SPATIALLY CONTROLLED RELAY BEAMFORMING

by

DIONYSIOS KALOGERIAS

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical Engineering

Written under the direction of

Athina P. Petropulu

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2017

ABSTRACT OF THE DISSERTATION

Spatially Controlled Relay Beamforming

by DIONYSIOS KALOGERIAS

Dissertation Director:
Athina P. Petropulu

This thesis is about fusion of optimal stochastic motion control and physical layer communications. Distributed, networked communication systems, such as relay beamforming networks (e.g., Amplify & Forward (AF)), are typically designed without explicitly considering how the positions of the respective nodes might affect the quality of the communication. Optimum placement of network nodes, which could potentially improve the quality of the communication, is not typically considered. However, in most practical settings in physical layer communications, such as relay beamforming, the Channel State Information (CSI) observed by each node, per channel use, although it might be (modeled as) random, it is both spatially and temporally correlated. It is, therefore, reasonable to ask if and how the performance of the system could be improved by (predictively) controlling the positions of the network nodes (e.g., the relays), based on causal side (CSI) information, and exploitting the spatiotemporal dependencies of the wireless medium. In this work, we address this problem in the context of AF relay beamforming networks. This novel, *cyber-physical system* approach to relay beamforming is termed as *"Spatially Controlled Relay Beamforming"*.

First, we discuss wireless channel modeling, however, in a rigorous, Bayesian framework. Experimentally accurate and, at the same time, technically precise channel modeling is absolutely essential for designing and analyzing spatially controlled communication systems. In this work, we are interested in two distinct *spatiotemporal* statistical models, for describing the behavior of the *log-scale magnitude* of the wireless channel:

1. *Stationary Gaussian Fields*: In this case, the channel is assumed to evolve as a stationary, Gaussian stochastic field in *continuous space* and *discrete time* (say, for instance, time slots). Under such assumptions, spatial and temporal statistical interactions are determined by a set of time and space invariant parameters, which completely determine the mean and covariance of the underlying Gaussian measure. This model is relatively simple to describe, and can be sufficiently characterized, at least for our purposes, both *statistically* and *topologically*. Additionally, the model is rather versatile and there is existing experimental evidence, supporting its practical applicability. Our contributions are summarized in properly formulating the whole *spatiotemporal* model in a completely rigorous mathematical setting, under a convenient measure theoretic framework. Such framework greatly facilitates formulation of meaningful stochastic control problems, where the wireless channel field (or a function of it) can be regarded as a *stochastic optimization surface*.

2. *Conditionally Gaussian Fields*, when conditioned on a *Markovian channel state*: This is a completely novel approach to wireless channel modeling. In this approach, the communication medium is assumed to behave as a *partially observable* (or *hidden*) *system*, where a hidden, global, *temporally varying* underlying stochastic process, called the *channel state*, affects the *spatial interactions* of the actual channel magnitude, evaluated at any set of locations in the plane. More specifically, we assume that, conditioned on the channel state, the wireless channel constitutes an observable, conditionally Gaussian stochastic process. The channel state evolves in time according to a *known*, possibly non stationary, non

Gaussian, low dimensional Markov kernel. Recognizing the intractability of general nonlinear state estimation, we advocate the use of grid based approximate nonlinear filters as an effective and robust means for recursive tracking of the channel state. We also propose a sequential spatiotemporal predictor for tracking the channel gains at any point in time and space, providing real time sequential estimates for the respective channel gain map. In this context, our contributions are multifold. Except for the introduction of the layered channel model previously described, this line of research has resulted in a number of general, *asymptotic convergence* results, advancing the theory of grid-based approximate nonlinear stochastic filtering. In particular, sufficient conditions, ensuring asymptotic optimality are relaxed, and, at the same time, the mode of convergence is strengthened. Although the need for such results initiated as an attempt to theoretically characterize the performance of the proposed approximate methods for statistical inference, in regard to the proposed channel modeling approach, they turn out to be of fundamental importance in the areas of nonlinear estimation and stochastic control. The experimental validation of the proposed channel model, as well as the related parameter estimation problem, termed as *"Markovian Channel Profiling (MCP)"*, fundamentally important for any practical deployment, are subject of current, ongoing research.

Second, adopting the first of the two aforementioned channel modeling approaches, we consider the spatially controlled relay beamforming problem for an AF network with a single source, a single destination, and multiple, controlled at will, relay nodes.

