A Variational Bayesian Framework Divergence Minimization and Its Application in CDMA Receivers

Hu, Bin

Publication date:
2010

Document Version
Accepted author manuscript, peer reviewed version

A Variational Bayesian Framework Divergence Minimization and Its Application in CDMA Receivers

THESIS

SUBMITTED FOR DEGREE OF DOCTOR OF PHILOSOPHY

BY

Bin Hu

DEPARTMENT OF ELECTRONIC SYSTEMS
AALBORG UNIVERSITY, DENMARK

Copenhagen, July 2010

Contents

Acknowledgments

I would have not finished this thesis without help. I am very grateful to my supervisor, Professor Bernard H. Fleury, for sharing his deep insights in many research fields, for setting up a high research standard for me, for always being patient, and for investing consistent efforts into making me a better technical writer. I am also very grateful to my co-supervisor Ingmar Land who has always been there to listen to me and who has helped me formulate my work in a more precise way. I am very grateful to Professor Lars K. Rasmussen from University of South Australia, Adelaide, who has given me invaluable guidance throughout my studies. In particular, I would like to thank him for supporting my research visit to University of South Australia. I would also like to thank my external advisor Klaus Albert from RTX Telecom A/S for his continuous encouragement during my studies and for his visionary industry guidance.

I would like to thank Romain Piton who worked very closely with me during most of my studies. He was there whenever I needed help. We could always discuss new immature ideas and open questions. Amongst many others I would like to thank especially the following people for their support: Troels Pedensen, Xuefeng Yin, Alexander Kocian and Gunvor Elisabeth Kirkelund. In addition, I would also like to thank Alex Grant, Joachim Wehinger, Lars Christensen, and Thomas Zemen for their helpful comments.

I would like to thank particularly my husband Yufeng for his support on this journey in both bright and dark moments. Thanks to my son Christian for giving me endless inspiration and energy to accomplish this work. Finally, I would like to thank my parents for their strong encouragement and their time spent on baby-sitting their grandson.

The work on this thesis was carried out during my study at Navigation and Communications Section, Institute of Electronics, Aalborg University and I am grateful for the financial support provided by the Program of Wireless Communications at the Institute of Electronics and by RTX Telecom A/S.

The thesis is completed more than two years after I started working in Modem Algorithm Design (MAD) team in Nokia Denmark. For this period, I would like to thank my boss Niels Mørch for his encouragement to finish the thesis work and for generous support. For the thesis review, I would like to thank my kind colleagues in MAD team: Maja Lončar, Lars P. Christensen, Morten Hansen and Craig Mitchell.

Abstract

This thesis deals with the problem of multiuser decoding in the presence of unknown channel parameters in the uplink of code-division multiple access (CDMA) systems. Inspirited by the variational Bayesian expectation-maximization algorithm, a theoretical framework that uses a divergence minimization (DM) criterion is derived. The DM method provides a unified approach to jointly design the main components of an iterative CDMA receiver: channel weight estimator, noise variance inverse estimator, multiuser detector and decoders. In the DM method, the Kullback Leibler (KL) divergence between a postulated auxiliary distribution and the original joint posterior distribution of the unknown parameters is defined and minimized in an iterative fashion. Compared to most heuristic approaches proposed in the literature, the advantage of this systematic approach is threefold: the components of the iterative receiver are optimized globally; the sequence of estimated auxiliary distributions is always guaranteed to converge in KL divergence; and the messages that are passed between the different components are derived within the framework.

In this thesis, we have selected two auxiliary distributions to design two receiver structures with different performances and complexities. The first receiver (DM3) performs joint-user channel weight estimation, noise variance inverse estimation, successive interference cancellation and single-user decoding in an iterative fashion. The second receiver structure (DM5) is similar to the first one except that it performs separate-user channel weight estimation, which results in a lower complexity.

The receiver structures derived within the DM framework differ from nowadays published related structures in the following characteristics: 1) The means and variances of the modulation symbols fed into the interference cancellation component are computed base on posterior distributions. This is in contrast to the previously published interference cancellation schemes using extrinsic values, as motivated by applying belief propagation on factor graphs. 2) The noise variance inverse estimator takes into account the uncertainty on the symbol estimates and the channel weight

estimates; thus, not only thermal noise but also residual interference are estimated. The noise variance inverse estimator is a novel alternative to the traditional instantaneous LMMSE filter applied after interference cancellation to remove the residual interference in current state-of-the-art solutions. 3) Single-user decoding employing the BCJR algorithm is a natural outcome from the derivation within the DM framework. Thus, the resulting iterative receiver with soft symbol feedback has a guaranteed convergence in KL divergence.

The DM receivers are compared to a start-of-the-art receiver. It is shown by means of Monte Carlo simulations, that in low load regime, the DM3 and state-of-the-art receivers have very similar bit-error-rate performance. In high load regime, however, the former receiver outperforms the latter even in the scenario where the state-of-the-art receiver knows the noise variance, while this information is not available to the DM3 receiver. As expected, the DM3 receiver outperforms the DM5 receiver in high load regime: the latter receiver supports fewer users than the former at the same signal-to-noise ratio.

Resumé

Denne afhandling omhandler dekoding af flere brugere i uplink code-division multiple access (CDMA) systemer under antagelse af ukendte kanalparamtre. Til løsning af dette problem udledes på baggrund af den variationelle Bayesianske expectation-maximization algoritme en teoretisk metode kaldet divergens minimering (DM). DM-metoden gør det muligt at opnå et samlet optimeringsmål for designet af hovedkomponenterne i en CDMA modtager: kanalkoefficient estimation, invers støjvarians estimation, flerbruger detektion samt kanaldekodning. I DM-metoden minimeres Kullback Leibler (KL) divergensen mellem en postuleret fordeling og den korrekte simultanfordeling for de ukendte paramtre ved hjælp af en iterativ procedure. Sammenlignet med de fleste andre heuristiske metoder foreslået i litteraturen har denne systematiske metode tre fordele: komponenterne i den iterative modtager optimeres samlet; den iterative optimering er garanteret at konvergere i KL divergensen; og beskederne der udveksles mellem komponenterne kan udledes direkte som en del af metoden.

I denne afhandling undersøges to fordelinger til design af modtagerstrukturer med forskellig ydeevne og implementeringskompleksitet. Den første modtagerstruktur (DM3) udfører iterativ flerbruger kanalestimation, invers støjvarians estimation, succetiv interferensfjernelse samt enkeltbruger dekodning. Den anden modtagerstruktur udledt fra DM metoden (DM5) er magen til den første bortset fra at den udfører separat enkeltbruger kanalestimation, hvilket resulterer i en lavere kompleksitet.

Modtagerstrukturerne udledt fra DM metoden udskiller sig fra andre relaterede offentliggjorte strukturer på disse punkter: 1) Middelværdi og varians af modulationssymbolerne som benyttes i komponenten til interferensfjernelse beregnes på baggrund af posteriorfordelinger. Dette står i kontrast til tidligere offentliggjorte resultater for interferensfjernelse, som er baseret på extrinsicfordelinger motiveret af belief propagation på faktorgrafer. 2) Estimationen af den inverse støjvarians medtager usikkerheden på symbolestimaterne og kanalkoefficienterne; dermed tages der

højde for både termisk støj samt residualinterferens. Den inverse støjvariansestimation er et nyt alternativ til det LMMSE filter der traditionelt anvendes efter interferensfjernelse i de bedste nuværende løsninger til at undertrykke evt. resterende interferens. 3) Enkeltbruger dekodning baseret på BCJR algoritmen er et naturligt resultat som kommer direkte ud af DM metoden. Den resulterende iterative modtagerstruktur med soft-symbol feedback er derfor garanteret at konvergere målt i KL divergensen.

De foreslåede DM modtagerstrukturer sammenlignes med en af de bedste eksisterende løsninger. Det illustreres ved hjælp af Monte Carlo simuleringer at når systembelastningen er lav opnår DM3 modtageren samme bitfejlsandsynlighed som andre af de bedste eksisterende løsninger. Når systembelastningen er høj opnår DM3 modtageren derimod bedre resultater end andre løsninger, selv når konkurrerende løsninger antages at kende støjvariansen og dette ikke er tilfældet for DM3 modtageren. Som forventet er DM3 modtageren bedre end DM5 modtageren når systembelastningen er stor; den sidstnævnte kan understøtte færre brugere end den førstnævnte ved et givent signalstøjforhold.

Acronyms

List of Acronyms

3GPP	3rd Generation Partnership Project
APP	*A Posteriori* Probability
BP	Belief Propagation
CDMA	Code-Division Multiple Access
CE	Channel Estimation
DM	Divergence Minimization
ECM	Expectation Conditioned Maximization
EDGE	Enhanced Data rates for GSM Evolution
EM	Expectation-Maximization
Est.	Estimated
E-step	Expectation step
GEM	Generalized Expectation-Maximization
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
HSDPA	High Speed Downlink Packet Access
IC	Interference Cancellation
ISI	Inter-Symbol Interference
KL	Kullback Leibler
LMMSE	Linear Minimum Mean-Square Error
LS	Least Squares
LTE	Long Term Evolution
MAI	Multiple Access Interference
MAP	Maximum *A Posteriori*
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
M-step	Maximization step
MSEE	Mean Square Estimation Error
NE	Noise matrix Estimation

OFDM	Orthogonal Frequency Division Multiplexing
PDA	Probabilistic Data Association
pdf	probability density function
PIC	Parallel Interference Cancellation
PSK	Phase-Shift Keying
r.h.s.	right-hand side
SAGE	Subspace-Alternating Generalized Expectation-maximization
SIC	Successive Interference Cancellation
SISO	Soft-Input Soft-Output
SNR	Signal-to-Noise Ratio
SU	Single User
SVD	Singular Value Decomposition
TDMA	Time-Division Multiple Access
TD-SCDMA	Time-Division Synchronous CDMA
VBEM	Variational Bayesian Expectation-Maximization
VEM	Variational Expectation-Maximization
w.r.t.	with respect to
WCDMA	Wideband CDMA
ZF	Zero Forcing

Chapter 1

Introduction

With the increasing demand for high data rates in wireless communication systems, the essential technologies that are employed in these systems are experiencing rapid evolution. The development of the cellular communications started with the global system for mobile communications (GSM) (2G) which paved later the path to general packet radio service (GPRS) system and enhanced data rates for GSM evolution (EDGE)(2.5G), all of which are based on time-division multiple access (TDMA). Recently, networks based on code-division multiple access (CDMA), wideband CDMA (WCDMA) or time-division synchronous CDMA (TD-SCDMA) (3G) and their evolution high speed downlink packet access (HSDPA) (3.5G), have become more and more commercially mature. Lately, the orthogonal frequency division multiplexing (OFDM) based long term evolution (LTE) (3.9G) system has been selected by the 3rd generation partnership project (3GPP) as the international standard of next generation wireless communications.

There are many ways to improve date rates in communication systems. Some researchers focus on network level studies, aiming at improving the overall capacity of the cellular network. Another parallel approach is to conduct link level studies, i.e., search for improvements in the performance at link level. This approach can be based on either joint transmitter and receiver study or on receiver study only.

In this thesis, I only consider the link level performance improvement arising from using more advanced receiver architectures. More specifically, the scope of the thesis is to investigate more efficient algorithms for channel parameter estimation and user data decoding in CDMA receivers. Here the term channel parameters include the channel weights and the noise covariance matrix. As a result, a holistic systematic variational Bayesian approach – divergence minimization (DM) – is formulated and

its application to the design of CDMA receivers is presented.

Even though this thesis deals with the problem of multiuser decoding in presence of unknown channels in CDMA systems, the results and the methodology used to solve this problem can be generalized to other applications that involve both error-control coding and multiple-access signaling. For example, the methods and the insights presented in the thesis can be applied to the problem of multiuser MIMO decoding in OFDM systems.

1.1 Background

The optimal solution to the problem of multiuser decoding in CDMA systems can be obtained by maximizing the posterior distribution of the code symbols given the observations or maximizing the likelihood function of the code symbols when the codewords are equiprobable. Maximum likelihood (ML) decoding of coded CDMA was first suggested in [21], where it was shown that the computational complexity of the decoding process grows exponentially with the product of the number of active users and the effective code constraint length. This complexity is prohibitive for any practical implementations. Therefore, much work has been devoted to search for suboptimal solutions.

1.1.1 Iterative Multiuser Decoding

Conventionally, the optimal multiuser decoder is replaced by a multiuser detector followed by a bank of single-user decoders. The multiuser detector and the single-user decoders operate in a sequential manner. One decade ago, the discovery of turbo codes and the turbo decoding method [7] has lead the design of multiuser decoders into a new era: the turbo era. Inspired by the decoding of serial concatenated interleaved codes, the most popular suboptimal approaches so far have been based on the so-called canonical iterative turbo multiuser decoding algorithm [50]. These schemes iterate between a soft-input soft-output (SISO) multiuser detector, which ignores coding constraints and individual posterior distribution decoders, which ignore residual multiuser interference. The attribute "soft" means that, instead of hard decisions on the symbols, the distributions or other quantities describing the state of knowledge of the symbols are used as tentative decisions over the iterations.

The task of the multiuser detector is to separate the users' signals in the received signal. Multiuser detection schemes for uncoded data have

been intensively studied in [57]. Some of the methods in [57] can be extended to accept soft inputs and provide soft outputs.

The optimal SISO multiuser detector is the so-called multiuser posterior distribution detector. The output of this detector are the marginals of the joint posterior distribution of all transmitted symbols computed while neglecting the code constraints, i.e., the code symbols are assumed to be independent. Iterative multiuser decoders that exchange soft information between the SISO multiuser posterior distribution detector and the single-user decoders are proposed in [45, 51]. However, the complexity of the multiuser posterior distribution detector is still exponential in the number of users. Thus, an abundance of suboptimal multiuser detectors have been studied for implementation in iterative multiuser decoders.

One family of suboptimal multiuser detectors applies linear filtering to remove MAI. The linear filter can be the matched filter, the linear minimum mean square error (LMMSE) filter or the least square (LS) (also known as zero forcing (ZF)) filter. These linear filters are applied to the received signal to separate the signals of individual users. The posterior distributions of the code symbols are computed based on the output of the linear filters, usually under the assumption that these outputs are Gaussian distributed. To calculate the LMMSE or the LS filter coefficients, a matrix of a size equal to the number of users, K, needs to be inverted. The complexity of the matrix inversion, in this case, is of the order $O(K^3)$, which is very high for any practical use. Furthermore, this family of detectors is not able to accept soft information of other users from previous iterations. Thus, these detectors cannot be extended to be SISO multiuser detectors.

Another family of suboptimal multiuser detectors is based on linear interference cancellation (IC). These detectors use some kind of tentative decisions on the code symbols of some of or all other users to calculate the interference signal for the user of interest and then suppress this signal from the received signal. The IC can be performed for all users at the same time, a process called parallel IC, or sequentially for each user, a process called successive IC. The interference cancellation can be performed only once per user or several times for each user to improve the detection accuracy. When the symbols of the same user are updated more than once, i.e., the symbols are estimated iteratively, the detectors are called iterative detectors. Iterative multiuser detectors were intensively studied in [18, 17, 16, 48]. In these works the hard decisions on the code symbols are fed back when calculating the MAI. These iterative multiuser detectors achieve performance similar to that of the previously discussed linear multiuser detectors, with complexity of the order $O(Km_{max})$, where m_{max}

is the maximum number of iterations [50]. One of the big advantages of the IC schemes is that they can easily accept soft information as input, e.g., the posterior distributions of the code symbols. This feature of the IC schemes makes iterative multiuser decoders exchanging soft information between an IC block and a bank of single-user decoders very popular.

Low complex receiver structures based on intuitive arguments have also been proposed as a tractable alternative to the receivers that iterate between the posterior distribution multiuser detector and the single-user decoders. As already mentioned, IC structures are well integrated in iterative joint multiuser decoding [2, 3, 46]. To further remove the residual interference due to the errors in the estimation of the soft symbols, LMMSE filtering was subsequently proposed in [58, 20, 10]. The resulting structure is linear IC followed by instantaneous LMMSE filtering [58]. Such a structure provides better performance at the expense of an increased complexity. This approach was justified based on probabilistic data association (PDA) arguments in [55, 56, 54]. Another method to improve the bit-error-rate performance of IC structure considers the correlation between the IC outputs in two consecutive iterations [15].

Multiuser decoding becomes more complicated when the channel parameters, e.g., channel weights and noise variance., need to be estimated. In this case, the optimal solution is to formulate the joint posterior distribution of the channel parameters and code symbols, marginalize this distribution with respect to the channel parameters and finally, apply a maximum *a posteriori* (MAP) method to the posterior probability of the code symbols.

The schemes suggested in [34, 41, 59, 1, 42] are based on a pre-conceived structure consisting of a channel estimator, a multiuser interference cancellation detector and a bank of single-user channel decoders. The iterative structure composed of these three components makes sense intuitively, but lacks a formal justification.

1.1.2 Theoretical Frameworks for Iterative Receiver Design

In very recent years, a few formal optimization frameworks have been applied to systematic design of iterative multiuser decoders. The ultimate goal of these frameworks is to approximate the posterior distribution of code symbols in such a way that the marginalization of the joint posterior distribution of the channel parameters and code symbols is easy to compute and the posterior distribution of the code symbols is easy to maximize.

Factor Graphs and Belief Propagation The joint posterior distribution of the channel parameters and code symbols can be represented on a factor graph. The marginalization with respect to either the channel parameters or the code symbols can be performed on the factor graph. When the factor graph has no cycle, the belief propagation (BP) algorithm can be used to obtain the exact marginal posterior distributions. However, for the problem of multiuser decoding the factor graph has cycles; thus, using the BP algorithm on this graph results in approximations of the marginal posterior distributions of the code symbols. The canonical decoders [45, 51] have been recognized to be approximations of the BP algorithm [39] applied to the factor graph for coded CDMA [9]. The graph related to the multiuser posterior distribution detector and the graph related to the single-user posterior distribution decoders appear to be two connected regions on the factor graph [9]. By passing extrinsic values between the detection and the decoding regions, the approximated marginal posterior distributions of the code symbols can be obtained. Thus the BP algorithm on factor graph provides justification for the canonical decoder structure. Usually for a large number of users (high system load), the BP algorithm is too complex for calculating the messages in the multiuser detection region on factor graph. The receiver structures [2, 58, 20] simplify the calculation of messages in the multiuser detection region.

EM and Related Methods The traditional expectation-maximization (EM) method and the space-alternating generalized EM (SAGE) method have been applied for many years to approximate the MAP or ML estimates. The sequence of likelihoods of the updated estimates computed in the EM algorithm is guaranteed to be non-decreasing over the iterations. Some degree of rigorous justification has been obtained by applying the EM and SAGE frameworks to the receiver design. In the applications of the EM and SAGE to CDMA receivers, the channel weights can be estimated either in the expectation (E)-step or the maximization (M)-step. When the channel weights are estimated in the E-step, as in [35, 37, 38], the resulting algorithms perform hard decisions due to the maximization of the cost function calculated in the E-step. When the channel is estimated in the M-step, as in [34], the E-step provides soft decisions on the code symbols. However, the complexity in this case is the same as that of the optimal multiuser decoder. Thus, some approximations have to be performed here to reduce the complexity. In conclusion, it is not possible to formally use soft symbols instead of the hard decisions with a

reasonable complexity within the rigorous EM/SAGE framework. The soft symbol versions of EM and SAGE are therefore based on modifications, violating the original framework in order to accommodate soft symbols [11, 28, 26]. As a consequence, none of these formal design frameworks can incorporate posterior distribution decoding as part of the rigorous optimization process.

Variational Inference Inference methods [32, 31] were proposed to approximate the joint posterior distribution of all unknown variables in a stochastic model. In contrast to the EM-like frameworks, which compute point estimates, the variational inference methods estimate the marginal posterior distributions of unknown variables. Approximations of the unknown distributions are computed by minimizing a variational free energy over the iterations. The variational free energy iteratively converges to a local minimum. The variational free energy is equivalent to the Kullback-Leibler (KL) divergence (also called information divergence or cross entropy). The KL divergence measures the similarity between a postulated auxiliary distribution and the desired joint posterior distribution. The resulting distribution minimizing the KL divergence also minimizes the variational free energy.

One approach of variational inference is based on the Bethe approximation. Yedidia *et. al.* [8] showed that, when the Bethe approximation for the variational free energy is considered, the zero-gradient points of the Bethe free energy are the stationary points of the BP algorithm [60, 61] applied to the corresponding factor graph. The Bethe free energy, however, is not a convex function of the marginal posterior distribution. Thus, the zero-gradient points are not guaranteed to minimize the variational free energy. In the problem of multiuser decoding in CDMA, the Bethe approximation can be applied for free energy minimization. However, as a result, the messages from the multiuser posterior distribution detector require the computation of the joint posterior distribution for the uncoded case which is intractable complexity-wise [9, 54, 56]. Further simplifications of the message computation were introduced in [54, 56], leading to messages being determined by a chip-based cancellation structure similar to the schemes proposed in [55].

When the postulated auxiliary distribution is restricted to distributions that can be factorized, the minimization of the variational free energy becomes particularly tractable. This is the so-called mean-field approximation. A generalized framework is obtained based on

variational Bayesian inference, which was proposed by Attias in [4] for model selection. Beal subsequently introduced the concept of complete data (known from the EM algorithm [14]) into the variational Bayesian framework to formulate the variational Bayesian EM (VBEM) algorithm in [6]. This variational Bayesian framework can be seen as an extension of the variational inference and the variational EM (VEM) algorithm proposed in [47]. The EM and SAGE algorithms, representing a formal optimization framework in their own rights, were shown to be special cases of the VEM/VBEM algorithms. For single-user MIMO systems, Christensen *et al.* proposed an iterative receiver based on the VBEM algorithm for data and channel parameter estimation [12]. In [43] the BP algorithm was applied to message passing in coded CDMA, where multiuser detection is based on the variational inference framework. Based on particular postulated auxiliary distributions, the structures in [2] and [58] were formally justified as solutions to variational energy minimization (equivalent to divergence minimization) approach.

In summary, two intuitive design principles have been studied: minimization of the Bethe free energy (or BP algorithms) and minimization of the variational free energy based on mean-field approximations. These principles can be used to justify the pre-conceived iterative receiver structures.

1.2 Outline of the Thesis

Motivated by the lack of a formal optimization framework for handling soft symbol processing and inspired by the development of the VBEM algorithm [6], this thesis proposes and develops such a design framework, the so-called divergence minimization (DM) method, based on the KL divergence.

First let us clarify some terminology used for the rest of the thesis. The channel parameters include the channel weights and the noise variance or noise covariance matrix. Channel weight estimation is referred to estimation of the channel weights. The noise estimators perform estimation of the noise variance inverse or the inverse of noise covariance matrix.

The joint posterior distribution of the channel parameters and code symbols is approximated with an auxiliary distribution which fulfills some pre-defined constraint. The KL divergence between these two distributions is minimized over the iterations in the iterative DM scheme. As for the VBEM framework, the auxiliary distribution is constrained to

some class of factorizable distributions. Depending on the selected factorizations of the auxiliary distribution, iterative receivers with different structures are obtained within the framework.

In the thesis we have selected two different factorizations of the auxiliary distribution. Receiver I, derived based on the first factorization, performs joint-user channel weight estimation, noise variance inverse estimation, successive interference cancellation and single-user decoding. Receiver II, obtained with the second factorization, is very similar to Receiver I, except that it performs separate-user channel weight estimation, instead of joint-user channel weight estimation. The thesis provides a rigorous derivation of the iterative structure for every choice of the factorization of the auxiliary distribution within the DM framework. It shall be noted that as a natural result from this derivation, posterior distribution based soft symbols are passed between the single-user decoders, the interference cancellation and the channel weight estimation components.

Our holistic approach is different from the approaches aiming at optimizing a pre-conceived structure, as suggested in [34, 41, 59]. Our resulting receiver has the same structural form as receivers previously suggested in literature. This result once again justifies the "well-established" pre-conceived structure that exchanges soft information among the channel parameter estimator, the multiuser detector and the single-user decoders. In addition to the justification of the structure, the type of the soft information is also theoretically justified. Receiver I is less complex and it achieves slightly better performance than the state-of-the-art structure suggested in [59].

1.3 Major Contributions of the Thesis

The main contribution in this thesis is to formalize a systematic, holistic framework for designing advanced iterative receivers that perform channel parameter estimation and multiuser decoding. The application of the framework to the design of CDMA receivers has the following remarkable advantages:

- The developed framework is based entirely on divergence minimization. It updates the auxiliary distribution to obtain a closer approximation to the desired posterior distribution at every iteration. No prior assumption is made regarding the receiver structure. The factorization constraint on the auxiliary distribution is the only prior

constraint required for applying the formal optimization framework. It determines the structure of the resulting receiver.

- The complexity of the resulting receiver can be adjusted by means of the degree of the selected factorization. In general, the deeper the level of factorization of the auxiliary distribution, the lower the complexity; however, at the cost of a performance loss. For example, the DM receiver with joint-user channel weight estimation outperforms the receiver with separate-user channel weight estimation. However, the former is more complex than the latter.
- The DM provides a formal optimization framework since the sequence of the divergences between the auxiliary distributions returned by the iterative structures and the true posterior distribution is non-decreasing over the iterations. Finally, the estimated auxiliary distribution typically converges to an approximation to the true posterior distribution.
- As a practical outcome of the formal optimization framework, we obtain iterative algorithms that perform joint channel weight estimation and decoding of coded CDMA, taking into account all the first and the second central moments of the unknown variables as well as the imposed code constraints.

The new proposed receivers provide new insights into iterative advanced multiuser decoding with parameter estimation:

- So far, when BP algorithm is applied on the factor graph, the extrinsic values shall be used for updating the messages [9]. Within the DM framework, this is not the case. As a result of applying the DM method, the posterior distributions of the code symbols, rather than the extrinsic values, are forwarded by the single-user posterior distribution decoders to the channel estimator, the estimator of the inverse of noise covariance matrix and the IC device, which in turn forwards extrinsic values to the single-user posterior distribution decoders.
- The residual interference after IC is implicitly handled in the estimation of the inverse of the noise covariance matrix. This again results directly from applying the formal DM framework. This noise estimator is an alternative solution to the commonly used LMMSE filter for residual interference mitigation after interference cancellation applied in other receivers proposed in the literature.
- The inference process includes the covariance matrices of the auxiliary distributions. In this way, the accuracy of the channel weight estimates and the code symbol estimates are taken into

account. Therefore, the DM receiver is a generalization of many receivers published in the literature, which make use of point estimates for these quantities and therefore discard the quality or accuracy of the estimates during the iterations.

- Finally, in the DM receivers the divergence between the auxiliary distribution and the target posterior distribution is guaranteed to converge. Previously suggested receiver structures that use soft-decision code symbols do not enjoy any guaranteed convergence property.

The thesis is partly based on the following publications:

- B. Hu, I. Land, L.K. Rasmussen, R. Piton and B. H. Fleury, A Variational Bayesian Inference Approach to Joint Multiuser Decoding for Coded CDMA. *IEEE Journal on Selected Areas in Communications - Special Issue on Multiuser Detection*, Mar. 2008. [29]

- B. Hu, I. Land, R. Piton and B. H. Fleury, A Bayesian Framework for Iterative Channel Estimation and Multiuser Decoding in Coded DS-CDMA. *Proceedings of the 50th IEEE Global Communications Conference*, Washington D.C., USA, Nov. 2007. [27]

- B. Hu, I. Land, R. Piton and B. H. Fleury, Iterative SAGE-Based Receivers for Synchronous Coded DS-CDMA. *Proceedings of the 66th Semi-Annual IEEE Vehicular Technology Conference (VTC)*, Baltimore, Maryland USA, Oct. 2007. [28]

- B. Hu, A. Kocian, R. Piton and B. H. Fleury, Performance of a SISO-SAGE Based Receiver for Coded CDMA. *Proceedings of the Winter School on Coding and Information Theory*, Bratislava, Feb. 2005. [25]

- B. Hu, A. Kocian, R. Piton, A. Hviid, B. H. Fleury and L.K. Rasmussen, Iterative Joint Channel Estimation and Interference Cancellation Using a SISO-SAGE Algorithm for Coded DS-CDMA. *Proceedings of the 38th IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2004. [26]

- B. Hu, A. Kocian, P. Sørensen, C. Rom and B. H. Fleury, Iterative Joint Data Detection and Channel Estimation of DS-CDMA Signals in Multipath Fading Using the SAGE Algorithm. *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2003. [36]

1.4 Organization of the Thesis

The remainder of the thesis is organized as follows: In Chapter 2 we discuss the basic concepts of the DM approach and its relationship to the VBEM, the EM and the SAGE algorithms. The system model for a synchronous CDMA system operating in flat fading channels is described in Chapter 3. In Chapter 4 we apply the DM algorithm with the first factorization constraint on the auxiliary distribution to derive the first receiver for the scenario described in Chapter 3. The resulting receiver performs joint-user channel weight estimation, noise covariance matrix inverse estimation, successive interference cancellation and single-user decoding. Its structure is compared to that of the state-of-the-art receiver [59]. The performance of both receivers are compared by means of Monte Carlo simulations. The receiver performing separate-user channel weight estimation generated based on the second constraint on the auxiliary distribution is derived and its performance is investigated in Chapter 5. Summary and conclusions are given in Chapter 6.

