

Statistical Signal Processing for Data Fusion

Domenico Ciunzo

November 27, 2012

Acknowledgements

First of all, my deep appreciation goes to my advisor, Prof. Francesco Palmieri, who played an important role for the development of the initial stage of my research projects, through its constant supervision and useful comments.

Secondly, my deep and sincere gratitude goes to Prof. Pierluigi Salvo Rossi and Prof. Gianmarco Romano for being precious friends and advisors at the same time; they have taught me how to efficiently interact with the scientific community and how to develop and conduct a research project in a careful and scrupolous way. They also avoided to let me become discouraged when difficulties showed up. I am deeply indebted to them. Furthermore, I would like to wish all the best to my colleagues (and friends) Giuseppe Papa and Pasquale Cuccaro for the Ph.D. they are working toward.

Also, I would like to thank Prof. Yaakov Bar-Shalom and Prof. Peter Willett (University of Connecticut, Storrs, US). They allowed me to spend nearly a year under their supervision; this will always remain an unforgettable experience throughout my life. They have been examples to me of how to perform excellent research and being humble at the same time; I learnt a lot each day I spent talking to them. I am sincerely grateful to them for their availability and I have to say I admire them for the contagious enthusiasm they show in doing research and disseminating knowledge. Again, I would like to wish all the best to Xiufeng Song, Gigi Belfadel, Ting Yuan, Ramona Georgescu, Sora Choi and Balakumar Balasingham for their future.

I would also like to record my gratitude to Mr. Steven Horn and Mrs. Karna Bryan (Center for Maritime Research and Experimentation, CMRE, formerly known as NURC, La Spezia, IT) which I had the pleasure to meet during my summer internship at NURC. I would like to thank them for their help and suggestions and for sharing with me their huge knowledge on real-world problems; it was my first (non-academic) job and i will remember it as a beautiful and intense experience. Furthermore, I would like to wish all the best to Dr. Salvatore Maresca, Dr. Bjoern Lund, Mr. Giuseppe Bombara and Mr. Ruben Carrasco.

I am also grateful to Dr. Aniello Buonanno and the whole Innovation Team (Selex Systems Integration, Giugliano, IT) for many helpful discussions and for sharing with me the “industry” point of view on research projects, which helped me understading the importance of practical aspects.

Special thanks go to Prof. Michael I. Jordan (U.C. Berkeley) and Prof. R. Schober (University of British Columbia, Canada) for their useful and helpful comments on preliminary versions of my research projects, which allowed me to improve my research significantly.

Finally, I want to deeply thank my loving and patient girlfriend Rosaria, my parents Lucia and Raffaele, my whole family and my friends for supporting me and making me happy even in the difficult moments I lived during these three years. Nothing could have happened without their support.

Contents

1	Introduction	5
1.1	Motivation	5
1.2	Publications	7
2	Data Fusion under uncertain models	11
2.1	Introduction	11
2.2	Inference for classification	14
2.2.1	Example 1	15
2.2.2	Example 2	15
2.3	Objective priors	17
2.3.1	Entropic priors	17
2.3.2	Maxinf priors	19
2.3.3	Back to Example 1	20
2.3.4	Back to Example 2	21
2.3.5	Entropic vs maxinf priors	22
2.4	Combinatorial justification for entropic priors (Wallis' idea)	22
2.5	Entropic priors for canonical likelihoods	26
2.6	Objective priors for sequences	29
2.6.1	Consistency for maxinf priors	30
2.6.2	Consistency for entropic priors	30
2.6.3	Consistency constraints on entropic priors	31
2.6.4	A simulation with a stationary i.i.d. Gaussian process	32
2.7	Conclusions	35
2.8	Appendix	35
2.8.1	Proof of entropic priors (2.3.2)	35
2.8.2	Proof of Proposition 1	36
2.8.3	Proof for consistent solution to entropic priors	36
3	Decision Fusion over MIMO with inst. CSI	39
3.1	Introduction	39
3.1.1	Motivation	39
3.1.2	Related Work	40
3.1.3	Main Results and Chapter Organization	41
3.2	System Model	42

3.3	Decode-and-Fuse	43
3.3.1	Optimum Rule	44
3.3.2	Maximum Ratio Combining (MRC)	44
3.3.3	Equal Gain Combining (EGC)	45
3.3.4	Max-Log Rule	45
3.4	Decode-then-Fuse	46
3.4.1	CV-ML	46
3.4.2	CV-MMSE	46
3.5	Implementation Issues	47
3.6	Simulation Results	48
3.7	Conclusions	52
3.8	Appendix	54
3.8.1	Proof of Proposition 1	54
3.8.2	Proof of Proposition 2	54
3.8.3	Proof of Proposition 3	55
3.8.4	CV-MMSE moments	55
3.8.5	Max-Log GSD derivation	56
4	MRC analysis over MIMO with inst. CSI	58
4.1	Introduction	58
4.2	System Model	59
4.2.1	WSN setting	59
4.2.2	Fusion Rules	60
4.3	MRC performance analysis	61
4.3.1	False Alarm and Detection probabilities computation	61
4.3.2	Deflection Coefficients	62
4.4	Simulations Results	64
4.5	Conclusions	65
4.6	Appendix	65
4.6.1	Proof of Proposition 7	65
5	Non-coherent Decision Fusion over MIMO	67
5.1	Introduction	67
5.2	System Model	69
5.2.1	WSN modeling	69
5.2.2	LLR	70
5.3	Optimality of Received Energy	70
5.4	Conclusions	74
5.5	Appendix	76
5.5.1	Proof of Proposition 1	76
5.5.2	Proof of Theorem 11	77
5.5.3	Proof of Theorem 12	78

6	Decision Fusion for detecting multiple targets	80
6.1	Introduction	80
6.2	System Model	82
6.2.1	The Overall System	82
6.2.2	WSN Local Decision Rule Model	84
6.3	Decision Fusion Algorithms	87
6.3.1	Joint Target - Optimal Decision Fusion (ODF) Approach	87
6.3.2	Separated Target - Decision Fusion Approaches	88
6.4	Simulations	90
6.5	Conclusions	95
6.6	Appendix	96
6.6.1	Maximum Detection Range	96
6.6.2	Time Complexity for (P)RLM	97
7	Data Fusion for inverse localization tasks	98
7.1	Introduction	98
7.1.1	Problem Motivation	98
7.1.2	Related Works	100
7.1.3	Main Results and Chapter Organization	101
7.2	System Model	102
7.2.1	Motion Models description	102
7.2.2	ML-PDA statistical assumptions and formulation	103
7.3	Objective Function Determination	105
7.4	“Geometry-driven” FIM-aided initial guess choice	108
7.4.1	Choice of $\hat{\mathbf{p}}_P(t_1)$ and $\hat{\mathbf{p}}_P(t_n)$	109
7.4.2	Choice of turning position vector $\hat{\mathbf{p}}_P(t_k)$	111
7.4.3	Remarks on α_θ	111
7.5	Simulation Results	112
7.6	Conclusions	120
7.7	Appendix	120
7.7.1	Proof of Proposition 13	120
7.7.2	Proof of Lemma 14	121
7.7.3	Proof of Proposition 15	122
7.7.4	Proof of Proposition 16	122
7.7.5	Choice of $\{\hat{r}_1, \hat{r}_n\}$	123
7.7.6	Choice of $\{\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_n\}$	124
7.7.7	Choice of $\hat{\mathbf{p}}_P(t_k)$	127
	Bibliography	129

Chapter 1

Introduction

1.1 Motivation

A formal definition of Data (or Information) Fusion, according to International Society of Information Fusion (ISIF), is

Definition. "Information fusion is the synergistic integration of information from different sources about the behavior of a particular system, to support decisions and actions relating to the system. Information fusion includes theory, techniques and tools for exploiting the synergy in the information acquired from multiple sources, for example sensors observing system behavior, databases storing knowledge about previous behavior, simulations predicting future behavior and information gathered by humans."

The problem of Data Fusion, from a practical point of view, can be explained as follows: "*We are drowning in information but starved for knowledge. This level of information is clearly impossible to be handled by present means. Uncontrolled and unorganized information is no longer a resource in an information society, instead it becomes the enemy.*" Data Fusion has recently gained attention as a widespread tool for a systematic analysis/design of several practical systems, such as:

- Wireless Sensor Networks systems for surveillance tasks;
- Cognitive Radio Networks for spectrum holes exploitation;
- Multi-Sensor Multi-target tracking systems, through the use of several and multi-modalities sensors;
- Enhanced Bio-medical systems, for effective healthcare assurance.

The classical models and algorithms derived in the single-source of information case need a re-thinking in the case of multi-dimensional (and possibly) heterogeneous observations coming from different sensors, typically distributed in a

wide geographical area. In fact, this indefinite growth of available information poses the serious problem of the "curse of dimensionality", thus affecting the scaling property of such complex systems. Therefore Data Fusion represents a cross-layer (w.r.t. the specific application and/or discipline) framework for enhancing, through an equal-balance between theoretical results and ready-to-implement practical (fusion) algorithms, the performance of such systems, while keeping the complexity under control.

In this dissertation we focus on statistical signal processing for Data Fusion, with a particular focus on wireless sensor networks. Six topics are studied: (i) Data Fusion for classification under model uncertainty (ii) Decision Fusion over coherent MIMO channels (iii) Performance analysis of Maximum Ratio Combining in MIMO decision fusion (iv) Decision Fusion over non-coherent MIMO channels (v) Decision Fusion for distributed classification of multiple targets (vi) Data Fusion for inverse localization problems, with application to wideband passive sonar platform estimation.

The first topic of this thesis addresses the problem of lack of knowledge of the prior distribution in classification problems that operate on small data sets that may make the application of Bayes' rule questionable. Uniform or arbitrary priors may provide classification answers that, even in simple examples, may end up contradicting our common sense about the problem. Entropic priors (EP), via application of the maximum entropy (ME) principle, seem to provide good objective answers in practical cases leading to more conservative Bayesian inferences. EP are derived and applied to classification tasks when only the likelihood functions are available. When inference is based only on one sample, we review the use of the EP also in comparison to priors that are obtained from maximization of the mutual information between observations and classes. This last criterion coincides with the maximization of the KL divergence between posteriors and priors that for large sample sets leads to the well-known reference (or Bernardo's) priors. The comparison on single samples considers both approaches in prospective and clarifies differences and potentials. A combinatorial justification for EP, inspired by Wallis' combinatorial argument for entropy definition, is also included. The application of the EP to sequences (multiple samples) that may be affected by excessive domination of the class with the maximum entropy is also considered with a solution that guarantees posterior consistency. An explicit iterative algorithm is proposed for EP determination solely from knowledge of the likelihood functions. Simulations that compare EP with uniform priors on short sequences are also included.

In the second topic, we study channel-aware binary-decision fusion over a shared Rayleigh flat-fading channel with multiple antennas at the Decision Fusion Center (DFC). We present the optimal rule and derive sub-optimal fusion rules, as alternatives with improved numerical stability, reduced complexity and lower system knowledge required. The set of rules is derived following both "Decode-and-Fuse" and "Decode-then-Fuse" approaches. Simulation results for performances are presented both under Neyman-Pearson and Bayesian frameworks. The effect of multiple antennas at the DFC for the presented rules is analyzed, showing corresponding benefits and limitations. Also, the effect on

performances as a function of the number of sensors is studied under a total power constraint.

The third part covers a theoretical performance analysis of the maximum ratio combining (MRC) rule for channel-aware decision fusion over multiple-input multiple-output channels, in the general case of both dependent and independent local decisions. The performances are evaluated through the closed form of the conditional moment generating function (MGF) of the MRC statistic, along with Gauss-Chebyshev quadrature rules. Also, the conditional MGF allows to derive the explicit expression of the deflection coefficients. Finally, all the theoretical results are confirmed through Monte Carlo simulations.

The fourth part of this dissertation presents a generalized optimality analysis of received-energy test for non-coherent decision fusion over a Rayleigh fading multiple access channel (MAC). More specifically, we provide a twofold generalization w.r.t the existing literature, allowing sensors to be non identical on one hand and introducing diversity on the other hand. Along with the derivation, we provide also a general tool to verify optimality of the the received energy test in scenarios with correlated sensor decisions. Finally, we derive an analytical expression of the effect of the diversity on the large-system performances, under both individual and total power constraints.

The fifth topic deals with the derivation of two sub-optimal decision fusion algorithms in the context of distributed classification of multiple moving targets, as a low complexity alternative to the optimal decision fusion. At the fusion center, all the the binary decisions coming from a wireless sensor network (WSN) designed for single target classification are exploited for a multiple classification task. Based on the concept of maximum detection range of each sensor and approximating the joint posterior as a product of the posterior marginal, we derive the RLM (Range Limited Marginalization) and PRLM (Parallel Range Limited Marginalization) algorithms. Comparison between these suboptimal algorithms and the optimal decision fusion are performed for different scenarios, in terms of probabilities of detection and false alarm and metrics related to complexity theory.

Finally we study an advanced topic in target motion analysis with wideband passive sonar. Maximum-likelihood probabilistic data-association represents an asymptotically efficient estimator for deterministic target motion, dealing with low-observable targets. Here we study the inverse problem, namely, how to identify the observing platform (following a "two-leg" motion model), from its results of the target estimation process, i.e. the estimated target state and the Fisher information matrix. We tackle the problem and we present observability properties, with supporting simulation results.

1.2 Publications

Publications directly related to the dissertation research:

Journal Papers

1. F. Palmieri and D. Ciuonzo, "Objective Priors from Maximum Entropy in Data Classification," *Information Fusion*, Elsevier, in press.
2. D. Ciuonzo, G. Romano and P. Salvo Rossi, "Channel-Aware Decision Fusion in Distributed MIMO Wireless Sensor Networks: Decode-and-Fuse vs. Decode-then-Fuse", *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2976-2985, August 2012.
3. D. Ciuonzo, G. Romano and P. Salvo Rossi, "Optimality of Received Energy in Decision Fusion over Rayleigh Fading Diversity MAC with Non-Identical Sensors", *IEEE Transactions on Signal Processing*, in press 2013.
4. D. Ciuonzo, P. K. Willett and Y. Bar-Shalom, "Tracking the Tracker from its Passive Sonar MLPDA Estimates", *IEEE Transactions on Aerospace and Electronics Systems*, major revision.
5. D. Ciuonzo, G. Romano and P. Salvo Rossi, "Performance analysis and Design of Maximum Ratio Combining in MIMO Decision Fusion", *IEEE Transactions on Wireless Communications*, submitted.

Conference Papers

1. F. Palmieri and D. Ciuonzo, "Data Fusion with Entropic Priors," in Proc. of the 20th Workshop on Neural Networks (WIRN 2010), pp. 107-114. ISBN 978-1-60750-691-1. Vietri sul Mare, Salerno, Italy. 27-29 May 2010. IOS press.
2. D. Ciuonzo and F. Palmieri, "Entropic Priors for Hidden-Markov Model Classification," in Proc. of 2011 IEEE Workshop on Statistical Signal Processing (SSP 2011), pp. 613-616. ISBN: 978-1-4577-0569-4. Hotel Negresco, Nice, France. 28-30 June 2011.
3. F. Palmieri and D. Ciuonzo, "Entropic Priors for Short-Term Stochastic Process Classification," in Proc. of 14th International Conference on Information Fusion (Fusion 2011), pp. 1-8. ISBN: 978-1-4577-0267-9. Chicago, IL, USA. 5-8 July 2011.
4. D. Ciuonzo, A. Buonanno, M. D'Urso and F. Palmieri, "Distributed Classification of Multiple Moving Targets with Binary Wireless Sensor Network," in Proc. of 14th International Conference on Information Fusion (Fusion 2011), pp. 1-8. ISBN: 978-1-4577-0267-9. Chicago, IL, USA. 5-8 July 2011.
5. F. Palmieri and D. Ciuonzo, "Consistency of Sequence Classification with Entropic Priors", 31th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2011), accepted. Waterloo, CAN. 10-15 July 2011.

6. D. Ciuonzo, G. Romano and P. Salvo Rossi, "Decision Fusion in MIMO Wireless Sensor Networks with Channel State Information", 2012 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2012), accepted. Hoboken - New Jersey, US. Jun. 17-20, 2012.
7. A. Buonanno, M. D'Urso, G. Prisco, M. Felaco, E. F. Meliadó, M. Mattei, F. Palmieri and D. Ciuonzo, "Mobile Sensor Networks Based on Autonomous Platforms for Homeland Security", 2012 IEEE Tyrrhenian Workshop on Advances in Radar and Remote Sensing (TYR 2012), accepted. Naples, IT. Sep. 12-14, 2012.

Other publications relevant to the Ph.D. research, but not otherwise reported in this dissertation include:

Journal Papers

1. P. Salvo Rossi, D. Ciuonzo, G. Romano and F. Palmieri, "On the Performance of Energy- Division Multiple Access over Fading Channels," *Wireless Personal Communications*, Springer, in press.
2. P. Salvo Rossi, D. Ciuonzo and G. Romano, "EDMA-based Schemes for Cognitive Radio Systems with Channel State Information", *Springer Wireless Personal Communications*, submitted.
3. G. Romano, D. Ciuonzo, P. Salvo Rossi and F. Palmieri "Low dominance based Sphere decoder", *IEEE Transactions on Communications*, submitted.
4. P. Salvo Rossi, D. Ciuonzo and G. Romano, "Orthogonality and Cooperation in Collaborative Spectrum Sensing through MIMO Decision Fusion", *IEEE Transactions on Wireless Communications*, submitted.

Conference Papers

1. G. Romano, D. Ciuonzo, P. Salvo Rossi, and F. Palmieri, "Tree-search ML detection for underdetermined MIMO systems with M-PSK constellations," in *Proc. of the 7th International Symposium on Wireless Communication Systems (ISWCS 2010)*, pp. 102-106. ISBN: 978-1-4244-6316-9. University of York, York, UK. 19-22 Sept. 2010.
2. F. Palmieri, D. Ciuonzo, D. Mattera, G. Romano and P. Salvo Rossi, "From Examples to Bayesian Inference," 21th Workshop on Neural Networks (WIRN 2011), accepted. Vietri sul Mare, Salerno, Italy. 3-5 June 2011.
3. P. Salvo Rossi, G. Romano, D. Ciuonzo and F. Palmieri, "Gain Design and Power Allocation for Overloaded MIMO-OFDM Systems with Channel State Information and Iterative Multiuser Detection," 8th IEEE Inter-

- national Symposium on Wireless Communication Systems (ISWCS 2011), accepted. RWTH Aachen University, Aachen, GER. 6-9 Nov. 2011.
4. G. Papa, D. Ciuonzo, G. Romano and P. Salvo Rossi, "Soft-Input Soft-Output King Decoder for Coded MIMO Wireless Communications," 8th IEEE International Symposium on Wireless Communication Systems (ISWCS 2011), accepted. RWTH Aachen University, Aachen, GER. 6-9 Nov. 2011.
 5. G. Romano, A. Drago and D. Ciuonzo, "Sub-optimal Importance Sampling for Fast Simulation of Linear Block Codes over BSC channels," 8th IEEE International Symposium on Wireless Communication Systems (ISWCS 2011), accepted. RWTH Aachen University, Aachen, GER. 6-9 Nov. 2011.
 6. D. Ciuonzo and S. Horn, "A New Hash Tree Based Approach for Totally Distributed TO-MHT Implementation", in Proc. of 10th IEEE Aerospace Conference (AEROCNF 2012), pp. 1-9. ISBN: 978-1-4577-0556-4. Big Sky - Montana, US. Mar. 3-10, 2012.
 7. P. Salvo Rossi, D. Ciuonzo and G. Romano, "Exploiting User Cooperation in MIMO Decision Fusion for Collaborative Spectrum Sensing", IEEE International Conference on Communications (ICC 2013), submitted.

Chapter 2

Data Fusion under uncertain models

2.1 Introduction

In many signal processing applications a system of probabilistic inference has to be designed according to a best guess about the model. More specifically, in applying Bayes' theorem, imperfect knowledge about likelihoods and priors may strongly affect the results of our classifications. The difficulty is particularly critical if we have to base our inferences on small sets of observations.

Controversial results on the application of Bayes' theorem have recently stimulated a growing interest in alternative theories to probability theory. For example, *evidence theory* ([1]; [2]) has been developed in the attempt of providing answers when direct application of Bayes' theorem seems to lead to solutions that are inadequate, or contradictory, with respect to our common sense. The Dempster-Shafer approach to data fusion ([3, 4, 5]) is becoming very popular in some data fusion areas ([6, 2]) and to many it seems to be a promising, useful alternative to the standard Bayesian framework.

We question whether remaining in the realm of classical probability and information theory ([7]) we can resolve some of the apparent difficulties in dealing with model uncertainties. We find that in most cases if we fill our lack of knowledge following the Maximum Entropy (ME) principle, all the contradictions seem resolved indicating that the problems that may have arisen are just the consequences of arbitrary implicit assumptions.

Of course, entropy maximization has a long history ([8, 9, 10]) both in signal processing and in more philosophical contexts. The literature is full of examples in which the ME principle has been translated into successful applications even though some general questions remain on its intrinsic nature.

In this chapter we concentrate on standard Bayesian inference in which parts of the model are not known. We search for "objective" solutions, which "represent the best method for objectively synthesizing and communicating the un-

certainties that arise in a specific scenario, but is not necessarily coherent in a more general sense,” or at least “a convention we should adopt in scenarios in which a subjective analysis is not tenable” ([11]).

We limit our attention only to unknown priors in data classification problems. Uncertainties in the likelihoods are also of great interest, but will be analyzed elsewhere.

Classical use of uniform priors that may be traced back to the work of Laplace ([12]), may result inappropriate in data fusion applications ([2]). Various attempts have appeared in the statistics literature in trying to obtain *objective alternatives* to uniform priors (for an excellent review of the various alternatives see ([13])). For example *Jeffreys’ priors* ([14]), named after Harold Jeffreys, are non-informative probabilities on the parameter space and are related to the square root of the determinant of Fisher information matrix. Jeffreys’ priors are defined for continuous parameter space and have the advantage of being independent upon the set of parameter variables. *Reference Priors* ([15, 16, 17]) are the most known and often used as an objective choice. They are derived for large sample sets from mutual information maximization and their definition is valid both for continuous and discrete parameter spaces. In the latter case the classic definition of reference priors leads trivially to discrete uniform priors ([16]). Reference priors have been recently redefined ([18]), with the objective of obtaining meaningful prior structures even in the discrete case. Other prior choices have been proposed in literature, such as the *Zellner’s Priors* ([19]) (MDI - Maximal Data Information priors). Their definition comes from the optimization of an ad-hoc information theory measure which is neither entropy nor mutual information.

Conjugate priors ([20]) are also very popular. They are introduced to force the posterior distribution to be in the same family as the prior distribution. The advantage is posterior’s tractability. A typical and well-known example is the Gaussian distribution which is conjugated to itself, i.e. when likelihood is Gaussian and the prior is chosen to be Gaussian, then also the posterior is Gaussian. Even if this choice has lead to closed-form priors, given a certain likelihood, their use is to be considered a “subjective” choice, because the structure and the parameters of these priors do not come from any supporting rationale except tractability. These priors were originally defined only for continuous parameters estimation, even though in recent years they have been applied in machine learning, through the so called “Dirichlet priors”. However, in the latter case, the distribution that “generates” the prior distribution (i.e. the Dirichlet distribution) depends on a certain number of hyper parameters, which again represents a “subjective” choice. Other priors, specific for given applications, can be derived from normalizations related to domain knowledge ([21]).

In this work we limit our attention to priors derived from the maximization of generative model entropy, i.e. the entropy of joint classification and (attributes) measure spaces distribution. The ME approach seems to be an excellent candidate for solving some of the inconsistencies related to model uncertainties (see ([22]) for a recent interesting discussion). After all, uniform priors represent the solution of maximum entropy maximization when no side information is given.

Well-known is the exponential solution to entropy maximization ([8, 10, 7]), also known as Boltzmann distribution, when constraints are imposed in the form of moments, either on the generative (corresp. joint distribution) or the discriminative (corresp. posterior conditional distribution) models. Even though attempts to impose moment constraints in discrete inferences have been reported in the literature, the solutions are somewhat cumbersome also because the nature of the classification problem may include both continuous (the measures) and discrete variables (the labels). Entropic Priors are not new ([23, 24, 25, 26, 27]), they have been introduced within the context of theoretical physics (see for example ([28]) and references therein) and their usage is still in its infancy also because of their possibly inconsistent behavior with repeated experiments ([25]), where an externally imposed ad-hoc constraint may have to be imposed. We address this issue in the second part of the chapter providing a self-contained solution and an algorithm for iterative prior computation.

The first original contribution of this chapter is the bottom-up approach from simple one-sample classification examples to compare posterior distributions when uniform, entropic priors and *maxinf* priors are used. The latter ones are priors derived from maximization of mutual information. Their analysis is justified by the fact that mutual information maximization asymptotically represents the formal definition of the reference priors for the discrete case. The comparison shows promising characteristics of the maximum entropy solution, that it is easy to compute and that seems to fix the inadequacies of Bayes' rule based on uniform, or other prior choices. After reviewing the idea of entropic priors, our second original contribution is a combinatorial justification of entropic priors for classification that extends Wallis' entropy derivation ([10]) to a multi-class context. To our knowledge the maximum entropy method has not been exploited much within the data classification context and it seems to qualify as a promising paradigm for many data fusion applications. The third original contribution of this work is in addressing the problem arising from excessive spreading of the entropy for multiple observations ([25]). To this purpose in the second part of the chapter, we first derive a condition on the likelihoods for posterior consistency, and successively when such a condition is not satisfied, we propose a solution for forcing consistency on the posterior sequence with an iterative algorithm that updates the entropic priors dynamically. Other domain-dependent attempts have been reported in the literature ([25]) to fix the consistency problem for continuous parameters. Our method focuses on discrete classes and it is totally self-contained within the entropy maximization criterion and requires only knowledge of the likelihood functions and their mutual Kullback-Leibler (KL) divergences.

Entropic priors have already shown very promising results in data fusion application. More specifically, in ([29]) we tested entropic priors for a target classification problem taken from ([6]) and ([2]) in comparison to the application of Dempster-Shafer (DS) theory. These experiences motivated our investigations on the method because the maximum entropy priors showed striking results in providing good common sense answers where Bayes' rule with uniform priors was claimed to be inadequate. In ([30]) we have applied entropic priors to

AR process classification; in ([31]) we have applied the same maximum entropy method to Hidden Markov models and in ([32]) we have addressed the issue of entropic prior consistency with reference to a binary sequence.

In Section 2.2 we describe the classification problem and propose two examples. In Section 2.3 we present entropic priors and maxinf priors. In Section 2.4 we present a combinatorial justification of entropic priors. In Section 2.5 we provide a list of entropic priors for canonical likelihoods and point to a geometric interpretation. In Section 2.6 entropic priors are applied to sequences where, after presenting a condition for consistency of Bayes' formula, we generalize the entropy maximization to sequences with the inclusion of the consistency condition. A set of simulations confirms the theory. Proofs and derivations are confined to the Appendix.

2.2 Inference for classification

In classification, or pattern recognition problems, we have available a set of d random attributes $\mathbf{X} = (X_1, X_2, \dots, X_d)$, directly linked to a discrete random variable S which takes values in the discrete alphabet $\mathcal{S} = \{s_1, s_2, \dots, s_M\}$. Vector \mathbf{X} is modeled as a d -dimensional random variable that belongs to a sample space \mathcal{X} , that can be finite, infinite, continuous or discrete. The elements of \mathbf{X} may be results of measurements, or be instances coming from other subsystems. Inference for classification consists in obtaining the degree of association between \mathbf{X} and S , i.e. from a realization $\mathbf{x} = (x_1, x_2, \dots, x_d)$, we want to obtain a degree of association to each one of the elements of \mathcal{S} .

Detailed inference about S can be obtained if the *a posteriori* probabilities (*posteriors*) are computed for each $s \in \mathcal{S}$ via the application of Bayes' theorem

$$P_r\{S = s | \mathbf{X} = \mathbf{x}\} = \frac{f(\mathbf{x}|s)P_r\{S = s\}}{\sum_{y \in \mathcal{S}} f(\mathbf{x}|y)P_r\{S = y\}}, \quad (2.2.1)$$

where $f(\mathbf{x}|s)$ are the *likelihood functions* and $P_r\{S = s\}$ are the *a priori probabilities (priors)*.

Unfortunately, in many practical cases the system designer has to operate with uncertainties about likelihoods and priors. His lack of knowledge about these two parts of the model may make the application of Bayes' theorem questionable. In our discussion, as stated in the introduction, we assume that we have the likelihood functions available, but we know nothing about the priors. In some critical decision-making contexts, such as when the classification output is a part of a more complex inference system, incorrect prior choice may propagate into the system and lead to inappropriate inference results, especially when the available data is scarce. Incorrect use of the priors sometimes even contradicts our most elementary common sense. These inconsistencies have often generated profound skepticism on the effectiveness of using Bayes' theorem in practical inferences in the presence of model uncertainties. Of course, assuming uniform priors is very common and it appears to be reasonable when such information is not available. After all, the well-known *maximum likelihood* (ML) method

makes decisions on the basis of uniform priors. Unfortunately, many cases have been shown in the literature ([6, 2, 28]) in which the use of uniform priors may show contradictory results when compared to our most intuitive understanding of the problem. The assumption of uniform priors may be sometimes too strong, or completely arbitrary. Let us start our discussion by looking at the following two examples.

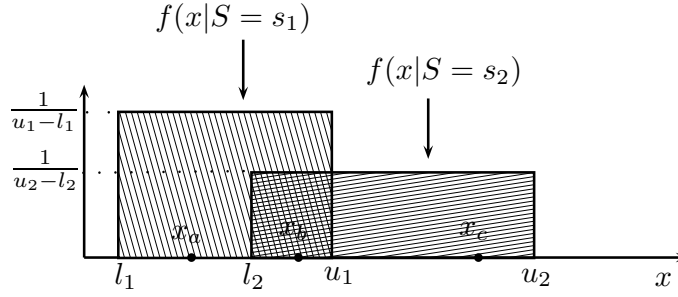
2.2.1 Example 1

Suppose that we are given a measurement of a continuous attribute X (length) ($d = 1$), that we know it can only come from one of two classes, s_1 or s_2 . Class s_i corresponds to lengths in the interval $[l_i, u_i]$, $i = 1, 2$. Assume that the two intervals have different sizes $u_1 - l_1 < u_2 - l_2$, and that they are partially overlapping $l_1 < l_2 < u_1 < u_2$. Now given a measurement $X = x$, we would like to infer about the class. Common sense dictates that: if $x \in [l_1, l_2]$ we declare class s_1 ; if $x \in [u_1, u_2]$ we declare class s_2 ; if $x \in [l_2, u_1]$ we cannot shift our belief towards one of the two classes and then we declare that both s_1 and s_2 should be equally likely. Instead by using Bayesian Standard Theory we set up two uniform likelihoods as $f(x|S = s_i) = \mathcal{U}(l_i, u_i)$, $i = 1, 2$, as shown in Fig. 2.2.1a and compute the posterior probabilities for each realization. If we observe $x_a \in [l_1, l_2]$, from Bayes' theorem the posterior probability is $P_r \{S = s_1|x_a\} = 1$, no matter what prior we choose, and similarly, when we observe $x_c \in [u_1, u_2]$, the posterior completely favors hypothesis s_2 , $P_r \{S = s_2|x_c\} = 1$, independently from the prior.

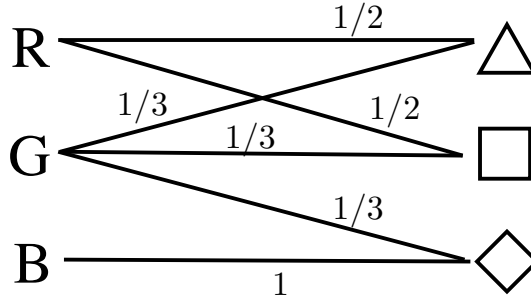
Quite surprisingly instead, with sample $x_b \in [l_2, u_1]$, if we use uniform priors (as in ML) $P_r \{S = s_1|x_b\} = \Delta_2/(\Delta_1 + \Delta_2)$, $P_r \{S = s_2|x_b\} = \Delta_1/(\Delta_1 + \Delta_2)$, $\Delta_1 \triangleq u_1 - l_1$, $\Delta_2 \triangleq u_2 - l_2$, and $P_r \{S = s_1|x_b\} > P_r \{S = s_2|x_b\}$: Bayes' theorem gives more confidence to hypothesis s_1 because its likelihood is the highest. If we look at this from outside the realm of probability theory, it appears somewhat contrary to our reasonable judgment: one would want to attribute the same confidence to s_1 and s_2 , as the value is compatible with both hypotheses and we know nothing about the phenomenon except the two intervals. Of course, from elementary probability theory, we know that the normalization property of distributions has raised the likelihood of s_1 , but nonetheless the result appears to our common sense as somewhat counter-intuitive.

2.2.2 Example 2

Suppose we have a set of coloured wooden objects $\mathcal{X} \triangleq \{\triangle, \square, \diamond\} = \{Triangle, Square, Diamond\}$, and we know that *Triangles* and *Squares* are painted only *Red* and *Green*, but not *Blue*. Also, *Diamonds* can only be *Green* or *Blue*. If we were blindfolded and handed one of the objects and asked to guess the colour, we would express equal confidence on *Red* and *Green* if we are given a *Triangle* or a *Square*, and equal confidence on *Green* and *Blue* if we are given a *Diamond*.



(a) The two likelihoods for Example 1



(b) The likelihoods of Example 2 in graphical form

Figure 2.2.1: Bayesian Inference Examples.

Instead let us see what happens if we set up the problem in the Bayesian framework. The set of colours is denoted $\mathcal{S} \triangleq \{R(ed), G(reen), B(lue)\}$ and we group in matrix form the obvious likelihoods

$$P_r(\mathbf{X}|S) = \begin{bmatrix} Pr\{X = \triangle | S = R\} & Pr\{X = \square | S = R\} & Pr\{X = \diamond | S = R\} \\ Pr\{X = \triangle | S = G\} & Pr\{X = \square | S = G\} & Pr\{X = \diamond | S = G\} \\ Pr\{X = \triangle | S = B\} & Pr\{X = \square | S = B\} & Pr\{X = \diamond | S = B\} \end{bmatrix} \quad (2.2.2)$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix}. \quad (2.2.3)$$

The likelihoods are better visualized in Fig. 2.2.1b. Since we do not have any knowledge on priors, we assume, as common in ML approach, equally likely priors on the colours $Pr\{S = R\} = Pr\{S = G\} = Pr\{S = B\} = \frac{1}{3}$, and

compute all the posteriors using Bayes' theorem

$$Pr_U(S|\mathbf{X}) = \begin{bmatrix} Pr\{S = R|X = \Delta\} & Pr\{S = R|X = \square\} & Pr\{S = R|X = \diamond\} \\ Pr\{S = G|X = \Delta\} & Pr\{S = G|X = \square\} & Pr\{S = G|X = \diamond\} \\ Pr\{S = B|X = \Delta\} & Pr\{S = B|X = \square\} & Pr\{S = B|X = \diamond\} \end{bmatrix} \quad (2.2.4)$$

$$= \begin{bmatrix} .60 & .60 & 0 \\ .40 & .40 & .25 \\ 0 & 0 & .75 \end{bmatrix}. \quad (2.2.5)$$

Bayes' theorem tells us to shift our belief towards *Red* if we observe a *Triangle* or a *Square*, and towards *Blue* if we observe a *Diamond*, rather than keep equal confidence on each pair. If we played this trivial guessing game with anyone, we would have a difficult time convincing him/her to shift his/her belief as in formula (2.2.4), perhaps attributing some philosophical rightness to the results of probability theory.

These counter-intuitive results come from the fact that we have inadvertently injected an arbitrary assumption into the model: the likelihoods are correct as they describe the way the data are composed, but *the uniform priors are totally arbitrary*. The main objective is then: search for a methodology to amend our lack of knowledge about the priors to get an answer from Bayes' rule that at least does not contradict our common sense, or alternatively that injects in the model specifications only the assumption made (through the likelihoods) by the system designer.

2.3 Objective priors

The problem we pose is one of a search for priors that are “objective”, in the sense that they are produced by a rationale synthesizing and communicating the uncertainties that arise in this scenario, i.e. a cautious inference is needed when few samples are observed. The search for objective priors has a long history and some of it has been reviewed in the introduction. In the following we discuss the method of maximum entropy for obtaining the entropic priors in inferences based on one realization and provide a comparison with the maxinf priors, that follow the same maximization criterion of the famous *reference* priors, but are considered here only for a discrete set and for a single sample. The comparison that follows should provide a better understanding in the classification context for each of these two approaches. The case of multiple observations is discussed in one of the following sections.

2.3.1 Entropic priors

Total unconstrained lack of knowledge about the distribution of a discrete parameter S defined on a finite set, is well described by the distribution that maximizes its marginal entropy $H(S)$. The solution is clearly the uniform distribution. Solutions to constrained entropy maximization problems are well

known in information theory when conditions on S are expressed in the form of moments, or functional expectations ([10]), ([7]). Unfortunately in our classification problem with no priors, it is not immediate to understand how to use moments or linear expectation to impose exact knowledge of the likelihood functions $f(\mathbf{x}|s)$, $s \in \mathcal{S}$. Therefore we adopt what appears to be the most sensible approach maximizing the joint entropy $H(\mathbf{X}, S)$ with respect to $P_r\{S = s\}$, keeping $f(\mathbf{x}|s)$ fixed. The joint distribution $f(\mathbf{x}, s)$ can be thought of as representing the “system model” and *Entropic priors* are the prior probabilities that, given the likelihood functions, maximize

$$H(\mathbf{X}, S) = \sum_{s \in \mathcal{S}} \int_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}|s) P_r\{S = s\} \log \frac{1}{f(\mathbf{x}|s) P_r\{S = s\}} d\mathbf{x} = E \left[\log \frac{1}{f(\mathbf{X}, S)} \right]. \quad (2.3.1)$$

The integral is replaced by a summation when the variable \mathbf{X} is discrete (log denotes the natural logarithm throughout the whole chapter).

Proposition. (*Entropic Priors*): *The prior distribution that maximizes the joint entropy $H(\mathbf{X}, S)$ for given likelihoods $f(\mathbf{X}|s)$, $s \in \mathcal{S}$, is*

$$P_{rE}\{S = s\} = \frac{e^{H(\mathbf{X}|s)}}{\sum_{y \in \mathcal{S}} e^{H(\mathbf{X}|y)}}, \quad (2.3.2)$$

where $H(\mathbf{X}|y)$ is the conditional entropy

$$H(\mathbf{X}|y) = \int_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}|y) \log \left[\frac{1}{f(\mathbf{x}|y)} \right] d\mathbf{x}. \quad (2.3.3)$$

The proof is straightforward and is reported in the appendix. Joint, conditional and marginal entropies with entropic priors are respectively $H_E(\mathbf{X}, S) = \log C$, $H_E(\mathbf{X}|S) = \frac{1}{C} \sum_s H(\mathbf{X}|s) e^{H(\mathbf{X}|s)}$ and $H_E(S) = \log C - \frac{1}{C} \sum_s H(\mathbf{X}|s) e^{H(\mathbf{X}|s)}$, where $C \triangleq \sum_s e^{H(\mathbf{X}|s)}$. Note that the entropies are expressed in *nats* for easier notation, but the base of the logarithm is totally irrelevant. Obviously if the likelihoods have the same conditional entropy, the solution is just the uniform distribution. Quite appealing is Bayes’ compact formula with entropic priors

$$P_{rE}\{S = s|\mathbf{x}\} = \frac{f(\mathbf{x}|s) e^{H(\mathbf{X}|s)}}{\sum_y f(\mathbf{x}|y) e^{H(\mathbf{X}|y)}}. \quad (2.3.4)$$

Entropic priors have been proposed in different forms ([25, 27, 24, 33, 26]) mostly within the context of theoretical physics. In ([25, 24, 33]) constraints on Fisher’s information matrix are also imposed for the continuous parameter case. In addition, to assure consistency of the classification, additional virtual constraints were added (the problems related to the consistency of this approach will be addressed in Section 2.6). Expressions similar to (2.3.2) have been derived in ([27]). The derivation reported here is slightly different from Neumann and Skilling’s because we maximize directly joint entropy, while they maximize prior entropy with a constraint on the average likelihood entropy. Their result

contains a scale parameter that needs to be resolved by other means. A similar derivation was reported also in ([23]) for obtaining maximum entropy marginals; the solution was applied to a system consisting of statistically dependent spins and to a thermodynamic model. Other ideas for entropic prior applications are in ([28]).

The rationale behind this approach is that entropy effectively represents the degree of uncertainty about the model ([8, 9, 10]) and it appears to be the natural choice when no further hypotheses can be injected into the system. Maximum entropy methods have been used extensively in various contexts showing great success in many application areas. In any case, to our knowledge, the real power of using entropic priors has not found full application yet in data classification problems.

2.3.2 Maxinf priors

Another appealing idea for obtaining objective priors may be to use the principle of Maximum Exchanged Information ([7]): the prior distribution is the result of maximization of the mutual information between S and \mathbf{X}

$$I(S, \mathbf{X}) = \sum_{s \in \mathcal{S}} \int_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x}|s) P_r\{S = s\} \log \frac{f(\mathbf{x}, s)}{f(\mathbf{x}) P_r\{S = s\}} d\mathbf{x}, \quad (2.3.5)$$

where the integral is replaced by a summation if variable \mathbf{X} is discrete. The mutual information can be rewritten as the expected KL divergence between the posterior $P(s|\mathbf{x}) \triangleq P_r\{S = s|\mathbf{x}\}$ and the prior $P(s) \triangleq P_r\{S = s\}$ (in the following when it is not necessary to specify the event we adopt the shorter notation without braces)

$$I(S, \mathbf{X}) = E \left[\log \frac{P(s|\mathbf{x})}{P(s)} \right]. \quad (2.3.6)$$

Mutual information is concave with respect to $P(s)$, therefore the *maxinf* solution $P_I(s)$ is unique and can be computed using the well-known Arimoto-Blahut algorithm ([7, p. 367]):

1. Start with an initial guess $P_0(s)$;
2. at a generic iteration step i apply Bayes' rule to get the current posterior:
$$Q_i(s|\mathbf{x}) = \frac{f(\mathbf{x}|s) P_{i-1}(s)}{\sum_c f(\mathbf{x}|c) P_{i-1}(c)};$$
3. replace the prior with the best distribution for the current posterior:
$$P_i(s) = \frac{e^{\int_{\mathbf{y}} f(\mathbf{y}|s) \log Q_i(s|\mathbf{y}) d\mathbf{y}}}{\sum_c e^{\int_{\mathbf{y}} f(\mathbf{y}|c) \log Q_i(c|\mathbf{y}) d\mathbf{y}}};$$
4. back to 2. and iterate.

The mutual information criterion is self-contained and elegant. The idea of maximizing average KL-distance between prior and posterior appears very appealing as a guideline for *non informative* priors. *Bernardo's priors*, or *reference priors* ([15, 16, 17]) are defined according to this criterion, but for continuous parameters and large sample sets. We discuss for now only the one-sample case delaying a further discussion on multiple samples to Section 2.6.

Mutual information maximization is well known in communication theory: if the variable S is the information source and \mathbf{X} models noisy channel outputs, the maximum of $I(S, \mathbf{X})$ with respect to the distribution of S is Shannon's *channel capacity* ([7]). The distribution in optimal conditions gives a criterion for the communication link designer to optimally "load" the channel. Therefore the *maxinf* criterion becomes a design tool for codes to provide a prior distribution that most "helps" the inference process, i.e. that maximize the mutual (exchange of) information between \mathbf{X} (the observed random vector) and S (the random variable to be inferred). Conversely, in data classification applications we are motivated to choose for priors that represent the worst-case for inference and that can provide the most cautious results. The following comparison should clarify differences and peculiarities for both approaches.

2.3.3 Back to Example 1

Entropic priors for the two uniform likelihoods of Ex. 1 are: $P_E(s_1) = \frac{\Delta_1}{\Delta_1 + \Delta_2}$, $P_E(s_2) = \frac{\Delta_2}{\Delta_1 + \Delta_2}$. Realizations x_a and x_c still give $P_E(s_1|x_a) = 1$ and $P_E(s_2|x_c) = 1$, no matter what priors we choose. Instead with sample x_b , Bayes' theorem with the entropic priors gives $P_E(s_j|x_b) = \frac{f(x_b|s_j)\Delta_j}{\sum_{i=1,2} f(x_b|s_i)\Delta_i} = \frac{1}{2}$; $j = 1, 2$. This appears as a more common sense answer with respect to the one obtained with uniform priors.

Maxinf priors, can be computed on this example, by applying Arimoto-Blahut algorithm of Section 2.3.2. At a generic step

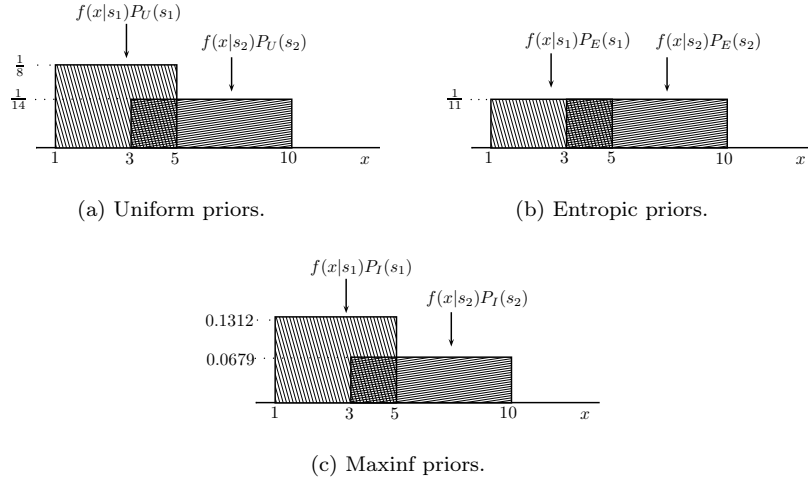
$$Q_i(s_1|x) = \begin{cases} 1, & l_1 < x < l_2, \\ \frac{P_{i-1}(s_1)\Delta_2}{P_{i-1}(s_1)\Delta_2 + P_{i-1}(s_2)\Delta_1}, & l_2 < x < u_1, \\ 0, & u_1 < x < u_2. \end{cases} \quad (2.3.7)$$

$$Q_i(s_2|x) = \begin{cases} 0, & l_1 < x < l_2, \\ \frac{P_{i-1}(s_2)\Delta_1}{P_{i-1}(s_1)\Delta_2 + P_{i-1}(s_2)\Delta_1}, & l_2 < x < u_1, \\ 1, & u_1 < x < u_2. \end{cases} \quad (2.3.8)$$

Since

$$\int_x f(x|s_1) \log Q_i(s_1|x) = -\frac{\Delta_{12}}{\Delta_1} \log \left(1 + \frac{P_{i-1}(s_2)\Delta_1}{P_{i-1}(s_1)\Delta_2} \right), \quad (2.3.9)$$

$$\int_x f(x|s_2) \log Q_i(s_2|x) = -\frac{\Delta_{12}}{\Delta_2} \log \left(1 + \frac{P_{i-1}(s_1)\Delta_2}{P_{i-1}(s_2)\Delta_1} \right), \quad (2.3.10)$$

Figure 2.3.1: The product $f(\mathbf{x}|s)P(s)$.

the complete recursion is

$$P_i(s_1) = \frac{e^{-\frac{\Delta_{12}}{\Delta_1} \log\left(1 + \frac{P_{i-1}(s_2)\Delta_1}{P_{i-1}(s_1)\Delta_2}\right)}}{e^{-\frac{\Delta_{12}}{\Delta_1} \log\left(1 + \frac{P_{i-1}(s_2)\Delta_1}{P_{i-1}(s_1)\Delta_2}\right)} + e^{-\frac{\Delta_{12}}{\Delta_2} \log\left(1 + \frac{P_{i-1}(s_1)\Delta_2}{P_{i-1}(s_2)\Delta_1}\right)}}; \quad P_i(s_2) = 1 - P_i(s_1), \quad (2.3.11)$$

where $\Delta_{12} \triangleq u_1 - l_2$. Fig. 2.3.1 shows the results for Example 1 for $(l_1, u_1) = (1, 5)$; $(l_2, u_2) = (3, 10)$. The product likelihood-prior $f(\mathbf{x}|s)P(s)$ is shown for uniform, entropic and maxinf priors respectively (recursion (2.3.11) converges in about ten steps). The three cases correspond to priors $(P_U(s_1), P_U(s_2)) = (\frac{1}{2}, \frac{1}{2})$, $(P_E(s_1), P_E(s_2)) = (\frac{4}{11}, \frac{7}{11})$ and $(P_I(s_1), P_I(s_2)) = (.5247, .4753)$, with posteriors at x_b respectively $(P_U(s_1|x_b), P_U(s_2|x_b)) = (\frac{7}{11}, \frac{4}{11})$, $(P_E(s_1|x_b), P_E(s_2|x_b)) = (\frac{1}{2}, \frac{1}{2})$ and $(P_I(s_1|x_b), P_I(s_2|x_b)) = (.6592, .3408)$. It is clear how the entropic priors equalize the inference by giving the least informative answer. Conversely, the maxinf priors tend to “load” more the first class because it is the least confused.

2.3.4 Back to Example 2

For the entropic priors in Ex. 2 we have $H(\mathbf{X}|s_1) = \log(2)$, $H(\mathbf{X}|s_2) = \log(3)$ and $H(\mathbf{X}|s_3) = 0$, i.e. $(P_E(s_1), P_E(s_2), P_E(s_3)) = (.3333, .5000, .1667)$. For the maxinf priors, Arimoto-Blahut converges in about 10-15 iterations and gives $(P_I(s_1), P_I(s_2), P_I(s_3)) = (.5, .0, .5)$ (also verifiable with direct substitution). The corresponding posteriors are

$$P_E(S|\mathbf{X}) = \begin{bmatrix} .5 & .5 & 0 \\ .5 & .5 & .5 \\ 0 & 0 & .5 \end{bmatrix}, \quad P_I(S|\mathbf{X}) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2.3.12)$$

It is clear how the response with the entropic prior reflects the desired common sense, while the maxinf solution assigns probability mass only to classes s_1 and s_3 because they are the most connected to the observations. Note that in the likelihoods s_2 is spread equally to all the three possible shapes and exchanges no information with the observations. Maxinf priors shift even more the inference belief in a polarized way in comparison to flat priors.

2.3.5 Entropic vs maxinf priors

When we look at the two approaches to prior's determination we see marked differences: in the maxinf approach more emphasis is put on the classes that are the easiest to invert. Vice versa in the entropic case we have more probability mass placed on the hypotheses that most mix the likelihoods. To visualize the differences we like to think in the maxinf case as if a "genie," hidden behind the "likelihood curtain," were working to help the inference process by placing more probability mass on the likelihoods that are the easiest to invert. Conversely, with the entropic priors we may imagine a "demon" that works against the inference process in misplacing the probability masses to make our Bayesian inversion as hard as possible.

The entropic priors seem to qualify as the most non-informative solution, as it has been argued from the examples. Vice-versa maxinf priors may be most useful when we can choose prior distribution to help the inference process. Entropic priors also seem to match better our quest for common sense responses. Much remains to be explored in the applications, also with the maxinf solution. In the following we focus mainly on entropic priors.

2.4 Combinatorial justification for entropic priors (Wallis' idea)

It is still intriguing to us why the method of maximum entropy provides problem solutions that make more sense when information about the model is scarce ([34]). An elegant answer to some of the general questions related to the ME methods can be built on a fundamental observation made by Wallis, and subsequently reported by Jaynes ([10]), that suggests how the entropy of a random variable could be derived as an asymptotic measure of complexity from elementary combinatorics. We still find remarkable that the entropy expression compactly measures the complexity of the phenomenon associated to a random variable. In the following, after reviewing Wallis' idea, we analyze the multi-class context as the limit of a combinatorial game.