We consider a time slotted system, where the relays update their positions before the beginning of each time slot. Under a general, rigorous and theoretically grounded framework, based on a version of the so-called *Fundamental Lemma of Stochastic Control*, we propose a novel, 2-stage stochastic programming formulation for specifying both beamforming weights and relay positions at each time slot. The objective is to maximize the *expected (long term) Quality-of-Service (QoS)* of the network, *at each time slot*, based on *causal* Channel State Information (CSI), while respecting a *total transmit power budget at the relays*. The resulting motion control problem is shown to

be equivalent to a set of much simpler, 2-dimensional subproblems, which may be solved in a distributed fashion, one at each relay. However, these problems are all nonconvex, and their objectives are impossible to evaluate analytically. Then, two methods are proposed, one based on the *Method of Statistical Differentials*, and one relying on the *multidimensional Gauss-Hermite Quadrature Rule (brute force)*. Both methods allow approximate, closed form evaluation of the aforementioned objectives, enabling the use of any preferable nonlinear solver, thus allowing the determination of approximately optimal relay controls.

Additionally, we show analytically that, although positions are optimized *myopically* at each time slot (based on all available CSI, though), the average network QoS is *nondecreasing across time slots*, as long as the temporal dependence of the communication medium is sufficiently strong. Synthetic numerical simulations are presented, confirming our theoretical predictions and corroborating the efficacy of the proposed approach.

The extension of this formulation, when the second of the channel modeling approaches presented above is adopted, in which the channel is modeled as partially observable Markovian system, is nontrivial and constitutes a subject of further research.

Acknowledgements

I would like to express my sincere gratitude to everyone who has contributed, willingly or not, to the completion of this thesis. To me, the priviledge of obtaining a Doctoral Degree is not typical or trivial; it is the result of a highly personally and socially demanding, life changing procedure. So far, it has been a wonderful and rewarding experience.

First, I am truly and deeply grateful to my advisor, professor Athina P. Petropulu, for her stimulating support and guidance throughout my past four-and-a-half year tenure as a PhD student at Rutgers. She has been a true mentor, from both a personal and professional perspectives. From the beginning, she has trusted and encouraged me in pursuing my own research directions, something that I think has fundamentally contributed in my development as an independent researcher.

My sincere thanks go to professors Waheed Bajwa, Narayan Mandayam, Anand Sarwate from the Department of Electrical & Computer Engineering, and to Michael Katehakis from the Business School here at Rutgers, for accepting to be members of my dissertation committee and for their valuable comments, which greatly helped improve my work. I would also like to thank the Graduate Director, professor Zoran Gajic, for his support and his patience.

Aside from the faculty, I am also indebted to many other colleagues from Rutgers, who have made my experience as a graduate student most pleasant and gratifying. A big "thank you" goes to Bo Li, Shunqiao Sun, Valerie Yang, Tasos Dimas, Zahra Shakeri, Haroon Raja, Talal Ahmed, Tong Wu, Mohsen Ghassemi and Hafiz Imtiaz.

A very special reference is due to the following *friends*: Nikos Chatzipanagiotis, with whom I have also had the most productive collaboration, and who essentially introduced me to the world of Mathematical Programming, Ioannis Manousakis, who has asked

me some of the most interesting applied probability questions, Manousos Maridakis, for being my "big" math friend, Kostantinos Nikolakakis, for being my "little" math friend, my long-time friend Ioannis Paraskevakos, for making things simple and providing quick answers to seemingly (to me) complicated questions, and my other long-time friend Ted Tsoulos, for his persistence.

Of course, I would also like to thank my parents, Spyros & Rena, for their unconditional support. Nothing would have been possible without it. In the same spirit, special thanks go to my sister, Mary, my brother-in-law Tasos, and my little niece, Anastasia.

Finally, I would like to express my deepest gratitude to my partner in life and everything, Ioanna Tsoulou, to whom this thesis is dedicated to.

Dedication

To Ioanna... & Uncertainty

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the first chapter of this work, we present the basic relay beamforming model, which will be of central interest in our subsequent developments. Based on this networking model, the problem of spatially controlled beamforming is naturally introduced, however in an informal and intuitive way. Additionally, the contributions and the outline of this dissertation are briefly presented, serving as a roadmap and connecting all theoretical and applied concepts, presented and developed in this work.

1.1 System Model

On a compact, square planar region $\mathcal{W} \subset \mathbb{R}^2$, we consider a wireless cooperative network consisting of one source, one destination and $R \in \mathbb{N}^+$ assistive relays, as shown in Fig. **??**. Each entity of the network is equipped with a single antenna, being able for both information reception and broadcasting/transmission. The source and destination are stationary and located at $\mathbf{p}_S \in \mathcal{W}$ and $\mathbf{p}_D \in \mathcal{W}$, respectively, whereas the relays are assumed to be mobile; each relay $i \in \mathbb{N}_R^+$ moves along a trajectory $\mathbf{p}_i(t) \in \mathcal{S} \subset \mathcal{W} - \{\mathbf{p}_S, \mathbf{p}_D\} \subset \mathcal{W}$, where, in general, $t \in \mathbb{R}_+$, and where $\mathcal{S}$ is compact. We also define the supervector $\mathbf{p}(t) \triangleq \left[\mathbf{p}_1^{\boldsymbol{T}}(t)\ \mathbf{p}_2^{\boldsymbol{T}}(t) \ldots \mathbf{p}_R^{\boldsymbol{T}}(t)\right]^{\boldsymbol{T}} \in \mathcal{S}^R \subset \mathbb{R}^{2R\times 1}$. Additionally, we assume that the relays can cooperate with each other, either by exchanging local messages, or by communicating with a local fusion center, through a dedicated channel. Hereafter, as already stated above, all probabilistic arguments made below presume the existence of a complete base probability space of otherwise completely arbitrary structure, prespecified by a triplet $(\Omega, \mathscr{F}, \mathcal{P})$. This base space models a universal source of randomness, generating all stochastic phenomena in our considerations.