Chapter 2

A Variational Bayesian Framework – Divergence Minimization

This chapter provides a self-contained introduction to the variational Bayesian framework and the so-called divergence minimization (DM) method. The DM method is a simple, efficient and general method for solving estimation problems in the presence of nuisance parameters.

2.1 Parameter Estimation in the Presence of Nuisance Parameters

In a stochastic model, the variables to be estimated are called the parameters of interest. The other unknown variables are called the nuisance parameters[1]. Usually, in order to estimate the parameters of interest based on a set of observations, the nuisance parameters are either estimated or integrated out.

Let $\boldsymbol{\theta}$ be the vector containing all the parameters of interest and $\boldsymbol{\eta}$ denote the vector of all nuisance parameters in a stochastic model. The parameters of interest and the nuisance parameters can be continuous variables or discrete variables or of both (mixed) types. In the rest of the thesis, we assume that all parameters in the stochastic model are continuous at the first place. The difference in the formulation arising from considering discrete variables will be pointed out only when necessary.

[1] In statistical inference, a nuisance parameter is a random variable which is fundamental to the probabilistic model, but which is not of particular interest in itself.

Finding the estimate of $\boldsymbol{\theta}$ which minimizes the error probability given the observation $\boldsymbol{r}$ is equivalent to maximizing the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{r})$, i.e., computing

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{r}). \tag{2.1}$$

Eq. (2.1) is the expression of the so-called MAP estimator of the parameters of interest $\boldsymbol{\theta}$.

When a model contains nuisance parameters, the MAP estimation of the parameter vector $\boldsymbol{\theta}$ should be conducted in three steps: I) The joint posterior distribution of all unknown random variables (including both the parameters of interest and the nuisance parameters) given the observation $\boldsymbol{r}$, $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{r})$, shall be obtained first; II) Marginalization over the nuisance parameters is performed to obtain the posterior distribution of the parameters of interest only. Mathematically, this step can be written as

$$p(\boldsymbol{\theta}|\boldsymbol{r}) = \int d\boldsymbol{\eta} p(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{r}); \tag{2.2}$$

III) Finally, the maximization step in (2.1) is conducted.

Both marginalization and maximization usually require extensive computational efforts. A variational Bayesian approach is introduced in the next section to iteratively approximate the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{r})$. As a by-product, this approach provides an approximation of the posterior marginal distribution of the parameters of interest, $p(\boldsymbol{\theta}|\boldsymbol{r})$ as well. Compared to the same operation in (2.1), the maximization of the approximated distribution is much more simplified.

2.2 Divergence Minimization

Dating back to 18^{th} century, Euler, Lagrange and others developed the calculus of variations, which is the core of the variational approaches. Recently in 1999, a variational approach was applied to the Bayesian framework for model selection in [32, 4]. A stochastic model is defined by certain parameters (including both the parameters of interest and the nuisance parameters) and by the dependence of the observation on these parameters. For each candidate model, all model parameters are inferred and the likelihood of each model is calculated. Eventually, the model with the largest likelihood is selected. The model selection process is called model inference.

For the problem of parameter estimation stated in Section 2.1, we would like to approximate the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{r})$ by an auxiliary distribution $q(\boldsymbol{\theta}, \boldsymbol{\eta})$. Thereby, differently from [4], we only need parameter inference since the model has already been specified by the parameters $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{r}$ and the posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{r})$. Inspired by the variational Bayesian expectation-maximization (VBEM) method [6] which provides analytical calculation of the posterior distribution of the parameters of interest and the nuisance parameters, we introduce a similar method called divergence minimization (DM) to obtain an estimate of the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{r})$ in an iterative way.

2.2.1 KL Divergence

The KL divergence is an appropriate, well-known criterion to measure the discrepancy between any two distributions. The KL divergence is also called cross-entropy, directed divergence or discrimination information [40]. In the field of physics, the terminology free energy is used.

The KL divergence between the auxiliary distribution $q(\boldsymbol{\eta}, \boldsymbol{\theta})$ and the target distribution $p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})$ is defined as [13]

$$D\Big(q(\boldsymbol{\eta}, \boldsymbol{\theta})||p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big) \triangleq \int_{\boldsymbol{\eta},\boldsymbol{\theta}} d\boldsymbol{\eta} d\boldsymbol{\theta}\; q(\boldsymbol{\eta}, \boldsymbol{\theta}) \log \frac{q(\boldsymbol{\eta}, \boldsymbol{\theta})}{p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})}. \tag{2.3}$$

An auxiliary distribution that is optimal in the sense of the KL divergence, minimizes the divergence in (2.3)[2]:

$$\hat{q}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \arg \min_{q(\boldsymbol{\eta},\boldsymbol{\theta})\in\mathcal{Q}} D\Big(q(\boldsymbol{\eta}, \boldsymbol{\theta})||p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big). \tag{2.4}$$

In (2.4) $\mathcal{Q}$ denotes a specified class of auxiliary distributions. The choice of $\mathcal{Q}$ reflects constraints imposed on the auxiliary distribution. Notice that any auxiliary distribution $q(\boldsymbol{\eta}, \boldsymbol{\theta})$ satisfies

$$\int_{\boldsymbol{\eta},\boldsymbol{\theta}} d\boldsymbol{\eta} d\boldsymbol{\theta} q(\boldsymbol{\eta}, \boldsymbol{\theta}) = 1 \quad \text{and} \quad q(\boldsymbol{\eta}, \boldsymbol{\theta}) \geq 0.$$

When $\mathcal{Q}$ is the set of all possible joint distributions of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, the divergence (2.3) is minimized for $q(\boldsymbol{\eta}, \boldsymbol{\theta}) = p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})$, in which case the divergence vanishes[3]. However, this solution does not simplify the

[2] A related method is information geometry, where the divergence $D(p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})||q(\boldsymbol{\eta}, \boldsymbol{\theta}))$ is minimized instead. For more details, we refer the reader to [30].

[3] The divergence between q and p satisfies

$$D(q||p) \geq 0$$

original problem. In order to reduce the complexity of the minimization problem in (2.4), we may impose additional constraints on the auxiliary distribution. For example, we may apply the Bethe approximation to the integral at the r.h.s. of (2.3) as shown in [61] or we may apply a structured mean-field approximation to the auxiliary distribution $q(\boldsymbol{\eta}, \boldsymbol{\theta})$, as shown in [6]. Using the Bethe approximation results in the BP algorithm, which still have a high computational complexity. With the structured mean-field approximation, the auxiliary distribution belongs to the class of distribution that factorize according to a certain scheme. As a result, the derivation of the divergence minimization is rather simple and systematic, and furthermore, the resulting algorithms have a low complexity and a good convergence property. Thus, we will focus on the structured mean-field approximations in our application.

2.2.2 Factorization of the Auxiliary Distribution

Adopting a structured mean-field approximation [6], the auxiliary distribution is constrained to be factorizable as

$$q(\boldsymbol{\eta}, \boldsymbol{\theta}) = q_\eta(\boldsymbol{\eta}) q_\theta(\boldsymbol{\theta}) \tag{2.5}$$

where both $q_\eta(\boldsymbol{\eta})$ and $q_\theta(\boldsymbol{\theta})$ are arbitrary distributions or are required to fulfill some constraints. Notice that for arbitrary distributions,

$$\int d\boldsymbol{\eta} q_\eta(\boldsymbol{\eta}) = 1,\ q_\eta(\boldsymbol{\eta}) \geq 0, \qquad \int d\boldsymbol{\theta} q_\theta(\boldsymbol{\theta}) = 1,\ q_\theta(\boldsymbol{\theta}) \geq 0.$$

On the one hand, the factorization implies that an independence assumption is imposed between the parameters of interest $\boldsymbol{\theta}$ and the nuisance parameters $\boldsymbol{\eta}$. The factorization in (2.5) makes the solution (2.4) suboptimal, i.e., the estimated distribution will not equal the true posterior distribution in most cases. Usually, the more factors the auxiliary distribution contains, the further the estimated distribution is away from the true distribution. An extreme case of a structured mean-field approximation is the so-called mean-field approximation or a naive mean-field approximation. In the mean-field approximation, the joint auxiliary distribution factorizes in the product of as many factors as the number of model parameters, with one factor per parameter. Thus, no further factorization is possible when the mean-field approximation is used. Compared to the mean-field approximation, the structured mean-field

with equality if and only if $q = p$.

approximation retains certain dependency between certain parameters by carefully selecting the factorization of the auxiliary distribution.

On the other hand, the advantage of the factorization of the auxiliary distribution is that it makes the computation of marginal distributions from the approximated joint distribution much easy to compute. For example, let $q_\eta(\boldsymbol{\eta})q_\theta(\boldsymbol{\theta})$ be the estimated distribution. Then the posterior distribution of the parameter $\boldsymbol{\theta}$ can be approximated by

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \int d\boldsymbol{\eta} p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{y}) \approx \int d\boldsymbol{\eta} q_\eta(\boldsymbol{\eta}) q_\theta(\boldsymbol{\theta}) = q_\theta(\boldsymbol{\theta}). \tag{2.6}$$

To further simplify the calculations, $q_\eta(\boldsymbol{\eta})$ and $q_\theta(\boldsymbol{\theta})$ may be further factorized. Different factorizations lead to different estimators with different trade-offs between estimation accuracy and computational complexity.

2.2.3 The DM Method

In this subsection, we derive the DM method considering the auxiliary distribution in (2.5). The generalization to an auxiliary distribution with more than two factors is straightforward.

Given the auxiliary distribution in (2.5), the KL divergence between the auxiliary distribution and the target distribution is

$$D\Big(q_\eta(\boldsymbol{\eta})q_\theta(\boldsymbol{\theta})\|p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big) = \int\int d\boldsymbol{\eta} d\boldsymbol{\theta} q_\eta(\boldsymbol{\eta}) q_\theta(\boldsymbol{\theta}) \log \frac{q_\eta(\boldsymbol{\eta})q_\theta(\boldsymbol{\theta})}{p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})}. \tag{2.7}$$

The goal of the DM method is to find $q_\eta(\boldsymbol{\eta})$ and $q_\theta(\boldsymbol{\theta})$ that minimize the above divergence in an iterative manner. First, an initial distribution $q_\eta(\boldsymbol{\eta})$ is selected and the divergence in (2.7) is minimized with respect to $q_\theta(\boldsymbol{\theta})$ for this initial setting. Then, by fixing the estimate of $q_\theta(\boldsymbol{\theta})$, the divergence is minimized with respect to $q_\eta(\boldsymbol{\eta})$. In this way, the estimate of $q_\eta(\boldsymbol{\eta})$ is updated. This iterative process is the so-called DM method.

Formally, at iteration $i+1$, $i > 0$, of the DM method the following two steps are performed:

Step 1 (*VB expectation step (E-step)* [6]) The auxiliary distribution of the parameters of interest is kept fixed to its current setting $q_\theta^{[i]}(\boldsymbol{\theta})$, and the auxiliary distribution of the nuisance parameters, $q_\eta(\boldsymbol{\eta})$, is estimated by solving

$$q_\eta^{[i+1]}(\boldsymbol{\eta}) = \arg\min \; D\Big(q_\theta^{[i]}(\boldsymbol{\theta})q_\eta(\boldsymbol{\eta})\|p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big) \tag{2.8}$$
$$s.t. \int d\boldsymbol{\eta} q_\eta(\boldsymbol{\eta}) = 1, \; q_\eta(\boldsymbol{\eta}) \geq 0.$$

Applying functional derivatives and Lagrange multipliers, one obtains the solution

$$q_\eta^{[i+1]}(\boldsymbol{\eta}) \propto \exp\Big[\int d\boldsymbol{\theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big]. \tag{2.9}$$

The basic definitions and formulas for functional derivatives are stated in Appendix A.1. The derivation steps from (2.8) to (2.9) are provided in Appendix A.2.

Step 2 (*VB maximization step (M-step)* [6]) [4] The auxiliary distribution of the nuisance parameters is kept fixed, and the auxiliary distribution of the parameters of interest, $q_\theta(\boldsymbol{\theta})$, is updated by solving

$$\begin{aligned} q_\theta^{[i+1]}(\boldsymbol{\theta}) = \arg\min\ & D\Big(q_\theta(\boldsymbol{\theta}) q_\eta^{[i+1]}(\boldsymbol{\eta}) \| p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big) \\ s.t.\ & \int d\boldsymbol{\theta} q_\theta(\boldsymbol{\theta}) = 1,\ q_\theta(\boldsymbol{\theta}) \geq 0. \end{aligned} \tag{2.10}$$

The solution is similar to (2.9):

$$q_\theta^{[i+1]}(\boldsymbol{\theta}) \propto \exp\Big[\int d\boldsymbol{\eta} q_\eta^{[i+1]}(\boldsymbol{\eta}) \log p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big]. \tag{2.11}$$

Since $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are symmetric in the minimization of the objective functional, the derivation of (2.11) is exactly the same as that of (2.9).

Since the divergence is minimized at each step, the divergence is non-increasing over the iterations:

$$\begin{aligned} D\Big(q_\eta^{[i]}(\boldsymbol{\eta}) q_\theta^{[i]}(\boldsymbol{\theta}) \| p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big) &\geq D\Big(q_\eta^{[i+1]}(\boldsymbol{\eta}) q_\theta^{[i]}(\boldsymbol{\theta}) \| p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big) \\ &\geq D\Big(q_\eta^{[i+1]}(\boldsymbol{\eta}) q_\theta^{[i+1]}(\boldsymbol{\theta}) \| p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big). \end{aligned} \tag{2.12}$$

Thus, the overall algorithm is guaranteed to converge in KL divergence.

2.3 DM and Expectation-Maximization (EM)

The DM method presented above approximates the joint posterior distribution iteratively. The EM algorithm [14] approximates the MAP or ML

[4] In the DM method, both steps involve expectation and maximization operations. Thus, it is inappropriate to call any step either the E-step or the M-step.

estimates iteratively. In this section, the DM method is applied in the case where the class Q of auxiliary distributions consists of Dirac delta functions. The concept of "complete data" is also introduced in the DM method. It is shown that the EM method is an instance of the DM method when the auxiliary distributions are constrained to Q.

2.3.1 The EM as an Instance of DM

Assume that the distribution of the parameters in (2.5) is a Dirac delta function:

$$q_\theta(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{2.13}$$

Notice that estimating this distribution is equivalent to estimating the value $\boldsymbol{\theta}_0$. Inserting (2.13) in (2.7), the two steps in (2.9) and (2.11) reduce to:
Step 1 (VB E-step):

$$q_\eta^{[i+1]}(\boldsymbol{\eta}) = p(\boldsymbol{\eta}|\boldsymbol{\theta}_0^{[i]}, \boldsymbol{r}); \tag{2.14}$$

Step 2 (VB M-step):

$$\boldsymbol{\theta}_0^{[i+1]} = \arg\min_{\boldsymbol{\theta}_0} \left(-\frac{1}{C_{\boldsymbol{\theta}}^{[i+1]}} \exp\left[\int d\boldsymbol{\eta} q_\eta^{[i]}(\boldsymbol{\eta}) \log p(\boldsymbol{\eta}, \boldsymbol{\theta}_0|\boldsymbol{r}) \right] \right) \tag{2.15}$$

where $C_{\boldsymbol{\theta}}^{[i]}$ is a non-negative constant, see more details in Appendix A.3. Notice that Step 2 can equivalently be written as

$$\boldsymbol{\theta}_0^{[i+1]} = \arg\max_{\boldsymbol{\theta}_0} E_{q_\eta^{[i]}} \left\{ \log p(\boldsymbol{\eta}, \boldsymbol{r}|\boldsymbol{\theta}_0) \right\} + \log p(\boldsymbol{\theta}_0). \tag{2.16}$$

2.3.2 EM Algorithm

The instance of the DM method in Subsection 2.3.1 is very similar to the EM algorithm [14]. In the EM algorithm, the concept of complete data is employed to help maximization of the likelihood function. Let $\boldsymbol{x}$ denote the complete data. Thus according to the definition of a complete data, the random vectors $\boldsymbol{r}$, $\boldsymbol{x}$ and $\boldsymbol{\theta}$ form a Markov chain, i.e., $p(\boldsymbol{r}, \boldsymbol{\theta}|\boldsymbol{x}) = p(\boldsymbol{\theta}|\boldsymbol{x})p(\boldsymbol{r}|\boldsymbol{x})$. To give an example, $\boldsymbol{x} = \{\boldsymbol{\eta}, \boldsymbol{r}\}$ is a complete data for estimating the parameters of interest. The EM algorithm computes, at each iteration, the following two steps:

E-step :

$$Q_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\boldsymbol{\theta}^{[i]}) = \int p(\boldsymbol{x}|\boldsymbol{\theta}_0^{[i]}, \boldsymbol{r}) \log p(\boldsymbol{x}|\boldsymbol{\theta}_0^{[i]}) d\boldsymbol{x}; \tag{2.17}$$

M-step :

$$\boldsymbol{\theta}_0^{[i+1]} = \arg\max_{\boldsymbol{\theta}_0} Q_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\boldsymbol{\theta}^{[i]}). \tag{2.18}$$

The selection of the complete data $\boldsymbol{x}$ guarantees that the likelihood function $p(\boldsymbol{r}|\boldsymbol{\theta}^{[i+1]})$ is non-decreasing over the iterations [14].

If the complete data is selected such that $\boldsymbol{x} = \{\boldsymbol{\eta}, \boldsymbol{r}\}$, the resulting estimates $\boldsymbol{\theta}_0^{[i]}$ for the DM in Subsection 2.3.1 and for the EM after each iteration are the same. Step 1 (VB E-step) provides the conditional probability distribution for the expectation calculated in Step 2 (VB M-step). In the original EM algorithm, this expectation is evaluated in the E-step. Therefore, for this particular selection of the complete data, the DM algorithm reduces to the EM algorithm provided the auxiliary distribution of the parameters of interest is restricted to be a Dirac delta function. The question is whether it is possible to show that for any selection of a complete data, the EM algorithm is an instance of the DM method? To verify this, we need to introduce the concept of complete data in the DM framework.

2.3.3 Complete Data in the DM Algorithm

For any random variables $\boldsymbol{x}$, the posterior distribution of the parameters can be obtained by computing

$$p(\boldsymbol{\theta}|\boldsymbol{r}) = \int d\boldsymbol{x} p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{r}). \tag{2.19}$$

In addition to approximating the joint posterior distribution of the parameters of interest and nuisance parameters, the DM method can be employed to approximate any joint posterior distribution. For example, the joint posterior distribution $p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{r})$ can be approximated by an auxiliary distribution

$$q(\boldsymbol{x}, \boldsymbol{\theta}) = q_{\mathrm{x}}(\boldsymbol{x}) q_{\theta}(\boldsymbol{\theta}). \tag{2.20}$$

Substituting the complete data $\boldsymbol{x}$ in (2.14) and (2.16), while making use of the Markov chain property of $\boldsymbol{\theta}$, $\boldsymbol{x}$ and $\boldsymbol{r}$ when $\boldsymbol{x}$ is a complete data in (2.16), yields

Step 1:

$$q_{\mathrm{x}}^{[i+1]}(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{\theta}_0^{[i]}, \boldsymbol{r}); \tag{2.21}$$

Step 2:

$$\boldsymbol{\theta}_0^{[i+1]} = \arg\max_{\boldsymbol{\theta}_0} E_{q_{\mathrm{x}}^{[i]}}\left\{\log p(\boldsymbol{x}|\boldsymbol{\theta}_0)\right\} + \log p(\boldsymbol{\theta}_0). \tag{2.22}$$

Step 1 provides the auxiliary distribution used to compute the expectation in Step 2. In the original EM method, this expectation is evaluated in the E-step, while the maximization in Step 2 is performed in the M-step. Thus the EM algorithm can be considered an instance of the DM method.

2.4 Related Theoretical Frameworks

In this section we present theoretical estimation frameworks closely related to the DM method. They include the EM, SAGE [19], variational EM (VEM)[47], generalized EM (GEM) [14] and expectation conditioned maximization (ECM) [44] methods. Fig. 2.1 depicts the relationship and the main differences between these frameworks.

(1) From VBEM to DM

The VBEM method is proposed for model selection in [6]. In order to select the model which gives the largest likelihood, all parameters in the model are inferred. In the DM method, the goal is to approximate the joint posterior distribution of all parameters, which implies that only one model is considered. The updating steps for inferring the parameters are the same in both methods.

(2) From DM to EM

As shown in Section 2.3, the EM algorithm results from the DM algorithm by constraining the auxiliary distribution $q_\theta(\boldsymbol{\theta})$ to belong to the class of Dirac delta function. As a result, the estimation of a distribution reduces to a point estimation. In that sense, the EM algorithm is an instance of the DM method.

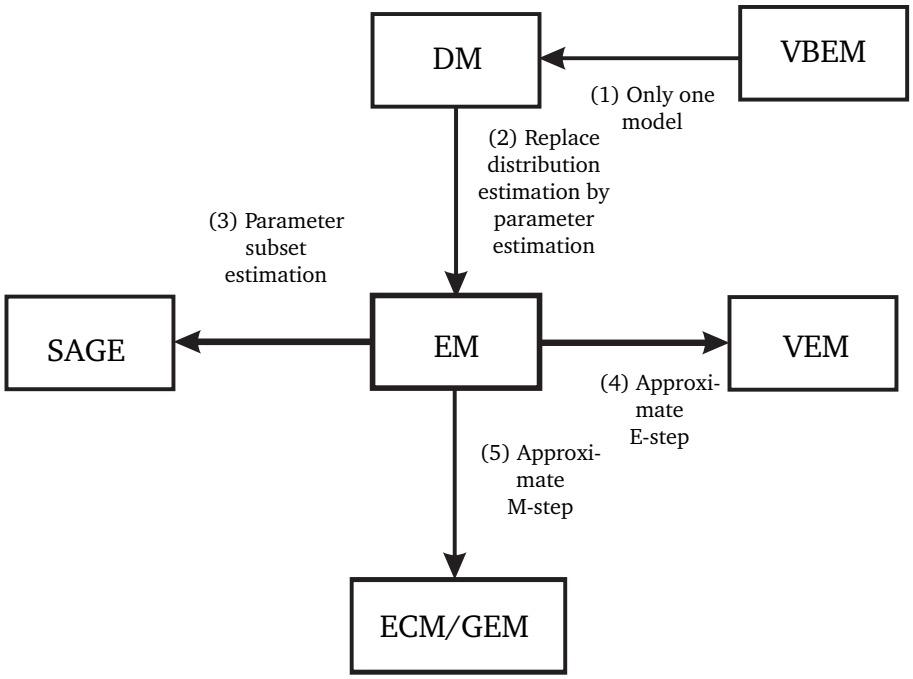

Figure 2.1: Overview of the estimation frameworks related to DM

(2) From EM to SAGE

When the process of updating all parameters at each iteration in the EM algorithm is computationally prohibitive, the SAGE method can be used to update only a subset of parameters at each iteration. In the SAGE method [19], the concept of admissible hidden data is introduced for each subset of the parameters based on the statistical structure of the likelihood function. To update a subset of the parameters at each iteration, we need to select corresponding admissible hidden data. The admissible hidden data for such a subset is a complete data for this subset when the other parameters are known. Then, the distribution of the admissible hidden data is estimated. Finally, the distribution of the admissible hidden data is maximized with respect to the subset of the parameters. Hidden data is selected in a way that the maximization of the estimated distribution can be easily performed. Similarly to the EM algorithm, the sequence of the likelihoods of the parameters estimates is non-decreasing over the iterations in the SAGE algorithm. More details about the SAGE algorithm can be found in [19]. Another difference between the EM and SAGE frameworks is that in the former the mapping of the complete data to

the incomplete (observation) data is deterministic, while in the latter the mapping of the admissible hidden data to the incomplete data may be stochastic.

(3) From EM to VEM

When the calculation of (2.17) in the E-step of the EM algorithm is too complex, the DM method can be employed to approximate $p(\boldsymbol{x}|\boldsymbol{\theta}^{[i]}, \boldsymbol{r})$.

Let $q'(\boldsymbol{x})$ denote an auxiliary distribution satisfying $\int_{\boldsymbol{x}} d\boldsymbol{x} q'(\boldsymbol{x}) = 1$ and $q'(\boldsymbol{x}) \geq 0$. Consider the KL divergence

$$D\Big(q'(\boldsymbol{x})||p(\boldsymbol{x}|\boldsymbol{\theta}^{[i]}, \boldsymbol{r})\Big) = \int_{\boldsymbol{x}} d\boldsymbol{x} q'(\boldsymbol{x}) \log \frac{q'(\boldsymbol{x})}{p(\boldsymbol{x}|\boldsymbol{\theta}^{[i]}, \boldsymbol{r})}. \tag{2.23}$$

With the auxiliary distribution restricted to factorize as

$$q'(\boldsymbol{x}) = \prod_i q'_i(x_i),$$

an approximation of $p(\boldsymbol{x}|\boldsymbol{\theta}^{[i]}, \boldsymbol{r})$ can be obtained by using the DM method in Subsection 2.2.3. Notice that there are two levels of iterations: the inner-level iteration to approximate the exact result of the E-step; the outer-level iteration to perform the approximated E-step and the exact M-step.

The VEM method does not converge in likelihood as the EM algorithm does, since the posterior distribution used in the E-step is not exact.

(4) From EM to GEM/ECM

When the maximization procedure in the standard M-step of the EM algorithm (2.18) is intractable (e.g., when the dimension of the parameters of interest is too large), one can resort to the GEM method [14]. This method obtains a new estimate by increasing the r.h.s. of (2.18) instead of maximizing it. The sequence of the likelihoods of the parameter estimates is also non-decreasing. However, the convergence rate of the GEM method is slower than that of the EM algorithm.

The ECM is a subclass of GEM. In the ECM algorithm, the M-step in (2.18) is replaced by a conditioned maximization step (CM-step). The basic idea of the CM-step is that the maximization is performed iteratively. The parameters of interest can be divided into more than one subsets. Instead of maximization over the entire set of parameters, the

maximization can be done over one subset of the parameters, conditioned on keeping the other subsets unchanged. After the maximization process have been performed with respect to all subsets of parameters, an update of the parameters of interest is obtained. Note that the objective function in the M-step is not always convex. Thus, the CM-step may not generate the minimum value of the divergence defined in (2.18). However, (2.18) is increased as in the GEM algorithm. More details about the ECM method can be found in [44].

Even though a subset of the parameters is updated at each iteration in the sub-maximization-steps of both the GEM/ECM methods and the SAGE algorithms, the calculations of their E-steps are different. As stated above, the selection of admissible hidden data in the E-step of the SAGE algorithm makes the sequence of likelihoods of the estimates returned by the method non-decreasing over the iterations. As a result, the sequence of likelihoods converges to a stable point much faster. This property is not necessarily fulfilled in the GEM and ECM algorithms.

In summary, several closely related theoretical estimation frameworks exist, which can serve to design iterative estimation and detection algorithms. Each can be suitable for specific applications. Apparently, the DM method is the most general tool among them. Thus, we use it in this thesis to design iterative receivers for CDMA.

Chapter 3

System Models for CDMA

This chapter describes the symbol notation and the signal model that will be used in the rest of the thesis.

3.1 Notation

In this thesis, we shall make use of the following notation. Vectors are denoted by boldface lowercase letters, e.g., $\boldsymbol{x}$, and matrices by boldface uppercase letters, e.g., $\boldsymbol{X}$. The i-th element of a vector $\boldsymbol{x}$ is denoted by either x_i or $\{\boldsymbol{x}\}_i$; the element in the i-th column and the j-th row of a matrix $\boldsymbol{X}$ reads either x_{ij} or $\{\boldsymbol{X}\}_{i,j}$. For scalars, $(\cdot)^*$ denotes complex conjugate, and for vectors and matrices, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote the transpose and the Hermitian transpose, respectively; $\mathrm{tr}\{\cdot\}$ denotes the trace, $\mathrm{diag}\{\boldsymbol{x}\}$ denotes a diagonal matrix with the elements of $\boldsymbol{x}$, and $\mathrm{Diag}\{\boldsymbol{X}\}$ denotes a diagonal matrix with the diagonal elements of $\boldsymbol{X}$. An estimate at iteration i is denoted by $(\cdot)^{[i]}$. Two kinds of proportionality are used: $x \propto y$ denotes $x = \alpha y$, and $x \propto^e y$ denotes $e^x = e^\beta e^y$, i.e., $x = \beta + y$ for real random variables x, y and arbitrary constants $\alpha, \beta \in \mathbb{R}$. Throughout the thesis, we use the natural logarithm, i.e., $\log x = \log_e x$. Finally, $E_{q_x}\{f(\boldsymbol{x})\}$ denotes the expectation of the function $f(\boldsymbol{x})$ with respect to the distribution $q_{\mathrm{x}}(\boldsymbol{x})$ of the random variable $\boldsymbol{x}$.