Consider a bucket that contains three different kinds of balls $\{1, 2, 3\}$. Let us denote with N the total number of balls and with N_i , the number of balls of type i , $i = 1, 2, 3$, as illustrated in Fig. 2.4.1.

The number of possible ways we can extract with no replacement the N balls

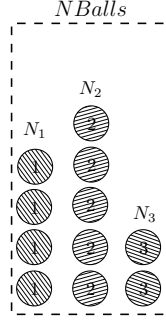


Figure 2.4.1: Wallis' Experiment: A bucket that contains different numbers N_i of three types $i = 1, 2, 3$ of balls.

from the bucket is given by the multinomial coefficient

$$W = \frac{N!}{N_1!N_2!N_3!}, \quad (2.4.1)$$

which is typically called the “multiplicity of the outcome”. Recall that a factorial $N!$ can be approximated for large N using Stirling's approximation as $\log N! = N \log N - N + \mathcal{O}(\log(N))$, where $\mathcal{O}(\log(N))$ denotes a term that grows as $\log(N)$ for $N \rightarrow +\infty$. Rather than W , we consider the monotonic function $\log(W)/N$ (recall the N is constant with respect to every possible assignment) and we take the limit for $N \rightarrow +\infty$, which leads to ([10])

$$\begin{aligned} \lim_{N \rightarrow +\infty} \frac{\log(W)}{N} &= \lim_{N \rightarrow +\infty} \frac{\log(N!) - \sum_{i=1}^3 \log(Np_i)!}{N} & (2.4.2) \\ &= \lim_{N \rightarrow +\infty} \frac{N \log(N) - N + \mathcal{O}(\log(N)) - \sum_{i=1}^3 [(Np_i) \log(Np_i) - (Np_i) + \mathcal{O}(\log(Np_i))]}{N} & (2.4.3) \end{aligned}$$

$$= \lim_{N \rightarrow +\infty} \left[\log(N) - \sum_{i=1}^3 p_i \log(Np_i) \right] + \lim_{N \rightarrow +\infty} \frac{\mathcal{O}(\log(N)) - \sum_{i=1}^3 \mathcal{O}(\log(Np_i))}{N} \quad (2.4.4)$$

$$= H(\pi) + 0 \quad (2.4.5)$$

where we have defined $(p_1, p_2, p_3) \triangleq (\frac{N_1}{N}, \frac{N_2}{N}, \frac{N_3}{N})$ as the limiting relative frequencies of the event associated to the extraction of a particular type of ball. We have exploited Stirling's approximation in the limit calculation and the normalization property $\sum_{i=1}^3 p_i = 1$. $H(\pi) \triangleq -\sum_{i=1}^3 p_i \log p_i$ is the entropy of assignment distribution and it represents compactly the degree of diversity in the experiment: the distribution π_{ME} that maximizes the entropy would favor a subdivision of the balls that maximizes the diversity in the ways the balls can be extracted from it. Obviously the uniform distribution achieves such a maximum, if there is no side information specified as a constraint.

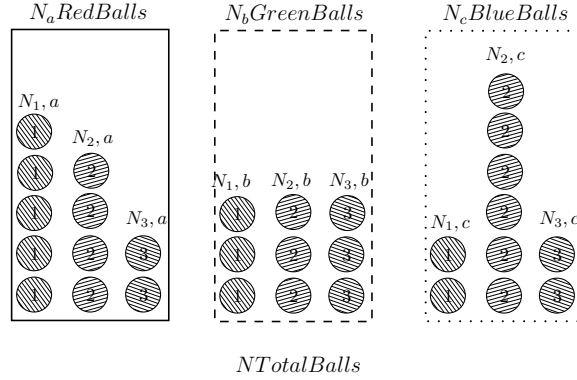


Figure 2.4.2: The Wallis' Experiment for classification tasks: a set of three different buckets $\mathcal{C} = (a, b, c)$ containing three different types $i = 1, 2, 3$ of balls $N_{i,col}$ in different numbers; in each bucket all the balls have the same colour.

The Wallis' combinatorial argument can be straightly extended to the classification context. Let us consider a set of three different buckets, namely $\mathcal{C} = (a, b, c)$, containing N_a Red, N_b Green and N_c Blue balls respectively. Each bucket has a different distribution of balls of types $(1, 2, 3)$, as shown in Fig. 2.4.2. More specifically, in the first bucket we have $N_{i,a}$ Red balls, in the second bucket $N_{i,b}$ Green balls and in the third bucket $N_{i,c}$ Blue balls of type i , for $i = 1, 2, 3$, respectively. The different colours can be thought of as classes and the types as attribute values. Each class has a different conditional distribution and the total number of balls is $N = \sum_{col \in \mathcal{C}} \sum_{i=1}^3 N_{i,col}$.

In this case the experiment consists in the extraction of the N balls from a bucket where all the balls of the three buckets have been mixed. The number of possible ways we could extract the N balls is given by the formula

$$W = W_a W_b W_c W_{colour} = \frac{N_a!}{\prod_{i=1}^3 N_{i,a}!} \frac{N_b!}{\prod_{i=1}^3 N_{i,b}!} \frac{N_c!}{\prod_{i=1}^3 N_{i,c}!} \frac{N!}{N_a! N_b! N_c!}, \quad (2.4.6)$$

where W_a, W_b and W_c are the number of permutations of Red, Green and Blue balls respectively, within the three types, and $W_{colour} = \frac{N!}{N_a! N_b! N_c!}$ is the number of permutations of colours. Considering again the ratio $\log(W)/N$, taking the limit $(N_a, N_b, N_c) \rightarrow +\infty$, which clearly implies $N \rightarrow +\infty$, and applying again

Stirling's approximation we have

$$\lim_{(N_a, N_b, N_c) \rightarrow +\infty} \frac{\ln(W)}{N} = \lim_{(N_a, N_b, N_c) \rightarrow +\infty} \frac{N_a \log(W_a)}{N} + \frac{N_b \log(W_b)}{N} + \frac{N_c \log(W_c)}{N} + \frac{\log(W_{colour})}{N} \quad (2.4.7)$$

$$= p_a H(T|a) + p_b H(T|b) + p_c H(T|c) + H(Colour) \quad (2.4.8)$$

$$= H(T|Colour) + H(Colour) \quad (2.4.9)$$

$$= H(T, Colour) \quad (2.4.10)$$

which represents the joint entropy of the model. Therefore, maximization of $H(T, Colour)$ with respect to $\pi_{colour} = (p_a, p_b, p_c)$, while keeping fixed the conditional distribution of each bucket, is equivalent to take balls from each bucket in a proportion (with respect N) that maximizes the number of possible ways we can extract all the balls from a unique bucket containing their mixture. The combinatorial arguments strongly support the use of joint entropy maximization as a means to avoid including implicitly any arbitrary assumptions about the colour distribution. In our genie-demon analogy, the (maxinf) genie loads the buckets in a way that favors inference, while the (entropic) demon loads them in the most confusing way.

Example: as a further experiment of comparison for prior distributions, we solve the inference example for the balls of Fig. 2.4.2. The balls correspond to the conditional distributions $f(t|a) = (\frac{5}{11}, \frac{4}{11}, \frac{2}{11})$; $f(t|b) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$; $f(t|c) = (\frac{1}{5}, \frac{3}{5}, \frac{1}{5})$. By looking at the colour proportions in the figure, the priors are $(P(a), P(b), P(c)) = (\frac{11}{30}, \frac{3}{10}, \frac{1}{3}) = (.3667, .3000, 0.3333)$. Entropic priors would rather suggest the different proportions $(P_E(a), P_E(b), P_E(c)) = (.3307, .3520, 0.3173)$ favoring the Green balls that are the most entropic. Maxinf priors, computed with Arimoto-Blahut algorithm, are $(P_I(a), P_I(b), P_I(c)) = (.2933, .0516, 0.6551)$; they attribute more emphasis on Red and Blue balls working in the opposite direction with respect to entropy maximization. Now if we pick a ball from a unique bucket reading only its type, we have colour posteriors (to read column wise)

$$P_{Tr}(Colour|t) = \begin{bmatrix} .5000 & .3077 & .2857 \\ .3000 & .2308 & .4286 \\ .2000 & .4615 & .2857 \end{bmatrix}; P_U(Colour|t) = \begin{bmatrix} .4601 & .2804 & .2542 \\ .3374 & .2570 & .4661 \\ .2025 & .4626 & .2797 \end{bmatrix};$$

$$P_E(Colour|t) = \begin{bmatrix} .4540 & .2810 & .2496 \\ .3544 & .2742 & .4870 \\ .1917 & .4448 & .2634 \end{bmatrix}; P_I(Colour|t) = \begin{bmatrix} .4735 & .2063 & .2646 \\ .0611 & .0333 & .0853 \\ .4654 & .7604 & .6501 \end{bmatrix},$$

respectively for true (Tr), uniform (U), entropic (E) and maxinf (I) priors. The ME criterion corresponds to the mixture which is the least favorable for the observations, while mutual information maximization works to provide priors which are more favorable for the identification.

$f(x s) s \in \{s_1, s_2, \dots, s_M\}$	$H(X s)$	$P_E(s)$
(Gaussian) $\frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{(x-\mu_s)^2}{2\sigma_s^2}}$	$\log(\sqrt{2\pi}e\sigma_s)$	$\frac{\sigma_s}{\sum_{y \in \mathcal{S}} \sigma_y}$
(Exponential) $\frac{1}{\lambda_s} e^{-\frac{x}{\lambda_s}}, x \geq 0$	$1 + \log(\lambda_s)$	$\frac{\lambda_s}{\sum_{y \in \mathcal{S}} \lambda_y}$
(Laplace) $\frac{1}{2\lambda_s} e^{-\frac{ x-\mu_s }{\lambda_s}}$	$1 + \log(2\lambda_s)$	$\frac{\lambda_s}{\sum_{y \in \mathcal{S}} \lambda_y}$
(Uniform) $\frac{1}{b_s-a_s}, a_s \leq x \leq b_s$	$\log(b_s - a_s)$	$\frac{(b_s-a_s)}{\sum_{y \in \mathcal{S}} (b_y-a_y)}$
(Rayleigh) $\frac{x}{\sigma_s^2} e^{-\frac{x^2}{2\sigma_s^2}}, x \geq 0$	$1 + \log(\frac{\sigma_s}{\sqrt{2}}) + \gamma/2$	$\frac{\sigma_s}{\sum_{y \in \mathcal{S}} \sigma_y}$
(Cauchy) $\frac{\lambda_s}{\pi(\lambda_s^2+x^2)}$	$\log(4\pi\lambda_s)$	$\frac{\lambda_s}{\sum_{y \in \mathcal{S}} \lambda_y}$
(Log-Normal) $\frac{1}{\sigma_s x \sqrt{2\pi}} e^{-\frac{(\log(x)-m_s)^2}{2\sigma_s^2}}, x > 0$	$m_s + \log(\sqrt{2\pi}e\sigma_s)$	$\frac{\sigma_s e^{m_s}}{\sum_{y \in \mathcal{S}} \sigma_y e^{m_y}}$
(Logistic) $\frac{e^{-\frac{(x-m_s)}{k_s}}}{k_s(1+e^{-\frac{(x-m_s)}{k_s}})^2}$	$2 + \log k_s$	$\frac{k_s}{\sum_{y \in \mathcal{S}} k_y}$
(Pareto) $\frac{a_s k_s^{a_s}}{x^{a_s+1}} x > k_s$	$\log(\frac{k_s}{a_s}) + 1 + \frac{1}{a_s}$	$\frac{e^{-a_s k_s/a_s}}{\sum_{y \in \mathcal{S}} e^{-a_y k_y/a_y}}$
(Triangular) $\begin{cases} \frac{2(x-a_s)}{(b_s-a_s)(c_s-a_s)} & a_s \leq x \leq c_s \\ \frac{2(b_s-x)}{(b_s-a_s)(b_s-c_s)} & c_s \leq x \leq b_s \end{cases}$	$\frac{1}{2} + \log(\frac{b_s-a_s}{2})$	$\frac{(b_s-a_s)}{\sum_{y \in \mathcal{S}} (b_y-a_y)}$

Table 2.1: Entropic Priors for canonical univariate conditional distributions $f(x|s)$.

2.5 Entropic priors for canonical likelihoods

Entropic priors can be computed numerically when we have empirical conditional distributions, or closed-form expressions can be obtained from Eq. (2.3.2) when $H(\mathbf{X}|s)$, $s \in \mathcal{S} = \{s_1, s_2, \dots, s_M\}$, are available analytically ([35, 36, 7]). In the following we report tables of entropic priors for common univariate and multivariate conditional distributions ($f(\mathbf{x}|s_i)$ is assumed to be completely known for all $i = 1, \dots, M$, which means that we have complete knowledge of all the density parameters for each class).

By rewriting Bayes' formula as

$$P(s|\mathbf{x}) = \frac{m(\mathbf{x}; s)}{\sum_c m(\mathbf{x}; c)}, \quad (2.5.1)$$

where $m(\mathbf{x}; s) = f(\mathbf{x}|s)P(s)$, $s \in \mathcal{S}$, represent the *likelihood-prior products*, we note that the effect of the entropic priors is to “equalize” the numerator, being the priors proportional to the relative support volume of the likelihoods. Recall

$f(\mathbf{x} s)$	$H(\mathbf{X} s)$	$P_E(s)$
(Gauss.) $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_s; \boldsymbol{\Sigma}_s)$	$\frac{1}{2} \log \{(2\pi e)^N \boldsymbol{\Sigma}_s \}$	$\frac{\sqrt{ \boldsymbol{\Sigma}_s }}{\sum_{y \in \mathcal{S}} \sqrt{ \boldsymbol{\Sigma}_y }}$
(Exp. Fam.) $\frac{1}{C(\theta_s)} e^{\sum_{i=1}^N \theta_{i,s} T_i(\mathbf{x})}$	$\log C(\theta_s) - \theta_s \nabla_{\theta} [\log C(\theta)]_{\theta=\theta_s}$	$\frac{C(\theta_s) e^{-\theta_s \nabla_{\theta} [\log C(\theta)]_{\theta=\theta_s}}}{\sum_{y \in \mathcal{S}} C(\theta_y) e^{-\theta_y \nabla_{\theta} [\log C(\theta)]_{\theta=\theta_y}}}$

Table 2.2: Entropic Priors for canonical multivariate conditional distributions $f(\mathbf{x}|s)$.

that $m(\mathbf{x}; s)$ are not densities, but they can be used immediately for obtaining the posteriors after normalization. For example, for conditional multivariate Gaussian densities (cfr. Table 2.2) with Entropic priors we get simply

$$m_E(\mathbf{x}; s) = \frac{1}{(2\pi)^{d/2} \sum_{y \in \mathcal{S}} \sqrt{|\boldsymbol{\Sigma}_y|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_s^{-1}(\mathbf{x}-\boldsymbol{\mu}_s)} \propto e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_s^{-1}(\mathbf{x}-\boldsymbol{\mu}_s)}. \quad (2.5.2)$$

Fig. 2.5.1 shows an example with three univariate Gaussian likelihoods where $m(\mathbf{x}; s)$ are compared for uniform, entropic and maxinf priors. In the bottom graphs the posterior probabilities are also shown (in these examples we have used discretized Gaussians on 100 points in the interval $[-10, 10]$, because we needed to compute numerically with Arimoto-Blahut algorithm the maxinf priors; in the graphs the density values are referred to truncated discrete densities and differ from the continuous version for a global scale of approximately $20/100 = 0.2$). Note how the entropic priors equalize the likelihoods providing a much more conservative decision in the superimposed regions and how the maxinf priors do the opposite by emphasizing the classes that exhibit the best separation. Other densities, such as those shown in the above tables, also in mixed configurations can be easily considered. The behaviors for uniform, entropic and maxinf priors shown in Fig. 2.5.1 are to be considered quite typical.

The likelihood-prior products obtained with maximum entropy method show striking similarities with the membership functions used in the fuzzy logic framework ([37]). Further comparison of the entropic prior approach with fuzzy logic is quite appealing, but it is beyond the scope of this dissertation and it will be addressed elsewhere.

Entropic priors have also a simple information-geometric interpretation. Recall that the volume of the support set of a random vector \mathbf{X} is approximately $V(\mathbf{X}) \approx e^{H(\mathbf{X})}$ ([7]). In our case we have the entropic priors as just the relative volumes of the conditional support sets, i.e.

$$P_E(s) \approx \frac{V(\mathbf{X}|s)}{\sum_{y \in \mathcal{S}} V(\mathbf{X}|y)}. \quad (2.5.3)$$

Then likelihoods with larger volumes have stronger entropic priors which balance the penalizing “spreading effect” observed in the intersection between conditional support sets, as motivated by Examples 1-2.

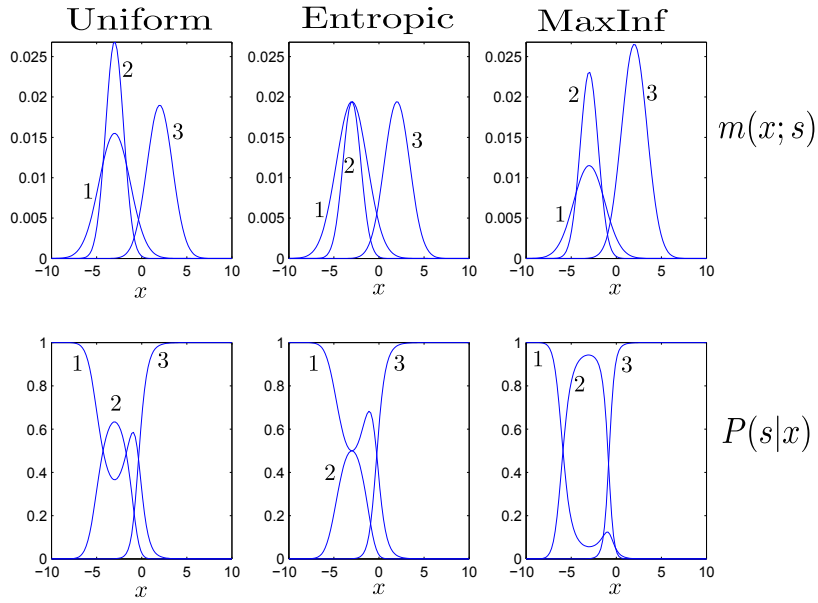


Figure 2.5.1: Comparison of $m(x; s)$ and $P(s|x)$ for uniform, entropic and maxinf priors; $f(x|s) = \mathcal{N}(x; \mu_s, \sigma_s)$, $s \in S = \{s_1, s_2, s_3\}$, means $(\mu_{s_1}, \mu_{s_2}, \mu_{s_3}) = (-3, -3, 2)$ and variances $(\sigma_{s_1}^2, \sigma_{s_2}^2, \sigma_{s_3}^2) = (3, 1, 2)$. Entropic priors are $(P_E(s_1), P_E(s_2), P_E(s_3)) = (.4177, .2412, .3411)$, and maxinf priors $(P_I(s_1), P_I(s_2), P_I(s_3)) = (.0476, .4611, .4913)$.

2.6 Objective priors for sequences

In the discussion presented above we have considered only one realization of a d -dimensional vector \mathbf{X} implicitly assuming that all our observations are condensed into one sample of \mathbf{X} . More in general when more samples of \mathbf{X} , possibly dependent, are available, we need to consider Bayes' inference for sequences in which we have available n d -dimensional random attributes $\mathbf{X}^n = (\mathbf{X}[1], \mathbf{X}[2], \dots, \mathbf{X}[n])$, all associated to one of the classes $\{s_1, s_2, \dots, s_M\}$. The set can represent n independent repeated trials of the same d -dimensional random variable, or more generally a dependent time series. Bayes' formula for the posterior probabilities at time n becomes

$$P_n(s_j|\mathbf{x}^n) = \frac{f(\mathbf{x}^n|s_j)P_n(s_j)}{f(\mathbf{x}^n)}, j = 1, \dots, M. \quad (2.6.1)$$

Objectivity for the priors in this more general scenario requires that at every time step a new set of priors is considered as a consequence of the new likelihoods $f(\mathbf{x}^n|s_j)$. Unfortunately, as n grows, a new problem must be considered, because if the data is generated from a specific class, the posteriors should asymptotically concentrate their probability mass onto that class (consistency). More specifically, prior determination has to account for *coherence* of the inference process, i.e. when we observe a sequence \mathbf{x}_j^n distributed according to the conditional distribution $f(\mathbf{x}^n|s_j)$ (which implies that it belongs to its typical set ([7, p. 57])), it must be recognized by Bayes' inference as belonging to the right class s_j . Any proposal for priors cannot avoid such a constraint. More formally, if the data sequence \mathbf{x}^n is generated from class s_j (denoted \mathbf{x}_j^n) we need to require that for $n \rightarrow \infty$

$$P_n(s_j|\mathbf{x}_j^n) \rightarrow 1, \quad P_n(s_i|\mathbf{x}_j^n) \rightarrow 0, \quad i \neq j. \quad (2.6.2)$$

Consistency constraints are made more explicit by the following proposition.

Proposition 1. *For a sequence \mathbf{X}^n extracted from a stationary ergodic process, sufficient condition for posterior consistency is that for large n the sequence of priors satisfy the condition $P_n(s_i) < P_n(s_j) e^{D_n(j;i)}$, $i \neq j$, $i, j = 1, \dots, M$, where $D_n(j;i) = \int_{\mathbf{x}^n} f(\mathbf{x}^n|s_j) \log \frac{f(\mathbf{x}^n|s_j)}{f(\mathbf{x}^n|s_i)} d\mathbf{x}^n$ is the KL-divergence between likelihood $f(\mathbf{x}^n|s_j)$ and $f(\mathbf{x}^n|s_i)$.*

The proposition states that if we control the priors' evolution, we avoid that possibly the wrong class takes over as n grows. This clearly depends on how much the class likelihoods are separated. The proof follows standard steps used in large deviation theory ([7]) and it is reported in the Appendix.

Note that when samples are independent $D_n(j;i) = nD(j;i)$ and we need to compute only the KL divergence between each pair of marginals $D(j;i) = \int_{\mathbf{x}} f(\mathbf{x}|s_j) \log \frac{f(\mathbf{x}|s_j)}{f(\mathbf{x}|s_i)} d\mathbf{x}$. Consistent special cases are: (a) uniform priors $P_n(s_j) = 1/M$, $j = 1, \dots, M$, because the condition $1 < e^{D_n(j;i)}$ will be satisfied for some $n > n_0$, because as n grows, $D_n(j;i)$ also grows; (b) constant priors $P_n(s_j) =$

$P(s_j)$, $j = 1, \dots, M$, because, similarly to (a), $P(s_i) < P(s_j)e^{D_n(j;i)}$ will be satisfied for some $n > n_1$ for all pairs because of the increasing behavior of $D_n(j; i)$ ([38]).

We adopt here the general condition $P_n(s_i) < P_n(s_j)e^{D_n(j;i)}$ for all n , even if the inequality has been derived as a sufficient condition for large n . The reason for this assumption is that even for small values of $n > 1$ exact time evolution of Bayes' formula can quickly become intractable even for small dimensionalities d . In practical application scenarios, in which we generally do not know how soon the law of the large numbers kicks in (it depends on the specific densities), if *for every* n we keep the condition satisfied we are guaranteed consistency within any chosen criterion for prior determination. We have verified in a number of practical examples that even for small n the condition on the time-varying priors (the entropic priors discussed in the following) nicely "drives" the posteriors towards consistent values keeping the entropy criterion maximized. More technical analyses of convergence both for small and large values of n would be desirable, but they are beyond the scope of this dissertation and will be explored elsewhere. We will confine our work to $n = 1$ and to the verification of convergence for large n .

2.6.1 Consistency for maxinf priors

Maxinf priors for large values of n , as mentioned above, become Reference, or Bernardo's priors. It is well known that reference priors in the discrete finite case tend to the uniform distribution as the sample size grows ([15]). They are obviously consistent. Hence in the comparisons that follow, we will concentrate only on entropic and uniform priors.

2.6.2 Consistency for entropic priors

Entropic priors for the more general case of sequences are the solution to the maximization of the model entropy $H(\mathbf{X}^n, S)$. The unconstrained solution is immediately (just as in (2.3.2))

$$P_{nE}(s_j) = \frac{e^{H(\mathbf{X}^n|s_j)}}{\sum_{l=1}^M e^{H(\mathbf{X}^n|s_l)}}, j = 1, \dots, M. \quad (2.6.3)$$

As it may appear a bit strange to have time-varying priors, true entropy maximization depends on $H(\mathbf{X}^n|s)$ and therefore it must depend also on n . Unfortunately as n grows, all the conditional entropies $H(\mathbf{X}^n|s)$ grow and tend to become dominated by the class (s_{max}) with the largest entropy, i.e. $\lim_{n \rightarrow \infty} P_{nE}(s) = \delta(s - s_{max})$. This problem has appeared as a fundamental limitation to the use of entropic priors for sequences, because the entropy spread may take control of the priors making Bayes' formula useless. Caticha and Preuss, that proposed entropic priors for continuous parameter spaces ([25]), pointed to such a problem and proposed a solution with a constrained average entropy for the priors to contain this saturation effect. Unfortunately such value

must be evaluated from specifics of the application domain. We would rather approach the problem in this discrete-class context recognizing that we are only interested in the posteriors and we do not necessarily care about the prior saturation effect, as far as Bayes' formula remains consistent. Then, substituting (2.6.3) in the condition of Proposition 1 we find that entropic priors (2.6.3) give consistent posteriors if

$$H(\mathbf{X}^n|s_i) < H(\mathbf{X}^n|s_j) + D_n(j; i), \quad i \neq j, \quad i, j = 1, \dots, M. \quad (2.6.4)$$

Inequality (2.6.4) is nice and compact and shows that even if some class entropies may grow more than others, we can still have correct inferences from entropic priors, if the class typical sets are sufficiently separated. There are many cases in which classes are well separated and unconstrained entropic priors for sequences can be used with no problems. We have applied entropic priors to AR Gaussian process classification ([30]) and found that unless the eigenvalues corresponding to the various classes are very similar, entropic priors can be used immediately for process classification. For i.i.d. sequences the condition for entropic priors consistency is simply $H(\mathbf{X}|s_i) < H(\mathbf{X}|s_j) + D(j; i)$, $i \neq j$, $i, j = 1, \dots, M$.

When the likelihoods do not satisfy condition (2.6.4) entropy maximization can be constrained as shown in the following section.

2.6.3 Consistency constraints on entropic priors

In many applications in which the class distributions may be poorly separated, perhaps because they have superimposed means, or similar covariances, etc., we should not give up our effort to obtain entropic objective priors just because condition (2.6.4) is not satisfied. We can upgrade our entropy maximization by appending the consistency constraints. Entropic priors $P_{nE}(s)$ definition at time n can be generalized to

$$\begin{cases} P_{nE}(s) = \arg \max_{P_n(s)} H(\mathbf{X}^n, S) \\ P_{nE}(s_i) - P_{nE}(s_j) e^{D_n(j; i) - \epsilon_n} \leq 0 \quad i \neq j \end{cases} \quad (2.6.5)$$

where ϵ_n is a small positive constant. Obviously, the problem is a bit more complicated because at each time step we must have available all the time-varying class entropies $H(X^n|s)$, and all the KL-divergences $D_n(i; j)$, and solve a constrained optimization problem. However, the conditions can be easily cast into a standard convex optimization problem with linear inequality constraints and solved with the help of Karhush-Khun-Tucker conditions ([39]). Entropic priors take the more general expression

$$P_{nE}(s_i) = \frac{e^{H(X^n|s_i) - \sum_{j=1, j \neq i}^M (\lambda_{ij}^n - \lambda_{ji}^n) e^{D_n(j; i) - \epsilon_n}}}{\sum_{l=1}^M e^{H(X^n|s_l) - \sum_{j=1, j \neq l}^M (\lambda_{lj}^n - \lambda_{jl}^n) e^{D_n(j; l) - \epsilon_n}}}, \quad i = 1, \dots, M, \quad (2.6.6)$$

where the constants $\lambda_{ij}^n \geq 0$, $i, j = 1, \dots, M$ are computed iteratively from the constraints at each time step n . The full derivation is reported in the Appendix

where an explicit algorithm for constant evaluation is also proposed. Essentially the entropy exponents are modified to ensure consistency: classes with large entropy must be contained in their prior value during sequence evolution while the prior for classes with small entropy is to be emphasized. The solution is quite general because when the class typical sets are sufficiently separated, many constants λ_{ij}^n (if not all) may vanish; when contribution (i, j) is present instead, either λ_{ij}^n , or λ_{ji}^n is null, as they assure mutual exclusive conditions (cfr. Eq.(2.6.5)), respectively adding or subtracting to the entropy exponent a positive value.

2.6.4 A simulation with a stationary i.i.d. Gaussian process

We have simulated class inference with entropic priors for various i.i.d. sequences where condition $H(\mathbf{X}|s_i) < H(\mathbf{X}|s_j) + D(j; i)$, $i \neq j$, is satisfied verifying that even if the class with the largest entropy takes over in the priors, posterior convergence is guaranteed. We show here instead a case in which the consistency condition is not satisfied and we have to resort to the constrained entropy maximization. More specifically we consider one-dimensional i.i.d. Gaussian likelihoods, i.e. $f(x^n|s) = \prod_{l=1}^n \mathcal{N}(x[l]; \mu_s, \sigma_s)$, with $M = 3$. The set of means and standard deviations are respectively $(\mu_{s_1}, \mu_{s_2}, \mu_{s_3}) = (0, 15, 0)$ and $(\sigma_{s_1}, \sigma_{s_2}, \sigma_{s_3}) = (40, 10, 5)$. Recall that the entropy of a Gaussian density $\mathcal{N}(x, \mu_i, \sigma_i)$ is $H(\mathbf{X}|s_i) = \log(\sqrt{2\pi e}\sigma_i)$ and KL-divergence between two Gaussian densities $\mathcal{N}(x, \mu_i, \sigma_i)$ and $\mathcal{N}(x, \mu_j, \sigma_j)$ is given by $D(j; i) = \frac{(\mu_j - \mu_i)^2 + \sigma_j^2 - \sigma_i^2}{2\sigma_i^2} + \log(\frac{\sigma_i}{\sigma_j})$. Eq. (2.6.4) in this case reduces to $(\mu_i - \mu_j)^2 > \sigma_i^2 - \sigma_j^2$, $i \neq j$. The parameters in the example do not satisfy the consistency condition. Therefore entropic priors are updated at each time step $n = 1, \dots, 20$, according to Eq. (2.6.6), through the algorithm presented in the Appendix. We set the parameter $\epsilon_n = (.05) \cdot n$ because it must be commensurate to the linear growth of entropy and KL divergence. At every step n the recursion (2.8.15) converges in less than five steps.

In Fig. 2.6.1 we report a comparison of posteriors evolution for a sequence extracted from the conditional density of class s_2 , i.e. $f(x^n|s_2)$ (see Fig. 2.6.1a) using uniform, unconstrained and constrained entropic priors. Class s_1 is the one with the largest entropy. Fig. 2.6.1b shows the entropic priors evolution with (Eq. (2.6.6)) and without (Eq. (2.6.3)) the consistency constraints. Note that the prior for the class with largest entropy (s_1) still dominates asymptotically even though its effect is contained in the constrained solution. Fig. 2.6.1c shows the posterior evolution with uniform priors. Fig. 2.6.1d is the posterior evolution with constrained and unconstrained entropic priors. Note that with unconstrained entropic priors class s_1 takes over because it has the largest entropy leading to incorrect inference. Conversely, after constraining the priors, the inference is asymptotically correct towards class s_2 . The evolution obtained from consistency-constrained entropic priors show a more cautious evolution towards proper inference as compared to uniform priors. Obviously many prior

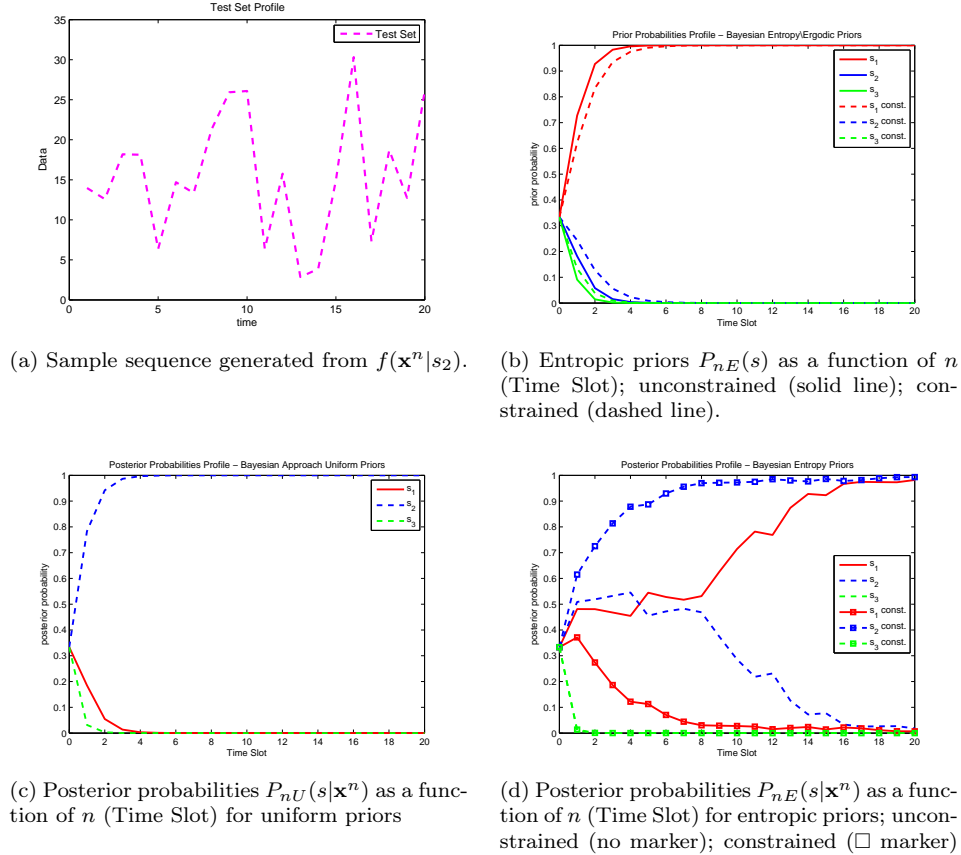
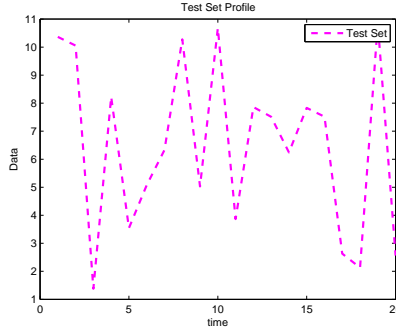


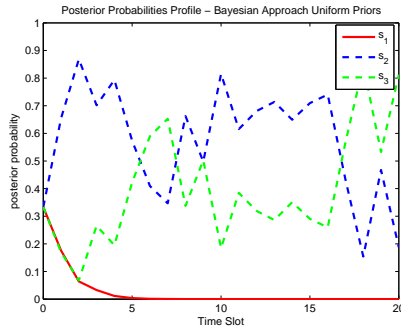
Figure 2.6.1: Comparison of uniform and entropic priors for a sequence generated according to $f(\mathbf{x}^n | s_2)$

choices may generate consistent posteriors as the priors become progressively irrelevant for large n . However, entropic priors that generate a posterior behavior that is to be considered more “objective” in comparison to uniform priors, can be very useful in critical inferences and decisions that have to be made on small values of n : once posterior consistency is ensured, the entropic solution may be the safest choice as compared to any other prior choice.

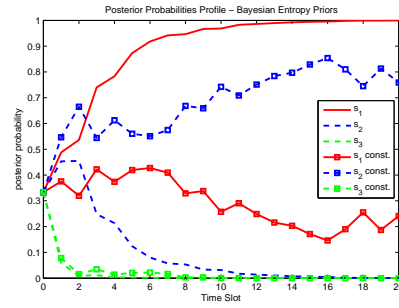
In a second set of simulations we have tested the application of entropic priors to a sample sequence that does not belong to any of the three classes included in the above model. This corresponds to a sequence that is non typical and may reflect an experiment in which we need to compute posteriors for an arbitrary sample. We wanted to test the claimed objectivity of the entropic priors in comparison to uniform priors. Recall that we started our discussion



(a) Sample sequence extracted from a uniform i.i.d. distribution in $[0.85, 10.85]$



(b) Posterior probabilities $P_{nU}(s|\mathbf{x}^n)$ as a function of n (Time Slot) for uniform priors



(c) Posterior probabilities $P_{nE}(s|\mathbf{x}^n)$ as a function of n (Time Slot) for entropic priors; unconstrained (no marker); constrained (\square marker)

Figure 2.6.2: Comparison of entropic and uniform priors over a non-typical set

at the beginning of this chapter with Examples 1-2 considering samples that belong to the intersection of more than one class. Fig. 2.6.2(a) shows a sequence of 20 samples from a uniform distribution in $[0.85, 10.85]$. The range is compatible with all the three classes and a uniform prior would favor classes s_2 and s_3 because $f(\mathbf{x}^n|s_1)$ is the lowest and the widest. Fig. 2.6.2(b) shows posterior evolution for uniform priors; Fig. 2.6.2(c) shows posterior evolution for unconstrained and constrained entropic priors. The unconstrained solution leads to a posterior that converges to class s_1 because it is the one with the largest entropy and it is certainly inappropriate. The consistency-constrained solution shows instead the maximum entropy cautious inference between class s_1 and s_2 reflecting the fact that the data points are most compatible with the ranges of $f(\mathbf{x}^n|s_1)$ and $f(\mathbf{x}^n|s_2)$, regardless of the priors. We would like to re-emphasize that the simulations are totally self-contained and no previous assumption except the likelihoods have been used.

2.7 Conclusions

Entropic priors seem to qualify as a cautious objective solution in Bayesian inference problems when only the likelihoods are known and we have available only small sample sets. The comparisons on inferences based on a single sample with maxinf priors shows in our examples that the ME method may be more appropriate when we have no previous model knowledge. Maxinf priors may be better used when the inference process can be controlled by emphasizing the classes that are the easiest to invert, as in communication channels. We believe that entropic priors have great potential in being used in critical pattern recognition applications when it is necessary to be very cautious about previous knowledge. Application to canonical densities shows intriguing connections to non-Bayesian methods, like Dempster-Shafer or Fuzzy Logic theory ([40]). We believe that this may be an interesting avenue of further investigation where the ME method to solve model indeterminacy is put in prospective with respect to theories alternative to probability. In this chapter we have also proposed a consistency-constrained entropic prior solution for sequences, where the spreading of the entropy for repeated experiments can be controlled to guarantee that in the transition from a few to many samples posterior inferences remain consistent. In such cases the priors are dynamically determined with an algorithm that needs only class entropies and mutual KL-divergences. We have verified the theory on a small set of typical examples in comparison to uniform priors. We believe that entropic priors show great potentials in the many signal processing applications areas. We have already applied the idea to kinematic classification for tracking ([29]), to Hidden Markov Model identification ([31]) and to AR process identification ([30]) and we are currently pursuing the application of the ME method to mixture models and to uncertain likelihoods.

2.8 Appendix

2.8.1 Proof of entropic priors (2.3.2)

For maximization of $H(\mathbf{X}, S)$, the Lagrange function is

$$\begin{aligned} \mathcal{L} &= H(\mathbf{X}, S) - \lambda(\sum_s P(s) - 1) = H(\mathbf{X}|S) + H(S) - \lambda(\sum_s P(s) - 1) \\ &= \sum_s H(\mathbf{X}|s)P(s) - \sum_s P(s) \log P(s) - \lambda(\sum_s P(s) - 1) \end{aligned} \quad (2.8.1)$$

where $P(s) = P_r\{S = s\}$ and $H(\mathbf{X}|s) = H(\mathbf{X}|S = s)$ and the only constraint is $\sum_s P(s) = 1$ (the inequality constraints $0 \leq P(s) \leq 1$ are verified a posteriori). Taking the derivative with respect to $P_r(s)$ we get $\frac{\partial \mathcal{L}}{\partial P(s)} = H(\mathbf{X}|s) - \log P(s) - 1 - \lambda$. Setting the derivative to zero we get $P(s) = e^{H(\mathbf{X}|s) - 1 - \lambda}$. By imposing the constraint we get $e^{-1-\lambda} = \frac{1}{\sum_y e^{H(\mathbf{X}|y)}}$, which gives the solution (2.3.2).

2.8.2 Proof of Proposition 1

Given an ergodic and stationary sequence \mathbf{x}_j^n coming from class s_j , as a consequence of the law of the large numbers, for large n

$$\log f(\mathbf{x}_j^n | s_j) \simeq -H(\mathbf{X}^n | s_j); \quad \log \frac{f(\mathbf{x}_j^n | s_j)}{f(\mathbf{x}_j^n | s_i)} \simeq D_n(j; i), \quad i \neq j. \quad (2.8.2)$$

When the samples are independent (i.i.d. sequence), $f(\mathbf{x}^n | s) = \prod_{l=1}^n f(\mathbf{x}[l] | s)$, $H(\mathbf{X}^n | s) = nH(\mathbf{X} | s)$, $D_n(j; i) = nD(j; i)$, Eq. (2.8.2) is the classic Asymptotic Equipartition Property (AEP) ([7, p. 245]). In general when the sequence has a dependence structure, there is a more technical derivation based a generalized version of the AEP with a sandwich argument to obtain (2.8.2) ([7, pp. 644-649]). A more detailed discussion of convergence is beyond the scope of this dissertation and will be reported elsewhere. The equivalent approximate expressions

$$f(\mathbf{x}_j^n | s_j) \simeq e^{-H(\mathbf{X}^n | s_j)}; \quad f(\mathbf{x}_j^n | s_i) \simeq e^{-H(\mathbf{X}^n | s_j) - D_n(j; i)}, \quad i \neq j \quad (2.8.3)$$

substituted in Bayes' formula, give the posteriors

$$P_n(s_j | \mathbf{x}_j^n) \simeq \frac{e^{-H(\mathbf{X}^n | s_j)} P_n(s_j)}{\sum_{l=1}^M e^{H(\mathbf{X}^n | s_l)} P_n(s_l)}, \quad (2.8.4)$$

$$P_n(s_i | \mathbf{x}_j^n) \simeq \frac{e^{-H(\mathbf{X}^n | s_j) - D_n(j; i)} P_n(s_i)}{\sum_{l=1}^M e^{H(\mathbf{X}^n | s_l)} P_n(s_l)}, \quad i \neq j, i = 1, \dots, M. \quad (2.8.5)$$

If the ratio $\frac{P_n(s_j | \mathbf{x}_j^n)}{P_n(s_i | \mathbf{x}_j^n)} \simeq \frac{P_n(s_j)}{P_n(s_i)} e^{D_n(j; i)}$ is kept > 1 , $P_n(s_j | \mathbf{x}_j^n) \rightarrow 1$ and $P(s_i | \mathbf{x}_j^n) \rightarrow 0$ $i \neq j$. The condition is then $P_n(s_i) < P_n(s_j) e^{D_n(j; i)}$.

2.8.3 Proof for consistent solution to entropic priors

We derive here the solution to (2.6.5) at time n by casting the problem into a standard constrained convex optimization problem. We simplify notation dropping index n and making $h_i \triangleq H(\mathbf{X}^n | s_i)$, $d_{ji} \triangleq D_n(j; i) - \epsilon_n$ and $p_i \triangleq P_n(s_i)$. The problem can then be written as

$$\begin{cases} \min_{p_i} (-\sum_{i=1}^M h_i p_i + \sum_{i=1}^M p_i \log p_i); \\ \sum_{i=1}^M p_i = 1; \\ -p_i \leq 0, \quad i = 1, \dots, M; \\ p_i - 1 \leq 0, \quad i = 1, \dots, M; \\ p_i - p_j e^{d_{ji}} \leq 0, \quad i \neq j, \quad i, j = 1, \dots, M. \end{cases} \quad (2.8.6)$$

The complete Lagrange function with all the constraints is

$$\begin{aligned} \mathcal{L} = & -\sum_{i=1}^M h_i p_i + \sum_{i=1}^M p_i \log p_i + \mu \left(\sum_{i=1}^M p_i - 1 \right) \\ & + \sum_{i=1}^M \sum_{j=1, j \neq i}^M \lambda_{ij} (p_i - p_j e^{d_{ji}}) - \sum_{i=1}^M \alpha_i p_i + \sum_{i=1}^M \beta_i (p_i - 1). \end{aligned} \quad (2.8.7)$$

There are in total: (1) one equality constraint for normalization (μ); (2) $\frac{M(M-1)}{2}$ inequality constraints ($\lambda_{ij}, i \neq j, i, j = 1, \dots, M$) for consistency; (3) M inequality constraints ($\alpha_i, i = 1, \dots, M$) for positivity; (4) M inequality constraints ($\beta_i, i = 1, \dots, M$) for bound to one. Taking the derivative with respect to p_i , we get

$$\frac{\partial \mathcal{L}}{\partial p_i} = -h_i + \log p_i + 1 + \mu + \sum_{j=1, j \neq i}^M (\lambda_{ij} - \lambda_{ji} e^{d_{ij}}) - \alpha_i + \beta_i. \quad (2.8.8)$$

Setting to zero for $i = 1, \dots, M$, the solution, the constraints and the coefficients, according to Karush–Kuhn–Tucker conditions ([39]) are written for $i, j = 1, \dots, M$, as

$$\begin{cases} p_i^* = e^{h_i - 1 - \mu - \sum_{j=1, j \neq i}^M (\lambda_{ij} - \lambda_{ji} e^{d_{ij}}) - \alpha_i + \beta_i}; & \sum_{i=1}^M p_i^* - 1 = 0; \\ \alpha_i, \beta_i \geq 0; & \lambda_{ij} \geq 0, i \neq j; \\ -p_i^* \leq 0; & p_i^* - 1 \leq 0; \\ p_i^* - p_j^* e^{d_{ji}} \leq 0, i \neq j; & \beta_i (p_i^* - 1) = 0; \\ \lambda_{ij} (p_i^* - p_j^* e^{d_{ij}}) = 0, i \neq j; \end{cases} \quad (2.8.9)$$

the last three equalities represents the complementary slackness conditions. Using the equality constraint we get

$$\sum_{i=1}^M p_i^* = e^{-1 - \mu} \left(\sum_{i=1}^M e^{h_i - \sum_{j=1, j \neq i}^M (\lambda_{ij} - \lambda_{ji} e^{d_{ij}}) - \alpha_i + \beta_i} \right) = 1. \quad (2.8.10)$$

Obtaining μ and substituting, the solution takes the form

$$p_i^* = \frac{e^{h_i - \sum_{j=1, j \neq i}^M (\lambda_{ij} - \lambda_{ji} e^{d_{ij}}) - \alpha_i + \beta_i}}{\sum_{l=1}^M e^{h_l - \sum_{j=1, j \neq l}^M (\lambda_{lj} - \lambda_{jl} e^{d_{lj}}) - \alpha_l + \beta_l}}, \quad i = 1, \dots, M. \quad (2.8.11)$$

From this expression all the p_i^* are automatically in $[0, 1]$. Therefore $\alpha_i = 0, \beta_i = 0, \forall i$. Solution and constraints simplify to

$$\begin{cases} p_i^* = \frac{e^{h_i - \sum_{j=1, j \neq i}^M (\lambda_{ij} - \lambda_{ji} e^{d_{ij}})}}{\sum_{l=1}^M e^{h_l - \sum_{j=1, j \neq l}^M (\lambda_{lj} - \lambda_{jl} e^{d_{lj}})}}; & \lambda_{ij} \geq 0, i \neq j; \\ p_i^* - p_j^* e^{d_{ji}} \leq 0, i \neq j; & \lambda_{ij} (p_i^* - p_j^* e^{d_{ij}}) = 0, i \neq j; \end{cases} \quad (2.8.12)$$

the inequality constraints can be re-written as

$$h_i - \sum_{n=1, n \neq i}^M (\lambda_{in} - \lambda_{ni} c_{in}) \leq h_j - \sum_{m=1, m \neq j}^M (\lambda_{jm} - \lambda_{mj} c_{jm}) + d_{ji}, \quad (2.8.13)$$

where $c_{ij} = e^{d_{ij}}$, to simplify notation. From the slackness conditions, if $\lambda_{ij} \neq 0$ (" λ_{ij} is active"), $p_i^* - p_j^* e^{d_{ji}} = 0$, otherwise $\lambda_{ij} = 0$. Extracting λ_{ij} from (2.8.13) we have

$$\begin{aligned} \lambda_{ij}(1 + c_{ji}) \geq & h_i - \sum_{n=1, n \neq i, j}^M \lambda_{in} + \sum_{n=1, n \neq i}^M \lambda_{ni} c_{in} - h_j \\ & + \sum_{m=1, m \neq j}^M \lambda_{jm} - \sum_{m=1, m \neq j, i}^M \lambda_{mj} c_{jm} - d_{ji}. \end{aligned} \quad (2.8.14)$$

Therefore an algorithm to get the coefficients can search iteratively for the coefficients applying the following formula for $i, j = 1, \dots, M; i \neq j$

$$\begin{aligned} \lambda_{ij} = \max\{0, & \frac{h_i - \sum_{n=1, n \neq i, j}^M \lambda_{in} + \sum_{n=1, n \neq i}^M \lambda_{ni} c_{in} - h_j}{1 + c_{ji}} \\ & + \frac{\sum_{m=1, m \neq j}^M \lambda_{jm} - \sum_{m=1, m \neq j, i}^M \lambda_{mj} c_{jm} - d_{ji}}{1 + c_{ji}}\}. \end{aligned} \quad (2.8.15)$$

We have tested the algorithm for small values of M . The unique solution is easily obtained in a few steps. Better algorithms may certainly be devised, but this issue is not central to the results of this dissertation.

Chapter 3

Decision Fusion over distributed MIMO channels with instantaneous channel state information

3.1 Introduction

3.1.1 Motivation

Decision Fusion (DF) in a wireless sensor network (WSN) consists in transmitting local decisions about an observed phenomenon from individual sensors to a DF Center (DFC) for a final decision. For sake of simplicity, the usual architecture assumes that each sensor communicates through a parallel access channel (PAC), which has to be implemented through time, code or frequency division schemes, since the wireless channel is “naturally” a broadcast medium [41, 42, 43, 44].

Recently it has been suggested to exploit the wireless medium as a multiple-access channel (MAC) for DF while coping with presence of intrinsic interference [45]. Furthermore, the problem of deep fading was addressed with multiple antennas at the DFC in order to ameliorate the fusion performances [46, 47, 45]. This choice demands only further complexity on DFC side and does not affect simplicity of sensors implementation. The result is a communication over a “virtual” Multiple-Input Multiple-Output (MIMO) channel between the sensors and the DFC, as shown in Fig. 3.1.1. Though appearing a stringent requirement, instantaneous channel state information (CSI) is often assumed [47, 46]. Design of channel-aware fusion rules and corresponding high performances motivates this assumption [42, 44, 48], fulfilled in many scenarios¹. For the same

¹We focus on the case in which the sensors and the DFC have minimal movement and

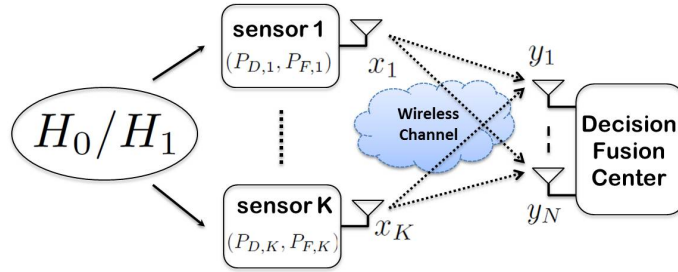


Figure 3.1.1: The Decision Fusion model in presence of a (virtual) MIMO channel.

reason, channel-aware fusion rules for coherent, non-coherent, and differential modulation were already proposed for PAC [42, 44, 48, 49].