Assuming that a direct link between the source and the destination does not exist,

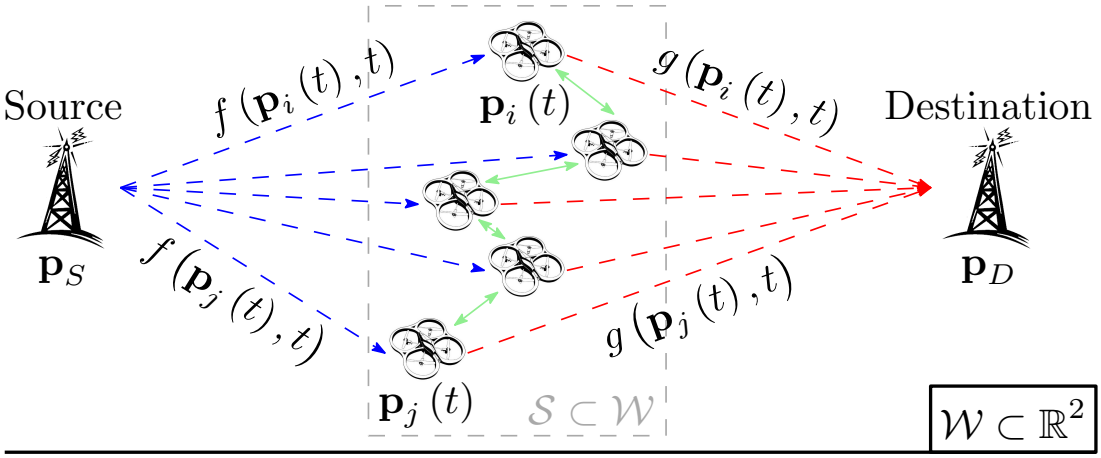


Figure 1.1: A schematic of the system model considered.

the role of the relays is determined to be assistive to the communication, operating in a two phase Amplify & Forward (AF) relaying mode [1–10]. Fix a $T > 0$, and *divide the time interval* $[0, T]$ *into* N_T *time slots*, with $t \in \mathbb{N}_{N_T}^+$ denoting the respective time slot. Let $s(t) \in \mathbb{C}$, with $\mathbb{E}\left\{|s(t)|^2\right\} \equiv 1$, denote the symbol to be transmitted at time slot t. Also, assuming a flat fading channel model, as well as channel reciprocity and quasistaticity in each time slot, let the sets $\{f_i(t) \in \mathbb{C}\}_{i \in \mathbb{N}_R^+}$ and $\{g_i(t) \in \mathbb{C}\}_{i \in \mathbb{N}_R^+}$ contain the *random, spatiotemporally varying* source-relay and relay-destination channel gains, respectively. These are further assumed to be *evaluations* of the *separable random channel fields* or *maps* $f(\mathbf{p}, t)$ and $g(\mathbf{p}, t)$, respectively, that is, $f_i(t) \equiv f(\mathbf{p}_i(t), t)$ and $g_i(t) \equiv g(\mathbf{p}_i(t), t)$, for all $i \in \mathbb{N}_R^+$ and for all $t \in \mathbb{N}_{N_T}^+$.

If $P_0 > 0$ denotes the transmission power of the source, during AF phase 1, the signals received at the relays can be expressed as

$$r_i(t) \triangleq \sqrt{P_0} f_i(t)\, s(t) + n_i(t) \in \mathbb{C}, \tag{1.1}$$

for all $i \in \mathbb{N}_R^+$ and for all $t \in \mathbb{N}_{N_T}^+$, where $n_i(t) \in \mathbb{C}$, with $\mathbb{E}\left\{|n_i(t)|^2\right\} \equiv \sigma^2$, constitutes a zero mean observation noise process at the i-th relay, independent across relays. During AF phase 2, all relays simultaneously retransmit the information received, each modulating their received signal by a weight $w_i(t) \in \mathbb{C}, i \in \mathbb{N}_R^+$. The signal received at the destination can be expressed as

