)^{(n)} \leftarrow \mathbf{U}(t)^{(n)}, \text{Weight } \omega(t)^{(n)} \right\}_{n=1}^N$$

4.4. METHODS

Figure 4.7 plots the simulation results of RB-LWF and the proposed method with $E_b/N_0 = 0$ dB, $f_D = 140$ Hz, $\sigma_\tau = 1.0$ μ s, and $N = 10$. The best regularization frequency determined by the pre-search was “once per 5 slots”; thus, the regularization timing in Algorithm 4.1 becomes “ $(t \bmod 5) = 0$.” The trajectories of RB-LWF and the proposed method diverge after the fifth slot, when the likelihood is first regularized. The proposed smoothing method clearly improves the accuracy of the converged estimator. Particularly, the proposed method avoids the degeneracy problem of the SIR smoother because it never resamples previous lagged particles.

To confirm that the proposed method generally improves the estimation accuracy of hyperparameters under wide-ranging conditions, we plot the cumulative distribution function of the hyperparameter estimation error for various combinations of $E_b/N_0 = \{0, 4, 8, 12, 16, 20\}$ dB, $f_D = 14 \times \{2, 4, 6, 8, 10\}$ Hz, and $\sigma_\tau = 1.0$ μ s. The plots are presented in Figure 4.8. According to Figure 4.8, the proposed method outperforms the conventional RB-LWF on average and generally improves the hyperparameter estimation; however, further improvement is possible.

We here mention related studies to our proposal. The fixed-lag particle filter and its application to hyperparameter estimation have been proposed by [4.34] and [4.35], respectively. These premise sequential importance sampling (SIS) and Markov chain Monte Carlo (MCMC), and differ from our proposal. SIS especially requires a lot of particles for certain estimation accuracy. Resampling process is crucial under limited number of particles. Regarding LWF which uses resampling, Prado and West [4.36] proposed a parameter smoothing method using LWF-estimated hyperparameters, but attempts to smooth the hyperparameters themselves have not been reported. Since our proposed method generally improves the LWF-based hyperparameter estimation accuracy under the Rao–Blackwellisation, it is

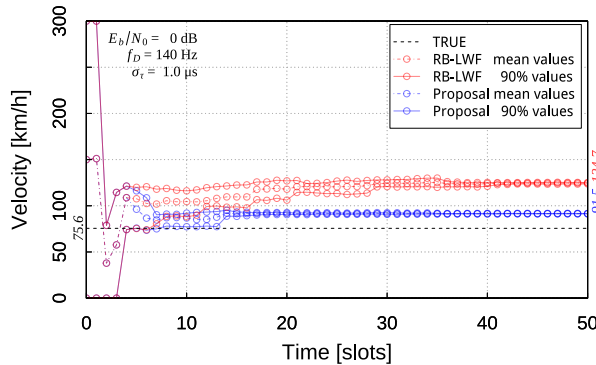


Figure 4.7: Hyperparameter estimation by RB-LWF and the proposal method with 10 particles.

CHAPTER 4. JOINT ESTIMATION

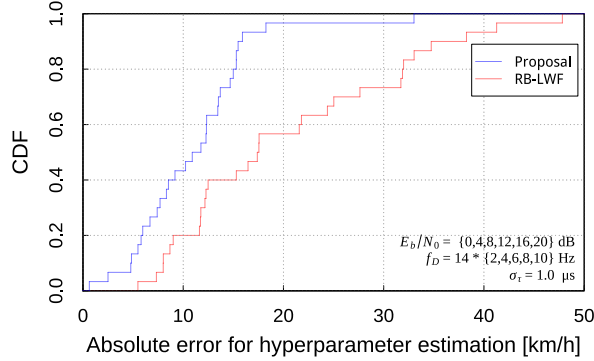


Figure 4.8: General performance of the proposed method.

referred to as the Rao–Blackwellised Liu and West *smoother* (RB-LWS).

4.5 Numerical Results and Discussion

Performances of the joint channel and hyperparameter estimation by RB-LWF and RB-LWS are numerically analyzed when a single hyperparameter (the velocity) is unknown ($\psi = \{f_D\}$) and when two hyperparameters (velocity and delay spread) are unknown ($\psi = \{f_D, \sigma_\tau\}$). The point estimator of the channel is derived by a maximum a posteriori probability (MAP) estimator of the hyperparameter.

4.5.1 The Case of a Single Unknown Hyperparameter (Velocity)

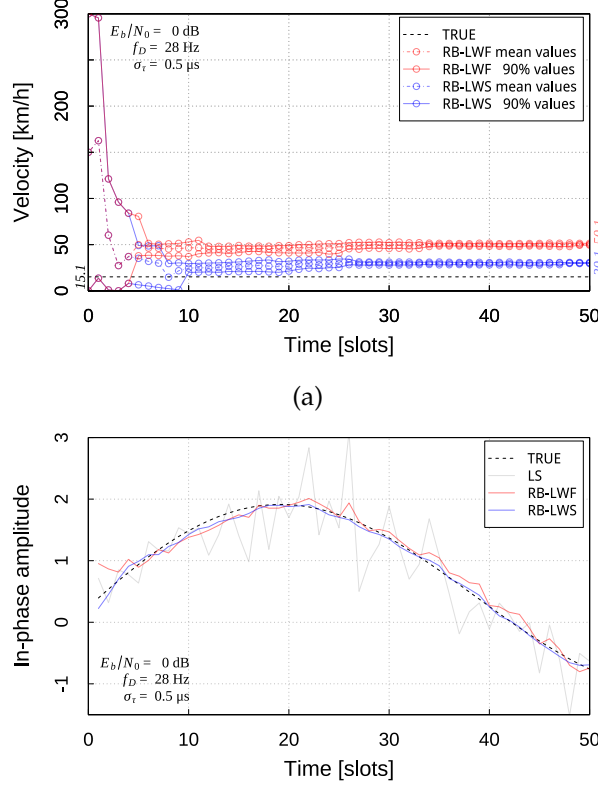
4.5.1.1 Number of Particles and Prior Distribution of Hyperparameters

In both RB-LWF and RB-LWS, 10 particles are employed and the prior distribution is a uniform distribution, as explained in Section 4.3.2.2.

4.5.1.2 Example Result

Figure 4.9 shows the simulation results for $E_b/N_0 = 0$ dB, $f_D = 28$ Hz, and $\sigma_\tau = 0.5$ μ s. The regularization timing in Algorithm 4.1 is set to “ $(t \bmod 5) = 0$,” as discussed in Section 4.4.3. The RB-LWF overestimates the converged hyperparameter due to low E_b/N_0 , but the RB-LWS estimate approaches the true value. Red and blue trajectories diverge after the fifth slot, when the likelihood is first regularized. Moreover, the channel accuracy is higher in RB-LWS than in RB-LWF. Note that the channel estimator based on RB-LWF tends to follow the noise because its hyperparameter is overestimated.

4.5. NUMERICAL RESULTS AND DISCUSSION



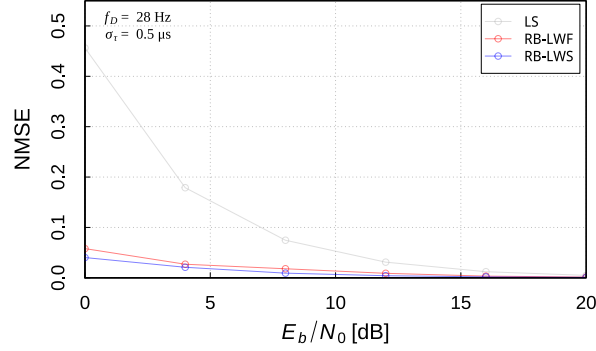
(b) The Y-axis denotes the in-phase amplitude of the narrow-band channel gain at pilot subcarrier index $i = 34$ ($i = 0, \dots, 49$). LS denotes the least-squares estimator $y_i(t)/s_i(t)$ with forcing the corresponding channel impulse response over GI to zero [4.11].

Figure 4.9: Simulated velocity (top) and channel estimation (bottom) in case of a single unknown hyperparameter.

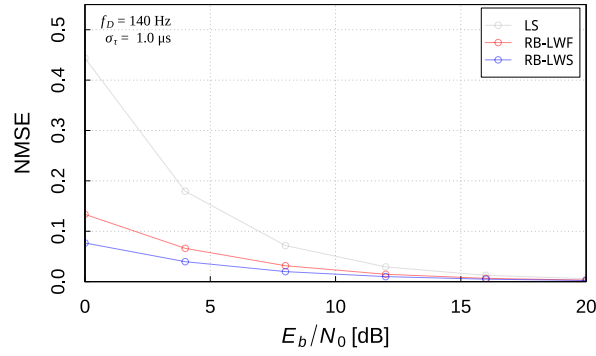
4.5.1.3 Normalized Mean Squared Error (NMSE) of Channel Estimator

Figure 4.10 plots the NMSE performances of the channel estimation for $f_D = 28$ and 140 Hz; $\text{NMSE} = \sum_{i=0}^{I-1} E[|\tilde{h}_i(t) - \hat{h}_i(t)|^2] / \sum_{i=0}^{I-1} E[|\tilde{h}_i(t)|^2]$. The hyperparameter is jointly estimated before 50 slots and thereafter converges to a fixed value. The RB-LWS outperforms the RB-LWF and LS in terms of NMSE. We find that the NMSE increases with f_D in both RB-LWF and RB-LWS. This occurs because the filtering/smoothing process produces an averaging effect in the time domain, which is less attractive at higher f_D ; the higher the f_D , the more frequent the channel temporal variations. Consequently, as the relative dt widens, the equivalent averaging samples decrease. Moreover, at higher f_D and lower E_b/N_0 , NMSEs of the RB-LWF and RB-LWS differ more widely. In this situation, the effect of smoothing in RB-LWS dominates the effect of filtering in RB-LWF for two reasons:

CHAPTER 4. JOINT ESTIMATION



(a)



(b)

Figure 4.10: NMSE of channel estimator with $f_D = 28$ Hz (top) and 140 Hz (bottom) in case of a single unknown hyperparameter.

a) as f_D increases, the averaging effect in the time domain becomes less desirable, and b) as E_b/N_0 reduces, the accuracy of the channel estimator becomes more compromised.

4.5.1.4 Bit Error Rate (BER) of Data Subcarriers

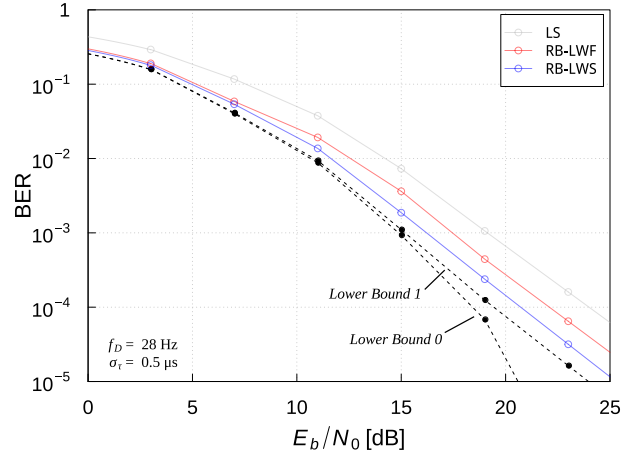
Figure 4.11 plots the BER performances of the data subcarriers for $f_D = 28$ and 140 Hz. The basic assumptions are given in Table 4.3. The BER performances of two lower bounds (Lower Bounds 0 and 1) are plotted. For Lower Bound 0, true channel values are applied to both pilot and data subcarriers; for Lower Bound 1, the true and interpolated channel values are applied to the pilot and data subcarriers, respectively.

4.5. NUMERICAL RESULTS AND DISCUSSION

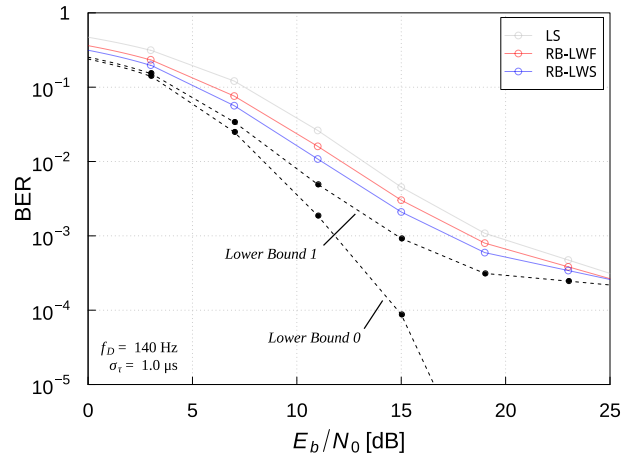
Table 4.3: Data subcarrier assumptions

Modulation	QPSK
Interpolation of channel estimator	• Frequency domain: Linear interpolation • Time domain: The nearest neighbor
Equalization	Minimum mean squared error (MMSE) [4.37]
Channel coding	• Convolutional codes – Rate: 1/2 – Constraint length: 7
Channel decoding	Soft-decision Viterbi algorithm
Interleaving	• Random interleaver – Unit: Symbol-by-symbol – Depth: 5 Resource blocks [4.15], which corresponds to 400 symbols

CHAPTER 4. JOINT ESTIMATION



(a)



(b)

Figure 4.11: BER of data subcarriers with $f_D = 28$ Hz (top) and 140 Hz (bottom) in case of a single unknown hyperparameter.

4.5. NUMERICAL RESULTS AND DISCUSSION

According to the NMSE results, the RB-LWS outperforms the RB-LWF and LS in terms of BER. Note that as f_D increases, the BER increases in both RB-LWF and RB-LWS. As explained in Section 4.5.1.3, the averaging effect in the time domain is less attractive at higher f_D . Moreover, at higher f_D and larger σ_τ , the BER floor appears earlier because the time and frequency selectivity of the channel increase. Therefore, the interpolated channel values of the data subcarriers diverge from their true values even at high E_b/N_0 . According to Figure 4.11, the coding gain improvements of RB-LWS over RB-LWF and LS are approximately 1 and 3 dB at most, respectively. Note that the throughput efficiency in terms of channel capacity is further considered in Appendix F; this trend is consistent with that of the BER results.