3.2 System Model

As mentioned in Chapter 1, in this thesis we deal with the problem of multiuser decoding in the presence of unknown channel parameters for CDMA systems. For a wideband system such as CDMA, the channel appears frequency selective in most cases. However, in this thesis, only

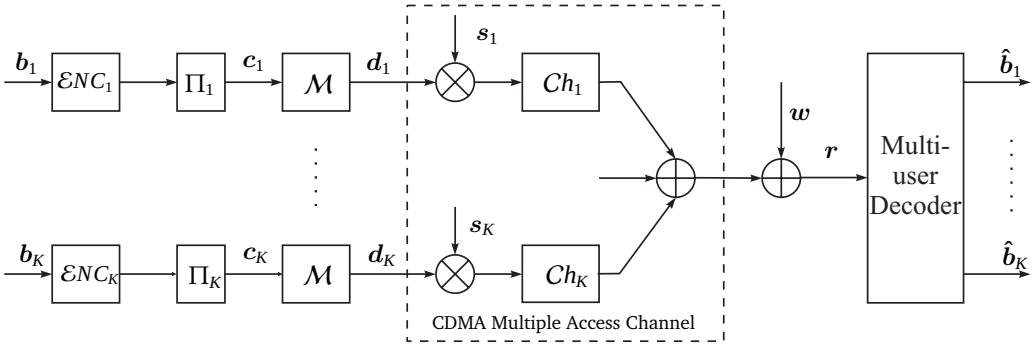

Figure 3.1: Block diagram of a DS-CDMA system ($\mathcal{E}NC_k$: encoder of user k; Π_k: interleaver of user k; $\mathcal{M}$: modulator; Ch_k: propagation channel of user k).

flat fading channel is considered. The reason for this choice is that this work is aiming at more theoretical findings, and simplifying the original problems can help gain more insights into the iterative process for data decoding and channel parameters.

In the rest of the thesis, we consider a direct sequence (DS)-CDMA system with K active users as shown in Fig. 3.1. Let $b_k[m] \in \{0,1\}$ denote the m-th information bit of user k, and let the column vector $\boldsymbol{b}_k = [b_k[0], \ldots, b_k[M-1]]^{\mathrm{T}}$ stand for the information bit sequence of user k. All information bits are independent and uniformly distributed. The information sequence $\boldsymbol{b}_k$ is convolutionally encoded and interleaved to yield the interleaved code sequence $\boldsymbol{c}_k$. Let C_k denote the set of interleaved codeword for user k, $\boldsymbol{c}_k \in C_k$. All K users employ codes with identical rate R_c but different interleavers. The interleaved code sequences are then mapped to the sequence of modulation symbol $\boldsymbol{d}_{k,c}$. Let $c_k[l] \in \{0,1\}$ and $d_k[l]$ denote the l-th code bit and the l-th transmission symbol of user k, respectively.

In the case of BPSK modulation, each interleaved code bit $c_k[l]$ is mapped onto a BPSK symbol $d_k[l] \in \{-1,+1\}$ with the mapping rule: $d_k[l] = 1 - 2c_k[l]$. Let $\mathcal{D}_k^B$ be the set of valid BPSK symbol sequences for user k, i.e., $\boldsymbol{d}_{k,c} \in \mathcal{D}_k^B$.

In the case of QPSK modulation, Gray mapping is considered. Each two interleaved code bits, $c_k[2l]$ and $c_k[2l+1]$, are mapped onto one QPSK modulation symbol $d_k[l] \in \{\pm 1 \pm j, \pm 1 \mp j\}$ using the mapping rule (see Fig. 3.2)

$$d_k[l] = (2c_k[2l] - 1) + j(2c_k[2l+1] - 1).$$

We denote the set of valid QPSK symbol sequence of user k by $\mathcal{D}_k^Q$, i.e., $\boldsymbol{d}_{k,c} \in \mathcal{D}_k^Q$.

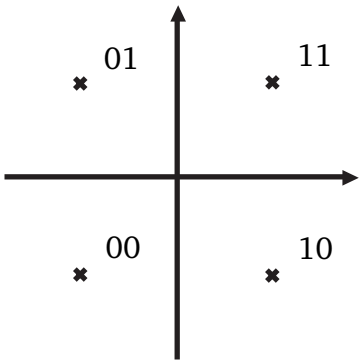

Figure 3.2: QPSK Gray mapping

Hereinafter, the generalized notation $\mathcal{D}_k$ is used for the set of modulation symbols for user k which could be either BPSK symbols or QPSK symbols. The modulation type will be specified when necessary. For the purpose of channel weight estimation, each (modulation) symbol sequence is multiplexed with L_p random pilot symbols placed at the beginning of the sequence. The column vector $\boldsymbol{d}_k = [\boldsymbol{d}_{k,p}\ \boldsymbol{d}_{k,c}]^{\mathrm{T}} = [d_k[0], \ldots, d_k[L-1]]^{\mathrm{T}}$ denotes the transmitted sequence of user k. The sequence $\boldsymbol{d}_k$ consists of the pilot symbol sequence $\boldsymbol{d}_{k,p}$[1] and the data symbol sequence $\boldsymbol{d}_{k,c}$ of length L_c, where $L = L_c + L_p$.

Each symbol modulates the signature waveform $s_k(t)$ with support $[0, T)$. The embedded spreading sequence $\boldsymbol{s}_k[l]$ for the l-th transmission symbol has the energy $\int_0^T |s_k(t)|^2 dt = 1$.

In the channel, each transmitted sequence of each user experiences block memoryless fading. The received signal is the superposition of the signals of all K users and white Gaussian noise.

Let the column vector $\boldsymbol{r}[l] = [r_1[l], \ldots, r_{N_c}[l]]^{\mathrm{T}}$ denote the output of the chip-matched filters at signaling interval l. Here N_c denotes the spreading factor. The (chip-rate) signal model is given by

$$\boldsymbol{r}[l] = \boldsymbol{S}[l]\boldsymbol{A}\boldsymbol{d}[l] + \boldsymbol{w}[l],\ l = 0, \ldots, L-1. \tag{3.1}$$

In this expression, $\boldsymbol{S}[l]$ is the spreading sequence matrix at the signaling interval l, i.e., the k-th column of $\boldsymbol{S}[l]$ is $\boldsymbol{s}_k[l] \in \{-1, +1\}^{N_c}$. Note that applying a real spreading sequence to complex QPSK symbols implies that the in-phase and the quadrature parts of these symbols are spread by the same sequence. In some applications, the in-phase and the quadrature parts of QPSK symbols are spread using different sequences,

[1] $\boldsymbol{d}_{k,p} \in \{-1, +1\}^{L_p}$ in case of BPSK and $\boldsymbol{d}_{k,p} \in \{\pm 1 \pm j, \pm 1 \mp j\}^{L_p}$ in case of QPSK.

which may improve the performance of the demodulator because of the extra diversity obtained from distinct spreading sequences. In that case, however, the demodulation is more complex. In the thesis, the same spreading sequence is used to modulate both in-phase and quadrature parts.

Let a_k denote the channel weight of user k for the current transmission block, and let $\boldsymbol{A} = \mathrm{diag}\{a_1, \ldots, a_K\}$. The column vector $\boldsymbol{d}[l] = [d_1[l], \ldots, d_K[l]]^{\mathrm{T}}$ contains the transmission symbols of all K users at signaling interval l. The noise vector $\boldsymbol{w}[l]$ is Gaussian complex circularly-symmetric with covariance matrix $\sigma^2 \boldsymbol{I}_{N_c}$, where $\sigma^2 = N_0/E_b$ and $\boldsymbol{I}_{N_c}$ is the identity matrix with dimensions $N_c \times N_c$. The symbol N_0 denotes the spectral height of white noise; E_b is the energy per BPSK symbol and $E_s = 2E_b$ is the energy per QPSK symbol.

In addition, an equivalent symbol-rate signal model is introduced for later convenience:

$$\boldsymbol{z}[l] = \boldsymbol{R}[l]\boldsymbol{D}[l]\boldsymbol{a} + \boldsymbol{n}[l], \; l = 0, \ldots, L-1, \tag{3.2}$$

where $\boldsymbol{D}[l] \triangleq \mathrm{diag}\{d_1[l], \ldots, d_K[l]\}$, $\boldsymbol{a} \triangleq [a_1, \ldots, a_K]^{\mathrm{T}}$ and $\boldsymbol{n}[l]$ is complex circularly symmetric Gaussian with covariance matrix $\sigma^2 \boldsymbol{R}[l]$, $\boldsymbol{R}[l] = \boldsymbol{S}[l]^{\mathrm{H}} \boldsymbol{S}[l]$. The i,j-th element $\{\boldsymbol{R}[l]\}_{ij} = \rho_{ij}[l]$ of the matrix $\boldsymbol{R}[l]$ is the cross-correlation between the signature waveforms of users i and j, i.e., $\rho_{ij}[l] = \boldsymbol{s}_i^{\mathrm{H}} \boldsymbol{s}_j$. Furthermore, we define for convenience: $\boldsymbol{z} \triangleq \Big[\boldsymbol{z}[0]^{\mathrm{T}}, \boldsymbol{z}[1]^{\mathrm{T}}, \ldots, \boldsymbol{z}[L-1]^{\mathrm{T}}\Big]^{\mathrm{T}}$, $\boldsymbol{R} \triangleq \mathrm{diag}\{\boldsymbol{R}[0], \ldots, \boldsymbol{R}[L-1]\}$, $\boldsymbol{D} \triangleq \Big[\boldsymbol{D}[0], \ldots, \boldsymbol{D}[L-1]\Big]^{\mathrm{T}}$.

Note that the signal models of (3.1) and (3.2) are related as follows,

$$\boldsymbol{z}[l] = \boldsymbol{S}[l]^{\mathrm{H}} \boldsymbol{r}[l].$$

Both $\boldsymbol{r}[l]$ and $\boldsymbol{z}[l]$ in (3.1) and (3.2), respectively, provide equivalent sufficient statistics[2] for decoding the information bits.

3.3 Optimal Multiuser Receiver

The optimal criterion for decoding the transmission bits $\boldsymbol{b}$ of all users is that the error probability conditioned on the observations $\boldsymbol{r}$ or $\boldsymbol{z}$ should

[2]The concept of sufficient statistic is, most generally, defined as follows: a statistic $T(x) = t$ is sufficient for underlying parameter θ if, and only if the conditional pdf of the data x, given the statistic $T(x)$, is independent of the parameter θ, i.e.

$$p(x|t, \theta) = p(x|t)$$

be minimized. The sequence error probability given the observation $\boldsymbol{r}$ is

$$P(\hat{\boldsymbol{b}} \neq \boldsymbol{b}|\boldsymbol{r}) = 1 - P(\hat{\boldsymbol{b}} = \boldsymbol{b}|\boldsymbol{r}),$$

where $\hat{\boldsymbol{b}}$ denote a decision on $\boldsymbol{b}$. The error probability is minimized by selecting the bit sequence estimate maximizing the *a posteriori* distribution:

$$\hat{\boldsymbol{b}}_{\text{MAP}} = \arg \max_{\boldsymbol{b}\in\{0,1\}^{KM}} p(\boldsymbol{b}|\boldsymbol{r}). \tag{3.3}$$

This decision rule is referred to as maximum *a posteriori* (MAP) sequence decoding. The above problem can be solved by exhaustive search methods. Solving (3.3) requires a computational complexity $O(2^{KM})$ which makes its practical implementation infeasible.

In the DS-CDMA system shown in Fig 3.1, the mappings $\boldsymbol{b}_k \rightarrow \boldsymbol{c}_k$ and $\boldsymbol{c}_k \rightarrow \boldsymbol{d}_k$ are one-to-one, i.e., one sequence $\boldsymbol{b}_k$ corresponds to a unique sequence $\boldsymbol{c}_k$ and one unique sequence $\boldsymbol{c}_k$ corresponds to a unique sequence $\boldsymbol{d}_k$. Thus, the optimal detector (3.3) is equivalent to

$$\hat{\boldsymbol{d}}_{\text{MAP}} = \arg \max_{\boldsymbol{d}\in\mathcal{D}} p(\boldsymbol{d}|\boldsymbol{r}), \tag{3.4}$$

where $\mathcal{D} := \left\{[\boldsymbol{d}_1, \ldots, \boldsymbol{d}_K] : \ \boldsymbol{d}_{c,k} \in \mathcal{D}_k, \ k = 1, \ldots, K\right\}$.

We assume that the input information sequence is uniformly distributed, i.e., $p(\boldsymbol{b}_k) = 1/2^M$. Due to the encoding and interleaving process, the distribution of the sequence of symbols at the output of the modulator is of the form

$$p(\boldsymbol{d}_k) = \begin{cases} 1/|\mathcal{D}_k| & \boldsymbol{d}_k \in \mathcal{D}_k \\ 0 & \boldsymbol{d}_k \notin \mathcal{D}_k, \end{cases} \tag{3.5}$$

where $|\mathcal{D}_k|$ denotes the cardinality of $\mathcal{D}_k$. Notice that the distribution of the modulated code symbol sequence is not uniform when considering all possible discrete sequences of length L. Further a specific encoder, an interleaver and a modulation scheme determine the set $\mathcal{D}_k$. Thus, (3.5) represents the constraints (so-called coding and modulation constraints) that are imposed by the channel encoder and the modulator on the symbol sequence $\boldsymbol{d}_k$.

Given that the transmission sequences are independent, the posterior distribution of $\boldsymbol{d}$ is proportional to the likelihood function of $\boldsymbol{d}$ in the range $\mathcal{D}$ and zero otherwise:

$$p(\boldsymbol{d}|\boldsymbol{r}) = p(\boldsymbol{r}|\boldsymbol{d}) \prod_k p(\boldsymbol{d}_k) \propto \begin{cases} p(\boldsymbol{r}|\boldsymbol{d}) & \boldsymbol{d} \in \mathcal{D}. \\ 0 & \boldsymbol{d} \notin \mathcal{D}. \end{cases} \tag{3.6}$$

Now assume that the channel weights $\boldsymbol{a}$ and the noise variance σ^2 are known. From (3.1), since noise is Gaussian distributed, the logarithm of the likelihood function of $\boldsymbol{d}$ can be written as

$$\log p(\boldsymbol{r}|\boldsymbol{d}) = \sum_{l=0}^{L-1} \log p(\boldsymbol{r}[l]|\boldsymbol{d}[l]) = L\log(\pi\sigma^{-2}) - \sigma^{-2}\sum_{l=0}^{L-1}\left\|\boldsymbol{r}[l] - \boldsymbol{S}[l]\boldsymbol{A}\boldsymbol{d}[l]\right\|^2$$

$$\propto 2\sigma^{-2}Re\left\{\sum_{l=0}^{L-1}\sum_{k=1}^{K} a_k d_k[l]\left(z_k[l]^* - \sum_{k'=1}^{K}\rho_{kk'}[l]a_{k'}^* d_{k'}[l]^*\right)\right\}. \tag{3.7}$$

Direct maximization of (3.7) with respect to the constraints $\boldsymbol{d} \in \mathcal{D}$ requires an exhaustive search, which is prohibitively complex. Therefore, we propose to use the DM method to find an estimate of the likelihood function (3.7).

3.4 Applications of the DM Method for CDMA Receivers

Many suboptimal multiuser decoders have been discussed in Chapter 1 where it was pointed out that the motivation of this thesis is the lack of theoretical justifications for the design of iterative receiver algorithms. In Chapter 2, the divergence minimization framework was introduced. The potential application of this framework to the design of CDMA receiver is as follows.

The divergence minimization framework is based on a factorization of the joint auxiliary distribution (or structured mean-field approximation). By employing different factorizations of the auxiliary distribution for the signal model of a coded CDMA system, different iterative receiver algorithms can be obtained.

In the next chapter we apply the DM method to the CDMA signal model (3.1) where the unknown parameters are the transmitted symbols $\boldsymbol{d}$, the channel weights $\boldsymbol{a}$ and the inverse of the covariance matrix $\boldsymbol{\Sigma}_w^{-1}$ of the additive noise. Due to the structure of the likelihood function, $p(\boldsymbol{r}|\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Sigma}_w)$, it is more convenient to work with the inverse of the noise covariance matrix than with the noise covariance matrix. During the optimization process in the DM method, approximates of the posterior marginal distributions are obtained. In the end, the estimates of the code bits can be obtained based on the approximated posterior marginal distribution of the code symbols.

To approximate the posterior joint distribution $p(\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Sigma}_w^{-1}|\boldsymbol{r})$, it is natural to consider the following factorization of the auxiliary distribution:

$$q(\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Sigma}_w^{-1}) = q_{\mathrm{a}}(\boldsymbol{a})q_{\Sigma_{\mathrm{w}}^{-1}}(\boldsymbol{\Sigma}_w^{-1})q_{\mathrm{d}}(\boldsymbol{d}). \tag{3.8}$$

However, the estimation of the distribution of all code sequences is still very complicated. Therefore it is necessary to further factorize $q_{\mathrm{d}}(\boldsymbol{d})$ in a product of auxiliary distributions of the symbol sequences of all K users. Doing so leads to

$$q(\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}) = q_{\mathrm{a}}(\boldsymbol{a}) q_{\Sigma_{\mathrm{w}}^{-1}}(\boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}) \prod_{k=1}^{K} q_{\mathrm{d}_{\mathrm{k}}}(\boldsymbol{d}_k). \tag{3.9}$$

In Chapter 4 we show how to obtain an iterative receiver based on the factorization in (3.9). The resulting receiver performs joint-user channel weight estimation, noise covariance matrix inverse estimation and successive interference cancellation and single-user decoding.

Starting from (3.9), a further factorization on the auxiliary distribution of the channel weights can be done:

$$q(\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}) = \prod_{k} q_{\mathrm{a}_{\mathrm{k}}}(a_k) q_{\Sigma_{\mathrm{w}}^{-1}}(\boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}) \prod_{k=1}^{K} q_{\mathrm{d}_{\mathrm{k}}}(\boldsymbol{d}_k). \tag{3.10}$$

Chapter 5 presents the derivation of an iterative receiver based on this factorization. Differently from the joint-user channel estimator obtained from (3.9), the channel weight estimation, in this case, is performed user-wise. Thus, the receiver in Chapter 5 performs iterative separate-user channel weight estimation, noise covariance matrix inverse estimation and successive interference cancellation and single-user decoding.

Chapter 4

DM Receiver with Joint-User Channel Estimation

In this chapter, an application of the DM method to a synchronous CDMA system operating in flat fading channels is described. In this application, the iterative receiver is derived based on the factorization in (3.9). The resulting iterative algorithm performs joint-user channel estimation, noise covariance matrix inverse estimation, successive interference cancellation and single-user decoding. The performance of the DM receiver is compared to that of a state-of-the-art ad-hoc receiver by means of Monte Carlo simulations.

For this particular application to CDMA receiver design, the algorithm is derived assuming that the variance of the white noise is unknown. The expression when the noise variance is known can easily be obtained. Interestingly, it will be shown that the receiver assuming unknown noise variance has a better BER performance than the one having knowledge of the variance.

4.1 Factorization of the Auxiliary Distribution

Let us define the new symbol $\boldsymbol{\Xi} \equiv \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}$. Then, we rewrite the auxiliary model in (3.9) as follows:

$$q(\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Xi}) = q_{\mathrm{a}}(\boldsymbol{a}) q_{\Xi}(\boldsymbol{\Xi}) \prod_{k=1}^{K} q_{\mathrm{d_k}}(\boldsymbol{d}_k),$$

where $\boldsymbol{a}$, $\boldsymbol{d}$ and Ξ denote the channel weights, the code sequences of all users and the inverse of the noise covariance matrix, respectively. The factor graph representation of the above auxiliary distribution is depicted

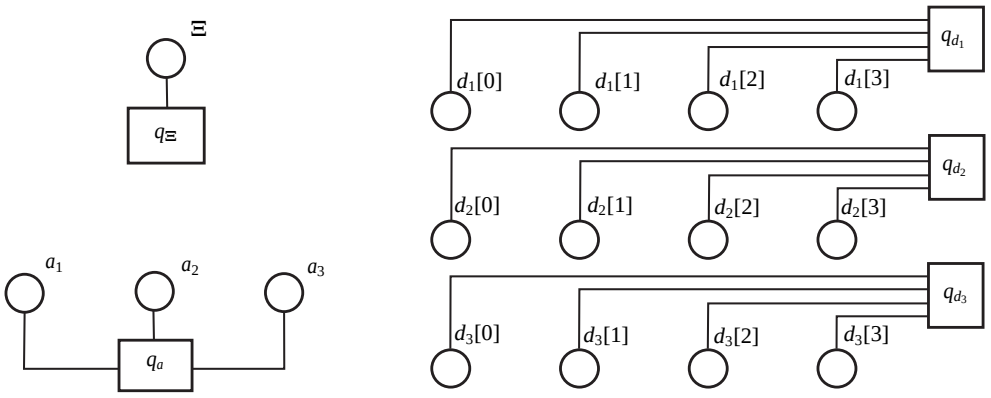

Figure 4.1: Factor graphs of the structured mean-field approximation implied by the factorization (3.9). Circles represent variable nodes; rectangles are functional nodes.

in Fig. 4.1. The following assumptions from (3.9) are reflected in the graph structure:

- The variable nodes for the transmitted symbols of different users are not connected, i.e., the multiple access constraint is relaxed.
- The variable nodes for the transmitted symbols of each user are connected, i.e., the coding constraints are taken into account:

$$q(\boldsymbol{d}_k)\begin{cases}\neq 0 & \boldsymbol{d}_k \in \mathcal{D}_k \\ = 0 & \boldsymbol{d}_k \notin \mathcal{D}_k.\end{cases} \tag{4.1}$$

- The variable nodes for the channel weights of all users are connected, i.e., the channel estimation is performed jointly over all users.

As stated in Chapter 2, the factorization (3.9) makes the integral of the joint auxiliary distribution with respect to the parameters very simple since $q(\boldsymbol{a})$, $q(\boldsymbol{\Xi})$, $q(\boldsymbol{d}_1)$, …, $q(\boldsymbol{d}_K)$ are marginals of $q(\boldsymbol{d}, \boldsymbol{a}, \boldsymbol{\Xi})$. Thus, the resulting $q(\boldsymbol{d}_k)$ for the k-th user is considered an estimate of $p(\boldsymbol{d}_k|\boldsymbol{r})$.

One might consider the full factorization as suggested in [49], i.e.,

$$q_{\mathrm{d_k}}(\boldsymbol{d}_k) = \prod_l q_{\mathrm{d_k[l]}}(d_k[l]),\ d_k[l] \in \{+1, -1\}, l = L_p, \ldots, L-1. \tag{4.2}$$

However, (4.2) implies that the elements $\{d_k[l] : l = L_p, \ldots, L-1\}$ are independent, i.e., discards the facts that the sequence $\boldsymbol{d}_k$ is a code

sequence, $\boldsymbol{d}_k \in \mathcal{D}_k$. Thus, the full factorization in (4.2) breaks down the code constraints and leads to the following inconsistency:

$$q_{\mathrm{d_k}}(\boldsymbol{d}_k) \neq \prod_l q_{\mathrm{d_k[l]}}(d_k[l]),\ \boldsymbol{d}_k \in \mathcal{D}_k,$$

i.e., if $q_{\mathrm{d_k}}(\boldsymbol{d}_k)$ is a valid distribution of the code sequence $\boldsymbol{d}_k$, $q_{\mathrm{d_k[l]}}(d_k[l])$ can not be a valid marginal distribution of the code symbol $d_k[l]$. Due to the above inconsistency, the full factorization will not be considered in this application.

4.2 Derivation of the DM Receiver

Given the auxiliary distribution in (3.9), the KL divergence to be minimized is

$$\begin{aligned} &D\Big(q(\boldsymbol{a},\boldsymbol{\Xi},\boldsymbol{d})\Big\|p(\boldsymbol{a},\boldsymbol{\Xi},\boldsymbol{d}|\boldsymbol{r})\Big) \\ &\qquad = D\Big(q_{\mathrm{a}}(\boldsymbol{a})q_{\Xi}(\boldsymbol{\Xi})\prod_{k=1}^{K} q_{\mathrm{d_k}}(\boldsymbol{d}_k)\Big\|p(\boldsymbol{a},\boldsymbol{\Xi},\boldsymbol{d}_1,\ldots,\boldsymbol{d}_K|\boldsymbol{r})\Big). \end{aligned} \tag{4.3}$$

Starting from some initial estimates of the auxiliary distributions, we update one of the distributions $q_{\mathrm{a}}(\boldsymbol{a})$, $q_{\Xi}(\boldsymbol{\Xi})$ and $q_{\mathrm{d_k}}(\boldsymbol{d}_k)$, $k = 1,\ldots,K$, at each step. The update step for $q_{\mathrm{a}}(\boldsymbol{a})$ performs joint-user estimation of the channel weights; the update step for $q_{\Xi}(\boldsymbol{\Xi})$ performs the estimation of the inverse of the noise covariance matrix and the update of $q_{\mathrm{d_k}}(\boldsymbol{d}_k)$ performs successive interference cancellation and single-user decoding. We now describe each of these update steps in detail.

4.2.1 Joint-user channel estimation

We first consider minimizing the divergence between (4.3) and $q_{\mathrm{a}}(\boldsymbol{a})$, while keeping $q_{\Xi}^{[i]}(\boldsymbol{\Xi})$ and $q_{\mathrm{d}}^{[i]}(\boldsymbol{d}) = \prod_{k=1}^{K} q_{\mathrm{d_k}}^{[i]}(\boldsymbol{d}_k)$ fixed, i.e.,

$$\begin{aligned} \text{minimize} \quad & D\Big(q_{\mathrm{a}}(\boldsymbol{a})q_{\Xi}^{[i]}(\boldsymbol{\Xi})q_{\mathrm{d}}^{[i]}(\boldsymbol{d})\Big\|p(\boldsymbol{a},\boldsymbol{\Xi},\boldsymbol{d}|\boldsymbol{r})\Big) \\ \text{subject to} \quad & \int d\boldsymbol{a} q_{\mathrm{a}}(\boldsymbol{a}) = 1 \\ & q_{\mathrm{a}}(\boldsymbol{a}) \geq 0. \end{aligned} \tag{4.4}$$

Solving this optimization problem, one obtains (see Appendix A.2)

$$q_{\mathrm{a}}^{[i+1]}(\boldsymbol{a}) \propto \exp\Big[\int d\boldsymbol{\Xi} q_{\Xi}^{[i]}(\boldsymbol{\Xi}) \sum_{\boldsymbol{d}\in\mathcal{D}} \prod_{k=1}^{K} q_{\mathrm{d_k}}^{[i]}(\boldsymbol{d}_k) \log p(\boldsymbol{a},\boldsymbol{\Xi},\boldsymbol{d}|\boldsymbol{r})\Big]. \tag{4.5}$$

Applying Bayes' rule

$$p(\boldsymbol{a}, \boldsymbol{\Xi}, \boldsymbol{d}|\boldsymbol{r}) = p(\boldsymbol{r}|\boldsymbol{a}, \boldsymbol{\Xi}, \boldsymbol{d})p(\boldsymbol{a})p(\boldsymbol{\Xi})p(\boldsymbol{d})/p(\boldsymbol{r}),$$

discarding the terms independent of $\boldsymbol{a}$, and moving the prior distribution $p(\boldsymbol{a})$ out of the argument of the exponential function in (4.5) yields

$$q_{\mathrm{a}}^{[i+1]}(\boldsymbol{a}) \propto p(\boldsymbol{a}) \exp\left[E_{q_{\Xi}^{[i]}}\left\{E_{q_{\mathrm{d}}^{[i]}}\left\{\log p(\boldsymbol{r}|\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Xi})\right\}\right\}\right]. \tag{4.6}$$

Thus, updating $q_{\mathrm{a}}(\boldsymbol{a})$ can be considered as scaling the prior distribution $p(\boldsymbol{a})$. If no prior information on $\boldsymbol{a}$ is available, the non-informative prior $p(\boldsymbol{a}) \propto 1$ can be used, which corresponds to dropping $p(\boldsymbol{a})$ in (4.6).

Since the noise in (3.1) is Gaussian, the log-likelihood function in (4.6) can be written as

$$\begin{aligned}&\log p(\boldsymbol{r}|\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Xi})\\ &\quad \propto^{e} L \log|\boldsymbol{\Xi}| - \mathrm{tr}\left\{\boldsymbol{\Xi}\sum_{l=1}^{L}\left(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}[l]\boldsymbol{a}\right)\left(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}[l]\boldsymbol{a}\right)^{\mathrm{H}}\right\}.\end{aligned} \tag{4.7}$$

In order to obtain (4.6), the expectation of the log-likelihood function with respect to the distributions $q_{\Xi}^{[i]}(\boldsymbol{\Xi})$ and $\prod_{k=1}^{K} q_{\mathrm{d}_k}^{[i]}(\boldsymbol{d}_k)$ must be computed.

Let us define the soft symbol $\tilde{d}_k^{[i]}[l] \triangleq E_{q_{\mathrm{d}}^{[i]}}\left\{d_k[l]\right\}$. Due to the factorization in (3.9), the second moments of the code symbols are computed to be

$$E_{q_{\mathrm{d}}^{[i]}}\left\{d_k[l]^* d_j[l]\right\} = \begin{cases}(\tilde{d}_k^{[i]}[l])^* \tilde{d}_j^{[i]}[l], & k \neq j\\ 1, & k = j\end{cases}, \tag{4.8}$$

$L = L_p, \ldots, L-1$. Notice that these moments are functions of the soft code symbols. The variances of the soft symbol estimates can be computed as

$$\sigma^2_{d_k^{[i]}[l]} = 1 - (\tilde{d}_k^{[i]}[l])^2 \tag{4.9}$$

for $k = 1, \ldots, K$ and $l = L_p, \ldots, L-1$. The computation of the soft symbols depends on the modulation scheme. The computations for BPSK and QPSK modulations can be found in Section 4.2.3.

Furthermore, we define $\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]} \triangleq E_{q_{\Xi}^{[i]}}\left\{\boldsymbol{\Xi}\right\}$ as the expected value of the inverse of the noise covariance matrix with respect to the distribution $q_{\Xi}^{[i]}(\boldsymbol{\Xi})$. The computation of $\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}$ is described in (4.18), Section 4.2.2.