$$y(t) \triangleq \sqrt{P_0} \sum_{i \in \mathbb{N}_R^+} w_i(t)\, g_i(t)\, r_i(t)$$

$$\equiv \underbrace{\sqrt{P_0}\sum_{i\in\mathbb{N}_R^+} w_i(t)\, g_i(t)\, f_i(t)\, s(t)}_{\text{signal (transformed)}} + \underbrace{\sum_{i\in\mathbb{N}_R^+} w_i(t)\, g_i(t)\, n_i(t) + n_D(t)}_{\text{interference + reception noise}} \in \mathbb{C}, \tag{1.2}$$

for all $i \in \mathbb{N}_R^+$ and $t \in \mathbb{N}_{N_T}^+$, where $n_D(t) \in \mathbb{C}$, with $\mathbb{E}\left\{|n_D(t)|^2\right\} \equiv \sigma_D^2$, constitutes a zero mean, spatiotemporally white noise process at the destination.

In the following, it is assumed that the channel fields $f(\mathbf{p},t)$ and $g(\mathbf{p},t)$ may be *statistically dependent both spatially and temporally*, and that, *as usual*, the processes $s(t)$, $[f(\mathbf{p},t)\ g(\mathbf{p},t)]$, $n_i(t)$ for all $i \in \mathbb{N}_R^+$, and $n_D(t)$ are mutually independent. Also, we will assume that, at each time slot t, CSI $\{f_i(t)\}_{i\in\mathbb{N}_R^+}$ and $\{g_i(t)\}_{i\in\mathbb{N}_R^+}$ is known *exactly* to all relays. This may be achieved through pilot based estimation.

1.2 AF Beamforming via Convex Optimization

In classical AF relay beamforming, choice of the relay retransmission weights $w_i(t), i \in \mathbb{N}_R^+$ constitutes a central problem. Of course, relay weights should be chosen *optimally*, according to some meaningful performance criterion. In the simple, one source, one destination setting, of interest in this work, we may identify two basic such criteria, which admit relatively simple solutions, and which, in a sense, may be considered dual to each other [1, 2, 9].

1.2.1 Network Quality-of-Service (QoS) Maximization

A fundamentally important beamforming criterion, which will be of central importance in our work, is that of enhancing network Quality-of-Service (QoS), or, in other words, maximizing the respective SINR at the destination, subject to a total power budget at the relays. At each time slot $t \in \mathbb{N}_{N_T}^+$, *given instantaneous* CSI and with $\boldsymbol{w}(t) \triangleq [w_1(t) \ \ldots \ w_R(t)]^{\boldsymbol{T}}$, this may be achieved by formulating the constrained optimization problem [2, 9]

$$\begin{aligned} \underset{\boldsymbol{w}(t)}{\text{maximize}} \quad & \frac{P_S(t)}{P_{I+N}(t)} \\ \text{subject to} \quad & P_R(t) \le P_c \end{aligned}, \tag{1.3}$$

where $P_R(t)$, $P_S(t)$ and $P_{I+N}(t)$ denote the random instantaneous power at the relays, that of the signal component and that of the interference plus noise component at the destination (see (1.2)), respectively and where $P_c > 0$ denotes the total available relay transmission power. Using the mutual independence assumptions regarding CSI related to the source and destination, respectively, (1.3) can be reexpressed analytically as [9]

$$\begin{aligned} \underset{\boldsymbol{w}(t)}{\text{maximize}} \quad & \frac{\boldsymbol{w}^{\boldsymbol{H}}(t)\,\mathbf{R}(\mathbf{p}(t),t)\,\boldsymbol{w}(t)}{\sigma_D^2 + \boldsymbol{w}^{\boldsymbol{H}}(t)\,\mathbf{Q}(\mathbf{p}(t),t)\,\boldsymbol{w}(t)} \\ \text{subject to} \quad & \boldsymbol{w}^{\boldsymbol{H}}(t)\,\mathbf{D}(\mathbf{p}(t),t)\,\boldsymbol{w}(t) \leq P_c \end{aligned}, \tag{1.4}$$

where, dropping the dependence on $(\mathbf{p}(t),t)$ or t for brevity,

$$\mathbf{D} \triangleq P_0 \text{diag}\left(\left[|f_1|^2\ |f_2|^2\ \ldots\ |f_R|^2\right]^{\boldsymbol{T}}\right) + \sigma^2\mathbf{I}_R \in \mathbb{S}_{++}^R, \tag{1.5}$$

$$\mathbf{R} \triangleq P_0\mathbf{h}\mathbf{h}^{\boldsymbol{H}} \in \mathbb{S}_{+}^R, \text{ with } \mathbf{h} \triangleq [f_1 g_1\, f_2 g_2\, \ldots\, f_R g_R]^{\boldsymbol{T}} \text{ and} \tag{1.6}$$

$$\mathbf{Q} \triangleq \sigma^2 \text{diag}\left(\left[|g_1|^2\ |g_2|^2\ \ldots\ |g_R|^2\right]^{\boldsymbol{T}}\right) \in \mathbb{S}_{++}^R. \tag{1.7}$$