Unfortunately, the optimal DF rule over MIMO channels with instantaneous CSI presents several difficulties in the implementation: *(i)* complete knowledge of channel parameters and sensors local performances; *(ii)* numerical instability of the expression, due to the presence of exponential functions with large dynamics; *(iii)* exponential growth of the complexity with the number of sensors. Design of sub-optimal DF rules with simple implementation and reduced system knowledge is then extremely desirable.

3.1.2 Related Work

A vast literature is present on DF, still growing in the context of WSNs. Several tutorial papers and books have been published on the topic, providing extensive references [50, 43, 51, 52]. Here we briefly discuss recent related work and focus on results needed for the work in this dissertation.

Sub-optimal rules for PAC scenario, presenting only issues (i) and (ii), were designed in [41, 42, 53, 48, 44, 49]. More specifically, the optimal rule was compared to Maximum Ratio Combining (MRC), Chair-Varshney Maximum-Likelihood (CV-ML), Equal Gain Combining (EGC) and Max-Log. MRC and CV-ML fusion rules approach optimum performance at very low and very high channel SNRs, respectively and they both suffer from a significant performance loss at moderate channel SNR [41, 42]. Instead EGC was shown to have robust performance for most SNR range [53]. Max-Log rule has been shown to outperform all the mentioned rules for all the SNR range [48]. Furthermore, robust sub-optimal fusion rules were obtained by fitting various non-linearities to the Log-Likelihood-Ratio (LLR) in [44]. The rules considered in [48] have also been derived and compared in the context of sensors differential encoding [48].

DF in distributed detection over MAC was first considered in [54], where an algorithm for the design of local-sensor quantizers was derived for the optimal

the environment changes slowly. More precisely, the coherence time of the wireless channel is assumed much longer than the time interval between two consecutive decisions made by the DFC, and instantaneous CSI can be obtained [47].

fusion rule. Also, several advanced transmission schemes have been proposed for MAC. In [55, 56] a fully-loaded code-division multiple access (CDMA) communication between sensors and the DFC is proposed: DF exploiting instantaneous CSI is performed and sub-optimal alternatives, based on the partitioning of transmitted symbols decoding and DF rule, are compared with the optimum. Also, in [57] a CDMA protocol is considered, but no decision is performed locally by the sensors, since they employ an *amplify-and-forward* scheme. In [58], the idea of the direct sequence spread spectrum is exploited to perform a communication based on the On-Off Keying (OOK) modulation with censoring; at the DFC the optimal rule and sub-optimal alternatives are derived on the basis only of statistical CSI.

An alternative and novel communication scheme, based on the method of types, called Type-Based Multiple Access (TBMA), has been proposed for distributed detection in [59, 60]. The scheme relies only on statistical CSI and presents limited bandwidth requirements. However, it has been shown in [61] that TBMA suffers from significant loss of performances in i.i.d. zero-mean fading channels.

DF over MIMO channels was firstly proposed in [47], focusing on J-Divergence optimal power allocation under non-identical local performances, which requires instantaneous CSI. DF rules over a matrix channel model, with only statistical CSI knowledge and non-coherent modulation have been studied in [45]. Distributed detection over MIMO with instantaneous CSI at the fusion center is tackled with the use of *amplify-and-forward* sensors in [46]; the optimum (data) fusion rule is derived and performance improvement is demonstrated when using multiple antennas at the fusion center.

3.1.3 Main Results and Chapter Organization

The main contributions of the chapter are summarized as follows.

- We study the design of channel-aware DF rules over MIMO channels, to best of our knowledge, for the first time. We discuss advantages and drawbacks for each rule in terms of *complexity*, *system knowledge required* and *performances*. The rules are grouped under *Decode-and-Fuse* (DaF) and *Decode-then-Fuse* (DtF) approaches. In the former case fusion is performed on the received signal, while in the latter case fusion is performed through CV rule, which processes ML or Minimum Mean Square Error (MMSE) estimate of the transmitted symbols. Multiple antennas at the DFC are shown to be beneficial independently of the considered (optimal or sub-optimal) fusion rule, however a (rule-dependent) saturation effect is present.
- Optimality properties are analytically demonstrated (in low and/or high SNR regime) for CV-ML, Max-Log and MRC over MIMO channel, in perfect analogy to the PAC case. Also, we formulate efficient Generalized Sphere Decoder (GSD) [62] implementation to tackle exponential complexity of the first two rules.

- We show that CV-MMSE outperforms CV-ML in low-to-moderate SNR regime, since it accounts for the correlation among decisions in the decoding stage. Furthermore, the co-channel interference of the MIMO channel makes MRC to outperform EGC under a Neyman-Pearson framework, in *disagreement* with PAC case [42].
- We show that each discussed rule, except for MRC, present an optimal number of sensors to be employed, under a total power constraint, to minimize the system probability of error. Further addition of sensors in the WSN *counter-intuitively* degrades the overall performances. Instead the MRC benefits from an indefinite increase of the number of sensors employed.

The chapter is organized as follows: Section 3.2 introduces the system model; Section 3.3 presents fusion rules based on DaF (including the optimum fusion rule) approach, while Section 3.4 presents fusion rules based on DtF approach; Section 3.6 presents an extensive set of simulations for performance comparison under different scenarios; some concluding remarks are given in Section 3.7; proofs and derivations are confined to Appendices.

Notation - Lower-case (resp. Upper-case) bold letters denote vectors (resp. matrices), with a_n (resp. $a_{n,m}$) representing the n th (resp. the (n,m) th) element of the vector \mathbf{a} (resp. matrix \mathbf{A}); upper-case calligraphic letters denote discrete and finite sets, with \mathcal{A}^K representing the k -ary Cartesian power of the set \mathcal{A} ; \mathbf{I}_N denotes the $N \times N$ identity matrix; $\mathbf{0}_N$ (resp. $\mathbf{1}_N$) denotes the null (resp. ones) vector of length N ; $\mathbb{E}\{\cdot\}$, $(\cdot)^*$, $(\cdot)^t$, $(\cdot)^\dagger$, $\Re(\cdot)$, $\angle(\cdot)$ and $\|\cdot\|$ denote expectation, conjugate, transpose, conjugate transpose, real part, phase and Frobenius norm operators; $P(\cdot)$ and $p(\cdot)$ are used to denote probabilities and probability density functions (pdf), in particular $P(A|B)$ and $p(a|b)$ represent the probability of event A conditioned on event B and the pdf of random variable a conditioned on random variable b , respectively; $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a circular symmetric complex normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; finally the symbols \sim and \propto mean “distributed as” and “proportional to” respectively.

3.2 System Model

We consider a distributed binary hypothesis test, where K sensors are used to discriminate between the hypotheses of the set $\mathcal{H} = \{H_0, H_1\}$. For example H_0 and H_1 may represent the absence and the presence of a specific target of interest, respectively. The *a priori* probability of hypothesis $H_i \in \mathcal{H}$ is denoted $P(H_i)$. The k th sensor, $k \in \mathcal{K} \triangleq \{1, 2, \dots, K\}$, takes a binary local decision $d_k \in \mathcal{H}$ about the observed phenomenon on the basis of its own measurements. The decision d_k is assumed independent of each other decisions d_ℓ , $\ell \in \mathcal{K}, \ell \neq k$, conditioned on $H_i \in \mathcal{H}$.

Each decision d_k is mapped to a symbol $x_k \in \mathcal{X} = \{-1, +1\}$ representing

BPSK modulation²: without loss of generality we assume that $d_k = H_0$ maps into $x_k = -1$ and $d_k = H_1$ into $x_k = +1$. The quality of the k th sensor decisions is characterized by the conditional probabilities $P(x_k|H_j)$. More specifically, we denote $P_{D,k} \triangleq P(x_k = 1|H_1)$ and $P_{F,k} \triangleq P(x_k = 1|H_0)$, respectively the probability of detection and false alarm of the k th sensor.

The sensors communicate with the DFC over a wireless flat-fading MAC, with i.i.d. Rayleigh fading coefficients of unitary mean power. The DFC is equipped with N receive antennas in order to exploit diversity and combat signal attenuation due to small-scale fading of the wireless medium; this configuration determines basically a distributed (or “virtual” [47]) MIMO channel, as shown in Fig. 3.1.1. Also, instantaneous CSI and perfect synchronization are assumed at the DFC as in [47]; note that multiple antennas at the DFC do not make these assumptions harder to verify w.r.t. (single antenna) MAC.

We denote: y_n the received signal at the n th receive antenna of the DFC after matched filtering and sampling; $h_{n,k} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ the fading coefficient between the k th sensor and the n th receive antenna of the DFC; w_n the additive white Gaussian noise at the n th receive antenna of the DFC. The vector model at the DFC is the following:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (3.2.1)$$

where $\mathbf{y} \in \mathbb{C}^N$, $\mathbf{H} \in \mathbb{C}^{N \times K}$, $\mathbf{x} \in \mathcal{X}^K$, $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma_w^2 \mathbf{I}_N)$ are the received signal vector, the channel matrix, the transmitted signal vector and the noise vector, respectively.

Remarks: The vector model in Eq. (3.2.1) can be underloaded ($K < N$), fully-loaded ($K = N$) or overloaded ($K > N$). While in MIMO communication systems all the three scenarios are of interest, in the specific case of WSN only the overloaded case is reasonable, as typically the number of sensors is typically much larger than the number of antennas that could be employed at the DFC (i.e. $K \gg N$). Throughout this chapter we will refer to the channel SNR as the ratio between the average total received energy from the WSN $\mathcal{E}_s = \mathbb{E} \left\{ \|\mathbf{H}\mathbf{x}\|^2 \right\}$ and the one-sided power spectral density of the continuous process noise σ_w^2 , that is $SNR \triangleq \mathcal{E}_s / \sigma_w^2 = KN / \sigma_w^2$. Note that the corresponding individual channel SNR for the k th sensor will be $SNR_k = N / \sigma_w^2$.

3.3 Decode-and-Fuse

In this case, see Fig. 3.2.1a, the DFC aims at detecting the presence of the target directly from the received signal vector without any intermediate step to decode the transmitted vector.

²Note that in case of an absence/presence task, where \mathcal{H}_0 is less probable, On-Off Keying (OOK) can be employed for energy efficiency purpose. In the following we will refer only to BPSK, however the results presented in this chapter apply straightforward to OOK.

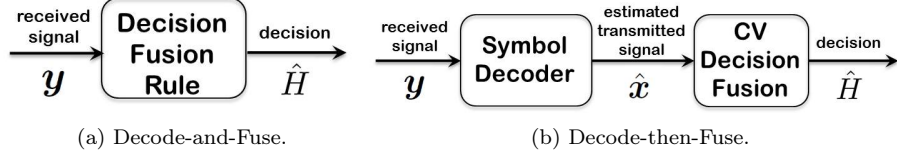


Figure 3.2.1: Fusion approaches.

3.3.1 Optimum Rule

The optimal test [63] for the considered problem can be formulated as

$$\Lambda_{opt} \triangleq \ln \left[\frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} \right] \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\geq}} \gamma \quad (3.3.1)$$

where \hat{H} , Λ_{opt} and γ denote the estimated hypothesis, the LLR (i.e. the optimal fusion rule, referred also as the “optimum” in the following) and the threshold to which the LLR is compared to. The threshold γ can be determined to assure a fixed system false-alarm rate, if a Neyman-Pearson detection is employed, or can be chosen to minimize the probability of error in Bayes detection. An explicit expression of the LLR from Eq. (3.3.1) is given by

$$\begin{aligned} \Lambda_{opt} &= \ln \left[\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} p(\mathbf{y}|\mathbf{x}) \prod_{k=1}^K P(x_k|H_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} p(\mathbf{y}|\mathbf{x}) \prod_{k=1}^K P(x_k|H_0)} \right] \\ &= \ln \left[\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_w^2}\right) \prod_{k=1}^K P(x_k|H_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_w^2}\right) \prod_{k=1}^K P(x_k|H_0)} \right] \end{aligned} \quad (3.3.2)$$

where we have exploited the conditional independence among x_k (given H_i), and of \mathbf{y} from H_i (given \mathbf{x}).

3.3.2 Maximum Ratio Combining (MRC)

The LLR of Eq. (3.3.2) can be simplified under the assumption of perfect sensors [48, 58], i.e. $(P_{D,k}, P_{F,k}) = (1, 0)$, $k \in \mathcal{K}$. In this case the transmitted vector $\mathbf{x} \in \{\mathbf{1}_K, -\mathbf{1}_K\}$ and the Eq. (3.3.2) reduces to:

$$\Lambda_{MRC} = \ln \left[\frac{\exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{1}_K\|^2}{\sigma_w^2}\right)}{\exp\left(-\frac{\|\mathbf{y}+\mathbf{H}\mathbf{1}_K\|^2}{\sigma_w^2}\right)} \right] \propto \Re(\mathbf{1}_K^t \mathbf{H}^\dagger \mathbf{y}) \quad (3.3.3)$$

where in the r.h.s. we have neglected the terms that can be incorporated in γ through Eq. (3.3.1). The following proposition states that as in PAC case [42] the MRC is the low-SNR approximation of the optimum of Eq. (3.3.2) when local performances of sensors are identical.

Proposition 2. For low SNR, if $(P_{D,k}, P_{F,k}) = (P_D, P_F)$, $k \in \mathcal{K}$: $\Lambda_{MRC} \approx \Lambda_{opt}$.

Proof. See Appendix 3.8.1. \square

3.3.3 Equal Gain Combining (EGC)

Motivated by the fact that Λ_{MRC} resembles a MRC statistics for diversity combining [64], cfr. Eq. (3.3.3), we propose a further rule in the simple form of an equal gain combiner:

$$\Lambda_{EGC} = \Re(\mathbf{z}^\dagger \mathbf{y}) \quad (3.3.4)$$

$$\mathbf{z} = e^{j \cdot \angle(\mathbf{H}\mathbf{1}_K)} \quad (3.3.5)$$

A similar expression was derived for MIMO beamforming and combining systems in [65].

3.3.4 Max-Log Rule

Let us first recall the Max-Log approximation known from turbo-codes literature [66]

$$\begin{aligned} \ln \left(\sum_{\ell=1}^L B_\ell e^{A_\ell} \right) &= \ln \left(\sum_{\ell=1}^L e^{A_\ell + \ln(B_\ell)} \right) \\ &\approx \max_{\ell \in \{1, 2, \dots, L\}} [A_\ell + \ln(B_\ell)] \end{aligned} \quad (3.3.6)$$

where $A_i \in \mathbb{R}$ and $B_i \in \mathbb{R}^+$. The approximation in Eq. (3.3.6) is accurate when one of the terms in the sum $\sum_{\ell=1}^L B_\ell e^{A_\ell}$ dominates over the remaining terms. LLR expression from Eq. (3.3.2) is in the same form of Eq. (3.3.6), thus using this approximation we obtain the following sub-optimal fusion rule:

$$\begin{aligned} \Lambda_{Max-Log} = \min_{\mathbf{x} \in \mathcal{X}^K} &\left[\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k | H_0) \right] \\ &- \min_{\mathbf{x} \in \mathcal{X}^K} \left[\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k | H_1) \right] \end{aligned} \quad (3.3.7)$$

which can be interpreted as the difference between *hypothesis prior-weighted* minimum distance searches.

Proposition 3. *Max-Log Properties:*

1. For low SNR, if $(P_{D,k} > \frac{1}{2}, P_{F,k} < \frac{1}{2})$: $\Lambda_{Max-Log} \approx \Lambda_{MRC}$.
 - (a) For low SNR, if $(P_{D,k}, P_{F,k}) = (P_D, P_F)$, $k \in \mathcal{K}$, and $(P_D > \frac{1}{2}, P_F < \frac{1}{2})$: $\Lambda_{Max-Log} \approx \Lambda_{opt}$.

(b) For high SNR : $\Lambda_{Max-Log} \approx \Lambda_{opt}$.

Proof. See Appendix 3.8.2. \square

The above proposition states that under particular circumstances Max-Log approximates the optimum both in the low and high SNR regime.

3.4 Decode-then-Fuse

In this case, see Fig. 3.2.1b, the DFC is based on the separation of decoding and fusing stages. Firstly, the decoding block computes an estimate of \mathbf{x} , denoted $\hat{\mathbf{x}}$ in the following, from \mathbf{y} . Finally, the global decision \hat{H} is taken on the basis of $\hat{\mathbf{x}}$ using the Chair-Varshney (CV) rule [67], i.e. the optimal fusion rule for noiseless channels, whose expression is given by

$$\Lambda_{CV} = \sum_{k=1}^K \hat{u}_k \ln \left(\frac{P_{D,k}}{P_{F,k}} \right) + (1 - \hat{u}_k) \ln \left(\frac{1 - P_{D,k}}{1 - P_{F,k}} \right), \quad (3.4.1)$$

where $\hat{u}_k \triangleq \frac{\hat{x}_k + 1}{2}$, $k \in \mathcal{K}$. Note that when local sensor performances are identical, i.e. $(P_{D,k}, P_{F,k}) = (P_D, P_F)$, Eq. (3.4.1) reduces to a simple counting rule [67].

3.4.1 CV-ML

In this case $\hat{\mathbf{x}}$ is obtained through ML decoder [64] as

$$\hat{\mathbf{x}}_{ML} = \arg \min_{\mathbf{x} \in \mathcal{X}^K} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \quad (3.4.2)$$

In analogy to PAC case [42], CV-ML is the high-SNR approximation of the optimum of Eq. (3.3.2), as stated through the following proposition.

Proposition 4. *For high SNR: $\Lambda_{CV-ML} \approx \Lambda_{opt}$.*

Proof. See Appendix 3.8.3. \square

3.4.2 CV-MMSE

In this case the ML decoder is replaced with a sub-optimal one presenting reduced complexity, obtained via the MMSE solution [64]. The rule however needs to take into account the correlation between symbols x_k , $k \in \mathcal{K}$, since they observe the same phenomenon. This issue was addressed in [56] but restricted to the case $P_{D,k} = 1 - P_{F,k}$. In the more general case, the following MMSE decoder should be considered [68]:

$$\hat{\mathbf{x}}_{MMSE} = \text{sign} \left[\bar{\mathbf{x}} + \mathbf{C}\mathbf{H}^\dagger (\mathbf{H}\mathbf{C}\mathbf{H}^\dagger + \sigma_w^2 \mathbf{I}_N)^{-1} (\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}) \right] \quad (3.4.3)$$

where $\bar{\mathbf{x}} \triangleq \mathbb{E}\{\mathbf{x}\}$ and $\mathbf{C} \triangleq \mathbb{E}\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^\dagger\}$ are the mean vector and the covariance matrix of \mathbf{x} , respectively. Their explicit expression is derived and reported in Appendix 3.8.4.

3.5 Implementation Issues

Practical implementation of LLR in Eq. (3.3.2) is difficult due to exponential functions with large dynamic range especially for moderate-to-high channel SNRs $K^N/\sigma_w^2 \gg 1$. This becomes a quite severe requirement for fixed point implementations [69, 48, 58]. All the proposed sub-optimal rules instead present numerical stability for realistic SNR values, although they require a different degree of system knowledge and they also differ in computational complexity. In Tab. 3.1 we report a complete comparison of the aspects mentioned. More specifically, for each fusion rule, we report the system parameters required for implementation (first column), the complexity w.r.t. both the number of sensors K and antennas N (second column), numerical stability (third column) and finally optimality properties under the assumption of identical sensors performance (last column). Also, it is apparent that the fusion rules requiring less a priori information and lower computational complexity are EGC and MRC, while Max-Log fusion rule is the only one exhibiting optimality at both low and high SNR.

Note that CV-ML requires $(P_{D,k}, P_{F,k})$, $k \in \mathcal{K}$, only if local performances of sensors are different, cfr. Eq. (3.4.1); this is not the case for CV-MMSE because local sensor performances are always required in the decoding stage of Eq. (3.4.3). Finally, MRC and EGC require only reduced knowledge of channel matrix.

The dependence of the complexity with respect to N is moderate for all the presented rules and it is polynomial in the worst-case (CV-MMSE³). This justifies the deployment of multiple antennas at DFC at the expenses of a slightly increase in the complexity burden. The dominant term of complexity for all the rules depends on the number of sensors K . The only exceptions are MRC and EGC, whose computational complexity, given by Eqs. (3.3.3) and (3.3.4), does not depend on the number of sensors K . In such a case, the dependence is only linear with respect to number of antennas, since during channel estimation step only the N -dimensional vector $\angle \mathbf{H}\mathbf{1}_K$ (corresp. $\mathbf{H}\mathbf{1}_K$), needs to be estimated.

Terms n_j , $j \in \{1, 2\}$, are inserted to underline that the *Exp*-complexity (with respect to K) of CV-ML and Max-Log can be mitigated by a GSD implementation [62]. In fact, for CV-ML, the equivalent problem $\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{X}^K} \|\mathbf{D}(\boldsymbol{\rho} - \mathbf{H}\mathbf{x})\|^2$ in place of Eq. (3.4.2) can be efficiently solved, with \mathbf{D} denoting the upper-triangular matrix deriving from the Cholesky Factorization of $\mathbf{G} \triangleq \mathbf{H}^\dagger \mathbf{H} + \beta \mathbf{I}_N$ (that is $\mathbf{G} = \mathbf{D}^\dagger \mathbf{D}$) and $\boldsymbol{\rho} \triangleq \mathbf{G}^{-1} \mathbf{H}\mathbf{y}$.

Differently, GSD implementation of Max-Log rule requires slight modifications

³It is worth remarking that CV-MMSE complexity in Tab. 3.1 has been derived under the assumption $K \gg N$.

Fusion Rule	required parameters	complexity	stability	optimality
Optimum	$(P_{D,k}, P_{F,k}), \mathbf{H}, \sigma_w^2$	$\mathcal{O}(2^K N)$	no	always
MRC	$\mathbf{H}\mathbf{1}_K$	$\mathcal{O}(N)$	yes	low SNR ¹
EGC	$\angle(\mathbf{H}\mathbf{1}_K)$	$\mathcal{O}(N)$	yes	never
Max-Log	$(P_{D,k}, P_{F,k}), \mathbf{H}, \sigma_w^2$	$\mathcal{O}(2^{(K-n_2)}N), n_2 > 0$	yes	low ² -high SNR
CV-ML	$(P_{D,k}, P_{F,k}), \mathbf{H}$	$\mathcal{O}(2^{(K-n_1)}N), n_1 > 0$	yes	high SNR
CV-MMSE	$(P_{D,k}, P_{F,k}), \mathbf{H}, \sigma_w^2$	$\mathcal{O}(K(K + NK + N^2))$	yes	never

Table 3.1: Comparison of the fusion rules. (1) if $P_D > P_F$; (2) if $(P_D > \frac{1}{2}, P_F < \frac{1}{2})$.

to the steps followed in [62]. The steps, reported in Appendix 3.8.5, lead to

$$\Lambda_{Max-Log} = \min_{\mathbf{x} \in \mathcal{X}^K} \left[\frac{\|\mathbf{D}(\boldsymbol{\rho} - \mathbf{H}\mathbf{x})\|^2}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k|H_0) \right] - \min_{\mathbf{x} \in \mathcal{X}^K} \left[\frac{\|\mathbf{D}(\boldsymbol{\rho} - \mathbf{H}\mathbf{x})\|^2}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k|H_1) \right] \quad (3.5.1)$$

The computation of Eq. (3.5.1) can be easily performed through a double search with GSD (one for each hypothesis) or with a more efficient single search, following the same approach in [70]. In both cases the complexity of Max-Log is always higher than CV-ML, that is $n_1 > n_2$. Detailed results on the complexity reduction deriving from the GSD implementations of minimum distance searches can be found in [62].

3.6 Simulation Results

In this section we compare the performances of the presented fusion rules in a WSN with sensors of identical local performances $(P_{D,k}, P_{F,k}) = (P_D, P_F)$, $k \in \mathcal{K}$. Unless differently stated, we assume $(P_D, P_F) = (0.5, 0.05)$ as adopted in [42, 44, 48, 53] for fusion rules comparison in PAC. The global performances are analyzed in terms of system probabilities of false alarm, detection and error, defined respectively as

$$P_{F_0} \triangleq P(\Lambda > \gamma|H_0), \quad P_{D_0} \triangleq P(\Lambda > \gamma|H_1), \quad (3.6.1)$$

$$P_{E_0} \triangleq \min_{\gamma} \{[1 - P_{D_0}(\gamma)]P(H_1) + P_{F_0}(\gamma)P(H_0)\}, \quad (3.6.2)$$

with Λ representing the decision statistics of a generic fusion rule. In the following figures, for comparison purposes, we report the (upper) ‘‘observation bound’’

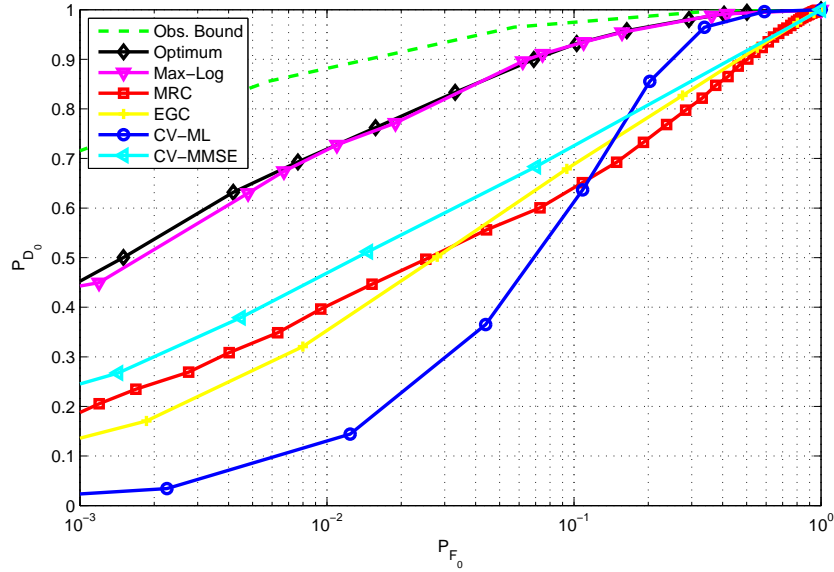


Figure 3.6.1: ROC for the presented rules. WSN with $K = 8$ sensors, $(P_{D,k}, P_{F,k}) = (0.5, 0.05)$, $k \in \mathcal{K}$. $N = 2$, $(SNR)_{dB} = 15$.

[45], i.e. the optimum performances over noiseless channel, given by:

$$P_{D_0}^{obs} = \sum_{i=K_\gamma}^K \binom{K}{i} (P_D)^i (1 - P_D)^{K-i}, \quad (3.6.3)$$

$$P_{F_0}^{obs} = \sum_{i=K_\gamma}^K \binom{K}{i} (P_F)^i (1 - P_F)^{K-i}. \quad (3.6.4)$$

where K_γ is a discrete threshold.

Receiver Operating Characteristic (ROC): In Fig. 3.6.1 we show the ROC (i.e. P_{D_0} vs P_{F_0}), for the presented rules in a WSN with $K = 8$ sensors and $N = 2$ antennas at the DFC, under a channel $(SNR)_{dB} = 15$ (corresp. $(SNR_k)_{dB} \approx 6$). It is apparent that Max-Log and Optimum ROCs are quite similar, but far from the observation bound. Instead ROCs of the MRC and EGC present a crossing point; the same happens between CV-ML and CV-MMSE. However while in the first case the result is independent of the specific channel SNR, in the latter case it depends on the poor performances of CV-ML statistics, due to the low channel SNR. This implies that when a fixed (low) P_{F_0} is imposed, as in typical Neyman-Pearson test [63], the MRC is a more attractive choice than EGC, in juxtaposition with the PAC case [42, 43]. The difference comes from the impossibility of vector \mathbf{z} in Eq. (3.3.4) to sum the contributions of all the sensors coherently.

P_{D_0} vs $(SNR)_{dB}$: In Fig. 3.6.2 we show, for the presented rules, P_{D_0} as a

function of the channel $(SNR)_{dB}$, under $P_{F_0} \leq 0.01$, in a WSN with $K = 8$ sensors (corresp. $(SNR_{R_k})_{dB} = (SNR)_{dB} - 10 \log_{10} K \approx (SNR)_{dB} - 9$); we plot the cases $N \in \{1, 2\}$ to investigate the effect on performances when two antennas are employed at the DFC. Firstly, numerical results confirm analytical derivations, i.e. CV-ML and MRC approach the optimum at high and low channel SNR, respectively, also in MIMO scenario. Max-Log is optimal at high SNR and also it strictly approaches the same performances as the optimum over all the SNR range considered (i.e. $[0, 30]_{dB}$); the pay-off is a high requirement on system knowledge and computational complexity (cfr. Tab. 3.1). It is worth noticing that curves corresponding to DtF rules in Fig. 3.6.2 exhibit jumpy-step and non-monotonic behaviors in the case of CV-ML (as pointed out in [53] for PAC) and CV-MMSE rules, respectively. Such phenomena are not surprising when related to the discrete nature of CV decision statistics as well as the operation point on the corresponding ROC. More specifically, the operation point does NOT show a fixed probability of false alarm with respect to SNR, and both probabilities of detection and false alarm are sometimes lowered in order to meet the constraint on the maximum allowed probability of false alarm. Finally CV-MMSE performs fairly better than CV-ML at low-medium SNR, as it exploits the local sensor performances information in the decoding stage. All the rules significantly benefit from the presence of two antennas at DFC (cfr. solid with dashed lines in Fig. 3.6.2). Max-Log (as the optimum) has the best improvement in the range $[5, 20]_{dB}$ and reaches the observation bound at $(SNR)_{dB} \approx 20$, instead of $(SNR)_{dB} \approx 30$ when $N = 1$ at the DFC. CV-ML rule needs higher SNR to get acceptable performances, but the case $N = 2$ still needs less energy to reach the observation bound (in fact if $N = 1$ the bound is reached at $(SNR)_{dB} > 30$, not visible in Fig. 3.6.2). Finally multiple antennas not only increase MRC, EGC and CV-MMSE performances at low-medium SNR, but also give better limiting performances.

P_{D_0} vs N : In Fig. 3.6.3, we show, for the presented rules, the P_{D_0} as a function of the number of antennas N , under $P_{F_0} \leq 0.01$; we plot the cases $(SNR)_{dB} \in \{5, 15\}$ to investigate the performances when N increases under realistic channel SNR values. It is apparent that adding more antennas at the DFC is beneficial for all the rules presented, however a saturation effect is present. The saturation depends on the SNR and the chosen fusion rule. In particular, specific configurations achieve the observation bound (e.g. Max-Log with $N = 4$ at $(SNR)_{dB} = 15$) while others (e.g. MRC with $N = 6$ at $(SNR)_{dB} = 5$) exploit all the diversity gain.

P_{E_0} vs K : In Figs. 3.6.4 and 3.6.5 we show, for the presented rules, the system probability of error P_{E_0} (under the assumption $P(H_i) = 1/2$, $H_i \in \mathcal{H}$) as a function of the number of sensors K ; we plot the case $(P_D, P_F) \triangleq (0.7, 0.05)$ to emphasize the results observed. We consider a WSN with $N = 1, 2$ (corresp. Fig. 3.6.4 and 3.6.5) antennas at DFC and a channel $(SNR)_{dB} = 15$. The latter assumption clearly represents a *total power constraint* (TPC) [57, 71] on the WSN, i.e. the individual channel SNR for the k th sensor is scaled as $SNR_k = \frac{SNR}{K}$. We also report the (upper) ‘‘communication’’ bound [45], i.e. P_{E_0} under the assumption of ideal sensors, i.e. $(P_D, P_F) = (1, 0)$. In this

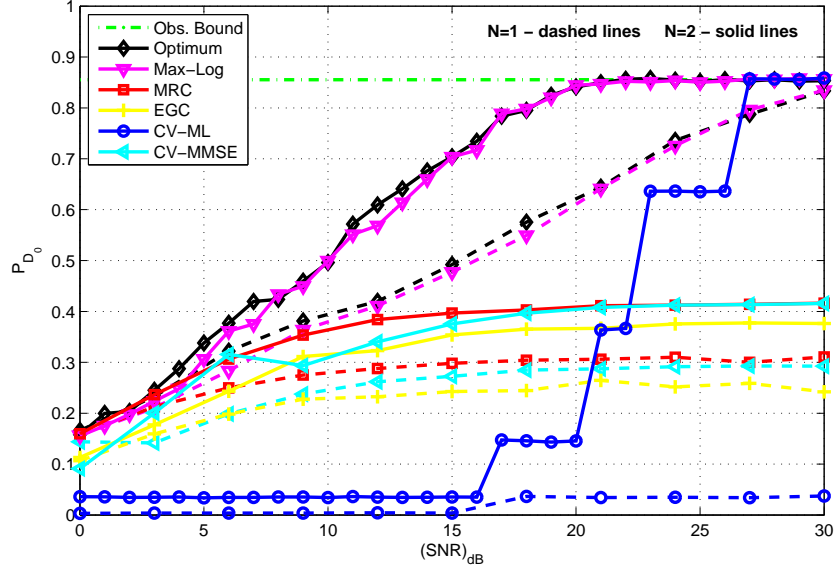


Figure 3.6.2: P_{D_0} vs channel $(SNR)_{dB}$ for the presented rules; $P_{F_0} \leq 0.01$. WSN with $K = 8$ sensors, $(P_{D,k}, P_{F,k}) = (0.5, 0.05)$, $k \in \mathcal{K}$. $N \in \{1, 2\}$.

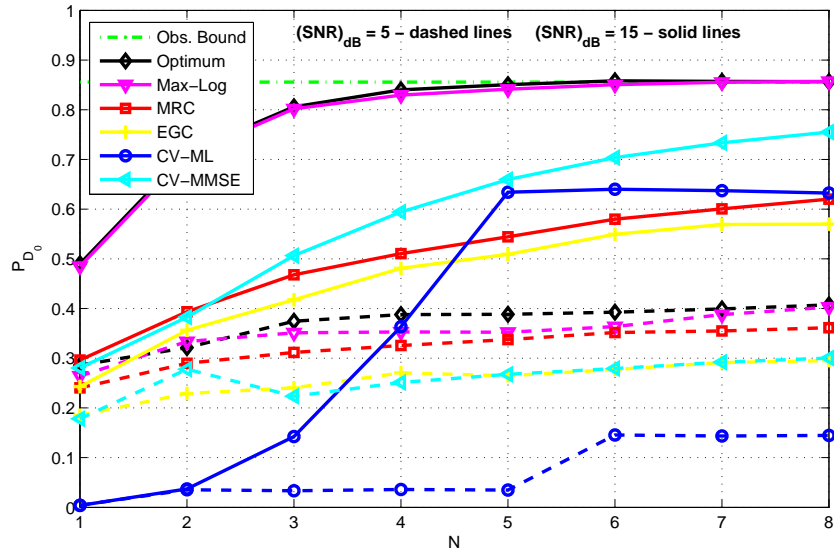


Figure 3.6.3: P_{D_0} vs N for the presented rules; $P_{F_0} \leq 0.01$. WSN with $K = 8$ sensors, $(P_{D,k}, P_{F,k}) = (0.5, 0.05)$, $k \in \mathcal{K}$. $(SNR)_{dB} \in \{5, 15\}$.

case the bound is represented by the symbol error probability of a N branch MRC combiner [72], that is $P_{E_0}^{comm} = \left(\frac{1-\mu}{2}\right)^N \sum_{l=0}^{N-1} \binom{N-1+l}{l} \left(\frac{1+\mu}{2}\right)^l$, where $\mu \triangleq \sqrt{\frac{SNR}{SNR+N}}$. Simulations show that all the rules, except for MRC, present a unimodal behaviour, i.e. there exists a finite number of sensors, under a fixed $(SNR)_{dB}$ value, which minimizes P_{E_0} . This implies that when fusing decisions using those rules, less sensors with a higher individual battery energy are more beneficial than several sensors with poor energy-powered batteries for improving performances. Instead, when fusing with MRC, the suggested trend is to divide the available energy among as many sensors as possible. Furthermore, some specific facts needs to be clarified:

- $P_{E_0}^{EGC}$ is lower than $P_{E_0}^{MRC}$ in Figs. 3.6.4,3.6.5; this result does not contradict Figs. 3.6.2,3.6.3 because the ROCs of the two rules present an intersection point, cfr. Fig. 3.6.1, so that the EGC is better than MRC from a Bayesian point of view. The optimal point on ROC which minimizes P_{E_0} is on the right of the intersection point.
- An increase in number of antennas N gives a decrease in minimum P_{E_0} attainable by every rule; the minimum is typically obtained with a larger number of sensors K . For example with CV-ML when $N = 1$ the minimum $P_{E_0}^{CV-ML} \approx 0.11$ is obtained with $K = 2$ sensors; however when $N = 2$ the minimum $P_{E_0}^{CV-ML} \approx 0.05$ is obtained with $K = 5$ sensors. Finally, the increase of N affects slope and limiting value of $P_{E_0}^{MRC}$ for MRC; this means that the same $P_{E_0}^{MRC}$ when using multiple antennas can be obtained with less equal-battery sensors.

3.7 Conclusions

In this chapter we addressed the design of sub-optimal fusion rules, more suitable for practical implementation than the exact LLR, for a DF task performed over a virtual MIMO channel. The study was motivated by the need of multiple antennas at the DFC to obtain a dramatic improvement in performances with a reduced WSN energy budget. The presented alternatives solve the issues about fixed point implementations and present a wide spectrum of choices for reduced complexity and lower system knowledge. Max-Log, MRC and CV-ML, as in the PAC case, confirm also their asymptotic optimality properties. Nonetheless, all these rules still significantly benefit from the addition of multiple antennas at the DFC, with a saturation on performance depending on the specific rule and channel SNR. Finally, all these rules present an optimal number of sensors to be employed to minimize the system error probability, under a total power constraint. The exception is represented by MRC, which shows decreasing system probability of error as a function of the number of sensors. The latter

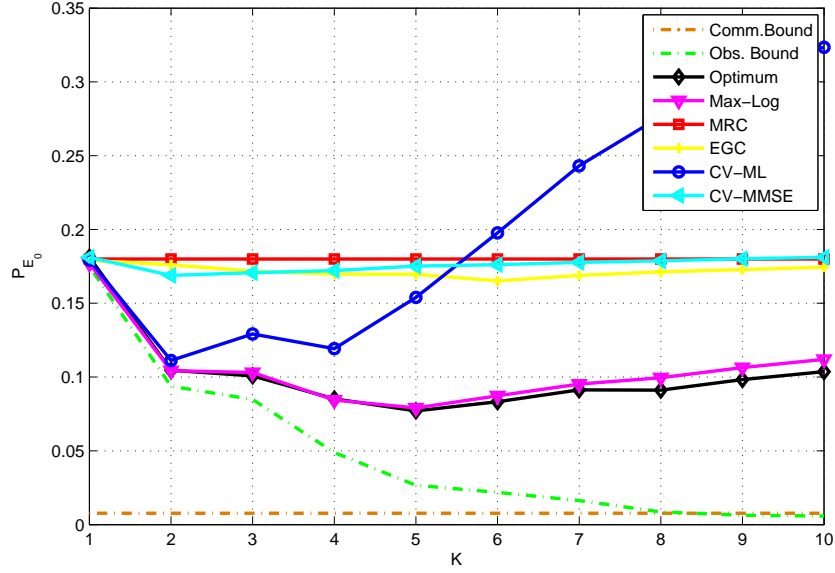


Figure 3.6.4: P_{E_0} vs K for all the rules presented; WSN with $N = 1$ antenna at DFC and channel $(SNR)_{dB} = 15$; $(P_{D,k}, P_{F,k}) = (0.7, 0.05)$, $k \in \mathcal{K}$.

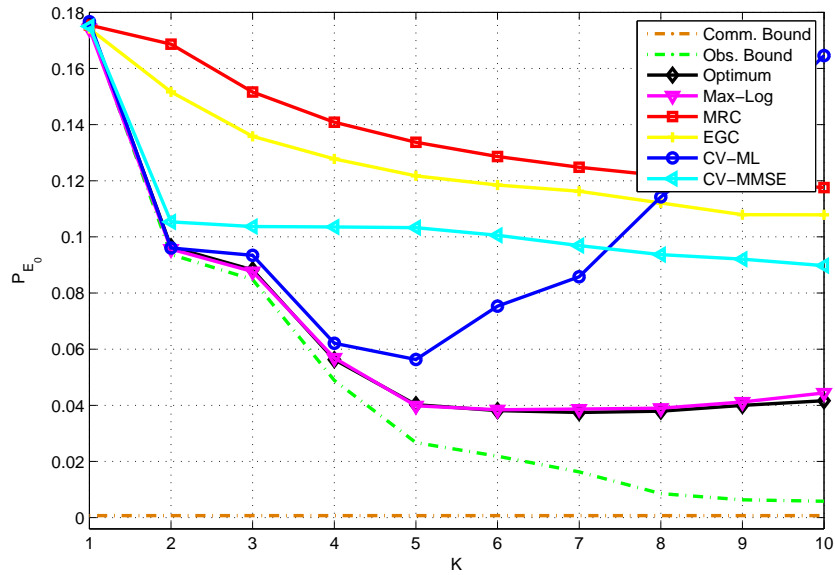


Figure 3.6.5: P_{E_0} vs K for all the rules presented; WSN with $N = 2$ antennas at DFC and channel $(SNR)_{dB} = 15$; $(P_{D,k}, P_{F,k}) = (0.7, 0.05)$, $k \in \mathcal{K}$.

property configures MRC as a more convenient choice when a huge number of low-powered sensors are available for the task.

3.8 Appendix

3.8.1 Proof of Proposition 1

From definition of exact LLR in Eq. (3.3.2) if we observe that for low SNR ($\mathbb{E}\{\|\mathbf{H}\mathbf{x}\|^2\}/\sigma_w^2 \ll 1$), $\exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_w^2}\right) \approx \exp\left(-\frac{\|\mathbf{y}\|^2}{\sigma_w^2}\right) \left(1 - \frac{\|\mathbf{H}\mathbf{x}\|^2 - 2\Re\{\mathbf{y}^\dagger \mathbf{H}\mathbf{x}\}}{\sigma_w^2}\right)$, we get

$$\Lambda_{opt} \approx \ln \left[\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} \left(1 - \frac{\|\mathbf{H}\mathbf{x}\|^2 - 2\Re\{\mathbf{y}^\dagger \mathbf{H}\mathbf{x}\}}{\sigma_w^2}\right) \prod_{k=1}^K P(x_k|H_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} \left(1 - \frac{\|\mathbf{H}\mathbf{x}\|^2 - 2\Re\{\mathbf{y}^\dagger \mathbf{H}\mathbf{x}\}}{\sigma_w^2}\right) \prod_{k=1}^K P(x_k|H_0)} \right] \quad (3.8.1)$$

Exploiting the normalization property $\sum_{\mathbf{x} \in \mathcal{X}^K} P(\mathbf{x}|H_i) = 1$, and using the approximation $\ln(1+x) \approx x$, when $x \ll 1$, we obtain:

$$\Lambda_{opt} \approx \frac{2\Re\{\mathbf{y}^\dagger \mathbf{H} (\mathbb{E}\{\mathbf{x}|H_1\} - \mathbb{E}\{\mathbf{x}|H_0\})\}}{\sigma_w^2} + \alpha \quad (3.8.2)$$

where α is a term not depending on \mathbf{y} . When local performances are identical, $\mathbb{E}\{\mathbf{x}|H_1\} = \mathbf{1}_K(2P_D - 1)$ and $\mathbb{E}\{\mathbf{x}|H_0\} = \mathbf{1}_K(2P_F - 1)$, the LLR reduces to:

$$\Lambda_{opt} \approx (P_D - P_F) \frac{4\Re\{\mathbf{y}^\dagger \mathbf{H} \mathbf{1}_K\}}{\sigma_w^2} + \alpha \quad (3.8.3)$$

Eq. (3.8.3) represents the same statistics as Eq. (3.3.3), since the terms $(P_D - P_F) \frac{4}{\sigma_w^2}$ (recall that $P_D > P_F$) and α can be incorporated in the threshold γ of Eq. (3.3.1).

3.8.2 Proof of Proposition 2

Part 1): Starting from the expression of Max-Log formula of Eq. (3.3.7) let us define:

$$\begin{aligned} \hat{\mathbf{x}}_i &\triangleq \arg \min_{\mathbf{x} \in \mathcal{X}^K} \left[\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k|H_i) \right], \quad H_i \in \mathcal{H} \\ &= \arg \min_{\mathbf{x} \in \mathcal{X}^K} \left[\frac{\|\mathbf{H}\mathbf{x}\|^2 - 2\Re\{\mathbf{y}^\dagger \mathbf{H}\mathbf{x}\}}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k|H_i) \right] \end{aligned} \quad (3.8.4)$$

where in the second line we have neglected the irrelevant term $\frac{\|\mathbf{y}\|^2}{\sigma_w^2}$. For low SNR ($\mathbb{E}\{\|\mathbf{H}\mathbf{x}\|^2\}/\sigma_w^2 \ll 1$), the first term of $\hat{\mathbf{x}}_i$ becomes irrelevant, so that the minimum

is determined by $\sum_{k=1}^K \ln P(x_k|H_i)$. If $P_{D,k} > \frac{1}{2}$ and $P_{F,k} < \frac{1}{2}$, $k \in \mathcal{K}$, then $\hat{\mathbf{x}}_0 \approx -\mathbf{1}_K$ and $\hat{\mathbf{x}}_1 \approx \mathbf{1}_K$. Using these approximations in Eq. (3.3.7) provides:

$$\Lambda_{Max-Log} \approx \frac{\|\mathbf{y} + \mathbf{H}\mathbf{1}_K\|^2}{\sigma_w^2} - \frac{\|\mathbf{y} - \mathbf{H}\mathbf{1}_K\|^2}{\sigma_w^2} + \delta \quad (3.8.5)$$

where $\delta \triangleq \ln \frac{\prod_{k=1}^K P_{D,k}}{\prod_{k=1}^K (1-P_{F,k})}$ is not dependent on \mathbf{y} . Eq. (3.8.5) is identical to MRC, cfr. Eq. (3.3.3), since δ can be easily incorporated in threshold of Eq. (3.3.1).

Part 2): The property is easily demonstrated by combining Propositions 2 and 3.1 .

Part 3): Starting from the expression of Max-Log formula of Eq. (3.3.7), for high SNR ($\mathbb{E}\{\|\mathbf{H}\mathbf{x}\|^2\}/\sigma_w^2 \gg 1$), we have that $\hat{\mathbf{x}}_i \approx \hat{\mathbf{x}}_{ML}$, since the first term in r.h.s. of Eq. (3.8.4) becomes dominant. Note that the term $\hat{\mathbf{x}}_{ML}$ is the same given by Eq. (3.4.2). Thus substituting the approximate expressions of $\hat{\mathbf{x}}_i$ in Eq. (3.3.7), we obtain the same rule as in Eq. (3.8.7), which is the CV-ML statistics. Since for high SNR, $\Lambda_{CV-ML} \approx \Lambda_{opt}$ (Proposition 4, Appendix 3.8.3), then $\Lambda_{Max-Log} \approx \Lambda_{CV-ML} \approx \Lambda_{opt}$.

3.8.3 Proof of Proposition 3

For high SNR ($\mathbb{E}\{\|\mathbf{H}\mathbf{x}\|^2\}/\sigma_w^2 \gg 1$), if we denote the true transmitted vector as \mathbf{x}_T , the corresponding value in Eq. (3.3.2) will be a dominating term of the sums at numerator and denominator; the LLR is then well approximated by

$$\begin{aligned} \Lambda_{opt} &\approx \ln \left[\frac{\exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}_T\|^2}{\sigma_w^2}\right) \prod_{k=1}^K P(x_{k,T}|H_1)}{\exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}_T\|^2}{\sigma_w^2}\right) \prod_{k=1}^K P(x_{k,T}|H_0)} \right] \\ &= \ln \left[\prod_{k=1}^K P(x_{k,T}|H_1) \right] - \ln \left[\prod_{k=1}^K P(x_{k,T}|H_0) \right]. \end{aligned} \quad (3.8.6)$$

Also, for high SNR, the ML estimate $\hat{\mathbf{x}}_{ML} = \arg \min_{\mathbf{x} \in \mathcal{X}^K} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \approx \mathbf{x}_T$, i.e. the ML decoder works near perfectly. Thus Eq. (3.8.6) reduces to

$$\Lambda_{opt} \approx \ln \left[\prod_{k=1}^K P(\hat{x}_{k,ML}|H_1) \right] - \ln \left[\prod_{k=1}^K P(\hat{x}_{k,ML}|H_0) \right] \quad (3.8.7)$$

which can be rearranged easily to obtain Eq. (3.4.1).

3.8.4 CV-MMSE moments

We derive explicit expressions for $\bar{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\}$ and $\mathbf{C} = \mathbb{E}\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^\dagger\}$ to be inserted in Eq. (3.4.3).