With $\tilde{d}_k^{[i]}[l]$, $\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}$ and (4.8) we can compute the expectation of (4.7) proceeding in this way yields

$$
\begin{aligned}
&E_{q_{\Xi}^{[i]}}\Big\{E_{q_{\mathrm{d}}^{[i]}}\Big\{\log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\Xi)\Big\}\Big\} \\
&\qquad \propto^{e} -\mathrm{tr}\Big\{\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}\Big(\sum_{l=0}^{L_p}(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a})(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a})^{\mathrm{H}} \\
&\qquad\quad + \sum_{l=1}^{L}(\boldsymbol{r}[l]-\boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{a})(\boldsymbol{r}[l]-\boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{a})^{\mathrm{H}} \\
&\qquad\quad - \sum_{l=1}^{L}\boldsymbol{S}[l]\boldsymbol{E}^{[i]}[l]\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{E}^{[i]}[l])^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Big)\Big\}, \quad (4.10)
\end{aligned}
$$

where $\boldsymbol{D}_p[l] \triangleq \mathrm{diag}\big\{d_{1,p}[l],\ldots,d_{K,p}[l]\big\}$, $l = 0,\ldots,L_p-1$ is the pilot symbol matrix. For $l = L_p,\ldots,L-1$, $\tilde{\boldsymbol{D}}^{[i]}[l] \triangleq \mathrm{diag}\big\{\tilde{d}_1^{[i]}[l],\ldots,\tilde{d}_K^{[i]}[l]\big\}$ is the soft symbol matrix. The error covariance matrix of $\boldsymbol{d}$ is defined as $\boldsymbol{E}^{[i]}[l]\boldsymbol{E}^{[i]}[l]^{\mathrm{H}}$ with $\boldsymbol{E}^{[i]}[l] \triangleq \mathrm{diag}\Big\{\sigma_{d_1^{[i]}[l]},\ldots,\sigma_{d_K^{[i]}[l]}\Big\}$ and $\sigma^2_{d_k^{[i]}[l]}$ defined in (4.9). The derivation of (4.10) is provided in Appendix B.1.

For Rayleigh fading channels, $\boldsymbol{a}$ is Gaussian distributed:

$$
p(\boldsymbol{a}) \propto \exp\Big\{-\boldsymbol{a}^{\mathrm{H}}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}\boldsymbol{a}\Big\}. \quad (4.11)
$$

Exploiting the properties of the trace operator in (4.10), inserting the prior distribution (4.11) into the update expression in (4.6), we obtain a Gaussian distribution for the update of the auxiliary distribution of the channel weight vector:

$$
q_{\mathrm{a}}^{[i+1]}(\boldsymbol{a}) \propto \exp\Big[-(\boldsymbol{a}-\boldsymbol{a}^{[i+1]})^{\mathrm{H}}(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i+1]})^{-1}(\boldsymbol{a}-\boldsymbol{a}^{[i+1]})\Big] \quad (4.12)
$$

with mean vector[1]

$$
\boldsymbol{a}^{[i+1]} = E_{q_{\mathrm{a}}^{[i+1]}}\{\boldsymbol{a}\} = (\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i+1]})^{-1}\sum_{l=0}^{L-1}(\tilde{\boldsymbol{D}}^{[i]}[l])^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}\boldsymbol{r}[l] \quad (4.13)
$$

[1]The mean vector (4.13) corresponds to the conditional mean vector from the MMSE channel estimator in [59].

and covariance matrix inverse

$$\left(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i+1]}\right)^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} + \sum_{l=0}^{L_p-1} (\tilde{\boldsymbol{D}}^{[i]}[l])^{\mathrm{H}} \boldsymbol{S}[l]^{\mathrm{H}} \boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]} \boldsymbol{S}[l] \tilde{\boldsymbol{D}}^{[i]}[l] + \sum_{l=L_p}^{L-1} (\tilde{\boldsymbol{D}}^{[i]}[l])^{\mathrm{H}} \boldsymbol{S}[l]^{\mathrm{H}} \boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]} \boldsymbol{S}[l] \tilde{\boldsymbol{D}}^{[i]}[l] + \sum_{l=L_p}^{L-1} (\boldsymbol{E}^{[i]}[l])^{\mathrm{H}} \mathrm{Diag}\{\boldsymbol{S}[l]^{\mathrm{H}} \boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]} \boldsymbol{S}[l]\} \boldsymbol{E}^{[i]}[l]. \quad (4.14)$$

Detailed derivations of (4.13) and (4.14) are found in Appendix B.1.

In (4.6) the update of the auxiliary distribution $q(\boldsymbol{a})$ is based on the current setting of the other auxiliary distributions. Since $q^{[i+1]}(\boldsymbol{a})$ is a Gaussian distribution, it is sufficient to compute its expectation (4.13) and covariance matrix (4.14). The computation module of (4.13) and (4.14) is depicted in Fig. 4.2. Note that the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i+1]}$ represents the precision in the estimation of the channel weights and $\boldsymbol{\Sigma}_{\boldsymbol{d}}^{[i]}$ represents the precision in the estimation of the code symbols. For constant envelope modulation schemes, $\boldsymbol{\Sigma}_{\boldsymbol{d}}^{[i]}$ can be obtained from $\tilde{\boldsymbol{d}}^{[i]}$.

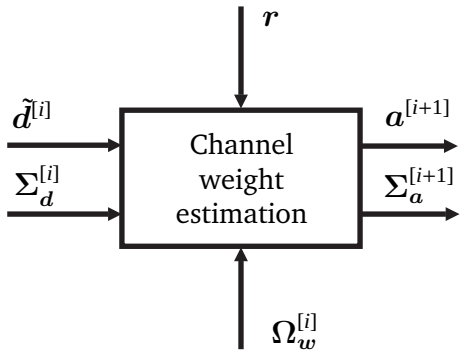

Figure 4.2: Channel weight estimator: Module for updating $q_{\mathrm{a}}(\boldsymbol{a})$.

4.2.2 Estimation of the Inverse of the Noise Covariance Matrix

When updating the distribution of $\boldsymbol{\Xi}$, the distributions $q_{\mathrm{a}}^{[i]}(\boldsymbol{a})$ and $q_{\mathrm{d}}^{[i]}(\boldsymbol{d})$ are kept fixed. The minimum KL divergence is achieved for the distribution

$$q_{\Xi}^{[i+1]}(\boldsymbol{\Xi}) \propto p(\boldsymbol{\Xi}) \exp\left[E_{q_{\mathrm{a}}^{[i]}}\left\{E_{q_{\mathrm{d}}^{[i]}}\{\log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\boldsymbol{\Xi})\}\right\}\right], \quad (4.15)$$

similarly to (4.6). To compute the argument of the exponential term in (4.15), we compute the expectation of (4.7) with respect to the distributions $q_{\mathrm{a}}^{[i]}(\boldsymbol{a})$ and $q_{\mathrm{d}}^{[i]}(\boldsymbol{d})$. The expectation with respect to $q_{\mathrm{d}}^{[i]}(\boldsymbol{d})$ makes use of the soft symbols $\tilde{d}_k[l]^{[i]}$, $k = 1, \ldots, K$, $l = L_p, \ldots, L-1$ and (4.8), while the expectation with respect to $q_{\mathrm{a}}^{[i]}(\boldsymbol{a})$ exploits the results from (4.13) and (4.14). Doing so yields

$$\exp\left[E_{q_{\mathrm{a}}^{[i]}}\left\{E_{q_{\mathrm{d}}^{[i]}}\left\{\log p\left(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\boldsymbol{\Xi}\right)\right\}\right\}\right] \propto |\boldsymbol{\Xi}|^{L} \exp\left[-\mathrm{tr}\left\{\boldsymbol{\Xi}\boldsymbol{B}^{[i]}\right\}\right] \tag{4.16}$$

with

$$\begin{aligned}
\boldsymbol{B}^{[i]} \triangleq & \sum_{l=0}^{L_p-1}\Big(\Big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a}^{[i]}\Big)\Big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a}^{[i]}\Big)^{\mathrm{H}} \\
& \qquad + \boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}\boldsymbol{D}_p[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Big) \\
& + \sum_{l=L_p}^{L-1}\Big(\Big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{a}^{[i]}\Big)\Big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{a}^{[i]}\Big)^{\mathrm{H}} \\
& \qquad + \boldsymbol{S}[l]\boldsymbol{E}^{[i]}[l]\boldsymbol{A}^{[i]}(\boldsymbol{A}^{[i]})^{\mathrm{H}}(\boldsymbol{E}^{[i]}[l])^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}} \\
& \qquad\quad + \boldsymbol{S}[l]\boldsymbol{E}^{[i]}[l]\mathrm{Diag}\{\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}\}(\boldsymbol{E}^{[i]}[l])^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}} \\
& \qquad\quad + \boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}(\tilde{\boldsymbol{D}}^{[i]}[l])^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Big).
\end{aligned} \tag{4.17}$$

The r.h.s. of (4.16) is (up to a proportionality constant) a the complex Wishart distribution[2].

Selecting the non-informative Wishart distribution $W_N(0, \mathbf{0})$ to be the prior distribution in (4.15) and inserting (4.16) into (4.15), it follows that $\boldsymbol{\Xi}$ is Wishart distributed:

$$\boldsymbol{\Xi} \sim W_N(L + N_c, (\boldsymbol{B}^{[i]})^{-1}).$$

The expectation of $\boldsymbol{\Xi}$ is then

$$\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i+1]} \triangleq E_{q_{\Xi}^{[i+1]}}\{\boldsymbol{\Xi}\} = \left(\frac{\boldsymbol{B}^{[i]}}{L+N_c}\right)^{-1}, \tag{4.18}$$

[2]If the distribution of a complex symmetric and positive definite random matrix $A_{p\times p}$ can be written as $f_A(A) \propto |A|^{n-p}\exp[-\mathrm{tr}\{V^{-1}A\}]$, where $n \geq p$ and $V_{p\times p}$ is a constant matrix, A is said to be complex Wishart distributed: $A \sim W_p(n, V)$; the expectation of A is nV [22].

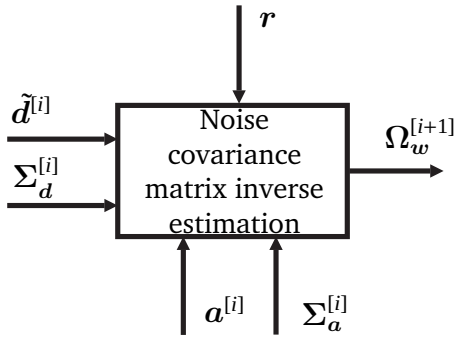

Figure 4.3: Estimator of the inverse of the noise covariance matrix: Module for updating $q_{\Xi}(\boldsymbol{\Xi})$. Note that the code symbol covariance $\boldsymbol{\Sigma}_{\boldsymbol{d}}^{[i]}$ can be obtained from $\tilde{\boldsymbol{d}}^{[i]}$.

where $\boldsymbol{B}^{[i]}$ is given by (4.17).

Since the expectation in (4.18) fully characterizes the distribution of the noise covariance matrix inverse in (4.15), to update the distribution of the noise covariance matrix inverse, the computation of (4.18) is sufficient. Note that the computation of (4.18) only requires the results of (4.13) and (4.14). The module computing $\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}$ is depicted in Fig. 4.3.

Simpler expressions for (4.18) are obtained for white Gausian noise. In this case, $\boldsymbol{\Xi} = \mathrm{diag}\{\sigma^{-2}, \ldots, \sigma^{-2}\}$ with the variance inverse σ^{-2} being chi-square distributed [23]:

$$\sigma^{-2} \sim \chi^2_{LN_c}.$$

We introduce a new symbol for the inverse of the noise variance: $\varsigma \equiv \sigma^{-2}$. The distribution of ς reads

$$q_{\varsigma}^{[i+1]}(\varsigma) \propto (\varsigma)^{LN_c} \exp\Big[-\varsigma \mathrm{tr}\{\boldsymbol{B}^{[i]}\}\Big]. \tag{4.19}$$

Here, we assume a non-informative prior distribution χ^2_0 for ς. With $\boldsymbol{B}^{[i]}$ given in (4.17), the expectation of ς is

$$E_{q_{\varsigma}^{[i+1]}}[\varsigma] = \left(\frac{\mathrm{tr}\{\boldsymbol{B}^{[i]}\}}{LN_c}\right)^{-1}. \tag{4.20}$$

Note that the estimates given in (4.18) and (4.20) coincide with the ML estimates.

4.2.3 Successive Interference Cancellation and Single-User Decoding

Similarly to (4.6) and (4.15), the minimization with respect to $q_{\mathrm{d_k}}(\boldsymbol{d}_k)$, while keeping the distributions $q_{\mathrm{a}}^{[i]}(\boldsymbol{a})$, $q_{\Xi}^{[i]}(\boldsymbol{\Xi})$ and $q_{\mathrm{d}_{\bar{\mathrm{k}}}}^{[i]}(\boldsymbol{d}) = \prod_{j\neq k} q_{\mathrm{d_j}}^{[i]}\left(\boldsymbol{d}_j\right)$ fixed, yields the update rule

$$q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k) \propto p(\boldsymbol{d}_k) \exp\left[E_{q_{\Xi}^{[i]}}\left\{E_{q_{\mathrm{d}_{\bar{\mathrm{k}}}}^{[i]}}\left\{E_{q_{\mathrm{a}}^{[i]}}\left\{\log p\left(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\boldsymbol{\Xi}\right)\right\}\right\}\right\}\right]. \tag{4.21}$$

In (4.21), the sequence of symbols of user k is updated based on the estimates of the channel weights, the estimated symbol sequences of the other users and the estimated noise variance inverse. The symbol sequences are updated user by user in a successive manner.

With the prior distribution given in (3.5) and making use of the results in (4.13), (4.14) and (4.17), (4.21) can be recasted as

$$q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k)\begin{cases}\propto \exp\left\{2\sum_{l=L_p}^{L-1} Re\left\{d_k[l]^{*}\gamma_k^{[i+1]}[l]\right\}\right\} & \boldsymbol{d}_k \in \mathcal{D}_k,\\ = 0 & \boldsymbol{d}_k \notin \mathcal{D}_k\end{cases}. \tag{4.22}$$

In the above expression,

$$\begin{aligned}\gamma_k^{[i+1]}[l] \triangleq{}& \left(a_k^{[i]}\right)^{*}\boldsymbol{s}_k[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}\boldsymbol{r}[l]\\ &-\sum_{\substack{j=1\\ j\neq k}}^{K}\tilde{d}_j^{[i]}[l]\left(a_k^{[i]}\right)^{*}a_j^{[i]}\boldsymbol{s}_k[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}\boldsymbol{s}_j[l]\\ &-\sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{j'=1}^{K}\left(\tilde{d}_j^{[i]}[l]\lambda_j^{[i+1]}u_{j'j}^{[i+1]}\left(u_{j'k}^{[i+1]}\right)^{*}\boldsymbol{s}_k[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}\boldsymbol{s}_j[l]\right)\\ ={}&\left(\left(a_k^{[i]}\right)^{*}\boldsymbol{s}_k[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}\right)\Bigg(\boldsymbol{r}[l]-\sum_{\substack{j=1\\ j\neq k}}^{K}\boldsymbol{s}_j[l]a_j^{[i]}\tilde{d}_j^{[i]}[l]\\ &\qquad-\sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{j'=1}^{K}\left(\left(a_k^{[i]^{*}}\right)^{-1}\boldsymbol{s}_j[l]\lambda_j^{[i+1]}u_{j'j}^{[i+1]}\left(u_{j'k}^{[i+1]}\right)^{*}\tilde{d}_j^{[i]}[l]\right)\Bigg).\end{aligned} \tag{4.23}$$

The first term on the r.h.s. of (4.23) represents the result of spreading code matched filtering. The second and third terms are the multiple

access interference. Thus, (4.23) performs interference cancellation, followed by linear filtering. It can be seen that the r.h.s. of (4.22) and (4.23) depend on the soft symbols. Calculation of the soft symbols corresponds to demodulation and decoding process as shown in the sequel. Further details of the derivations of (4.22) and (4.23) are provided in Appendix B.2.

In (4.22), $u_{j'j}^{[i+1]}$ denote the element at the j'-th row and the j-column of the lower triangular matrix $\boldsymbol{U}^{[i+1]}$ (see (B.5) in Appendix B.2) which is obtained by a singular value decomposition of the positive semi-definite symmetric error covariance matrix $\Sigma_{\boldsymbol{a}}^{[i+1]}$:

$$\Sigma_{\boldsymbol{a}}^{[i+1]} = \boldsymbol{U}^{[i+1]}\Lambda^{[i+1]}\left(\boldsymbol{U}^{[i+1]}\right)^{\mathrm{H}},$$

with $\Lambda^{[i+1]} = \operatorname{diag}\left\{\lambda_1^{[i+1]}, \ldots, \lambda_K^{[i+1]}\right\}$ denoting the diagonal matrix of the eigenvalues of $\Sigma_{\boldsymbol{a}}^{[i+1]}$. The last term in (4.23), relying on singular value decomposition, is a correction term to the interference cancellation process originating from the channel estimation covariance matrix. This term accounts for the accuracy of the channel weight estimates. Since the correction is relatively minor compared to the interference, becoming negligible after only a few iterations, it can be omitted without noticeable performance penalty.

A. Computation of the Soft Symbols

The soft symbol $\tilde{d}_k^{[i+1]}[l]$ is defined as the expectation of $d_k[l]$ with respect to the updated distribution in (4.22), i.e., $\tilde{d}_k^{[i+1]}[l] \triangleq E_{q_{\mathrm{d}}^{[i+1]}}\left\{d_k[l]\right\}$. As mentioned in Chapter 3, the mapping of code bits to modulation symbols is one-to-one. Thus, we can replace $q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k)$ by $q_{\mathrm{c_k}}^{[i+1]}(\boldsymbol{c}_k)$ when $\boldsymbol{c}_k$ is mapped to $\boldsymbol{d}_k$. The soft symbols for BPSK and QPSK are given below.

BPSK modulation One interleaved code bit is mapped onto one BPSK modulation symbol. Thus the expectation of $d_k[l]$ is

$$\tilde{d}_k^{[i+1]}[l] \triangleq E_{q_{\mathrm{d}}^{[i+1]}}\left\{d_k[l]\right\} = E_{q_{\mathrm{d_k}}^{[i+1]}}\left\{d_k[l]\right\} = \sum_{\substack{\boldsymbol{d}_k \in \mathcal{D}_k^B \\ d_k[l]=1}} q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k) - \sum_{\substack{\boldsymbol{d}_k \in \mathcal{D}_k^B \\ d_k[l]=-1}} q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k). \tag{4.24}$$

QPSK modulation Two interleaved code bits are mapped onto one QPSK symbol. The expectation of a QPSK symbol is computed via the soft code bit as

$$\tilde{d}_k^{[i+1]}[l] = E_{q_{\mathrm{d_k}}^{[i+1]}}\left\{d_k[l]\right\} = \sum_{\boldsymbol{d}_k \in \mathcal{D}_k^D} d_k[l] q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k) = \tilde{c}_k^{[i+1]}[2l] + j\tilde{c}_k^{[i+1]}[2l+1] \tag{4.25}$$

with the $2l$-th and $2l+1$-th interleaved soft code bit defined as

$$\tilde{c}_k^{[i+1]}[2l] = \sum_{\substack{\boldsymbol{c}_k \in \mathcal{C}_k \\ c_k[2l]=1}} q_{\mathrm{c_k}}^{[i+1]}(\boldsymbol{c}_k) - \sum_{\substack{\boldsymbol{c}_k \in \mathcal{C}_k \\ c_k[2l]=0}} q_{\mathrm{c_k}}^{[i+1]}(\boldsymbol{c}_k) \tag{4.26}$$

and

$$\tilde{c}_k^{[i+1]}[2l+1] = \sum_{\substack{\boldsymbol{c}_k \in \mathcal{C}_k \\ c_k[2l+1]=1}} q_{\mathrm{c_k}}^{[i+1]}(\boldsymbol{c}_k) - \sum_{\substack{\boldsymbol{c}_k \in \mathcal{C}_k \\ c_k[2l+1]=0}} q_{\mathrm{c_k}}^{[i+1]}(\boldsymbol{c}_k) \tag{4.27}$$

respectively. Note that the calculation in (4.25) includes two steps: a demodulation step relating modulation symbols to interleaved code bits and a decoding step in which expectations are calculated based on sequence probabilities.

B. Single-User Decoding

Conceptually, the distribution of the codeword vector $\boldsymbol{d}_k \in \mathcal{D}_k$ is estimated in (4.22). However, since the calculations of (4.18), (4.13), (4.14) and (4.23) only require the soft symbols given in (4.24) and (4.25) for BPSK and QPSK, respectively, the soft symbols calculated based on this distribution are sufficient to perform the other steps.

BPSK modulation When BPSK modulation is used, Eq. (4.22) can be rewritten as

$$q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k) \begin{cases} \propto \exp\left\{\sum_{l=L_p}^{L-1} d_k[l] 2Re\left\{\gamma_k^{[i+1]}[l]\right\}\right\}, & \boldsymbol{d}_k \in \mathcal{D}_k^B \\ = 0, & \boldsymbol{d}_k \notin \mathcal{D}_k^B \end{cases}. \tag{4.28}$$

The quantity $\gamma_k^{[i+1]}[l]$ is computed as in (4.23). Comparing the structure of (4.28) to the maximum posterior distribution algorithm in [24, 53]

(called Log-MAP therein) gives rise to an interesting interpretation. The value $\gamma_k^{[i+1]}[l]$ may be related to the L-value[3] for $d_k[l]$ based on the "observation" $\gamma_k^{[i+1]}[l]$, i.e.,

$$\mathcal{L}(\gamma_k^{[i+1]}[l]|d_k[l]) = 4Re\left\{\gamma_k^{[i+1]}[l]\right\} \tag{4.29}$$

for $k = 1, \ldots, K$ and $l = L_p, \ldots, L-1$.

With the L-values defined in (4.29), $\sum_{\substack{\boldsymbol{d}_k \in \mathcal{D}_k^B \\ d_k[l]=1}} q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k)$ in (4.24) gives the posterior probability of $d_k[l] = 1$ given the observations $\gamma_k^{[i+1]}[l]$, $l = L_p, \ldots, L-1$. The BCJR algorithm [5] can be applied to compute the posterior distributions of the symbols in (4.24) [24, 53]. Note that $\tilde{d}_k^{[i+1]}[l]$ can be viewed as a posterior soft code symbol. Using this interpretation and the well-known rules for conversion between L-values and soft symbols, extrinsic soft code symbols can be defined in a straightforward manner. The extrinsic soft code symbols will only be used in Section 4.5 for the purpose of comparison, since the formal DM framework dictates the use of posterior soft code symbols.

QPSK modulation When QPSK modulation is used, (4.22) yields

$$q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k) \begin{cases} \propto \exp\left\{2\sum_{l=L_p}^{L-1} Re\left\{d_k[l]^* \gamma_k^{[i+1]}[l]\right\}\right\}, & \boldsymbol{d}_k \in \mathcal{D}_k^Q \\ = 0, & \boldsymbol{d}_k \notin \mathcal{D}_k^Q \end{cases} \tag{4.30}$$

Inserting $d_k[l] = 2(c_k[2l] - 0.5) + j2(c_k[2l+1] - 0.5)$ into (4.30) yields

$$q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k) \propto \exp\left\{2\sum_{l=L_p}^{L-1}\left(2(c_k[2l] - 0.5)Re\left\{\gamma_k^{[i+1]}[l]\right\} + 2(c_k[2l+1] - 0.5)Im\left\{\gamma_k^{[i+1]}[l]\right\}\right)\right\} \tag{4.31}$$

for $\boldsymbol{c}_k \in C_k$. Similarly to the case of BPSK symbols, the L-value of the $2l$-th interleaved code bit is defined as

$$\mathcal{L}(\gamma_k^{[i+1]}[l]|c_k[2l]) = 4Re\{\gamma_k^{[i+1]}[l]\} \tag{4.32}$$

and the L-value of the $(2l+1)$-th interleaved code bit is defined as

$$L(\gamma_k^{[i+1]}[l]|c_k[2l+1]) = 4Im\{\gamma_k^{[i+1]}[l]\}. \tag{4.33}$$

[3] For a binary variable, the L-value is defined as the logarithmic ratio between the probabilities of the variable being 0 and 1, respectively.

With the L-values (4.32) and (4.33), we can compute the soft interleaved code bits in (4.26) and (4.27) and thereby compute the soft modulation symbols. The module for computing the soft code symbols of user k is depicted in Fig. 4.4.

After the last iteration, the decisions on the information bits are obtained by making hard decisions on the posterior information bit distributions provided by the single-user decoders.

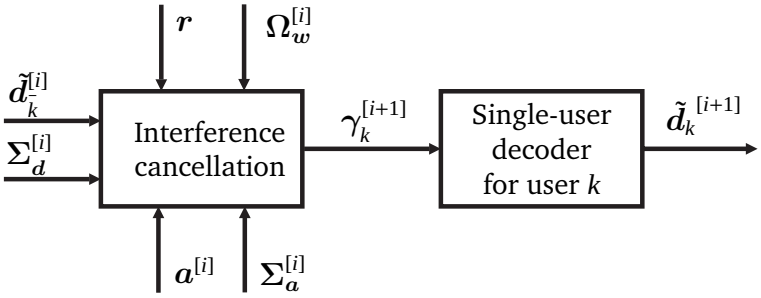

Figure 4.4: Interference cancellation device and single-user decoder: Module for updating $q_{\mathrm{d_k}}(\boldsymbol{d}_k)$. The code symbol covariance $\boldsymbol{\Sigma}_{\boldsymbol{d}}^{[i]}$ can be obtained from $\tilde{\boldsymbol{d}}^{[i]}$.

So far, the derivations of the three components: channel weights estimator, noise covariance matrix inverse estimator, and user sequence decoders, are described. For the seek of simplicity, we only consider white Gaussian noise from now on. However, the question that arises now is how these three components are connected and in which order they are activated. The scheduling is discussed in the next section.

4.3 Scheduling in the DM Receiver

Different versions of the iterative DM receiver can be obtained by using different scheduling schemes for updating the channel weights, the inverse of the noise variance and the users' soft symbols. No analytical approach for determining the optimal scheduling order has been proposed yet. In this thesis, we consider four versions of the DM receiver with scheduling rules listed in Table 4.3. We evaluate their performance by means of Monte Carlo simulations.

The computational complexities of the channel estimation, the noise variance inverse estimation and the interference cancellation versus the system setting are given by $\mathcal{O}_{CE}(K^3)$, $\mathcal{O}_{NE}(N^2)$ and $\mathcal{O}_{IC}(K^3)$, respectively. One stage is considered completed when the soft code symbols of all the users have been updated once. For the different versions of the DM

DM1	$q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\varsigma}(\varsigma) \to q_{\mathrm{d}_1}(\boldsymbol{d_1}) \ldots \to q_{\mathrm{d}_K}(\boldsymbol{d_K}) \to q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\varsigma}(\varsigma) \to q_{\mathrm{d}_1}(\boldsymbol{d_1}) \ldots$
DM2	$q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\varsigma}(\varsigma) \to q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\mathrm{d}_1}(\boldsymbol{d_1}) \to q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\mathrm{d}_2}(\boldsymbol{d_2}) \ldots$
DM3	$q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\varsigma}(\varsigma) \to q_{\mathrm{d}_1}(\boldsymbol{d_1}) \to q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\varsigma}(\varsigma) \to q_{\mathrm{d}_2}(\boldsymbol{d_2}) \ldots .$
DM4	$q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\varsigma}(\varsigma) \to q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\mathrm{d}_1}(\boldsymbol{d_1}) \to q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\varsigma}(\varsigma) \to q_{\mathrm{a}}(\boldsymbol{a}) \to q_{\mathrm{d}_2}(\boldsymbol{d_2}) \ldots$

Table 4.1: The update scheduling schemes considered for the DM receiver.

	CE	NE	IC
DM1	1	1	K
DM2	$K+1$	1	K
DM3	K	K	K
DM4	$2K$	K	K

Table 4.2: Number of module activations per stage; CE: channel estimation, NE: noise variance inverse estimation, IC: interference cancellation.

receiver listed in Table 4.3 the corresponding complexities of one stage, in terms of module (CE, NE and IC) activations, are listed in Table 4.3. For example, for the DM1 receiver, CE is activated once, NE is activated once and IC is activated K times per stage.

Although convergence is guaranteed for all four versions of the DM receiver, they may converge to different stationary points, leading to different performances in terms of bit error rate (BER). The performances of the four versions of the DM receiver are considered in Section 4.5.