Note that the program (1.4) is *always feasible, as long as P_c is nonnegative*. It is well known that the optimal value of (1.4) can be expressed in closed form as [9]

$$V_t \equiv V(\mathbf{p}(t),t) \triangleq P_c\lambda_{max}\left(\left(\sigma_D^2\mathbf{I}_R + P_c\mathbf{D}^{-1/2}\mathbf{Q}\mathbf{D}^{-1/2}\right)^{-1}\mathbf{D}^{-1/2}\mathbf{R}\mathbf{D}^{-1/2}\right), \tag{1.8}$$

for all $t \in \mathbb{N}_{N_T}^+$. Exploiting the structure of the matrices involved, V_t may also be expressed *analytically* as [2]

$$\begin{aligned} V_t &\equiv \sum_{i\in\mathbb{N}_R^+} \frac{P_c P_0 |f(\mathbf{p}_i(t),t)|^2 |g(\mathbf{p}_i(t),t)|^2}{P_0\sigma_D^2 |f(\mathbf{p}_i(t),t)|^2 + P_c\sigma^2 |g(\mathbf{p}_i(t),t)|^2 + \sigma^2\sigma_D^2} \\ &\triangleq \sum_{i\in\mathbb{N}_R^+} V_I(\mathbf{p}_i(t),t), \quad \forall t \in \mathbb{N}_{N_T}^+. \end{aligned} \tag{1.9}$$

1.2.2 Total Relay Power Minimization

The second relay beamforming criterion of interest in this work is that of minimizing the transmission power at the relays, while satisfying a user specified QoS demand at the destination. This demand is quantified via an explicit lower bound on the achieved SINR, which has to be strictly respected at the destination. In a sense, this problem

constitutes a dual version of the SINR maximization problem, previously considered in Section 1.2.1. Its structure, though, is somewhat more complicated, as we shall shortly see. At each time $t \in \mathbb{N}_{N_T}^{+}$ and *given instantaneous* CSI, transmission power minimization at the relays this may be achieved by formulating the constrained optimization problem [2, 9]

$$\begin{array}{ll} \underset{\boldsymbol{w}(t)}{\text{minimize}} & P_R(t) \\ \text{subject to} & \dfrac{P_S(t)}{P_{I+N}(t)} \geq \zeta \end{array}, \tag{1.10}$$

where $P_R(t)$, $P_S(t)$ and $P_{I+N}(t)$ denote the same quantities as in Section 1.2.1, respectively and where $\zeta > 0$ denotes the *time slot independent* QoS requirement at the destination, translated in a hard SINR constraint. As with (1.3), (1.10) can be reexpressed analytically as [9]

$$\begin{array}{ll} \underset{\boldsymbol{w}(t)}{\text{minimize}} & \boldsymbol{w}^{\boldsymbol{H}}(t)\,\mathbf{D}(\mathbf{p}(t),t)\,\boldsymbol{w}(t) \\ \text{subject to} & \dfrac{\boldsymbol{w}^{\boldsymbol{H}}(t)\,\mathbf{R}(\mathbf{p}(t),t)\,\boldsymbol{w}(t)}{\sigma_D^2 + \boldsymbol{w}^{\boldsymbol{H}}(t)\,\mathbf{Q}(\mathbf{p}(t),t)\,\boldsymbol{w}(t)} \geq \zeta \end{array}, \tag{1.11}$$

where the matrices $\mathbf{D}(\mathbf{p}(t),t)$, $\mathbf{R}(\mathbf{p}(t),t)$ and $\mathbf{Q}(\mathbf{p}(t),t)$ are defined as in (1.5),(1.6) and (1.7), respectively. Note that, contrary to (1.4), program (1.11) will not be feasible for any choice of $\zeta > 0$. This is due to the fact that the constraint in (1.11) may be expressed equivalently as

$$\boldsymbol{w}^{\boldsymbol{H}}(\mathbf{R} - \zeta\mathbf{Q})\,\boldsymbol{w} \geq \zeta\sigma_D^2, \tag{1.12}$$

from where it trivially follows that, if the matrix $\mathbf{R}-\zeta\mathbf{Q}$ is strictly negative definite, then there is no $\boldsymbol{w}$ such that $\boldsymbol{w}^{\boldsymbol{H}}(\mathbf{R} - \zeta\mathbf{Q})\,\boldsymbol{w}$ takes a positive value. This fact is important, as it introduces considerable complications when one is interested in a motion aware formulation. For now, however, let us assume that ζ is chosen sufficiently small, such that (1.11) is feasible; we will revisit this issue later on. Under these circumstances, the optimal value of (1.4) can be expressed in closed form as [9]