Proposition 5. *The explicit expressions of $\bar{\mathbf{x}}$ and \mathbf{C} are given by:*

$$\bar{x}_k = P_{F,k} + P_{D,k} - 1 \quad k \in \mathcal{K} \quad (3.8.8)$$

$$c_{\ell,j} = r_{\ell,j} - \bar{x}_\ell \bar{x}_j \quad \ell, j \in \mathcal{K} \quad (3.8.9)$$

$$r_{\ell,j} = \begin{cases} 1 & \ell = j \\ \frac{(2P_{F,\ell}-1)(2P_{F,j}-1)+(2P_{D,\ell}-1)(2P_{D,j}-1)}{2} & \ell \neq j \end{cases} \quad (3.8.10)$$

Proof. The k th element of the mean vector \bar{x}_k can be expressed as :

$$\bar{x}_k = \mathbb{E}\{x_k\} = \sum_{i=0,1} \mathbb{E}\{x_k|H_i\}P(H_i) \quad (3.8.11)$$

Assuming equally likely priors (i.e. $P(H_i) = 1/2$, $H_i \in \mathcal{H}$), observing that $\mathbb{E}\{x_k|H_0\} = 2P_{F,k} - 1$ and $\mathbb{E}\{x_k|H_1\} = 2P_{D,k} - 1$, and substituting these expressions in Eq. (3.8.11) gives Eq. (3.8.8). Define $r_{\ell,j}$ as the (ℓ, j) th element of the correlation matrix $\mathbf{R} = \mathbb{E}\{\mathbf{x}\mathbf{x}^\dagger\}$, i.e.

$$r_{\ell,j} = \mathbb{E}\{x_\ell x_j\} = \sum_{i=0,1} \mathbb{E}\{x_\ell x_j|H_i\}P(H_i) \quad (3.8.12)$$

$$= \begin{cases} \frac{1}{2} \sum_{i=0,1} \mathbb{E}\{x_\ell x_j|H_i\} & \ell \neq j \\ \frac{1}{2} \sum_{i=0,1} \mathbb{E}\{x_\ell^2|H_i\} & \ell = j \end{cases} \quad (3.8.13)$$

Exploiting the conditional independence (given H_i) of x_ℓ and x_j we have that $\mathbb{E}\{x_\ell x_j|H_i\} = \mathbb{E}\{x_\ell|H_i\} \cdot \mathbb{E}\{x_j|H_i\}$ and observing that $\mathbb{E}\{x_\ell^2|H_i\} = 1$, $H_i \in \mathcal{H}$, we obtain Eq. (3.8.8). \square

3.8.5 Max-Log GSD derivation

Proposition 6. *The Max-Log formula of Eq. (3.3.7) can be expressed in the equivalent form:*

$$\Lambda_{Max-Log} = \min_{\mathbf{x} \in \mathcal{X}^K} \left[\frac{\|\mathbf{D}(\boldsymbol{\rho} - \mathbf{H}\mathbf{x})\|^2}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k|H_0) \right] - \min_{\mathbf{x} \in \mathcal{X}^K} \left[\frac{\|\mathbf{D}(\boldsymbol{\rho} - \mathbf{H}\mathbf{x})\|^2}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k|H_1) \right] \quad (3.8.14)$$

Proof. To prove this proposition we follow similar steps as in [62]. Starting from the expression in Eq. (3.3.7), adding and subtracting the constant term $\beta \frac{\mathbf{x}^\dagger \mathbf{x}}{\sigma_w^2}$ (recall that BPSK is a constant modulus modulation, i.e. $\mathbf{x}^\dagger \mathbf{x} = K$) and

simplifying the term $\frac{\|\mathbf{y}\|^2}{\sigma_w^2}$ we get:

$$\begin{aligned} \Lambda_{Max-Log} = \min_{\mathbf{x} \in \mathcal{X}^K} & \left[\frac{\mathbf{x}^\dagger (\mathbf{H}^\dagger \mathbf{H} + \beta \mathbf{I}_N) \mathbf{x} - \mathbf{y}^\dagger \mathbf{H} \mathbf{x}}{\sigma_w^2} \right. \\ & \left. - \frac{\mathbf{x}^\dagger \mathbf{H}^\dagger \mathbf{y}}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k | H_0) \right] - \min_{\mathbf{x} \in \mathcal{X}^K} \left[\frac{\mathbf{x}^\dagger (\mathbf{H}^\dagger \mathbf{H} + \beta \mathbf{I}_N) \mathbf{x}}{\sigma_w^2} \right. \\ & \left. - \frac{\mathbf{y}^\dagger \mathbf{H} \mathbf{x} - \mathbf{x}^\dagger \mathbf{H}^\dagger \mathbf{y}}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k | H_1) \right] \end{aligned} \quad (3.8.15)$$

Denote \mathbf{D} the upper-triangular matrix deriving from the Cholesky Factorization of $\mathbf{G} \triangleq \mathbf{H}^\dagger \mathbf{H} + \beta \mathbf{I}_N$ (that is $\mathbf{G} = \mathbf{D}^\dagger \mathbf{D}$) and $\boldsymbol{\rho} \triangleq \mathbf{G}^{-1} \mathbf{H} \mathbf{y}$. We can add and subtract the term $\frac{\boldsymbol{\rho}^\dagger \mathbf{D}^\dagger \mathbf{D} \boldsymbol{\rho}}{\sigma_w^2}$ to both the minimum problems and use the relationships now defined to straightly obtain Eq. (3.8.14). The new squared radius choice r_{new}^2 for pruning is given by

$$r_{new}^2 = r^2 + \frac{\beta \mathbf{x}^\dagger \mathbf{x} + \mathbf{y}^\dagger \mathbf{H} \mathbf{G}^{-1} \mathbf{H}^\dagger \mathbf{y} - \mathbf{y}^\dagger \mathbf{y}}{\sigma_w^2} \quad (3.8.16)$$

where r represents the radius of the original test $\left[\frac{\|\mathbf{y} - \mathbf{H} \mathbf{x}\|^2}{\sigma_w^2} - \sum_{k=1}^K \ln P(x_k | H_i) \right] < r^2$, $H_i \in \mathcal{H}$. \square

The computation of the radius in Eq. (3.8.16) can be avoided if depth-first Schnorr-Euchner enumeration strategy is adopted in the pruning process [73].

Chapter 4

Performance analysis of MRC over distributed MIMO channels with instantaneous CSI

4.1 Introduction

Channel-aware decision fusion (DF) in wireless sensor networks (WSNs) is a challenging task and, due to the numerical instability and strong requirements on the system knowledge required by the log-likelihood ratio (LLR) test, several sub-optimal alternatives have been analyzed in the recent literature, such as maximum ratio combining (MRC), equal gain combining and Chair-Varshney maximum likelihood [43]. Max-Log rule has been studied in [48] and shown to outperform other sub-optimal rules though exhibiting higher requirements on system knowledge, e.g. the channel variance. All the mentioned rules were derived in the parallel-access channel scenario and, for such a case, a theoretical performance analysis was also conducted in [48].

Recently, DF exploiting the interfering nature of the broadcast wireless medium is becoming more attractive for spectral-efficiency purposes. Non-coherent modulation and censoring over parallel- and multiple-access channels (MAC) have been analyzed in [45] with emphasis on processing gain and combining loss. Techniques borrowed from direct sequence spread spectrum systems are combined with on-off keying (OOK) modulation and censoring for DF in scenarios with statistical channel state information (CSI) [58]. DF with a multiple-input multiple-output (MIMO) wireless channel model has been first studied in [47], with a focus on power allocation design based on instantaneous CSI, under the framework of J-divergence. Data fusion over MIMO channels with amplify and forward sensors was then studied in [74]. It is worth noting

that in a DF scenario the LLR suffers also from the exponential growth of the computational complexity with respect to (w.r.t.) the number of sensors. Various sub-optimal fusion rules for channel-aware DF in the same MIMO scenario are proposed in [75], where decode-and-fuse and decode-then-fuse approaches are compared. More specifically, it has been shown that the MRC fusion rule in MIMO scenario has the following appealing properties [75]: (i) even if sub-optimal, it exploits efficiently diversity from multiple antennas; (ii) it achieves optimality at low-SNR; (iii) its complexity is linear w.r.t. the number of antennas and independent of the number of sensors, both in the fusion and channel estimation stages.

In this chapter we provide a theoretical performance analysis of the MRC rule over MIMO channels. More specifically, we derive the closed form of the conditional moment generating function (MGF) and exploit Gauss-Chebyshev (GC) quadrature rules to efficiently evaluate the MRC performance in terms of the probability of detection and false alarm. The MGF also allows to evaluate the (modified) deflection coefficient in closed form, allowing practical system-design considerations. The obtained results are quite general, as they hold for scenarios with both dependent and independent local decisions.

The chapter is organized as follows: Sec. 4.2 introduces the system model; in Sec. 4.3 we present the theoretical performance analysis of MRC, confirmed in Sec. 4.4 through simulation results; in Sec. 4.5 we draw some concluding remarks; proofs are confined to Appendices.

Notation - Lower-case (resp. Upper-case) bold letters denote vectors (resp. matrices), with a_n (resp. $a_{n,m}$) representing the n th (resp. the (n,m) th) element of \mathbf{a} (resp. \mathbf{A}); upper-case calligraphic letters denote finite sets, with \mathcal{A}^K representing the k -ary Cartesian power of \mathcal{A} ; \mathbf{I}_N denotes the $N \times N$ identity matrix; $\mathbf{0}_N$ (resp. $\mathbf{1}_N$) denotes the null (resp. ones) vector of length N ; $\mathbb{E}\{\cdot\}$, $\text{var}\{\cdot\}$, $(\cdot)^t$, $(\cdot)^\dagger$, $\Re(\cdot)$, $\|\cdot\|$, $\lceil \cdot \rceil$ and $\det(\cdot)$ denote expectation, variance, transpose, conjugate transpose, real part, Frobenius norm, upper integer rounding and matrix determinant operators respectively; $P(\cdot)$ and $p(\cdot)$ are used to denote probability mass functions (pmf) and probability density functions (pdf), while $P(\cdot|\cdot)$ and $p(\cdot|\cdot)$ their corresponding conditional counterparts; $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a circularly symmetric complex normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; the symbol \sim means “distributed as”.

4.2 System Model

4.2.1 WSN setting

We consider a distributed binary hypothesis test, where K sensors are used to discriminate between the hypotheses of the set $\mathcal{H} = \{H_0, H_1\}$. For example H_0 and H_1 may represent the absence and the presence of a specific target of interest, respectively. The k th sensor, $k \in \mathcal{K} \triangleq \{1, 2, \dots, K\}$, takes a binary local decision $d_k \in \mathcal{H}$ about the observed phenomenon on the basis of its own measurements. Differently from [75], here *we do not make any conditional* (given

$H_i \in \mathcal{H}$) *mutual independence assumption* on d_k . Each decision d_k is mapped to a symbol $x_k \in \mathcal{X} = \{-1, +1\}$ representing a BPSK modulation¹: w.l.o.g. we assume that $d_k = H_i$ maps into $x_k = 2i - 1$, $i \in \{0, 1\}$. The quality of the k th sensor decisions is characterized by the conditional probabilities $P(x_k|H_j)$. More specifically, we denote $P_{D,k} \triangleq P(x_k = 1|H_1)$ and $P_{F,k} \triangleq P(x_k = 1|H_0)$ the probability of detection and false alarm of the k th sensor, respectively.

The sensors communicate with the decision fusion center (DFC) over a wireless flat-fading MAC, with i.i.d. Rayleigh fading coefficients of unitary mean power. The DFC is equipped with N receive antennas in order to exploit diversity and combat signal attenuation due to small-scale fading; this configuration determines a distributed or “virtual” MIMO channel [47, 75]. Also, instantaneous CSI and perfect synchronization² are assumed at the DFC as in [47, 75].

We denote: y_n the received signal at the n th receive antenna of the DFC after matched filtering and sampling; $h_{n,k} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ the fading coefficient between the k th sensor and the n th receive antenna of the DFC; w_n the additive white Gaussian noise at the n th receive antenna of the DFC. The vector model at the DFC is the following:

$$\mathbf{y} = \frac{1}{\sqrt{KN}} \mathbf{H} \mathbf{x} + \mathbf{w} \quad (4.2.1)$$

where $\mathbf{y} \in \mathbb{C}^N$, $\mathbf{H} \in \mathbb{C}^{N \times K}$, $\mathbf{x} \in \mathcal{X}^K$, $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma_w^2 \mathbf{I}_N)$ are the received signal vector, the channel matrix, the transmitted signal vector and the noise vector, respectively. Finally, we define the random variable $\ell(\mathbf{x})$ denoting the number of sensors deciding H_1 and the set $\mathcal{L} \triangleq \{0, \dots, K\}$, denoting the outcomes of $\ell(\mathbf{x})$.

We refer to the channel SNR as the ratio between the average total received energy from the WSN $\mathcal{E}_s \triangleq \mathbb{E} \left\{ \frac{\|\mathbf{H}\mathbf{x}\|^2}{KN} \right\}$ and the one-sided power spectral density of the continuous process noise σ_w^2 , i.e. $\text{SNR} \triangleq \mathcal{E}_s / \sigma_w^2 = \frac{1}{\sigma_w^2}$. The scale factor $\frac{1}{\sqrt{KN}}$ in Eq. (4.2.1) is added to keep the channel SNR independent on N and K , therefore the individual channel SNR for the k th sensor will be $\text{SNR}_k = \frac{1}{K\sigma_w^2}$.

4.2.2 Fusion Rules

Optimum Rule: The optimal test [63] for the considered problem can be formulated as

$$\Lambda_{opt} \triangleq \ln \left[\frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} \right] \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\geq}} \gamma \quad (4.2.2)$$

where \hat{H} , Λ_{opt} and γ denote the estimated hypothesis, the LLR (i.e. the optimal fusion rule) and the threshold to which the LLR is compared to. The threshold

¹In the case of an absence/presence task, where \mathcal{H}_0 is less probable, OOK can be employed for energy efficiency purposes. In the following we will refer only to BPSK, however the results apply straightforward to OOK.

²Multiple antennas at the DFC do not make these assumptions harder to verify w.r.t. a single antenna MAC.

γ can be determined to assure a fixed system false-alarm rate (Neyman-Pearson approach), or can be chosen to minimize the probability of error (Bayesian approach) [63]. Exploiting the independence of \mathbf{y} from H_i , given \mathbf{x} , an explicit expression of the LLR from Eq. (4.2.2) is given by

$$\Lambda_{opt} = \ln \left[\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} p(\mathbf{y}|\mathbf{x})P(\mathbf{x}|H_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} p(\mathbf{y}|\mathbf{x})P(\mathbf{x}|H_0)} \right] = \ln \left[\frac{\sum_{\mathbf{x} \in \mathcal{X}^K} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_w^2}\right) P(\mathbf{x}|H_1)}{\sum_{\mathbf{x} \in \mathcal{X}^K} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_w^2}\right) P(\mathbf{x}|H_0)} \right]. \quad (4.2.3)$$

MRC: The LLR of Eq. (4.2.3) can be simplified under the assumption of perfect sensors [48, 58], i.e. $P(\mathbf{x} = \mathbf{1}_K|H_1) = P(\mathbf{x} = -\mathbf{1}_K|H_0) = 1$. In this case $\mathbf{x} \in \{\mathbf{1}_K, -\mathbf{1}_K\}$ and Eq. (4.2.3) reduces to [75]:

$$\Lambda_{MRC} \triangleq \Re(\mathbf{z}_{MRC}^\dagger \mathbf{y}) \quad \text{with} \quad \mathbf{z}_{MRC} \triangleq \mathbf{H}\mathbf{1}_K \quad (4.2.4)$$

where irrelevant terms have been incorporated in γ through Eq. (4.2.2). It has been proved in [75] that MRC is the low-SNR approximation of the optimum of Eq. (4.2.3) when local performances of sensors are identical³.

4.3 MRC performance analysis

4.3.1 False Alarm and Detection probabilities computation

The system probabilities of false alarm and detection are

$$P_{F_0} \triangleq P(-\Lambda < -\gamma|H_0), \quad P_{D_0} \triangleq P(-\Lambda < -\gamma|H_1), \quad (4.3.1)$$

with Λ representing the decision statistic of a generic fusion rule. Although it is often difficult to derive the conditional pdf $p_{-\Lambda}(\lambda|H_i)$, $H_i \in \mathcal{H}$, of the negative LLR $-\Lambda$, the corresponding Laplace transform $\Phi_{-\Lambda}(s|H_i)$ (i.e. the MGF of Λ) is usually easier to obtain. Using the relationship $p_{-\Lambda}(\lambda|H_i) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_{-\Lambda}(s|H_i) \exp(\lambda s) ds$, where c is a small positive constant in the region of convergence of the integral, both probabilities in Eq. (4.3.1) can be rewritten as

$$\int_{-\infty}^{-\gamma} p_{-\Lambda}(\lambda|H_i) d\lambda = \int_{c-j\infty}^{c+j\infty} \Phi_{-\Lambda}(s|H_i) \cdot \frac{\exp(-\gamma s)}{2\pi j} \cdot \frac{ds}{s} \quad (4.3.2)$$

Based on Eq. (4.3.2), P_{F_0} and P_{D_0} can be calculated for any fusion rule provided that the integral in Eq. (4.3.2) can be solved efficiently and the

³Even if in [75] conditional mutual independence of local decisions was assumed, it can be shown by inspection of the derivation that such an assumption is not necessary in proving MRC optimality at low-SNR.

corresponding Laplace Transform $\Phi_{-\Lambda}(s|H_i)$ can be derived in closed form. The integral in Eq. (4.3.2) can be solved numerically through GC quadrature rules (as in Sec. 4.4) or exactly using the residue approach [76, 77]. Differently, $\Phi_{-\Lambda}(s|H_i)$ can be expanded as

$$\Phi_{-\Lambda}(s|H_i) = \sum_{\mathbf{x} \in \mathcal{X}^K} \Phi_{-\Lambda}(s|\mathbf{x})P(\mathbf{x}|H_i) \quad H_i \in \mathcal{H}. \quad (4.3.3)$$

We derive here $\Phi_{-\Lambda_{MRC}}(s|\mathbf{x})$ in closed form, as summarized by the following proposition.

Proposition 7. *The Laplace Transform $\Phi_{-\Lambda_{MRC}}(s|\mathbf{x})$ of $-\Lambda_{MRC}|\mathbf{x}$ is given in closed form as*

$$\Phi_{-\Lambda_{MRC}}(s|\mathbf{x}) = \frac{1}{\left\{ \left[1 + \frac{1}{2\sqrt{KN}} \cdot (K\xi^+ - 2\ell(\mathbf{x}))s \right] \left[1 + \frac{1}{2\sqrt{KN}} \cdot (K\xi^- - 2\ell(\mathbf{x}))s \right] \right\}^N}. \quad (4.3.4)$$

where $\xi^+ \triangleq 1 + \sqrt{1 + \frac{N}{\text{SNR}}}$ and $\xi^- \triangleq 1 - \sqrt{1 + \frac{N}{\text{SNR}}}$.

Proof. The proof is confined to Appendix 4.6.1. \square

It is worth noting that, in this particular case, Eq. (4.3.4) depends on \mathbf{x} only through $\ell(\mathbf{x})$, i.e. $\Phi_{-\Lambda}(s|\mathbf{x}) = \Phi_{-\Lambda}(s|\ell(\mathbf{x}))$. This allows a very efficient computation of Eq. (4.3.3) through

$$\Phi_{-\Lambda_{MRC}}(s|H_i) = \sum_{\ell(\mathbf{x})=0}^K \Phi_{-\Lambda_{MRC}}(s|\ell(\mathbf{x}))P(\ell(\mathbf{x})|H_i), \quad (4.3.5)$$

i.e. requiring only a sum over $(K+1)$ terms ($\ell(\mathbf{x}) \in \mathcal{L}$), instead of 2^K (cfr. with Eq. (4.3.3)). Also only $P(\ell(\mathbf{x})|H_i)$ instead of $P(\mathbf{x}|H_i)$ is needed.

4.3.2 Deflection Coefficients

Although efficient computation of P_{F_0} and P_{D_0} can be obtained through Eqs. (4.3.2), (4.3.4) and (4.3.5), a compact and concise parameter for performance analysis (independent on γ) would be desirable. The deflection coefficient D_0 and its modified version D_1 are commonly considered as synthetic performance measures; they are defined as follows [78, 79]:

$$D_i \triangleq \frac{[\mathbb{E}\{\Lambda|H_1\} - \mathbb{E}\{\Lambda|H_0\}]^2}{\text{var}\{\Lambda|H_i\}}, \quad i \in \{0, 1\}. \quad (4.3.6)$$

The deflection coefficients are widely accepted parameters⁴ for design and analysis purposes [78], while requiring only the first two order conditional moments. The expressions of $\mathbb{E}\{\Lambda_{MRC}|H_i\}$ and $\text{var}\{\Lambda_{MRC}|H_i\}$ are given in closed form by the following proposition.

⁴Also, in the specific case of a Gaussian shift-in-mean hypothesis testing, they represent the SNR of the statistic under Neyman-Pearson framework [63].

Proposition 8. *The mean and the variance of $\Lambda_{MRC}|H_i$ are:*

$$\mathbb{E}\{\Lambda_{MRC}|H_i\} = 2\sqrt{\frac{N}{K}} \mathbb{E}\{\ell|H_i\} - \sqrt{KN} \quad (4.3.7)$$

$$\text{var}\{\Lambda_{MRC}|H_i\} = K\left(1 + \frac{N}{2\text{SNR}}\right) - 2\mathbb{E}\{\ell|H_i\} + \frac{2}{K}\mathbb{E}\{\ell^2|H_i\} + \frac{4N}{K}\text{var}\{\ell|H_i\} \quad (4.3.8)$$

Proof. The moments of $\Lambda_{MRC}|H_i$ are obtained from the MGF definition [80]:

$$\mathbb{E}\{(\Lambda_{MRC})^m|H_i\} = \left. \frac{d^m [\Phi_{-\Lambda_{MRC}}(s|H_i)]}{ds^m} \right|_{s=0} \quad (4.3.9)$$

Hence, the first two order moments are obtained by setting $m = 1$ and $m = 2$, respectively. Finally, the variance is computed as $\text{var}\{\Lambda_{MRC}|H_i\} = \mathbb{E}\{\Lambda_{MRC}^2|H_i\} - \mathbb{E}\{\Lambda_{MRC}|H_i\}^2$. \square

As a corollary we also report D_i^{MRC} explicitly in the case of conditionally i.i.d. sensor decisions, i.e. $P(\mathbf{x}|H_i) = \prod_{k=1}^K P(x_k|H_i)$ and $P_D = P_{D,k}$, $P_F = P_{F,k}$.

Corollary 9. *The deflection coefficients for conditionally i.i.d. sensor decisions are:*

$$D_i^{MRC} = \frac{4NK(P_D - P_F)^2}{K\left(1 + \frac{N}{2\text{SNR}}\right) + 2(2N + 1 - K)\rho_i}, \quad (4.3.10)$$

where $\rho_0 \triangleq P_D(1 - P_D)$ and $\rho_1 \triangleq P_F(1 - P_F)$. Note that D_i^{MRC} are strictly increasing with K and therefore the large system (i.e. $K \rightarrow +\infty$) deflections, $D_i^{MRC} = \frac{4N(P_D - P_F)^2}{(1 + \frac{N}{2\text{SNR}}) - 2\rho_i}$, represent the maximum attainable, given the other parameters, as shown in [75] through simulations. For such a reason it is interesting to obtain a coarse indication on the minimum K needed to achieve the maximum deflection with a *fixed loss*:

$$K_{\epsilon_r, i} : \frac{D_i^{MRC}(K_{\epsilon_r, i} + 1) - D_i^{MRC}(K_{\epsilon_r, i})}{D_i^{MRC}(K_{\epsilon_r, i})} \leq \epsilon_r \quad (4.3.11)$$

where ϵ_r represents a fixed tolerance on the relative increment of the deflection coefficients w.r.t. K . The dependence of D_i^{MRC} w.r.t. K is stressed in Eq. (4.3.11). If we define $b_i \triangleq \left[1 + \frac{N}{2\text{SNR}}\right] - 2\rho_i$ and $c_i \triangleq [2(2N + 1)\rho_i]$ it can be shown that

$$K_{\epsilon_r, i} = \left\lceil \frac{\sqrt{\epsilon_r^2(b_i + c_i)^2 + 4\epsilon_r c_i b_i} - \epsilon_r(b_i + c_i)}{2\epsilon_r b_i} \right\rceil. \quad (4.3.12)$$

The proof is not reported here for brevity.

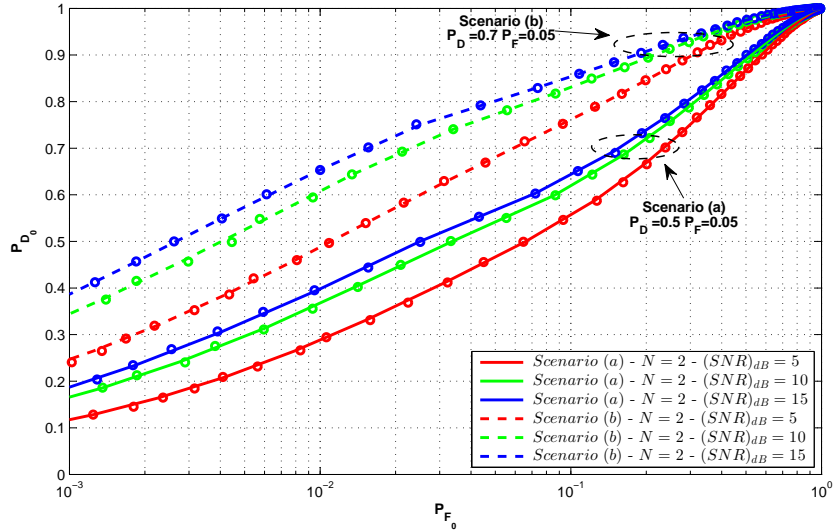


Figure 4.4.1: MGF-based (lines) and simulation-based (circles) ROC. WSN with $K = 8$, $N = 2$, $(SNR)_{dB} \in \{5, 10, 15\}$. $(P_{D,k}, P_{F,k}) = (0.5, 0.05)$ for Scenario (a); $(P_{D,k}, P_{F,k}) = (0.7, 0.05)$ for Scenario (b).

4.4 Simulations Results

In this section we show the perfect agreement between the results (in terms of P_{F_0} and P_{D_0}) obtained by evaluating the integration of $\Phi_{-\Lambda_{MRC}}(s|H_i)$ in Eq. (4.3.2), through GC rules (with 500 nodes) [76, 77], and the results obtained through Monte Carlo simulations (with 10^5 runs). Also, we show the benefits of using of D_i as a performance metric for MRC. For simplicity we consider conditionally i.i.d. sensor decisions, i.e. $(P_{D,k}, P_{F,k}) = (P_D, P_F)$, $k \in \mathcal{K}$.

Fig. 4.4.1 shows the Receiver Operating Characteristics (ROC) of the MRC rule in a WSN with $K = 8$ and $N = 2$ at the DFC in two scenarios with different P_D (we fix $P_F = 0.05$): (a) $P_D = 0.5$; (b) $P_D = 0.7$. For each scenario we report the performance at three different SNR values, i.e. $(SNR)_{dB} \in \{5, 10, 15\}$. The lines represent the performance evaluated with the proposed approach, while the circle markers represent the corresponding simulations. It is apparent how the proposed approach perfectly matches the simulation results, while requiring dramatically minimal computational resources.

Fig. 4.4.2 shows D_i vs K , evaluated according to Eq. (4.3.10), with $N \in \{1, 2\}$ at the DFC and for $(SNR)_{dB} \in \{5, 15\}$. It is apparent how the performance saturates, under a total power constraint (cfr. Eq. (4.2.1)), with K , as well as the beneficial effect of the increased number of antennas. In the same figure we also report, for each plot, the corresponding value of K_{ϵ_r} with $\epsilon_r \triangleq 0.01$,

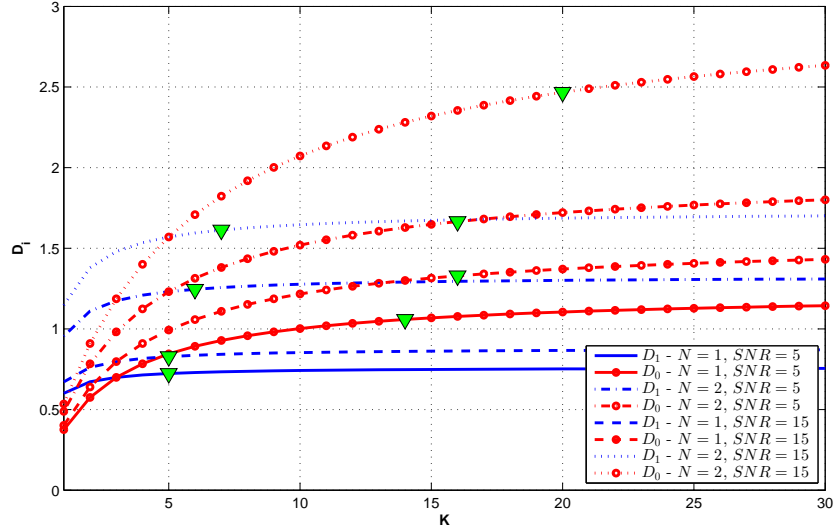


Figure 4.4.2: D_i vs K ; WSN with $(P_{D,k}, P_{F,k}) = (0.5, 0.05)$; $N \in \{1, 2\}$; $(SNR)_{dB} \in \{5, 15\}$. K_{ϵ_r} is reported with a triangle marker, $\epsilon_r \triangleq 0.01$.

obtained through Eq. (4.3.12); the 1% limit on the relative increment of D_i provides insights on the number of sensors required to attain the maximum with a negligible loss.

4.5 Conclusions

In this chapter we exploited the MGF framework for performance evaluation of MRC DF rule over MIMO channels. We derive the MGF of the MRC over MIMO channels in closed form in order to compute effectively the system probabilities of false alarm and detection via GC quadrature rules (numerically). The explicit expression of conditional MGF is also exploited to obtain the deflection coefficients in closed form, as a compact measure of the statistic performance, thus allowing practical system design.

4.6 Appendix

4.6.1 Proof of Proposition 7

We derive here the closed form of $\Phi_{-\Lambda_{MRC}}(s|\mathbf{x})$ in Eq. (4.3.4). Using a similar approach as in [81], where the symbol error probability in a fading environment

with antenna diversity was obtained, we express $-\Lambda_{MRC}|\mathbf{x}$ as follows

$$-\Lambda_{MRC}|\mathbf{x} = \sum_{n=1}^N (\mathbf{v}_n|\mathbf{x})^\dagger \cdot \mathbf{F} \cdot (\mathbf{v}_n|\mathbf{x}), \quad (4.6.1)$$

where the Gaussian vectors $(\mathbf{v}_n|\mathbf{x})$, $n \in \{1, \dots, N\}$, and the deterministic matrix \mathbf{F} have the explicit expressions:

$$\mathbf{v}_n|\mathbf{x} \triangleq [y_n|\mathbf{x} \quad \mathbf{h}_{r,n}\mathbf{1}_K]^t \quad \mathbf{F} \triangleq \begin{bmatrix} 0 & -\frac{1}{2} \\ -\frac{1}{2} & 0 \end{bmatrix} \quad (4.6.2)$$

with $\mathbf{h}_{r,n}$ denoting the n th row of \mathbf{H} . Note that Eq. (4.6.1) is a sum of Hermitian quadratic forms of complex Gaussian vectors $\mathbf{v}_n|\mathbf{x}$. Since $\mathbf{v}_n|\mathbf{x}$ are i.i.d. vectors, the Laplace transform of Eq. (4.6.1) has the following closed form [80]:

$$\Phi_{-\Lambda_{MRC}}(s|\mathbf{x}) = \left[\frac{1}{\det(\mathbf{I} + s\mathbf{L}(\mathbf{x}))} \right]^N \quad (4.6.3)$$

where $\mathbf{L}(\mathbf{x}) \triangleq (\mathbf{R}|\mathbf{x}) \cdot \mathbf{F}$ and $\mathbf{R}|\mathbf{x} \triangleq \mathbb{E}\{(\mathbf{v}_n|\mathbf{x})(\mathbf{v}_n|\mathbf{x})^\dagger\}$, i.e. the covariance matrix of $\mathbf{v}_n|\mathbf{x}$, since it can be shown that $\mathbb{E}\{(\mathbf{v}_n|\mathbf{x})\} = \mathbf{0}_2$. The explicit expression of $\mathbf{R}|\mathbf{x}$ is:

$$\mathbf{R}|\mathbf{x} = \mathbb{E}\{(\mathbf{v}_n|\mathbf{x})(\mathbf{v}_n|\mathbf{x})^\dagger\} = \begin{bmatrix} \frac{1}{N} + \sigma_w^2 & \frac{2\ell(\mathbf{x})-K}{\sqrt{KN}} \\ \frac{2\ell(\mathbf{x})-K}{\sqrt{KN}} & K \end{bmatrix}. \quad (4.6.4)$$

Denoting $\lambda_i(\mathbf{x})$, $i \in \{1, 2\}$, the two eigenvalues of $\mathbf{L}(\mathbf{x})$ we have that

$$\Phi_{-\Lambda_{MRC}}(s|\mathbf{x}) = \left[\frac{1}{(1 + s\lambda_1(\mathbf{x}))(1 + s\lambda_2(\mathbf{x}))} \right]^N. \quad (4.6.5)$$

Evaluation of $\lambda_i(\mathbf{x})$, through $\det(s\mathbf{I} - \mathbf{L}(\mathbf{x})) = 0$, gives:

$$\lambda_i(\mathbf{x}) = \frac{K - 2\ell(\mathbf{x})}{2\sqrt{KN}} \pm \frac{1}{2} \sqrt{K \left(\frac{1}{N} + \frac{1}{\text{SNR}} \right)}, \quad (4.6.6)$$

where we have exploited that $\text{SNR} = 1/\sigma_w^2$. Direct substitution of explicit expression of $\lambda_i(\mathbf{x})$ in Eq. (4.6.5) provides the result.

Chapter 5

Non-coherent Decision Fusion over distributed MIMO channels

5.1 Introduction

Starting from classical distributed detection [50], large efforts in the recent literature have been devoted to the implementation of distributed detection in wireless sensor networks (WSNs) [43, 51, 52]. Local decisions in a WSN are usually transmitted to a decision fusion center (DFC) in order to improve reliability of geographically distributed sensing through central processing. Common system architectures make reference to the availability of parallel (non-interfering) channels from the sensors to the DFC [42, 53, 48]. However, more sophisticated setups have been investigated, where the intrinsically interfering nature of the wireless channel is exploited and not combated [45, 47, 75].

Recently, in [45] and [82], the received-energy test was studied for non-coherent decision fusion over a multiple access channel (MAC). More specifically, in [45] the received energy was claimed as optimal for the no-diversity case with conditionally (given the phenomenon) mutually independent and identically distributed (i.i.d.) sensor local decisions, as long as the probability of false alarm of the generic sensor is lower than the corresponding probability of detection. Also, analytical performances of the received-energy test in the diversity scenario were derived. However, optimality property of the test was not investigated. The optimality of the test for the no-diversity case with conditionally i.i.d. sensor local decisions was proven in [82]. Only the case with sensors whose probability of false alarm is lower than the corresponding probability of detection was considered. Nonetheless, the diversity case was still ignored in the optimality analysis.

The main contributions of this chapter are:

- a rigorous proof of the *optimality* of the received-energy test for non-coherent decision fusion¹ over a Rayleigh fading MAC with *arbitrary order of diversity* and with conditionally mutually independent *but non identically distributed* (i.n.i.d.) sensor decisions, as long as *each* sensor probability of false alarm is lower than the correspondent probability of detection;
- as a side result, a sufficient condition on the log-likelihood ratio (LLR) of the number of active sensors suited for testing received energy optimality in scenarios with correlated local decisions;
- analytical derivation of large-system performances for conditionally i.i.d. sensor local decisions as a function of the order of diversity, where two different scenarios are considered: (a) sensors with an individual power constraint (IPC); (b) sensors with a total power constraint (TPC).

It is worth noting that in [75] a different scenario was analyzed, where: (i) conditionally i.i.d. sensor decisions were considered, and (ii) instantaneous channel state information (CSI) at the DFC was assumed. The focus was on the performance analysis of several sub-optimal fusion rules in terms of complexity, required knowledge, probability of detection and false alarm.

The chapter is organized as follows: Sec. 5.2 introduces the system model; in Sec. 5.3 we present the main results of this chapter, while in Sec. 5.4 we draw some concluding remarks; proofs are confined to Appendices.

Notation - Lower-case (resp. Upper-case) bold letters denote vectors (resp. matrices), with a_n (resp. $a_{n,m}$) representing the n th (resp. the (n,m) th) element of the vector \mathbf{a} (resp. matrix \mathbf{A}); upper-case calligraphic letters, e.g. \mathcal{A} , denote discrete and finite sets; \mathbf{I}_N denotes the $N \times N$ identity matrix; $\mathbf{0}_N$ (resp. $\mathbf{1}_N$) denotes the null (resp. ones) vector of length N ; $\mathbb{E}\{\cdot\}$, $(\cdot)^t$, $(\cdot)^\dagger$, $\Re(\cdot)$, $\Im(\cdot)$ and $\|\cdot\|$ denote expectation, transpose, conjugate transpose, real part, imaginary part and Frobenius norm operators; $P(\cdot)$ and $p(\cdot)$ are used to denote probability mass functions (pmf) and probability density functions (pdf), while $P(\cdot|\cdot)$ and $p(\cdot|\cdot)$ their corresponding conditional counterparts; $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (resp. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) denotes a circular symmetric complex (resp. real) normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\mathcal{B}(k, p)$ denotes a binomial distribution of k trials with probability of success p and χ_L^2 denotes a chi-square distribution with L degrees of freedom; $(a * b)(\ell)$ denotes the convolution between series $a(\ell)$ and $b(\ell)$; finally the symbols \sim and \xrightarrow{d} mean “distributed as” and “convergence in distribution”.

¹Although, energy receiver and non-coherent are not synonyms, in the chapter we will confuse them. In the related literature, such a misuse is common due to the fact that the energy detector is the default receiver adopted for non-coherent decision fusion.

5.2 System Model

5.2.1 WSN modeling

We consider a distributed binary hypothesis test, where K sensors are used to discriminate between the hypotheses of the set $\mathcal{H} = \{H_0, H_1\}$, representing, not necessarily, the absence (H_0) or the presence (H_1) of a specific target of interest. The k th sensor, $k \in \mathcal{K} \triangleq \{1, 2, \dots, K\}$, takes a binary local decision $d_k \in \mathcal{H}$ about the observed phenomenon on the basis of its own measurements.

Each decision d_k is mapped to a symbol $x_k \in \mathcal{X} = \{0, 1\}$ representing an On-Off Keying (OOK) modulation: without loss of generality we assume that $d_k = H_i$ maps into $x_k = i$, $i \in \{0, 1\}$. The quality of the k th sensor decisions is characterized by the conditional probabilities $P(x_k|H_j)$. More specifically, we denote $P_{D,k} \triangleq P(x_k = 1|H_1)$ and $P_{F,k} \triangleq P(x_k = 1|H_0)$, respectively the probability of detection and false alarm of the k th sensor.

The sensors communicate with the DFC over a wireless flat-fading MAC, modeled through i.i.d. Rayleigh fading coefficients with equal mean power. The DFC employs an N -diversity approach in order to combat signal attenuation due to small-scale fading of the wireless medium. The diversity can be accomplished with time, frequency, code or antenna diversity (as recently proposed in [47, 46]). Statistical CSI is assumed at the DFC, i.e. only the pdf of each fading coefficient is available.

We denote: y_n the received signal at the n th diversity branch of the DFC after matched filtering and sampling; $h_{n,k} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_h^2)$ the fading coefficient between the k th sensor and the n th diversity branch of the DFC²; w_n the additive white Gaussian noise at the n th diversity branch of the DFC. The vector model at the DFC is the following:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (5.2.1)$$

where $\mathbf{y} \in \mathbb{C}^N$, $\mathbf{H} \in \mathbb{C}^{N \times K}$, $\mathbf{x} \in \mathcal{X}^K$, $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma_w^2 \mathbf{I}_N)$ are the received signal vector, the channel matrix, the transmitted signal vector and the noise vector, respectively. Finally, we define the random variable $\ell \triangleq \ell(\mathbf{x}) = \sum_{k=1}^K x_k$, representing the number of active sensors and the set $\mathcal{L} \triangleq \{0, \dots, K\}$ of possible realizations of ℓ .

²It is worth noting that assuming an asymmetric model for channel coefficient statistics would be more realistic. However, this would make the results much more dependent on the specific scenario without adding any significant insight from a theoretical point of view. A symmetric model for channel coefficient statistics is assumed for a two-fold reason: on one side it can be considered as a starting point before analyzing more realistic application-dependent scenarios; on the other side a symmetric scenario could represent scenarios in which power control is considered.

5.2.2 LLR

The optimal test [50, 63] for the considered problem can be formulated as

$$\Lambda_{opt} \triangleq \ln \left[\frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} \right] \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\geq}} \gamma \quad (5.2.2)$$

where \hat{H} , Λ_{opt} and γ denote the estimated hypothesis, the Log-Likelihood-Ratio (LLR, i.e. the optimal fusion rule) and the threshold to which the LLR is compared to. The threshold γ can be determined to assure a fixed system false-alarm rate (Neyman-Pearson approach) or can be chosen to minimize the probability of error (Bayes approach) [50, 63]. An explicit expression of the LLR from Eq. (5.2.2) is given by

$$\Lambda_{opt} = \ln \left[\frac{\sum_{\ell=0}^K p(\mathbf{y}|\ell)P(\ell|H_1)}{\sum_{\ell=0}^K p(\mathbf{y}|\ell)P(\ell|H_0)} \right] \quad (5.2.3)$$

$$= \ln \left[\frac{\sum_{\ell=0}^K \frac{1}{(\sigma_w^2 + \ell\sigma_h^2)^N} \exp\left(-\frac{\|\mathbf{y}\|^2}{\sigma_w^2 + \ell\sigma_h^2}\right) P(\ell|H_1)}{\sum_{\ell=0}^K \frac{1}{(\sigma_w^2 + \ell\sigma_h^2)^N} \exp\left(-\frac{\|\mathbf{y}\|^2}{\sigma_w^2 + \ell\sigma_h^2}\right) P(\ell|H_0)} \right] \quad (5.2.4)$$

where we have exploited the conditional independence of \mathbf{y} from H_i (given ℓ).

In the case of conditionally (given H_i) i.i.d. sensor decisions ($(P_{D,k}, P_{F,k}) = (P_D, P_F)$, $k \in \mathcal{K}$) we have that $\ell|H_1 \sim \mathcal{B}(K, P_D)$ and $\ell|H_0 \sim \mathcal{B}(K, P_F)$. Differently, when local sensor decisions are conditionally i.n.i.d. the pmfs $P(\ell|H_i)$ are represented by the more general *Poisson-Binomial* distribution [83, 84, 85], with expressions given by:

$$P(\ell|H_1) = \sum_{\mathbf{x}: \ell(\mathbf{x})=\ell} \prod_{k=1}^K (P_{D,k})^{x_k} \prod_{s=1}^K (1 - P_{D,s})^{(1-x_s)} \quad (5.2.5)$$

$$P(\ell|H_0) = \sum_{\mathbf{x}: \ell(\mathbf{x})=\ell} \prod_{k=1}^K (P_{F,k})^{x_k} \prod_{s=1}^K (1 - P_{F,s})^{(1-x_s)} \quad (5.2.6)$$

It is worth noting that Eq. (5.2.5) requires sums which are infeasible to compute in practice unless the number of sensors K is small. For this reason different methods have been proposed in literature for its efficient evaluation. The alternatives include fast convolution of individual Bernoulli pmfs [83], recursive approaches [84] and a Discrete Fourier Transform (DFT) based computation [85].

5.3 Optimality of Received Energy

As already stated in [45], the received energy $\psi \triangleq \|\mathbf{y}\|^2$ is a sufficient statistic for the LLR, since Eq. (5.2.3) depends on \mathbf{y} only through ψ . However, *sufficiency*

alone does not guarantee that the test

$$\psi \underset{\hat{H}=H_0}{\overset{\hat{H}=H_1}{\geq}} \gamma' \quad (5.3.1)$$

is equivalent to Eq. (5.2.2). As shown in [82], the test in Eq. (5.3.1) is optimal iff $\Lambda_{opt}(\psi)$ is a strictly increasing function of ψ . If this property is satisfied, the test in Eq. (5.2.2) is equivalent to Eq. (5.3.1) by simply setting $\gamma' = \Lambda_{opt}^{-1}(\gamma)$. For this purpose in the following we first introduce a general optimality test (in the form of a sufficient condition) which relates the pmfs $P(\ell|H_i)$, $H_i \in \mathcal{H}$, to assure that $\Lambda_{opt}(\psi)$ is strictly increasing in the case of an N -diversity MAC.

Proposition 10. *A sufficient condition for $\Lambda_{opt}(\psi)$ to be a strictly increasing function of ψ is given by:*

$$\lambda(\ell) > \lambda(\ell - 1), \quad \ell \in \mathcal{L} \setminus \{0\} \quad (5.3.2)$$

where $\lambda(\ell) \triangleq \ln \left[\frac{P(\ell|H_1)}{P(\ell|H_0)} \right]$.

Proof. The proof is reported in Appendix 5.5.1. \square

The above proposition states that strictly increasing property of $\lambda(\ell)$ assures optimality of the test in Eq. (5.3.1). We will refer to $\lambda(\ell)$ as the ℓ -LLR hereinafter³. It is worth noting that such a sufficient condition is independent of the order of diversity N . Also, Eq. (5.3.2) depends on the WSN model only through the number of active sensors ℓ (given H_i) and does not require any specific assumption on $P(\mathbf{x}|H_i)$, e.g. conditional mutual independence of local sensor decisions, i.e. $P(\mathbf{x}|H_i) = \prod_{k=1}^K P(x_k|H_i)$. This means that Eq. (5.3.2) plays the role of a general property for received energy optimality, to be verified even in the case of conditionally correlated local sensor decisions.

In the simplest case of conditionally i.i.d. local sensor decisions, $(P_{D,k}, P_{F,k}) = (P_D, P_F)$, $k \in \mathcal{K}$, as assumed in [45, 82], the strictly increasing property of $\lambda(\ell)$, $\ell \in \mathcal{L}$, is equivalent to

$$\frac{\binom{K}{\ell} P_D^\ell (1 - P_D)^{K-\ell}}{\binom{K}{\ell} P_F^\ell (1 - P_F)^{K-\ell}} > \frac{\binom{K}{\ell-1} P_D^{\ell-1} (1 - P_D)^{K-\ell+1}}{\binom{K}{\ell-1} P_F^{\ell-1} (1 - P_F)^{K-\ell+1}}, \quad (5.3.3)$$

that reduces to $P_D > P_F$. This result, not only confirms theoretical findings for optimality of Eq. (5.3.1) when $N = 1$ as in [45, 82], but it also proves the optimality of the test over the *Diversity MAC* (i.e. $N > 1$) used in [45]; this result shows the effectiveness and simplicity of Proposition 10 w.r.t. the approach taken in [82].

Differently, when sensor decisions are conditionally i.n.i.d. (i.e. the case of a heterogeneous WSN), the following theorem generalizes the result in Eq. (5.3.3).

³Note that we will not consider in Eq. (5.3.2) (and throughout the chapter) the case $\ell = 0$ when testing ℓ -LLR strictly increasing property, since $\lambda(-1)$ has no physical meaning.

Theorem 11. If $P(\mathbf{x}|H_i) = \prod_{k=1}^K P(x_k|H_i)$, $H_i \in \mathcal{H}$, and $P_{D,k} > P_{F,k}$, $k \in \mathcal{K}$, the ℓ -LLR satisfies the strictly increasing property described in Eq. (5.3.2) and thus Eq. (5.3.1) is the optimal test when an N -diversity MAC is employed.

Proof. The proof is reported in Appendix 5.5.2. \square

Theorem 11 states that under non identical sensors ($P_{D,k}, P_{F,k}$) and N -diversity MAC the received energy ψ is again the optimal test. Also, Theorem 11 relies on a sufficient condition, i.e. specific WSN configurations not satisfying the assumption $P_{D,k} > P_{F,k}$, $k \in \mathcal{K}$, but still verifying Eq. (5.3.2) may exist. Although such a case is not “typically” of interest, since the condition $P_{D,k} \leq P_{F,k}$ for k th sensor is not realistic in practical scenarios (i.e. sensors operating under nominal conditions), it proves the robustness of the received energy in scenarios with some faulty (or byzantine) sensors⁴.

We finally evaluate analytically the performances as the number of sensors goes large, in the case of conditionally i.i.d. sensors. Both IPC and TPC on the WSN and arbitrary diversity N are considered here. This result generalizes [82], where no-diversity ($N = 1$) and IPC assumptions were made in deriving formulas. We define

$$\mathbf{z} \triangleq \frac{1}{\sqrt{P_F K \sigma_h^2}} \left(\frac{\mathbf{H}\mathbf{x}}{\sqrt{N}} + \mathbf{w} \right) \quad (5.3.4)$$

where, compared to Eq. (5.2.1), $\frac{1}{\sqrt{P_F K \sigma_h^2}}$ is a merely scaling factor and $\mathbf{H}\mathbf{x}$ is replaced with $\frac{\mathbf{H}\mathbf{x}}{\sqrt{N}}$ in \mathbf{z} in order to keep a fixed amount of average energy $\varepsilon \triangleq E\{\|\mathbf{H}\mathbf{x}\|^2\}$ w.r.t. N . Then we define the system probabilities of false alarm and detection, respectively P_{F_0} and P_{D_0} , as:

$$P_{F_0} \triangleq P(\|\mathbf{z}\|^2 \geq \bar{\gamma} | H_0), \quad P_{D_0} \triangleq P(\|\mathbf{z}\|^2 \geq \bar{\gamma} | H_1), \quad (5.3.5)$$

Eqs. (5.3.4) and (5.3.5) hold for TPC scenario when replacing \mathbf{z} with $\tilde{\mathbf{z}} \triangleq \frac{1}{\sqrt{P_F \sigma_h^2}} \left(\frac{\mathbf{H}\mathbf{x}}{\sqrt{KN}} + \mathbf{w} \right)$. In this case the average energy is kept fixed w.r.t. both K and N .

Theorem 12. If σ_h^2 and σ_w^2 are finite, as $K \rightarrow +\infty$

$$\mathbf{z}|H_0 \xrightarrow{d} \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}_N, \frac{1}{N} \mathbf{I}_N \right), \quad \mathbf{z}|H_1 \xrightarrow{d} \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}_N, \frac{P_D}{P_F N} \mathbf{I}_N \right), \quad (5.3.6)$$

$$\tilde{\mathbf{z}}|H_0 \xrightarrow{d} \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}_N, \frac{1}{\alpha_F} \frac{1}{N} \mathbf{I}_N \right), \quad \tilde{\mathbf{z}}|H_1 \xrightarrow{d} \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}_N, \frac{1}{\alpha_D} \frac{P_D}{P_F N} \mathbf{I}_N \right), \quad (5.3.7)$$

where $\alpha_F \triangleq \frac{P_F \sigma_h^2}{P_F \sigma_h^2 + \sigma_w^2 N}$ and $\alpha_D \triangleq \frac{P_D \sigma_h^2}{P_D \sigma_h^2 + \sigma_w^2 N}$. Then the large-system

⁴A WSN with $K = 3$ sensors such that $(P_{D,1}, P_{F,1}) = (0.5, 0.05)$, $(P_{D,2}, P_{F,2}) = (0.4, 0.1)$ and $(P_{D,3}, P_{F,3}) = (0.3, 0.4)$ verifies the property in Eq. (5.3.2).

$(P_{D_0-IPC}^*, P_{F_0-IPC}^*)$ are given by:

$$P_{F_0-IPC}^*(\bar{\gamma}) = \exp(-\bar{\gamma}N) \times \sum_{n=0}^{N-1} \frac{1}{n!} (\bar{\gamma}N)^n; \quad (5.3.8)$$

$$P_{D_0-IPC}^*(\bar{\gamma}) = \exp\left(-\frac{\bar{\gamma}N}{\frac{P_D}{P_F}}\right) \times \sum_{n=0}^{N-1} \frac{1}{n!} \left(\frac{\bar{\gamma}N}{\frac{P_D}{P_F}}\right)^n; \quad (5.3.9)$$

while $(P_{D_0-TPC}^*, P_{F_0-TPC}^*)$ are given by:

$$P_{F_0-TPC}^*(\bar{\gamma}) = \exp(-\bar{\gamma}N\alpha_F) \times \sum_{n=0}^{N-1} \frac{1}{n!} (\bar{\gamma}N\alpha_F)^n; \quad (5.3.10)$$

$$P_{D_0-TPC}^*(\bar{\gamma}) = \exp\left(-\frac{\bar{\gamma}N\alpha_D}{\frac{P_D}{P_F}}\right) \times \sum_{n=0}^{N-1} \frac{1}{n!} \left(\frac{\bar{\gamma}N\alpha_D}{\frac{P_D}{P_F}}\right)^n. \quad (5.3.11)$$

Proof. The proof is reported in Appendix 5.5.3 for the IPC case; performance in the TPC scenario can be derived in a similar fashion. \square

As expected, if $N = 1$ the result of Eqs. (5.3.8) and (5.3.9) coincides with the one given in [82, Sec. IV].