4.5.2 The Case of Two Unknown Hyperparameters (Velocity and Delay Spread)

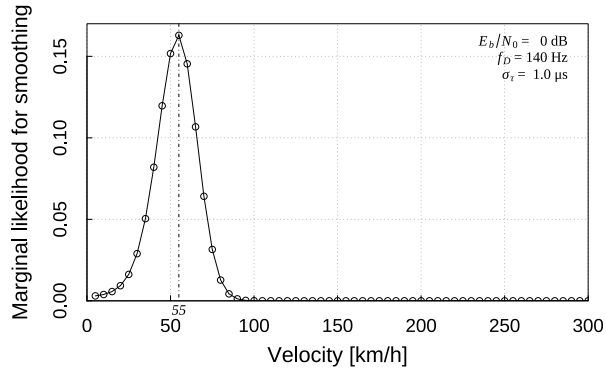
4.5.2.1 Number of Particles and Prior Distribution of Hyperparameters

When estimating multiple hyperparameters, the increase in complexity requires careful consideration as both RB-LWF and RB-LWS are stochastic methods. Therefore, in this subsection, the number of particles is first increased from 10 to 100 to retain the particle diversity. Next, the precision of the prior distribution is improved by Kalman smoothing of the initial 5 slots. Specifically, the prior distribution is a two-dimensional truncated normal distribution, with lower and upper bounds set to the values in Table 4.1, and means and variances determined through the marginal likelihood of the initial 5-slot smoothing. This marginal likelihood is derived as follows:

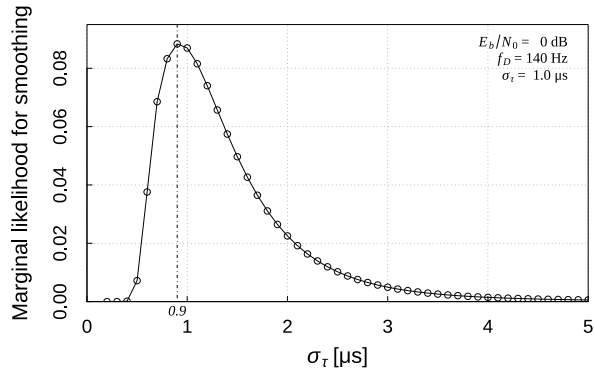
- Velocity: The Kalman smoother is applied with the delay spread set to 0 μs and the velocities set to 5 km/h interval values.
- Delay spread: The Kalman smoother is reapplied with the velocity set to the previous derived MAP estimator and the delay spread set to 0.1 μs interval values.

This prior distribution refinement is confirmed in simulations with $E_b/N_0 = 0$ dB, $f_D = 140$ Hz, and $\sigma_\tau = 1.0$ μs . Figure 4.12 plots the marginal likelihoods of the velocity and delay spread. The mean and standard deviation of the prior velocity distribution are set to the MAP estimator 55 km/h and 10 km/h (covering 67% of the likelihood), respectively. In the prior distribution of the delay spread, the mean is set to the MAP estimator (0.9 μs) and the standard deviation is 0.5 μs (covering 65% of the likelihood).

CHAPTER 4. JOINT ESTIMATION



(a)



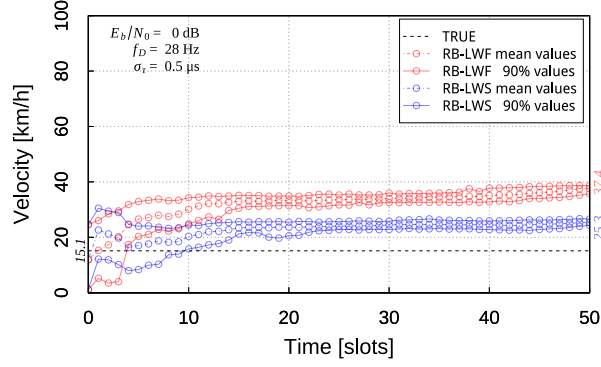
(b)

Figure 4.12: Marginal likelihood of velocity (top) and delay spread (bottom) with initial 5-slot smoothing.

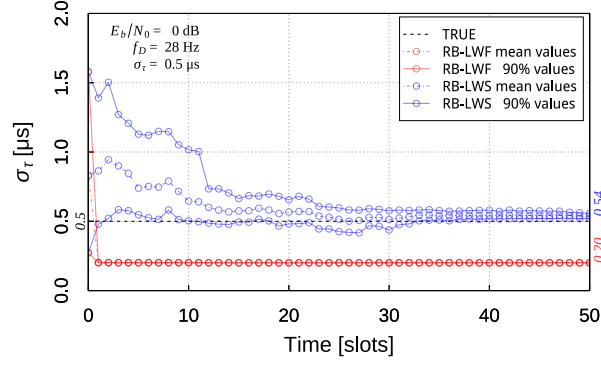
4.5.2.2 Example Result

Figure 4.13 shows the simulation results for $E_b/N_0 = 0$ dB, $f_D = 28$ Hz, and $\sigma_\tau = 0.5$ μ s.

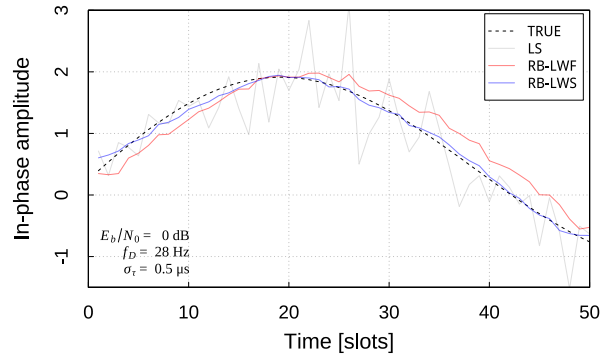
4.5. NUMERICAL RESULTS AND DISCUSSION



(a)



(b)



(c) The Y-axis denotes the in-phase amplitude of the narrow-band channel gain at pilot subcarrier index $i = 34$ ($i = 0, \dots, 49$).

Figure 4.13: Simulated velocity (top), delay spread (center), and channel estimation (bottom) in case of two unknown hyperparameters.

CHAPTER 4. JOINT ESTIMATION

The velocity estimation in RB-LWF is improved compared with Figure 4.9(a) according to its prior distribution refinement. The delay-spread estimation fails in RB-LWF, because most of the diversity in this hyperparameter is rapidly reduced during the first filtering slot and it cannot be recovered by the LWF mechanism. Joint estimation of channel and multiple hyperparameters is a difficult task and failures sometimes happen. As seen later, inadequate hyperparameter estimation strongly degrades the channel estimation. This situation could be improved by a straightforward approach (more particles), but this increases the complexity and memory requirements. To overcome this dilemma, we note that RB-LWS slightly increases the frequency of likelihood regularization and regularizes the likelihood at the first slot. By setting the regularization timing in Algorithm 4.1 to “ $t = 1$ or $(t \bmod 5) = 0$,” we can improve the resampling accuracy in the first slot and retain the particle diversity in the delay spread. With this regularization timing, RB-LWS adequately converges the delay spread and velocity toward their true values, without requiring more particles. In the channel estimation (bottom panel of Figure 4.13), RB-LWS is more accurate than RB-LWF. Compared with Figure 4.9(b), RB-LWF significantly degrades estimation accuracy due to the estimation failure of delay spread, whereas RB-LWS can maintain a comparable good performance.

4.5.2.3 NMSE of Channel Estimator

Figure 4.14 plots the NMSE performances of the channel estimation with $f_D = 28$ and 140 Hz. The hyperparameter is jointly estimated before 50 slots and thereafter converges to a constant. The NMSE is lower in RB-LWS than in RB-LWF and LS. Similarly to Section 4.5.1.3, we observe that at higher f_D , the NMSE increases in both RB-LWF and RB-LWS; moreover, at higher f_D and lower E_b/N_0 , the NMSE differs more widely between the two methods. However, the performance of RB-LWF significantly degrades and is sometimes worse than that of LS, because of delay spread estimation failure. As previously mentioned, joint channel and multiple hyperparameter estimation is much more difficult than single hyperparameter case and cannot always be accomplished by RB-LWF even at high E_b/N_0 . Conversely, RB-LWS adequately converges multiple hyperparameters with a channel estimation accuracy comparable to that of Section 4.5.1.3.

4.5. NUMERICAL RESULTS AND DISCUSSION

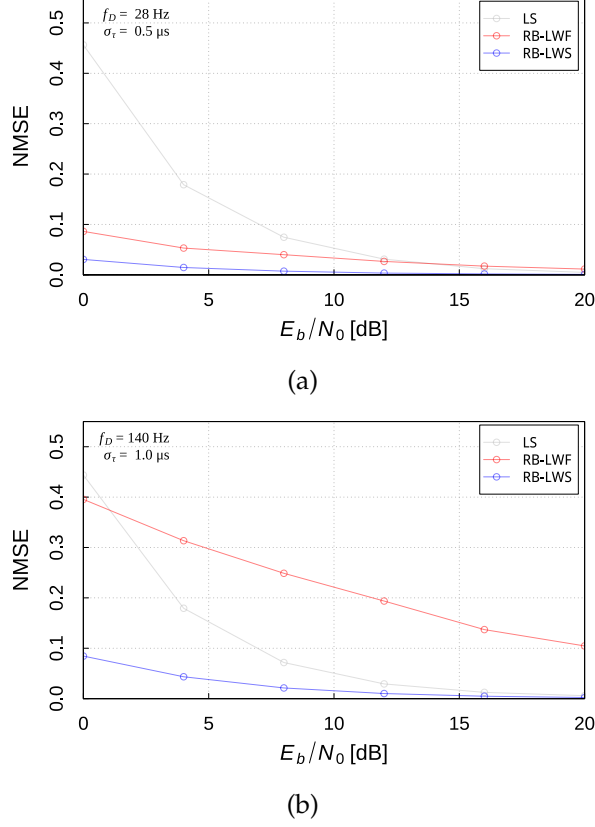
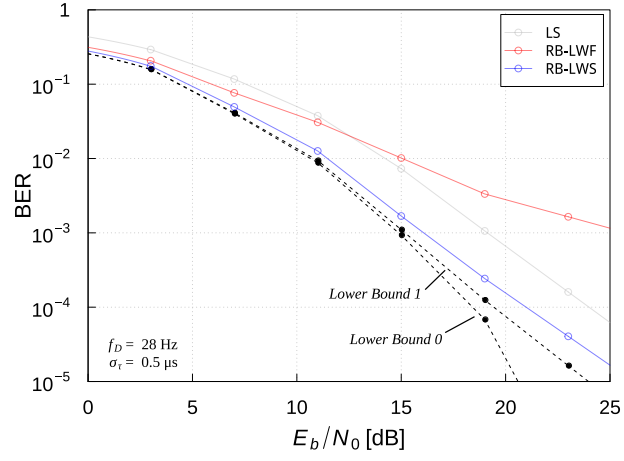


Figure 4.14: NMSE of channel estimator with $f_D = 28$ Hz (top) and 140 Hz (bottom) in case of two unknown hyperparameters.

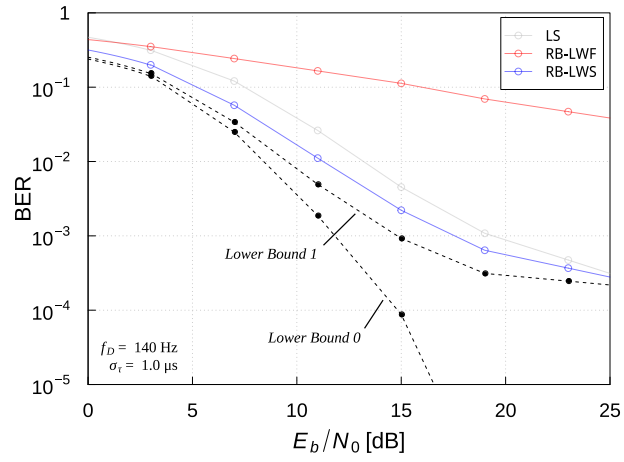
4.5.2.4 BER of Data Subcarriers

Figure 4.15 plots the BER performances of the data subcarriers for $f_D = 28$ and 140 Hz. The basic assumptions are given in Table 4.3. Again, the BER performances of Lower Bounds 0 and 1 are plotted. According to the NMSE results, RB-LWS improves the BER over RB-LWF and LS. Similar to Section 4.5.1.4, the higher the f_D , the higher the BER of both RB-LWF and RB-LWS; moreover, the BER floor appears earlier as f_D and σ_τ increase. Different from Section 4.5.1.4 but identically to Section 4.5.2.3, the performance of RB-LWF can be worse than that of LS. Again, this degradation is caused by delay spread estimation failure in RB-LWF. According to Figure 4.15, the coding gain improvements of RB-LWS over RB-LWF and LS approximate 10 and 3 dB at most, respectively. Note that the throughput efficiency in terms of channel capacity is further considered in Appendix F; this trend is consistent with that of the BER results.

CHAPTER 4. JOINT ESTIMATION



(a)



(b)

Figure 4.15: BER of data subcarriers with $f_D = 28$ Hz (top) and 140 Hz (bottom) in case of two unknown hyperparameters.

4.6 Summary

This study investigates a real-time joint channel and hyperparameter estimation method for OFDM mobile communications. To estimate the channel frequency response of the pilot subcarrier and its fixed hyperparameters (such as channel statistics), the method adopts LWF, a modern and effective technique based on the state-space model and sequential Monte Carlo method. For the first time, to our knowledge, we demonstrate biased hyperparameter estimation using the conventional LWF and show that bias results from poor likelihood estimation caused by overfitting in noisy environments. This problem, which cannot be overcome by conventional smoothing techniques, is here corrected by a novel method that modifies the conventional LWF and regularizes the likelihood using a Kalman smoother. The effectiveness of the proposed method is confirmed in numerical analyses assuming CRSs in 3GPP LTE as pilot subcarriers. The following conclusions were drawn.

First, when the hyperparameter f_D is unknown, the NMSE and BER of conventional RB-LWF are acceptable, but the proposed RB-LWS outperforms the RB-LWF and LS, and achieves good performance near the lower bound. Regarding the BER performance, the proposed method improves the coding gain to RB-LWF and LS by approximately 1 and 3 dB at most, respectively.

Second, when two hyperparameters (f_D and σ_τ) are unknown, LS sometimes outperforms the conventional RB-LWF, because of hyperparameter estimation failure in the latter method. The proposed RB-LWS overcomes this problem by avoiding such failures. We emphasize that joint channel and multiple hyperparameters estimation is much more difficult than single hyperparameter case. The proposed RB-LWS outperforms the RB-LWF and LS, and achieves good performance near the lower bound. Regarding the BER performance, the proposed method improves the coding gain to RB-LWF and LS by approximately 10 and 3 dB at most, respectively.

Clearly, the proposed method offers higher channel and hyperparameter estimation accuracy than the conventional RB-LWF.