$$\begin{aligned} V_t &\equiv V(\mathbf{p}(t),t) \\ &\triangleq \frac{\zeta\sigma_D^2}{\lambda_{max}\left(\mathbf{D}^{-1/2}(\mathbf{p}(t),t)\left(\mathbf{R}(\mathbf{p}(t),t) - \zeta\mathbf{Q}(\mathbf{p}(t),t)\right)\mathbf{D}^{-1/2}(\mathbf{p}(t),t)\right)}, \end{aligned} \tag{1.13}$$

for all $t \in \mathbb{N}_{N_T}^{+}$. It is very important to observe that, in this case, V_t *does not admit any known analytical expression*, simply because the eigenstructure of the matrix

$\mathbf{D}^{-1/2}\left(\mathbf{R}-\zeta\mathbf{Q}\right)\mathbf{D}^{-1/2}$, being a sum of a full rank and a rank-1 matrices, is, in general, unknown.

1.3 Spatially Controlled Wireless Communications: Basic Concepts

In both relay beamforming formulations previously presented in Section 1.2, as well as in distributed, networked communication systems, in general, spatial placement of network nodes is rarely taken into account as a system design factor, although it is clear that it could potentially improve the quality of the communication. This is due to the widely accepted and experimentally validated fact that, at least in most practical settings in physical layer communications, including relay beamforming, the Channel State Information (CSI) observed by each node, per channel use, is both spatially and temporally correlated. It is, therefore, reasonable to ask if and how the performance of the system could be improved by (predictively) controlling the positions of the network nodes (e.g., the relays), based on causal side (CSI) information, and exploiting the spatiotemporal dependencies of the wireless medium.

In this work, this problem is addressed in the context of AF relay beamforming networks. This novel, *cyber-physical system* approach to relay beamforming is thereafter termed as *"Spatially Controlled Relay Beamforming"*. Intuitively, if, at each time slot t, we are given *causal information* about the current and past positions of the relays and the observed CSI, then, naturally, the high-level goal is to try to *predictively* decide on the *future* positions of the relays, which will in some sense further optimize our agreed *future* beamforming objective. This should be considered as a *temporally dynamic* procedure, as the actions should be determined *sequentially* across time slots, depending on what has been observed so far.

There are multiple interesting questions arising in this general concept of spatially controlled relay beamforming (and wireless communications, in general), and which need to be addressed. These questions may be categorized as follows:

1. *Dynamic Wireless Channel Modeling*: We have to agree on a specific model for describing the spatiotemporal interactions of the communication medium. This

model might be either stochastic or deterministic, either hierarchical (layered) or simply parametric, and should exhibit certain properties, which will allow precise problem formulation and performance analysis. Although spatial and temporal wireless channel models are separately available in the literature, development of accurate representations for *joint* spatiotemporal channel modeling constitutes an active research direction.

2. *Scheduling of Communications and Controls*: We have to agree on the temporal order of the different tasks need to be carried out by the network. Should the relays move-and-beamform (or communicate, in general), beamform-and-move, or beamform while moving?

3. *Formulation*: There are multiple choices for meaningful problem formulation. For example, we might be interested in a multistage or myopic, fully or partially dynamic, single-period or multi-period formulation. Additionally, what causal information will be used for updating the locations of the relays?

4. *Optimality and Guarantees*: It will be readily apparent that rigorous analysis of such systems is far from trivial. Depending on the particular problem formulation adopted, the resulting system inherits unique properties, translated into useful theoretical guarantees. Such guarantees are extremely important in practice, because they provide performance quantification *in advance*, prior to numerical simulation or practical deployment.

Our contributions are generally concentrated around the various subproblems induced by each of the aforementioned categories. Additionally, this work presents a holistic treatment of the spatially controlled relay beamforming problem, which rigorously combines relay communications and stochastic motion control into a unified framework. This framework is sufficient for formulating other interesting and/or extended problems in spatially controlled communications, as well; several of them are the subject of our current and future research.

1.4 Contributions of the Dissertation

1.4.1 Performance Analysis of Grid-Based Approximate Nonlinear Stochastic Filtering

We consider the problem of approximating optimal in the Minimum Mean Squared Error (MMSE) sense nonlinear filters in a general, *discrete time* setting. More specifically, we consider a class of nonlinear, partially observable stochastic systems, comprised by a (possibly nonstationary) hidden stochastic process (the state), observed through another conditionally Gaussian stochastic process (the observations). Under general assumptions, we show that, given an approximating process which, for each time step, is stochastically convergent to the state process under an appropriate sense, an approximate filtering operator can be defined, which converges to the true optimal nonlinear filter of the state in a strong and well defined sense. In particular, the convergence is compact in time and uniform in a *completely characterized* measurable set of probability measure almost unity, also providing a purely quantitative justification of Egoroff's Theorem for the problem at hand. The results presented in this work can form a common basis for the analysis and characterization of a number of heuristic approaches for approximating optimal nonlinear filters, such as approximate grid based techniques, known to perform well in a variety of applications.