It is worth remarking that, in both IPC and TPC scenarios with diversity, Neyman-Pearson and Bayesian error exponents are zero (cfr. with [82]), because the large-system ROC can not be driven toward the point $(P_{D_0}^*, P_{F_0}^*) = (1, 0)$ by increasing the number of sensors, as long as the diversity N is kept finite. This intuition is confirmed by the non-zero values assumed under an IPC and a TPC by the large-system J-Divergence, $J(p(\mathbf{z}|H_0), p(\mathbf{z}|H_1)) = N \times \left[\left(\frac{P_D}{P_F} + \frac{P_F}{P_D} \right) - 2 \right]$, $J(p(\tilde{\mathbf{z}}|H_0), p(\tilde{\mathbf{z}}|H_1)) = N \times \left[\left(\frac{P_D\alpha_F}{P_F\alpha_D} + \frac{P_F\alpha_D}{P_D\alpha_F} \right) - 2 \right]$, which represents a lower-bound for the system error probability [86], thus enforcing a zero Bayesian error exponent. Differently, the Neyman-Pearson error exponent is given by $\lim_{K \rightarrow +\infty} -\frac{\ln[1 - P_{D_0}(\bar{\gamma}, K)]}{K}$, under $P_{F_0}(\bar{\gamma}, K) \leq \alpha$. If we choose $\bar{\gamma}_\alpha$ such that $P_{F_0}^*(\bar{\gamma}_\alpha) = \alpha$, then $\lim_{K \rightarrow +\infty} -\ln[1 - P_{D_0}(\bar{\gamma}_\alpha, K)] = -\ln[1 - P_{D_0}^*(\bar{\gamma}_\alpha)] < +\infty$, giving again a zero error exponent.

Note that the performance in TPC scenarios differ through the ratio $(\alpha_F/\alpha_D) < 1$ (cfr. Eqs. (5.3.8) and (5.3.9) with Eqs. (5.3.10) and (5.3.11)) which represents the *performance reduction factor* w.r.t. IPC scenarios. Note that α_F/α_D : (i) is an increasing function of the ratio $\frac{\sigma_h^2}{\sigma_w^2}$ (i.e. the received SNR), with limiting value equal to one; (ii) is a decreasing function of N , meaning a diverging separation in performance between IPC and TPC as N increases.

The diversity affects in a significant way the large-system probabilities of detection and false alarm, under an IPC, by shifting the Receiver Operating Characteristic (ROC) toward the upper-left corner, as shown in Fig. 5.3.1, meaning a performance improvement. Differently, it can be seen how a different effect is present in the TPC case, where an increase of N does not always coincide with performance improvement, but rather an optimal N , depending

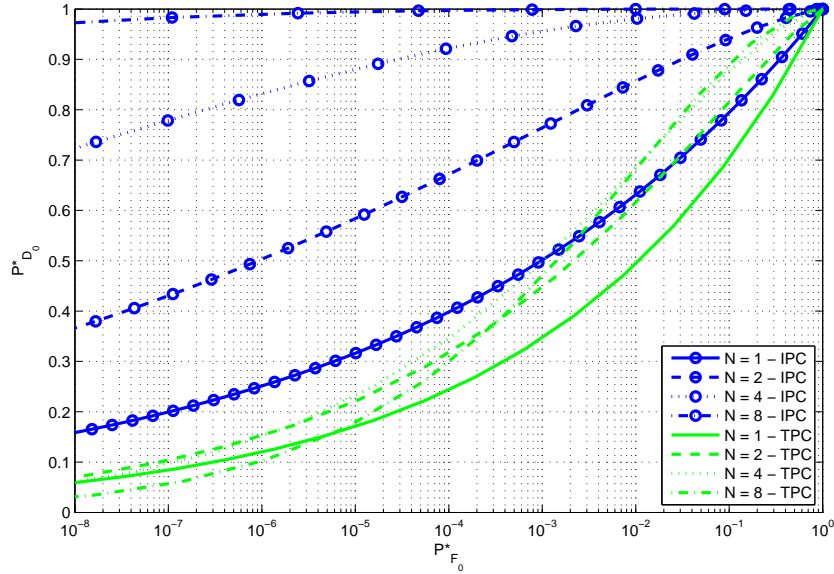


Figure 5.3.1: Effect of diversity N on large-system ROC $(P_{D_0}^*, P_{F_0}^*)$ under both IPC and TPC; WSN with sensor characteristics $(P_{D,k}, P_{F,k}) = (P_D, P_F) = (0.5, 0.05)$; $(\sigma_h^2/\sigma_w^2)_{dB} = 15$.

on $(P_D, P_F, \frac{\sigma_h^2}{\sigma_w^2})$, exists. This effect was already noticed in [45] and it is due to non-coherent combining loss of branch contributions.

Finally, in Figs. 5.3.2 and 5.3.3 we verify, through simulations, the convergence of the ROC to the large system expression ($K \rightarrow +\infty$) given by Eqs. (5.3.8) and (5.3.9) (resp. Eqs. (5.3.10) and (5.3.11)), under IPC (resp. TPC). It is apparent that the convergence under the TPC is faster w.r.t. the IPC case, because in both cases the large system ROC expressions rely on the Gaussian approximation of the Gaussian mixture given by Eq. (5.2.3). For such a reason, for a given K , imposing a TPC on the WSN assures a better matching w.r.t. to the IPC case, since all the components of the mixture will be more concentrated.

5.4 Conclusions

In this chapter we showed the optimality of the received-energy test for decision fusion performed over a non-coherent diversity MAC with conditionally i.n.i.d.

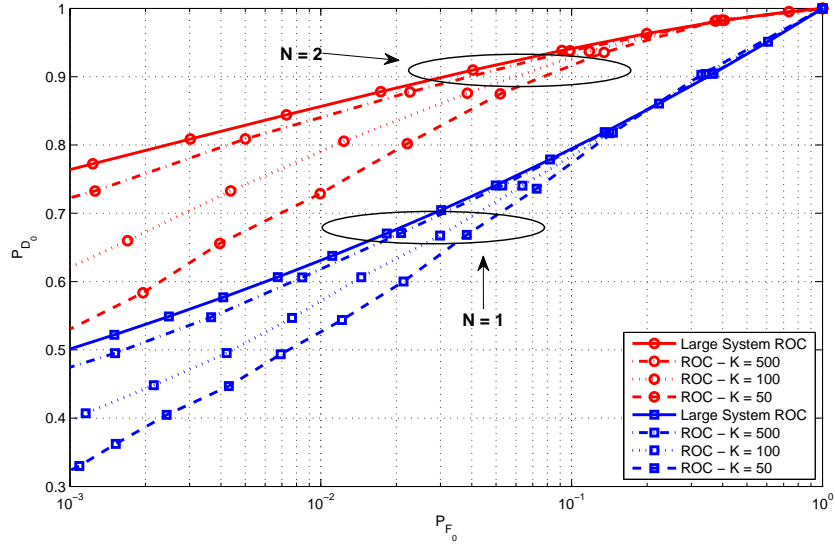


Figure 5.3.2: ROC comparison: large system vs finite number of sensors ($K \in \{50, 100, 500\}$) under IPC. $N \in \{1, 2\}$, $(\sigma_h^2/\sigma_w^2)_{dB} = 15$.

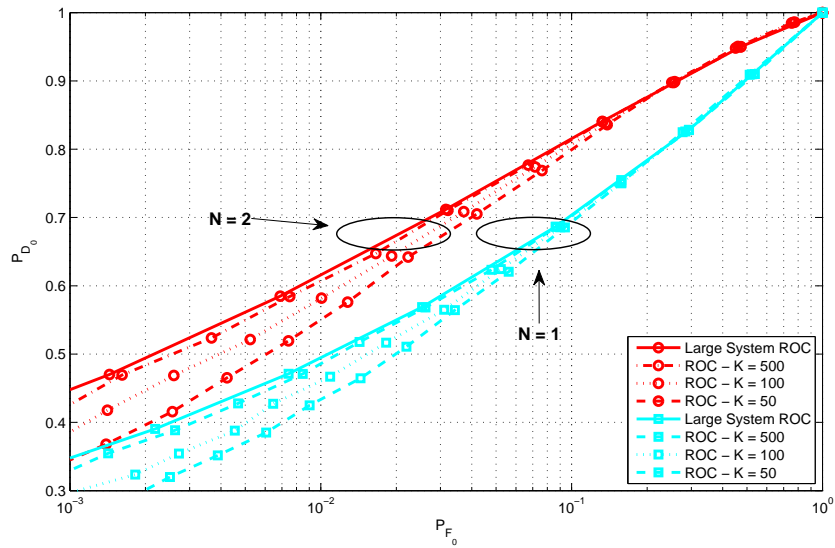


Figure 5.3.3: ROC comparison: large system vs finite number of sensors ($K \in \{50, 100, 500\}$) under TPC. $N \in \{1, 2\}$, $(\sigma_h^2/\sigma_w^2)_{dB} = 15$.

sensor decisions. We derived a sufficient condition on the LLR of the number of active sensors which can be applied to test the received-energy optimality in WSN with conditionally correlated sensor decisions. Finally, we showed, through analytical results, how the diversity in a WSN with conditionally i.i.d sensor decisions affects the large-system performance under both IPC and TPC.

5.5 Appendix

5.5.1 Proof of Proposition 1

We prove in this section that Eq. (5.3.2) is a sufficient condition for the optimality of ψ test. From Eq. (5.2.3), looking at the LLR as a function of ψ , we get:

$$\begin{aligned} \frac{\partial \Lambda_{opt}(\psi)}{\partial \psi} &= \frac{1}{\alpha(\psi)} \left[\sum_{\ell_1=0}^K \frac{\partial g(\psi, \ell_1)}{\partial \psi} P(\ell_1|H_1) \times \sum_{\ell_2=0}^K g(\psi, \ell_2) P(\ell_2|H_0) \right] \\ &\quad - \frac{1}{\alpha(\psi)} \left[\sum_{\ell_2=0}^K \frac{\partial g(\psi, \ell_2)}{\partial \psi} P(\ell_2|H_0) \times \sum_{\ell_1=0}^K g(\psi, \ell_1) P(\ell_1|H_1) \right] \end{aligned} \quad (5.5.1)$$

where we denoted $g(\psi, \ell) \triangleq \frac{1}{(\sigma_w^2 + \ell \sigma_h^2)^N} \exp\left(-\frac{\psi}{\sigma_w^2 + \ell \sigma_h^2}\right)$ and $\alpha(\psi)$ indicates a positive function of ψ (i.e. $\alpha(\psi) > 0, \forall \psi \in \mathbb{R}^+$). Strictly increasing property of LLR is guaranteed if $\frac{\partial \Lambda_{opt}(\psi)}{\partial \psi} > 0, \forall \psi \in \mathbb{R}^+$, thus manipulations from Eq. (5.5.1) lead to

$$\sum_{\ell_1=0}^K \sum_{\ell_2=0}^{\ell_1-1} k(\ell_1, \ell_2) \times \left[\frac{\partial g(\psi, \ell_1)}{\partial \psi} g(\psi, \ell_2) - \frac{\partial g(\psi, \ell_2)}{\partial \psi} g(\psi, \ell_1) \right] > 0 \quad (5.5.2)$$

where $k(\ell_1, \ell_2) \triangleq [P(\ell_1|H_1)P(\ell_2|H_0) - P(\ell_2|H_1)P(\ell_1|H_0)]$. In deriving Eq. (5.5.2) we could express the double sums in Eq. (5.5.1) as a function only of the indices $\ell_1 > \ell_2$, since the term in bracket in Eq. (5.5.2) equals to zero when $\ell_1 = \ell_2$. Noting that $\frac{\partial g(\psi, \ell)}{\partial \psi} = -\frac{1}{(\sigma_w^2 + \ell \sigma_h^2)} g(\psi, \ell)$ the condition is rewritten as

$$\sum_{\ell_1=0}^K \sum_{\ell_2=0}^{\ell_1-1} k(\ell_1, \ell_2) g(\psi, \ell_1) g(\psi, \ell_2) \left[\frac{\sigma_h^2(\ell_1 - \ell_2)}{(\sigma_w^2 + \ell_1 \sigma_h^2)(\sigma_w^2 + \ell_2 \sigma_h^2)} \right] > 0 \quad (5.5.3)$$

Since both $g(\psi, \ell)$ and the term in square brackets are positive (note that indices in the sums are such that $\ell_1 > \ell_2$), the term $k(\ell_1, \ell_2)$ is responsible for the sign of each term in the sum. Then a *sufficient condition* for Eq. (5.5.3) is obtained assuming that each of those terms is positive. This is achieved if the following property holds

$$k(\ell_1, \ell_2) > 0, \quad \ell_1 > \ell_2. \quad (5.5.4)$$

It is easy to demonstrate that the condition $k(\ell, \ell - 1) > 0$, $\ell \in \mathcal{L} \setminus \{0\}$, representing the strictly increasing property of ℓ -LLR, i.e. $\frac{P(\ell|H_1)}{P(\ell|H_0)} > \frac{P(\ell-1|H_1)}{P(\ell-1|H_0)}$, is equivalent to Eq. (5.5.4). In fact Eq. (5.5.4) implies that ℓ -LLR is strictly increasing; this is verified just substituting $\ell_2 = \ell_1 - 1$. Differently, we can show that ℓ -LLR strictly increasing property implies Eq. (5.5.4) by constructing the chain of inequalities $\frac{P(\ell_1|H_1)}{P(\ell_1|H_0)} > \frac{P(\ell_1-1|H_1)}{P(\ell_1-1|H_0)} > \dots > \frac{P(\ell_2|H_1)}{P(\ell_2|H_0)}$, all deriving from ℓ -LLR strictly increasing property.

5.5.2 Proof of Theorem 11

We prove the strictly increasing property of ℓ -LLR by induction. Let us assume there exists a set of $(t - 1)$ sensors with local performances $(P_{D,k}, P_{F,k})$, $k \in \{1, \dots, t - 1\}$. The number of active sensors in this case is denoted $\ell_{t-1} \triangleq \sum_{k=1}^{t-1} x_k$, $\ell_{t-1} \in \mathcal{L}_{t-1}$. We denote the probability of ℓ active sensors *out-of-* $(t - 1)$, given H_i , as $P_{t-1}(\ell|H_i)$, $H_i \in \mathcal{H}$ and the corresponding ℓ -LLR as $\lambda_{t-1}(\ell)$.

Initialization: the strictly increasing property of ℓ -LLR in single sensor case $\lambda_1(\ell_1) > \lambda_1(\ell_1 - 1)$, $\ell_1 \in \mathcal{L}_1 \setminus \{0\}$, is straightly verified when $P_{D,1} > P_{F,1}$.

Induction: Let us assume that for a specific configuration of $(t-1)$ sensors the ℓ -LLR $\lambda_{t-1}(\ell_{t-1})$ satisfies the strictly increasing property, that is $\lambda_{t-1}(\ell_{t-1}) > \lambda_{t-1}(\ell_{t-1} - 1)$, $\ell_{t-1} \in \mathcal{L}_{t-1} \setminus \{0\}$. If we add the t th sensor satisfying $P_{D,t} > P_{F,t}$ and we prove that the new ℓ -LLR $\lambda_t(\ell_t) > \lambda_t(\ell_t - 1)$, $\ell_t \in \mathcal{L}_t \setminus \{0\}$, i.e. it retains strictly increasing property, then the proof is complete.

To proceed let us first define $a(\ell) \triangleq P_{t-1}(\ell|H_1)$, $b(\ell) \triangleq P_{t-1}(\ell|H_0)$, $c(\ell) \triangleq P_1(\ell|H_1)$ and $d(\ell) \triangleq P_1(\ell|H_0)$.

The number of sensors transmitting when the t th sensor is added is then given by $\ell_t = \sum_{k=1}^t x_k = \ell_{t-1} + x_t$. The pmfs $P_t(\ell_t|H_0)$ and $P_t(\ell_t|H_1)$ are then given by [87]

$$P_t(\ell_t|H_0) = (b * d)(\ell_t) \quad P_t(\ell_t|H_1) = (a * c)(\ell_t) \quad (5.5.5)$$

The LLR strictly increasing condition is then formulated as follows

$$\exp[\lambda_t(\ell_t)] = \frac{(a * c)(\ell_t)}{(b * d)(\ell_t)} > \frac{(a * c)(\ell_t - 1)}{(b * d)(\ell_t - 1)} = \exp[\lambda_t(\ell_t - 1)] \quad (5.5.6)$$

By exploiting the support set of $c(\ell)$ and $d(\ell)$ we can rewrite Eq. (5.5.6) as follows

$$\frac{\sum_{k \in \{0,1\}} c(k)a(\ell_t - k)}{\sum_{k \in \{0,1\}} d(k)b(\ell_t - k)} > \frac{\sum_{k \in \{0,1\}} c(k)a(\ell_t - 1 - k)}{\sum_{k \in \{0,1\}} d(k)b(\ell_t - 1 - k)} \quad (5.5.7)$$

where obviously $a(t) = b(t) = 0$. Exploiting $c(0) + c(1) = (1 - P_{D,t}) + P_{D,t} = 1$, $d(0) + d(1) = (1 - P_{F,t}) + P_{F,t} = 1$, we obtain

$$\frac{[1 - c(1)]a(\ell_t) + c(1)a(\ell_t - 1)}{[1 - d(1)]b(\ell_t) + d(1)b(\ell_t - 1)} > \frac{[1 - c(1)]a(\ell_t - 1) + c(1)a(\ell_t - 2)}{[1 - d(1)]b(\ell_t - 1) + d(1)b(\ell_t - 2)} \quad (5.5.8)$$

The condition expressed in Eq. (5.5.8) can be rewritten as:

$$\begin{aligned} & \{[1 - c(1)][1 - d(1)][a(\ell_t)b(\ell_t - 1) - a(\ell_t - 1)b(\ell_t)]\} + \\ & \quad \{c(1)d(1)[a(\ell_t - 1)b(\ell_t - 2) - a(\ell_t - 2)b(\ell_t - 1)]\} + \\ & \quad + \{c(1)[1 - d(1)][a(\ell_t - 1)b(\ell_t - 1) - a(\ell_t - 2)b(\ell_t)]\} \\ & \quad - \{[1 - c(1)]d(1)[a(\ell_t - 1)b(\ell_t - 1) - a(\ell_t)b(\ell_t - 2)]\} > 0 \quad (5.5.9) \end{aligned}$$

Since $\frac{a(\ell_t)}{b(\ell_t)} > \frac{a(\ell_t - 1)}{b(\ell_t - 1)} > \frac{a(\ell_t - 2)}{b(\ell_t - 2)}$, we have that:

$$[a(\ell_t)b(\ell_t - 1) - a(\ell_t - 1)b(\ell_t)] > 0 \quad (5.5.10)$$

$$[a(\ell_t - 1)b(\ell_t - 2) - a(\ell_t - 2)b(\ell_t - 1)] > 0 \quad (5.5.11)$$

$$[a(\ell_t)b(\ell_t - 2) - a(\ell_t - 2)b(\ell_t)] > 0 \quad (5.5.12)$$

The condition in Eq. (5.5.9) is satisfied since positivity of the first two terms follows from the inequalities in Eq. (5.5.10), and the difference of the third and fourth terms in Eq. (5.5.9) is positive because $c(1)[1 - d(1)] > [1 - c(1)]d(1)$ (since $P_{D,t} > P_{F,t}$) and exploiting the inequality in Eq. (5.5.12). This concludes the proof.

5.5.3 Proof of Theorem 12

The proof follows in the first part similar steps as in [82]; for this reason we will only sketch it and underline the substantial differences. We use here the *characteristic function* of the vector $\mathbf{z}|H_i$, $i \in \{1, 2\}$, denoted as $\Phi_{\mathbf{z}|H_i}(\mathbf{t})$, to easily evaluate the limit for $K \rightarrow +\infty$. We then use this result, in conjunction with *Levy's Continuity Theorem* [87] to demonstrate the convergence in distribution of large-system $p(\mathbf{z}|H_i)$. Let us now write the characteristic function of $\mathbf{z}|H_0$ as a function of $\mathbf{t} = (t_1^t, t_2^t)^t$:

$$\Phi_{\mathbf{z}|H_0}(\mathbf{t}) = \mathbb{E}_{\mathbf{z}|H_0} \{ \exp(j\mathbf{t}_1^t \mathbf{z}_1 + j\mathbf{t}_2^t \mathbf{z}_2) \} \quad (5.5.13)$$

$$= \int \cdots \int \exp(j\mathbf{t}_1^t \mathbf{z}_1 + j\mathbf{t}_2^t \mathbf{z}_2) \times \sum_{\ell=0}^K p(\mathbf{z}_1, \mathbf{z}_2|\ell) P(\ell|H_0) d\mathbf{z}_1 d\mathbf{z}_2 \quad (5.5.14)$$

where $\mathbf{z}_1 \triangleq \Re\{\mathbf{z}\}$ and $\mathbf{z}_2 \triangleq \Im\{\mathbf{z}\}$ (with \mathbf{t}_i , $i \in \{1, 2\}$, representing the index-corresponding dual vectors over Fourier domain). Following analogous steps as in [82], exploiting: *i*) conditional independence assumptions such as $p(\mathbf{z}_1, \mathbf{z}_2|\ell) = p(\mathbf{z}_1|\ell)p(\mathbf{z}_2|\ell)$ and $p(\mathbf{z}_i|\ell) = \prod_{s=1}^N p(z_{i,s}|\ell)$, $i \in \{1, 2\}$; *ii*) the characteristic func-

tion of $x \sim \mathcal{N}(0, \sigma^2)$ is given by $\Phi_x(t) = \exp(-\frac{\sigma^2 t^2}{2})$ [87]; we get

$$\begin{aligned} \Phi_{\mathbf{z}|H_0}(\mathbf{t}) &= \sum_{\ell=0}^K P(\ell|H_0) \times \exp \left[-\frac{1}{4} \sum_{s=1}^N (t_{1,s}^2 + t_{2,s}^2) \left(\frac{\ell}{NK P_F} + \frac{\sigma_w^2}{\sigma_h^2 K P_F} \right) \right] \\ &= \exp \left[-\frac{1}{4} \sum_{s=1}^N (t_{1,s}^2 + t_{2,s}^2) \frac{\sigma_w^2}{\sigma_h^2 K P_F} \right] \times \end{aligned} \quad (5.5.15)$$

$$\left\{ P_F \exp \left[-\frac{1}{4} \sum_{s=1}^N \frac{(t_{1,s}^2 + t_{2,s}^2)}{NK P_F} \right] + (1 - P_F) \right\}^K \quad (5.5.16)$$

where in the last line we exploited $\ell|H_0 \sim \mathcal{B}(K, P_F)$. Also, exploiting similar noteworthy limits as in [82], eventually we have that $\lim_{K \rightarrow +\infty} \Phi_{\mathbf{z}|H_0}(\mathbf{t}) = \exp \left[-\frac{1}{2} \sum_{s=1}^N \frac{(t_{1,s}^2 + t_{2,s}^2)}{2N} \right]$. Applying the Continuity Theorem [87], we obtain $\mathbf{z}|H_0 \xrightarrow{d} \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \frac{1}{N} \mathbf{I}_N)$. In a similar way it can be shown that $\mathbf{z}|H_1 \xrightarrow{d} \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \frac{P_D}{P_F N} \mathbf{I}_N)$.

The last part consists in proving Eqs. (5.3.8) and (5.3.9). The large-system probabilities of false alarm and detection can be expressed in the equivalent form:

$$P_{F_0}^*(\bar{\gamma}) = P(\|\mathbf{z}\|^2 \geq \bar{\gamma}|H_0) = P\left(\frac{1}{2N}\xi \geq \bar{\gamma}|H_0\right) \quad (5.5.17)$$

$$P_{D_0}^*(\bar{\gamma}) = P(\|\mathbf{z}\|^2 \geq \bar{\gamma}|H_1) = P\left(\frac{P_D}{2P_F N}\xi \geq \bar{\gamma}|H_1\right) \quad (5.5.18)$$

where $\xi \sim \chi_{(2N)}^2$. The probabilities are then easily calculated evaluating the cumulative distribution function of ξ [87]:

$$P_{F_0}^*(\bar{\gamma}) = \int_{2\bar{\gamma}N}^{+\infty} p(\xi) d\xi \quad P_{D_0}^*(\bar{\gamma}) = \int_{2\bar{\gamma}N \frac{P_E}{P_D}}^{+\infty} p(\xi) d\xi \quad (5.5.19)$$

which provides the result.

Chapter 6

Decision Fusion for distributed classification of multiple targets with binary sensors

6.1 Introduction

We consider a surveillance task that consists in tracking and classifying (as depicted in Fig.6.1.1 for the binary classification case) N different targets moving into a given area of interest. One of multiple hypotheses is chosen for each target on the basis of a particular set of features. A wireless sensor network (WSN), deployed in the surveillance area, typically transmits hard decisions (to reduce the communication burden) on each target class via a wireless channel to a fusion center (FC) that forms a global situation assessment. This task is known in the literature as a distributed *decision fusion for multiple target classification* (multi-classification).

In [88] the problem has been studied with particular attention to the target-tracking task where, to simplify the distributed multi-classification algorithm, it is assumed that each sensor has only one target in its range at each time sample; thus a separate classification problem for each target is considered. This assumption is removed in [89] where all the sensor decisions depend on all the targets to be classified; the classification of each target is restricted to a binary hypothesis testing (as we will do throughout this chapter), which is equivalent to an M -ary hypothesis-testing problem with $M = 2^N$. Each sensor is designed to transmit N bits, where the n -th bit represents a local decision on the class of target n . Sub-optimal local decision rules for sensors have been derived and compared with the local optimum decision. The FC is simply chosen to be the joint optimum, as shown in [90], given each local decision

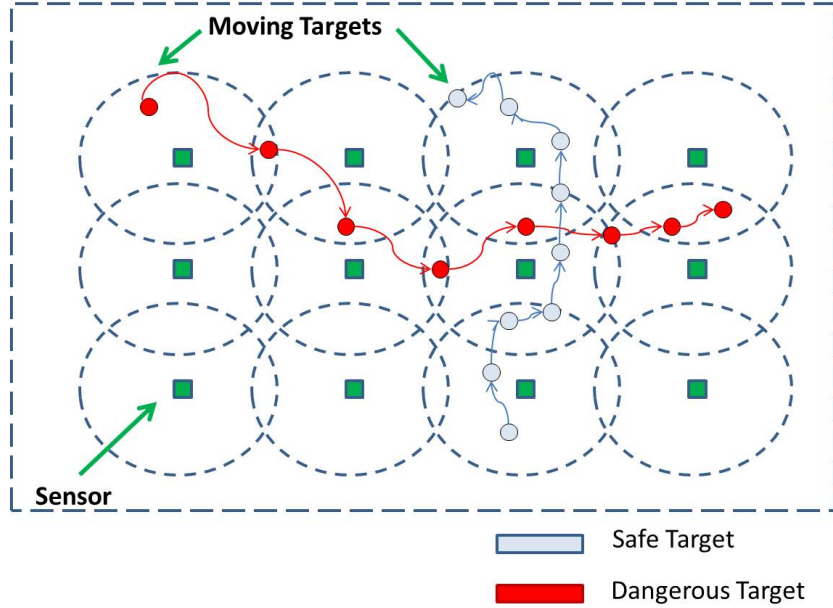


Figure 6.1.1: Graphical Description of the problem

rule. Alternative approaches have been presented for the more general M -ary hypothesis-testing problem in [91], where local decision rules and the decision fusion rule are designed as binary decision trees (BDT); the global decision at the FC is implemented as a sequence of $\log_2 M$ binary classifications, each of them representing the decision between two subsets obtained by the split of the hypothesis set being considered at each stage.

The approaches described for the binary multi-classification require N bits to be sent by each sensor. However, to further reduce the communication burden, it is often assumed that also in this context a sensor transmits only one bit, as in the classical binary classification task [67, 92]. Results on related problems are available in literature. In [93] it is demonstrated how a WSN with local binary decision rules could be effectively used in conjunction with the optimal FC (given these rules) to perform M -ary hypothesis testing and getting asymptotic detection of correct hypothesis, as some mild conditions on local transition probabilities are verified. Also some methods for local decision thresholds design, through a genetic algorithm optimization, are presented in [93]. In [94] a way of choosing the local binary decision rules is suggested and a sub-optimal fusion center that requires only the computation of binary operations instead of real numbers is presented. The complexity of this approach depends directly on the dimension M of the hypothesis space.

Note that many times the local binary decision rules are already completely designed with reference to a single target classification task. Furthermore the

hypothesis space grows exponentially with the number of targets ($M = 2^N$) and the optimal FC may become too complex even for small values of N . Therefore sub-optimal FC rules with reduced complexity are considered in this chapter under the assumption that the DF is aware of the information coming from the multi-target tracking system and the local decision model of the WSN. We assume in this work that the targets to be classified are moving, but we do not consider the tracking issues and the related uncertainties. In addition we assume an ideal noiseless link for bits transmitted from the sensors to the FC. The contributions of the chapter are the following:

- We propose two sub-optimal decision fusion algorithms (Range Limited Marginalization - RLM and the parallelized form PRLM) as reduced complexity alternatives to the joint Optimal Decision Fusion (ODF). The ideas behind these algorithms are: a) the exploitation of the concept that each sensor has a maximum detection range; b) the fusion center classifies each target separately by accounting approximately about the dependence of each local decision with respect to all the targets.
- We show, through several simulations, the mean speed-up factor (Mean Time Complexity Gain) and verify the guaranteed memory reduction (Space Complexity Gain) achieved by (P)RLM over ODF; we also analyze the consequent loss in terms of performances of these sub-optimal approaches (measured through generalized user-averaged Probabilities of False Alarm and Missed Detection) deriving from the approximations.

The remainder of the chapter is organized as follows. In Section 6.2 we introduce and formalize the system model. In Section 6.3 we recall the optimal centralized decision fusion and we propose our two alternative inference algorithms with reduced complexity. In Section 6.4 we compare the performances and the complexity of these algorithms with a set of simulations. Finally in Section 6.5 we discuss some concluding remarks and future developments.

6.2 System Model

6.2.1 The Overall System

As illustrated in Fig. 6.2.1 we consider N targets which move in a surveilled area (in this chapter we will analyze only targets moving in a closed box; issues about the entrance and the exit of targets are not considered here) and which are tracked on the basis of a multi-sensor multi-target tracking system that computes the position estimate $\hat{\mathbf{x}}_n[i]$ of each target $n = 1, \dots, N$ at time i . We assume perfect position estimation, that is $\hat{\mathbf{x}}_n[i] = \mathbf{x}_n[i]$. Each target n is associated to a class T_n , where $T_n = 0$ stands for a “safe target” and $T_n = 1$ for an “enemy” or “dangerous target” and the whole set $\mathcal{T} = \{T_1, \dots, T_N\} \in \{0, 1\}^N$ is estimated at time k on the basis of all the binary decisions (we consider only decisions as possible inputs from the WSN

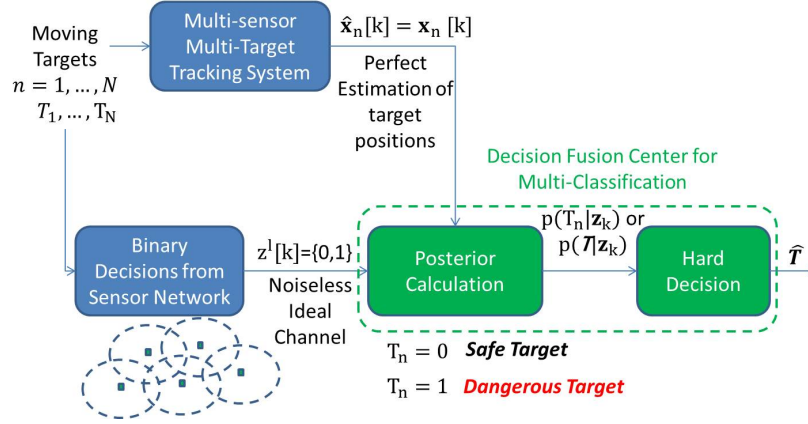


Figure 6.2.1: Block diagram of the considered system.

for our fusion task; for a detailed discussion on various potential fusion architectures refer to [95]) $\mathbf{z}_k = [\mathbf{z}[1]^T, \mathbf{z}[2]^T, \dots, \mathbf{z}[k]^T]^T$, where $\mathbf{z}[i]$, $i = 1, \dots, k$ represents the column vector of the decisions coming from the sensor network at time i . We assume the WSN to be composed by L sensors, which means that $\mathbf{z}[i] = [z^1[i], \dots, z^L[i]]^T$, with $z^\ell[i]$ representing the decision coming from sensor ℓ at time i . The sensors are deployed in the area in a static way and their absolute position \mathbf{x}^ℓ is known in advance. We assume ideal noiseless communication channel, that is the binary decisions $z^\ell[i]$ travel in a transparent way over the wireless channel. The decision $z^\ell[i] = \{0, 1\}$ has a clear meaning in the case of a single target classification, but it does not have a direct mapping to the multi-classification task. A decision fusion algorithm at time k may compute the posterior probability (or an approximation) of the binary hypotheses: (A) for each target given all the decisions observed till that time, that is $p(T_n|\mathbf{z}_k)$; (B) for the whole joint set $\mathcal{T} = \{T_1, \dots, T_N\}$, that is $p(\mathcal{T}|\mathbf{z}_k)$. From the posteriors hard decisions can be taken by the respective following rules:

$$(A) \quad \begin{cases} \hat{T}_n[k] = 1, & \frac{p(T_n=1|\mathbf{z}_k)}{p(T_n=0|\mathbf{z}_k)} \geq \Gamma, \\ \hat{T}_n[k] = 0, & \text{otherwise;} \end{cases} \quad (6.2.1)$$

$$(B) \quad \hat{\mathcal{T}}[k] = \arg \max_{\mathcal{T} \in \{0,1\}^N} p(\mathcal{T}|\mathbf{z}_k); \quad (6.2.2)$$

where Γ is a threshold to be chosen, that represents a free parameter in the system. Note that biases can be added to (B) to prioritize certain hypotheses, however the choice of these terms is not intuitive and so they cannot be used to perform a Neyman-Pearson Test on each target, as done with (A). The posterior probability computation is possible because we have exact knowledge of the sensor local decision rules, and all the mutual distances, $r_n^\ell[i] = \|\mathbf{x}_n[i] - \mathbf{x}^\ell\|$, $\forall n, \ell, i$ ($\|\cdot\|$ stands for the Euclidean distance).

6.2.2 WSN Local Decision Rule Model

Joint Decision Rule Model

In this chapter we assume that each decision $z^\ell[i]$ is conditionally (given the class of each target T_n , $n = 1, \dots, N$) independent from all the other decisions at a different time slot $z^c[j]$, $c = 1, \dots, L$, $j \neq i$ and from all the other decisions at the same time slot, but coming from the other sensors $z^m[i]$, $m \neq \ell$ (for a detailed discussion on the requirements needed to assure these independency conditions see [96]).

Also we assume, for simplicity of notation and without loss of generality, that each sensor decision follows the same probabilistic model. This means that the joint likelihood of all the decisions till time k is given by the product of the single likelihoods of each decision $z^\ell[i]$. Due to the space-temporal conditional independence assumption in fact we have:

$$\begin{aligned} p(\mathbf{z}_k|T_1, \dots, T_N) &= p(\mathbf{z}[0], \dots, \mathbf{z}[k]|T_1, \dots, T_N) \\ &= \prod_{i=0}^k \prod_{\ell=1}^L p(z^\ell[i]|T_1, \dots, T_N) \end{aligned} \quad (6.2.3)$$

In the next two sub-sections we will concentrate on the model of the single decision $z^\ell[i]$, i.e. the conditional likelihood $p(z^\ell[i]|T_1, \dots, T_N)$.

Local Decision Rule for Single Target

As pointed out in the introduction, we consider sensors which are designed and whose model is available for single-target classification only. More specifically we consider sensors with a binary outcome $z^\ell[i] = \{0, 1\}$, which in the case of the presence of a single target n , indicates a hard estimate of its target class T_n . The likelihoods of the generic sensor ℓ are given by the following typical expressions strictly related to binary hypothesis testing [67, 92, 97]:

$$p(z^\ell[i] = 1|T_n = 0) = p_{fa} \quad (6.2.4)$$

$$p(z^\ell[i] = 1|T_n = 1) = p_d(r_n^\ell[i]) \quad (6.2.5)$$

where p_{fa} represents the *Probability of False Alarm* of the sensor, given a safe target ($T_n = 0$) and $p_d(r_n^\ell[i])$ represents the *Probability of Detection* of the sensor, given an enemy target ($T_n = 1$). The $p_d(\cdot)$ is modeled as a decreasing function of $r_n^\ell[i] = \|\mathbf{x}_n[i] - \mathbf{x}^\ell\|$, which represents the *sensor-target distance* at time i . As we can see from Eqs. (6.2.4) and (6.2.5) the likelihoods for single-target classification are completely known, due to the fact that we have perfect position knowledge. More specifically $p_d(\cdot)$ is assumed to follow the Gaussian-law:

$$p_d(r_n^\ell[i]) = (1 - p_{fa})e^{-\frac{(r_n^\ell[i])^2}{\sigma^2}} + p_{fa} \quad (6.2.6)$$

which is illustrated in Fig. 6.2.2 for $\sigma = 2$ meters and $p_{fa} = 10^{-2}$.

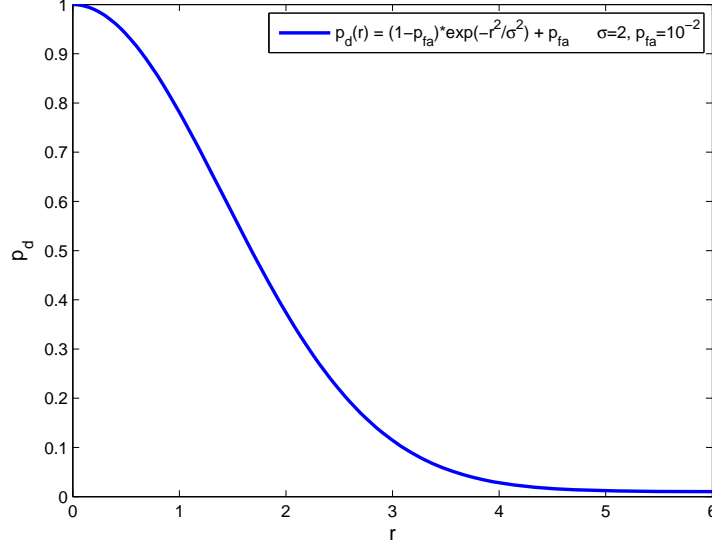


Figure 6.2.2: Probability of Detection as a function of sensor-target distance

Parameter σ is intended to give, in conjunction with p_{fa} , a gross and essential information on the sensor detection range. For example, if $p_{fa} = 10^{-2}$ and $\sigma = 2$ meters, it could be assumed that the *maximum detection range* is $r_{max} = 4$ meters, due to the fact that $p_d(r_n^\ell[i] = 4) \simeq 2 \cdot p_{fa}$, which implies a very low Hellinger distance between the two hypotheses [98], as further clarified in the Appendix A.

This type of sensor modeling has a broad analytical justification from the fact that it fits well the detection process under different conditions. However if there is a more detailed knowledge about the underlying process of detection, a more precise expression of p_d can be obtained, such as in [97]. However a different expression for p_d (as a function of $r_n^\ell[i]$) would not alter the generality of the approach presented in this chapter, thus for reasons of clarity in the following we will consider only the model specified by Eq. (6.2.6) to extend the classification to multiple targets.

Interpretation of Local Decision Rule for Multiple Targets

In the case of multiple targets in the area of the sensor, we need to generalize the model presented in the previous subsection, to give an interpretation of the decision $z^\ell[i]$ as an estimation of a subspace of the joint feature set \mathcal{T} . To do this we assume a signal (given by the coherent sum of “emissions” coming from each dangerous target) plus noise model for multiple target classification. More specifically the received power $P_n^\ell[i]$ of target n from sensor ℓ at time i is given

by a square law attenuation model:

$$P_n^\ell[i] = \frac{P_0}{r_n^\ell[i]^2} \quad (6.2.7)$$

where P_0 is the “signal power” emitted by the target, which is assumed identical for all targets. If we substitute the Eq. (6.2.7) in the Eq. (6.2.6) we obtain p_d as a function of the received power $P_n^\ell[i]$ as:

$$p_d(P_n^\ell[i]) = (1 - p_{fa})e^{-\frac{(P_0/P_n^\ell[i])}{\sigma^2}} + p_{fa}. \quad (6.2.8)$$

It is assumed that in presence of more than one “enemy” targets the emission powers are trivially added, e.g. for the case of two enemy targets m and n ($T_m = T_n = 1$) at distances respectively $r_m^\ell[i]$ and $r_n^\ell[i]$ the received power at the sensor ℓ is given by:

$$P_{m+n}^\ell[i] = \frac{P_0}{r_m^\ell[i]^2} + \frac{P_0}{r_n^\ell[i]^2}. \quad (6.2.9)$$

The probability of detection of the sensor in presence of enemy targets $T_m = 1$ and $T_n = 1$ is obtained by substituting Eq. (6.2.9) in Eq. (6.2.8)

$$\begin{aligned} p_d(P_{m+n}^\ell[i]) &= (1 - p_{fa})e^{-\frac{P_0}{\sigma^2 P_{m+n}^\ell[i]}} + p_{fa} \\ &= (1 - p_{fa})e^{-\frac{r_m^\ell[i]^2 \parallel r_n^\ell[i]^2}{\sigma^2}} + p_{fa} \end{aligned} \quad (6.2.10)$$

where \parallel indicates the parallel operator between the squared distances $r_m^\ell[i]^2$ and $r_n^\ell[i]^2$ ($a \parallel b = \frac{ab}{a+b}$). This argument can be generalized by iterating the parallel rule for more than two enemy targets. So the likelihoods of in presence N targets with a joint feature set $\mathcal{T} = \{T_1, T_2, \dots, T_N\}$ is given by:

$$p(z^\ell[i] = 1 | T_1, \dots, T_N) = p_d((r_{\mathcal{T}_E}^\ell[i])^2) \quad (6.2.11)$$

where $(r_{\mathcal{T}_E}^\ell[i])^2$ needs to be interpreted as an equivalent squared distance obtained iterating the parallel on the squared distances of the targets that belong to \mathcal{T}_E , which is the subset of dangerous targets of \mathcal{T} . If the set \mathcal{T}_E is empty the probability value reduces to p_{fa} . It is worth noting that the set \mathcal{T}_E is assumed to be time-invariant (i.e. we assume that a target n cannot change its status T_n), however this is not true for the equivalent squared distance $r_{\mathcal{T}_E}^\ell[i]^2$, because the targets considered have time-varying distances with respect to the sensor ℓ . As it is evident from (6.2.11) the likelihood is now defined on a space of 2^N possible combinations and can be used to get an information on the estimated joint feature set.

However the likelihood of the sample at time i of each sensor ℓ $p(z^\ell[i] | T_1, \dots, T_N)$ from which the detection $z^\ell[i]$ is generated could be well approximated by:

$$p(z^\ell[i] | T_1, \dots, T_N) \simeq p(z^\ell[i] | \mathcal{T}^\ell[i]), \quad (6.2.12)$$

where $\mathcal{T}^\ell[i]$ denotes a subset of \mathcal{T} which is defined as

$$\mathcal{T}^\ell[i] = \{T_n \in \mathcal{T} : \|\mathbf{x}_n[i] - \mathbf{x}^\ell\| < r_{max}\}.$$

The set $\mathcal{T}^j[k]$ groups the targets which are inside the maximum detection range r_{max} of sensor ℓ at time i . The approximation is reasonable if we consider that the targets that are outside r_{max} don't contribute much to the density of Eq. (6.2.11) with their received power at sensor ℓ . It is also worth noting that Eq. (6.2.12) means that if $T_n \notin \mathcal{T}^\ell[i]$ the decision coming from sensor ℓ at time i does not represent a valuable information for classification of target n .

The approximation derived in Eq. (6.2.12) will be used in the following to derive alternative fusion rules for efficient posterior update in time-space (samples-sensors) dimensions.

6.3 Decision Fusion Algorithms

We define three metrics for the analysis of computational complexity of the Decision Fusion Algorithms [99]:

- The *Space Complexity* C_S of the decision fusion algorithm, which is the number of probabilities to be memorized by the decision fusion algorithm at the end of each time slot k ;
- The *Time Complexity* at time k $C_T[k]$, which is the number of elementary operations to be performed till time k to obtain the computation of posterior probabilities at time k ;
- The *Overall Complexity at time k* , that is $C_O[k] = C_S \cdot C_T[k]$.

6.3.1 Joint Target - Optimal Decision Fusion (ODF) Approach

The optimal decision fusion algorithm for multi-target classification requires computation of the Bayes recursive update as [89]:

$$\begin{aligned} p(T_1, \dots, T_N | \mathbf{z}_k) &\propto p(T_1, \dots, T_N | \mathbf{z}_{k-1}) p(\mathbf{z}[k] | T_1, \dots, T_N) \\ &= p(T_1, \dots, T_N | \mathbf{z}_{k-1}) \prod_{l=1}^L p(z^\ell[k] | T_1, \dots, T_N). \end{aligned} \quad (6.3.1)$$

It is clear from the formula that with the ODF approach, the algorithm needs to memorize all the joint posterior probabilities $p(T_1, \dots, T_N | \mathbf{z}_k)$ at each time k . At each time k (and also for each sensor l) 2^N likelihood probabilities $P(z^\ell[k] | T_1, \dots, T_N)$ need to be computed. So the Space Complexity C_s for ODF grows exponentially with the number of users, that is $C_s^{ODF} = O(2^N)$. For example the presence of 30 targets in the area of interest would lead to the necessity of storing 2^{30} posterior probabilities during the entire fusion process. In addition

the Time Complexity $C_T[k]$ is given by the 2^N update of posterior probabilities every time a new decision $z^\ell[i]$ arrives, which leads to $C_T^{ODF}[k] = O(k \cdot L \cdot 2^N)$. The Overall Complexity $C_O^{ODF}[k]$ is $C_O^{ODF}[k] = O(k \cdot L \cdot 2^{2N})$, which is not feasible when the number of targets to be classified is significant.

In addition one of the main drawbacks of the joint decision fusion is the fact that it is not easily compatible with the integration between fusion centers. Consider for example two adjacent areas covered by different sensor networks and controlled by different fusion centers. A target n that exits from the first area (controlled by the FC-1) and enters in the other one (controlled by the FC-2) implies the necessity of sending the acquired inference about the classification of target n from the FC-1 to the FC-2. But this information it is not directly available at fusion center, as it stores $P(T_1, \dots, T_n, \dots, T_N | \mathbf{z}_k)$. The relevant information to be sent to FC-2 is the posterior marginal $P(T_n | \mathbf{z}_k)$, which needs to be calculated by a marginalization [2] (which is a complex operation) from the joint posterior of FC-1. At the other end the FC-2 cannot use the marginal $p(T_n | \mathbf{z}_k)$ in a direct way, as it stores the joint posterior of the targets of its area, so that an operation of extension [2] has to be performed.

It is also worth remarking in this context that Eq. (6.2.12) cannot be used to simplify this decision fusion process, as the reduced joint likelihood of the decision $z^\ell[i]$ cannot be used to update directly the joint posterior, but has to be extended. We intend to propose reduced complexity fusion algorithms which update themselves in a separate way for each user directly on the respective posterior marginal $p(T_n | \mathbf{z}_k)$ and which are clearly sub-optimal in terms of performance related to error probability, but that can be implemented in practice.

6.3.2 Separated Target - Decision Fusion Approaches

To deal with complexity we start from the *exact* computation formula of the N marginal posteriors through the Bayes' update obtained by the conditioning chain rule [7]:

$$\begin{aligned} p(T_n | \mathbf{z}_k) &\propto p(T_n | \mathbf{z}_{k-1}) p(\mathbf{z}[k] | \mathbf{z}_{k-1}, T_n) \\ &= p(T_n | \mathbf{z}_{k-1}) \prod_{\ell=1}^L p(z^\ell[k] | \mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k], T_n); \end{aligned} \quad (6.3.2)$$

where $\mathbf{z}^{1:(\ell-1)}[k] = [z^1[k], \dots, z^{\ell-1}[k]]$. Eq. (6.3.2) can be equivalently written as a sum update for log-likelihood ratio $\Lambda_n[k] = \ln \left[\frac{P(T_n=1 | \mathbf{z}_k)}{P(T_n=0 | \mathbf{z}_k)} \right]$ of the target n :

$$\Lambda_n[k] = \Lambda_n[k-1] + \sum_{\ell=1}^L \ln \left[\frac{P(z^\ell[k] | \mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}, T_n = 1)}{P(z^\ell[k] | \mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}, T_n = 0)} \right]. \quad (6.3.3)$$

The conditioned likelihoods $p(z^\ell[k]|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k], T_n)$ are obtained by the following marginalization:

$$p(z^\ell[k]|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k], T_n) = \sum_{\mathcal{T}-T_n} p(z^\ell[k]|\mathcal{T})p(\mathcal{T} \setminus T_n|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k], T_n); \quad (6.3.4)$$

where $p(z^\ell[k]|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k], \mathcal{T}) = p(z^\ell[k]|\mathcal{T})$, due to space-time independency. Eq. (6.3.4) requires for its computation the additional knowledge of $p(\mathcal{T}|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])$, as $p(\mathcal{T} \setminus T_n|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k], T_n) = p(\mathcal{T}|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])/p(T_n|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])$. This is clearly not affordable as requires the additional computational of the joint posterior, which we don't want to compute. In the following subsections we make additional assumption to simplify the computation and get a complexity-tractable sub-optimal Bayes update.

(Parallel) Range Limited Marginalization (P)RLM

If we start from the conditional distribution of Eq. (6.3.4) we can use the following approximations

$$\begin{aligned} & \sum_{\mathcal{T}-T_n} p(z^\ell[k]|\mathcal{T}) \frac{p(\mathcal{T}|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])}{p(T_n|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])} \\ & \simeq \sum_{\mathcal{T}-T_n} p(z^\ell[k]|\mathcal{T}) \frac{\prod_{s=1}^N p(T_s|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])}{p(T_n|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])} \\ & = \sum_{\mathcal{T}-T_n} p(z^\ell[k]|\mathcal{T}) \prod_{s=1, s \neq n}^N p(T_s|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k]) \\ & \simeq \sum_{\mathcal{T}^\ell[k]-T_n} p(z^\ell[k]|\mathcal{T}^\ell[k]) \prod_{s \in \mathcal{T}^\ell[k], s \neq n}^N p(T_s|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k]), \quad (6.3.5) \end{aligned}$$

where in the first line we have used the definition of conditional probability $p(A|B) = P(A, B)/P(B)$, in the second line we have approximated the joint posterior $p(\mathcal{T}|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])$ as a product of its posterior marginals $\prod_{s=1}^N p(T_s|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k])$ and in the final line we used the simplification given by Eq. (6.2.12); we call this algorithm *Range Limited Marginalization* (RLM). The approximation of the second line is motivated by the fact that, given a joint pdf of r.v. A, B and C , $p(A, B, C)$, the product of its marginals $p(A) \cdot p(B) \cdot p(C)$ is the product of pdfs which is the nearest in Kullback-Leibler (KL) Divergence sense to $p(A, B, C)$ [7].

If we adopt the further approximation $p(T_s|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k]) \simeq p(T_s|\mathbf{z}_{k-1})$,

we obtain the expression:

$$p(z^\ell[k]|\mathbf{z}_{k-1}, \mathbf{z}^{1:(\ell-1)}[k], T_n) \simeq \sum_{\mathcal{T}^\ell[k]-T_n} p(z^\ell[k]|\mathcal{T}^\ell[k]) \prod_{s \in \mathcal{T}^\ell[k], s \neq n}^N p(T_s|\mathbf{z}_{k-1}); \quad (6.3.6)$$

which implies that each conditional distribution for each sensor could be implemented in a parallel way. We call this alternative algorithm *Parallel Range Limited Marginalization* (PRLM).

It is worth remarking that the algorithms now derived, unlike the joint ODF, are perfectly compatible with respect to the integration of more FC to cover different zones of the surveilled area, because the approximated posterior marginals $p_{(P)RLM}(T_n|\mathbf{z}_k)$ computed by (P)RLM already represent the direct information to be transmitted in the case of a target switching to a different FC-controlled zone. This represents a more flexible network architecture.