Our method can be improved further. In principle, our proposal could be adapted to particle learning [4.38], which tracks sufficient statistics of the target distribution and is almost superior to LWF. Such an extension and investigation is planned for future studies.

CHAPTER 4. JOINT ESTIMATION

Bibliography

- [4.1] Y. R. Zheng and C. Xiao, "Mobile speed estimation for broadband wireless communications over Rician fading channels," *IEEE Trans. Wireless Communications*, vol. 8, no. 1, pp. 1–5, Jan. 2009.
- [4.2] W. Zhou and W. H. Lam, "A novel method of Doppler shift estimation for OFDM systems," in *Proc. IEEE Military Communications Conference (MILCOM)*, San Diego, U.S.A., Nov. 2008, pp. 1–7.
- [4.3] J. Hagiwara, "An OFDM channel estimation method based on a state-space model that appropriately considers frequency correlation," *IEICE Trans. Fundamentals*, vol. E98-A, no. 2, pp. 537–548, Feb. 2015.
- [4.4] Y. Huang, P. M. Djurić, and J. Zhang, "Joint velocity estimation and symbol detection in non-stationary fading channels by particle filtering," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 4, San Francisco, U.S.A., Dec. 2003, pp. 2269–2273.
- [4.5] C. Nemeth, P. Fearnhead, and L. Mihaylova, "Sequential Monte Carlo methods for state and parameter estimation in abruptly changing environments," *IEEE Trans. Signal Processing*, vol. 62, no. 5, pp. 1245–1255, Mar. 2014.
- [4.6] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo Methods in Practice*. Springer, 2001, pp. 197–223.
- [4.7] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4.8] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. John Wiley & Sons, 2003.
- [4.9] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Communications*, vol. 3, no. 1, pp. 315–325, Jan. 2004.
- [4.10] D. Larsson, "Analysis of channel estimation methods for OFDMA," Master's thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, Dec. 2006.
- [4.11] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 2, pp. 18–48, Apr. 2007.

BIBLIOGRAPHY

- [4.12] W. C. Jakes, Jr., *Microwave Mobile Communications*. Wiley, New York, 1974, [reprinted by IEEE Press].
- [4.13] H. Iwai, *Radio Propagation in Mobile Communications — Fundamental Knowledge for Simulations Analysis of Wireless Communications —*. Corona Publishing, 2012, [in Japanese].
- [4.14] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering, Transactions of the ASME*, vol. 82 (Series D), no. 1, pp. 35–45, Mar. 1960.
- [4.15] “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation,” The 3rd Generation Partnership Project (3GPP), TS 36.211, Dec. 2009.
- [4.16] L. Wang, G. Libert, and P. Manneback, “Kalman filter algorithm based on singular value decomposition,” in *Proc. 31st IEEE Conference on Decision and Control (CDC)*, Tucson, U.S.A., Dec. 1992, pp. 1224–1229.
- [4.17] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, “Interacting multiple model methods in target tracking: A survey,” *IEEE Trans. Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 103–123, Jan. 1998.
- [4.18] C. C. Logothetis, W. Yao, and A. Logothetis, “Joint channel and velocity estimation,” in *Proc. IEEE International Conference on Networking, Sensing and Control (ICNSC)*, London, U.K., Apr. 2007, pp. 781–786.
- [4.19] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.
- [4.20] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation,” *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2, pp. 107–113, Apr. 1993.
- [4.21] G. Kitagawa, “Monte Carlo filter and smoother for non-Gaussian nonlinear state space models,” *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, Mar. 1996.
- [4.22] T. Katayama, *Nonlinear Kalman Filter*. Asakura Publishing, 2011, [in Japanese].
- [4.23] G. Kitagawa, “A self-organizing state-space model,” *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1203–1215, Sep. 1998.

CHAPTER 4. JOINT ESTIMATION

- [4.24] M. West, “Approximating posterior distributions by mixture,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 55, no. 2, pp. 409–422, 1993.
- [4.25] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic Linear Models with R*. Springer, 2009.
- [4.26] M. P. Rios and H. F. Lopes, “The extended Liu and West filter: Parameter learning in Markov switching stochastic volatility models,” in *State-Space Models: Applications in Economics and Finance*. Springer, 2013, pp. 23–61.
- [4.27] T. Higuchi, G. Ueno, S. Nakano, K. Nakamura, and R. Yoshida, *Introduction to Data Assimilation*, T. Higuchi, Ed. Asakura Publishing, 2011, [in Japanese].
- [4.28] G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*. Springer-Verlag New York, 1996, vol. 116.
- [4.29] “Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN),” The 3rd Generation Partnership Project (3GPP), TR 25.913, Mar. 2007.
- [4.30] O. Cappé, S. J. Godsill, and E. Moulines, “An overview of existing methods and recent advances in sequential Monte Carlo,” *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, May 2007.
- [4.31] S. Särkkä, *Bayesian Filtering and Smoothing*, ser. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2013.
- [4.32] H. E. Rauch, F. Tung, and C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, Aug. 1965.
- [4.33] J. Hagiwara, T. Nishimura, T. Ohgane, and Y. Ogawa, “Comparison of Kalman smoother to Wiener smoother in practical OFDM channel estimation,” in *Proc. 21st Asia-Pacific Conference on Communications (APCC)*, Kyoto, Japan, Oct. 2015, pp. 454–459.
- [4.34] T. C. Clapp and S. J. Godsill, “Fixed-lag smoothing using sequential importance sampling,” in *Bayesian Statistics*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. Smith, Eds. Oxford University Press, 1999, vol. VI, pp. 743–752.
- [4.35] N. G. Polson, J. R. Stroud, and P. Müller, “Practical filtering with sequential parameter learning,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 2, pp. 413–428, 2008.

BIBLIOGRAPHY

- [4.36] R. Prado and M. West, *Time Series: Modeling, Computation, and Inference*. CRC Press, 2010.
- [4.37] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [4.38] C. Carvalho, M. S. Johannes, H. F. Lopes, and N. G. Polson, “Particle learning and smoothing,” *Statistical Science*, vol. 25, no. 1, pp. 88–106, Feb. 2010.

Chapter 5

Detecting Abrupt Changes

5.1 Preliminary Remarks

As previously confirmed in Chapters 3 and 4, channel statistics such as velocity and delay spread are required for accurate channel estimation. These are rarely known, but can be estimated through various methods. For instance, a real-time joint channel and its statistics estimation method were proposed in [5.1]. Regarding temporal change in the channel statistics, [5.1] supposes a quickly repeated process to track the changes. Assuming that the joint estimation converges within 10 ms at most, for example, the estimation process is repeated every 10 ms. When temporal changes in the channel statistics occur at a slower rate than the repeat interval, this approach is sufficiently valid. However, the statistics may abruptly change faster than the repeat interval. For example, a behavior such as turning the street may cause a switch from line-of-sight to non-line-of-sight and a sudden increase in the delay spread. The quickly repeated process approach mentioned above cannot always follow such abrupt changes and may degrade the estimation accuracy. Universal modeling for the temporal changes in channel statistics is inherently difficult as discussed in [5.1]; however, detecting only the abrupt changes will enable the resetting of the channel estimation process and triggering of an immediate follow up. This chapter focuses on detecting abrupt changes in the channel statistics to avoid degradation in the channel estimation accuracy. Regarding related studies, [5.2] and [5.3] proposed an empirical method that detects abrupt changes according to time differences in the estimated channel values. This thesis proposes a novel detection method that relies on the state-space model and makes use of the Kalman filter. Through computer simulation using the delay spread as the channel statistics, we confirm that the proposed method adequately detects abrupt changes based on statistical theory.

The remainder of this chapter is structured as follows. Section 5.2

CHAPTER 5. DETECTING ABRUPT CHANGES

provides background on the study (problem formulation). Section 5.3 introduces the theory and proposes the novel method for abrupt change detection. Section 5.4 analyzes the numerical results, and Section 5.5 summarizes the discussion.

5.2 Background

5.2.1 Problem Formulation

OFDM channel can be estimated by various models [5.4–5.8]. This study adopts the state-space model of [5.5], but excludes the extended formulation for simplicity. The observation and state equations in this model are expressed as [5.5]:

$$\mathbf{y}(t) = \mathbf{F}(t) \boldsymbol{\theta}(t) + \mathbf{v}(t), \quad (5.1)$$

$$\boldsymbol{\theta}(t) = \mathbf{G}(t) \boldsymbol{\theta}(t-1) + \mathbf{w}(t), \quad (5.2)$$

where (for the pilot subcarriers at the t -th symbol) $\mathbf{y}(t)$ and $\boldsymbol{\theta}(t) = [\theta_0(t), \dots, \theta_i(t), \dots, \theta_{I-1}(t)]^T$ denote the $I \times 1$ observation vector and the $I \times 1$ state vector, respectively, and $\mathbf{F}(t)$ and $\mathbf{G}(t)$ denote the $I \times I$ observation matrix and $I \times I$ transition matrix, respectively. For the pilot subcarriers at the t -th symbol, the covariance matrix in the $I \times 1$ observation noise vector $\mathbf{v}(t) \sim \mathcal{CN}(\mathbf{0}_I, \mathbf{V})$ is set to $\mathbf{V} = \text{diag}(V_0, \dots, V_i, \dots, V_{I-1})$, while the covariance matrix in the $I \times 1$ state noise vector $\mathbf{w}(t) \sim \mathcal{CN}(\mathbf{0}_I, \mathbf{W})$ is set to $\mathbf{W} = \text{diag}(W_0, \dots, W_i, \dots, W_{I-1})$. $\mathbf{y}(t)$ and $\mathbf{v}(t)$ in Equation (5.1) are identical with them in Equation (2.4).

The terms in the observation equation are specified below:

$$\tilde{\mathbf{h}}(t) = \boldsymbol{\Sigma} \boldsymbol{\theta}(t), \quad (5.3)$$

$$\mathbf{F}(t) = \mathbf{S}(t) \boldsymbol{\Sigma}, \quad (5.4)$$

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Omega}_f \right)^{1/2}, \quad (5.5)$$

$$\left(\boldsymbol{\Omega}_f \right)_{\text{row, col}} = \rho(\Delta t = 0, \Delta f = (\text{col} - \text{row})df), \quad (5.6)$$

$$\rho(\Delta t = 0, \Delta f) = \frac{1 + j2\pi\Delta f\sigma_\tau}{1 + (2\pi\Delta f\sigma_\tau)^2}, \quad (5.7)$$

where $\rho(\Delta t, \Delta f)$ denotes the time versus frequency correlation coefficient of the narrow-band channel gain, Δt and Δf denote the time and frequency differences, respectively, and σ_τ is the channel delay spread.

The terms in the state equation are given as

$$\mathbf{G}(t) = \mathbf{E}_I, \quad (5.8)$$

$$\mathbf{W} = 2l^2(1 - \rho(\Delta t = 1dt, \Delta f = 0))\mathbf{E}_I, \quad (5.9)$$

$$\rho(\Delta t, \Delta f = 0) = J_0(2\pi f_D \Delta t), \quad (5.10)$$

5.3. THEORY AND METHOD

l^2 denotes the average power of the narrow-band channel gain of each pilot subcarrier, J_0 represents a zero-order Bessel function of the first kind [5.9], and f_D is the maximum Doppler frequency of the channel.

In this model, the narrow-band channel gains correspond to parameter $\theta(t)$ and the channel statistics correspond to hyperparameters σ^2 , σ_τ , l^2 , and f_D . For simplicity, we assume that all hyperparameters are known and time invariant except for the delay spread (delay spread may change abruptly). Note that the delay spread as abrupt changeable channel statistics is just one example and essentially the proposed method is applicable to abrupt changes in other channel statistics.

5.3 Theory and Method

When the hyperparameters in Section 5.2.1 are specified, the linear Gaussian state-space model can be solved using a Kalman filter [5.10]. The recursion process of the Kalman filter at the t -th symbol is presented as Algorithm 2.1. In the Kalman filtering process, one-step-ahead prediction is sequentially corrected by received data and one-step-ahead prediction error $\mathbf{e}(t) = \mathbf{y}(t) - \mathbf{f}(t)$ is derived. Term $\mathbf{e}(t)$ is also referred to as *innovations* [5.11] and has been used for model diagnosis [5.12] or anomaly detection [5.13] because it has an independent identical complex normal distribution. A similar approach is also applicable to detecting abrupt changes in channel statistics and this thesis proposes a method according to $\mathbf{e}(t)$. In more detail, when the following score value in Equation (5.12) derived from $\mathbf{e}(t)$ exceeds a certain threshold, an abrupt change in the channel statistics is detected:

$$\tilde{\mathbf{e}}(t) = [\tilde{e}_0(t), \dots, \tilde{e}_i(t), \dots, \tilde{e}_{I-1}(t)]^T = \left(\mathbf{Q}(t)^{1/2}\right)^{-1} \mathbf{e}(t), \quad (5.11)$$

$$\text{Score}(t) = \sqrt{1/I \sum_{i=0}^{I-1} |\tilde{e}_i(t)|^2}, \quad (5.12)$$

where $\mathbf{Q}(t)^{1/2}$ is derived using singular value decomposition (SVD) in this study. $\mathbf{Q}(t)$ means the covariance of innovations, and the detection score in Equation (5.12) corresponds to the root mean square of the standardized and decorrelated one-step-ahead prediction error; the value is normally approximate one and increases significantly when an abrupt change occurs.