In particular, we then revisit the development of grid based recursive approximate filtering of general Markov processes in discrete time, partially observed in conditionally Gaussian noise. The grid based filters considered rely on two types of state quantization: The *Markovian* type and the *marginal* type. We propose a set of novel, relaxed sufficient conditions, ensuring strong and fully characterized pathwise convergence of these filters to the respective MMSE state estimator. In particular, for marginal state quantizations, we introduce the notion of *conditional regularity of stochastic kernels*, which, to the best of our knowledge, constitutes the most relaxed condition proposed, under which asymptotic optimality of the respective grid based filters is guaranteed. Further, we extend our convergence results, including filtering of bounded and continuous functionals of the state, as well as recursive approximate state prediction. For both

Markovian and marginal quantizations, the whole development of the respective grid based filters relies more on linear-algebraic techniques and less on measure theoretic arguments, making the presentation considerably shorter and technically simpler.

This work has been published in:

- D. S. Kalogerias and A. P. Petropulu, "Asymptotically Optimal Discrete Time Nonlinear Filters From Stochastically Convergent State Process Approximations," *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3522 – 3536, July 2015.

- D. S. Kalogerias and A. P. Petropulu, "Grid-Based Filtering of Markov Processes Revisited: Recursive Estimation & Asymptotic Optimality," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4244 - 4259, July 2016.

1.4.2 Hierarchical Wireless Channel Modeling & Markovian Channel Profiling (MCP)

We propose a nonlinear filtering framework for approaching the problems of channel state tracking and spatiotemporal channel gain prediction in mobile wireless networks, in a Bayesian setting. We assume that the (log-scale) wireless channel constitutes an observable (by the sensors/network nodes), spatiotemporal, conditionally Gaussian stochastic process, which is statistically dependent on a set of *hidden* channel parameters, called the *channel state*. The channel state evolves in time according to a known, *non stationary, nonlinear and/or non Gaussian* Markov stochastic kernel. Recognizing the intractability of general nonlinear state estimation, we advocate the use of grid based approximate filters as an effective and robust means for recursive tracking of the channel state. We propose a sequential spatiotemporal predictor for tracking the channel gains at any point in time and space, providing real time sequential estimates for the respective channel gain map, for each sensor in the network. Non trivial, real time recursive estimators of the variance of the channel gain map predictions are also developed. Additionally, we show that *all three* estimators converge towards the true respective MMSE optimal estimators, in a common, relatively strong sense. Numerical

simulations corroborate the effectiveness of the proposed approach.

This work has been published in:

- D. S. Kalogerias and A. P. Petropulu, "Nonlinear SpatioTemporal Channel Gain Map Tracking in Mobile Cooperative Networks," *16th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2015)*, Stockholm, Sweden, June/July 2015.

- D. S. Kalogerias and A. P. Petropulu, "Sequential channel state tracking & spatiotemporal channel prediction in mobile wireless sensor networks," *CSPL Technical Report*, Rutgers, The State University of New Jersey, 2015. Available at Arxiv.

- A. Dimas, C. Koumpouzi, D. S. Kalogerias and A. P. Petropulu, "Markovian Channel Profiling: Parameter Estimation & Space-Time Recursive Tracking", *under preparation*, 2017.

1.4.3 Spatially Controlled Relay Beamforming via Distributed Stochastic Programming

The problem of enhancing Quality-of-Service (QoS) in power constrained, mobile relay beamforming networks, by optimally and dynamically controlling the motion of the relaying nodes, is considered, in a dynamic channel environment. We assume a time slotted system, where the relays update their positions before the beginning of each time slot. Modeling the wireless channel as a Gaussian spatiotemporal stochastic field, we propose a novel 2-stage stochastic programming problem formulation for optimally specifying the positions of the relays at each time slot, such that the expected QoS of the network is maximized, based on causal Channel State Information (CSI) and under a total relay transmit power budget. This results in a schema where, at each time slot, the relays, apart from optimally beamforming to the destination, also optimally, predictively decide their positions at the next time slot, based on causally accumulated experience. Exploiting either the Method of Statistical Differentials, or

the multidimensional Gauss-Hermite Quadrature Rule, the stochastic program considered is shown to be approximately equivalent to a set of simple subproblems, which are solved in a distributed fashion, one at each relay. Optimality and performance of the proposed spatially controlled system are also effectively assessed, under a rigorous technical framework; strict optimality is rigorously demonstrated via the development of an original version of the Fundamental Lemma of Stochastic Control, and, performance-wise, it is shown that, quite interestingly, the optimal average network QoS exhibits an increasing trend across time slots, despite our myopic problem formulation. Numerical simulations are presented, experimentally corroborating the success of the proposed approach and the validity of our theoretical predictions.