The two algorithms have the same complexity metrics. The Space Complexity for (P)RLM is obviously given by $C_s^{(P)RLM} = O(N)$, as only the posterior marginals $p(T_n|\mathbf{z}_k)$ needs to be memorized (except some temporary cells needed to compute the distribution reported in Eq. (6.3.5) in the case of RLM and to compute the distribution reported in Eq. (6.3.6) to compute PRLM).

The calculus of Time Complexity is rather tedious and it is reported in the Appendix B. Complexity analysis demonstrates that RLM and PRLM have $C_T^{(P)RLM}[k] = O(\sum_{i=1}^k \sum_{\ell=1}^L |T^\ell[i]|^2 \cdot 2^{|T^\ell[i]|})$, where $|T^\ell[i]|$ denotes the cardinality of the set $T^\ell[i]$, which in the worst case leads to $O(k \cdot L \cdot N^2 \cdot 2^N)$, which is worse than Time Complexity of the ODF. However the cardinality of the sets $T^\ell[i]$ is typically far lower than N and so there is a substantially Time Complexity reduction which is verified through a mean analysis presented in the simulation results of Section 4. The Overall Complexity instead is given by $C_O^{(P)RLM}[k] = O(\sum_{i=1}^k \sum_{\ell=1}^L T^\ell[i]^2 \cdot 2^{T^\ell[i]} \cdot N)$, which in the worst case leads to $O(k \cdot L \cdot N^3 \cdot 2^N)$.

6.4 Simulations

The setting considered in this section is a surveillance task in a square area of $[15m, 15m]$ with a network of $L = 16$ binary sensors displaced by the following grid rule:

$$\mathbf{x}^l = \begin{bmatrix} x^l \\ y^l \end{bmatrix} = \begin{bmatrix} \Delta \cdot i \\ \Delta \cdot j \end{bmatrix} \quad i, j = 1, \dots, 4;$$

where $\Delta = 3$ meters represents the grid spacing. We consider in this case sensors with the same performance characteristics, with $\sigma = 2$ meters and $p_{fa} = 10^{-2}$. As discussed in Section 2, we can assume that, given these parameters, the maximum detection range is $r_{max} = 4$ meters. The trajectories of the targets moving in the defined box are generated by applying the acceptance-rejection

method to a first order AR classical motion model [100]. The performance and the complexity metrics analyzed are obtained by averaging them over 10^4 different realizations of trajectories configurations, each one of them associated to a different configuration map of T_n , $n = 1, \dots, N$, generated by an equi-probable distribution on the whole joint set \mathcal{T} . The performance metrics considered for the comparison are the *User Averaged - System Probability of Missed Detection* $\bar{p}_{md}[k]$ and the *User Averaged - System Probability of false alarm* $\bar{p}_{fa}[k]$ which at time k are given by:

$$\bar{p}_{md}[k] = \frac{\sum_{n=1}^N p_{md,n}[k]}{N}, \quad (6.4.1)$$

$$\bar{p}_{fa}[k] = \frac{\sum_{n=1}^N p_{fa,n}[k]}{N}, \quad (6.4.2)$$

where $p_{md,n}[k]$ and $p_{fa,n}[k]$ represent the probabilities of missed detection and false alarm of the target n till time k , i.e. $p_{md,n}[k] = p(\hat{T}_n[k] = 0 | T_n = 1)$ and $p_{fa,n}[k] = p(\hat{T}_n[k] = 1 | T_n = 0)$. The couple $(\bar{p}_{md}[k], \bar{p}_{fa}[k])$ represents a synthetic performance information on how the decision fusion center is capable of correctly classifying a generic target n moving in the surveilled area as a function of the time k . Complexity metrics used for comparison are based on the expressions obtained for $C_T[k]$ and $C_O[k]$ for ODF and (P)RLM:

$$g_T[k] = \frac{C_T^{ODF}[k]}{E[C_T^{(P)RLM}[k]]}, \quad (6.4.3)$$

$$g_O[k] = \frac{C_O^{ODF}[k]}{E[C_O^{(P)RLM}[k]]}, \quad (6.4.4)$$

where $g_T[k]$ and $g_O[k]$ are respectively the *Mean Time Complexity Gain* (essentially a speed-up factor) and the *Mean Overall Complexity Gain* of the (P)RLM over ODF. We consider in this chapter three different target population scenarios, exactly the case with $N = 3, 5, 7$ targets. The ODF is compared with (P)RLM for different values of the threshold $\gamma = \ln \Gamma$ on the basis of the performance and complexity metrics introduced.

From the analysis of Figs. (6.4.1,6.4.2,6.4.3,6.4.4,6.4.5,6.4.6) we can see substantially that in the considered scenario the performances of RLM and PRLM are essentially the same for all the three populations configurations. However in the case of a widely deployed wireless sensor network it is expected that the performances differ in a significant way, because when the number of sensors increases approximation of Eq. (6.3.6) becomes less tight.

The comparison between ODF and (P)RLM is very interesting. From the Figs. (6.4.1,6.4.3,6.4.5) it can be seen how RLM-PRLM with $\gamma = 0$ presents a loss of performance in terms of $\bar{p}_{fa}[k]$, which presents a weak negative error exponent. However by comparing the Figures (6.4.2,6.4.4,6.4.6) we can see how (P)RLM approaches the same $\bar{p}_{md}[k]$ as the ODF.

If we raise the value of γ in an accurate way we can obtain a lower $\bar{p}_{fa}[k]$; the degradation in terms of $\bar{p}_{md}[k]$ is fairly acceptable as a change of γ does

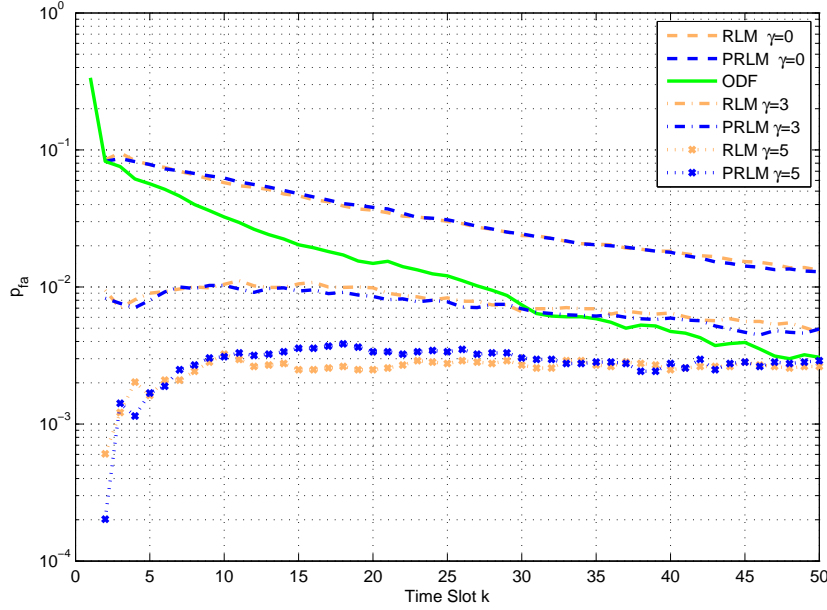


Figure 6.4.1: Averaged User System Probability of False Alarm $\bar{p}_{fa}[k]$ for ODF, RLM and PRLM. $N = 3$ targets, $k = 1, \dots, 50$.

not alter substantially the error exponent of the $\bar{p}_{md}[k]$ for the RLM/PRLM approach, which means that we can wait for a number of time slots to obtain the same $\bar{p}_{md}[k]$ of the ODF.

Regarding the complexity we refer to Fig. 6.4.7 where we report the Time and Overall Complexity Gain of the (P)RLM over ODF $g_T[k]$ and $g_O[k]$ as functions of time k and for the three different population scenarios considered, i.e. $N = 3, 5, 7$.

By analyzing the Fig. 6.4.7 we can see how $g_T[k]$ and $g_O[k]$ are time-stationary, that is $g_T[k] = g_T$ and $g_O[k] = g_O$ which means that the (P)RLM gives a reduction of complexity which is independent of the observation time being considered. Furthermore by considering g_T even only in the case of $N = 3$

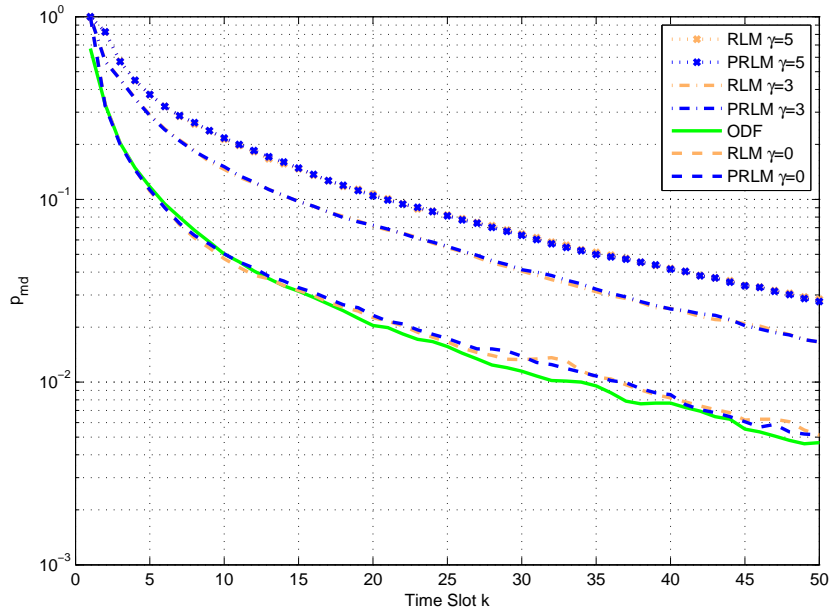


Figure 6.4.2: Averaged User System Probability of False Alarm $\bar{p}_{md}[k]$ for ODF, RLM and PRLM. $N = 3$ targets, $k = 1, \dots, 50$.

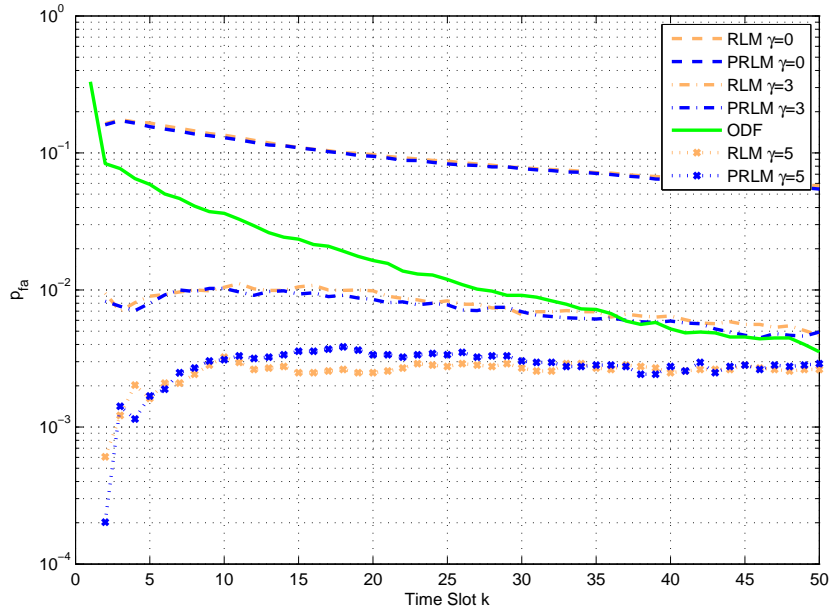


Figure 6.4.3: Averaged User System Probability of False Alarm $\bar{p}_{fa}[k]$ for ODF, RLM and PRLM. $N = 5$ targets, $k = 1, \dots, 50$.

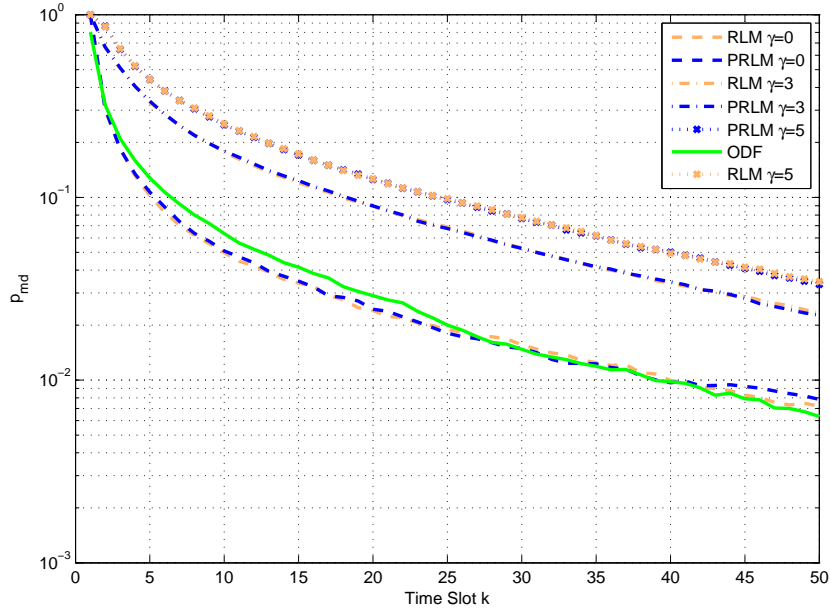


Figure 6.4.4: Averaged User System Probability of False Alarm $\bar{p}_{md}[k]$ for ODF, RLM and PRLM. $N = 5$ targets, $k = 1, \dots, 50$.

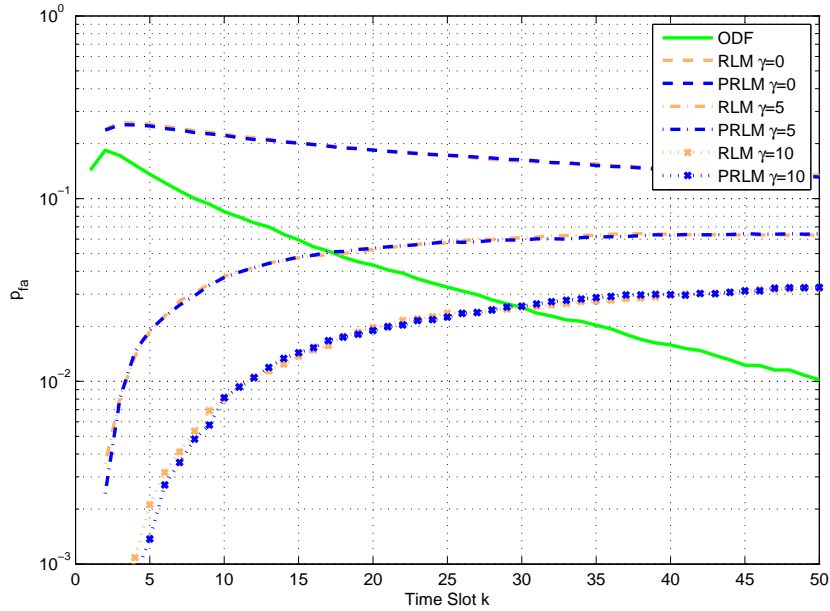


Figure 6.4.5: Averaged User System Probability of False Alarm $\bar{p}_{fa}[k]$ for ODF, RLM and PRLM. $N = 7$ targets, $k = 1, \dots, 50$.

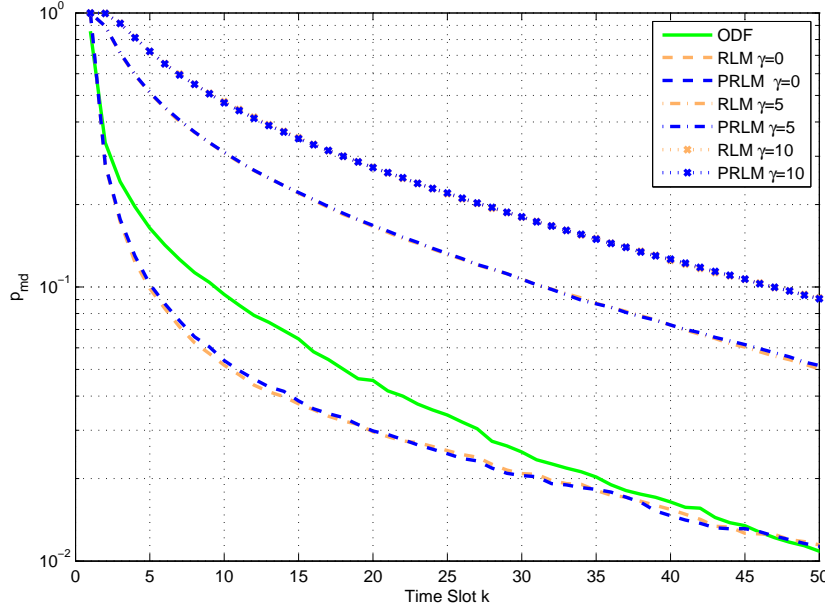


Figure 6.4.6: Averaged User System Probability of False Alarm $\bar{p}_{md}[k]$ for ODF, RLM and PRLM. $N = 7$ targets, $k = 1, \dots, 50$.

targets we get a speed-up factor of about 2.5 times for the (P)RLM algorithms. This speed-up grows with the number of targets being considered, as it is evident from the figure, because in the case of $N = 5$ and $N = 7$ we got respectively a speed-up of approximately 3 and 5 times. Conversely g_O is far higher, as it accounts also the Space Complexity reduction given by the (P)RLM. In fact we get respectively Overall Complexity Gain for the three cases of $g_O = 7$, $g_O = 20$ and $g_O = 100$.

6.5 Conclusions

In this chapter we presented two sub-optimal fusion alternatives to ODF in the context of distributed multi-classification with a binary wireless sensor network. It was shown that the complexity of these approaches is lower both in Time and Space dimensions with respect to ODF, even though the trade-off is a loss in terms of the Probability of False Alarm. However, due to the separated target approach of (P)RLM we can choose the threshold in an accurate way to obtain a desired false alarm rate, but degrading the Probability of Detection. Nonetheless it was shown experimentally that moving the threshold does not change the Probability of Detection negative error exponent, which is essentially

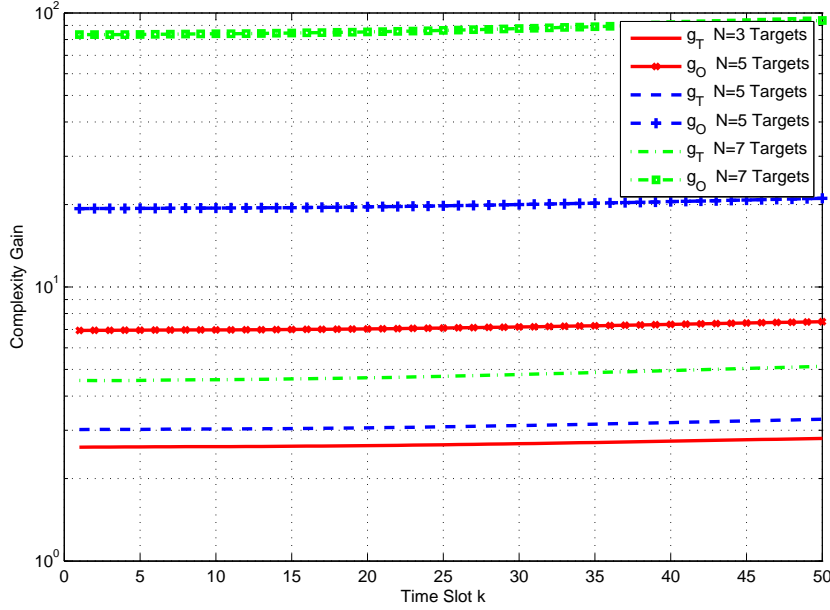


Figure 6.4.7: Gain Complexity (Time and Overall) of (P)RLM over ODF.

the same as the ODF and so only more samples are required by (P)RLM to obtain the same detection performances of the optimum approach.

Further Developments about this work will be the analysis and the comparison between ODF and (P)RLM over non-ideal communication channel models (such as AWGN or Rayleigh Flat Fading), and the consequent loss of performance of these approaches at low SNR. Finally also the uncertainties deriving from non-perfect track estimation of multi-sensor multi-target tracking system will be analyzed both in loss of performances and implications that will eventually imply a possible slight modification of the algorithms.

6.6 Appendix

6.6.1 Maximum Detection Range

In this paragraph we explain the argument behind the concept of maximum detection range r_{max} . Let us first recall the formula of Hellinger Distance (HD) for discrete distributions. Given two distributions $p(x_i)$ and $q(x_i)$ defined on the same discrete support $x_i \in \mathcal{X}$, HD [98] is defined by $d_H(p, q) = \frac{1}{2} \sum_{x_i \in \mathcal{X}} (\sqrt{p(x_i)} - \sqrt{q(x_i)})^2$; the HD satisfies the inequalities $0 \leq d_H(p, q) \leq 1$, where $d_H(p, q) = 0$ means that $p(x_i) = q(x_i), \forall x_i \in \mathcal{X}$ and $d_H(p, q) = 1$ implies a complete separation between distributions. In our case when $p_d(r_n^l[i]) \simeq 2 \cdot p_{fa}$, the HD between the two distributions $p(z^\ell[i]|T_n = 0)$ and $p(z^\ell[i]|T_n = 1)$ is given by $d_H(\cdot) = \frac{1}{2} [(\sqrt{p_{fa}} - \sqrt{2p_{fa}})^2 + (\sqrt{1 - p_{fa}} - \sqrt{1 - 2p_{fa}})^2]$ which under the

assumption that $p_{fa} \ll 1$ can be well approximated as $\frac{1}{2} [(1 - \sqrt{2})^2 p_{fa}] \simeq 0.085 p_{fa}$. So considering higher $r_n^\ell[i]$ implies obtaining inference on two hypothesis which are Hellinger-separated by a factor of $0.085 p_{fa}$, which does not give a valuable information.

6.6.2 Time Complexity for (P)RLM

In this paragraph we derive the formula for Time Complexity $C_T^{(P)RLM}[k]$ for (P)RLM sub-optimal fusion algorithms. At each time i and for each sensor ℓ we need to compute Eq. (6.3.5) for RLM or Eq. (6.3.6) for PRLM. This marginalization requires $|\mathcal{T}^\ell[i]| 2^{|\mathcal{T}^\ell[i]|-1}$ operations (products) to obtain the terms of the sum and $2^{|\mathcal{T}^\ell[i]|-1}$ operations (sums) to complete marginalization. This marginalization has to be repeated two times ($T_n = \{0, 1\}$) to obtain the log-ratio of the (6.3.3), which has to be computed for each user belonging to $|\mathcal{T}^\ell[i]|$, so that the total number of operations at time k for the sensor ℓ is $N_o = |\mathcal{T}^\ell[i]| (|\mathcal{T}^\ell[i]| 2^{|\mathcal{T}^\ell[i]|} + 2^{|\mathcal{T}^\ell[i]|}) = |\mathcal{T}^\ell[i]| (|\mathcal{T}^\ell[i]| + 1) (2^{|\mathcal{T}^\ell[i]|})$. By considering the dominant terms and summing over all the sensors ℓ and over all the time slots i till time k , we get $C_T[k] = O(\sum_{i=1}^k \sum_{\ell=1}^L T^\ell[i]^2 2^{T^\ell[i]})$.

Chapter 7

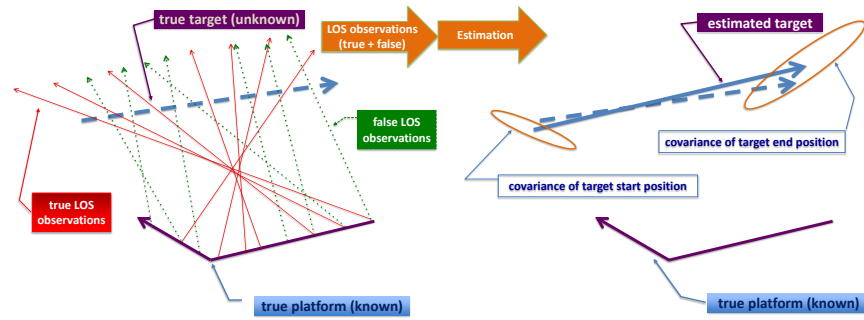
Data Fusion for inverse localization tasks: application to wideband passive sonar estimation with ML-PDA

7.1 Introduction

7.1.1 Problem Motivation

Target motion analysis (TMA) with bearings-only measurements is a well-understood and -studied problem (see Fig. 7.1.1). It has been shown in the literature that as long as the platform is outmaneuvering the target, observability of the latter is assured and its motion can be inferred, even from very noisy measurements. However, since the bearing observations depend on the relative positions between the target and the platform, it is useful to understand whether, given the results of TMA estimation, it could be possible to identify, completely or at least partially, the trajectory of the *observing platform*.

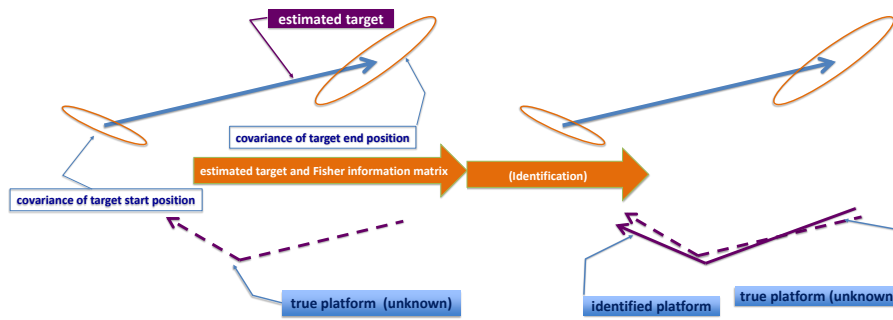
More specifically, the problem arises when a “target-friendly” entity, as opposed to being cooperative with the platform, intercepts the results of the target estimation performed by the platform (see Fig. 7.1.2); the question is whether this entity can identify, partially or totally, the trajectory traced by the platform. The feasibility of this problem is of twofold interest: *(i)* it verifies the utility of intercepting communications (containing TMA-related information) between the platform and platform-cooperative entities, because this information would be useful; and *(ii)* it motivates, at the platform side, the need for secure and encrypted transmission of the TMA estimation results.



(a) Collecting measurements.

(b) Estimating target trajectory.

Figure 7.1.1: The target estimation process by the platform side.



(a) Intercepting target estimation results from the platform side.

(b) Identifying the platform trajectory.

Figure 7.1.2: The platform identification process by the target-friendly side.

7.1.2 Related Works

Seminal results on the continuous-time observability of the target motion, through a wideband passive sonar, were derived in [101, 102]. In fact, by successive differentiations of the measurement function, a necessary condition was derived and it was shown that a platform maneuver is a needed prerequisite to ensure observability of the target; however unobservable maneuvering-platform trajectories could exist. This analysis was rigorously extended, in the form of a necessary and sufficient condition, in [103, 104, 105, 106]; a comprehensive analysis of observability related to practical scenarios was also conducted in [106]. These results were also demonstrated in discrete-time in [107] via linear algebra; observability insights in different scenarios were presented, and also a stochastic observability (and estimability) analysis was performed. In [108] a theoretical connection between the invertibility of the Fisher information matrix (FIM) and target (local) observability was established. In [109] (and references therein) the optimal platform maneuver was designed, in the sense of the best estimation accuracy in terms of the FIM.

There are three common approaches to standard TMA (with or without Doppler measurements): *maximum likelihood* (ML), *pseudolinear* (PL) and *instrumental variables* (IV), estimation. Although the first approach is asymptotically efficient, it is complex and therefore sub-optimal solutions are desirable. PL estimation has the advantage of being in closed form and of easy computation, however it can lead to severe bias even in favorable conditions [110]. Consequently IV estimation, presenting reduced bias estimates, has seen recent attention [111, 112, 110, 113, 114].

Alternative bearings-only TMA scenarios have been studied recently in [115, 116, 117] and ML batch estimators have been proposed. More specifically, in [115] the problem of bearings-only TMA for conditionally-deterministic target motion and with operational constraints on the platform is tackled with Markov chain Monte Carlo methods. In [116] TMA of a maneuvering target and non-maneuvering platform is studied; observability is established and a batch estimator is proposed. The concept was later applied to the scenario of a circular constant-speed target and a non-maneuvering target in [117].

The estimators proposed in these references do not deal with the problem of false measurements (clutter) and less-than-unity probability of detection. The seminal work in [118] derived a ML estimate of target parameters for both wideband and narrowband passive sonars in the presence of false detections (clutter), based on probabilistic data association (ML-PDA); the performance of the estimator was evaluated in terms of the Cramér-Rao lower bound (CRLB). It was shown that the effect of the clutter on the performance through the CRLB was simply via a product with a less-than-unity scalar value, called the information reduction factor (IRF).

The ML-PDA was extended by incorporating amplitude information to enhance performance in the scenarios of “low-observable” (i.e. low Signal-to-Noise ratio) targets in [119]; improved accuracy and superior global convergence were demonstrated. In [120] ML-PDA was applied to the problem of low-observable

target estimation using electro-optical sensors; also a sliding-window batch approach for ML-PDA estimation was derived, capable of dealing with temporary disappearance of targets and/or targets with velocities changing over time. ML-PDA was also successfully applied to active sonar tracking in [121], where also an efficient computation of the ML estimate, namely, directed subspace search (DSS), was derived.

7.1.3 Main Results and Chapter Organization

The main contributions of the chapter are summarized as follows.

- We study the “inverse” problem of identifying the *platform* trajectory, following a “two-leg” motion model, through its ML-PDA estimation results on a target; to the best of our knowledge, this is addressed here for the first time.
- We derive and study the objective function to be optimized for identifying the platform trajectory; it is shown that the optimization of this function depends on neither the IRF nor the measurement variance at the platform side; that is, the exact¹ information to identify the platform trajectory is unnecessary. Also it is demonstrated that the platform trajectory is *unobservable* unless it keeps a constant speed during its two different legs.
- We use an efficient and practical algorithm, based on *derivative-free* local search, to solve the nonlinear problem associated with the identification task. The local optimization routine is initialized from geometric considerations and exploiting the structure of the observed FIM.

The chapter is organized as follows: In Sec. 7.2 we introduce the model for passive wideband-sonar localization and we give the background on the ML-PDA approach; in Sec. 7.3 we formulate the problem of inverse localization and we show some important identification properties; in Sec. 7.4 we devise a procedure to compute a good initial guess as input for the local optimization routines, while in Sec. 7.5 we show, by simulation, the performance of the proposed solution; some concluding remarks and future research are given in Sec. 7.6; proofs and derivations are confined to the Appendices.

Notation - Lower-case (resp. Upper-case) bold letters denote vectors (resp. matrices), with a_n (resp. $A_{n,m}$) representing the n th (resp. the (n, m) th) element of the vector \mathbf{a} (resp. matrix \mathbf{A}); upper-case calligraphic letters and braces denote finite sets, with $[a : b]$, $a \leq b$ representing the set $\{a, a + 1, \dots, b\}$; \mathbf{I}_N denotes the $N \times N$ identity matrix, while $\text{diag}(\mathbf{t})$ is a diagonal matrix with diagonal equal to \mathbf{t} ; $(\cdot)^t$, $\|\cdot\|$ and $\langle \cdot \rangle$ denote transpose, Frobenius norm and inner product operators, respectively; $\nabla_{\mathbf{t}}(\cdot)$ denotes the gradient operator w.r.t. the vector \mathbf{t} ; $\mathbf{e}^j(\mathbf{A}, i)$ denotes the unit eigenvector of a symmetric and diagonalizable

¹We will show however that for devising an efficient local-optimization algorithm a range of variability should be given; however the width of this range does not affect significantly the performance.

matrix \mathbf{A} (of size $[r \times r]$) corresponding to the eigenvalue λ_i , $i \in \mathcal{R} \triangleq \{1, \dots, r\}$, where $\lambda_s > \lambda_{s+1}$ $s \in \mathcal{R}/\{r\}$ and $j \in \{-1, 1\}$ denotes the sign ambiguity in the eigenvector formula; $\arctan_2(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, denotes the four-quadrant inverse tangent with argument $\frac{x_1}{x_2}$; $P(\cdot)$ is used to denote probability mass functions (pmf) or probability density functions (pdf), while $P(\cdot|\cdot)$ the corresponding conditional counterpart; $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a real normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; finally the symbols \rightarrow , \ni , \sim , and \perp mean “maps to”, “such that”, “distributed as” and “orthogonal”, respectively.

7.2 System Model

7.2.1 Motion Models description

The system model is described graphically in Fig. 7.2.1. We assume that the target is observed by the platform at n time samples, i.e. $t \in \mathcal{T} \triangleq \{t_1, \dots, t_n\}$; also we define the set of indices $\mathcal{I} \triangleq \{1, \dots, n\}$. In the following we will explicitly list all the assumptions made, starting from the motion models of the platform and the target.

Assumption I: We assume that the target moves according to a constant velocity (CV) motion model [100]. For this reason we define $\mathbf{p}_T(t_i) \triangleq [\xi_T(t_i) \ \eta_T(t_i)]^t$ and $\mathbf{v}_T \triangleq [\dot{\xi}_T \ \dot{\eta}_T]^t$ as the position at t_i and the (constant) velocity 2-D vector of the target; ξ and η are used to denote the east and north directions. Given the CV assumption, $\{\mathbf{p}_T(t_1), \mathbf{p}_T(t_n)\}$ uniquely define the state of the target at $t_i \in \mathcal{T}$. Therefore we append them in $\mathbf{x}_T \triangleq [\mathbf{p}_T(t_1)^t \ \mathbf{p}_T(t_n)^t]^t$, which represents the true target state vector, unknown at the platform side. The target motion model has the explicit expression:

$$\begin{bmatrix} \xi_T(t_i) \\ \eta_T(t_i) \end{bmatrix} = \mathbf{p}_T(t_1) + (t_i - t_1) \cdot \mathbf{v}_T \quad (7.2.1)$$

$$= \mathbf{p}_T(t_1) + \underbrace{\frac{(t_i - t_1)}{(t_n - t_1)}}_{\triangleq \alpha_i} [\mathbf{p}_T(t_n) - \mathbf{p}_T(t_1)], \quad t_i \in \mathcal{T} \quad (7.2.2)$$

Assumption II: We assume a platform moving according to a “two-legs” motion model; this requirement not only ensures observability of the target from the platform point of view [101, 122], but also represents the easiest trajectory that can be followed by the platform. We denote $\mathbf{p}_P(t_i) \triangleq [\xi_P(t_i) \ \eta_P(t_i)]^t$, $\mathbf{v}_{P,1} \triangleq [\dot{\xi}_{P,1} \ \dot{\eta}_{P,1}]^t$ and $\mathbf{v}_{P,2} \triangleq [\dot{\xi}_{P,2} \ \dot{\eta}_{P,2}]^t$, as the position at $t_i \in \mathcal{T}$, first-leg and second-leg velocity vectors; also we group $\mathbf{x}_P \triangleq [\mathbf{p}_P(t_1)^t \ \mathbf{v}_{P,1}^t \ \mathbf{v}_{P,2}^t]^t$ into the platform state vector, representing the unknowns at the target-friendly side. Note that \mathbf{x}_P does not *uniquely* define the platform trajectory, since the turning time t_k is also needed.

Assumption III: Throughout this chapter we will make the simplifying assumption that t_k is known at the target-friendly side. In fact it is reasonable

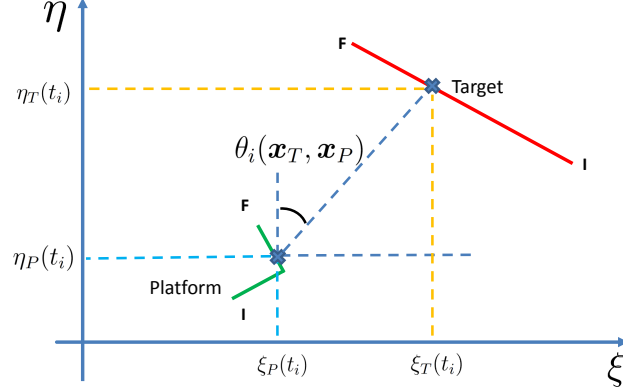


Figure 7.2.1: System model considered for our application. The crosses refer to the bearing measurement at snapshot $t_i \in \mathcal{T}$.

to assume that t_k in practice will happen nearly the middle of the observation interval, i.e., $t_k \approx \frac{t_n - t_1}{2}$, in order to assure a good degree of observability². This assumption will be relaxed in Sec. 7.5, where a sensitivity analysis with respect to (w.r.t.) the timing-uncertainty on t_k will be shown. Therefore, once t_k is assumed to be known, the platform motion model is explicitly described as

$$\begin{bmatrix} \xi_P(t_i) \\ \eta_P(t_i) \end{bmatrix} = \begin{cases} \mathbf{p}_P(t_1) + (t_i - t_1)\mathbf{v}_{P,1} & t_i \in \mathcal{T}, \ni t_i < t_k \\ \mathbf{p}_P(t_1) + (t_k - t_1)\mathbf{v}_{P,1} + (t_i - t_k)\mathbf{v}_{P,2} & t_i \in \mathcal{T}, \ni t_i \geq t_k \end{cases} \quad (7.2.3)$$

Note that here the magnitudes of $\mathbf{v}_{P,1}$ and $\mathbf{v}_{P,2}$ are arbitrary. As shown later, unique identifiability of the platform trajectory (our goal) requires these magnitudes to be the same, i.e., the platform speed should be constant.

7.2.2 ML-PDA statistical assumptions and formulation

The statistical assumptions on the measurements are summarized as follows.

²In fact, a turn at the beginning or the end of the observation interval would result in a platform trajectory similar to a single leg (CV model), thus leading to a nearly singular FIM.

Assumption I: The bearing (true) measurement κ_i , collected by the platform at t_i , follows the model

$$\kappa_i = \theta_i(\mathbf{x}_T, \mathbf{x}_P) + n_i = \arctan\left(\frac{\xi_T(t_i) - \xi_P(t_i)}{\eta_T(t_i) - \eta_P(t_i)}\right) + n_i \quad (7.2.4)$$

where $\theta_i(\mathbf{x}_T, \mathbf{x}_P)$ denotes the noise-free bearing (we stress the dependence on both platform and target state vectors) and $n_i \sim \mathcal{N}(0, \sigma_\theta^2)$. For notational convenience we also define here the range $r_i(\mathbf{x}_T, \mathbf{x}_P)$ as

$$r_i(\mathbf{x}_T, \mathbf{x}_P) \triangleq \|\mathbf{p}_T(t_i) - \mathbf{p}_P(t_i)\| \quad (7.2.5)$$

Assumption II: We assume, as in realistic environments, that a passive sonar at $t_i \in \mathcal{T}$ collects a set of measurements $\mathbf{z}(i)$, due to clutter and non-perfect detection. More specifically, we have

$$\mathbf{z}(i) \triangleq \{z_j(i)\}_{j=1}^{m_i} \quad (7.2.6)$$

where m_i denotes the number of collected measurements at $t_i \in \mathcal{T}$. The statistical assumptions over the set in Eq. (7.2.6) are: (i) the true measurement κ_i can be detected at most only once, with probability P_D ; (ii) the number of false measurements at $t_i \in \mathcal{T}$ follows a known probability mass function $\mu_F(\cdot)$, given by a Poisson law with known expected number of false alarms per unit of volume λ ; therefore the false measurements are distributed uniformly in the surveillance region (in the bearing space).

Assumption III: We assume conditional mutual independence among the set of measurements, that is $P(\mathbf{z}(i_1), \mathbf{z}(i_2)|\mathbf{x}_T, \mathbf{x}_P) = P(\mathbf{z}(i_1)|\mathbf{x}_T, \mathbf{x}_P)P(\mathbf{z}(i_2)|\mathbf{x}_T, \mathbf{x}_P)$, $\forall i_1 \neq i_2$.

Under these assumptions and denoting $\check{\mathbf{x}}_T$ (resp. $\check{\mathbf{x}}_P$) as the true target (resp. platform) state vector, the ML-PDA estimate $\hat{\mathbf{x}}_T$ is obtained as

$$\hat{\mathbf{x}}_T \triangleq \arg \max_{\mathbf{x}_T} \prod_{i=0}^n P(\mathbf{z}(i)|\mathbf{x}_T, \check{\mathbf{x}}_P) \quad (7.2.7)$$

where the likelihood $P(\mathbf{z}(i)|\mathbf{x}_T, \check{\mathbf{x}}_P)$ is given in explicit form as follows:

$$\begin{aligned} P(\mathbf{z}(i)|\mathbf{x}_T, \check{\mathbf{x}}_P) &= u^{-m_i} (1 - P_D) \mu_F(m_i) \\ &+ \frac{u^{1-m_i} P_D \mu_F(m_i - 1)}{m_i} \sum_{j=1}^{m_i} \frac{1}{\sqrt{2\pi}\sigma_\theta} \times \exp\left(-\frac{1}{2} \left(\frac{z_j(i) - \theta_i(\mathbf{x}_T, \check{\mathbf{x}}_P)}{\sigma_\theta}\right)^2\right) \end{aligned} \quad (7.2.8)$$

Note that $\check{\mathbf{x}}_P$ is assumed known at the platform side. It was shown numerically in [118] that the covariance matrix of the ML-PDA estimator *essentially* attains the CRLB and therefore it can be regarded “practically” as an efficient estimator. For this reason, the covariance matrix is approximated by the inverse of the FIM given by

$$\mathbf{J}(\mathbf{x}_T, \mathbf{x}_P, \alpha_\theta) \triangleq \alpha_\theta \underbrace{\sum_{i=0}^n \nabla_{\mathbf{x}_T}(\theta_i(\mathbf{x}_T, \mathbf{x}_P)) \nabla_{\mathbf{x}_T}^t(\theta_i(\mathbf{x}_T, \mathbf{x}_P))}_{\triangleq \mathbf{J}_u(\mathbf{x}_T, \mathbf{x}_P)} \quad (7.2.9)$$

where α_θ is defined as³

$$\alpha_\theta \triangleq \frac{q_2}{\sigma_\theta^2} \quad (7.2.10)$$

with q_2 representing the IRF [118]. Note that

$$q_2 = q_2(\lambda v_g, P_D, g) \quad (7.2.11)$$

where v_g and g denote the volume of the validation region and the gating threshold, respectively. It can be shown, after some manipulations, that $\nabla_{\mathbf{x}_T}(\theta_i(\mathbf{x}_T, \mathbf{x}_P))$ has the explicit expression

$$\nabla_{\mathbf{x}_T}(\theta_i(\mathbf{x}_T, \mathbf{x}_P)) = \frac{1}{r_i(\mathbf{x}_T, \mathbf{x}_P)} \begin{bmatrix} (1 - \alpha_i) \cos(\theta_i(\mathbf{x}_T, \mathbf{x}_P)) \\ -(1 - \alpha_i) \sin(\theta_i(\mathbf{x}_T, \mathbf{x}_P)) \\ \alpha_i \cos(\theta_i(\mathbf{x}_T, \mathbf{x}_P)) \\ -\alpha_i \sin(\theta_i(\mathbf{x}_T, \mathbf{x}_P)) \end{bmatrix} \quad (7.2.12)$$

The FIM at the platform side is necessarily evaluated as $\mathbf{J}^{obs} \triangleq \mathbf{J}(\check{\mathbf{x}}_P, \hat{\mathbf{x}}_T, \check{\alpha}_\theta)$, where $\check{\alpha}_\theta$ denotes the true α_θ , known at the platform side. Given the results of the estimation process at the platform side, that is $\{\hat{\mathbf{x}}_T, \mathbf{J}^{obs}\}$, our task can be summarized as follows.

We wish to identify the platform state, represented by \mathbf{x}_P , by observing only the estimation results of the ML-PDA, that is $\{\hat{\mathbf{x}}_T, \mathbf{J}^{obs}\}$. It is worth remarking that the unknowns of this deterministic problem are represented by $\{\check{\mathbf{x}}_P, \check{\alpha}_\theta\}$. In fact, even if $\check{\alpha}_\theta$ does not contribute to specifying the platform trajectory, it has to be identified to solve this task.

The first important remark is that the identification problem is function only of $\hat{\mathbf{x}}_T$ rather than the true trajectory $\check{\mathbf{x}}_T$. This has an important consequence: *the identification of the platform trajectory does not depend on the true target trajectory $\check{\mathbf{x}}_T$; however we will show that due to the sensitivity w.r.t. the true platform parameters, a “larger” FIM matrix will lead to an easier identification in terms of local optimization routines.*

7.3 Objective Function Determination

As a starting point of the identification problem, it would be natural to solve the non-linear equation

$$\mathbf{J}(\hat{\mathbf{x}}_T, \mathbf{x}_P, \alpha_\theta) = \mathbf{J}^{obs} \quad (7.3.1)$$

³Note that α_θ should not be confused with $\alpha_i = \frac{t_i - t_1}{t_n - t_1}$ defined in Eq. (7.2.2).

for the variables \mathbf{x}_P and α_θ . However, as stated by the following proposition, we will show that this system is *unobservable*, since there exists an infinite number of solutions satisfying Eq. (7.3.1).

Proposition 13. *If $\{\mathbf{x}_P^*, \alpha_\theta^*\}$ is a solution of (7.3.1), then each $\{\mathbf{x}'_P, \alpha'_\theta\}$ generated by the subspace*

$$\{\mathbf{x}'_P, \alpha'_\theta\} = \{\beta\mathbf{x}_P^* + (1 - \beta)\bar{\mathbf{x}}_E, \beta^2\alpha_\theta^*\}, \quad \beta \in \mathbb{R} \quad (7.3.2)$$

$$\bar{\mathbf{x}}_E \triangleq [\hat{\mathbf{p}}_T(t_1)^t \quad \hat{\mathbf{v}}_T^t \quad \hat{\mathbf{v}}_T^t]^t \quad (7.3.3)$$

is also a solution of (7.3.1).

Proof. The proof is given in Appendix 7.7.1. □

The above proposition states that *no unambiguous identification of the platform trajectory $\check{\mathbf{x}}_P$ is possible* when the platform is following a trajectory according to Eq. (7.2.3) with different speed in each leg. The explanation of this is given by the fact that a convex combination (through β) of the platform and the (augmented) target state vector would produce the same FIM with an α_θ scaled by β^2 . At this point it is worth pointing out the *difference* between this requirement and the target estimation, *in which the platform maneuver is the only prerequisite* to ensure observability of the target.

When both legs of the platform trajectory are constrained to have the same speed s , that is

$$s \triangleq \|\mathbf{v}_{P,1}\| = \|\mathbf{v}_{P,2}\| \quad (7.3.4)$$

it can be shown that the subspace described by Eq. (7.3.2) violates constraint (7.3.4) (see Appendix 7.7.1), hence identification of the platform trajectory is possible. More specifically, the constraint in Eq. (7.3.4) represents a *necessary condition* for *observability* of the platform trajectory. On the basis of this constraint we *define* a new platform-state vector \mathbf{x}_P^s (and we denote the true platform-state vector as $\check{\mathbf{x}}_P^s$) as follows

$$\mathbf{x}_P^s \triangleq [\xi_P(t_1) \quad \eta_P(t_1) \quad s \quad \phi_1 \quad \phi_2]^t \quad (7.3.5)$$

$$\phi_i \triangleq \arctan_2(\mathbf{v}_{P,i}), \quad i \in \{1, 2\} \quad (7.3.6)$$

Thus Eq. (7.3.1) becomes

$$\mathbf{J}(\mathbf{x}_P^s, \alpha_\theta) = \mathbf{J}^{obs} \quad (7.3.7)$$

with unknowns \mathbf{x}_P^s and α_θ (starting from here we drop the dependence on $\hat{\mathbf{x}}_T$ to keep the notation simple). In general, the non-linear system described by Eq. (7.3.7) can still admit multiple solutions, since there is no theoretical proof that the constraint in Eq. (7.3.4) is also a sufficient condition for observability of $\{\mathbf{x}_P^s, \alpha_\theta\}$. This is because proving that $\{\mathbf{x}_P^s, \alpha_\theta\} \rightarrow \mathbf{J}(\mathbf{x}_P^s, \alpha_\theta)$ is a *one-to-one mapping* is an extremely difficult task. Nonetheless, we will show, through simulations in Sec. 7.5, that this property seems to be satisfied and that \mathbf{x}_P^s can be identified.

To solve Eq. (7.3.7) in an efficient way we consider the search for the minimum of the Frobenius norm $\mathcal{F}(\mathbf{x}_P^s, \alpha_\theta)$, namely,

$$\{\hat{\mathbf{x}}_P^s, \hat{\alpha}_\theta\} = \arg \min_{\{\mathbf{x}_P^s, \alpha_\theta\}} \underbrace{\|\mathbf{J}^{obs} - \mathbf{J}(\mathbf{x}_P^s, \alpha_\theta)\|_2}_{\triangleq \mathcal{F}(\mathbf{x}_P^s, \alpha_\theta)} \quad (7.3.8)$$

It is easy to see that the global minimum (corresponding to a zero value) of Eq. (7.3.8) corresponds to the solution of Eq. (7.3.7). Although studying the function in Eq. (7.3.8) appears as arbitrary (in fact other matrix distance norms can be considered), there is an important reason behind this choice: we will show in the following that $\mathcal{F}(\mathbf{x}_P^s, \alpha_\theta)$ (see (7.3.11)) can be expressed in terms of a *weighted-square-distance*, and an important property of weighted non-linear least squares problems can be exploited [68].

The first step to express Eq. (7.3.8) in terms of a convenient weighted-least squares problem is to search for independent entries of $\mathbf{J}(\mathbf{x}_P^s, \alpha_\theta)$. The following Lemma will be used.

Lemma 14. *The FIM $\mathbf{J}(\mathbf{x}_P^s, \alpha_\theta)$ has only 9 independent entries.*

Proof. The proof is given in Appendix 7.7.2. \square

Exploiting Lemma 14, we can express Eq. (7.3.8) in terms of a weighted-square distance, as stated by the following proposition.

Proposition 15. *The norm in Eq. (7.3.8) can be equivalently written in the form*

$$\mathcal{F}(\mathbf{x}_P^s, \alpha_\theta) = [\mathbf{j}^{obs} - \mathbf{j}(\mathbf{x}_P^s, \alpha_\theta)]^t \mathbf{W} [\mathbf{j}^{obs} - \mathbf{j}(\mathbf{x}_P^s, \alpha_\theta)] \quad (7.3.9)$$

where $\mathbf{j}^{obs} \in \mathbb{R}^9$ and $\mathbf{j}(\mathbf{x}_P^s, \alpha_\theta) \in \mathbb{R}^9$ are obtained by stacking the independent components of $\mathbf{J}(\mathbf{x}_P^s, \alpha_\theta)$ and \mathbf{J}^{obs} , respectively and \mathbf{W} is a diagonal weighting matrix, defined as follows:

$$\mathbf{W} \triangleq \text{diag} \left([1 \ 1 \ 1 \ 1 \ 2 \ 2 \ 4 \ 2 \ 2]^t \right) \quad (7.3.10)$$

Proof. The proof is given in Appendix 7.7.3. \square

Note that $\mathbf{j}(\mathbf{x}_P^s, \alpha_\theta)$ retains the same factorization as $\mathbf{J}(\mathbf{x}_P^s, \alpha_\theta)$, that is, $\mathbf{j}(\mathbf{x}_P^s, \alpha_\theta) = \alpha_\theta \mathbf{j}_u(\mathbf{x}_P^s)$, where $\mathbf{j}_u(\mathbf{x}_P^s)$ is defined accordingly to Eq. (7.2.9). Hence Eq. (7.3.9) can be rewritten as

$$\mathcal{F}(\mathbf{x}_P^s, \alpha_\theta) = [\mathbf{j}^{obs} - \alpha_\theta \mathbf{j}_u(\mathbf{x}_P^s)]^t \mathbf{W} [\mathbf{j}^{obs} - \alpha_\theta \mathbf{j}_u(\mathbf{x}_P^s)] \quad (7.3.11)$$

Note that the minimization of the objective function in Eq. (7.3.11) is in the standard form of non-linear weighted least squares [68]. Additionally in this case the non-linear weighted square distance is linear in some of the parameters to be estimated, in this case α_θ , and thus this non-linear problem can be solved in a reduced dimension space; the details are given by the following proposition.