5.4 Numerical Results and Discussion

5.4.1 Simulation Configuration

To verify the effectiveness of the proposed method, a computer simulation is performed. In the simulations, we used 3GPP LTE specifications

CHAPTER 5. DETECTING ABRUPT CHANGES

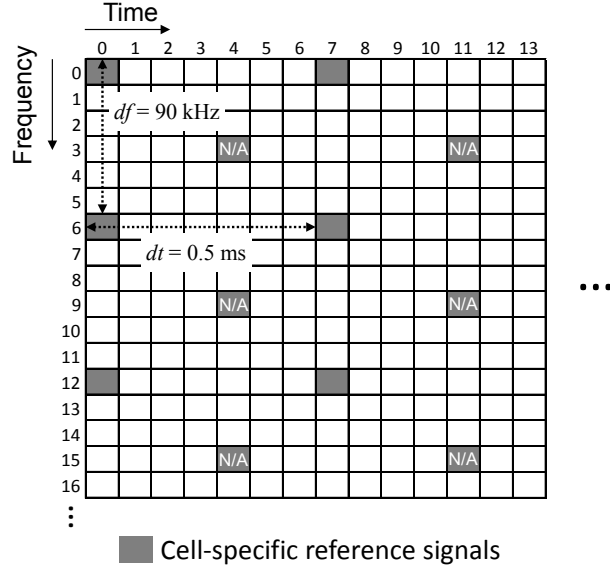


Figure 5.1: Allocation of scattered pilot subcarriers.

with a 5 MHz bandwidth. The frequency interval between each subcarrier was 15 kHz, and the symbol length (excluding the GI) was 66.7 μ s. The fast Fourier transform (FFT) size was 512, wherein 300, 1, and 211 were the sizes for the active, direct current (DC), and guard subcarriers, respectively. The guard band was $5 - (300+1) \times 0.015 = 0.485$ MHz. As the pilot subcarriers, we adopted cell-specific reference signals (CRSs) [5.14], allocated as shown in Figure 5.1. As shown in the figure, the CRSs at timeslots 4, 11, $\dots$ were excluded for simplicity. We set $dt = 0.5$ ms, $df = 90$ kHz and modulated the specific pseudorandom sequence for the CRSs by quadrature phase-shift keying (QPSK). There were 250 data subcarriers and 50 pilot subcarriers for one symbol duration. An isolated cell and single-user environment without inter-cell interference and multiple access were assumed. The other basic assumptions of the computer simulation are summarized in Table 5.1.

Regarding performance, the normalized mean squared error ($\text{NMSE} = \sum_{i=0}^{I-1} E[|\tilde{h}_i(t) - \hat{\tilde{h}}_i(t)|^2] / \sum_{i=0}^{I-1} E[|\tilde{h}_i(t)|^2]$) of the channel estimator and bit error rate (BER) of the data subcarriers are examined in the following. The data subcarrier assumptions are given in Table 5.2. The BER lower bound, where the respective true and its interpolated channel values are applied to the pilot and data subcarriers, is also derived for reference. Note that the numerical results using the Kalman smoother as well as the Kalman filter are confirmed. For the Kalman smoother, the Rauch–Tung–Striebel (RTS) algorithm [5.16] with the time lag of 20 slots [5.17] is applied (one slot corresponds to one dt).

5.4. NUMERICAL RESULTS AND DISCUSSION

Table 5.1: Basic computer simulation assumptions

Carrier frequency	2 GHz
Maximum Doppler frequency: f_D	14 Hz (i.e., 7.65 km/h)
Delay profile	• 24-path exponential – Delay spread: 0.2 μs to 5.0 μs (Extreme) 0.5 μs to 1.0 μs (Moderate) – Maximum delay: 4.7 μs (within GI)
Numerical computation of Kalman filter/smoothen	A square root Kalman filter/smoothen based on SVD [5.15] is applied to suppress the degradation in numerical accuracy.

Table 5.2: Data subcarrier assumptions

Modulation	QPSK
Interpolation of channel estimator	• Frequency domain: Linear interpolation • Time domain: The nearest neighbor
Equalization	• Zero forcing (ZF) [5.18] for least-squares (LS) channel estimator $y_i(t)/s_i(t)$ • Minimum mean squared error (MMSE) [5.18] for the other channel estimators
Channel coding	• Convolutional codes – Rate: 1/2 – Constraint length: 7
Channel decoding	Soft-decision Viterbi algorithm
Interleaving	• Random interleaver – Unit: Symbol-by-symbol – Depth: 5 Resource blocks [5.14], which corresponds to 400 symbols

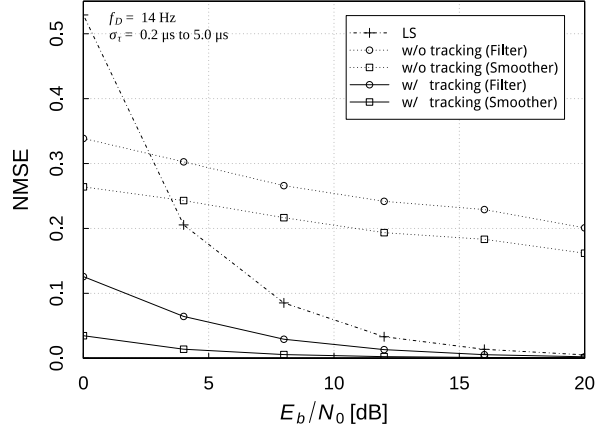
CHAPTER 5. DETECTING ABRUPT CHANGES

An abrupt change is supposed to occur at the middle point during one trial of 200 slots. The delay spread varies from $0.2\ \mu\text{s}$ to $5.0\ \mu\text{s}$ and from $0.5\ \mu\text{s}$ to $1.0\ \mu\text{s}$ for the extreme and moderate changes, respectively. In both changes, a true delay spread is applied to the Kalman filter/smoothers during the first half, while the true or false (the same as in the first half) delay spread is applied to the Kalman filter/smoothers during the last half; “w/” or “w/o” tracking the abrupt change. Note that the NMSE and BER for one trial are calculated over all 200 slots.

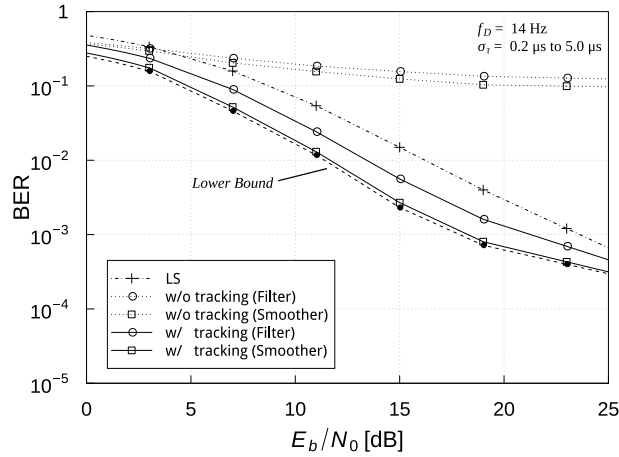
5.4.2 Extreme Change in Delay Spread

First, we verify how the tracking of an abrupt change impacts on the channel estimation performance. Figure 5.2 shows that the NMSE and BER degrade severely without tracking.

5.4. NUMERICAL RESULTS AND DISCUSSION



(a) NMSE of channel estimator



(b) BER of data subcarriers

Figure 5.2: Performance for extreme change in delay spread.

The Kalman filter/smoothing without tracking sometimes degrades the performance to a level lower than that for LS, while the Kalman filter/smoothing with tracking achieves good performance near the lower bound. These results suggest the importance of the tracking.

Next, in order to confirm the effectiveness of the proposed detection method, the score value in Equation (5.12) is verified when an abrupt change is not tracked. Figure 5.3 shows that the score value is approximately one during the first half, then increases significantly at an abrupt change and remains high during the last half.

CHAPTER 5. DETECTING ABRUPT CHANGES

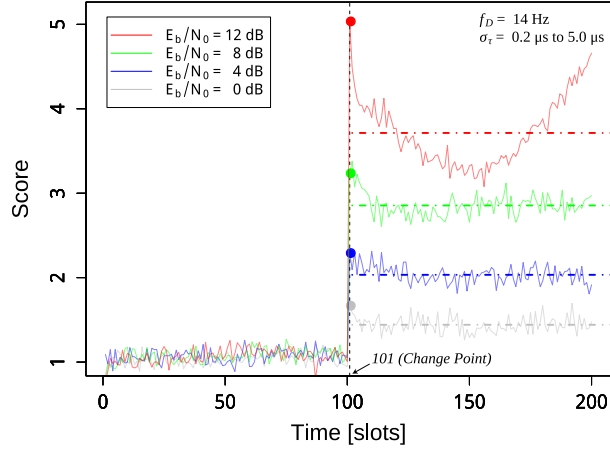


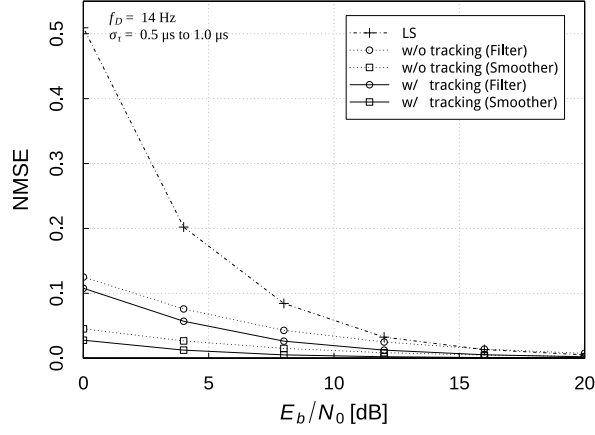
Figure 5.3: Score values for extreme change in delay spread.

Figure 5.3 also shows that the average score value during the last half (thick dotted-dashed line) varies according to E_b/N_0 ; the higher the E_b/N_0 , the higher the average level. As noise is reduced, the prediction error essentially yielded by the inconsistency in the channel statistics is more clearly revealed. According to Figure 5.3, we recognize that an abrupt change can be basically detected, while the threshold value for the detection score is a tuning matter. When the threshold value is set to 2, for example, an abrupt change can be detected at $E_b/N_0 \geq 4$ dB.

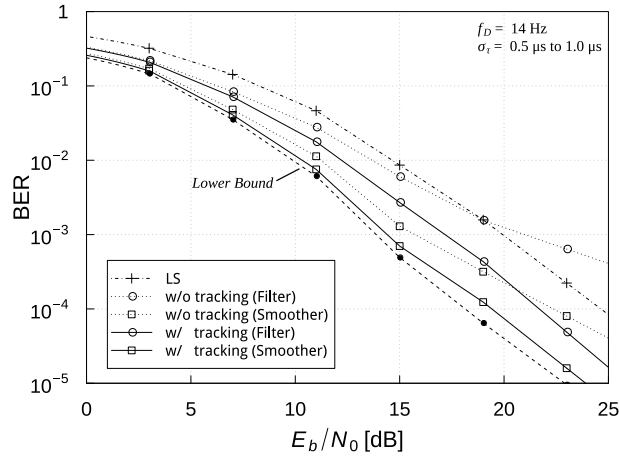
5.4.3 Moderate Change in Delay Spread

First, we verify the performance impact on the tracking of an abrupt change. Similar to Section 5.4.2, Figure 5.4 shows that both the NMSE and BER degrade without tracking.

5.4. NUMERICAL RESULTS AND DISCUSSION



(a) NMSE of channel estimator



(b) BER of data subcarriers

Figure 5.4: Performance for moderate change in delay spread.

Compared to Section 5.4.2, a lower degree of deviance between the delay spreads incurs less performance degradation.

Next, the score value in Equation (5.12) is verified when an abrupt change is not tracked. Similar to Section 5.4.2, Figure 5.5 shows that the score value increases during the last half according to E_b/N_0 .

CHAPTER 5. DETECTING ABRUPT CHANGES

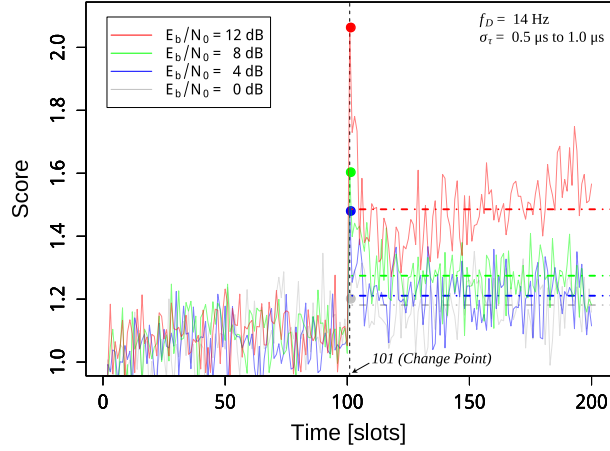


Figure 5.5: Score values for moderate change in delay spread.

Compared to Section 5.4.2, a lower degree of deviance between the delay spreads makes a lower increase in score. We also recognize that an abrupt change can be basically detected according to Figure 5.5, while the threshold value for the detection score is a tuning matter. When the threshold value is set to 2, for example, an abrupt change can be detected at $E_b/N_0 \geq 12 \text{ dB}$.

5.5 Summary

This chapter investigated a method for detecting abrupt changes in the channel statistics to maintain accurate channel estimation. The proposed novel detection method relies on the state-space model and makes use of the Kalman filter. The effectiveness of the proposed method is verified based on computer simulation using the delay spread as the channel statistics. The proposed method basically detects abrupt changes through extreme and moderate changes in delay spread, while the detection threshold is a tuning matter. We also confirmed that the greater the deviation between delay spreads and the higher the E_b/N_0 value, the easier the detection. In future studies, we plan to clarify the applicable conditions and improve the detection precision.

Bibliography

- [5.1] J. Hagiwara, "Real-time joint channel and hyperparameter estimation using sequential Monte Carlo methods for OFDM mobile communications," *IEICE Trans. Communications*, vol. E99-B, no. 8, pp. 1655–1668, Aug. 2016.
- [5.2] H. Wang and T. Li, "Signal detection and abrupt-channel tracking for MC-CDMA," in *Proc. 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, May 2006, pp. 21–24.
- [5.3] —, "Channel tracking and signal detection for MC-CDMA in a time-varying environment," *IEEE Trans. Vehicular Technology*, vol. 56, no. 6, pp. 3613–3620, Nov. 2007.
- [5.4] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. John Wiley & Sons, 2003.
- [5.5] J. Hagiwara, "An OFDM channel estimation method based on a state-space model that appropriately considers frequency correlation," *IEICE Trans. Fundamentals*, vol. E98-A, no. 2, pp. 537–548, Feb. 2015.
- [5.6] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Communications*, vol. 3, no. 1, pp. 315–325, Jan. 2004.
- [5.7] D. Larsson, "Analysis of channel estimation methods for OFDMA," Master's thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, Dec. 2006.
- [5.8] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 2, pp. 18–48, Apr. 2007.
- [5.9] W. C. Jakes, Jr., *Microwave Mobile Communications*. Wiley, New York, 1974, [reprinted by IEEE Press].
- [5.10] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering, Transactions of the ASME*, vol. 82 (Series D), no. 1, pp. 35–45, Mar. 1960.
- [5.11] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall Upper Saddle River, NJ, 2000.