4.4 Receiver Architecture

Fig. 4.5 shows a block diagram of the receiver architecture. Let us consider iteration i where the soft symbols of user k are updated. The soft symbols and the estimate of the inverse of the noise variance from the previous iteration $i-1$ are used to perform joint-user channel weight estimation, which provides the mean and covariance matrix of the channel weight vector as outputs of (4.13) and (4.14). The noise variance inverse estimator utilizes the output of the channel weight estimator as

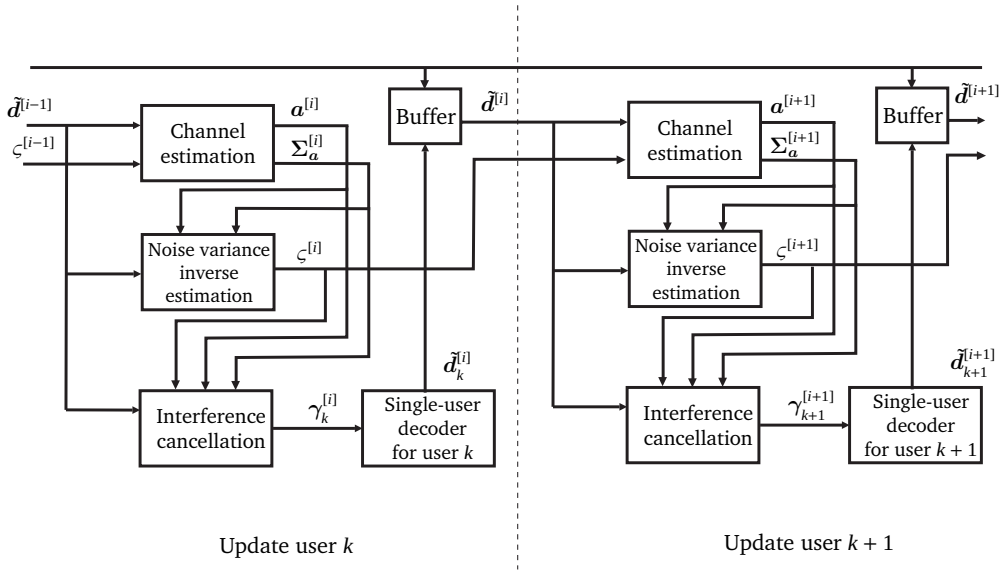

Figure 4.5: The receiver architecture for user k at iteration i.

well as the soft symbols at iteration $i-1$ to update the mean of the noise variance inverse according to (4.20). Subsequently, the updated mean and the covariance matrix of the channel weight vector and the updated mean of the noise variance inverse are fed as input to the interference cancellation device that generates the decoding metrics used to update the soft symbols of user k according to (4.23). The updated soft symbols are stored for use in the next iteration $i+1$. At iteration $i+1$, the soft symbols of user $k+1$ are updated in a similar way as above and so on.

In summary, the DM algorithm applied in the considered coded CDMA system performs the following steps:

1. *Initialization of the iterative process:* Calculate $\boldsymbol{a}^{[0]}$, $\tilde{\boldsymbol{d}}^{[0]}$ and $\varsigma^{[0]}$.
2. *Channel estimation:* Calculate $\boldsymbol{a}^{[i]}$ and $\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}$ using (4.13) and (4.14). Perform singular value decomposition of $\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}$.
3. *Noise variance inverse estimation:* Calculate $\varsigma^{[i]}$ using (4.20).
4. *Select the user to be updated:* Let k indicate the selected user. The users can be for example ordered according to the magnitudes of the estimated channel weights.
5. *Interference cancellation:* Calculate the sequence $\boldsymbol{\gamma}_k = \left[\gamma_k[L_p], \ldots, \gamma_k[L-1]\right]$ for user k using (4.23) and map the components of the vector $\boldsymbol{\gamma}_k$ to L-values.
6. *Single-user decoding:* Use the BCJR algorithm to calculate the posterior distributions of the modulation symbols based on the

L-values computed in Step 5. Compute the soft symbols $\tilde{d}_k[l]^{[i]}$ $l = L_p, \ldots, L-1$ for user k based on the posterior distributions.

7. Go back to step 2 if the iteration process is not terminated[4].

Notice that the above steps are referred as to the DM3 scheme in Table 4.3. For other versions of the DM scheme in Table 4.3, Step 7 should be changed accordingly.

4.5 Simulation Results

In this section we evaluate the performance of the four versions of the DM receiver by means of Monte Carlo simulations. Each version is specified by its update scheduling scheme specified in Table 4.3. The system model detailed in Section 3 is considered. More specifically, all users apply the same rate $R_c = 1/2$ terminated convolution code with generator polynomials $(5,7)_8$. The generated codewords have length $L_c = 320$ code symbols, corresponding to information sequences of length $M = 158$ information symbols. Random signature sequences of length $N_c = 8$ chips are assigned to the users. Each codeword is multiplexed with L_p random pilot symbols and each block of $L = L_c + L_p$ symbols is transmitted through a block fading channel. The effective signal-to-noise ratio is defined as $E_b/N_0 = L/(L_c R_c) \cdot E_s/N_0$, where E_s is the energy per code symbol and E_b is the energy per information bit and N_0 is the power spectral hight. In the simulations, we set $N_0 = 1$. All users have the same E_b/N_0.

The performance of the DM receiver depends on the initialization. Thus it is important to initialize the iterative process properly. A reasonable compromise is to let the initial estimates of the channel weights be determined by a linear least-squares estimator based on the pilot symbols. To get initial estimates of the input probabilities to the single-user decoders, a decorrelating filter is applied to the received signal vector. Together with the initial estimates of the channel weights, the initial input probabilities are found. Similarly, an initial estimate of the noise variance inverse is computed from (4.17) and (4.18), given the initial channel weight estimates and the initial symbol probabilities. The iterative process is terminated after 10 stages. This figure was found to be sufficient for convergence (see e.g. Fig. 4.8).

[4]The iterative process terminates when the divergence in (2.3) does not decrease significantly with further iterations. In practice, a few iterations suffice for this to happen.

The averaged BER returned by the simulations ($\overline{BER}$) is obtained by averaging the BERs over all users. Whenever appropriate, the simulated BER performance is compared to the performance of a single-user (SU) system with known channel weights and known noise variance.

Pilot symbol overhead First, we investigate how many pilot symbols are required for a DM receiver to have satisfactory performance. Clearly, a large number of pilot symbols creates an overhead which is not preferred. In Fig. 4.6, the averaged BER performance of the DM3 receiver (see Table 4.3.) is plotted for $K = 32$ users as a function of E_b/N_0 with the number of pilot symbols as a parameter. As the number of pilot symbols increases from $L_p = 4$ to $L_p = 6$, the BER of the DM3 receiver approaches the SU performance. No additional improvement can be observed when more than 6 pilot symbols are employed. The total transmission overhead with $L_p = 6$ is $L_p/L \approx 1.8\%$ which is sufficiently low. Therefore, we choose $L_p = 6$ in the subsequent examples.

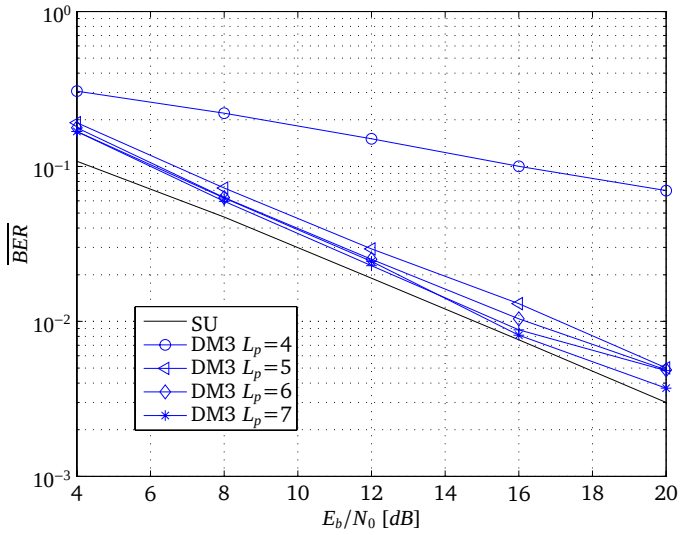

Figure 4.6: Averaged BER performance of the DM3 receiver for $K = 32$ users with respect to the number L_p of pilot symbols.

Scheduling and sorting order The different versions of the DM receiver performing the update scheduling schemes given in Table 4.3 have different complexities. In the following we investigate the impacts of the update scheduling schemes on the BER performance and on the convergence rate. In addition, the particular order of updating users

also influences the performance. The best user schedule sorts users in descending order according to the magnitudes of the estimated channel weights, where the strongest user is processed first. For comparison, we also consider a user update schedule based on sorting users in ascending order of the magnitudes of the estimated channel weights.

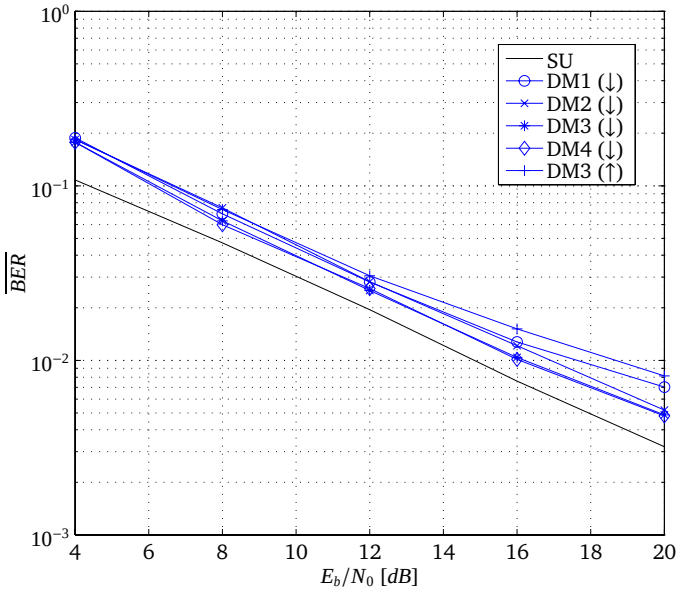

Figure 4.7: Averaged BER performance of the four versions of the DM receiver for $K = 32$ users. The users are updated in increasing (↑) or decreasing (↓) order of magnitudes of the estimated channel weights.

In Fig. 4.7 and Fig. 4.8, we compare the performance of the different versions of the DM receiver for $K = 32$ users. The four versions of the DM receiver update the users according to descending order of magnitudes of the estimated channel weights (denoted by (↓)). For comparison, we also consider the DM3 receiver with a user schedule based on ascending order of magnitudes of the estimated channel weights (denoted by (↑)). We observe in Fig. 4.7 that the DM3 (↓) receiver outperforms the DM3 (↑) receiver. We also observe that all the simulated receivers perform close to the SU case with the DM4 receiver exhibiting the best performance. The DM3 and DM4 receivers outperform the DM1 and DM2 receivers, since the channel weight estimation and the estimation of the noise variance inverse are performed more often at each stage. However, the DM3 and DM4 receivers have higher complexity per stage compared to the DM1 and DM2 receivers.

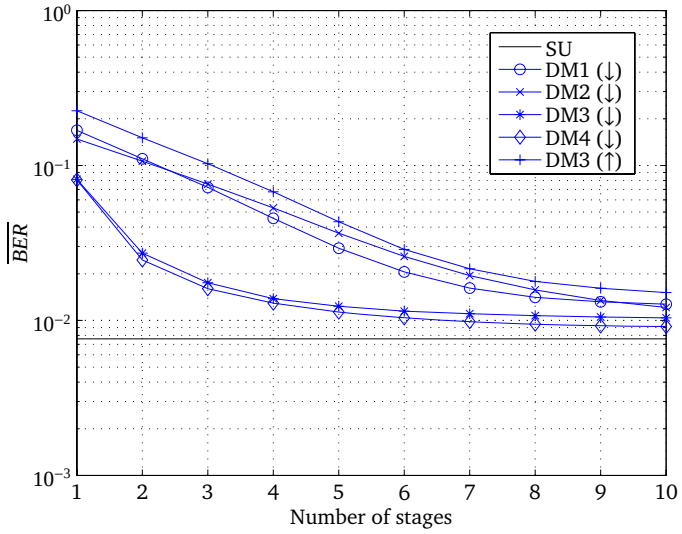

Figure 4.8: Averaged BER performance versus number of stages at $E_b/N_0 = 16$ dB of the four versions of the DM receiver for $K = 32$ users. The users are updated in an ascending (↑) or a descending (↓) order of the magnitude of the estimated channel weights.

In Fig. 4.8, the BER performance of the four versions of the DM receiver are shown versus the number of stages. The improvement on $\overline{BER}$ is marginal after 5-6 stages for the DM3 and DM4 receivers, and after 9-10 stages for the DM1 and DM2 receivers. The BERs of the DM3 and DM4 receivers decrease at a similar rate.

Regarding Fig. 4.7 and Fig. 4.8, the DM3 and DM4 receivers have very similar BER performance. However, the DM3 receiver is less complex than the DM4 receiver. Thus we consider the DM3 (↓) receiver in the following simulations. In addition, the descending order of absolute estimated channel weights is used in the rest of the simulations.

Exchanged information In the DM receiver, the CE, NE and IC modules accept soft symbols as inputs. These soft symbols are calculated based on the posterior probabilities (APPs) of the code symbols. In related works, it is suggested to feed soft symbols computed based on extrinsic (EXT) probability [43, 34] or hard-decision symbols [37] to the CE and IC modules. In the following simulations, we investigate the impacts on performance of feeding different types of information into the CE, the NE and the IC device of the DM3 receiver.

In Fig. 4.9, we compare the DM3 receivers with different types of

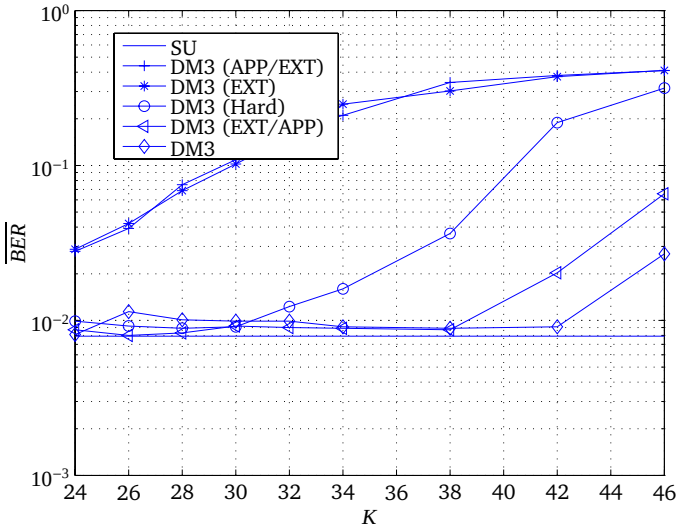

Figure 4.9: Averaged BER versus the number of active users at $E_b/N_0 = 16$ dB for the DM3 receivers using different types of information for CE, NE and IC.

information passed into the CE, NE and IC modules. The performances of the following schemes are compared:

- "DM3": APP-based soft symbols passed to CE, NE and IC;
- "DM3 (EXT/EXT)": EXT-based soft symbols passed to CE, NE and IC;
- "DM3 (Hard)": Hard-decision symbols passed to CE, NE and IC;
- "DM3 (EXT/APP)": EXT-based symbols passed to CE and NE and APP-based soft symbols to IC;
- "DM3 (APP/EXT)": APP-based soft symbols passed to CE and NE and EXT-based soft symbols passed to IC.

Fig. 4.9 illustrates the BER performance versus the number of active users for the four DM3 receivers at $E_b/N_0 = 16$dB. We observe that the DM3 receiver using APP-based soft symbols for CE, NE and IC has the best BER performance, especially when the number of active users is high. The receiver using EXT-based soft symbols for CE and NE, and APP for IC has a slightly worse performance. The DM3 receivers using EXT-based soft symbols for IC "DM3 (EXT)" and "DM3 (APP/EXT)" perform considerably worse than the other receivers. The DM3 receiver using hard-decisions coincides with the one derived from the EM framework [37]. It has a reasonably good performance for $K \leq 32$. When $K > 32$, the hard-decision-based receiver performs better than the ones using EXT-based soft symbols for IC and worse than the ones using APPs for IC.

Note that for this application, APP is the type of information indicated by the DM framework. The above simulation results confirm that APPs rather than EXT information should be used for computing the soft symbols for CE, NE and IC in the DM receiver.

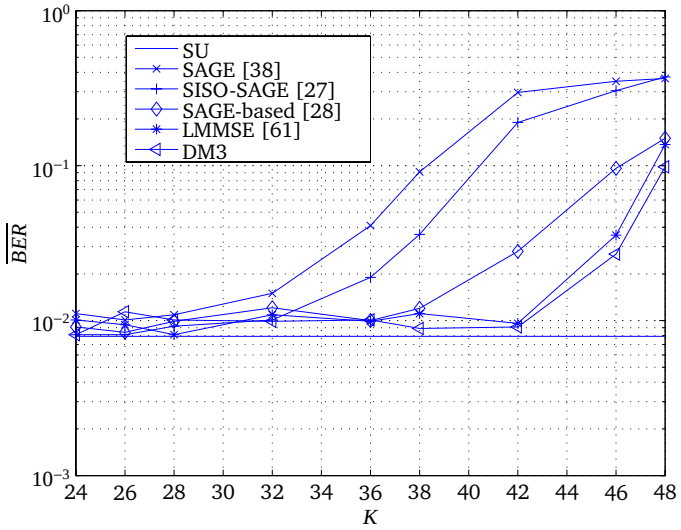

Figure 4.10: Averaged BER performance versus the number of users K at E_b/N_0 = 16dB of the DM3 receiver, the LMMSE receiver [59], the SAGE-based receiver [28], the SISO-SAGE receiver [26], and the SAGE receiver [38].

Comparison of the DM3 receiver with the related works In Fig. 4.10, we compare the BER performance of the DM3 receiver with some related receivers [59, 38, 28, 26]. The average BER is plotted as a function of the system load in terms of numbers of users, at E_b/N_0 = 16 dB. The DM3 receiver, the LMMSE receiver [59], the SAGE-based receiver [28], and the SISO-SAGE [26] exhibit BER performance close to the SU performance for up to $K = 32$ users. For $K > 32$ users, the performance of the hard-decision-based DM3 and the SISO-SAGE receiver deteriorates rapidly. Similar behavior is observed for the DM3 receiver and the LMMSE receiver: the BER performances of the two receivers starts to degrade when $K > 42$. For $K = 46$, the BER performance of the DM3 receiver is slightly better than that of the LMMSE receiver. The SAGE-based receiver performs slightly worse, but enjoys a similarly graceful BER degradation. The SAGE receiver is competitive up to $K = 32$, after which it suffers a rapid performance degradation. Note that the performance of the LMMSE receiver in [59] is obtained by assuming perfect knowledge of

the noise variance, while in the DM3 receiver no prior information about the noise variance is available. As shown in Fig. 4.10, the DM3 receiver and the LMMSE receiver [59] exhibit very similar performances. Further investigations to compare these two receivers are carried out later on.

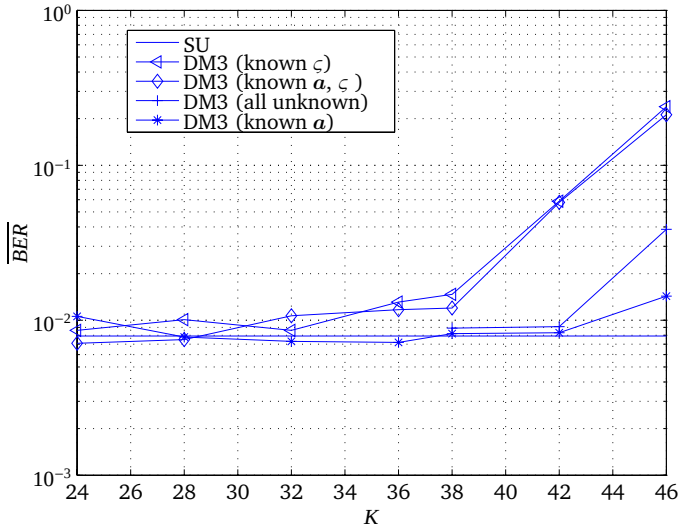

Figure 4.11: Averaged BER performance versus the number for users K of the DM3 receiver with different levels of channel knowledge. The effective SNR is $E_b/N_0 = 16$ dB.

The DM3 receiver with different levels of channel knowledge In the following, we investigate the BER performance of the DM3 receiver for different levels of channel knowledge. In principle, the more the receiver knows about the channel state, the better its performance is expected to be. However, through the simulations reported here, we observe a somewhat intriguing behaviour.

Fig. 4.11 illustrates the BER performance of the DM3 receiver for different levels of channel information at $E_b/N_0 = 16$ dB. It is shown that the receiver estimating the noise variance inverse in the iterative process always outperforms the receiver that does not update the noise variance inverse. Two effects are observed in Fig. 4.11. The first is that the receiver with unknown ς and $\boldsymbol{a}$ has a better performance than the one with known ς and unknown $\boldsymbol{a}$. The second observation is that the receiver with unknown ς and $\boldsymbol{a}$ has a better performance than the one with known ς and $\boldsymbol{a}$. Although surprising at first, this behavior is in fact to be expected. With an increasing number of users in the system,

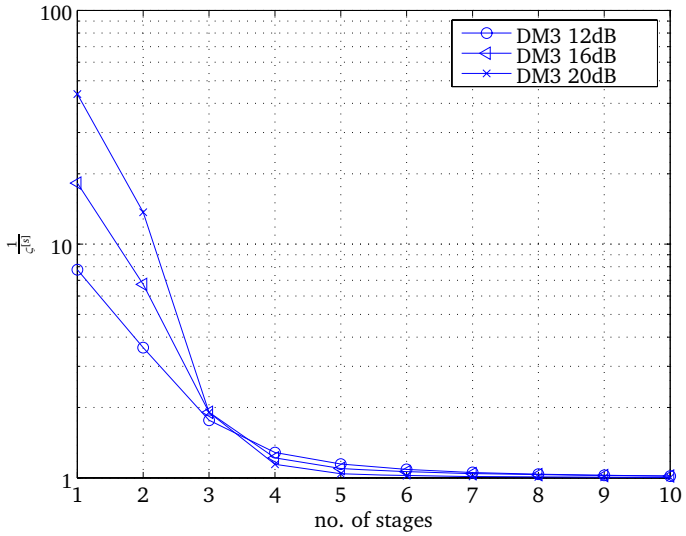

Figure 4.12: Estimated noise variance of the DM3 receiver versus the number of stages for $K = 32$ users, at $E_b/N_0 = \{12, 16, 20\}$ dB. The estimate is the inverse of the noise variance inverse estimate returned by the DM3 receiver.

the amount of multiple-access interference (MAI) increases. Beyond a certain level of MAI, the interference cancellation process is overwhelmed, resulting in an increasing level of residual interference. The residual interference can be considered as additional additive noise, which can be estimated and accounted for together with the additive white Gaussian noise, leading to better performance. Therefore, it is very important to include the noise variance estimation in the iterative process.

Since we assume $\sigma^2 = 1$ in the simulations reported here, the noise covariance matrix is an identity matrix. The estimated noise variance versus the number of stages is plotted in Fig. 4.12 for $K = 32$ and different E_b/N_0. The estimate is obtained by taking the inverse of the estimated noise variance inverse returned by the DM3 receiver. As mentioned previously, one stage is considered completed when the soft code symbols of all users have been updated once. At the first few stages, a significant part of the estimated noise variance is contributed by the residual interference. As more stages are performed, less residual interference is left. The estimated noise variance finally converges to the true noise variance after 9-10 stages.

4.6 Comparison Of the DM Receiver and the LMMSE-Based Receiver [59]

In this section we interpret our DM algorithm and compare it to the LMMSE-based receiver structure proposed in [59]. We consider the latter receiver as the current state-of-the-art receiver.

4.6.1 Short Description of the LMMSE-Based Receiver [59]

The LMMSE-based receiver [59] consists of three components: a channel estimator, an interference-cancellation multiuser detector, and a bank of single-user posterior distribution decoders. In the following, the receiver is described shortly.

The design of the LMMSE-based receiver [59] is based on the following assumptions: i) The thermal noise is zero-mean additive white Gaussian noise with known variance σ^2; ii) The vector of channel weights is zero-mean Gaussian with diagonal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{a}}$; iii) Only the code-symbol distributions are updated over the iterations.

With these assumptions, the vector of channel weights is estimated using a LMMSE filter, which outputs

$$\boldsymbol{a}_{LMMSE}^{[i+1]} = \left(\boldsymbol{B}_{\text{ce}}^{[i]}\right)^{-1} \varsigma \sum_{l=0}^{L-1} \left(\tilde{\boldsymbol{D}}^{[i]}[l]\right)^{\text{H}} \boldsymbol{S}[l]^{\text{H}} \boldsymbol{r}[l] \tag{4.34}$$

with the covariance matrix

$$\begin{aligned} \boldsymbol{B}_{\text{ce}}^{[i]} = \varsigma \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \Big(& \sum_{l=0}^{L-1} \boldsymbol{S}[l] \tilde{\boldsymbol{D}}^{[i]}[l] \boldsymbol{\Sigma}_{\boldsymbol{a}} \left(\tilde{\boldsymbol{D}}^{[i]}[l]\right)^{\text{H}} \boldsymbol{S}[l]^{\text{H}} \\ & + \sum_{l=L_p}^{L-1} \boldsymbol{S}[l] \boldsymbol{E}^{[i]}[l] \boldsymbol{\Sigma}_{\boldsymbol{a}} \left(\boldsymbol{E}^{[i]}[l]\right)^{\text{H}} \boldsymbol{S}[l]^{\text{H}} + \boldsymbol{I} \Big). \end{aligned} \tag{4.35}$$

The multiuser detector performs interference cancellation first:

$$\boldsymbol{y}_{k,LMMSE}^{[i+1]}[l] = \boldsymbol{r}[l] - \sum_{\substack{j=1 \\ j \neq k}}^{K} \boldsymbol{s}_j[l] a_j^{[i]} \tilde{d}_j^{[i]}[l] \tag{4.36}$$

for user k and $l = 0, \ldots, L-1$. To further suppress the residual interference in the signal on the l.h.s of (4.36), a linear LMMSE filter [58] is applied:

$$\gamma_{k,LMMSE}^{[i+1]}[l] = \left(\boldsymbol{g}_{\text{ic},k}^{[i+1]}[l]\right)^{\text{H}} \boldsymbol{y}_{k,LMMSE}^{[i+1]}[l]. \tag{4.37}$$

The filter coefficients are defined as

$$\boldsymbol{g}_{\mathrm{ic},k}^{[i+1]}[l]^{\mathrm{H}} = \left(a_k^{[i+1]}\right)^* \boldsymbol{s}_k[l]^{\mathrm{H}} \left(\boldsymbol{B}_{\mathrm{ic}}^{[i]}\right)^{-1} \tag{4.38}$$

with the covariance matrix

$$\boldsymbol{B}_{\mathrm{ic}}^{[i]} = \frac{1}{L}\sum_{l=0}^{L-1}\left(\boldsymbol{S}[l]\boldsymbol{E}^{[i]}[l]\boldsymbol{A}^{[i]}(\boldsymbol{A}^{[i]})^{\mathrm{H}}\left(\boldsymbol{E}^{[i]}[l]\right)^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}} + \sigma^2\boldsymbol{I}\right). \tag{4.39}$$

Thus, applying the LMMSE filter (4.38) to the interference-mitigated signal (4.36) yields the symbol-wise input to the single-user decoders (4.37).

4.6.2 Discussion and Comparison of the Two Structures

Both the DM and LMMSE-based receivers perform channel weight estimation, interference-cancellation, and single-user APP decoding. However, the mathematical expressions for channel weight estimation and interference cancellation are different in the two receivers. In the LMMSE receiver, each component is designed independently based on some assumptions on the statistics of the component input. In contrast, our receiver algorithm is based on updating the auxiliary distributions for the channel coefficient vector, the noise variance inverse, and the codewords.

In the following, we compare the blocks performing channel weight estimation, multiuser detection and single-user decoding of the DM and LMMSE receivers.

Channel estimation The expectation of the channel estimates of the two receivers are reported in Table 4.3. The expressions (4.13) and (4.34) for the estimated mean have the same structural form. However, the noise variance inverse in the two estimates are different (see Table 4.3). The differences stem from the fact that the noise variance inverse is assumed to be known in the LMMSE receiver while it is estimated in the DM receiver. If the estimate of the noise covariance matrix is replaced by a diagonal matrix with the diagonal elements equal to the known noise variance σ^2 in (4.13) and (4.14), (4.34) reduces to the equation of the classical LMMSE receiver.