Proposition 16. *The minimization of the 6-dimensional objective $\mathcal{F}(\mathbf{x}_P^s, \alpha_\theta)$ is equivalent to maximization of the objective $\mathcal{G}(\mathbf{x}_P^s)$, defined as*

$$\mathcal{G}(\mathbf{x}_P^s) \triangleq \langle \mathbf{j}^{obs}, \mathbf{c}(\mathbf{x}_P^s) \rangle^2 \quad (7.3.12)$$

$$\mathbf{c}(\mathbf{x}_P^s) \triangleq \frac{\mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)}{[\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)]^{1/2}} \quad (7.3.13)$$

Proof. The proof is given in Appendix 7.7.4. \square

The maximization of $\mathcal{G}(\mathbf{x}_P^s)$ has the advantage of reducing the search space from \mathbb{R}^6 to \mathbb{R}^5 . The exclusion of α_θ also allows a search *only* in the subspace of variables determining the platform trajectory. At this point some observations on the objective function in Eq. (7.7.19) are in order:

- The function $\mathcal{G}(\mathbf{x}_P^s)$ is highly non-linear in \mathbf{x}_P^s , therefore no closed form solution to optimization of the Eq. (7.3.12) exists, thus numerical optimization procedures need to be used;
- Local optimization routines can get stuck in local maxima or other (non-equilibrium) points⁴.

An important issue in local optimization routines that determines their success, is the choice of a good initial guess. In the following we will suggest a good start, based on the FIM and geometry.

7.4 “Geometry-driven” FIM-aided initial guess choice

In this section we show how a good initial guess, denoted as $\hat{\mathbf{x}}_P^{s,0}$, can be chosen to help convergence of local optimization routines. To accomplish this task let us define:

$$\mathbf{p}_P \triangleq \begin{cases} \left[\begin{array}{ccc} \mathbf{p}_P(t_1)^t & \mathbf{p}_P(t_k)^t & \mathbf{p}_P(t_n)^t \end{array} \right]^t \\ \ni \frac{\|\mathbf{p}_P(t_k) - \mathbf{p}_P(t_1)\|}{t_k - t_1} = \frac{\|\mathbf{p}_P(t_n) - \mathbf{p}_P(t_k)\|}{t_n - t_k} \end{cases} \quad (7.4.1)$$

Since it is easy to show that there is a *one-to-one mapping* between \mathbf{p}_P and \mathbf{x}_P^s , we will search for a $\hat{\mathbf{p}}_P^0$ close to $\check{\mathbf{p}}_P$ (even though the latter is not known), as the equivalent input of the local optimization routine. For this reason, we first will find good approximations $\hat{\mathbf{p}}_P^0(t_1)$ and $\hat{\mathbf{p}}_P^0(t_n)$. After this, we will give the details on how to find $\hat{\mathbf{p}}_P^0(t_k)$, under the constraint of Eq. (7.4.1). The following considerations are based on the assumption that $\check{\alpha}_\theta$ is known; at the end of the section we will remove this restriction. In the following, for the sake of simplicity, we will use the short-hand notations $r_i \triangleq r_i(\hat{\mathbf{x}}_T, \check{\mathbf{x}}_P^s)$ for the range, $\theta_i \triangleq \theta_i(\hat{\mathbf{x}}_T, \check{\mathbf{x}}_P^s)$ for the bearing, $\mathbf{i}_i \triangleq [\sin(\theta_i) \quad \cos(\theta_i)]^t$ for the bearing unit vector, and we will drop the zero superscript in $\hat{\mathbf{p}}_P^0$ and $\hat{\mathbf{p}}_P^0(t_i)$.

⁴In fact, several local optimization routines are not based on the evaluation of the gradient vector of the objective function at each iteration, thus a “null gradient” condition is not ensured when a stopping condition is met.

7.4.1 Choice of $\hat{\mathbf{p}}_P(t_1)$ and $\hat{\mathbf{p}}_P(t_n)$

It can be easily shown that $\{\check{\mathbf{p}}_P(t_1), \check{\mathbf{p}}_P(t_n)\}$ can be expressed in the form:

$$\check{\mathbf{p}}_P(t_1) = \hat{\mathbf{p}}_T(t_1) + r_1 \hat{\mathbf{i}}_1 \quad (7.4.2)$$

$$\check{\mathbf{p}}_P(t_n) = \hat{\mathbf{p}}_T(t_n) + r_n \hat{\mathbf{i}}_n \quad (7.4.3)$$

In order to obtain good estimates $\{\hat{\mathbf{p}}_P(t_1), \hat{\mathbf{p}}_P(t_n)\}$ we need to find good approximations $\{\hat{r}_1, \hat{r}_n\}$ and $\{\hat{\mathbf{i}}_1, \hat{\mathbf{i}}_n\}$. These issues can be tackled separately.

Choice of $\{\hat{r}_1, \hat{r}_n\}$

A coarse approximation of the ranges r_1 and r_n is given by:

$$\hat{r}_1 \triangleq \sqrt{\frac{\check{\alpha}_\theta \sum_{i=1}^n (1 - \alpha_i)^2}{\text{tr}(\mathbf{J}^{obs}[1, 1])}}; \quad \hat{r}_n \triangleq \sqrt{\frac{\check{\alpha}_\theta \sum_{i=1}^n \alpha_i^2}{\text{tr}(\mathbf{J}^{obs}[2, 2])}} \quad (7.4.4)$$

where $\mathbf{J}^{obs}[m, n]$ represents the (m, n) th $[2 \times 2]$ block matrix of \mathbf{J}^{obs} , whose explicit expression in block form (exploiting the derivation in Appendix 7.7.2) is given as follows:

$$\mathbf{J}^{obs} = \begin{bmatrix} \mathbf{J}^{obs}[1, 1] & \mathbf{J}^{obs}[1, 2] \\ \mathbf{J}^{obs}[2, 1] & \mathbf{J}^{obs}[2, 2] \end{bmatrix} \quad (7.4.5)$$

$$= \alpha_\theta \begin{bmatrix} \sum_{i=1}^n (1 - \alpha_i)^2 \check{\mathbf{y}}_i \check{\mathbf{y}}_i^t & \sum_{i=1}^n \alpha_i (1 - \alpha_i) \check{\mathbf{y}}_i \check{\mathbf{y}}_i^t \\ \sum_{i=1}^n \alpha_i (1 - \alpha_i) \check{\mathbf{y}}_i \check{\mathbf{y}}_i^t & \sum_{i=1}^n \alpha_i^2 \check{\mathbf{y}}_i \check{\mathbf{y}}_i^t \end{bmatrix} \quad (7.4.6)$$

where $\check{\mathbf{y}}_i \triangleq \frac{1}{r_i(\hat{\mathbf{x}}_T, \hat{\mathbf{x}}_P^s)} [\cos(\theta_i(\hat{\mathbf{x}}_T, \hat{\mathbf{x}}_P^s)) \quad -\sin(\theta_i(\hat{\mathbf{x}}_T, \hat{\mathbf{x}}_P^s))]^t$. The term $\text{tr}(\mathbf{J}^{obs}[1, 1])$ (resp. $\text{tr}(\mathbf{J}^{obs}[2, 2])$) represents the sum of the eigenvalues of the FIM of $\mathbf{p}_T(t_1)$ (resp. $\mathbf{p}_T(t_n)$), that is, the sum of the magnitudes of the corresponding ellipse. Also, the terms in the numerators of Eq. (7.4.4) are correction factors which avoid biased estimates \hat{r}_1 and \hat{r}_n . The derivation of Eq. (7.4.4) is given in Appendix 7.7.5.

It is worth remarking that Eq. (7.4.4) represents a rough approximation of $\{r_1, r_n\}$; however, as we will show in Sec. 7.5, the accuracy obtained is sufficient to determine values close to $\check{\mathbf{p}}_P(t_1)$ and $\check{\mathbf{p}}_P(t_n)$ in all the practical scenarios considered.

Choice of $\{\hat{\mathbf{i}}_1, \hat{\mathbf{i}}_n\}$

By defining $\mathbf{C}^{obs} \triangleq (\mathbf{J}^{obs})^{-1}$ and denoting $\mathbf{C}^{obs}[\ell, m]$ as the (ℓ, m) th $[2 \times 2]$ block matrix of \mathbf{C}^{obs} we have that:

$$\hat{\mathbf{i}}_1^p \triangleq \mathbf{e}^p(\mathbf{C}^{obs}[1, 1]; 1), \quad \hat{\mathbf{i}}_n^q \triangleq \mathbf{e}^q(\mathbf{C}^{obs}[2, 2]; 1); \quad p, q \in \{-1, 1\} \quad (7.4.7)$$

Note that $\mathbf{C}^{obs}[1, 1]$ (resp. $\mathbf{C}^{obs}[2, 2]$) is a lower bound of the covariance matrix of any unbiased estimator of $\mathbf{p}_T(t_1)$ (resp. $\mathbf{p}_T(t_n)$) and thus the eigenvector

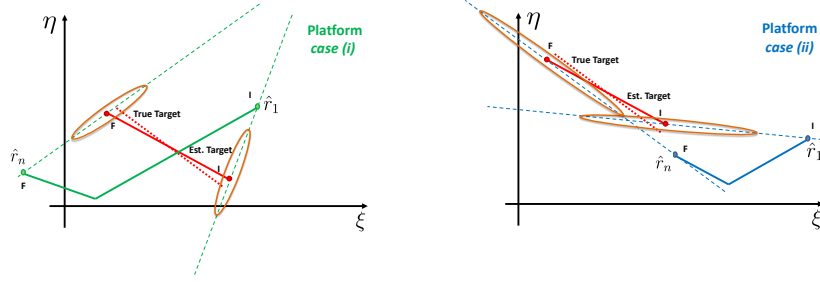


Figure 7.4.1: Graphical description of Eq. (7.4.8): cases (i) and (ii) satisfying it.

corresponding to the largest eigenvalue represents the axis of maximum uncertainty, i.e., the one along the range between the platform and the target. The indices $\{p, q\}$ underline the incomplete information about $\{\hat{i}_1, \hat{i}_n\}$ contained in the FIM, which intrinsically leads to a sign ambiguity (because if \mathbf{e} is an eigenvector, so is $-\mathbf{e}$). The derivation of Eq. (7.4.7) is given in Appendix 7.7.6.

It is apparent that the sign ambiguity in Eq. (7.4.7) would lead to four possible pairs $\{\hat{i}_1^p, \hat{i}_n^q\}$. However, in practical scenarios the sign-ambiguity leads to only two pairs if we assume that the following vector equation *has no solution*:

$$\zeta \cdot \hat{\mathbf{p}}_T(t_1) + (1 - \zeta) \cdot \hat{\mathbf{p}}_T(t_n) = \mathbf{p}_P(t_i); \quad \forall t_i \in \mathcal{T}, \zeta \in \mathbb{R} \quad (7.4.8)$$

Eq. (7.4.8) admits solutions only in the following cases⁵: (i) the platform trajectory and the target trajectory have at least one crossing point; (ii) the platform is crossing the line representing the direction of the CV target trajectory, determined by its velocity vector $\hat{\mathbf{v}}_T$. These scenarios are graphically depicted in Fig. 7.4.1. While the former case represents a totally unrealistic scenario (the platform and the target would be too near), it can be shown that the latter is of little interest, since such a platform trajectory would lead to poor observability of the target [122].

Therefore the only two admissible pairs are obtained as follows: let us define a vector $\{\mathbf{u} \in \mathbb{R}^2 : \mathbf{u} \perp \hat{\mathbf{v}}_T, \|\mathbf{u}\| = 1\}$ and take the pairs $\{\hat{i}_1^p, \hat{i}_n^q\}$ verifying

$$\text{sign} \langle \hat{i}_1^p, \mathbf{u} \rangle = \text{sign} \langle \hat{i}_n^q, \mathbf{u} \rangle, \quad p, q \in \{-1, 1\} \quad (7.4.9)$$

In the following we denote as $\{\hat{i}_1^g, \hat{i}_n^g\}$, $g \in \{1, 2\}$, the resulting two pairs. Thus exploiting Eqs. (7.4.4), (7.4.7) and (7.4.9) we obtain:

$$\hat{\mathbf{p}}_P^g(t_1) = \hat{\mathbf{p}}_T(t_1) + \hat{r}_1 \cdot \hat{i}_1^g, \quad (7.4.10)$$

$$\hat{\mathbf{p}}_P^g(t_n) = \hat{\mathbf{p}}_T(t_n) + \hat{r}_n \cdot \hat{i}_n^g, \quad g \in \{1, 2\} \quad (7.4.11)$$

⁵Although Eq. (7.4.8) refers to $\hat{\mathbf{p}}_T(t_i)$, the same applies to $\mathbf{p}_T(t_i)$ under the assumption that the platform obtains a reasonably “confident” estimate of the target.

7.4.2 Choice of turning position vector $\hat{\boldsymbol{p}}_P(t_k)$

As pointed out previously, Eqs. (7.4.10) and (7.4.11) define two pairs $\{\hat{\boldsymbol{p}}_P^g(t_1), \hat{\boldsymbol{p}}_P^g(t_n)\}$, $g \in \{1, 2\}$. Hence for each $\{\hat{\boldsymbol{p}}_P^g(t_1), \hat{\boldsymbol{p}}_P^g(t_n)\}$ a corresponding $\hat{\boldsymbol{p}}_P^g(t_k)$ needs to be computed. In this case it can be shown that $\hat{\boldsymbol{p}}_P^g(t_k)$ cannot be chosen only relying on the geometric properties of the FIM, as opposed to $\hat{\boldsymbol{p}}_P(t_1)$ and $\hat{\boldsymbol{p}}_P(t_n)$. Rather a $\pm \frac{\pi}{2}$ platform turn initial assumption is made and the turn-sign ambiguity is solved exploiting the coarse information in the FIM. By defining $t_m \triangleq \min_{i \in \mathcal{I}} \|t_i - \frac{t_n - t_1}{2}\|$, such a vector is obtained as (we drop the superscript g)

$$\hat{\boldsymbol{p}}_P(t_k) = \boldsymbol{\rho}_q(t_k) \quad (7.4.12)$$

$$\boldsymbol{\rho}_\ell(t_k) \triangleq \hat{\boldsymbol{p}}_P(t_1) + \frac{(t_k - t_1)}{(t_n - t_k)} \|\hat{\boldsymbol{p}}_P(t_n) - \hat{\boldsymbol{p}}_P(t_1)\| \cos(\nu) \begin{bmatrix} \sin(\psi_\ell) \\ \cos(\psi_\ell) \end{bmatrix} \quad (7.4.13)$$

$$\psi_\ell \triangleq \arctan_2(\hat{\boldsymbol{p}}_P(t_n) - \hat{\boldsymbol{p}}_P(t_1)) + \ell \cdot \left(\frac{\pi}{2} - \nu\right) \quad (7.4.14)$$

$$\nu \triangleq \arctan\left(\frac{t_k - t_1}{t_n - t_k}\right) \quad (7.4.15)$$

$$q \triangleq \arg \min_{\ell \in \{-1, 1\}} \|\tilde{\boldsymbol{p}}_P(t_m) - \boldsymbol{\rho}_\ell(t_m)\| \quad (7.4.16)$$

where $\boldsymbol{\rho}_\ell(t_m)$ is the position vector at t_m of the two-leg trajectory described by $\{\hat{\boldsymbol{p}}_P(t_1), \boldsymbol{\rho}_\ell(t_k), \hat{\boldsymbol{p}}_P(t_n)\}$, and $\tilde{\boldsymbol{p}}_P(t_m)$ is given by

$$\tilde{\boldsymbol{p}}_P(t_m) \triangleq \hat{\boldsymbol{p}}_T(t_m) + \tilde{r}_m \cdot \tilde{\boldsymbol{i}}_m \quad (7.4.17)$$

$$\tilde{r}_m \triangleq \sqrt{\frac{\check{\alpha}_\theta \sum_{i=1}^n \alpha_i (1 - \alpha_i)}{\text{tr}(\boldsymbol{J}^{obs}[1, 2])}} \quad (7.4.18)$$

$$\tilde{\boldsymbol{i}}_m \triangleq e^q(\boldsymbol{J}^{obs}[1, 2]; 1) \quad (7.4.19)$$

$$q \ni \text{sign} \langle e^j(\boldsymbol{J}^{obs}[1, 2]; 1), \boldsymbol{u} \rangle = \text{sign} \langle \hat{\boldsymbol{i}}_1, \boldsymbol{u} \rangle, \quad j \in \{-1, 1\}. \quad (7.4.20)$$

The complete derivation of the selection of $\hat{\boldsymbol{p}}_P^g(t_k)$ is given in Appendix 7.7.7.

7.4.3 Remarks on α_θ

The presented method computes $\hat{\boldsymbol{p}}_P^g = [\hat{\boldsymbol{p}}_P^g(t_1)^t \quad \hat{\boldsymbol{p}}_P^g(t_k)^t \quad \hat{\boldsymbol{p}}_P^g(t_n)^t]^t$, $g \in \{1, 2\}$, under the assumption that $\check{\alpha}_\theta$ is known at the target-friendly side. However, since $\check{\alpha}_\theta$ is known exactly only at the platform side, we will replace $\check{\alpha}_\theta$ with the variable α_θ in Eq. (7.4.4), thus leading to $\hat{\boldsymbol{p}}_P^g = \hat{\boldsymbol{p}}_P^g(\alpha_\theta)$, i.e., a continuum of initial guesses. Therefore, in order to obtain a (small) finite set of initial guesses we apply the following steps:

1. define a uniform grid of N_θ values on α_θ , constrained in the interval $[\alpha_{\theta, \min}, \alpha_{\theta, \max}]$ ⁶ and, after denoting the m -th value as $\alpha_\theta[m] = \alpha_{\theta, \min} +$

⁶The values $\{\alpha_{\theta, \min}, \alpha_{\theta, \max}\}$ are obtained automatically, once a reasonable range is given for parameters concurring with the definition of α_θ (e.g. σ_θ); an example will be given in Section 7.5.

Table 7.1: Parameters known at the platform side.

Parameter	Value	Unit
$t_n - t_1$	800	s
$t_i - t_{i-1}, i \in \mathcal{I} \setminus \{1\}$	4	s
$\check{\mathbf{x}}_P^s$ in scenario (i)	$[10^4 \quad 2 \cdot 10^4 \quad 7.1 \quad \frac{3}{4}\pi \quad \frac{\pi}{4}]^t$	$[m \quad m \quad m/s \quad rad \quad rad]^t$
$\check{\mathbf{x}}_P^s$ in scenario (ii)	$[10^4 \quad 2 \cdot 10^4 \quad 7.1 \quad -\frac{\pi}{4} \quad \frac{\pi}{4}]^t$	$[m \quad m \quad m/s \quad rad \quad rad]^t$
$q_2(\lambda v_g, P_D, g)$	$q_2(0.3, 0.9, 5) = 0.814$	dimensionless
σ_θ	1 (0.0175)	deg. (rad)
$\check{\alpha}_\theta = q_2/\sigma_\theta^2$	$2.6580 \cdot 10^3$	rad ⁻²

$\frac{m-1}{N_\theta-1}(\alpha_{\theta,\max} - \alpha_{\theta,\min})$, $m \in \mathcal{S}_\theta \triangleq \{1, \dots, N_\theta\}$, we compute $\hat{\mathbf{p}}_P^g(\alpha_\theta[m])$, $g \in \{1, 2\}$, $m \in \mathcal{S}_\theta$;

2. split up the set $\hat{\mathbf{p}}_P^g(\alpha_\theta[m])$, $m \in \mathcal{S}_\theta, g \in \{-1, 1\}$ into three subsets, corresponding to zones (a), (b) and (c), defined as follows⁷: (i) we split $\hat{\mathbf{p}}_P^g(\alpha_\theta[m])$ into two sets $\hat{\mathbf{p}}_P^{g-1}(\alpha_\theta[m])$ and $\hat{\mathbf{p}}_P^{g+1}(\alpha_\theta[m])$; (ii) by defining $\Gamma_{m,g} \triangleq \arctan_2(\hat{\mathbf{p}}_P^g(t_n, \alpha_\theta[m]) - \hat{\mathbf{p}}_P^g(t_1, \alpha_\theta[m]))$ we seek for $\{m^*, g^*\} \ni \text{sign}(\Gamma_{m^*,g^*}) \neq \text{sign}(\Gamma_{(m^*+1),g^*})$; (iii) we split the set corresponding to g^* as $\hat{\mathbf{p}}_P^{g^*}(\alpha_\theta[1 : m^*])$ and $\hat{\mathbf{p}}_P^{g^*}(\alpha_\theta[(m^* + 1) : N_\theta])$;
3. choose, into each defined subset, the guess corresponding to the maximum of Eq. (7.3.12), and we denote the obtained triple as $\{\hat{\mathbf{p}}_P^A, \hat{\mathbf{p}}_P^B, \hat{\mathbf{p}}_P^C\}$, which is fed (possibly in parallel) to a local optimization routine.

A good choice of N_θ can be obtained as follows. We first observe that the worst-case error in grid sampling of α_θ is given by $\Delta\varepsilon_\theta = \frac{(\alpha_{\theta,\max} - \alpha_{\theta,\min})}{2 \cdot (N_\theta - 1)}$; such error is negligible in Eq. (7.4.4) when $\check{\alpha}_\theta \gg \Delta\varepsilon_\theta$, thus leading to $N_\theta \gg \frac{\alpha_{\theta,\max} - \alpha_{\theta,\min}}{2\check{\alpha}_\theta} + 1$. However, since $\check{\alpha}_\theta$ is not known, we can consider the conservative inequality $\alpha_{\theta,\min} \gg \Delta\varepsilon_\theta$, which leads to $N_\theta \gg \frac{1}{2} \frac{\alpha_{\theta,\min} + \alpha_{\theta,\max}}{\alpha_{\theta,\min}}$.

7.5 Simulation Results

In this section we consider two scenarios, taken from [100], to corroborate the theoretical results presented and show the performance of the geometry-driven initial guess procedure. In both scenarios (i) and (ii) we assume the same $\hat{\mathbf{x}}_T$ (since the absolute position is irrelevant, only the relative geometry), while we consider two different $\check{\mathbf{x}}_P^s$ (see Table 7.1). In Table 7.1 we report the list of parameters known at the platform side; it is worth noting that $\check{\alpha}_\theta$ is completely specified by q_2 (obtained as in [118, Table II]) and σ_θ .

In Table 7.2 we report the list of parameters known at the target-friendly side, after intercepting the ML-PDA estimates. The interval $[\alpha_{\theta,\min}, \alpha_{\theta,\max}]$

⁷Note that the definition of the zones is arbitrary.

Table 7.2: Parameters known at the target-friendly side.

Parameter	Value	Unit
$t_n - t_1$	800	s
$t_i - t_{i-1}, i \in \mathcal{I} \setminus \{1\}$	4	s
$\hat{\mathbf{x}}_T$	$[15 \times 10^3 \quad 35 \times 10^3 \quad -10 \quad 5]^t$	$[m \quad m \quad m/s \quad m/s]^t$
k	101	dimensionless
q_2^*	$\in [0.652, 0.982]$	dimensionless
σ_θ^*	$\in [1, 2]$ ($\in [0.0175, 0.0350]$)	deg. (rad)
α_θ^*	$\in [532.2449, 3206.5]$	rad^{-2}

(needed to obtain the initial guesses in the three zones) is obtained as follows. We assume that the target-friendly entity possesses the coarse information $\sigma_\theta \in [1, 2]^\circ$ and $q_2 \in [0.652, 0.982]$ (under the assumptions⁸ that $\lambda v_g \in [0.1, 0.5]$, $P_D \in [0.8, 1]$, $g = 5$, cf. [118, Table II]), thus leading to $\alpha_\theta \in [532.2449, 3206.5] \text{rad}^{-2}$.

The guesses $\{\hat{\mathbf{p}}_P^A, \hat{\mathbf{p}}_P^B, \hat{\mathbf{p}}_P^C\}$, obtained with the approach described in Sec. 7.4 and setting $N_\theta = 5^9$, are given as input to the *Nelder–Mead simplex* method [123], which seeks for a local maximum of the objective function in Eq. (7.3.12). This local optimization routine has been chosen because it is a *derivative-free* method, as opposed to Newtonian and quasi-Newtonian local optimization routines. Such a choice avoids the evaluation of the Jacobian matrix (and also of the Hessian matrix, in the Newtonian approaches), which is composed of 45 entries, thus requiring high complexity resources.

Fig. 7.5.1 presents a plot of the platform and target trajectories for scenario (i), while Fig. 7.5.5 shows them for scenario (ii). It is worth noting that scenario (i) represents a low-observability case, while scenario (ii) has good observability, as shown through the 95% confidence ellipses of $\hat{\mathbf{p}}_T(t_1)$ and $\hat{\mathbf{p}}_T(t_n)$ in Figs. 7.5.1 and 7.5.5.

In Fig. 7.5.2 (resp. Fig. 7.5.6) we show the initial guesses corresponding to zones (a), (b) and (c) (defined in step 2) of Subsec. 7.4.3) of the platform trajectory (7.3.12). It is apparent that the initial guess corresponding to the true zone is near the true platform trajectory in both cases; incidentally a degree of similarity is also present in the guess corresponding to a different zone in scenario (ii) (zone (b), cf. Fig. 7.5.6). Nonetheless, in both cases the procedure produces a very good initial guess corresponding to the zone where $\check{\mathbf{x}}_P^s$ belongs, which bodes well for a local optimization routine.

The convergence properties of the algorithm are illustrated in Figs. 7.5.3 and 7.5.7 for the three different initial guesses, and the performance is analyzed

⁸Note that in Table 7.2 the asterisk indicates that those parameters are only needed to compute wise initial guesses for the local optimization procedure, but not for the maximization of Eq. (7.3.12).

⁹It is worth remarking that in such a case $\left[\frac{\alpha_{\theta, \max} - \alpha_{\theta, \min}}{2\alpha_\theta} + 1\right] \approx 1.5$, thus $N_\theta = 5$ for *this particular scenario* is a fair choice.

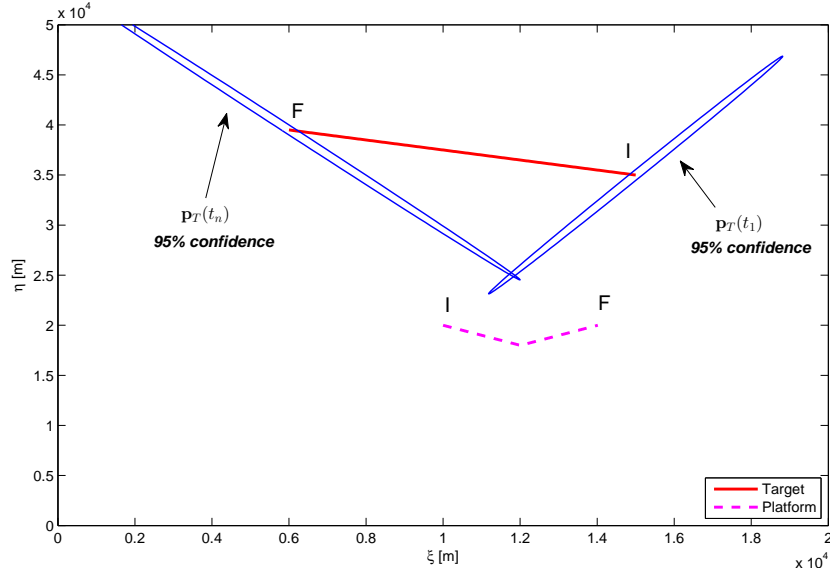


Figure 7.5.1: Platform and target trajectories considered for scenario (i).

in terms of the time-averaged root-square position-error (RSPE) defined as

$$\text{RSPE}(m) \triangleq \frac{1}{n} \sum_{i=1}^n \|\hat{\mathbf{p}}_P^m(t_i) - \check{\mathbf{p}}_P(t_i)\| \quad (7.5.1)$$

with $\hat{\mathbf{p}}_P^m(\cdot)$ denoting the output of the local optimization routine after m iterations. It is apparent that there is no monotonic decrease of the RSPE, since the maximization of the objection function is conducted w.r.t. the vector \mathbf{x}_P^s ; however convergence is observed with an acceptable number of iterations (recall that there is no need to compute Jacobian at each iteration and so each iteration is very light from a numerical point of view). In Fig. 7.5.4 (resp. Fig. 7.5.8) we finally report the platform corresponding to the maximum of the three outputs obtained with the Nelder-Mead method and the three different initial guesses. It is apparent that in both the scenarios the true platform trajectory is identified exactly (recall that this is a deterministic problem); such results confirm our conjecture on the uniqueness of the Eq. (7.3.7).

Turning time t_k sensitivity analysis

In this paragraph we will remove the assumption that the turning time t_k is known at the target-friendly side and we will show the effects of a grid search on the time index, denoted as k , for both scenarios (i) and (ii). Figs. 7.5.9 and 7.5.10 show the RSPE (obtained as the minimum along the three zones) as a function of the assumed index k . It is apparent that the true platform trajectory is still identifiable in both scenarios; however the RSPE w.r.t. to the index k

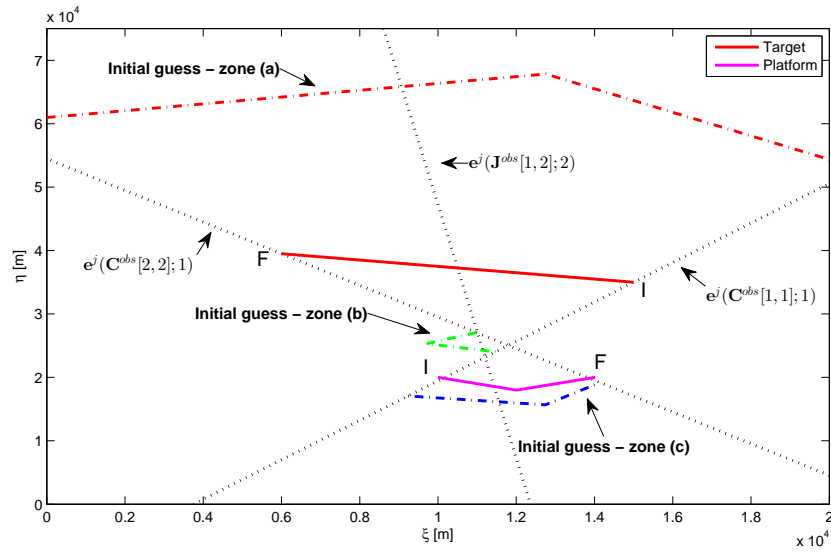


Figure 7.5.2: Geometry-driven initial guess procedure for platform trajectory in scenario (i).

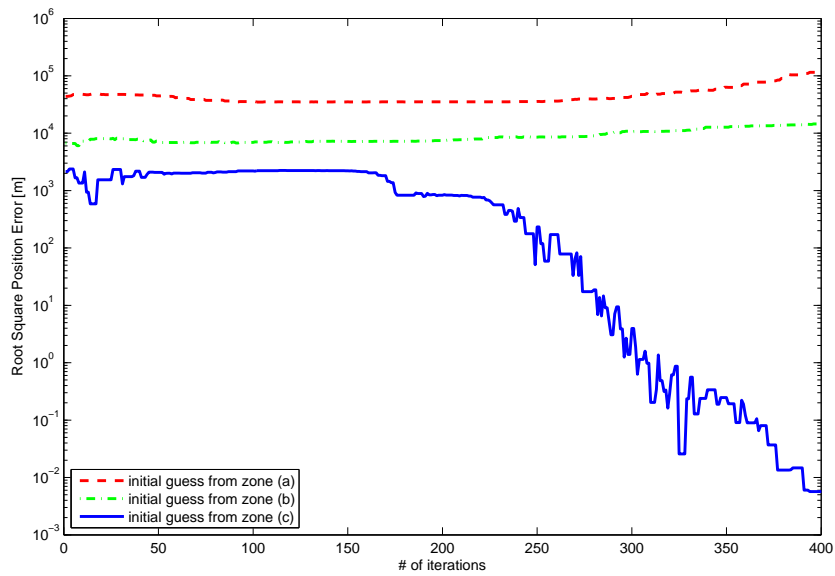


Figure 7.5.3: RSPE as a function of the number of iterations for scenario (i).

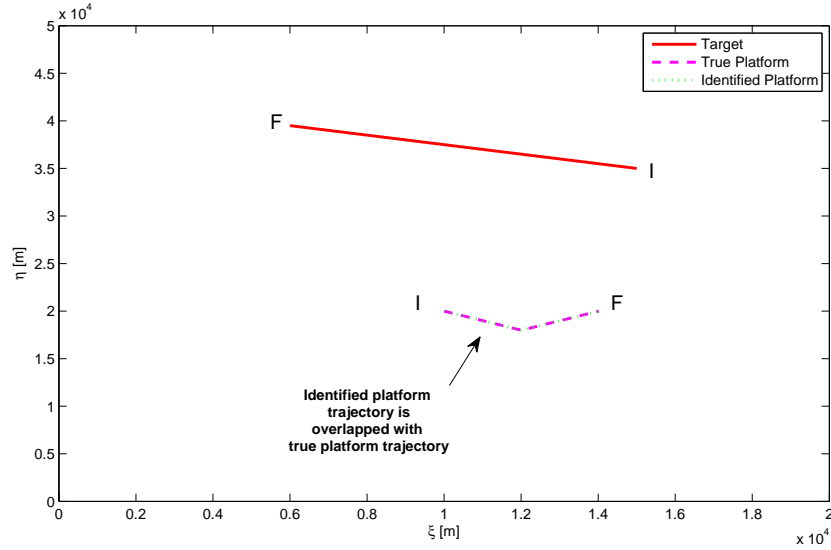


Figure 7.5.4: Identification results for scenario (i).

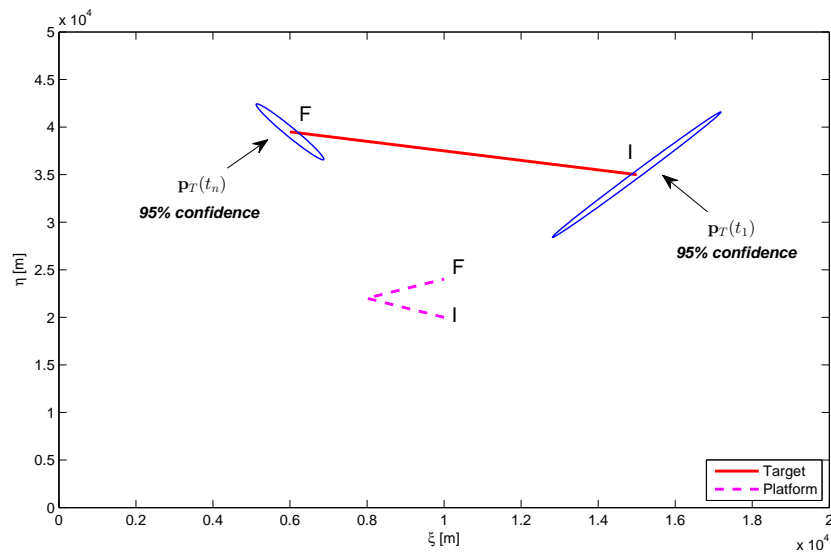


Figure 7.5.5: Platform and target trajectories considered for scenario (ii).

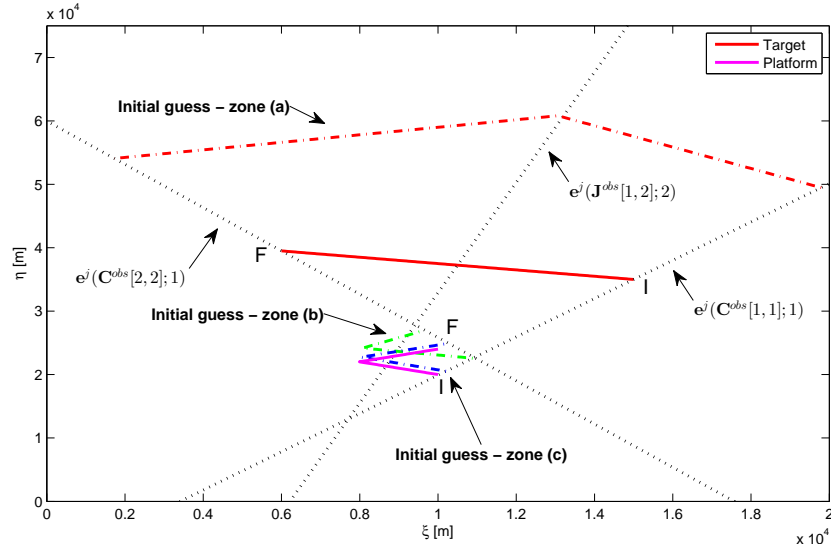


Figure 7.5.6: Geometry-driven initial guess procedure for platform trajectory in scenario (ii).

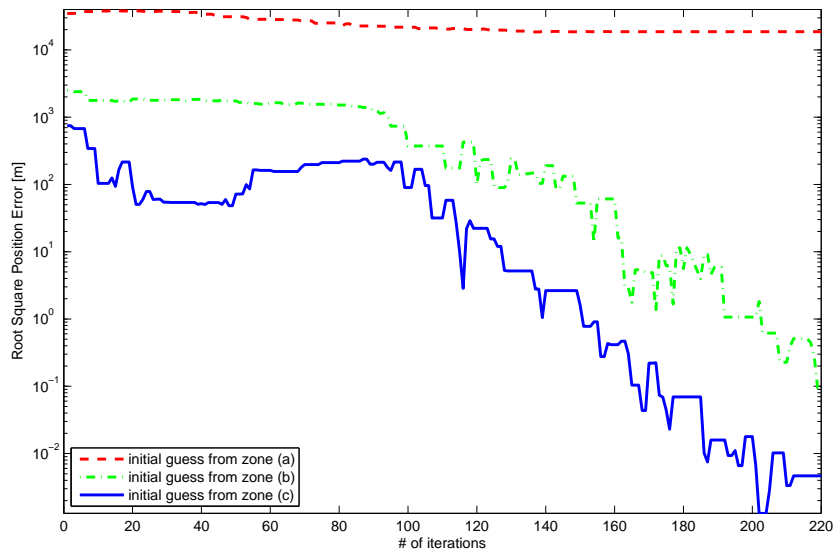


Figure 7.5.7: RSPE as a function of the number of iterations for scenario (ii).

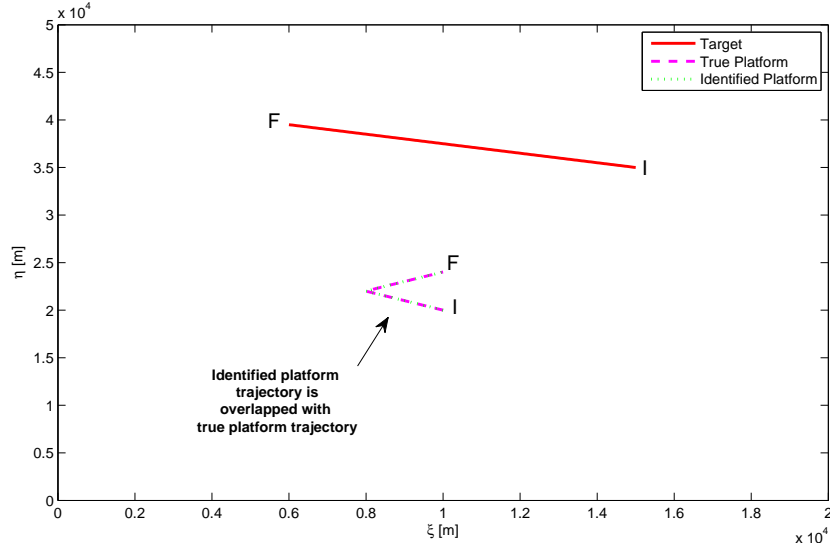


Figure 7.5.8: Identification results for scenario (ii).

is not a unimodal (discrete) function and therefore no “naïve” golden-search method can be applied to identify the platform; rather a parallel approach is needed.

Finally it is worth noting that the overall complexity Ξ of the proposed approach is given by:

$$\Xi \triangleq \sum_{k=1}^n \omega_a(k) + \omega_b(k) + \omega_c(k) \quad (7.5.2)$$

$$= \bar{\omega} \sum_{k=1}^n \chi_a(k) + \chi_b(k) + \chi_c(k) \quad (7.5.3)$$

$$\omega_i(k) \triangleq \bar{\omega} \cdot \chi_i(k) \quad (7.5.4)$$

where $\omega_i(k)$ denotes the complexity of the Nelder-Mead method with initial guess belonging to zone (i) and assumed index k . Note that $\omega_i(k)$ is simply given by the complexity of the single iteration $\bar{\omega}$, multiplied by the number of iterations $\chi_i(k)$. It is worth noting that there is no convergence theory to support an analysis providing an estimate for the number of iterations required to satisfy any reasonable accuracy constraint, given as a stopping condition [124]. Instead, regarding the complexity of the single iteration, it has been proved in [124] that $\bar{\omega}$ has a complexity function which is only dependent on the dimension of the search space, which is five-dimensional search in our case (since $\mathbf{x}_P^s \in \mathbb{R}^5$).

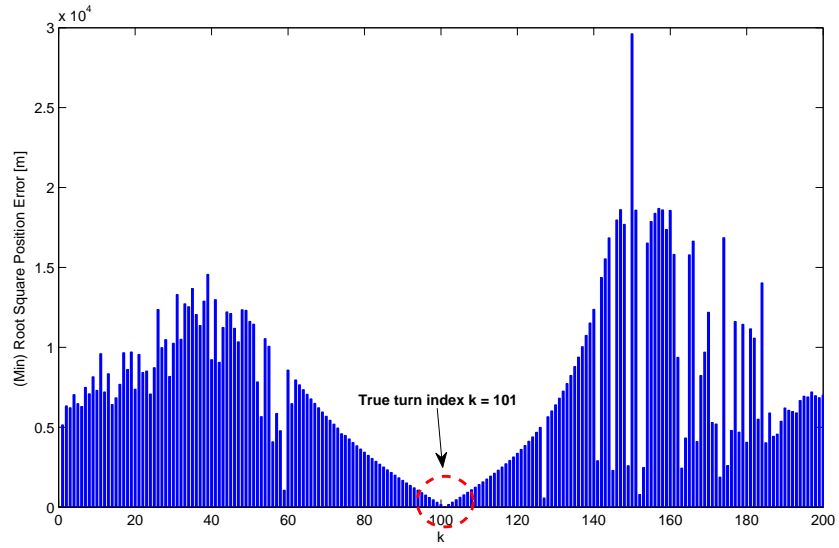


Figure 7.5.9: Sensitivity analysis in scenario (i).

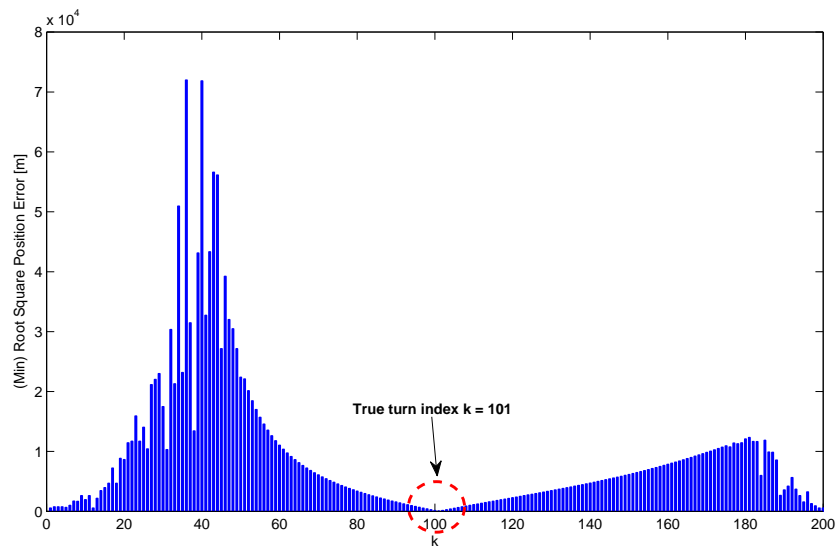


Figure 7.5.10: Sensitivity analysis in scenario (ii).

7.6 Conclusions

In this chapter we studied the problem of identifying the platform motion from its ML-PDA estimation results on an observed target—the estimation of the stealthy estimator. We have addressed only the common “two-leg” platform trajectory (see Figs. 7.1.1, 7.1.2, 7.2.1, 7.4.1 etc.); a similar analysis could be performed to other platform trajectories (such as a constant speed turn). We demonstrated that even a general “two-leg” platform motion model can lead to ambiguity in the identification; however, imposition of the constraint of constant speed motion ensures observability of the platform. We modelled the problem as a Frobenius norm minimization and we found a convenient objective function, exploiting the FIM repeated elements due to its symmetry and independent on the platform measurement-related parameters, namely α_θ . Also, we devised a procedure for the choice of a very small set of initial guesses as input for every numerical optimization procedure, on the basis of theoretical considerations on the geometry of the FIM. Finally, we corroborated the theoretical findings and we shown the effectiveness of the approach for the choice of the initial guesses, through simulation results. Future research will tackle incomplete (and noisy) intercepted information and different platform motion models. Note that although we have taken the ML-PDA as the underlying algorithm this is only in attempt at generality. In fact, the results presented here apply also to situations with any degree of measurement origin uncertainty, including of course the situation of a deterministic target observed via “clean” measurements without false alarms or missed detections.

7.7 Appendix

7.7.1 Proof of Proposition 13

To prove this proposition let us assume that there exists a solution $\{\mathbf{x}_P^*, \alpha_\theta^*\}$ such that $\mathbf{J}(\hat{\mathbf{x}}_T, \mathbf{x}_P^*, \alpha_\theta^*) = \mathbf{J}^{obs}$. Now let us consider the subspace $\{\mathbf{x}_P', \alpha_\theta'\}$ defined as

$$\{\mathbf{x}_P', \alpha_\theta'\} = \{\beta \mathbf{x}_P^* + (1 - \beta) \bar{\mathbf{x}}_E, \beta^2 \alpha_\theta^*\}, \quad \beta \in \mathbb{R} \quad (7.7.1)$$

where $\bar{\mathbf{x}}_E \triangleq [\hat{\mathbf{p}}_T(t_1)^t \quad \hat{\mathbf{v}}_T^t \quad \hat{\mathbf{v}}_T^t]^t$. The subspace contains the set of platform trajectories whose velocity and position vectors are linear combinations of the ones of the platform trajectory defined by \mathbf{x}_P^* and the estimated target trajectory.

In this case the corresponding $\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P')$ and $r_i(\hat{\mathbf{x}}_T, \mathbf{x}_P')$ have the explicit expressions

$$\begin{aligned} \theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P') &= \arctan \left(\frac{\hat{\xi}_T(t_i) - (\beta \xi_P(t_i) + (1 - \beta) \hat{\xi}_T(t_i))}{\hat{\eta}_T(t_i) - (\beta \eta_P(t_i) + (1 - \beta) \hat{\eta}_T(t_i))} \right) \\ &= \theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^*) \end{aligned} \quad (7.7.2)$$

$$\begin{aligned} r_i(\hat{\mathbf{x}}_T, \mathbf{x}'_P) &= \sqrt{(\hat{\xi}_T(t_i) - (\beta\xi_P(t_i) + (1-\beta)\hat{\xi}_T(t_i)))^2 + (\hat{\eta}_T(t_i) - (\beta\eta_P(t_i) + (1-\beta)\hat{\eta}_T(t_i)))^2} \\ &= \beta r_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^*) \end{aligned} \quad (7.7.3)$$

By plugging Eqs. (7.7.2) and (7.7.3) into Eq. (7.2.12) we obtain the equality

$$\nabla_{\mathbf{x}_T}(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}'_P)) = \frac{1}{\beta} \nabla_{\mathbf{x}_T}(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^*)) \quad (7.7.4)$$

Using Eq. (7.7.4) we can express $\mathbf{J}(\hat{\mathbf{x}}_T, \mathbf{x}'_P, \alpha'_\theta)$ as

$$\begin{aligned} \mathbf{J}(\hat{\mathbf{x}}_T, \mathbf{x}'_P, \alpha'_\theta) &= (\beta^2 \alpha_\theta^*) \sum_{i=0}^n \frac{1}{\beta} \nabla_{\mathbf{x}_T}(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^*)) \frac{1}{\beta} \nabla_{\mathbf{x}_T}^t(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^*)) \\ &= \mathbf{J}(\hat{\mathbf{x}}_T, \mathbf{x}_P^*, \alpha_\theta^*) \end{aligned} \quad (7.7.6)$$

Since $\{\mathbf{x}_P^*, \alpha_\theta^*\}$ and $\{\mathbf{x}'_P, \alpha'_\theta\}$ (independently of β) lead both to \mathbf{J}^{obs} , $\mathbf{J}(\cdot)$ does not represent a *one-to-one mapping*, which makes the platform-state vector \mathbf{x}_P unidentifiable.

7.7.2 Proof of Lemma 14

We start by observing that $\mathbf{J}(\hat{\mathbf{x}}_T, \mathbf{x}_P^s, \alpha_\theta)$ and $\nabla_{\mathbf{x}_T}(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^s))$ can be expressed similarly as Eqs. (7.2.9) and (7.2.12). Also, for notational convenience let us define

$$\mathbf{y}_i \triangleq \frac{1}{r_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^s)} \begin{bmatrix} \cos(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^s)) & -\sin(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^s)) \end{bmatrix}^t \quad (7.7.7)$$

and rewrite $\nabla_{\mathbf{x}_T}(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^s))$ as

$$\nabla_{\mathbf{x}_T}(\theta_i(\hat{\mathbf{x}}_T, \mathbf{x}_P^s)) = \begin{bmatrix} (1 - \alpha_i) \mathbf{y}_i \\ \alpha_i \mathbf{y}_i \end{bmatrix} \quad (7.7.8)$$

where α_i has been defined in Eq. (7.2.2). Substituting Eq. (7.7.8) into the explicit form of $\mathbf{J}(\hat{\mathbf{x}}_T, \mathbf{x}_P^s, \alpha_\theta)$, we get

$$\mathbf{J}(\hat{\mathbf{x}}_T, \mathbf{x}_P^s, \alpha_\theta) \triangleq \begin{bmatrix} \mathbf{J}[1, 1] & \mathbf{J}[1, 2] \\ \mathbf{J}[2, 1] & \mathbf{J}[2, 2] \end{bmatrix} \quad (7.7.9)$$

$$= \alpha_\theta \begin{bmatrix} \sum_{i=1}^n (1 - \alpha_i)^2 \mathbf{y}_i \mathbf{y}_i^t & \sum_{i=1}^n \alpha_i (1 - \alpha_i) \mathbf{y}_i \mathbf{y}_i^t \\ \sum_{i=1}^n \alpha_i (1 - \alpha_i) \mathbf{y}_i \mathbf{y}_i^t & \sum_{i=1}^n \alpha_i^2 \mathbf{y}_i \mathbf{y}_i^t \end{bmatrix} \quad (7.7.10)$$

where we have underlined the block-decomposition arising from Eq. (7.7.8) into $[2 \times 2]$ matrices $\mathbf{J}[\ell, m]$, $\ell, m \in \{1, 2\}$. Since we have that $\mathbf{J}[1, 2] = \mathbf{J}[2, 1]$, only three matrices contain non-repeated entries (i.e., the 4 entries of $\mathbf{J}[1, 2]$ or $\mathbf{J}[2, 1]$ can be neglected). Also, since $\mathbf{J}[1, 1]$, $\mathbf{J}[2, 1]$ and $\mathbf{J}[2, 2]$ are symmetric matrices, there is a repeated entry in each of them, thus leading to 3 other dependent entries. Therefore the 16 entries of the FIM actually contain only 9 independent elements. W.l.o.g. we consider here (and throughout the chapter), the following independent entries (we drop the dependence w.r.t. $\hat{\mathbf{x}}_T$, \mathbf{x}_P^s and α_θ) and we denote with \mathcal{C} the corresponding set of indices:

$$\{J_{1,1}, J_{2,2}, J_{3,3}, J_{4,4}, J_{1,2}, J_{1,3}, J_{1,4}, J_{2,4}, J_{3,4}\} \quad (7.7.11)$$

7.7.3 Proof of Proposition 15

We have shown, through Lemma 14, that $\mathbf{J}(\mathbf{x}_P^s, \alpha_\theta)$ has only 9 independent entries. Also, the dependent entries are simple repetitions of the corresponding independent entries, due to particular symmetry structure of the FIM (cfr. Eq. (7.7.9)).