CHAPTER 5. DETECTING ABRUPT CHANGES

- [5.12] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*, 2nd ed. Oxford University Press, 2012.
- [5.13] T. Idé, *Introduction to Anomaly Detection using Machine Learning — A Practical Guide with R*. Corona Publishing, 2015, [in Japanese].
- [5.14] “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation,” The 3rd Generation Partnership Project (3GPP), TS 36.211, Dec. 2009.
- [5.15] L. Wang, G. Libert, and P. Manneback, “Kalman filter algorithm based on singular value decomposition,” in *Proc. 31st IEEE Conference on Decision and Control (CDC)*, Tucson, U.S.A., Dec. 1992, pp. 1224–1229.
- [5.16] H. E. Rauch, F. Tung, and C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, Aug. 1965.
- [5.17] J. Hagiwara, T. Nishimura, T. Ohgane, and Y. Ogawa, “Comparison of Kalman smoother to Wiener smoother in practical OFDM channel estimation,” in *Proc. 21st Asia-Pacific Conference on Communications (APCC)*, Kyoto, Japan, Oct. 2015, pp. 454–459.
- [5.18] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

Chapter 6

Wiener and Kalman Filters: Practical Comparison

6.1 Preliminary Remarks

In Chapters 3, 4, and 5, we proposed a sequential OFDM channel estimation method based on the Kalman filter [6.1]. On the other hand, the batch OFDM channel estimation method is also applicable to the same problem when fading is stationary and latency is tolerated. The Wiener filter¹[6.2] is the most popular batch method. Literature such as [6.3] sometimes describes the Wiener filter as theoretically equivalent to the Kalman filter under some conditions. However, it is considered that there is no literature that discusses a detailed comparison in a practical environment. Therefore, this chapter compares both methods in practical OFDM channel estimation and summarizes the resulting knowledge. For comparison, we assume that the fading channel obeys a stationary stochastic process in this chapter. This implies that hyperparameters such as the Doppler frequency, delay spread, and E_b/N_0 are time-invariant during the study period. Furthermore, true hyperparameters are assumed to be known for simplicity.

The rest of this chapter is structured as follows. Section 6.2 describes the background of the study (problem formulation). Section 6.3 describes the conditions for fair comparison. Section 6.4 presents numerical analysis results for the comparison. Section 6.5 summarizes the discussion.

¹This is also called a linear minimum mean squared error (LMMSE) filter.

6.2 Background

6.2.1 Problem Formulation

For the OFDM channel estimation, various formulations of the Kalman and Wiener filters have been proposed [6.4–6.8]. This thesis supposes the following formulations that are considered to be the most basic.

6.2.1.1 Kalman Filter

The formulation of the Kalman filter is based on the state-space model according to [6.5], but the extended formulation for complexity reduction is not considered. Observation and state equations for the state-space model are expressed as

$$\mathbf{y}(t) = \mathbf{F}(t) \boldsymbol{\theta}(t) + \mathbf{v}(t), \quad (6.1)$$

$$\boldsymbol{\theta}(t) = \mathbf{G}(t) \boldsymbol{\theta}(t-1) + \mathbf{w}(t), \quad (6.2)$$

where $\mathbf{y}(t)$ denotes the I -by-1 observation vector for pilot subcarriers at the t -th symbol (as in Equation (2.4)), $\mathbf{F}(t)$ denotes the I -by- I observation matrix for pilot subcarriers at the t -th symbol, $\boldsymbol{\theta}(t) = [\theta_0(t), \dots, \theta_i(t), \dots, \theta_{I-1}(t)]^T$ denotes the I -by-1 state vector for pilot subcarriers at the t -th symbol, and $\mathbf{G}(t)$ denotes the I -by- I transition matrix for pilot subcarriers at the t -th symbol. $\mathbf{v}(t) \sim \mathcal{CN}(\mathbf{0}_I, \mathbf{V})$ denotes the I -by-1 observation noise vector for pilot subcarriers at the t -th symbol; its covariance matrix is set to $\mathbf{V} = \text{diag}(V_0, \dots, V_i, \dots, V_{I-1})$ (as in Equation (2.4)). $\mathbf{w}(t) \sim \mathcal{CN}(\mathbf{0}_I, \mathbf{W})$ denotes the I -by-1 state noise vector for pilot subcarriers at the t -th symbol; its covariance matrix is set to $\mathbf{W} = \text{diag}(W_0, \dots, W_i, \dots, W_{I-1})$.

Detailed definitions regarding the observation equation are given below:

$$\tilde{\mathbf{h}}(t) = \boldsymbol{\Sigma} \boldsymbol{\theta}(t), \quad (6.3)$$

$$\mathbf{F}(t) = \mathbf{S}(t) \boldsymbol{\Sigma}, \quad (6.4)$$

$$\boldsymbol{\Sigma} = (\boldsymbol{\Omega}_f)^{1/2}, \quad (6.5)$$

$$\left(\boldsymbol{\Omega}_f \right)_{\text{row, col}} = \rho(\Delta t = 0, \Delta f = (\text{col} - \text{row})df), \quad (6.6)$$

$$\rho(\Delta t = 0, \Delta f) = \frac{1 + j2\pi\Delta f\sigma_\tau}{1 + (2\pi\Delta f\sigma_\tau)^2}, \quad (6.7)$$

where $\rho(\Delta t, \Delta f)$ denotes the time and frequency correlation coefficient of the narrow-band channel gain, Δt denotes the time difference, Δf denotes the frequency difference, and σ_τ denotes the channel delay spread.

6.3. CONDITIONS FOR FAIR COMPARISON

Detailed definitions regarding the state equation are given below:

$$\mathbf{G}(t) = \mathbf{E}_I, \quad (6.8)$$

$$\mathbf{W} = 2l^2(1 - \rho(\Delta t = 1dt, \Delta f = 0))\mathbf{E}_I, \quad (6.9)$$

$$\rho(\Delta t, \Delta f = 0) = J_0(2\pi f_D \Delta t), \quad (6.10)$$

where l^2 denotes the average power of the narrow-band channel gain for each pilot subcarrier, J_0 represents Bessel functions of the first kind of order zero [6.9], and f_D denotes the channel maximum Doppler frequency.

Thus, the Kalman filter algorithm in Algorithm 2.1 is applied to the above linear Gaussian state-space model.

6.2.1.2 Wiener Filter

The formulation of the Wiener filter is based on [6.8, eq. (41)]:

$$\hat{\mathbf{H}} = [\hat{\mathbf{h}}(t-L)^T, \dots, \hat{\mathbf{h}}(t)^T, \dots, \hat{\mathbf{h}}(t+L)^T]^T, \quad (6.11)$$

$$= \mathbf{\Omega}_{tf} \left(\mathbf{\Omega}_{tf} + \frac{\beta}{\text{SNR}} \mathbf{E}_I \right)^{-1} \hat{\mathbf{H}}_{\text{LS}}, \quad (6.12)$$

$$\mathbf{\Omega}_{tf} = \begin{cases} \mathbf{\Omega}_t \otimes \mathbf{\Omega}_f & \text{for upper diagonal elements of } \mathbf{\Omega}_t, \\ \mathbf{\Omega}_t \otimes (\mathbf{\Omega}_f)^H & \text{otherwise,} \end{cases} \quad (6.13)$$

$$(\mathbf{\Omega}_t)_{\text{row, col}} = \rho(\Delta t = |\text{col} - \text{row}|dt, \Delta f = 0), \quad (6.14)$$

$$\beta = E[|s_i(t)|^2] E[1/|s_i(t)|^2], \quad (6.15)$$

$$\hat{\mathbf{H}}_{\text{LS}} = [\hat{\mathbf{h}}_{\text{LS}}(t-L)^T, \dots, \hat{\mathbf{h}}_{\text{LS}}(t)^T, \dots, \hat{\mathbf{h}}_{\text{LS}}(t+L)^T]^T, \quad (6.16)$$

$$\hat{\mathbf{h}}_{\text{LS}}(t) = \mathbf{S}(t)^{-1} \mathbf{y}(t), \quad (6.17)$$

where the batch time period corresponds to $2L + 1$ symbols.

6.3 Conditions for Fair Comparison

6.3.1 Type of Estimation

When the data at the t -th symbol are estimated using the data up to the k -th symbol, the estimation is generally classified into the following three types depending on the relationship between t and k :

$$\begin{cases} \text{Filtering} & \text{if } t = k, \\ \text{Prediction} & \text{if } t > k, \\ \text{Smoothing} & \text{if } t < k. \end{cases}$$

CHAPTER 6. WIENER AND KALMAN FILTERS: PRACTICAL COMPARISON

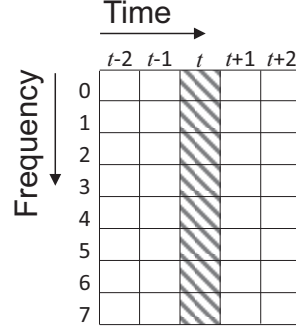


Figure 6.1: Estimated pilot signals for smoothing ($L = 2$ and $I = 8$).

This thesis supposes *smoothing* for the ease of comparison. Regarding the Wiener filter, $\hat{\mathbf{h}}(t)$ in Section 6.2.1.2 straightforwardly corresponds to a smoothing estimator. For example, when $L = 2$ and $I = 8$, estimated pilot signals for smoothing are represented by the hatched area in Figure 6.1. The batch smoothing process in this thesis is hereafter referred to as the Wiener smoother. On the other hand, adapted to the Wiener smoother, the Kalman filter requires a backward smoothing process in addition to that in Section 6.2.1.1. When the popular Rauch-Tung-Striebel (RTS) algorithm [6.10] is applied to the backward smoothing process as in Algorithm 2.2, the smoothed state (mean) in this study is expressed as

$$\mathbf{u}(t) = \mathbf{m}(t) + \mathbf{C}(t)\mathbf{R}(t+1)^{-1}[\mathbf{u}(t+1) - \mathbf{a}(t+1)]. \quad (6.18)$$

The sequential smoothing process in this thesis is hereafter referred to as the Kalman smoother.

6.3.2 Fixed Time Lag L

The fixed time lag, L , in both the Kalman and Wiener smoothers must be specified in a practical smoothing process. Lag L should be the required minimum because sequential processing may yield cumulative error unlike batch processing. To determine this value, a computer simulation is performed using 3GPP LTE specifications with a 5 MHz bandwidth. The frequency interval between each subcarrier is 15 kHz, and the symbol length (except for the GI) corresponds to 66.7 μ s. The fast Fourier transform (FFT) size is 512, wherein 300, 1, and 211 are the sizes for the active, direct current (DC), and guard subcarriers, respectively. The guard band is $5 - (300+1) \times 0.015 = 0.485$ MHz. Cell-specific reference signals (CRSs) [6.11] are regarded as pilot subcarriers. Figure 6.2 shows the CRS allocation. In the figure, CRSs at timeslots 4, 11, $\dots$ are not used in the simulation, $dt = 0.5$ ms, $df = 90$ kHz, and a specific pseudorandom sequence for the CRSs is modulated by QPSK. Thus, $\beta/\text{SNR} = 1/(2E_b/N_0)$

6.3. CONDITIONS FOR FAIR COMPARISON

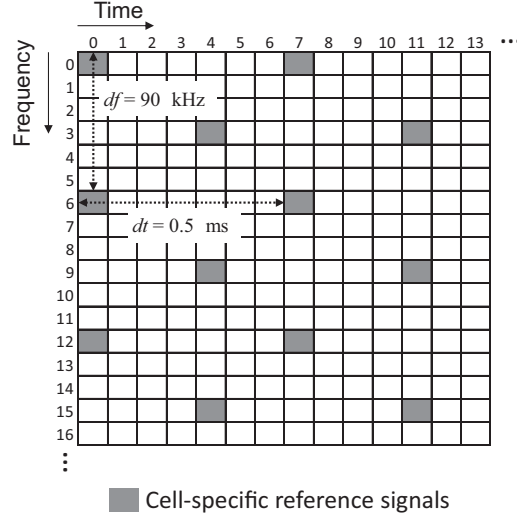


Figure 6.2: Scattered pilot subcarrier allocation.

in Equation (6.12). There are 250 data subcarriers and 50 pilot subcarriers for one symbol duration. An isolated cell and single-user environment without inter-cell interference and multiple access are assumed. Table 6.1 shows the other assumptions used for the computer simulation.

Fig 6.3 shows the impact of L on the cumulative error in the Kalman smoother.

CHAPTER 6. WIENER AND KALMAN FILTERS: PRACTICAL COMPARISON

Table 6.1: Computer simulation assumptions

Carrier frequency	2 GHz
Maximum Doppler frequency: f_D	14 Hz (i.e., 7.6 km/h)
Delay profile	• 24-path exponential – Delay spread: 1 μs – Maximum delay: 4.7 μs (within GI)
Numerical computation of Kalman filter/smoothen	A square root Kalman filter/smoothen based on singular value decomposition [6.12] is applied to suppress the degradation in numerical accuracy.

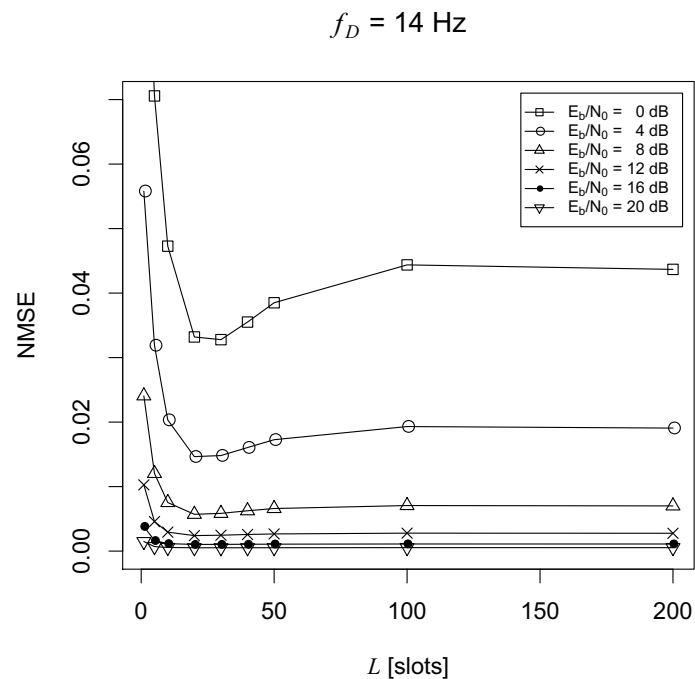


Figure 6.3: Impact of L on the cumulative error in Kalman smoother.