Multiuser detection The expressions of the interference cancellation schemes of the two receivers are reported in Table 4.4. We compare (4.36) and the last factors of the product in (4.23). In (4.23), the

The LMMSE receiver [59]	The DM receiver
$$\boldsymbol{a}_{LMMSE}^{[i+1]} = \left(\boldsymbol{B}_{\text{ce}}^{[i]}\right)^{-1} \cdot \varsigma \sum_{l=0}^{L-1} \tilde{\boldsymbol{D}}^{[i]}[l]^{\text{H}} \boldsymbol{S}[l]^{\text{H}} \boldsymbol{r}[l]$$ (4.34)	$$\boldsymbol{a}^{[i+1]} = (\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i+1]})^{-1} \varsigma^{[i]} \sum_{l=0}^{L-1} (\tilde{\boldsymbol{D}}^{[i]}[l])^{\text{H}} \boldsymbol{S}[l]^{\text{H}} \boldsymbol{r}[l]$$ (4.13)
$$\boldsymbol{B}_{\text{ce}}^{[i]} = \varsigma \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \cdot \left(\sum_{l=0}^{L-1} \boldsymbol{S}[l] \tilde{\boldsymbol{D}}^{[i]}[l] \boldsymbol{\Sigma}_{\boldsymbol{a}} \tilde{\boldsymbol{D}}^{[i]}[l]^{\text{H}} \boldsymbol{S}[l]^{\text{H}} + \sum_{l=L_p}^{L-1} \boldsymbol{S}[l] \boldsymbol{E}^{[i]}[l] \boldsymbol{\Sigma}_{\boldsymbol{a}} \boldsymbol{E}^{[i]}[l]^{\text{H}} \boldsymbol{S}[l]^{\text{H}} + \boldsymbol{I} \right)$$ (4.35)	$$\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i+1]} = \left(\varsigma^{[i]} \sum_{l=0}^{L-1} (\tilde{\boldsymbol{D}}^{[i]}[l])^{\text{H}} \boldsymbol{S}[l]^{\text{H}} \boldsymbol{S}[l] \tilde{\boldsymbol{D}}^{[i]}[l] + \varsigma^{[i]} \sum_{l=L_p}^{L-1} (\boldsymbol{E}^{[i]}[l])^{\text{H}} \text{Diag}\{\boldsymbol{S}[l]^{\text{H}} \boldsymbol{S}[l]\} \boldsymbol{E}^{[i]}[l] + \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \right)^{-1}$$ (4.14)

Table 4.3: Comparison of channel weight estimation in the LMMSE receiver [59] and the DM receiver.

post-cancellation linear filter has the same structural form as that of the LMMSE filter in (4.39). However, the covariance matrices in the post-cancellation filters are different (see the second row in Table 4.4). Compared to (4.36), there is an additional term in the last line in (4.23). This term accounts for the covariance matrix representing the precision of the channel estimates. Thus, the interference calculated in the DM receiver takes into account the accuracy in the channel weight estimation. As a result, the interference cancellation should be more efficient in the DM receiver.

In both receivers, post-cancellation filters are applied after interference cancellation. Here we compare the covariance matrices in (4.39) and (4.18) together with (4.17). Consider the terms in the r.h.s. of (4.17) together with (4.18). The first term is an estimate of the covariance matrix of the residual interference, i.e., it includes the effects of both thermal noise and residual multiple-access interference. The remaining three terms are the covariance matrix representing the precision of the channel vector estimate, the codewords and the combined effect of the two. Compared to (4.17), the covariance matrix (4.39) comprises only two of these terms. Since the distribution of channel weights is not updated in the LMMSE-based estimator, the terms containing the covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}$ are missing. More importantly, the covariance

The LMMSE receiver [59]	The DM receiver
$\boldsymbol{r}[l] - \sum_{\substack{j=1 \\ j\neq k}}^{K} \boldsymbol{s}_j[l] a_j^{[i]} \tilde{d}_j^{[i]}[l]$ (4.36)	$\boldsymbol{r}[l] - \sum_{\substack{j=1 \\ j\neq k}}^{K} \boldsymbol{s}_j[l] a_j^{[i]} \tilde{d}_j^{[i]}[l]$ $- \sum_{\substack{j=1 \\ j\neq k}}^{K} \sum_{j'=1}^{K} \left(\left(a_k^{[i]*} \right)^{-1} \boldsymbol{s}_j[l] \lambda_j^{[i+1]} u_{j'j}^{[i+1]} \left(u_{j'k}^{[i+1]} \right)^* \tilde{d}_j^{[i]}[l] \right)$ (4.23)
$\boldsymbol{B}_{\text{ic}}^{[i]}$ $= \frac{1}{L} \sum_{l=0}^{L-1} \Big(\boldsymbol{S}[l] \boldsymbol{E}^{[i]}[l] \boldsymbol{A}^{[i]} (\boldsymbol{A}^{[i]})^{\text{H}} \boldsymbol{E}^{[i]}[l]^{\text{H}} \boldsymbol{S}[l]^{\text{H}}$ $+ \sigma^2 \boldsymbol{I} \Big)$ (4.39)	$\boldsymbol{B}^{[i]}$ $= \sum_{l=0}^{L-1} \Big(\big(\boldsymbol{r}[l] - \boldsymbol{S}[l] \tilde{\boldsymbol{D}}^{[i]}[l] \boldsymbol{a}^{[i]} \big) \big(\boldsymbol{r}[l] - \boldsymbol{S}[l] \tilde{\boldsymbol{D}}^{[i]}[l] \boldsymbol{a}^{[i]} \big)^{\text{H}}$ $+ \boldsymbol{S}[l] \boldsymbol{E}^{[i]}[l] \boldsymbol{A}^{[i]} (\boldsymbol{A}^{[i]})^{\text{H}} \boldsymbol{E}^{[i]}[l]^{\text{H}} \boldsymbol{S}[l]^{\text{H}}$ $+ \boldsymbol{S}[l] \boldsymbol{E}^{[i]}[l] \text{Diag}\{\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}\} \boldsymbol{E}^{[i]}[l]^{\text{H}} \boldsymbol{S}[l]^{\text{H}}$ $+ \boldsymbol{S}[l] \tilde{\boldsymbol{D}}^{[i]}[l] \boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]} \tilde{\boldsymbol{D}}^{[i]}[l]^{\text{H}} \boldsymbol{S}[l]^{\text{H}} \Big)$ (4.18) and (4.17)
$\gamma_{k,LMMSE}^{[i+1]}[l]$ $= \left(a_k^{[i+1]} \right)^* \boldsymbol{s}_k[l]^{\text{H}} \left(\boldsymbol{B}_{\text{ic}}^{[i]} \right)^{-1} \cdot$ $\left(\boldsymbol{r}[l] - \sum_{\substack{j=1 \\ j\neq k}}^{K} \boldsymbol{s}_j[l] a_j^{[i]} \tilde{d}_j^{[i]}[l] \right)$ (4.37)	$\gamma_k^{[i+1]}[l]$ $= \left(\left(a_k^{[i]} \right)^* \boldsymbol{s}_k[l]^{\text{H}} \varsigma^{[i]} \right) \cdot$ $\Big(\boldsymbol{r}[l] - \sum_{\substack{j=1 \\ j\neq k}}^{K} \boldsymbol{s}_j[l] a_j^{[i]} \tilde{d}_j^{[i]}[l] -$ $\sum_{\substack{j=1 \\ j\neq k}}^{K} \sum_{j'=1}^{K} \left(\left(a_k^{[i]*} \right)^{-1} \boldsymbol{s}_j[l] \lambda_j^{[i+1]} u_{j'j}^{[i+1]} \left(u_{j'k}^{[i+1]} \right)^* \tilde{d}_j^{[i]}[l] \right) \Big)$. (4.23)

Table 4.4: The interference cancellation for the LMMSE receiver [59] and the DM receiver.

matrix of the residual interference is estimated as the first term in (4.17). With perfect cancellation the residual interference is nothing but the thermal noise; however, a significant level of residual multiple-access interference is present in the early iterations. The way that the residual interference is accounted for in (4.17) makes the post-cancellation filter more efficient in the DM receiver as shown in the simulation results later.

Single-user decoding Single-user decoders in the DM receiver outputs posterior probabilities, while the corresponding modules in [59] as well as in [43, 34, 41] output extrinsic probabilities. This is another significant difference between the two receiver structures. As previously mentioned, the exchange of extrinsic probabilities is an immediate consequence of applying BP. In contrast, the formal DM framework investigated here specifies the single-user decoders to compute posterior probability output from the single-user decoders.

Note that the noise variance is assumed to be known and no solution

to the unknown variance case is suggested in [59]. In the following we add a ML noise variance estimator to the receiver in [59] and compare it to our receiver. The simulation settings used to generate the following figures are provided in Section 4.5.

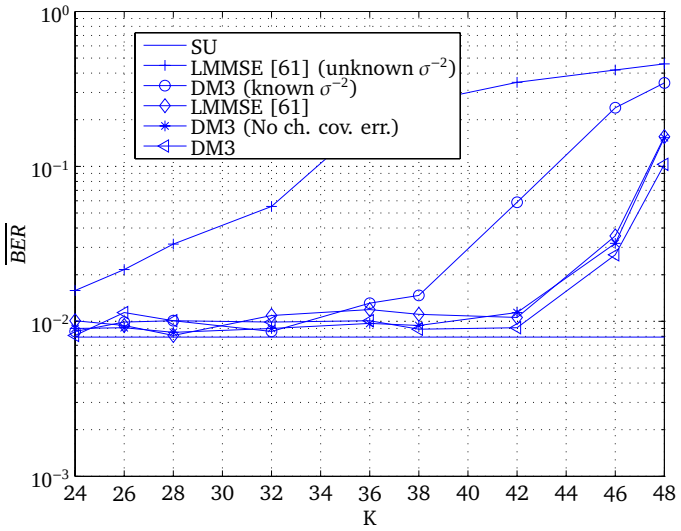

Figure 4.13: Performance of the LMMSE receiver [59] and the DM3 receiver versus the system load when the receivers have and do not have the knowledge of the noise variance; $E_b/N_0 = 16$dB.

In Fig. 4.13, we investigate the BER performance of the LMMSE receiver and the DM3 receiver in the cases where the noise variance is unknown or known to these receivers. Consider first the BER performance of the LMMSE receiver [59] and the DM3 receiver assuming known noise variance (even though, in practice, this assumption is unrealistic). In this case, the BER performance of [59] is much better than that of the DM3 receiver at high system load $K \geq 38$. This performance degradation of the DM3 receiver can be easily explained. Assuming known noise variance is equivalent to switching off the noise variance estimation, which leads to switching off the inherent estimation of the residual interference in the DM3 receiver. In [59] on the other hand, this residual interference is handled in the LMMSE filter (see (4.35)) no matter whether noise variance estimation is performed or not. Thus, the performance degradation of the modified DM3 receiver is caused by the lack of residual interference cancellation.

Next, let us compare the BER performance of the LMMSE receiver [59] and the DM3 receiver in the more realistic case where the noise variance is

unknown. Here, the later receiver performs much better than the former at high system load $K \geq 24$. These different behaviours can be explained by the distinct ways interference cancellation is performed in the various parts of the receivers. The ML estimate of the noise variance in the augmented LMMSE receiver includes also the residual interference. Thus, in this receiver, the residual interference is estimated twice, once in the LMMSE filter and once in the ML noise estimator. This redundant residual interference cancellation eventually causes performance degradation.

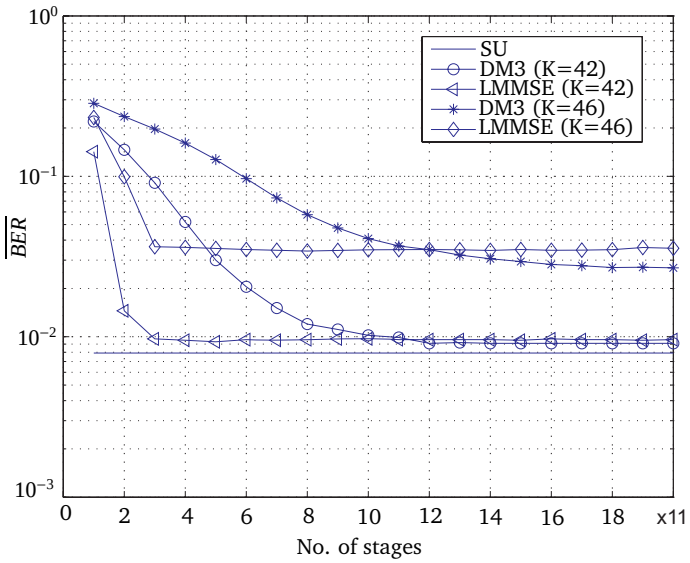

Figure 4.14: BER performance of the LMMSE and DM3 receivers versus the number of stages; E_b/N_0 = 16dB.

Since the covariance matrix representing the accuracy of the channel weights estimate is not used in [59], we also investigate the performance of the DM3 receiver when this matrix is discarded. As shown in Fig. 4.13, the DM3 receiver performs better when it accounts for this matrix even though the resulting gain is not significant.

In Fig. 4.14, we compare the convergence behaviour of the LMMSE receiver [59] and the DM3 receiver. Convergence is achieved when the $\overline{BER}$ is not decreasing any further with additional stages. The convergence rate is investigated for two high-load cases: $K = 42$ and $K = 46$. We observe that in both cases, the BER of the LMMSE receiver decreases very rapidly within few stages. After 4 stages, the BER of the LMMSE stabilizes. However, the DM3 receiver requires 10 stages for $K = 42$ and more than 20 stages for $K = 46$ to stabilize. Since the noise variance is assumed to be known in the LMMSE receiver, the residual interference is computed

more accurately in the first few stages. However, in the DM3 receiver, the noise variance estimate is affected by the residual interference in the first stages. Without knowing the noise variance, the noise and the residual interference cannot be distinguished from each other, which makes interference cancellation less efficient.

In summary, the DM and LMMSE receivers are very similar from structural viewpoint. However, there are several fundamental differences that make the performance and the complexity of the receivers different. In the LMMSE receiver, extrinsic soft symbols are used for interference cancellation, while APPs are used in the DM receiver. The residual interference is handled with a linear LMMSE filter in the LMMSE receiver, while the DM receiver handles it by estimating the inverse of the noise covariance matrix. In contrast to the DM receiver, the covariance matrix representing the accuracy of the channel weight estimate is not included when estimating the noise and decoding the code symbols in the LMMSE receiver. The DM3 receiver outperforms the LMMSE receiver in the case of unknown noise variance (see Fig. 4.13). The LMMSE receiver assuming known noise variance is less complex then the DM3 receiver since it requires fewer stages to converge to a stable BER performance (see Fig. 4.14).

Chapter 5

DM Receiver with Separate-User Channel Estimation

In this chapter, the DM method is applied to the same problem as in Chapter 4 with the difference that the auxiliary distribution of the vector of channel weights is further factorized. As a result of this additional factorization, the channel estimator performs separate-user channel weight estimation. Compared to the joint-user channel weight estimation, the separate-user channel weight estimation has lower complexity. However, the overall receiver performance with separate-user channel weight estimation is worse. The applications described in this chapter and in the previous chapter illustrate that the complexity of the resulting receiver can be adjusted by the degree of the factorization, and that the performance is related to the selected factorization.

5.1 Factorization of the Auxiliary Distribution

When the complexity of joint-user channel weight estimation needs to be reduced, the auxiliary distribution of the channel weights can be further factorized as follows (see also (3.10))

$$q(\boldsymbol{a}, \boldsymbol{d}, \Xi) = q_{\Xi}(\Xi) \prod_{k=1}^{K} q_{\mathrm{a}_k}(a_k) \prod_{k=1}^{K} q_{\mathrm{d}_k}(\boldsymbol{d}_k). \tag{5.1}$$

When choosing the auxiliary distribution (5.1), the resulting receiver performs separate-user channel weight estimation in contrast to the receiver derived based on (3.9) which implements joint-user channel weight estimation. The derivations based on (5.1) are shown in the

following sections. The factor graph representation of the auxiliary model (5.1) is illustrated in Fig. 5.1.

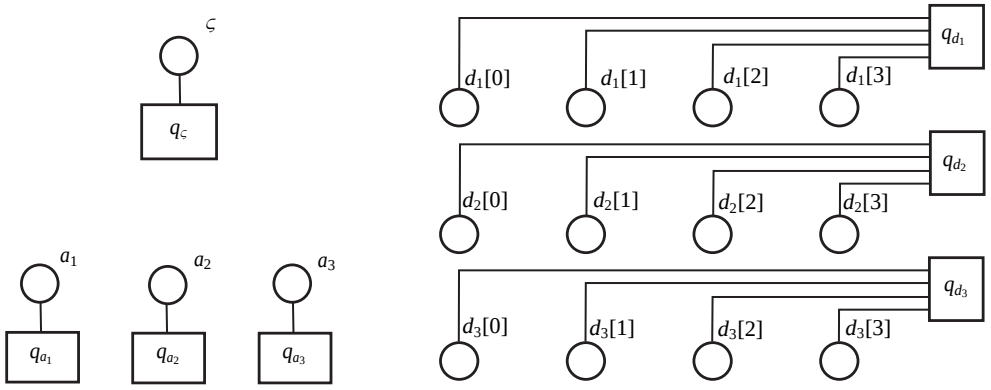

Figure 5.1: Dependency among the variables under the structured mean-field approximation in (3.10). Circles denote variable nodes; rectangles denote functional nodes.

Compared to Fig. 4.1, the variable nodes of the channel weights of different users are disjoint, i.e., the channel weights will not be updated jointly. The coding constraints on the transmitted code symbols are the same as in Fig. 4.1.

Given the auxiliary distribution (5.1), the KL divergence to be minimized is

$$D\left(q_{\Xi}(\boldsymbol{\Xi})\prod_{k=1}^{K}q_{\mathrm{a_k}}(a_k)\prod_{k=1}^{K}q_{\mathrm{d_k}}(\boldsymbol{d}_k)\,\middle\|\,p(\boldsymbol{a},\boldsymbol{\Xi},\boldsymbol{d}|\boldsymbol{r})\right). \tag{5.2}$$

The KL divergence (5.2) can be minimized in an iterative way by using the DM method. The auxiliary distributions $q_{\mathrm{a_k}}(a_k)$, $q_{\mathrm{d_k}}(\boldsymbol{d}_k)$, $q_{\Xi}(\boldsymbol{\Xi})$ are updated serially in successive steps.

5.2 Components of the DM receiver

In this section, the receiver components performing separate-user channel weight estimation, noise covariance matrix inverse estimation and interference cancelation followed single-user decoder are derived. They are compared to the corresponding components described in Chapter 4.

5.2.1 Separate-User Channel Estimation

When updating $q_{\mathrm{a_k}}(a_k)$, $k = 1,...,K$, separate-user channel weight estimation is performed. Similarly to (4.4), minimizing the divergence in (5.2) while keeping $q_{\Xi}^{[i]}(\boldsymbol{\Xi})$, $q_{\mathrm{a_j}}^{[i]}(a_j)$, $j \neq k$, $j = 1,\ldots,K$ and $q_{\mathrm{d}}^{[i]}(\boldsymbol{d}) = \prod_{k=1}^{K} q_{\mathrm{d_k}}^{[i]}(\boldsymbol{d}_k)$ fixed, yields

$$\begin{aligned} q_{\mathrm{a_k}}^{[i+1]}(a_k) &\propto \exp\left[\int d\boldsymbol{\Xi} q_{\Xi}^{[i]}(\boldsymbol{\Xi}) \int da_1 \ldots da_{k-1} da_{k+1} \ldots da_K \prod_{j\neq k}^{K} q_{\mathrm{a_j}}^{[i]}(a_j) \right. \\ &\qquad \left. \cdot \sum_{\boldsymbol{d}\in C} q_{\mathrm{d}}^{[i]}(\boldsymbol{d}) \log p(\boldsymbol{a},\boldsymbol{d}|\boldsymbol{r})\right] \\ &\propto \exp\left[\int d\boldsymbol{\Xi} q_{\Xi}^{[i]}(\boldsymbol{\Xi}) \int da_1 \ldots da_{k-1} da_{k+1} \ldots da_K \prod_{j\neq k}^{K} q_{\mathrm{a_j}}^{[i]}(a_j) \right. \\ &\qquad \cdot \sum_{\boldsymbol{d}\in C} q_{\mathrm{d}}(\boldsymbol{d}) \log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\boldsymbol{\Xi}) \\ &\qquad \left. + \int da_1 \ldots da_{k-1} da_{k+1} \ldots da_K \prod_{j\neq k}^{K} q_{\mathrm{a_j}}^{[i]}(a_j) \log p(\boldsymbol{a})\right] \end{aligned} \tag{5.3}$$

The log-likelihood function in (5.3) reads (see (4.7))

$$\begin{aligned} &\log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\boldsymbol{\Xi}) \\ &\qquad \propto^{e} L \log|\boldsymbol{\Xi}| - \sum_{l=1}^{L}\Big(\boldsymbol{r}[l] - \boldsymbol{S}[l]\boldsymbol{D}[l]\boldsymbol{a}\Big)^{\mathrm{H}} \boldsymbol{\Xi}\Big(\boldsymbol{r}[l] - \boldsymbol{S}[l]\boldsymbol{D}[l]\boldsymbol{a}\Big). \end{aligned} \tag{5.4}$$

In this update step, the expectation of the log-likelihood function in (5.4) with respect to $q_{\Xi}^{[i]}(\boldsymbol{\Xi})$, $\prod_{k=1}^{K} q_{\mathrm{d_k}}^{[i]}(\boldsymbol{d}_k)$ and $q_{a_j}^{[i]}(a_j)$, $j = 1,\ldots,K, j \neq k$ needs to be derived. Finally, the following expectation needs to be computed

$$E_{q_{\bar{a}_k}^{[i]}}[a_{k'}] \triangleq \int a_{k'} da_1 \ldots da_{k-1} da_{k+1} \ldots da_K \prod_{j\neq k}^{K} q_{a_j}^{[i]}(a_j) = \begin{cases} a_k & k' = k \\ \hat{a}_{k'}^{[i]} & k' \neq k. \end{cases} \tag{5.5}$$

The prior distribution of the channel weights is the same as in (4.11). Based on (5.5), we can compute the expectation of the sum of (5.4) and

(4.11) with respect to $q_{a_j}^{[i]}(a_j)$, $j = 1, \ldots, K$ and $j \neq k$,

$$E_{q_{\bar{a}_k}^{[i]}}\left[\log p(\boldsymbol{r}|\boldsymbol{a}, \boldsymbol{d}, \boldsymbol{\Xi})\right] \propto \sum_{l=0}\left(2Re\left\{a_k^* d_k^*[l]\left(\left(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{r}[l]\right)_k - \sum_{j\neq k}\hat{a}_j^{[i]} d_j[l]\left(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l]\right)_{kj}\right)\right\} + |a_k|^2|d_k[l]|^2\left(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l]\right)_{kk}\right) \quad (5.6)$$

and

$$E_{q_{\bar{a}_k}}\left[\log p(\boldsymbol{a})\right] \propto -\sigma_{a_k}^{-2}|a_k|^2. \quad (5.7)$$

Note that in (5.6), $|d_k[l]|^2 = 1$ for constant envelop modulation. Inserting (5.6) and (5.7) into (5.3), applying the results in (4.24), (4.8), and the definition $\boldsymbol{\Omega}_{\boldsymbol{w},s}^{[i]} \triangleq E_{q_{\Xi}^{[i]}}\left\{\boldsymbol{\Xi}\right\}$, the resulting distribution of the channel weight of user k is Gaussian. More specifically,

$$q_{\mathrm{a_k}}^{[i+1]}(a_k) \propto \exp\left[-\left|\frac{a_k - \mu_{a_k}^{[i+1]}}{\sigma_{a_k}^{[i+1]}}\right|^2\right] \quad (5.8)$$

with mean value

$$\mu_{a_k}^{[i+1]} = \left(\left(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w},s}^{[i]}\boldsymbol{S}[l]\right)_{kk} + \sigma_{a_k}^{-2}\right)^{-1} \cdot Re\left\{\tilde{d}_k^{[i]}[l]^*\left(\left(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w},s}^{[i]}\boldsymbol{r}[l]\right)_k - \sum_{j\neq k}\hat{a}_j^{[i]}\tilde{d}_j^{[i]}[l]\left(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w},s}^{[i]}\boldsymbol{S}[l]\right)_{kj}\right)\right\} \quad (5.9)$$

and variance

$$(\sigma_{a_k}^{[i+1]})^2 = \left(\left(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w},s}^{[i]}\boldsymbol{S}[l]\right)_{kk} + \sigma_{a_k}^{-2}\right)^{-1}. \quad (5.10)$$

Detailed derivations of (5.9) and (5.10) are provided in Appendix B.3.

5.2.2 Estimation of the Inverse of the Noise Covariance Matrix

Updating $q_{\Xi}(\boldsymbol{\Xi})$ results in the estimation of the noise covariance matrix inverse. Similarly to (4.15), the distribution of $\boldsymbol{\Xi}$ is updated according to

$$q_{\Xi}^{[i+1]}(\boldsymbol{\Xi}) \propto p(\boldsymbol{\Xi}) \exp\left[\int da_1 \dots da_K \prod_k q_{a_k}^{[i]}(a_k) \sum_{\boldsymbol{d}\in C} q_{\mathrm{d}}^{[i]}(\boldsymbol{d}) \cdot \log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\boldsymbol{\Xi})\right]. \quad (5.11)$$

To obtain the exponential term in (5.11), we compute the expectation in (5.4) with respect to $q_{\mathrm{a_k}}^{[i]}(a_k)$, $k = 1,\dots,K$ and $q_{\mathrm{d}}^{[i]}(\boldsymbol{d})$. Doing so yields

$$\exp\left[\int da_1 \dots da_K \prod_k q_{\mathrm{a_k}}^{[i]}(a_k) \sum_{\boldsymbol{d}\in C} q_{\mathrm{d}}^{[i]}(\boldsymbol{d}) \log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\boldsymbol{\Xi})\right] \propto |\boldsymbol{\Xi}|^L \exp\left[-\mathrm{tr}\left\{\boldsymbol{\Xi}\boldsymbol{B}_s^{[i]}\right\}\right] \quad (5.12)$$

with

$$\begin{aligned}\boldsymbol{B}_s^{[i]} \triangleq \sum_{l=0}^{L_p-1}\Big(\Big(\boldsymbol{r}[l] - \boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a}^{[i]}\Big)\Big(\boldsymbol{r}[l] - \boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a}^{[i]}\Big)^{\mathrm{H}}\Big) \\ + \sum_{l=L_p}^{L-1}\Big(\Big(\boldsymbol{r}[l] - \boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{a}^{[i]}\Big)\Big(\boldsymbol{r}[l] - \boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{a}^{[i]}\Big)^{\mathrm{H}} \\ + \boldsymbol{S}[l]\boldsymbol{E}^{[i]}[l]\boldsymbol{A}^{[i]}(\boldsymbol{A}^{[i]})^{\mathrm{H}}\Big(\boldsymbol{E}^{[i]}[l]\Big)^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Big). \quad (5.13)\end{aligned}$$

Similarly to (4.16), the right-hand side of (5.12) is (up to a proportionality constant) a complex Wishart distribution.

Comparing (5.13) and (4.17), the terms containing the covariance matrix of the channel weight estimates does not exist in (5.13). This is due to the factorization of the auxiliary distribution of the channel weights in (5.1). Similarly to (4.18), the expectation of $\boldsymbol{\Xi}$ is

$$\boldsymbol{\Omega}_{\boldsymbol{w},s}^{[i+1]} \triangleq E_{q_{\Xi}^{[i+1]}}\left\{\boldsymbol{\Xi}\right\} = \left(\frac{\boldsymbol{B}_s^{[i]}}{L+N_c}\right)^{-1}. \quad (5.14)$$

When the noise is white and Gaussian, the mean of the noise variance inverse is given by

$$E_{q_{\varsigma}^{[i+1]}}[\varsigma] = \left(\frac{\mathrm{tr}\{\boldsymbol{B}_s^{[i]}\}}{LN_c}\right)^{-1}. \quad (5.15)$$

5.2.3 Interference Cancellation and Single-User Decoding

The process of updating the symbol sequence of user k, $q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k)$ is carried out in two functional blocks: interference cancellation and single-user decoding.

Similarly to (4.21), the minimization yields the update rule

$$q_{\mathrm{d_k}}^{[i+1]}(\boldsymbol{d}_k) \;=\; q_{\mathrm{d_{k,c}}}^{[i+1]}(\boldsymbol{d}_{k,c}) \;\propto\; p(\boldsymbol{d}_k)\exp\left[E_{q_{\Xi}^{[i]}}\left\{E_{q_{\mathrm{d}_{\bar{k}}}^{[i]}}\left\{E_{q_{\mathrm{a}}^{[i]}}\left\{\log p\left(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\Xi\right)\right\}\right\}\right\}\right]. \tag{5.16}$$

Inserting the prior distribution (3.5) into (5.16) and performing the expectation in (5.16) yields the same result as (4.22), however, with the metrics

$$\gamma_k^{[i+1]}[l] \quad \triangleq \quad 2\mathrm{Re}\left\{\left(\left(a_k^{[i]}\right)^{*}\boldsymbol{s}_k[l]^{\mathrm{H}}\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}\right)\left(\boldsymbol{r}[l] \;-\; \sum_{\substack{j=1\\ j\neq k}}^{K}\boldsymbol{s}_j[l]a_j^{[i]}\tilde{d}_j^{[i]}[l]\right)\right\}. \tag{5.17}$$

The formula in (5.17) is much simpler than (4.23). It can be observed that the terms containing the second cross-moments of distinct channel weights and the variance representing the precision of the code symbol estimates do not appear in (5.17).

5.3 Scheduling

Similarly to the DM receiver discussed in Chapter 4, different scheduling schemes can be applied in the DM receiver considered here. Every scheduling scheme leads to a guaranteed convergence in divergence. However, the iterative process with different scheduling schemes may end in different local optima. It was shown in the previous chapter that the DM3 receiver achieves the best performance-complexity trade-off. In the following, we only consider the scheduling scheme originally used in the DM3 receiver, where the channel weights, the noise variance inverse and data symbols are estimated sequentially. The scheduling of the two versions of the DM receiver - DM5 and DM6 - is reported in Table 5.1.