The same argument extends to $\mathbf{D}(\mathbf{x}_P^s, \alpha_\theta) \triangleq (\mathbf{J}^{obs} - \mathbf{J}(\mathbf{x}_P^s, \alpha_\theta))$ and its entry-wise squared version $\bar{\mathbf{D}}(\mathbf{x}_P^s, \alpha_\theta)$ (i.e. $\bar{D}_{\ell,m}(\mathbf{x}_P^s, \alpha_\theta) = D_{\ell,m}^2(\mathbf{x}_P^s, \alpha_\theta)$), iff \mathbf{J}^{obs} retains the same property of symmetry; this is accomplished if \mathbf{J}^{obs} is a noise-free observed FIM, that is $\mathbf{J}^{obs} = \mathbf{J}(\check{\mathbf{x}}_P^s, \check{\alpha}_\theta)$.

By construction, the following equality holds

$$\|\mathbf{J}^{obs} - \mathbf{J}(\mathbf{x}_P^s, \alpha_\theta)\|_2 = \sum_{\ell=1}^4 \sum_{m=1}^4 \bar{D}_{\ell,m}(\mathbf{x}_P^s, \alpha_\theta) \quad (7.7.12)$$

This sum can be efficiently evaluated by considering only $(\ell, m) \in \mathcal{C}$, with \mathcal{C} being the set of the independent entries according to Eq. (7.7.11), and weighting them by the number of times they are repeated in the matrix $\bar{\mathbf{D}}$. Thus the l.h.s. of Eq. (7.7.12) can be rewritten as

$$\sum_{(\ell,m) \in \mathcal{C}} \sum_{(\ell,m) \in \mathcal{C}} W_{\ell,m} \bar{D}_{\ell,m}(\mathbf{x}_P^s, \alpha_\theta) = \sum_{(\ell,m) \in \mathcal{C}} \sum_{(\ell,m) \in \mathcal{C}} W_{\ell,m} (J_{\ell,m}^{obs} - J_{\ell,m}(\mathbf{x}_P^s, \alpha_\theta))^2 \quad (7.7.13)$$

By stacking the elements $J_{\ell,m}^{obs}$ (resp. $J_{\ell,m}(\mathbf{x}_P^s, \alpha_\theta)$), $(\ell, m) \in \mathcal{C}$, into the vector \mathbf{j}^{obs} (resp. $\mathbf{j}(\mathbf{x}_P^s, \alpha_\theta)$), and defining the diagonal matrix \mathbf{W} with elements $W_{\ell,m}$, we obtain the weighted form of Eq. (7.3.9). Finally, it is straightforward to show that \mathbf{W} has the expression

$$\mathbf{W} = \text{diag} \left([1 \ 1 \ 1 \ 1 \ 2 \ 2 \ 4 \ 2 \ 2]^t \right) \quad (7.7.14)$$

where the weight equal to 4 accounts for the repeated elements along the minor diagonal, while the weights equal to 2 account for the remaining symmetric entries into each block matrix $\mathbf{J}[\ell, m]$, $\ell, m \in \{1, 2\}$, and between the matrices $\mathbf{J}[1, 2]$ and $\mathbf{J}[2, 1]$.

7.7.4 Proof of Proposition 16

Let us rewrite Eq. (7.3.11) here for convenience:

$$\mathcal{F}(\mathbf{x}_P^s, \alpha_\theta) = [\mathbf{j}^{obs} - \mathbf{j}_u(\mathbf{x}_P^s)\alpha_\theta]^t \mathbf{W} [\mathbf{j}^{obs} - \mathbf{j}_u(\mathbf{x}_P^s)\alpha_\theta] \quad (7.7.15)$$

For a fixed \mathbf{x}_P^s it is easy to check that the observation model in Eq. (7.7.15) is linear in α_θ . Therefore given \mathbf{x}_P^s , $\hat{\alpha}_\theta$ can be found as the solution of a standard least squares problem [68] as

$$\hat{\alpha}_\theta = \frac{\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}^{obs}}{\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)} \quad (7.7.16)$$

Then substituting Eq. (7.7.16) in Eq. (7.7.15) we obtain the expression for $\mathcal{F}(\mathbf{x}_P^s, \alpha_\theta)$

$$\mathcal{F}(\mathbf{x}_P^s, \hat{\alpha}_\theta) = \left[\mathbf{j}^{obs} - \mathbf{j}_u(\mathbf{x}_P^s) \frac{\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}^{obs}}{\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)} \right]^t \mathbf{W} \left[\mathbf{j}^{obs} - \mathbf{j}_u(\mathbf{x}_P^s) \frac{\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}^{obs}}{\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)} \right] \quad (7.7.17)$$

Exploiting the *orthogonality principle* of linear least squares [68], Eq. (7.7.17) reduces to

$$\mathcal{F}(\mathbf{x}_P^s, \hat{\alpha}_\theta) = (\mathbf{j}^{obs})^t \left(\mathbf{W} - \frac{\mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s) \mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W}}{\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)} \right) \mathbf{j}^{obs} \quad (7.7.18)$$

Since $(\mathbf{j}^{obs})^t \mathbf{W} \mathbf{j}^{obs}$ in Eq. (7.7.18) does not depend on \mathbf{x}_P^s , it is irrelevant in the minimization. Thus neglecting it and considering the opposite of the remaining term, we can equivalently maximize:

$$\mathcal{G}(\mathbf{x}_P^s) \triangleq (\mathbf{j}^{obs})^t \frac{\mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s) \mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W}}{\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)} \mathbf{j}^{obs} \quad (7.7.19)$$

Finally, by defining $\mathbf{c}(\mathbf{x}_P^s) \triangleq \frac{\mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)}{[\mathbf{j}_u(\mathbf{x}_P^s)^t \mathbf{W} \mathbf{j}_u(\mathbf{x}_P^s)]^{1/2}}$, we can rewrite Eq. (7.7.19) as

$$\mathcal{G}(\mathbf{x}_P^s) = (\mathbf{j}^{obs})^t \mathbf{c}(\mathbf{x}_P^s) \mathbf{c}(\mathbf{x}_P^s)^t \mathbf{j}^{obs} \quad (7.7.20)$$

$$= \langle \mathbf{j}^{obs}, \mathbf{c}(\mathbf{x}_P^s) \rangle^2 \quad (7.7.21)$$

which concludes the proof.

7.7.5 Choice of $\{\hat{r}_1, \hat{r}_n\}$

In this section we will derive the expressions in Eq. (7.4.4). For the sake of simplicity we will use the short-hand notations $r_i \triangleq r_i(\hat{\mathbf{x}}_T, \check{\mathbf{x}}_P^s)$ and $\theta_i \triangleq \theta_i(\hat{\mathbf{x}}_T, \check{\mathbf{x}}_P^s)$. We start by considering the block decomposition of \mathbf{J}^{obs} , through Eq. (7.7.9). Let us focus in particular on $\mathbf{J}^{obs}[1, 1]$ and $\mathbf{J}^{obs}[2, 2]$, i.e. the diagonal blocks, whose explicit expressions are given by

$$\mathbf{J}^{obs}[1, 1] = \check{\alpha}_\theta \sum_{i=1}^n \frac{(1 - \alpha_i)^2}{r_i^2} \underbrace{\begin{bmatrix} \cos^2(\theta_i) & -1/2 \sin(2\theta_i) \\ -1/2 \sin(2\theta_i) & \sin^2(\theta_i) \end{bmatrix}}_{\triangleq \mathbf{D}_i} \quad (7.7.22)$$

$$\mathbf{J}^{obs}[2, 2] = \check{\alpha}_\theta \sum_{i=1}^n \frac{\alpha_i^2}{r_i^2} \underbrace{\begin{bmatrix} \cos^2(\theta_i) & -1/2 \sin(2\theta_i) \\ -1/2 \sin(2\theta_i) & \sin^2(\theta_i) \end{bmatrix}}_{\triangleq \mathbf{D}_i} \quad (7.7.23)$$

note that an analogous expression holds for $\mathbf{J}^{obs}[1, 2]$, i.e. $\mathbf{J}^{obs}[1, 2] = \check{\alpha}_\theta \sum_{i=1}^n \frac{\alpha_i(1 - \alpha_i)}{r_i^2} \mathbf{D}_i$. It can be readily shown, through Eqs. (7.7.22) and (7.7.23), that $\text{tr}(\mathbf{J}^{obs}[\ell, \ell])$,

$\ell \in \{1, 2\}$, are given by

$$\text{tr}(\mathbf{J}^{obs}[1, 1]) = \check{\alpha}_\theta \left[\frac{1}{r_1^2} + \sum_{i=2}^n \frac{(1 - \alpha_i)^2}{r_i^2} \right] \quad (7.7.24)$$

$$\text{tr}(\mathbf{J}^{obs}[2, 2]) = \check{\alpha}_\theta \left[\frac{1}{r_n^2} + \sum_{i=1}^{n-1} \frac{\alpha_i^2}{r_i^2} \right] \quad (7.7.25)$$

From inspection of Eqs. (7.7.24) and (7.7.25), it is apparent that each trace is a weighted sum (scaled by $\check{\alpha}_\theta$) of $\frac{1}{r_i^2}$. Note that, by definition, the following inequalities hold $\forall i \in \mathcal{I}$:

$$(1 - \alpha_i)^2 < (1 - \alpha_{i+1})^2, \quad \alpha_{i+1}^2 < \alpha_i^2 \quad (7.7.26)$$

By exploiting them, we have that in $\text{tr}(\mathbf{J}^{obs}[1, 1])$ (resp. $\text{tr}(\mathbf{J}^{obs}[2, 2])$) the term $\frac{1}{r_1^2}$ (resp. $\frac{1}{r_n^2}$) receives the highest weight, while the term $\frac{1}{r_n^2}$ (resp. $\frac{1}{r_1^2}$) contributes to the sum with zero weight. To obtain a good approximation (and avoid biased estimates) of $\{r_1, r_n\}$ we first consider convex combination counterparts of Eqs. (7.7.24) and (7.7.25), since the positive weights in $\text{tr}(\mathbf{J}^{obs}[1, 1])$ and $\text{tr}(\mathbf{J}^{obs}[2, 2])$ do not satisfy the normalization property (i.e. $\sum_{i=1}^n (1 - \alpha_i)^2 \neq 1$ and $\sum_{i=1}^n \alpha_i^2 \neq 1$). The reasons are twofold: (i) the weights α_i and $(1 - \alpha_i)$ appear in squared form in Eqs. (7.7.24) and (7.7.25); and (ii) even α_i and $(1 - \alpha_i)$ do not sum to one¹⁰. For such a reason we normalize $\text{tr}(\mathbf{J}^{obs}[1, 1])$ (resp. $\text{tr}(\mathbf{J}^{obs}[2, 2])$) by $\sum_{i=1}^n (1 - \alpha_i)^2$ (resp. $\sum_{i=1}^n \alpha_i^2$). Then, we note that if

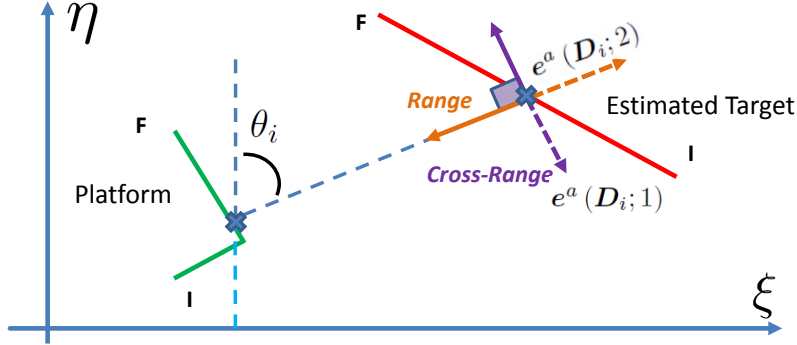
$$\begin{aligned} (a) \quad r_i &\approx r_1, \quad i \ni \alpha_i^2 \ll 1 \\ (b) \quad r_j &\approx r_n, \quad j \ni (1 - \alpha_j)^2 \ll 1 \end{aligned} \quad (7.7.27)$$

the term $\frac{\text{tr}(\mathbf{J}^{obs}[1, 1])}{\sum_{i=1}^n (1 - \alpha_i)^2}$ (resp. $\frac{\text{tr}(\mathbf{J}^{obs}[2, 2])}{\sum_{i=1}^n \alpha_i^2}$) well approximates $\frac{\check{\alpha}_\theta}{r_1^2}$ (resp. $\frac{\check{\alpha}_\theta}{r_n^2}$), thus leading to Eq. (7.4.4); therefore the value of $\text{tr}(\mathbf{J}^{obs}[1, 1])$ (resp. $\text{tr}(\mathbf{J}^{obs}[2, 2])$) will be high when the platform trajectory is near to the target at the beginning (resp. at the end) of the observation time. The conditions in Eq. (7.7.27) reflect the assumption that ranges at early (resp. late) time samples have non-negligible weights in Eq. (7.7.24) (resp. Eq. (7.7.25)) but they are very similar to r_1 (resp. r_n), which is a reasonable assumption in sonar tracking. Also, note that the mentioned condition is weaker than $r_i \approx r, \forall i \in \mathcal{I}$, (i.e. an approximately constant range assumption) and includes it as a more restrictive case.

7.7.6 Choice of $\{\hat{\mathbf{i}}_1, \hat{\mathbf{i}}_n\}$

The purpose of this Appendix is to show that: (i) \mathbf{J}^{obs} contains only *incomplete* information about $\{\mathbf{i}_1, \mathbf{i}_n\}$, meaning that it is not possible to extract these

¹⁰For example in the case of uniform sampling it holds $\sum_{i=1}^n \alpha_i = \sum_{i=1}^n (1 - \alpha_i) = \frac{n}{2}$; thus sum grows proportionally with the sample rate. The growth of $\sum_{i=1}^n \alpha_i$ and $\sum_{i=1}^n (1 - \alpha_i)$ with the sample rate is present also under the more general non-uniform sampling assumption, but in the latter case $\sum_{i=1}^n \alpha_i \neq \sum_{i=1}^n (1 - \alpha_i) \neq \frac{n}{2}$.


 Figure 7.7.1: Eigenvectors ambiguity in the choice of $\{\hat{\mathbf{i}}_1, \hat{\mathbf{i}}_n\}$.

vectors without ambiguity; (ii) a good approach to extract such information is represented by Eq. (7.4.7).

For this purpose, let us consider the definition of \mathbf{D}_i in Eqs. (7.7.22) and (7.7.23). It can be readily shown that each \mathbf{D}_i has eigenvalues $\{\lambda_{1,i}, \lambda_{2,i}\} = \{1, 0\}$ and that the corresponding (orthogonal) eigenvectors are

$$\mathbf{e}^a(\mathbf{D}_i; 1) = a \begin{bmatrix} \cos(\theta_i) \\ -\sin(\theta_i) \end{bmatrix}, \quad \mathbf{e}^a(\mathbf{D}_i; 2) = a \begin{bmatrix} \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix}, \quad a \in \{-1, 1\} \quad (7.7.28)$$

The pair $\{\lambda_{1,i}, \mathbf{e}^a(\mathbf{D}_i; 1)\}$ corresponds to the *cross-range direction* at t_i , while $\{\lambda_{2,i}, \mathbf{e}^a(\mathbf{D}_i; 2)\}$ corresponds to the *range direction* at t_i . In fact, recall that each \mathbf{D}_i has an *informative contribution* along the *cross-range* direction, thus $\lambda_{1,i} = 1$ ¹¹; on the other hand each \mathbf{D}_i has *no informative contribution* along the *range* direction (since the range is estimated with at least two bearing measurements), that is $\lambda_{2,i} = 0$. For this reason the information related to \mathbf{i}_1 (resp. \mathbf{i}_n) is contained in $\mathbf{e}^a(\mathbf{D}_1; 2)$ (resp. $\mathbf{e}^a(\mathbf{D}_n; 2)$). Note that however the sign-ambiguity in the definition of $\mathbf{e}^a(\mathbf{D}_1; 2)$ (resp. $\mathbf{e}^a(\mathbf{D}_n; 2)$) denotes the impossibility to recover exactly \mathbf{i}_1 (resp. \mathbf{i}_n), even if \mathbf{D}_1 (resp. \mathbf{D}_n) would have been perfectly available (a graphical description is given in Fig. (7.7.1)); this proves (i).

Before proceeding in the proof, it is worth noting that $\mathbf{e}^a(b_i \mathbf{D}_i; t) = \mathbf{e}^a(\mathbf{D}_i; t)$, $t \in \{1, 2\}$, $b_i \in \mathbb{R}^+$; for such a reason in the following w.l.o.g. we will search for a matrix \mathbf{K}_i which approximates well \mathbf{D}_i , except for a scale factor b_i , i.e. $\mathbf{K}_i \approx b_i \mathbf{D}_i$; once \mathbf{K}_i is obtained, the pairs $\{\hat{\mathbf{i}}_1^p, \hat{\mathbf{i}}_n^q\}$, $p, q \in \{-1, 1\}$, are simply evaluated by considering the least informative eigenvectors of \mathbf{K}_1 and \mathbf{K}_n , respectively.

¹¹Note that even if the informative contribution of \mathbf{D}_i is always $\lambda_{1,i} = 1$ along the cross-range direction, its contribution in the FIM is weighted by $\frac{1}{r_i^2}$.

To obtain good estimates of $b_1 \mathbf{D}_1$ and $b_n \mathbf{D}_n$ we can consider

$$\mathbf{K}_1 \triangleq \mathbf{J}^{obs}[1, 1] = \frac{\check{\alpha}_\theta}{r_1^2} \mathbf{D}_1 + \check{\alpha}_\theta \sum_{i=2}^{n-1} \frac{(1 - \alpha_i)^2}{r_i^2} \mathbf{D}_i \quad (7.7.29)$$

$$\mathbf{K}_n \triangleq \mathbf{J}^{obs}[2, 2] = \frac{\check{\alpha}_\theta}{r_n^2} \mathbf{D}_n + \check{\alpha}_\theta \sum_{i=1}^{n-1} \frac{\alpha_i^2}{r_i^2} \mathbf{D}_i \quad (7.7.30)$$

where in Eq. (7.7.29) (resp. Eq. (7.7.30)) we stress the (scaled) contribution of \mathbf{D}_1 (resp. \mathbf{D}_n) w.r.t. the spurious terms, i.e. \mathbf{D}_i , $i \in \mathcal{I} \setminus \{1\}$ (resp. $i \in \mathcal{I} \setminus \{n\}$). Exploiting again the inequalities among α_i and the assumptions of Eq. (7.7.27), it can be shown that $\frac{\check{\alpha}_\theta}{r_1^2} \mathbf{D}_1$ (resp. $\frac{\check{\alpha}_\theta}{r_n^2} \mathbf{D}_n$) is well approximated by Eq. (7.7.29) (resp. Eq. (7.7.30)); note that in this case convex combination counterparts are not needed because eigenvectors are not changed by a scaling factor. However, we will show hereinafter that a better estimate of $b_i \mathbf{D}_i$ can be obtained.

In fact let us consider $\mathbf{C}^{obs} \triangleq (\mathbf{J}^{obs})^{-1}$ and denote $\mathbf{C}^{obs}[\ell, m]$ as the (ℓ, m) th $[2 \times 2]$ block matrix of \mathbf{C}^{obs} . By exploiting the block-wise inversion formula [125] we obtain

$$(\mathbf{C}^{obs}[1, 1])^{-1} = \mathbf{J}^{obs}[1, 1] - \mathbf{J}^{obs}[1, 2] \mathbf{J}^{obs}[2, 2]^{-1} \mathbf{J}^{obs}[1, 2] \quad (7.7.31)$$

$$(\mathbf{C}^{obs}[2, 2])^{-1} = \mathbf{J}^{obs}[2, 2] - \mathbf{J}^{obs}[1, 2] \mathbf{J}^{obs}[1, 1]^{-1} \mathbf{J}^{obs}[1, 2] \quad (7.7.32)$$

Exploiting the expression for $\mathbf{J}[\ell, m]$, $\ell, m \in \{1, 2\}$ as in Eqs. (7.7.22) and (7.7.23) and putting in evidence the scaled contribution of \mathbf{D}_1 (resp. \mathbf{D}_n), we get:

$$(\mathbf{C}^{obs}[1, 1])^{-1} = \frac{\check{\alpha}_\theta}{r_1^2} \mathbf{D}_1 + \check{\alpha}_\theta \left\{ \sum_{i=2}^{n-1} \frac{1}{r_i^2} \mathbf{D}_i [(1 - \alpha_i)^2 \mathbf{I}_2 - \alpha_i(1 - \alpha_i) \mathbf{T}_1] \right\} \quad (7.7.33)$$

$$(\mathbf{C}^{obs}[2, 2])^{-1} = \frac{\check{\alpha}_\theta}{r_n^2} \mathbf{D}_n + \check{\alpha}_\theta \left\{ \sum_{i=2}^{n-1} \frac{1}{r_i^2} \mathbf{D}_i [\alpha_i^2 \mathbf{I}_2 - \alpha_i(1 - \alpha_i) \mathbf{T}_n] \right\} \quad (7.7.34)$$

where \mathbf{T}_1 and \mathbf{T}_n are defined respectively as

$$\mathbf{T}_1 \triangleq \left[\left(\sum_{\ell=2}^n \frac{\alpha_\ell^2}{r_\ell^2} \mathbf{D}_\ell \right)^{-1} \sum_{j=2}^{n-1} \frac{\alpha_j(1 - \alpha_j)}{r_j^2} \mathbf{D}_j \right] \quad (7.7.35)$$

$$\mathbf{T}_n \triangleq \left[\left(\sum_{\ell=1}^{n-1} \frac{(1 - \alpha_\ell)^2}{r_\ell^2} \mathbf{D}_\ell \right)^{-1} \sum_{j=2}^{n-1} \frac{\alpha_j(1 - \alpha_j)}{r_j^2} \mathbf{D}_j \right] \quad (7.7.36)$$

It is apparent how each spurious term in the braces of Eqs. (7.7.33) and (7.7.34) (cf. with Eqs. (7.7.29) and (7.7.30)) is now filtered through the matrix gain \mathbf{T}_1 (resp. \mathbf{T}_n) with weight $\alpha_i(1 - \alpha_i)$. The matrix \mathbf{T}_1 (resp. \mathbf{T}_n) represents a “smoothing” factor (independent of t_i , cf. with Eqs. (7.7.29) and (7.7.30)). The

weights $\alpha_i(1 - \alpha_i)$ are such that there is a higher correction w.r.t. \mathbf{D}_i corresponding to the middle of the observation interval, while a little correction at the beginning or the end of the observation interval. However, while in the first case the correction tends to reduce the spurious term, in the latter case there is an increase of the error given by the spurious term.

To gain an intuition about the effect of matrix \mathbf{T}_1 (same considerations apply to \mathbf{T}_n in Eq. (7.7.34)) on Eq. (7.7.33) let us consider the case $r_i \approx r, \forall i \in \mathcal{I}$. In this case we have that $\mathbf{T}_1 \approx (\sum_{\ell=2}^n \alpha_\ell^2 \mathbf{D}_\ell)^{-1} \sum_{j=2}^{n-1} \alpha_j(1 - \alpha_j) \mathbf{D}_j$ and the sum of the spurious terms in Eq. (7.7.33) reduces to

$$\check{\alpha}_\theta \left\{ \frac{1}{r^2} \sum_{i=2}^{n-1} \mathbf{D}_i [(1 - \alpha_i)^2 \mathbf{I}_2 - \alpha_i(1 - \alpha_i) \mathbf{T}_1] \right\} \quad (7.7.37)$$

Thus the magnitude of \mathbf{T}_1 will depend on the ratio of the concentration of t_i at the middle of the observation interval by the concentration of t_i at the end of the observation interval; in fact a higher ratio will imply a lower distortion in the smoothing of residual terms.

Remark: Note that Eqs. (7.7.33) and (7.7.34) cannot be exploited to obtain better estimates of \hat{r}_1 and \hat{r}_n than the ones in Eq. (7.4.4), even if the spurious terms are smoothed in such a case. The reason is that a proper normalization factor to obtain a convex combination cannot be found, since it can be shown that such a value would be dependent on $\mathbf{D}_i, i \in \mathcal{I}$, which are clearly not available.

7.7.7 Choice of $\hat{\mathbf{p}}_p(t_k)$

In order to obtain $\hat{\mathbf{p}}_P(t_k)$ we will first seek an approximation of $\mathbf{p}_P(t_m)$ (denoted as $\check{\mathbf{p}}_P(t_m)$), where $t_m \triangleq \min_{i \in \mathcal{I}} \|t_i - \frac{t_n - t_1}{2}\|$, i.e. the nearest t_i to the middle of the observation interval. Similarly as $\{\check{\mathbf{p}}_P(t_1), \check{\mathbf{p}}_P(t_n)\}$ (cf. Eq. (7.4.2)), we can express $\check{\mathbf{p}}_P(t_m)$ as

$$\check{\mathbf{p}}_P(t_m) = \hat{\mathbf{p}}_T(t_m) + r_m \mathbf{i}_m \quad (7.7.38)$$

The estimate of r_m , denoted as \tilde{r}_m , is obtained, in analogy to Eq. (7.4.4), as

$$\tilde{r}_m \triangleq \sqrt{\frac{\check{\alpha}_\theta \sum_{i=1}^n \alpha_i(1 - \alpha_i)}{\text{tr}(\mathbf{J}^{obs}[1, 2])}} \quad (7.7.39)$$

$$\text{tr}(\mathbf{J}^{obs}[1, 2]) = \check{\alpha}_\theta \left[\frac{\alpha_m(1 - \alpha_m)}{r_m^2} + \sum_{i=1, i \neq m}^n \frac{\alpha_i(1 - \alpha_i)}{r_i^2} \right] \quad (7.7.40)$$

As opposed to the case of $\{\mathbf{i}_1, \mathbf{i}_n\}$, the estimate of \mathbf{i}_m , denoted as $\tilde{\mathbf{i}}_m$, is obtained exploiting $\mathbf{J}^{obs}[1, 2]$ as a rough estimate of $b_m \mathbf{D}_m$, similarly as in Eqs. (7.7.29)

and (7.7.30):

$$\tilde{\mathbf{i}}_m \triangleq \mathbf{e}^q (\mathbf{J}^{obs}[1, 2]; 2) \quad (7.7.41)$$

$$\mathbf{J}^{obs}[1, 2] = \check{\alpha}_\theta \left[\frac{\alpha_m(1-\alpha_m)}{r_m^2} \mathbf{D}_m + \sum_{i=1, i \neq m}^n \frac{\alpha_i(1-\alpha_i)}{r_i^2} \mathbf{D}_i \right] \quad (7.7.42)$$

$$q \ni \text{sign} \langle \mathbf{e}^j (\mathbf{J}^{obs}[1, 2]; 1), \mathbf{u} \rangle = \text{sign} \langle \hat{\mathbf{i}}_1, \mathbf{u} \rangle, \quad j \in \{-1, 1\} \quad (7.7.43)$$

The last line accounts for the sign ambiguity, in analogy to Eq. (7.4.9). The explicit form of $\tilde{\mathbf{p}}_P(t_m)$ is obtained by replacing $\{r_m, \mathbf{i}_m\}$ with $\{\tilde{r}_m, \tilde{\mathbf{i}}_m\}$ in Eq. (7.7.38). Some important remarks, about $\tilde{\mathbf{p}}_P(t_m)$, are in the following:

- Eq. (7.7.40) represents a weighted combination of the inverse squared ranges, in analogy to Eq. (7.4.4). However it is apparent that the weight corresponding to $\frac{1}{r_m^2}$, that is $\check{\alpha}_\theta \alpha_m (1-\alpha_m)$, is weaker in this case w.r.t. the weights of the spurious terms. This leads not only to a poorer estimate \tilde{r}_m , as opposed to \hat{r}_1 and \hat{r}_n , but also \tilde{r}_m will be biased toward the line between r_0 and r_n , leading to an estimated low-observable platform trajectory (which is not desirable as an initial guess);
- In Eq. (7.7.42) $\mathbf{J}^{obs}[1, 2]$ is used as a $b_m \mathbf{D}_m$ estimate, since a better estimate cannot be found, differently from $\{\mathbf{D}_1, \mathbf{D}_n\}$. In fact it can be shown that $(\mathbf{C}^{obs}[1, 2])^{-1}$ does not represent a better estimate of $b_m \mathbf{D}_m$. Therefore $\mathbf{J}^{obs}[1, 2]$ represents the only existing approximation of $b_m \mathbf{D}_m$.

The above considerations suggest that one select a reasonable estimate of $\check{\mathbf{p}}_P(t_k)$ according to a different approach, as described in the following.

In fact, we assume that the two legs form a $\pm \frac{\pi}{2}$ angle, since the platform needs to perform a maneuver that guarantees a good degree of observability; therefore $\{\check{\mathbf{p}}_P(t_1), \check{\mathbf{p}}_P(t_k), \check{\mathbf{p}}_P(t_n)\}$ will form a right triangle. Furthermore the sign ambiguity in the turn leads to the definition of two specular vectors, denoted as $\boldsymbol{\rho}_\ell(t_k)$, $\ell \in \{-1, 1\}$.

Since t_k is assumed known and s is constant during the two legs, it can be shown, after geometric considerations, that such vectors are given by:

$$\boldsymbol{\rho}_\ell(t_k) \triangleq \hat{\mathbf{p}}_P(t_1) + \frac{(t_k - t_1)}{(t_n - t_k)} \|\hat{\mathbf{p}}_P(t_n) - \hat{\mathbf{p}}_P(t_1)\| \cos(\nu) \begin{bmatrix} \sin(\psi_\ell) \\ \cos(\psi_\ell) \end{bmatrix} \quad (7.7.44)$$

$$\psi_\ell \triangleq \arctan_2(\hat{\mathbf{p}}_P(t_n) - \hat{\mathbf{p}}_P(t_1)) + \ell \cdot \left(\frac{\pi}{2} - \nu \right) \quad (7.7.45)$$

$$\nu \triangleq \arctan \left(\frac{t_k - t_1}{t_n - t_k} \right) \quad (7.7.46)$$

where ν and $\frac{(t_k - t_1)}{(t_n - t_k)} \|\hat{\mathbf{p}}_P(t_n) - \hat{\mathbf{p}}_P(t_1)\| \cos(\nu)$ represent the angle whose vertex is $\hat{\mathbf{p}}_P(t_n)$ and the distance between $\hat{\mathbf{p}}_P(t_1)$ and $\boldsymbol{\rho}_\ell(t_k)$, respectively. Finally, the unit vector $\begin{bmatrix} \sin(\psi_\ell) & \cos(\psi_\ell) \end{bmatrix}^t$ represents the direction from $\hat{\mathbf{p}}_P(t_1)$ to $\boldsymbol{\rho}_\ell(t_k)$ and accounts for the sign ambiguity in the turn, through the angle ψ_ℓ .

The ambiguity is resolved by exploiting the coarse information given by Eqs. (7.7.40) and (7.7.42):

$$\hat{\boldsymbol{p}}_P(t_k) = \boldsymbol{\rho}_q(t_m) \quad (7.7.47)$$

$$q = \arg \min_{\ell \in \{-1, 1\}} \|\tilde{\boldsymbol{p}}_P(t_m) - \boldsymbol{\rho}_\ell(t_m)\| \quad (7.7.48)$$

where $\boldsymbol{\rho}_\ell(t_m)$ denotes the position vector at t_m of the two-leg trajectory described by $\{\hat{\boldsymbol{p}}_P(t_1), \boldsymbol{\rho}_\ell(t_k), \hat{\boldsymbol{p}}_P(t_n)\}$.

Bibliography

- [1] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.
- [2] A. Benavoli, L. Chisci, B. Ristic, A. Farina, and A. Graziano, *Reasoning under uncertainty: from Bayes to Valuation Based Systems. Applications to target classification and threat evaluation*. Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze, Italy, Selex Sistemi Integrati, Sept. 2007, ISBN: 978-888665843-0.
- [3] P. Smets, “Belief functions: The disjunctive rule of combination and the generalized bayesian theorem,” *International Journal of Approximate Reasoning*, vol. 9, no. 1, pp. 1–35, August 1993.
- [4] —, “Decision making in the tbm: the necessity of the pignistic transformation,” *International Journal of Approximate Reasoning*, vol. 38, no. 2, pp. 133–147, February 2005.
- [5] B. Ristic and P. Smets, “Belief function theory on the continuous space with an application to model based classification,” in *Proceedings of 10th International Conference IPMU 2004*, July 2004, pp. 1119–1126.
- [6] —, “Target classification approach based on the belief function theory,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 2, pp. 574–583, April 2005.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.
- [8] E. T. Jaynes, “Information theory and statistical mechanics,” *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, May 1957.
- [9] —, “Information theory and statistical mechanics. ii,” *Phys. Rev.*, vol. 108, no. 2, pp. 171–190, Oct 1957.
- [10] —, *Probability Theory : The Logic of Science*. Cambridge University Press, 2003.
- [11] J. O. Berger, “The case for objective bayesian analysis,” *Bayesian Analysis*, vol. 3, pp. 385–402, 2006.

- [12] P. S. Laplace, *Théorie analytique des Probabilités*. Veuve Courcier, Paris, 1812.
- [13] R. E. Kass and L. Wasserman, “The selection of prior distributions by formal rules,” *Journal of the American Statistical Association*, vol. 91, no. 435, pp. 1343–1370, Sep. 1996.
- [14] H. Jeffreys, “An invariant form for the prior probability in estimation problems,” in *Proc. Royal Soc. A*, vol. 186, no. 1007, 1946, pp. 453–461.
- [15] J. M. Bernardo, “Reference posterior distributions for bayesian inference,” *J. R. Statist. Soc. B*, vol. 41, no. 2, pp. 113–147, 1979.
- [16] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. John Wiley & Sons, 1994.
- [17] J. O. Berger, J. M. Bernardo, and D. Sun, “The formal definition of reference priors,” *The Annals of Statistics*, vol. 37, no. 2, pp. 905–938, 2009.
- [18] ———, “Objective priors for discrete parameter spaces,” *submitted to J. Amer. Statist. Assoc.*, 2011.
- [19] A. Zellner, “Models, prior information, and bayesian analysis,” *Models, prior information, and Bayesian analysis*, vol. 75, no. 1, pp. 51 – 68, 1996.
- [20] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, H. U. Division of Research, Graduate School of Business Administration, Ed., 1961.
- [21] R. E. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley-Interscience, 2000.
- [22] A. Caticha, “Entropic inference,” *arXiv:1011.0723v1*, 2 Nov 2010.
- [23] V. Majerník, “Marginal probability distribution determined by the maximum entropy method,” *Rep. on Math. Ph.*, vol. 45, no. 2, pp. 171–181, 2000.
- [24] A. Caticha, “Maximum entropy, fluctuations and priors,” in *20th Int. Work. on Bayesian. Inf. and Max. Ent.*, vol. 568. AIP, 2001, pp. 94–105.
- [25] A. Caticha and R. Preuss, “Maximum entropy and bayesian data analysis: Entropic priors,” *Physical Review*, vol. E 70, pp. pp. 046 127:1–12, 2004.
- [26] A. Giffin and A. Caticha, “Updating probabilities with data and moments,” in *27th Int. Work. on Bayesian Inf. and Max. Ent.*, vol. 954, no. 1. AIP, 2007, pp. 74–84.

- [27] T. Neumann, “Bayesian inference featuring entropic priors,” in *27th Int. Work. on Bayesian Inf. and Max. Ent.*, vol. 954, no. 1. AIP, 2007, pp. 283–292.
- [28] B. J. Brewer and M. J. Francis, “Entropic priors and bayesian model selection,” *29th Int. Work. on Bayesian Inference and Maximum Entropy Meth. in Sc. and Eng., July 5-10, Oxford, Mississippi, USA, AIP*, pp. 179–186, 2009.
- [29] F. Palmieri and D. Ciuonzo, “Data fusion with entropic priors,” in *Proc. of 20th Work. on Neural Networks, WIRN 2010.*, ser. Frontiers in Artificial Intelligence and Applications, vol. Volume 226, Vietri sul Mare, Italy, 2010, pp. 107–114.
- [30] F. A. N. Palmieri and D. Ciuonzo, “Entropic priors for short-term stochastic process classification,” in *14th Int. Conf. on Information Fusion.*, Chicago, IL., July 5-8 2011.
- [31] D. Ciuonzo and F. A. N. Palmieri, “Entropic priors for hidden-markov model classification,” in *2011 IEEE Statistical Signal Processing Workshop (SSP)*, june 2011, pp. 613–616.
- [32] F. Palmieri and D. Ciuonzo, “Consistency of sequence classification with entropic priors,” in *Proc. of MAXENT2011 , 31st Int. Work. on Bayesian Inf. and Max. Ent. Meth. in Sc. and Eng.* Waterloo, Canada: AIP, July 10-15 2011.
- [33] C. C. Rodriguez, “Entropic priors for discrete probabilistic networks and for mixtures of gaussians models,” in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 617. AIP Conference Proceedings, 2002, pp. 410–432.
- [34] J. B. Paris, “Common sense and maximum entropy,” *Synthese*, vol. 117, no. 1, pp. 75–93, 1999.
- [35] A. V. Lazo and P. Rathie, “On the entropy of continuous probability distributions,” *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 120 – 122, jan 1978.
- [36] G. A. Darbellay and I. Vajda, “Entropy expressions for multivariate continuous distributions,” *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 709 –712, mar 2000.
- [37] J. M. V. Novak, I. Perfilieva, *Mathematical Principles of Fuzzy Logic*. Kluwer Academic Press, 1999.
- [38] M. I. Jordan, “Stat260 lecture notes,” *U. C. Berkeley.*, vol. <http://www.cs.berkeley.edu/~jordan/courses>, 2010.
- [39] M. Avriel, *Nonlinear Programming: Analysis and Methods*. Dover, 2003.

- [40] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, Inc., 2007.
- [41] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Fusion of decisions transmitted over fading channels in wireless sensor networks," in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers.*, vol. 2, Nov. 2002, pp. 1184–1188.
- [42] —, "Channel aware decision fusion in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3454–3458, Dec. 2004.
- [43] B. Chen, L. Tong, and P. K. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16–26, Jul. 2006.
- [44] R. Jiang, S. Misra, B. Chen, and A. Swami, "Robust suboptimal decision fusion in wireless sensor networks," in *IEEE Military Communications Conference, 2005. MILCOM 2005.*, vol. 4, Oct. 2005, pp. 2107–2113.
- [45] J. O. Berger, J. M. Bernardo, and D. Sun, "Reference priors for discrete parameter spaces," Technical Report, <http://www.uv.es/bernardo/publications.html>, Tech. Rep., 2010.
- [46] M. K. Banavar, A. D. Smith, C. Tepedelenlioglu, and A. Spanias, "Distributed detection over fading MACs with multiple antennas at the fusion center," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Mar. 2010, pp. 2894–2897.
- [47] X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection over MIMO channels in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4124–4140, Sep. 2008.
- [48] A. Lei and R. Schober, "Coherent Max-Log decision fusion in wireless sensor networks," *IEEE Transactions on Communications*, vol. 58, no. 5, pp. 1327–1332, May 2010.
- [49] R. Jiang and B. Chen, "Fusion of censored decisions in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2668–2673, Nov. 2005.
- [50] P. K. Varshney, *Distributed Detection and Data Fusion*, 1st ed. Springer-Verlag New York, Inc., 1996.
- [51] M. Gastpar, M. Vetterli, and P. Dragotti, "Sensing reality and communicating bits: a dangerous liaison," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 70–83, Jul. 2006.

- [52] J. F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 16–25, May 2007.
- [53] R. Niu, B. Chen, and P. K. Varshney, "Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1018–1027, Mar. 2006.
- [54] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 2, pp. 469–476, Apr. 1998.
- [55] S. K. Jayaweera, "Optimal bayesian data fusion and low-complexity approximations for distributed DS-CDMA wireless sensor networks in Rayleigh fading," in *Proceedings of 2005 International Conference on Intelligent Sensing and Information Processing (ICISIP)*., Jan. 2005, pp. 19–24.
- [56] J. S. Dyer and B. Natarajan, "Improved low-complexity fusion receivers for synchronous DS-CDMA sensor networks," in *Proceedings of 16th International Conference on Computer Communications and Networks. ICCCN 2007.*, Aug. 2007, pp. 1123–1126.
- [57] S. K. Jayaweera, "Large system decentralized detection performance under communication constraints," *IEEE Communications Letters*, vol. 9, no. 9, pp. 769–771, Sep. 2005.
- [58] S. Yiu and R. Schober, "Nonorthogonal transmission and noncoherent fusion of censored decisions," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 263–273, Jan. 2009.
- [59] K. Liu and A. M. Sayeed, "Asymptotically optimal decentralized type-based detection in wireless sensor networks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)*., vol. 3, may 2004, pp. 873–876.
- [60] —, "Optimal distributed detection strategies for wireless sensor networks," in *in 42nd Annual Allerton Conf. on Commun., Control and Comp.*, 2004.
- [61] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access over multiaccess fading channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 1081–1092, Mar. 2007.
- [62] T. Cui and C. Tellambura, "An efficient generalized sphere decoder for rank-deficient MIMO systems," *IEEE Communications Letters*, vol. 9, no. 5, pp. 423–425, May 2005.

- [63] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice Hall PTR, Jan. 1998.
- [64] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and V. Poor, *MIMO Wireless Communications*. Cambridge University Press, 2007.
- [65] D. J. Love and R. W. Heath Jr., “Equal gain transmission in multiple-input multiple-output wireless systems,” *IEEE Transactions on Communications*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.
- [66] P. Robertson, E. Villebrun, and P. Hoeher, “A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain,” in *IEEE International Conference on Communications (ICC '95)*, vol. 2, Seattle, Jun. 1995, pp. 1009–1013.
- [67] Z. Chair and P. K. Varshney, “Optimal data fusion in multiple sensor detection systems,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.
- [68] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*. Prentice Hall PTR, 1993.
- [69] A. Lei and R. Schober, “Coherent max-log decision fusion in wireless sensor networks,” *IEEE Transactions on Communications*, vol. 58, no. 5, pp. 1327–1332, may 2010.
- [70] C. Studer and H. Bölcskei, “Soft-input soft-output single tree-search sphere decoding,” vol. 56, pp. 4827–4842, 2010.
- [71] S. K. Jayaweera, “Bayesian fusion performance and system optimization for distributed stochastic gaussian signal detection under communication constraints,” *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1238–1250, april 2007.
- [72] M. K. Simon and M. S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, 1st ed. New York: Wiley, 2000.
- [73] M. O. Damen, H. El Gamal, and G. Caire, “On maximum-likelihood detection and the search for the closest lattice point,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.
- [74] M. K. Banavar, A. D. Smith, C. Tepedelenlioglu, and A. Spanias, “On the effectiveness of multiple antennas in distributed detection over fading MACs,” vol. 11, no. 5, pp. 1744–1752, May 2012.
- [75] D. Ciuonzo, G. Romano, and P. Salvo Rossi, “Channel-aware decision fusion in distributed MIMO wireless sensor networks: Decode-and-Fuse vs. Decode-then-Fuse,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2976–2985, Aug. 2012.

- [76] E. Biglieri, G. Caire, G. Taricco, and J. Ventura-Traveset, "Simple method for evaluating error probabilities," *Electronics Letters*, vol. 32, no. 3, pp. 191–192, Feb. 1996.
- [77] A. Annamalai, C. Tellambura, and V. K. Bhargava, "Efficient computation of MRC diversity performance in Nakagami fading channel with arbitrary parameters," *Electronics Letters*, vol. 34, no. 12, pp. 1189–1190, Jun. 1998.
- [78] B. Picinbono, "On deflection as a performance criterion in detection," vol. 31, no. 3, pp. 1072–1081, Jul. 1995.
- [79] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, Feb. 2008.
- [80] M. Schwarz, W. R. Bennet, and S. Stein, *Communication Systems and Techniques*. New York: McGraw-Hill, 1966.
- [81] X. Zhu and R. D. Murch, "Performance analysis of maximum likelihood detection in a MIMO antenna system," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 187–191, Feb. 2002.
- [82] F. Li, J. S. Evans, and S. Dey, "Decision fusion over noncoherent fading multiaccess channels," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4367–4380, Sep. 2011.
- [83] L. A. Belfore, "An $O(n(\log_2(n))^2)$ algorithm for computing the reliability of k -out-of- n : G & k -to- l -out-of- n : G systems," *IEEE Transactions on Reliability*, vol. 44, no. 1, pp. 132–136, Mar. 1995.
- [84] S. X. Chen and J. S. Liu, "Statistical applications of the Poisson-Binomial and conditional Bernoulli distributions," *Statistica Sinica*, vol. 7, no. 4, 1997.
- [85] M. Fernandez and S. Williams, "Closed-form expression for the Poisson-Binomial probability density function," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 2, pp. 803–817, Apr. 2010.
- [86] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag New York, Inc., 1994.
- [87] A. F. Karr, *Probability*. Springer-Verlag New York, Inc., 1993.
- [88] R. R. Brooks, P. Ramanathan, and A. M. Sayeed, "Distributed target classification and tracking in sensor networks," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1163–1171, Aug. 2003.
- [89] J. H. Kotecha, V. Ramachandran, and A. M. Sayeed, "Distributed multi-target classification in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 703–713, Apr. 2005.

- [90] W. Baek and S. Bommareddy, "Optimal m-ary data fusion with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 3, pp. 1150–1152, Jul. 1995.
- [91] Q. Zhang and P. Varshney, "Decentralized m-ary detection via hierarchical binary decision fusion," *Information Fusion*, vol. 2, no. 1, pp. 3 – 16, 2001.
- [92] B. Dasarathy, *Decision Fusion*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1994.
- [93] X. Zhu, Y. Yuan, C. Rorres, and M. Kam, "Distributed m-ary hypothesis testing with binary local decisions," *Information Fusion*, vol. 5, no. 3, pp. 157 – 167, 2004.
- [94] T. Wang, Y. S. Han, P. K. Varshney, and P. Chen, "Distributed fault-tolerant classification in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 724–734, Apr. 2005.
- [95] B. Dasarathy, "Sensor fusion potential exploitation-innovative architectures and illustrative applications," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 24 –38, jan 1997.
- [96] A. D'Costa, V. Ramachandran, and A. M. Sayeed, "Distributed classification of gaussian space-time sources in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1026–1036, Aug. 2004.
- [97] R. Niu, P. Varshney, and Q. Cheng, "Distributed detection in a large wireless sensor network," *Information Fusion*, vol. 7, no. 4, pp. 380 – 394, 2006, special Issue on the Seventh International Conference on Information Fusion-Part I, Seventh International Conference on Information Fusion.
- [98] A. van der Vaart, *Asymptotic Statistics*. Cambridge University Press, 2000.
- [99] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*, C. U. Press, Ed. Cambridge, 2009.
- [100] Y. Bar-Shalom, T. Kirubarajan, and X. R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: John Wiley & Sons, Inc., 2002.
- [101] S. C. Nardone and V. J. Aidala, "Observability criteria for bearings-only target motion analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-17, no. 2, pp. 162–166, Mar. 1981.
- [102] S. E. Hammel and V. J. Aidala, "Observability requirements for three-dimensional tracking via angle measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-21, no. 2, pp. 200–207, Mar. 1985.

- [103] A. N. Payne, "Observability conditions for angles-only tracking," in *Twenty-Second Asilomar Conference on Signals, Systems and Computers, 1988.*, vol. 1, 1988, pp. 451–457.
- [104] E. Fogel and M. Gavis, "Nth-order dynamics target observability from angle measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 3, pp. 305–308, May 1988.
- [105] K. Becker, "Simple linear theory approach to TMA observability," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 2, pp. 575–578, Apr. 1993.
- [106] C. Jauffret and D. Pillon, "Observability in passive target motion analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1290–1300, Oct. 1996.
- [107] J.-P. Le Cadre and C. Jauffret, "Discrete-time observability and estimability analysis for bearings-only target motion analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 1, pp. 178–201, Jan. 1997.
- [108] C. Jauffret, "Observability and Fisher information matrix in nonlinear regression," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 2, pp. 756–759, Apr. 2007.
- [109] J. M. Passerieux and D. V. Cappel, "Optimal observer maneuver for bearings-only tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 3, pp. 777–788, Jul. 1998.
- [110] K. Doğançay, "On the bias of linear least squares algorithms for passive target localization," *Signal Processing*, vol. 84, no. 3, pp. 475–486, Mar. 2004.
- [111] J.-P. Le Cadre and C. Jauffret, "On the convergence of iterative methods for bearings-only tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 3, pp. 801–818, Jul. 1999.
- [112] K. Doğançay, "Passive emitter localization using weighted instrumental variables," *Signal Processing*, vol. 84, no. 3, pp. 487–497, 2004.
- [113] —, "On the efficiency of a bearings-only instrumental variable estimator for target motion analysis," *Signal Processing*, vol. 85, no. 3, pp. 481–490, Mar. 2005.
- [114] Y. J. Zhang and G. Z. Xu, "Bearings-only target motion analysis via instrumental variable estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5523–5533, Nov. 2010.

- [115] F. Bavencoff, J. M. Vanpeperstraete, and J.-P. Le Cadre, "Constrained bearings-only target motion analysis via Markov chain Monte Carlo methods," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 4, pp. 1240–1263, Oct. 2006.
- [116] C. Jauffret, D. Pillon, and A.-C. Pignol, "Bearings-only maneuvering target motion analysis from a nonmaneuvering platform," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 4, pp. 1934–1949, Oct. 2010.
- [117] J. Clavard, D. Pillon, A.-C. Pignol, and C. Jauffret, "Bearings-only target motion analysis of a source in a circular constant speed motion from a non-maneuvering platform," in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, Jul. 2011, pp. 1–8.
- [118] C. Jauffret and Y. Bar-Shalom, "Track formation with bearing and frequency measurements in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 6, pp. 999–1010, Nov. 1990.
- [119] T. Kirubarajan and Y. Bar-Shalom, "Low observable target motion analysis using amplitude information," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1367–1384, Oct.. 1996.
- [120] M. R. Chummun, Y. Bar-Shalom, and T. Kirubarajan, "Adaptive early-detection ML-PDA estimator for LO targets with EO sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 694–707, Apr. 2002.
- [121] W. R. Blanding, P. K. Willett, Y. Bar-Shalom, and R. Lynch, "Directed subspace search ML-PDA with application to active sonar tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 201–216, Jan. 2008.
- [122] S. C. Nardone, A. Lindgren, and K. Gong, "Fundamental properties and performance of conventional bearings-only target motion analysis," *IEEE Transactions on Automatic Control*, vol. 29, no. 9, pp. 775–787, Sep. 1984.
- [123] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [124] S. Singer and S. Singer, "Complexity analysis of Nelder-Mead search iterations," in *Proceedings of the first Conference on Applied Mathematics and Computation*, Sep. 1999, pp. 185–196.
- [125] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Feb. 2005.