6.4. NUMERICAL ANALYSIS

In the figure, L (one slot corresponds to one dt) is denoted on the X-axis and the normalized mean squared error (NMSE) = $\sum_{i=0}^{L-1} E[|\tilde{h}_i(t) - \hat{\tilde{h}}_i(t)|^2] / \sum_{i=0}^{L-1} E[|\tilde{h}_i(t)|^2]$ is denoted on the Y-axis. The figure shows that L becomes optimal at around 20, so L is set to 20 slots (i.e., 10 ms) in the following investigation. This optimal L value corresponds to the time lag where the time correlation coefficient decreases from 1 to approximately 0.8 for $f_D = 14$ Hz.

6.4 Numerical Analysis

6.4.1 Performance Comparison

To compare the performance of NMSE and the bit error rate (BER), a computer simulation is performed under the same conditions as in Section 6.3.2.

6.4.1.1 NMSE Performance

Figure 6.4 shows the NMSE performance for $f_D = 14$ Hz. In the figure, E_b/N_0 is denoted on the X-axis and the NMSE is denoted on the Y-axis. Both the Kalman and Wiener smoothers at $L = 20$ exhibit good performance at almost the same level. However, the Wiener smoother at $L = 20$ slightly outperforms the Kalman smoother at $L = 20$. For example, at $E_b/N_0 = 0$ dB, the Wiener smoother at $L = 20$ improves the NMSE by 0.02 compared to the Kalman smoother at $L = 20$. Cumulative error in the sequential processing yields this difference. The higher E_b/N_0 becomes, the smaller this cumulative error becomes. Thus, at $E_b/N_0 = 20$ dB, the NMSE difference decreases to the negligible value of 0.0003.

The other performance levels such as the least-squares (LS) estimator, $y_i(t)/s_i(t)$, Wiener smoother at $L = 0$, and the Kalman filter are also plotted in Figure 6.4 for deeper understanding. The LS estimator and Wiener smoother at $L = 0$ estimate a fading channel using not past and future information but current information only. The Wiener smoother at $L = 0$ further considers *frequency correlation* and outperforms the LS estimator. The Kalman filter estimates the fading channel using *past* and current information considering frequency correlation. Thus, the Kalman filter outperforms the Wiener smoother at $L = 0$. The Kalman and Wiener smoothers at $L = 20$ estimate the fading channel using all the past, current, and relative *future* information considering frequency correlation. Thus, the Kalman and Wiener smoothers at $L = 20$ outperform the Kalman filter. As a result, the more information that is used for the estimation, the higher the level of performance that can be archived essentially.

CHAPTER 6. WIENER AND KALMAN FILTERS: PRACTICAL COMPARISON

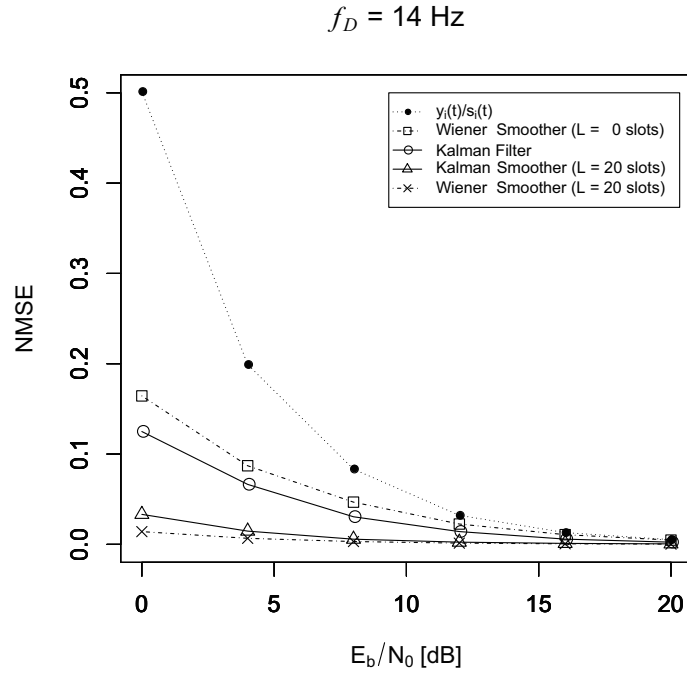


Figure 6.4: NMSE performance.

6.4.1.2 BER Performance

Figure 6.5 shows the BER performance for $f_D = 14 \text{ Hz}$ using the data subcarrier assumptions given in Table 6.2.

6.4. NUMERICAL ANALYSIS

Table 6.2: Data subcarrier assumptions

Modulation	QPSK
Interpolation of channel estimator	• Frequency domain: Linear insertion • Time domain: The nearest neighbor
Equalization	• Zero forcing (ZF) [6.13] for LS channel estimator $y_i(t)/s_i(t)$ • Minimum Mean Squared Error (MMSE) [6.13] for the other channel estimators
Channel coding	• Convolutional codes – Rate: 1/2 – Constraint length: 7
Channel decoding	Soft-decision Viterbi algorithm
Interleaving	• Random interleaver – Unit: Symbol-by-symbol – Depth: 5 Resource blocks (RBs) [6.11], which corresponds to 400 symbols

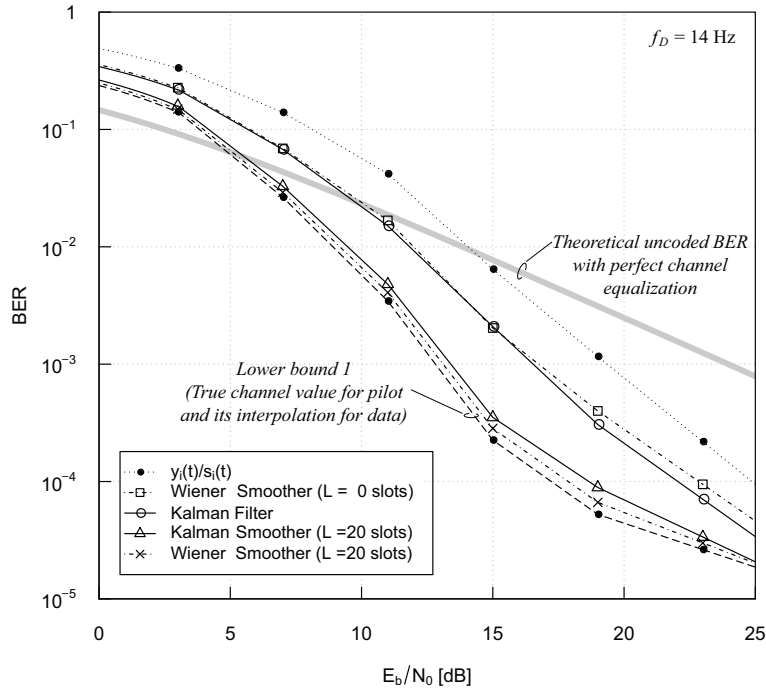


Figure 6.5: BER performance.

CHAPTER 6. WIENER AND KALMAN FILTERS: PRACTICAL COMPARISON

The BER and E_b/N_0 are denoted on the Y-axis and X-axis, respectively. The BER performance improves in the same order as the NMSE performance in Figure 6.4. Both the Kalman and Wiener smoothers at $L = 20$ exhibit almost the same performance close to Lower Bound 1. However, the Wiener smoother at $L = 20$ slightly outperforms the Kalman smoother at $L = 20$ due to the avoidance of cumulative error in the sequential processing. For example, at the BER of 10^{-4} , the Wiener smoother at $L = 20$ improves the coding gain by 0.8 dB compared to the Kalman smoother at $L = 20$.

6.4.2 Complexity Comparison

Complexity in this thesis refers to the number of multiplications and divisions occurring during the smoothing process which estimates the pilot signals at the t -th symbol. We assume pure matrix operations, and that the complexities of the inverse matrix derivation and Cholesky factorization are I^3 and $I^3/6$, respectively. In addition, we assume that the time and frequency correlation coefficients have been calculated in advance. Table 6.3 gives the complexity for the matrices prepared before the smoothing process. Note that $\mathcal{O}(1)$ complexity such as the setting of $\mathbf{V}$ and $\mathbf{W}$ is omitted from Table 6.3 because it has negligible impact. Table 6.4 also gives the complexity for the smoothing process itself. According to Tables 6.3 and 6.4, the complexity for the Kalman smoother is [Preparation of $\mathbf{F}(t)$] + $(2L + 1)$ [Kalman filter for 1 step] + L [RTS algorithm for 1 step] + [Derivation of $\hat{\mathbf{h}}(t)$ from state for the t -th symbol] = $(12L + 5 + 1/6)I^3 + (7L + 4)I^2$, and that for the Wiener smoother is [Preparation of $\mathbf{\Omega}_{tf}$] + [Wiener smoother] = $2(2L + 1)^3I^3 + (2L + 1)^2I^2 + 2(2L + 1)I$. Thus, the complexity for the Kalman smoother is always lower than that for the Wiener smoother for any integer $L \geq 0$ and $I \geq 1$. The major reason for this result depends on the presence or absence of a large matrix operation. In particular, the coefficient of I^3 for the Wiener smoother reaches $\mathcal{O}(L^3)$, while that for the Kalman smoother remains at $\mathcal{O}(L)$. According to this fact, the longer L becomes, the higher the complexity ratio of the Wiener smoother to the Kalman smoother becomes. For example, the ratio reaches 556 for $L = 20$ and $I = 50$.

Table 6.3: Complexity for matrices prepared before smoothing process

	Complexity	Note
$\mathbf{F}(t)$	$(2L + 1)I^2 + I^3/6$	Refer to Equations (6.4) and (6.5)
$\mathbf{\Omega}_{tf}$	$(2L + 1)I$	Refer to Toeplitz and Hermitian Equation (6.13)

Table 6.4: Complexity for smoothing process

	Complexity	Note
Kalman filter	$5I^3 + 2I^2$ (for 1 step)	Refer to Table 1 in [6.5]
RTS algorithm	$I^3 + I^3 + I^2$ (for 1 step)	Refer to Equation (6.18)
Derivation of $\tilde{\mathbf{h}}(t)$ from state	I^2 (for the t -th symbol)	Refer to Equation (6.3)
Wiener smoother	$2((2L+1)I)^3 + ((2L+1)I)^2 + (2L+1)I$	Refer to Equation (6.12)

6.5 Summary

This chapter compared the Kalman smoother to the Wiener smoother for stationary processes in terms of practical OFDM channel estimation on the receiver side. These two methods are theoretically equivalent but practically different. First, the conditions for fair comparison were discussed. Regarding the smoothing process, the fixed time lag should be set to the required minimum, because sequential processing yields cumulative error. The optimal value corresponds to the time lag where the time correlation coefficient decreases from 1 to approximately 0.8 for $f_D = 14$ Hz. The results of numerical analysis lead to the following conclusions.

- Performance (NMSE and BER): Both the Kalman and Wiener smoothers exhibit almost the same good performance levels. However, the Wiener smoother slightly outperforms the Kalman smoother because it avoids cumulative error in the sequential processing. The NMSE and coding gain improvements using the Wiener smoother reach, at most, 0.02 and 0.8 dB, respectively, compared to those for the Kalman smoother.
- Complexity: Regarding the number of multiplications and divisions occurring during the smoothing process which estimates the pilot signals at some symbols, the complexity of the Kalman smoother is always lower than that for the Wiener smoother because there is no large matrix operation. For example, the complexity ratio of the Wiener smoother to the Kalman smoother reaches 556 for $L = 20$ and $I = 50$.

According to the above, we recognize that the Kalman and Wiener smoothers for stationary processes have different advantages; the Kalman filter has a lower complexity level whereas the Wiener filter yields better performance. Note that the performance of the Wiener filter for non-stationary processes degrades compared with that for stationary processes; thus, the performance advantage of the Wiener filter is limited to stationary processes.

CHAPTER 6. WIENER AND KALMAN FILTERS: PRACTICAL COMPARISON

Bibliography

- [6.1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering, Transactions of the ASME*, vol. 82 (Series D), no. 1, pp. 35–45, Mar. 1960.
- [6.2] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Technology Press and Wiley, NY, 1949.
- [6.3] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall Upper Saddle River, NJ, 2000.
- [6.4] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. John Wiley & Sons, 2003.
- [6.5] J. Hagiwara, "An OFDM channel estimation method based on a state-space model that appropriately considers frequency correlation," *IEICE Trans. Fundamentals*, vol. E98-A, no. 2, pp. 537–548, Feb. 2015.
- [6.6] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Communications*, vol. 3, no. 1, pp. 315–325, Jan. 2004.
- [6.7] D. Larsson, "Analysis of channel estimation methods for OFDMA," Master's thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, Dec. 2006.
- [6.8] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 2, pp. 18–48, Apr. 2007.
- [6.9] W. C. Jakes, Jr., *Microwave Mobile Communications*. Wiley, New York, 1974, [reprinted by IEEE Press].
- [6.10] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, Aug. 1965.
- [6.11] "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation," The 3rd Generation Partnership Project (3GPP), TS 36.211, Dec. 2009.
- [6.12] L. Wang, G. Libert, and P. Manneback, "Kalman filter algorithm based on singular value decomposition," in *Proc. 31st IEEE Conference on Decision and Control (CDC)*, Tucson, U.S.A., Dec. 1992, pp. 1224–1229.

BIBLIOGRAPHY

- [6.13] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

Chapter 7

Conclusions

For contributing to the significantly growing mobile communication field, this study investigated a method for wireless OFDM channel estimation that uses a stochastic approach to improve channel estimation accuracy and clarify characteristics based on estimation processing.

Chapter 1 presents an introduction of this study, including a general background, purpose, and structure.

Chapter 2 provides the preliminary background of the study, including investigation assumptions, the formulation in OFDM transmission, and the fundamentals of stochastic filtering.

Chapter 3 proposes a novel scheme for sequential OFDM channel estimation; it comprises two methods: Method 1, which improves estimation accuracy, and Method 2, which reduces computational complexity. Method 1 appropriately considers frequency correlation based on a state-space model. Hyperparameters in the state-space model are specified by the ML method, assuming these hyperparameters do not change during a short training period before channel estimation; the ML specification error does not have a dominant influence on performance. Method 2 is based on Method 1 and forces the observation matrix into a sparse bidiagonal matrix. The effect of the proposed scheme is confirmed on the basis of numerical analysis that assumes CRSs in 3GPP LTE as pilot subcarriers. The results of that analysis lead to the following conclusions regarding performance:

- NMSE
 - The NMSE is improved in order of the LS estimator $y_i(t)/s_i(t)$, the conventional Kalman filter (which does not consider frequency correlation), Method 2, and Method 1.
 - NMSE improvements reach, at most, about 50% compared with the conventional Kalman filter.