Compared to the DM3 receiver, the DM5 receiver differs only in the channel weight estimation; the former receiver performs joint-user channel weight estimation, while the latter estimates the channel for individual users separately. The consequence of per-user channel weight estimation is twofold:

DM5	$q_{\mathrm{a}_1}(a_1) \ldots \rightarrow q_{\mathrm{a}_K}(a_K) \rightarrow q_\varsigma(\varsigma) \rightarrow q_{\mathrm{d}_1}(\boldsymbol{d_1}) \ldots$ $\rightarrow q_{\mathrm{a}_1}(a_1) \ldots \rightarrow q_{\mathrm{a}_K}(a_K) \rightarrow q_\varsigma(\varsigma) \rightarrow q_{\mathrm{d}_K}(\boldsymbol{d_K})$
DM6	$q_{\mathrm{a}_1}(a_1) \rightarrow q_\varsigma(\varsigma) \rightarrow q_{\mathrm{d}_1}(\boldsymbol{d_1}) \ldots$ $\rightarrow q_{\mathrm{a}_K}(a_K) \rightarrow q_\varsigma(\varsigma) \rightarrow q_{\mathrm{d}_K}(\boldsymbol{d_K})$

Table 5.1: Selected update scheduling schemes for the DM receiver with separate-user channel weight estimation.

- The covariance matrix representing the precision of the channel weight estimates is missing in the noise covariance inverse estimation and in the data decoding process.
- The matrix inverse in the joint-user channel weight estimation is not required here so that the complexity of the channel weight estimation is reduced. The complexity of the matrix inverse is $\mathcal{O}(K^3)$. The complexity of the channel weight estimation for all users in the DM5 receiver is $\mathcal{O}(K^2)$.

In the DM6 receiver, the complexity of the channel weight estimation is reduced even further, since only the channel of one user is updated before the user's data is decoded.

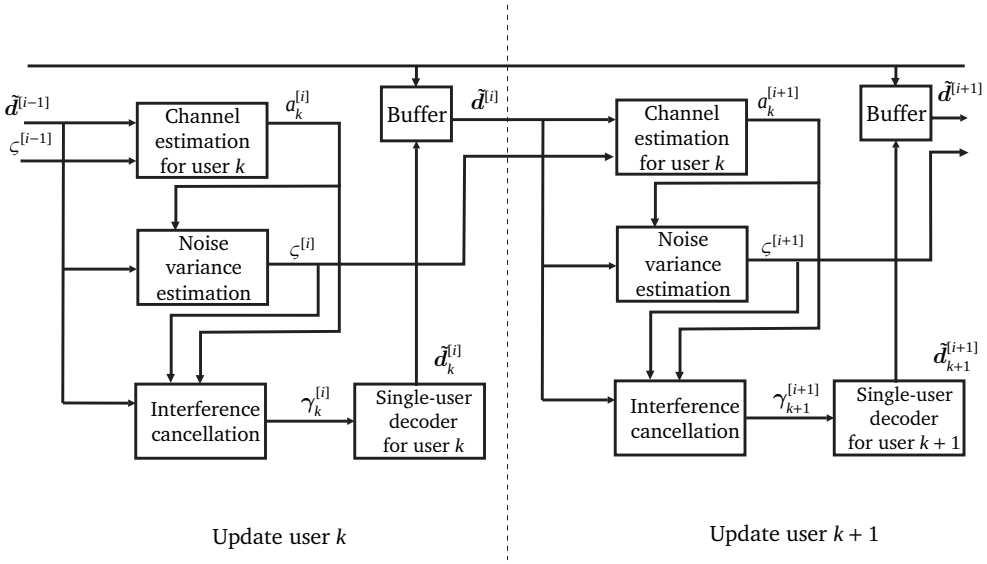

Figure 5.2: Receiver architecture for the DM6 receiver with separate-user channel weight estimation.

The architecture of the DM6 receiver is depicted in Fig. 5.2. The channel weights of one user are estimated first and then the estimation of

the noise variance inverse and the interference cancellation are performed before single-user decoding. The same procedure is repeated for the next user. In the DM5 receiver, the channel weight estimation block for user k is replaced by K parallel blocks, and each block estimates a channel weight for each individual user.

5.4 Simulation Results

The same simulation settings are used as in Chapter 4. The system model detailed in Section 3 is considered. All users employ the same rate $R_c = 1/2$ terminated convolution code with generators $(5, 7)_8$. The generated codewords have length $L_c = 320$ code bits, corresponding to information sequences of length $M = 158$ information bits. Random signature sequences of length $N_c = 8$ chips are assigned to the users. Each codeword is multiplexed with L_p random pilot symbols and each block of $L = L_c + L_p$ symbols is transmitted across a block-fading channel. The effective signal-to-noise ratio is defined as $E_b/N_0 = L/(L_c R_c) \cdot E_s/N_0$, where E_s is the energy per code symbol and E_b is the energy per information bit and $N_0 = 1$. All users have the same E_b/N_0. We assume additive white Gaussian noise. In the simulations, the performance of the DM5 and DM6 receivers are evaluated and compared to that of the DM3 receiver derived in Chapter 4.

As Table 5.1 shows, it is obvious that the DM6 receiver has a lower convergence rate than the DM5 receiver. This can be explained by the different scheduling schemes. In the DM5 receiver, the channel weights of all users are updated before decoding the information of a single-user information, while only the channel weight of one user is updated in the DM6 receiver. We plot the BER performance of the DM6 receiver versus the number of stages in Fig. 5.3. In one stage, the data symbol estimates of all users are updated once. It can be seen that in a low load system, e.g., $K = 32$, $K = 34$ and $K = 36$, most performance improvement is achieved when the number of implemented stages is 12. For a high load system, e.g., $K = 38$ and $K = 40$, no further improvement is observed after $s = 15$ stages. The BER performance is enhanced with the number of performed stages. In Figure 5.4 the performances of the different versions of the DM receiver are compared when the numbers of stages implemented are $s = 8$ and $s = 15$.

In Fig. 5.4 the BER performance of the DM6 receiver after $s = 8$ and $s = 15$ stages under different conditions are plotted. In both plots, the DM6 receiver with unknown noise variance and known channel exhibits

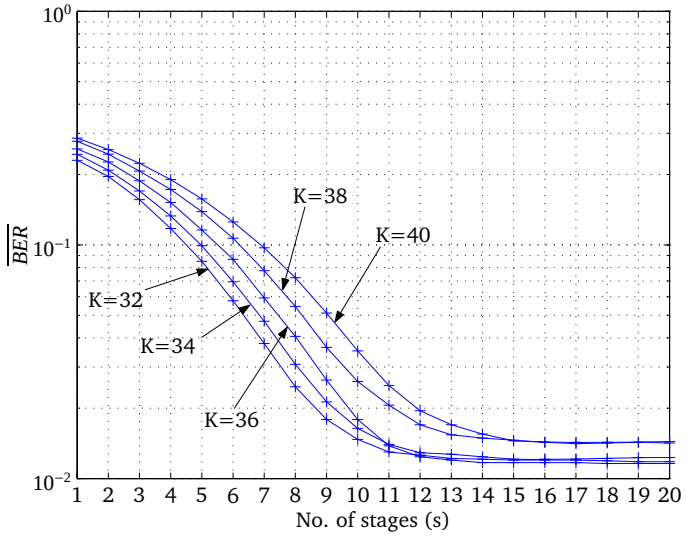

Figure 5.3: Convergence rate of the DM6 receiver versus the number of stages with the number of users as a parameter. For one stage, the symbol estimates of all users are updated once; $E_b/N_0 = 16$ dB.

the best performance. It can also be seen that the performance of the DM6 receiver operating under these conditions is very good after 8 stages, which indicates a fast convergence rate. The DM6 receiver operating without knowledge of the noise variance and channel weights performs worse when $K < 32$. However, after $s = 15$ stages, it performs better than the DM6 receiver that knows the noise variance. We observe that the DM6 receiver needs more stages to converge when it does not know the noise variance than when it has this information available. Provided the DM receiver with unknown noise variance performs a sufficiently large number of stages, its BER performance is better than when the noise variance is known. This phenomenon is similar to that observed for the DM3 receiver and the explanation can be found in Chapter 4. In addition, we also observe that the DM3 receiver performs better when it knows the channel weights than when it does not have this information and therefore has to estimate these weights.

The BER performance of the DM5 receiver implementing $s = 8$ and $s = 15$ stages under different conditions is depicted in Fig. 5.5. Again in both plots, the DM5 receiver with unknown noise variance and known channel weights shows the best performance. The order of the different curves in Fig. 5.5 is very similar to those obtained for the DM6 receiver in Fig. 5.4.

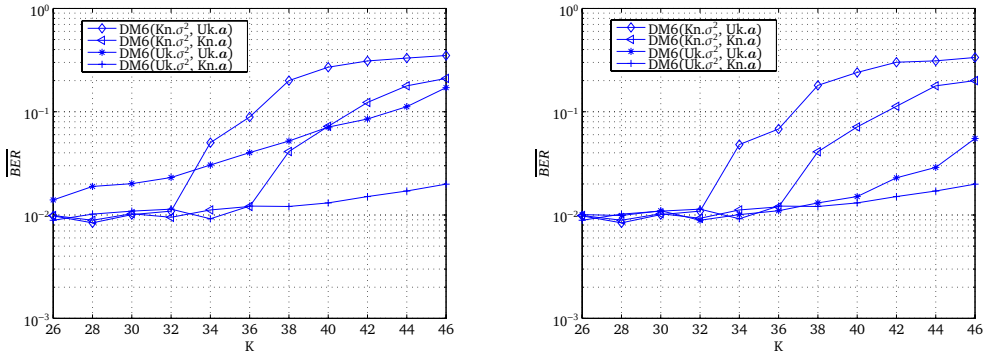

Figure 5.4: The BER performance of the DM6 receiver versus the number of users after $s = 8$ (left) and $s = 15$ stages (right) for $Eb/N_0 = 16$ dB. ("Kn" stands for known channel and "Uk" stands for unknown channel).

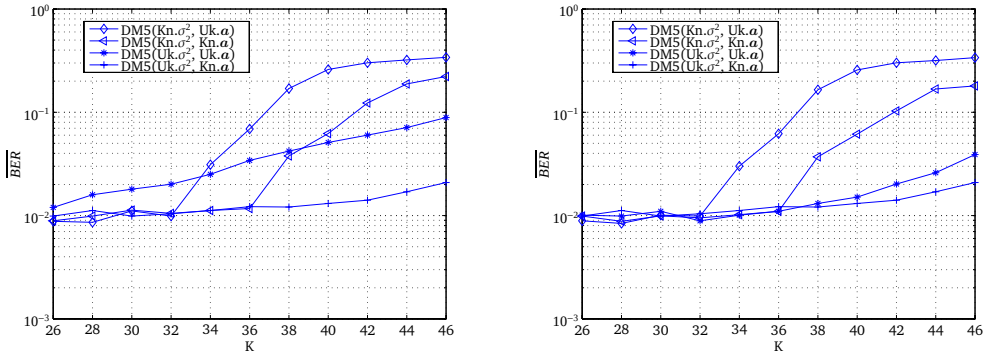

Figure 5.5: The BER performance of the DM5 receiver versus the number of users after $s = 8$ (left) and $s = 15$ stages (right) for $Eb/N_0 = 16$ dB. ("Kn" stands for known channel and "Uk" stands for unknown channel).

Comparing Fig. 5.5 with Fig. 5.4, it can be seen that the DM5 receiver outperforms the DM6 receiver after 8 stages. It also can be seen that at a high SNR regime, the performance of the DM5 receiver is slightly better than the performance of the DM6 receiver after 15 stages. This also indicates that the DM6 receiver has a slower convergence rate than the DM5 receiver.

In Fig. 5.6, the BER performance of the DM3, DM5 and DM6 receivers implementing $s = 15$ stages is evaluated and compared at $E_b/N_0 = 16$ dB. We observe that with an increasing number of users, the DM3 receiver has the best performance. The performance of the DM5 receiver is the second best. However, from a complexity viewpoint, the order is reversed. The DM6 receiver has the lowest complexity and the third best in terms of

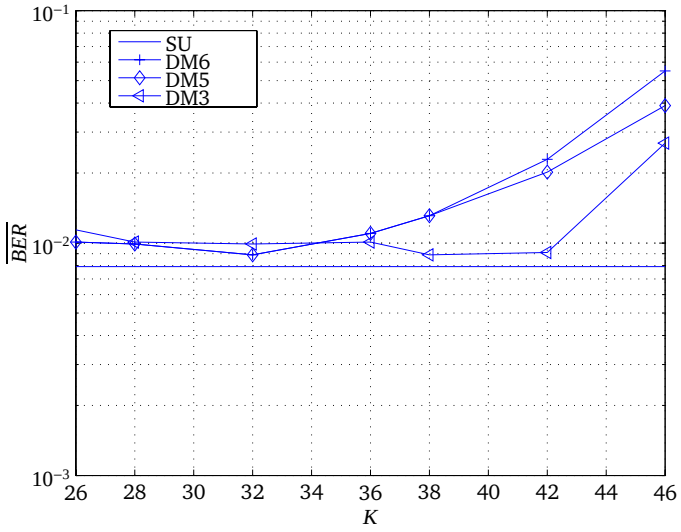

Figure 5.6: Comparison of the performance of the DM3, the DM5 and the DM6 receivers after $s = 15$ stages for $E_b/N_0 = 16$dB.

performance is the DM5 receiver.

To summarize this chapter, we have applied a further factorization of the auxiliary distribution of the channel weights in the DM method. This has resulted in an iterative receiver performing separate-user channel weight estimation, noise variance inverse estimation, interference cancellation followed by a bank of single-user decoders. As expected and shown in the simulation results, the performance of the resulting receiver is not as good as the one in Chapter 4. The complexity is reduced at the cost of performance degradation. From this example, we can see that the DM method is a flexible tool to trade-off performance for the complexity when needed.

Chapter 6

Summary and Conclusions

This chapter summarizes the main findings and results of this thesis and draws conclusions based on these outcomes.

6.1 Summary

Many different iterative receiver structures have been suggested for iterative multiuser decoding in recent years. Most of them are partially or purely based on heuristic approaches or arguments. As pointed out in Chapter 1, there is a lack of a formal optimization framework for designing iterative receivers that exchange soft symbols. This serves as a main motivation for this thesis. The goal of this thesis is to provide a systematic and holistic framework for the design of iterative receivers that operate with soft symbols.

A theoretical DM framework inspired by the VBEM approach is presented in Chapter 2 for the problem of multiuser decoding and parameter estimation. The DM method approximates the target posterior joint distribution by an auxiliary distribution in an iterative manner. To make the complexity of the iterative process tractable the auxiliary distribution is constrained to factorize in a pre-determined manner. As a by-product, the factors of the auxiliary distribution can be seen as approximations of the marginal posterior distributions of some model parameters. The KL divergence between the auxiliary distribution and the target joint distribution is guaranteed to be non-decreasing over the iterations. The DM method is closely related to other algorithms such as EM, SAGE, GEM, etc., since it can be seen as a generalization of these methods.

The signal model for a CDMA system operating in flat fading channels and preliminaries of multiuser decoding are described in Chapter 3.

In order to design a CDMA receiver for data decoding in case of unknown noise variance and channel weights, we consider two specific factorizations of the auxiliary distributions leading to the two iterative receiver structures described in Chapter 4 and Chapter 5.

In Chapter 4, by using the first factorization, the resulting CDMA receiver performs joint-user channel weight estimation, noise variance inverse estimation, interference cancellation and single-user decoding in a iterative manner.

The resulting estimator of the channel weights is a linear MMSE estimator which depends on the posterior means and variances of the code symbols and the mean of the noise variance inverse. The channel estimator provides not only estimates of the channel weights but also a covariance matrix which is an indicator of the estimation accuracy.

The interference cancellation structure results naturally within the DM framework. The interference is calculated based on the posterior means and variances of the code symbols, for all interfering users, as well as the channel weight estimates and the covariance matrix of these estimates. The output of the interference cancellation scheme provides extrinsic values for the single-user decoders.

The single-user decoders compute the probabilities of the individual code symbols based on the conventional maximum *a posteriori* algorithm such as the BCJR algorithm.

The above components are obtained purely based on systematic and strict derivations within the DM framework. Thus, once the components are activated sequentially, the KL divergence between the auxiliary distribution and the target posterior distribution is guaranteed to be non-decreasing.

Clearly, different versions of a given iterative receiver structure are generated by activating its components in different orders (scheduling scheme). We are lacking a criterion for optimizing the scheduling scheme; thus, four different scheduling schemes are chosen heuristically. These different versions of the DM receiver are evaluated by means of Monte Carlo simulations. Among these four versions, the DM3 receiver, which first performs channel weight estimation for all users, followed by noise variance inverse estimation and interference cancellation, has the best performance in terms of both BER and complexity.

The DM3 receiver is interpreted and compared to the start-of-the-art receiver [59] for multiuser decoding and channel parameter estimation both in terms of structural feature and simulated performance. Structure-wise, both of them consist of channel weight estimation, successive interference cancellation and single-users decoders. Besides, the DM3

receiver estimate the noise variance inverse, and the latter mitigates the residual interference using a LMMSE filter. The simulation results show that the DM3 receiver outperforms the receiver [59], even if the noise variance is unknown to the former receiver, while the latter has this information available.

In Chapter 5 the DM method is applied to derive a CDMA receiver with separate-user channel weight estimation. The main difference of this DM receiver compared to the one described in Chapter 4 is that it performs channel weight estimation separately for each individual user. In the single-user channel estimator, firstly the interference is estimated based on the posterior means and variances of the code symbols, the channel estimates of the interfering users and the estimated mean of the noise variance inverse. This interference is subtracted from the received signal, and then a least squares estimator is applied to the resulting signal. Except for the channel estimator, the remaining components in this DM receiver have a structure very similar to the corresponding components described in Chapter 4. Compared to the DM receiver with joint-user channel weight estimation, the DM receiver with separate-user channel weight estimation exhibits a performance loss; however, it complexity is reduced.

We have applied the DM method to the problem of joint multiuser decoding, channel weights and noise variance inverse estimation based on two selections of the auxiliary distribution. However, the choice of the auxiliary distribution is not limited to these two.

6.2 Conclusions

The DM method considered in this thesis is an effective tool for iterative receiver design. It is a purely systematic and holistic approach to design each individual component in iterative receivers. The only assumption upon which it relies is the factorization of the auxiliary distribution. The proposed receivers are derived rigorously within the DM framework. The DM design approach is flexible in the way that different receiver architectures can be obtained by selecting different factorization forms. For the resulting iterative receivers the KL divergence between the auxiliary distribution and the target posterior distribution is guaranteed to be non-decreasing over the iterations.

Through the work done in this thesis, we have gained many new insights on iterative receiver design:

Soft symbols As a result of the expectation performed in every updating step of the DM method, soft code symbols are naturally used to estimate the channel parameters and to perform interference cancellation. This application of the DM method provides a theoretical foundation for receivers that pass through soft information. Regarding the type of the soft information, extrinsic values for code symbols are forwarded to the single-user decoders, which is in accordance with the messages computed using BP. However, the single-user decoders feed posterior probabilities to the interference cancellation device and the channel parameter estimator. This is different from the extrinsic values generated by BP. The BP algorithm guarantees that the exact marginals can be computed for graphs without circles. The factor graph representing the signal model of a multi-user system like the one considered here contains many short cycles. Therefore, approximated marginals are obtained by using extrinsic values in this case. Simulation results show that in high SNR regime the DM receiver with posterior probabilities fed back for interference cancellation can support more users than the receiver providing extrinsic values for interference cancellation.

Uncertainty of the parameter estimates At each iteration the distribution of the parameters is updated. Thus, not only point estimates, which usually coincide with the first moments of the unknown variables, but also the second central moment estimates are provided to the other components. These central moments provide a measure of the accuracy of the estimates. For example, in our application of the DM method, the auxiliary distribution of the channel weights turns out to be a Gaussian distribution determined by its mean and covariance matrix. Both quantities are provided by the channel estimator to the other components.

Mitigation of residual interference Conventionally, the residual interference after interference cancellation is mitigated by using a LMMSE filter. In the DM receiver residual interference suppression is embedded in the noise variance inverse estimation. Thus, it is not necessary to implement the additional LMMSE filter after the interference cancellation module.

Annealing effect The simulation results shows that in high SNR regime, the DM receiver having no knowledge of the noise variance can support more users than the DM receiver which knows the noise variance. This

effect is similar to the annealing phenomenon in physics. The explanation lies in the fact that the proposed receivers are sub-optimal and can only approach a local optimum. The final result after convergence depends on the shape of the cost function and the initialization. Additional unknown parameters may change the shape of the joint posterior function and thus make the suboptimal solutions close to the optimal one.

6.3 Future Work

There are several open topics worth a further study. First of all, the application of the DM method for CDMA receivers under more realistic channels, for example, frequency selective fading channels, shall be investigated. Other parameters assumed known in the investigation, for example, channel correlation matrix, can be seen as additional unknown quantities to be estimated using the DM method. Other factorization of the auxiliary distributions can also be investigated for the problem at hand.

The DM method has been applied in this work to design iterative receivers for multi-user DS-CDMA systems. However, the developed theoretical framework is generic and can potentially be applied to any wireless multi-user access system. A promising application is to MIMO-OFDM, where estimation of the time-variant multi-dimensional transfer functions of the channels of multiple users is a challenging task, especially in fast time-varying scenarios. As a matter of fact, some follow-up works on this topic, which have been strongly inspired by the results of this thesis, have been recently published [52][33].

Another interesting topic is the optimization of scheduling schemes of the identified iterative structures. The same DM algorithm with different scheduling schemes can converge to different local optima. An interesting open issue is whether there is any criterion that can capture the difference between the local optima?

Appendix A

Theoretical Results

A.1 Functionals and Functional Derivatives

In this section, the concepts of functional and functional derivative are described in order to understand the derivation of the updating steps in the DM algorithm.

Let $y(x)$ be a function of a variable $x \in [a, b]$, $a, b \in \mathbb{R}$ and $a \leq b$. Here, a function is a mapping from a set of real values to another set of real values.

Functional A functional is a mapping from a class $\mathbb{E}$ of functions Y onto the set of real numbers,

$$F : Y \rightarrow \mathbb{R}, \; y \rightarrow F[y].$$

The class $\mathbb{E}$ can be, for instance, a class of functions defined on an interval $[a, b]$ that are integrable.

Note that $F[y]$ is a function of all the values of $y(x)$ when x ranges in the interval $[a, b]$. A functional takes as input a function in its domain - not the value $y(x)$ of the function at a specific point x. The output of a functional is a number.

Functional derivative Consider a functional $F[y]$. The functional derivative at y^0 is defined as

$$dF = F(y^0 + dy) - F(y^0) = F'(y)\Big|_{y^0} dy$$

where y^0 is an arbitrary function in the domain Y.

When considering a functional $F[y_1, y_2]$ of two variables y_1, y_2, the derivative of F is given by

$$dF = \frac{\partial F}{\partial y_1}\bigg|_{y_1^0} dy_1 + \frac{\partial F}{\partial y_2}\bigg|_{y_2^0} dy_2 \tag{A.1}$$

where y_1^0 and y_2^0 are arbitrary functions in the domain of y_1 and y_2, respectively. The expression in (A.1) can be generalized to a function of N variables as follows

$$dF = \sum_{n=1}^{N} \frac{\partial F}{\partial y_n}\bigg|_{y_n^0} dy_n = \sum_{n=1}^{N} \alpha \left(\frac{1}{\alpha} \frac{\partial F}{\partial y_n}\bigg|_{y_n^0} \right) dy_n \tag{A.2}$$

where α denotes a real number.

Functional derivative In case that $N \to \infty$ and $\alpha \to 0$[1], the left-hand side of (A.2) can be rewritten as[2]

$$dF = \int_a^b dx \frac{\delta F}{\delta y(x)}\bigg|_{y^0(x)} \delta y(x) \tag{A.3}$$

where $y^0(x)$ is an arbitrary value of the function $y(x)$ in the domain. The definition of the *functional derivative* $\frac{\delta F}{\delta y(x)}$ can be found in (A.3). The functional derivative can be seen as a response of the functional F to a small change in the function y.

The Euler equation When the functional is an integral form, the Euler's equation gives a powerful formula for quick calculation of the functional derivative. Let

$$F[y] = \int L(x, y(x)) dx. \tag{A.4}$$

The functional due to an infinitesimal change $\delta y(x)$ is calculated as

$$\begin{aligned} F[y + \delta y] &= \int L(x, y + \delta y) dx \\ &= \int \left(L(x, y) + \frac{\partial L(x, y)}{\partial y} \delta y \right) dx. \end{aligned} \tag{A.5}$$

[1] N and α converge in a coordinate way to infinity and not independently.

[2] The definition of an integral is $\int_a^b dx f(x) = \lim_{\epsilon \to 0} \sum_{n=1}^{N} \alpha f(x_n)$, $x_n = a + n\alpha$. Note that $\frac{1}{\alpha}$ is absorbed into $\frac{\delta F}{\delta y(x)}$ and dy_n is replaced by $\delta y(x)$ in (A.3).

The second derivative and higher order derivatives with respect to $y(x)$ are discarded in the r.h.s. of (A.5). Referring to the definition in (A.3), the functional derivative of F in (A.4) with respect to $y(x)$ reads

$$\frac{\delta F}{\delta y(x)} = \frac{\partial L(x,y)}{\partial y}. \tag{A.6}$$

To better understand how to calculate the functional derivative, two examples are given in the following.

Example 1 Let

$$F[y] = \int_0^1 \left(y(x)\right)^2 dx. \tag{A.7}$$

With an infinitesimal change $\delta y(x)$, the functional is calculated as

$$\begin{aligned} F[y+\delta y] &= \int_0^1 [y(x)+\delta y(x)]^2 dx \\ &= \int_0^1 [\left(y(x)\right)^2 + 2y(x)\delta y(x) + \left(\delta y(x)\right)^2] dx \\ &= F[y] + \int_0^1 2y(x)\delta y(x) dx. \end{aligned} \tag{A.8}$$

In the last line, the term $(\delta y)^2$ is discarded. From (A.8) the change of the functional dF reads

$$dF = F[y+\delta y] - F[y] = \int_0^1 2y(x)\delta y(x) dx. \tag{A.9}$$

According to (A.3), the functional derivative with respect to $y(x)$ is

$$\frac{\delta F}{\delta y(x)} = 2y(x). \tag{A.10}$$

Alternatively, letting $L(x, y(x)) = \left(y(x)\right)^2$ and inserting into the Euler's formula (A.6) results in

$$\frac{\delta F}{\delta y(x)} = \frac{\partial y(x)^2}{\partial y(x)} = 2y(x). \tag{A.11}$$

Example 2 A functional can depend on more than one function. For instance,

$$F[y(x), z(x)] = \int_0^1 y(x)^2 z(x)^3 dx. \tag{A.12}$$

The functional derivatives with respect to functions $y(x)$ and $z(x)$ are

$$\frac{\delta F}{\delta y(x)} = 2y(x)z(x)^3 \tag{A.13}$$

$$\frac{\delta F}{\delta z(x)} = 3y(x)^2 z(x)^2. \tag{A.14}$$

A.2 Derivation of Step 1 in the DM Method

In the following, the Lagrange multiplier method is employed to minimize the objective function (2.8) and obtain (2.9) in Step 1.

Note that in the below derivation, $\boldsymbol{\eta}$ is a vector of continuous variables and $\boldsymbol{\theta}$ is a vector of discrete variables. However, $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ are not restricted to either continuous or discrete. With the above choice, the updating steps for both continuous and discrete variables are described in below.

The optimization problem for Step 1 needs

$$q_\eta^{[i+1]}(\boldsymbol{\eta}) = \arg\min_{q_\eta(\boldsymbol{\eta})} \; D\Big(q_\theta^{[i]}(\boldsymbol{\theta}) q_\eta(\boldsymbol{\eta}) \| p(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{y})\Big)$$
$$\text{s.t.} \int d\boldsymbol{\eta} q_\eta(\boldsymbol{\eta}) = 1, \; q_\eta(\boldsymbol{\eta}) \geq 0.$$

Introducing a Lagrange multiplier to account for the above constraint yields the objective functional

$$L(q_\eta(\boldsymbol{\eta})) = \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \int d\boldsymbol{\eta} q_\eta(\boldsymbol{\eta}) \Big(\log q_\eta(\boldsymbol{\eta}) + \log q_\theta^{[i]}(\boldsymbol{\theta}) - \log p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r}) \Big)$$
$$+ \lambda \Big(\int d\boldsymbol{\eta} q_\eta(\boldsymbol{\eta}) - 1 \Big). \tag{A.15}$$

To find the solution (2.8), we use the necessary condition that the first order functional derivative of (A.15) vanishes for this solution:

$$\frac{\partial L(q_\eta(\boldsymbol{\eta}))}{\partial q_{\mathrm{x}}(\boldsymbol{\eta})} = \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \Big(\log q_\eta(\boldsymbol{\eta}) + 1 + \log q_\theta^{[i]}(\boldsymbol{\theta}) - \log p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r}) \Big) + \lambda = 0. \tag{A.16}$$

Solving (A.16) for $q_\eta(\boldsymbol{\eta})$ yields

$$q_\eta(\boldsymbol{\eta}) = \exp\Big[\sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r}) - 1 - \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log q_\theta^{[i]}(\boldsymbol{\theta}) - \lambda\Big]. \tag{A.17}$$

Making use of the condition $\int d\boldsymbol{\eta} q_\eta(\boldsymbol{\eta}) = 1$, we obtain

$$\int d\boldsymbol{\eta} \exp\Big[\sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big] = \frac{1}{C_1} \tag{A.18}$$

where the constant C_1 is defined as

$$C_1 \triangleq \exp\Big[-1 - \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log q_\theta^{[i]}(\boldsymbol{\theta}) - \lambda\Big]. \tag{A.19}$$

Inserting (A.19) in (A.17) yields

$$q_\eta(\boldsymbol{\eta}) = C_1 \exp\Big[\sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r})\Big]. \tag{A.20}$$

The second order functional derivative is computed as

$$\frac{\partial \dfrac{\partial L(q_\eta(\boldsymbol{\eta}))}{\partial q_{\mathrm{x}}(\boldsymbol{\eta})}}{\partial q_{\mathrm{x}}(\boldsymbol{\eta})} = \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \frac{1}{q_\eta(\boldsymbol{\eta})} > 0. \tag{A.21}$$

Notice that $q_\eta(\boldsymbol{\eta})$ in (A.20) fulfills the other constraint $q_\eta(\boldsymbol{\eta}) \geq 0$. Thus, the objective function is minimized by (A.20).