CHAPTER 7. CONCLUSIONS

- The lower the E_b/N_0 value becomes, the more effective the proposed methods are at improving the NMSE.
- BER
 - The BER is improved in the same order as the NMSE.
 - Coding gain improvements reach, at most, about 1 dB compared with the conventional Kalman filter.
 - The condition where the proposed methods are effective, compared with the conventional Kalman filter, is when there is a small frequency interval, df , between adjacent pilot subcarriers (which is equivalent to a small channel delay spread, σ_τ). When df is reduced to half, coding gain improvements with true hyperparameters reach, at most, about 3 dB compared with the conventional Kalman filter and the BER performance is closer to the lower bound.
- Computational complexity
 - The proposed methods' complexity is higher than that of the conventional Kalman filter.
 - Method 2 reduces complexity by about 50% compared with Method 1 and suppresses increased complexity up to a few percents compared with the conventional Kalman filter.

Thus, the proposed scheme improves channel estimation accuracy and reduces computational complexity.

Chapter 4 investigates a real-time joint channel and hyperparameter estimation method for OFDM mobile communications. To estimate the channel frequency response of the pilot subcarrier and its fixed hyperparameters (such as channel statistics), the method adopts LWF, a modern and effective technique based on the state-space model and sequential Monte Carlo method. For the first time, to our knowledge, we demonstrate biased hyperparameter estimation using the conventional LWF and show that bias results from poor likelihood estimation caused by overfitting in noisy environments. This problem, which cannot be overcome by conventional smoothing techniques, is here corrected by a novel method that modifies the conventional LWF and regularizes the likelihood using a Kalman smoother. The effectiveness of the proposed method is confirmed in numerical analyses assuming CRSs in 3GPP LTE as pilot subcarriers. The following conclusions were drawn. First, when the hyperparameter f_D is unknown, the NMSE and BER of conventional RB-LWF are acceptable, but the proposed RB-LWS outperforms the RB-LWF and LS, and achieves good performance near the lower bound. Regarding the BER performance, the proposed method improves the coding gain to RB-LWF and

LS by approximately 1 and 3 dB at most, respectively. Second, when two hyperparameters (f_D and σ_τ) are unknown, LS sometimes outperforms the conventional RB-LWF, because of hyperparameter estimation failure in the latter method. The proposed RB-LWS overcomes this problem by avoiding such failures. We emphasize that joint channel and multiple hyperparameters estimation is much more difficult than single hyperparameter case. The proposed RB-LWS outperforms the RB-LWF and LS, and achieves good performance near the lower bound. Regarding the BER performance, the proposed method improves the coding gain to RB-LWF and LS by approximately 10 and 3 dB at most, respectively. Clearly, the proposed method offers higher channel and hyperparameter estimation accuracy than the conventional RB-LWF.

Chapter 5 investigated a method for detecting abrupt changes in the channel statistics to maintain accurate channel estimation. The proposed novel detection method relies on the state-space model and makes use of the Kalman filter. The effectiveness of the proposed method is verified based on computer simulation using the delay spread as the channel statistics. The proposed method basically detects abrupt changes through extreme and moderate changes in delay spread, while the detection threshold is a tuning matter. We also confirmed that the greater the deviation between delay spreads and the higher the E_b/N_0 value, the easier the detection.

Chapter 6 compared the Kalman smoother to the Wiener smoother for stationary processes in terms of practical OFDM channel estimation. These two methods are theoretically equivalent but practically different. First, the conditions for fair comparison were discussed. Regarding the smoothing process, the fixed time lag should be set to the required minimum, because sequential processing yields cumulative error. The optimal value corresponds to the time lag where the time correlation coefficient decreases from 1 to approximately 0.8 for $f_D = 14$ Hz. The results of numerical analysis lead to the following conclusions.

- Performance (NMSE and BER): Both the Kalman and Wiener smoothers exhibit almost the same good performance levels. However, the Wiener smoother slightly outperforms the Kalman smoother because it avoids cumulative error in the sequential processing. The NMSE and coding gain improvements using the Wiener smoother reach, at most, 0.02 and 0.8 dB, respectively, compared to those for the Kalman smoother.
- Complexity: Regarding the number of multiplications and divisions occurring during the smoothing process which estimates the pilot signals at some symbols, the complexity of the Kalman smoother is always lower than that for the Wiener smoother because there is no large matrix operation. For example, the complexity ratio of the

CHAPTER 7. CONCLUSIONS

Wiener smoother to the Kalman smoother reaches 556 for $L = 20$ and $I = 50$.

According to the above, we recognize that the Kalman and Wiener smoothers for stationary processes have different advantages; the Kalman filter has a lower complexity level whereas the Wiener filter yields better performance. Note that the performance of the Wiener filter for non-stationary processes degrades compared with that for stationary processes; thus, the performance advantage of the Wiener filter is limited to stationary processes.

Mobile communication researchers have recognized the importance of channel statistics in channel estimation for broadband wireless transmission. However, channel statistics are not appropriately considered, assumed to be known or extractable by other methods. Our precise and pragmatic proposal methods should considerably improve the quality of mobile communications. Moreover, our proposal is mathematically formulated and is applicable to not only wireless technology but also related areas such as artificial intelligence and econometrics.

Multiple-input multiple-output (MIMO [7.1,7.2])-OFDM schemes are nowadays state of the art for modern communication systems. The space correlation as well as time and frequency correlation should be considered for single-user MIMO-OFDM channel estimation [7.3]. From point of view of multi-user MIMO-OFDM channel estimation, precoding is required in the downlink process to eliminate inter-user interference, and channel prediction based on the estimator obtained at the base station is important [7.4]. Relevant extensions to MIMO systems will be examined in future work.

Bibliography

- [7.1] T. Ohgane and Y. Ogawa, *Introduction to MIMO Sytem Technology*. Ohm Publishing, 2009, [in Japanese].
- [7.2] K. Nishimori, *Basic Theory and Performance on Multi-user MIMO*. Corona Publishing, 2014, [in Japanese].
- [7.3] T. Roman, M. Enescu, and V. Koivunen, "Recursive estimation of time-varying channel and frequency offset in MIMO OFDM systems," in *Proc. 14th IEEE Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 2, Beijing, China, Sep. 2003, pp. 1934–1938.
- [7.4] K. Yamaguchi, H. P. Bui, Y. Ogawa, T. Nishimura, and T. Ohgane, "Channel prediction techniques for a Multi-User MIMO system in time-varying environments," *IEICE Trans. Communications*, vol. E97-B, no. 12, pp. 2747–2755, 2014.

Appendix A

Derivation of Covariance $\mathbf{W}$ in State Noise

From Equation (3.5), $\mathbf{W} = \text{Var}[\mathbf{w}(t)] = \text{Var}[\boldsymbol{\theta}(t) - \boldsymbol{\theta}(t-1)]$. $\boldsymbol{\theta}(t)$ has no frequency correlation from its definition, so $\mathbf{W}$ becomes a diagonal matrix. Furthermore, the WSSUS condition results in all diagonal elements of $\mathbf{W}$ to be the same. When $\theta_i(t)$ denotes the i -th element of $\boldsymbol{\theta}(t)$, the diagonal value is as follows:

$$\begin{aligned} \text{Var}[\theta_i(t) - \theta_i(t-1)] &= E[\{(\theta_i(t) - \theta_i(t-1)) - (0 - 0)\} \\ &\quad \cdot \{(\theta_i(t) - \theta_i(t-1)) - (0 - 0)\}^*] \\ &= E[(\theta_i(t) - \theta_i(t-1))(\theta_i^*(t) - \theta_i^*(t-1))] \\ &= E[\theta_i(t)\theta_i^*(t)] + E[\theta_i(t-1)\theta_i^*(t-1)] \\ &\quad - E[\theta_i(t)\theta_i^*(t-1)] - E[\theta_i(t-1)\theta_i^*(t)], \quad (\text{A.1a}) \end{aligned}$$

where, by the WSSUS condition,

$$= 2E[\theta_i(t)\theta_i^*(t)] - 2E[\theta_i(t)\theta_i^*(t-1)], \quad (\text{A.1b})$$

where, by definition,

$$\begin{aligned} &= 2l^2 - 2l^2\rho(\Delta t = 1dt, \Delta f = 0) \\ &= 2l^2(1 - \rho(\Delta t = 1dt, \Delta f = 0)). \quad (\text{A.1c}) \end{aligned}$$

Thus, $\mathbf{W} = 2l^2 (1 - \rho(\Delta t = 1dt, \Delta f = 0)) \mathbf{E}_I$.

Appendix B

Kalman Filter: Another Representation for Complexity Investigation

To aid in the complexity investigation, in Chapter 3, Kalman filter update process at the t -th symbol in Algorithm 2.1 is slightly modified for avoiding redundant calculation:

- One-step-ahead prediction of state

$$\text{(Mean)} \quad \mathbf{a}(t) \leftarrow \mathbf{G}(t)\mathbf{m}(t-1) \quad (\text{B.1})$$

$$\text{(Variance)} \quad \mathbf{R}(t) \leftarrow \mathbf{G}(t)\mathbf{C}(t-1)\mathbf{G}(t)^H + \mathbf{W} \quad (\text{B.2})$$

- One-step-ahead prediction of observation

$$\begin{aligned} \text{(Variance)} \quad \mathbf{q}(t) &\leftarrow \mathbf{R}(t)\mathbf{F}(t)^H \\ \mathbf{Q}(t) &\leftarrow \mathbf{F}(t)\mathbf{q}(t) + \mathbf{V} \end{aligned} \quad (\text{B.3})$$

- Kalman gain

$$\mathbf{K}(t) \leftarrow \mathbf{q}(t)\mathbf{Q}(t)^{-1} \quad (\text{B.4})$$

- Filtered state

$$\text{(Mean)} \quad \mathbf{m}(t) \leftarrow \mathbf{a}(t) + \mathbf{K}(t)[\mathbf{y}(t) - \mathbf{F}(t)\mathbf{a}(t)] \quad (\text{B.5})$$

$$\text{(Variance)} \quad \mathbf{C}(t) \leftarrow \mathbf{R}(t) - \mathbf{K}(t)\mathbf{q}(t)^H \quad (\text{B.6})$$

, being Hermitian $\mathbf{R}(t)$.

Appendix C

How to Derive the Nearest Positive Definite Matrix for a Forced Tridiagonal Correlation Matrix

As mentioned in Section 3.4.2, several approaches were proposed for deriving the nearest positive definite matrix; however, this thesis proposes an analytical approach, making use of the tridiagonal feature. Here, the individual conditions for the positive definite matrix or the nearest matrix are first discussed and then their simultaneous conditions are considered.

For any Hermitian matrix, its positive definiteness is equivalent to the condition where its eigenvalues are all positive, so if all eigenvalues are obtained, the condition for positive definiteness is clear. The eigenvalue of a general matrix whose order is five and more cannot be obtained analytically [C.1]. However, all eigenvalues of a tridiagonal correlation matrix can be obtained analytically even in an arbitrary order. When the tridiagonal correlation matrix whose order is I is expressed as Equation (C.1), its eigenvalues are obtained as Equation (C.2). This derivation can be shown, for example, through the fact that the characteristic polynomial of Equation (C.1) becomes the Chebyshev polynomials of the second kind [C.2] using cofactor expansion.

$$\begin{bmatrix} 1 & g & & & \\ g^* & 1 & g & & \\ & g^* & 1 & \ddots & \\ & & \ddots & \ddots & g \\ & & & g^* & 1 \end{bmatrix}, \quad (\text{C.1})$$

where g is a frequency correlation coefficient between adjacent pilot sub-

APPENDIX C. HOW TO DERIVE THE NEAREST POSITIVE DEFINITE MATRIX FOR A FORCED TRIDIAGONAL CORRELATION MATRIX

carriers.

$$1 - 2|g| \cos\left(\frac{i+1}{I+1}\pi\right), \quad (\text{C.2})$$

where $i = 0, \dots, I-1$.

Assuming Equation (C.2) is positive, the condition for positive definiteness is

$$|g| < \frac{1}{2 \cos(\pi/(I+1))}. \quad (\text{C.3})$$

On the other hand, the nearest matrix is defined here as the closest matrix from the viewpoint of its Frobenius norm, according to [C.3]. For matrix Equation (C.1), the condition for the nearest matrix corresponds to that where value $|g|$ is the closest to the original value.

From the above, how to derive the nearest positive definite matrix for the forced tridiagonal correlation matrix defined in Equation (C.1) is as follows:

- When Equation (C.3) is satisfied (positive definiteness is already satisfied), g is not changed.
- Otherwise, g should be changed to $(\frac{1}{2 \cos(\pi/(I+1))} - \epsilon) \exp\{j \arg(g)\}$ (where ϵ denotes a very small numerical value, such as a machine epsilon [C.4]).

Bibliography

- [C.1] F. Chatelin, *Eigenvalues of Matrices*, Revised ed., ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2012, no. 71.
- [C.2] D. Fuchs and S. Tabachnikov, *Mathematical Omnibus: Thirty Lectures on Classic Mathematics*. American Mathematical Society, 2007.
- [C.3] N. J. Higham, "Computing the nearest correlation matrix — a problem from finance," *IMA Journal of Numerical Analysis*, vol. 22, no. 3, pp. 329–343, 2002.
- [C.4] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. Society for Industrial and Applied Mathematics, 1999.

Appendix D

RB-APF

Algorithm D.1 summarizes the recursion process of RB-APF at the t -th symbol. The target distribution is the filtered distribution $p(\psi(t) \mid \mathbf{y}(1:t))$.

Algorithm D.1 RB-APF

0. Filtered state at the $t-1$ -th symbol:

$$\left\{ \text{Realization } \psi(t-1)^{(n)}, \mathbf{m}(t-1)^{(n)}, \mathbf{C}(t-1)^{(n)}, \text{Weight } \omega(t-1)^{(n)} \right\}_{n=1}^N$$
1. Update process at the t -th symbol
 - *Resampling*
 Draw auxiliary index k_n from a set $\{1, \dots, N\}$ with $P(k_n = n) \propto \omega(t-1)^{(n)} p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(n)})$, where
 - $\hat{\psi}(t)^{(n)} = E[\psi(t) | \psi(t-1)^{(n)}]$,
 - $p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(n)})$ is equal to

$$\mathcal{L}(\text{hyp}=\hat{\psi}(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(n)}, \mathbf{C}(t-1)^{(n)}\}, \text{dat}=\mathbf{y}(t)).$$
 - **for** $n = 1$ **to** N **do**
 - a-1. *Realization (for non-linear and non-Gaussian part)*
 Draw $\psi(t)^{(n)}$ from

$$\text{transition distribution } p(\psi(t) | \psi(t-1)^{(k_n)}).$$
 - a-2. *Realization (for linear and Gaussian part)*
 Derive $\{\mathbf{m}(t)^{(n)}, \mathbf{C}(t)^{(n)}\}$ via $\mathcal{KF}(\text{hyp}=\psi(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(k_n)}, \mathbf{C}(t-1)^{(k_n)}\}, \text{dat}=\mathbf{y}(t))$.
 - b. *Weight*

$$\omega(t)^{(n)} \leftarrow \frac{p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \psi(t)^{(n)})}{p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(k_n)})}, \text{ where}$$
 - $p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \psi(t)^{(n)})$ is equal to

$$\mathcal{L}(\text{hyp}=\psi(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(k_n)}, \mathbf{C}(t-1)^{(k_n)}\}, \text{dat}=\mathbf{y}(t)).$$
 - **end for**
 - *Normalization of the weights:* $\omega(t)^{(n)} \leftarrow \omega(t)^{(n)} / \sum_{n=1}^N \omega(t)^{(n)}.$
2. Filtered state at the t -th symbol:

$$\left\{ \text{Realization } \psi(t)^{(n)}, \mathbf{m}(t)^{(n)}, \mathbf{C}(t)^{(n)}, \text{Weight } \omega(t)^{(n)} \right\}_{n=1}^N$$

Appendix E

RB-APF with the Kernel Smoothing Technique

Algorithm E.1 summarizes the recursion process of RB-APF with the kernel smoothing technique at the t -th symbol. The target distribution is the filtered distribution $p(\psi(t) \mid \mathbf{y}(1:t))$.

APPENDIX E. RB-APF WITH THE KERNEL SMOOTHING TECHNIQUE

Algorithm E.1 RB-APF with the kernel smoothing technique

0. Filtered state at the $t-1$ -th symbol:

$$\left\{ \text{Realization } \psi(t-1)^{(n)}, \mathbf{m}(t-1)^{(n)}, \mathbf{C}(t-1)^{(n)}, \text{Weight } \omega(t-1)^{(n)} \right\}_{n=1}^N$$
 1. Update process at the t -th symbol
 - *Artificial moving average for hyperparameters*

$$\mu^{(n)} \leftarrow a\psi(t-1)^{(n)} + (1-a)E_{\omega(t-1)^{(n)}}[\psi(t-1)^{(n)}],$$

$$\Gamma \leftarrow (1-a^2)\text{Var}_{\omega(t-1)^{(n)}}[\psi(t-1)^{(n)}].$$
 - *Resampling*

Draw auxiliary index k_n from a set $\{1, \dots, N\}$ with $P(k_n = n) \propto \omega(t-1)^{(n)} p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(n)})$, where

 - $\hat{\psi}(t)^{(n)} = \mu^{(n)}$,
 - $p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(n)})$ is equal to

$$\mathcal{L}(\text{hyp}=\hat{\psi}(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(n)}, \mathbf{C}(t-1)^{(n)}\}, \text{dat}=\mathbf{y}(t)).$$
 - **for** $n = 1$ **to** N **do**
 - a-1. *Realization* (for nonlinear and non-Gaussian part)

Draw $\psi(t)^{(n)}$ from

$$\text{continuous distribution (Mean}=\mu^{(k_n)}, \text{Variance}=\Gamma).$$
 - a-2. *Realization* (for linear and Gaussian part)

Derive $\{\mathbf{m}(t)^{(n)}, \mathbf{C}(t)^{(n)}\}$ via $\mathcal{KF}(\text{hyp}=\psi(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(k_n)}, \mathbf{C}(t-1)^{(k_n)}\}, \text{dat}=\mathbf{y}(t))$.
 - b. *Weight*

$$\omega(t)^{(n)} \leftarrow \frac{p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \psi(t)^{(n)})}{p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \hat{\psi}(t)^{(k_n)})}, \text{ where}$$
 - $p(\mathbf{y}(t) | \mathbf{y}(1:t-1), \psi(t)^{(n)})$ is equal to

$$\mathcal{L}(\text{hyp}=\psi(t)^{(n)}, \text{pri}=\{\mathbf{m}(t-1)^{(k_n)}, \mathbf{C}(t-1)^{(k_n)}\}, \text{dat}=\mathbf{y}(t)).$$
 - *Normalization of the weights:* $\omega(t)^{(n)} \leftarrow \omega(t)^{(n)} / \sum_{n=1}^N \omega(t)^{(n)}.$
 2. Filtered state at the t -th symbol:

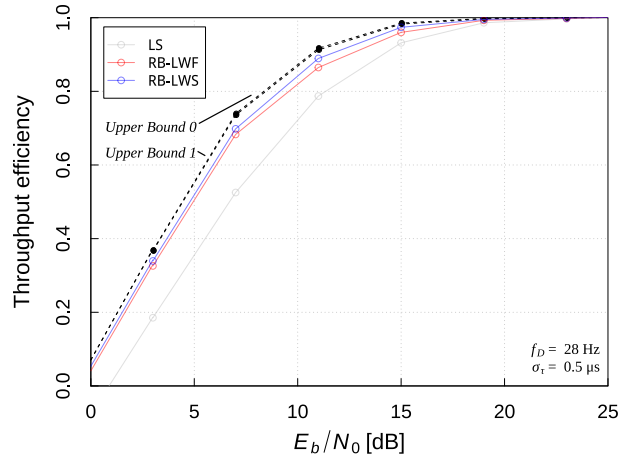
$$\left\{ \text{Realization } \psi(t)^{(n)}, \mathbf{m}(t)^{(n)}, \mathbf{C}(t)^{(n)}, \text{Weight } \omega(t)^{(n)} \right\}_{n=1}^N$$
-

Appendix F

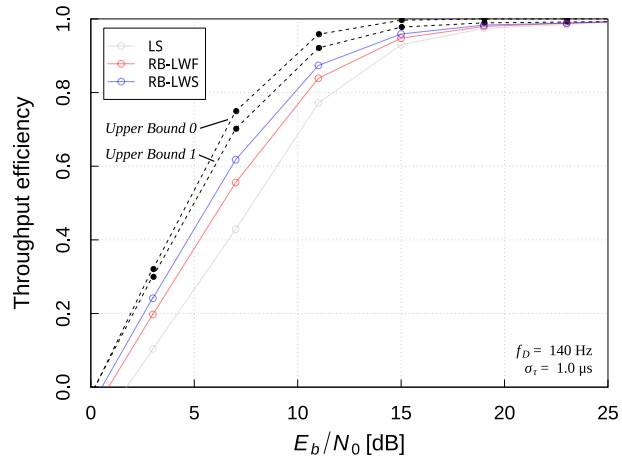
Throughput Efficiency

When a pure selective repeat type of automatic repeat request (ARQ) is applied, the throughput efficiency is expressed as $1 - \text{packet error rate}$ [F.1]. Assuming the coding block in Table 4.3 to be a packet, the throughput efficiency performances are shown in Figures F.1 and F.2. In the figures, the Upper Bound 0 and 1 correspond to the Lower bound 0 and 1 in Section 4.5, respectively.

APPENDIX F. THROUGHPUT EFFICIENCY

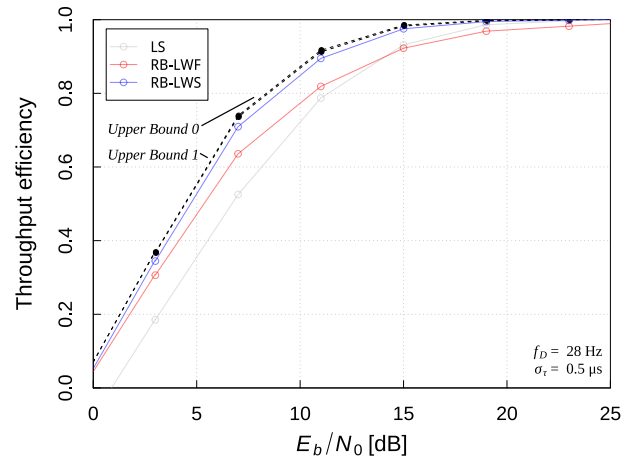


(a)

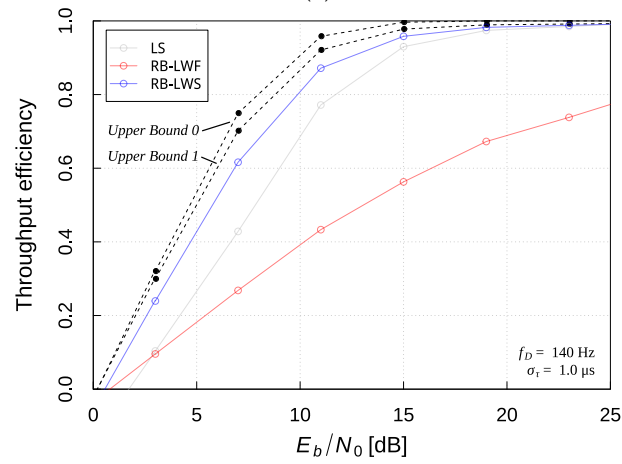


(b)

Figure F.1: Throughput efficiency with $f_D = 28$ Hz (top) and 140 Hz (bottom) in case of a single unknown hyperparameter.



(a)



(b)

Figure F.2: Throughput efficiency with $f_D = 28$ Hz (top) and 140 Hz (bottom) in case of two unknown hyperparameters.

APPENDIX F. THROUGHPUT EFFICIENCY

Bibliography

- [F.1] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Pearson Prentice Hall, 2004.

Acknowledgements

I sincerely thank my supervisor Professor Takeo Ohgane of the Graduate School of Information Science and Technology, Hokkaido University for his kind guidance and support.

In addition to my supervisor, I am grateful to the sub-supervisors of my thesis, Professor Yoshikazu Miyanaga and Professor Kunimasa Saito of the Graduate School of Information Science and Technology, Hokkaido University, for their insightful comments and suggestions.

I would like to extend my appreciation to Professor Emeritus Yasutaka Ogawa and Assistant Professor Toshihiko Nishimura of the Graduate School of Information Science and Technology, Hokkaido University for their constructive discussions and encouragement.

I also thank Mr. Seizo Onoe, Executive General Manager of the R&D Innovation Division, NTT DOCOMO, INC., Dr. Narumi Umeda, Advisor of the Research and Development Division, Japan Radio Co., Ltd., and Mr. Akihiro Maebara, General Manager of the Radio Access Network Development Department, NTT DOCOMO, INC. for their grant and financial support towards my endeavor.

I am indebted to Dr. Takahiro Asai, Executive Research Engineer of the Research Laboratories, NTT DOCOMO, INC., Dr. Yusuke Ohwatari, Assistant Manager of the R&D Strategy Department, NTT DOCOMO, INC., Mr. Yousuke Sano of the 5G Laboratory, NTT DOCOMO, INC., and Mr. Daisuke Koizumi, Assistant Manager of the Service Innovation Department, NTT DOCOMO, INC. for their very valuable advice and cooperation regarding numerical evaluation in my thesis.

I would further like to thank Dr. Yoshihiro Ishikawa, Senior Manager of the Radio Access Network Development Department, NTT DOCOMO, INC., Dr. Kenji Shimizu, Senior Research Engineer of the Network Innovation Laboratories, NTT, INC., and Mr. Akira Yamada, Executive Research Engineer of the Research Laboratories, NTT DOCOMO, INC. for their help and encouragement.

Finally, I truly thank my family for all the happy memories.

Publications

Journal Paper

- [1] J. Hagiwara, “An OFDM channel estimation method based on a state-space model that appropriately considers frequency correlation,” *IEICE Trans. Fundamentals*, vol. E98-A, no. 2, pp. 537–548, Feb. 2015.
- [2] —, “Real-time joint channel and hyperparameter estimation using sequential Monte Carlo methods for OFDM mobile communications,” *IEICE Trans. Communications*, vol. E99-B, no. 8, pp. 1655–1668, Aug. 2016.

International Conference

- [1] J. Hagiwara, T. Nishimura, T. Ohgane, and Y. Ogawa, “Comparison of Kalman smoother to Wiener smoother in practical OFDM channel estimation,” in *Proc. 21st Asia-Pacific Conference on Communications (APCC)*, Kyoto, Japan, Oct. 2015, pp. 454–459.
- [2] —, “Detecting abrupt changes in delay spread for OFDM channel estimation,” in *Proc. IEEE VTS 13th Asia-Pacific Wireless Communications Symposium (APWCS)*, Tokyo, Japan, Aug. 2016, pp. 479–483.

Others

- [1] J. Hagiwara, Y. Ogawa, H. Kitajima, and T. Shirakawa, “Fading analysis in land mobile reception of TV waves,” *The Journal of the Institute of Television Engineers of Japan*, vol. 46, no. 12, pp. 1625–1634, Dec. 1992, [in Japanese].
- [2] T. Matsuda, S. Tomisato, M. Hata, H. Fujii, and J. Hagiwara, “Peak reduction improvement in iterative clipping and filtering with a graded band-limiting filter for OFDM transmission,” *IEICE Trans. Fundamentals*, vol. E90-A, no. 7, pp. 1362–1365, Jul. 2007.

PUBLICATIONS

- [3] K. Yagyu, T. Nakamori, H. Ishii, M. Iwamura, N. Miki, T. Asai, and J. Hagiwara, "Investigation on signaling overhead for mobility management with carrier aggregation in LTE-Advanced," *IEICE Trans. Communications*, vol. E94-B, no. 12, pp. 3335–3345, Dec. 2011.
- [4] —, "Investigation on mobility management for carrier aggregation in LTE-Advanced," in *Proc. IEEE Vehicular Technology Conference (VTC)-Fall*, Sep. 2011, pp. 1–5.
- [5] M. Shirakabe, Y. Ofuji, A. Hanaki, N. Miki, T. Asai, and J. Hagiwara, "Study on resource allocation for VoIP traffic with subframe bundling in LTE uplink," in *Proc. IEEE 14th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Oct. 2011, pp. 1–5.
- [6] J. Kim, Y. Morioka, and J. Hagiwara, "An optimized seamless IP flow mobility management architecture for traffic offloading," in *Proc. IEEE Network Operations and Management Symposium (NOMS)*, Apr. 2012, pp. 229–236.
- [7] Y. Oyabu, M. Terada, T. Yamaguchi, S. Iwasawa, J. Hagiwara, and D. Koizumi, "Evaluating reliability of mobile spatial statistics," *NTT DOCOMO Technical Journal*, vol. 14, no. 3, pp. 16–23, Jan. 2013.