Furthermore, if $\boldsymbol{\eta}$ is complete data, $p(\boldsymbol{r}|\boldsymbol{\eta}, \boldsymbol{\theta}) = p(\boldsymbol{r}|\boldsymbol{\eta})$. Using this property and Bayes' rule yields

$$p(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{r}) = \frac{p(\boldsymbol{r}|\boldsymbol{\eta}) p(\boldsymbol{\eta}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\boldsymbol{r})}. \tag{A.22}$$

Let's consider the deterministic mapping $\boldsymbol{r} = f(\boldsymbol{\eta})$, i.e., $p(\boldsymbol{r}|\boldsymbol{\eta}) = \delta(\boldsymbol{r} - f(\boldsymbol{\eta}))$. Inserting (A.22) with this setting into (A.20) yields

$$\begin{aligned} q_\eta(\boldsymbol{\eta}) &= C_1 \exp\Big[\sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) + \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\eta}|\boldsymbol{\theta}) - \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{r})\Big] \\ &= C_1 C_2 \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\eta}|\boldsymbol{\theta}) \\ &= C_3 \sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\eta}|\boldsymbol{\theta}) \end{aligned} \tag{A.23}$$

where the constant C_2 is defined as

$$C_2 = \exp\Big[\sum_{\boldsymbol{\theta}\in\Theta} q_\theta^{[i]}(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) - \log p(\boldsymbol{r})\Big] \tag{A.24}$$

and $C_3 = C_1 C_2$.

A.3 Proof that the E-step and the M-step in (2.14) and (2.15) respectively are an Instance of the DM Method

E-step: Inserting (2.13) in (2.9) yields (2.14).

M-step: Computa

$$D\Big(q_\theta(\boldsymbol{\theta}) q_\eta^{[i]}(\boldsymbol{\eta}) \| p(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{y})\Big)\Big|_{q_\theta(\boldsymbol{\theta})=\delta(\boldsymbol{\theta}-\boldsymbol{\theta}_0)} \propto -\exp\Big[\int d\boldsymbol{\eta} q_\eta^{[i]}(\boldsymbol{\eta}) \log p(\boldsymbol{\eta}, \boldsymbol{\theta}_0|\boldsymbol{r})\Big]. \tag{A.25}$$

Minimizing the l.h.s. of (A.25) is equivalent to maximizing the integral in the r.h.s. of (A.25). The latter operation is equivalent to (2.15).

Appendix B

Specific Derivations for the DM Receivers

B.1 Derivation of the Updating Step for the Auxiliary Distribution of the Channel Weights of All Users

From (4.7), the expectation in (4.10) with respect to $q_{\mathrm{d}}^{[i]}(\boldsymbol{d})$ and $q_{\Xi}^{[i]}(\boldsymbol{\Xi})$ is computed as follows:

$$\begin{aligned} & E_{q_{\Xi}^{[i]}}\left\{E_{q_{\mathrm{d}}^{[i]}}\left\{\log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\boldsymbol{\Xi})\right\}\right\} \propto^{e} \\ & \quad -\mathrm{tr}\Bigg\{\left(\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]}\right)^{-1}\Bigg(\sum_{l=0}^{L_p-1}(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a})(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a})^{\mathrm{H}} \\ & \qquad + E_{q_{\mathrm{d}}^{[i]}}\Bigg\{\sum_{l=L_p}^{L-1}\boldsymbol{r}[l]\boldsymbol{r}[l]^{\mathrm{H}} - \boldsymbol{r}[l]\boldsymbol{a}^{\mathrm{H}}\boldsymbol{D}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}} \\ & \qquad\quad - \boldsymbol{S}[l]\boldsymbol{D}[l]\boldsymbol{a}\boldsymbol{r}[l]^{\mathrm{H}} + \boldsymbol{S}[l]\boldsymbol{D}[l]\boldsymbol{a}\boldsymbol{a}^{\mathrm{H}}\boldsymbol{D}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Bigg\}\Bigg)\Bigg\}. \end{aligned} \tag{B.1}$$

Exchanging the order of summation and expectation in the last four terms on the r.h.s., the expectation (B.1) results in

$$\begin{aligned} & \boldsymbol{r}[l]\boldsymbol{r}[l]^{\mathrm{H}} - \boldsymbol{r}[l]\boldsymbol{a}^{\mathrm{H}}\tilde{\boldsymbol{D}}^{[i]}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}} - \boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{a}\boldsymbol{r}[l]^{\mathrm{H}} \\ & \qquad + \boldsymbol{S}[l]\boldsymbol{A}\,E_{q_{\mathrm{d}}^{[i]}}\left\{\boldsymbol{d}[l]\boldsymbol{d}[l]^{\mathrm{H}}\right\}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}} \\ = & \boldsymbol{r}[l]\boldsymbol{r}[l]^{\mathrm{H}} - \boldsymbol{r}[l]\boldsymbol{a}^{\mathrm{H}}\tilde{\boldsymbol{D}}^{[i]}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}} - \boldsymbol{S}[l]\tilde{\boldsymbol{D}}^{[i]}[l]\boldsymbol{a}\boldsymbol{r}[l]^{\mathrm{H}} \\ & \qquad + \boldsymbol{S}[l]\boldsymbol{A}\tilde{\boldsymbol{d}}^{[i]}[l]\tilde{\boldsymbol{d}}^{[i]}[l]^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}} \\ & \qquad + \boldsymbol{S}[l]\boldsymbol{A}\boldsymbol{E}^{[i]}[l]\boldsymbol{E}^{[i]}[l]^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}. \end{aligned}$$

Inserting in (B.1) yields (4.10). The expectation in (4.16) resulting in (4.17) is determined in a similar manner.

B.2 Derivation of the Updating Step for the Codeword Distribution

B.2.1 Computing the soft QPSK symbols The soft QPSK symbols are defined as

$$\tilde{d}_k[l]^{[i]} \triangleq \sum_{\boldsymbol{d}_k \in \mathcal{D}_k^Q} d_k[l] q_{\mathrm{d_k}}^{[i]}(\boldsymbol{d}_k) = \sum_{\boldsymbol{d}_k \in \mathcal{D}_k^Q} \left[(2c_k[2l]-1) + j(2c_k[2l+1]-1)\right] q_{\mathrm{d_k}}^{[i]}(\boldsymbol{d}_k). \tag{B.2}$$

Since the mapping $\boldsymbol{c}_k \to \boldsymbol{d}_k$ is one-to-one, (B.2) can be rewritten as

$$\begin{aligned} \tilde{d}_k[l]^{[i]} &= \sum_{\boldsymbol{c}_k \in C_k} \left[(2c_k[2l]-1) + j(2c_k[2l+1]-1)\right] q_{\mathrm{c_k}}^{[i]}(\boldsymbol{c}_k) \\ &= \tilde{c}_k[2l]^{[i]} + j\tilde{c}_k[2l+1]^{[i]}. \end{aligned} \tag{B.3}$$

Note that as a result of the expectation in (B.2), the soft symbols are the means of the modulation symbols given the code bit probabilities.

B.2.2 Computing the input for the single-user decoder Using the results in (4.24) and (4.8) for codewords and the results in (4.13) and (4.14) for the channel weights, we compute the expectation in (4.21) as follows

$$\begin{aligned} &E_{\Xi^{[i]}}\left\{E_{q_{\mathrm{a}}^{[i]}}\left\{E_{q_{\mathrm{d}_{\bar{k}}}^{[i]}}\left\{\log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d},\Xi)\right\}\right\}\right\} \\ &\quad \propto^{e} -\mathrm{tr}\Bigg\{\boldsymbol{\Omega}_{\boldsymbol{w}}^{[i]} \sum_{l=L_p}^{L-1}\Big(\big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\tilde{\boldsymbol{D}}_k^{[i]}[l]\boldsymbol{a}^{[i]}\big)\big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\tilde{\boldsymbol{D}}_k^{[i]}[l]\boldsymbol{a}^{[i]}\big)^{\mathrm{H}}\Big) \\ &\qquad\qquad + \boldsymbol{S}[l]\tilde{\boldsymbol{D}}_k^{[i]}[l]\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}\tilde{\boldsymbol{D}}_k^{[i]}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Bigg\}, \end{aligned} \tag{B.4}$$

where $\tilde{\boldsymbol{D}}_k^{[i]}[l] = \tilde{\boldsymbol{D}}^{[i]}[l](\boldsymbol{I} - \mathrm{diag}\{\boldsymbol{e}_k\}) + \boldsymbol{D}[l]\mathrm{diag}\{\boldsymbol{e}_k\}$ and $\boldsymbol{e}_k$ is a unit vector with element k equal to 1. Terms independent of $d_k[l]$, $l = L_p, \ldots, L-1$ are discarded in (B.4).

The compact expression in (4.22) is obtained from (B.4) by expanding the matrix-vector multiplications and discarding terms irrelevant for determining the codeword distribution of user k. The trace operator can be eliminated by exploiting its cyclic property and decomposing the error

covariance matrix. In particular, since the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}$ is always positive semi-definite we can employ a singular value decomposition, $\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]} = \boldsymbol{U}^{[i]}\Lambda^{[i]}(\boldsymbol{U}^{[i]})^{\mathrm{H}}$, to obtain the last term in (4.23):

$$\begin{aligned}
&\mathrm{tr}\Big\{\Omega_{\boldsymbol{w}}^{[i]}\sum_{l=L_p}^{L-1}\boldsymbol{S}[l]\tilde{\boldsymbol{D}}_k^{[i]}[l]\boldsymbol{\Sigma}_{\boldsymbol{a}}^{[i]}\tilde{\boldsymbol{D}}_k^{[i]}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Big\}\\
&\qquad=\mathrm{tr}\Big\{\Omega_{\boldsymbol{w}}^{[i]}\sum_{l=L_p}^{L-1}\boldsymbol{S}[l]\tilde{\boldsymbol{D}}_k^{[i]}[l]\boldsymbol{U}^{[i]}\Lambda^{[i]}(\boldsymbol{U}^{[i]})^{\mathrm{H}}\tilde{\boldsymbol{D}}_k^{[i]}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Big\}\\
&\qquad=\sum_{l=L_p}^{L-1}\mathrm{tr}\Big\{(\boldsymbol{U}^{[i]})^{\mathrm{H}}\tilde{\boldsymbol{D}}_k^{[i]}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\Omega_{\boldsymbol{w}}^{[i]}\boldsymbol{S}[l]\tilde{\boldsymbol{D}}_k^{[i]}[l]\boldsymbol{U}^{[i]}\Lambda^{[i]}\Big\}\\
&\qquad\qquad\propto^{e}\sum_{l=L_p}^{L-1}d_k[l]2\mathrm{Re}\Big\{\sum_{j\neq k}(\boldsymbol{u}_k^{[i]})^{H}\boldsymbol{s}_k^{\mathrm{H}}[l]\boldsymbol{\Omega}^{[i]}\boldsymbol{s}_j\boldsymbol{u}_j^{[i]}\lambda_j^{[i]}\Big\}. \quad \text{(B.5)}
\end{aligned}$$

In this expression $\boldsymbol{u}_k^{[i]}$ is the k-th column of eigenvector matrix $\boldsymbol{U}^{[i]}$ and $\lambda^{[i]}$ is the j-th eigenvalue of the diagonal matrix $\Lambda^{[i]}$.

B.2.3 Initialization of the iterative process The vector of channel weights $\boldsymbol{a}^{[0]}$ is initialized based on the pilot symbols $\boldsymbol{D}_p[0],\ldots,\boldsymbol{D}_p[L_p-1]$ using a least-square estimator

$$\boldsymbol{a}^{[0]}=\Big(\sum_{l=0}^{L_p-1}(\boldsymbol{D}_p[l])^{\mathrm{H}}\boldsymbol{R}[l]\boldsymbol{D}_p[l]\Big)^{-1}\sum_{l=0}^{L_p-1}(\boldsymbol{D}_p[l])^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{r}[l].$$

Initial symbol estimates are obtained also by means of a least-square estimator

$$\check{\boldsymbol{d}}^{[0]}=\Big(\sum_{l=0}^{L_p-1}\boldsymbol{A}^{[0]}[l]^{\mathrm{H}}\boldsymbol{R}[l]\boldsymbol{A}^{[0]}[l]\Big)^{-1}\sum_{l=0}^{L_p-1}\boldsymbol{A}^{[0]}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{r}[l].$$

Note that $\boldsymbol{A}^{[0]}[l]$ is a diagonal matrix representation of the vector $\boldsymbol{a}^{[0]}[l]$. The initial noise variance inverse estimate is then

$$\begin{aligned}
\varsigma^{[0]}=\Big(LN_c+2\Big)\Big(&\sum_{l=0}^{L_p-1}\Big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a}^{[0]}\Big)^{\mathrm{H}}\Big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\boldsymbol{D}_p[l]\boldsymbol{a}^{[0]}\Big)\\
&+\sum_{l=L_p}^{L-1}\Big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\check{\boldsymbol{D}}^{[0]}[l]\boldsymbol{a}^{[0]}\Big)^{\mathrm{H}}\Big(\boldsymbol{r}[l]-\boldsymbol{S}[l]\check{\boldsymbol{D}}^{[0]}[l]\boldsymbol{a}^{[0]}\Big)\Big).
\end{aligned}$$

B.3 Derivation for Updating the Auxiliary Distribution of the Channel Weight of User k

Letting $\tilde{a}_j^{[i]} \triangleq E_{q_{\tilde{a}_k}^{[i]}}\left[a_j\right]$, the expectation of w.r.t. reads

$$\begin{aligned} E_{q_{\tilde{a}_k}^{[i]}}\Big[\log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d})\Big] \propto -\sum_l (-a_k^*(\boldsymbol{D}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{r}[l])_k - (\boldsymbol{r}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l]\boldsymbol{D}[l])_k a_k \\ + \sum_{j\neq k} a_k^*(\boldsymbol{D}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{D}[l]\boldsymbol{S}[l])_{kj}\tilde{a}_j^{[i]} \\ + \sum_{j\neq k} (\tilde{a}_j^{[i]})^*(\boldsymbol{D}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l]\boldsymbol{D}[l])_{kj}a_k \\ + |a_k|^2(\boldsymbol{D}[l]^{\mathrm{H}}\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l]\boldsymbol{D}[l])_{kk}). \end{aligned} \tag{B.6}$$

The above equation can be further reduced to

$$\begin{aligned} E_{q_{\tilde{a}_k}^{[i]}}\Big[\log p(\boldsymbol{r}|\boldsymbol{a},\boldsymbol{d})\Big] &\propto \sum_l \Big(2Re\Big\{a_k^* d_k^*(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{r}[l])_k\Big\} \\ &- 2Re\Big\{\sum_{j\neq k} a_k^* d_k^*[l]d_j[l]\tilde{a}_j^{[i]}(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l])_{kj}\Big\} - |a_k|^2 d_k^*[l]d_k[l](\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l])_{kk}\Big) \\ &= \sum_l 2Re\Big\{a_k^* d_k^*[l]((\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{r}[l])_k - \sum_{j\neq k}\tilde{a}_j^{[i]}d_j[l](\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l])_{kj})\Big\} \\ &\quad - |a_k|^2\sum_l |d_k[l]|^2(\boldsymbol{S}[l]^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{S}[l])_{kk}. \end{aligned} \tag{B.7}$$

B.4 Derivation of the LMMSE Channel Estimator

The LMMSE channel estimate [59] assumming white noise is given by

$$\begin{aligned} \hat{\boldsymbol{a}} &= \boldsymbol{\Sigma}_{\boldsymbol{a}}^H\tilde{\boldsymbol{X}}^H\boldsymbol{S}^H(\boldsymbol{S}(\tilde{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{a}}\tilde{\boldsymbol{X}}^H + \boldsymbol{\Lambda})\boldsymbol{S}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{r} \\ &= \boldsymbol{\Sigma}_{\boldsymbol{a}}^H\tilde{\boldsymbol{X}}^H\boldsymbol{S}^H(\boldsymbol{S}\tilde{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{a}}\tilde{\boldsymbol{X}}^H\boldsymbol{S}^H + \boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{r} \end{aligned} \tag{B.8}$$

with $\boldsymbol{\Lambda} = [\sigma_a^2(1-\sigma_{d_1}^2),\ldots,\sigma_a^2(1-\sigma_{d_K}^2)]$.

Using the matrix inversion lemma

$$\boldsymbol{P}\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{P}\boldsymbol{B}^T + \boldsymbol{R}) = (\boldsymbol{P}^{-1} + \boldsymbol{B}^T\boldsymbol{R}^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{R}^{-1} \tag{B.9}$$

the LMMSE channel estimate can be recasted as

$$\hat{\boldsymbol{a}} = (\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} + \tilde{\boldsymbol{X}}^H\boldsymbol{S}^H(\boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{S}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^H\boldsymbol{S}^H(\boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{r}. \tag{B.10}$$

Applying the matrix inversion lemma

$$\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B}+\boldsymbol{I}) = (\boldsymbol{A}+\boldsymbol{B}\boldsymbol{B}^T)^{-1}\boldsymbol{B} \tag{B.11}$$

yields

$$\boldsymbol{S}^H(\boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^H+\sigma^2\boldsymbol{I})^{-1} = (\sigma^2\boldsymbol{I}+\boldsymbol{R}\boldsymbol{\Lambda})^{-1}\boldsymbol{S}^H. \tag{B.12}$$

Inserting (B.12) into (B.10), we obtain

$$\hat{\boldsymbol{a}} = (\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}+\tilde{\boldsymbol{X}}^H(\boldsymbol{R}\boldsymbol{\Lambda}+\sigma^2\boldsymbol{I})^{-1}\boldsymbol{R}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^H(\boldsymbol{R}\boldsymbol{\Lambda}+\sigma^2\boldsymbol{I})^{-1}\boldsymbol{S}^H\boldsymbol{r}. \tag{B.13}$$

Bibliography

[1] P. D. Alexander and A. J. Grant. Iterative channel and information sequence estimation in CDMA. In *Proc. IEEE 6th Int. Symp. on Spread-Spectrum Tech. & App.*, pages 593–597, New Jersey, USA, September 2000.

[2] P. D. Alexander, A. J. Grant, and M. C. Reed. Iterative detection in code-division multiple-access with error control coding. *Euro. Trans. Telecommunications*, 9:419–425, September/October 1998.

[3] P. D. Alexander, M. C. Reed, J. A. Asenstorfer, and C. B. Schlegel. Iterative multiuser interference reduction: Turbo CDMA. *IEEE Trans. Commun.*, 47:1008–1014, July 1999.

[4] H. Attias. Inferring parameters and structure of latent variable models by variational Bayes. In *Proc. 15th Conf. on Uncertainty in Artificial Intelligence*, 1999.

[5] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. Inf. Theory*, 20:284–287, March 1974.

[6] M. Beal. *Variational Algorithms For Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, London's Global University, UK, 2003.

[7] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon-limit error-correcting coding and decoding: Turbo-codes. In *Proc. IEEE Intern. Conf. on Commun. 1993 (ICC '93)*, pages 1064–1070, Geneva, Switzerland, May 1993.

[8] H. A. Bethe. Statistical Theory of Superlattices. In *Roy. Soc. London*, page 552, 1935.

[9] J. Boutros and G. Caire. Iterative multiuser joint decoding: Unified framework and asymptotic analysis. *IEEE Trans. Inf. Theory*, 48:1772–1793, July 2002.

[10] G. Caire, R. Müller, and T. Tanaka. Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation. *IEEE Trans. Inf. Theory*, 50:1950–1972, September 2004.

[11] E. Chiavaccini and G. M. Vitetta. MAP symbol estimation on frequency-flat Rayleigh fading channels via a Bayesian EM algorithm. *IEEE Trans. Commun.*, 49:1869–1872, November 2001.

[12] L. P. B. Christensen and J. Larsen. On data and parameter estimation using the variational Bayesian EM algorithm for block-fading frequency selective MIMO channels. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 4, pages 465–468, 2006.

[13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[14] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc., Ser. B*, 39:1–38, January 1977.

[15] D. Divsalar, M. K. Simon, and D. Raphaeli. Improved parallel interference cancellation for CDMA. *IEEE Trans. Commun.*, 46:258–268, February 1998.

[16] A. Duel-Hallen. A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels. *IEEE Trans. Commun.*, 43:421–434, Feb./Mar./Apr. 1995.

[17] A. Duel-Hallen, J. Holtzman, and Z. Zvonar. Multiuser detection for CDMA systems. *IEEE Pers. Commun.*, 2:46–57, April 1995.

[18] U. Fawer and B. Aazhang. A multiuser receiver for code division multiple access communications over multipath channels. *IEEE Trans. Commun.*, 43:1556–1565, February 1995.

[19] J. A. Fessler and A. O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Trans. Signal Processing*, 42:2664–2677, October 1994.

[20] H. El Gamal and E. Geraniotis. Iterative multiuser detection for coded CDMA signals in AWGN and fading channels. *IEEE J. Sel. Areas Commun.*, 18:30–41, January 2000.

[21] T. R. Giallorenzi and S. G. Wilson. Multiuser ML sequence estimator for convolutionally coded asynchronous DS-CDMA systems. *IEEE Trans. Commun.*, 44:997–1008, August 1996.

[22] N. R. Goodman. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Statist.*, 34:152–177, 1963.

[23] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman & Hall/CRC, 2000.

[24] J. Hagenauer. Iterative decoding of binary block and convolutional codes. *IEEE Trans. Inf. Theory*, 42:429–445, March 1996.

[25] B. Hu, A. Kocian, R. Piton, and B. H. Fleury. Performance of a SISO-SAGE based receiver for coded CDMA. In *In Proc. of Winter School on Coding and Inf. Theory*, Feb 2005.

[26] B. Hu, A. Kocian, R. Piton, A. Hviid, B. H. Fleury, and L. K. Rasmussen. Iterative joint channel estimation and successive interference cancellation using a SISO-SAGE algorithm for coded CDMA. In *Proc. IEEE 38th Asilomar Conference on Signals, Systems and Computers*, pages 622–626, Pacific Grove, CA, November 2004.

[27] B. Hu, I. Land, R. Piton, and B. H. Fleury. A Bayesian framework for iterative channel estimation and multiuser decoding in coded DS-CDMA. In *Proceeding of the 50th IEEE Global Comm. Conf.*, Washington D.C., USA, Nov 2007.

[28] B. Hu, I. Land, R. Piton, and B. H. Fleury. Iterative SAGE-based receivers for synchronous coded DS-CDMA. In *Proc. IEE 66th Semi-Annual Vehicular Tech. Conf.*, 2007.

[29] B. Hu, I. Land, L.K. Rasmussen, R. Piton, and B. H. Fleury. A variational Bayesian inference approach to joint multiuser decoding for coded CDMA. *IEEE J. on Sel. Areas in Commun. - Special Issue on Multiuser Detection*, March 2008.

[30] S. Ikeda, T. Tanaka, and S. Amari. Stochastic reasoning, free energy, and information geometry. *Neural Computation*, 16(9):1779–1810, 2004.

[31] T. Jaakkola. *Advanced Mean Field Methods: Theory and Practice.*, chapter Tutorial on variational approximation methods. MIT Press., 2000.

[32] M. I. Jordan, T. S. Jaakkola Z. Ghahramani, and L. K. Saul. *Learning in Graphical Models*, chapter An introduction to variational methods for graphical models., pages 105–162. Kluwer Academic Publisher, 1998.

[33] G. E. Kirkelund, C. N. Manchon, L. P. B. Christensen, E. Riegler, and B. H. Fleury. Variational message-passing for joint channel estimation and decoding in MIMO-OFDM. In *Proc. of the IEEE Global Communications Conference (Globecom 2010)*, Florida, USA, December 2010, to appear.

[34] M. Kobayashi, J. Boutros, and G. Caire. Successive interference cancellation with SISO decoding and EM channel estimation. *IEEE J. Sel. Areas Commun.*, 19:1450–1460, August 2001.

[35] A. Kocian and B. H. Fleury. EM-based joint data detection and channel estimation of DS-CDMA signals. *IEEE Trans. Commun.*, 51:1709–1720, October 2003.

[36] A. Kocian, B. Hu, P. Sørensen, C. Rom, B. H. Fleury, and E. K. Poulsen. Iterative joint data detection and channel estimation of DS/CDMA signals in multipath fading using the SAGE algorithm. In *Proc. 37th IEEE Asilomar Conference on Signals, Systems, and Computers*, pages 443–447, Pacific Grove, CA, USA, November 2003.

[37] A. Kocian, I. Land, and B. H. Fleury. Optimal weighting of soft-information in a SAGE-based iterative receiver for coded CDMA. In *Proc. IEEE Global Telecomms. Conf.*, volume 3, pages 1555–1559, November 2005.

[38] A. Kocian, I. Land, and B.H. Fleury. Joint channel estimation, partial successive interference cancellation, and data decoding for DS-CDMA based on the SAGE algorithm. *IEEE Trans. commun.*, 2007.

[39] F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inf. Theory*, 47:498–518, 2001.

[40] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[41] A. Lampe. Iterative multiuser detection with integrated channel estimation for coded DS-CDMA. *IEEE Trans. Commun.*, 50:1217–1223, August 2002.

[42] H. Li, S. M. Betz, and H. V. Poor. Performance analysis of iterative channel estimation and multiuser detection in multipath DS-CDMA channels. *IEEE Trans. Signal Processing*, 55(5):1981–1993, May 2007.

[43] D. D. Lin and T. J. Lim. A variational free energy minimization interpretation of multiuser detection in CDMA. In *Proc. IEEE Global Telecomm. Conf.*, pages 1570–1576, November 2005.

[44] X. L. Meng and D. B. Rubin. Maximum likelihood via the ECM algorithm: a general framework. *Biometrika*, 80:267–278, 1993.

[45] M. Moher. An iterative multiuser decoder for near-capacity communications. *IEEE Trans. Commun.*, 46:870–880, July 1998.

[46] R. Müller and J. B. Huber. Iterated soft decision interference cancellation for CDMA. In *Proc. 9th Tyrrhenian Int. Workshop Digital Commun.*, 1997.

[47] R. M. Neal and G. E. Hinton. *Learning in Graphical Models*, chapter A View of the EM Algorithm that justifies incremental, sparse and other variants., pages 355–370. Kluwer Academic Publisher, 1998.

[48] L. B. Nelson and H. V. Poor. Iterative multiuser receivers for CDMA channels: An EM-based approach. *IEEE Trans. Commun.*, 44:1700–1710, December 1996.

[49] Mauri Nissilä. *Iterative Receivers for Digital Communications via Variational Inference and Estimation*. PhD thesis, Faculty of Technology, University of Oulu, 2008.

[50] H. V. Poor. Iterative multiuser detection. *IEEE Signal Processing Mag.*, 21(1):81–88, January 2004.

[51] M. C. Reed, C. B. Schlegel, P. D. Alexander, and J. A. Asenstorfer. Iterative multiuser detection for CDMA with FEC: Near-single-user performance. *IEEE Trans. Commun.*, 46:1693–1699, December 1998.

[52] E. Riegler, G. E. Kirkelund, C. N. Manchon, and B. H. Fleury. Merging belief propagation and the mean field approximation: a free energy approach. In *Proc. 6th International Symposium on Turbo Codes and Iterative Information Processing*, Brest, France, 2010.

[53] P. Robertson, P. Hoeher, and E. Villebrun. Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding. *Euro. Trans. Telecomms.*, 2:119–125, March/April 1997.

[54] P. H. Tan. *Simplified Graphical Approaches for CDMA Multi-user Detection, Decoding and Power Control*. PhD thesis, Department of Computer Science and Engineering, Chalmers University of Technology, Sweden, 2005.

[55] P. H. Tan and L. K. Rasmussen. Asymptotically optimal nonlinear MMSE multiuser detection based on multivariate Gaussian approximation. *IEEE Trans. Commun.*, 54:1427–1438, August 2006.

[56] P. H. Tan and L. K. Rasmussen. Belief propagation for coded multiuser detection. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 1919–1923, Seattle, Washington, USA, July 2006.

[57] S. Verdú. *Multiuser Detection*. Cambridge University Press, 1998.

[58] X. Wang and H. V. Poor. Iterative (Turbo) soft interference cancellation and decoding for coded CDMA. *IEEE Trans. Commun.*, 47:1046–1061, July 1999.

[59] J. Wehinger and C. F. Mecklenbräuker. Iterative CDMA multiuser receiver with soft decision-directed channel estimation. *IEEE Trans. Signal Processing*, 54:3922–3934, October 2006.

[60] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. *Advances in Neural Information Processing Systems (NIPS)*, 13:689–695, December 2000.

[61] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Inf. Theory*, 51(7):2282–2312, July 2005.