This work has been published in:

- D. S. Kalogerias and A. P. Petropulu, "*Mobile Beamforming & Spatially Controlled Relay Communications*," 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014), Shanghai, China, March 2016 (*invited*, selected as the "*Best Paper of the Special Sessions*").

- D. S. Kalogerias and A. P. Petropulu, "*Enhancing QoS in Spatially Controlled Beamforming Networks via Distributed Stochastic Programming*," 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017), New Orleans, LA, USA, March 2017.

- D. S. Kalogerias and A. P. Petropulu, "*Spatially Controlled Relay Beamforming, Part I: 2-Stage QoS Enhancement Policies*," IEEE Transactions on Signal Processing, to be submitted in early 2017.

- D. S. Kalogerias and A. P. Petropulu, "*Spatially Controlled Relay Beamforming, Part II: Optimality Analysis & Extensions*," IEEE Transactions on Signal Processing, to be submitted in early 2017.

1.5 Outline of the Dissertation

The rigorous investigation of the spatially controlled beamforming problem has led to the development of fundamental and generally applicable theoretical contributions, as well, especially in the context of approximate Bayesian inference and stochastic nonlinear filtering. Therefore, we have decided to dedicate separate chapters to the presentation of general theoretical results. Those chapters will be usually presented prior to the relevant chapters discussing wireless channel modeling, and the actual formulation of the spatially controlled relay beamforming problem.

In Chapter 2, we present a complete and rigorous analysis of approximate nonlinear filtering in discrete time, with a focus on grid-based approximate nonlinear filtering of Markov processes, observed in conditionally Gaussian noise. First, a general asymptotic result is established, regarding general, possibly non-Markovian approximate filters. Subsequently, this result is exploited in order to provide strong asymptotic optimality guarantees for grid-based, *recursive* approximate filters, under various settings. The analysis is theoretically precise and is presented under a convenient measure-theoretic framework.

Chapter 3 is devoted to the discussion of the two distinct channel modeling approaches advocated and exploited in the dissertation. The two models are presented under the assumption of a set of possibly mobile, single antenna nodes in the space, communicating with some spatially fixed, reference base station. The first of these models is based on parametric *spatiotemporal* Gaussian random fields. We present a complete formulation of the model, its basic functional characteristics, as well as its fundamental topological properties. The second channel model constitutes a hierarchical Bayesian modeling approach, where the communication medium is described as a partially observable Markovian system. The hidden layer of the system, called the *channel state*, which models the *temporal* changes in the surrounding environment, affects the spatiotemporally interacting, actual channel magnitudes in a nonlinear way. Leveraging our results from Chapter 2, we present a rigorous treatment of the practically important problems of spatiotemporal channel state estimation and channel

map tracking/prediction, under a grid-based approximate nonlinear filtering framework. Synthetic numerical simulations are also presented, confirming the effectiveness of our approach.

Chapter 4 deals with the main problem of interest in the dissertation, namely, that of properly formulating the spatially controlled relay beamforming problem under reasonable assumptions, as well as providing a viable and effective solution. In this chapter, we adopt exclusively the parametric spatiotemporal random field channel model presented in Chapter 3, mainly because of its practical relevance and simple structure. The beamforming under consideration is selected to be the QoS maximization problem, introduced earlier in Section 1.2.1. The main idea in Chapter 4 is to formulate the spatially controlled relay beamforming problem, under a fully dynamic stochastic environment. In such a dynamic formulation, and assuming a time slotted system, the relays are expected to update their positions at each time slot, based on their experience accumulated so far, in order to further enhance the QoS of the network. We propose a myopic stochastic programming framework, where, at each time slot, the relays optimize their positions at the next time slot, based on CSI accumulated up to the current time slot. Optimality and performance of the proposed system is technically justified, under a rigorous, measure theoretic framework. Additionally, representative numerical simulations are presented, confirming the success of the proposed approach, as well as our theoretical predictions.

Finally, Chapter 5 concludes the dissertation and provides various directions for further research and development.

1.6 Notation

It will be soon noticeable to the reader that mathematical notation in each of the chapters to follow is somewhat intense. We will freely *redefine* variables, but all notation will be consistent within *sections*. However, same basic conventions will be universal throughout the main body of the dissertation, as follows.

Whenever applicable, the *state vector* will be represented as X_t, its approximations

as $X_t^{L_S}$. All other matrices and vectors, either random or not, will be denoted by boldface letters (to be clear by the context). Real valued random variables and abstract random elements will be denoted by uppercase letters. Calligraphic letters and formal script letters will denote sets and σ-algebras, respectively. Table 1.1 shows a list of common, additional conventions, which have been globally adopted in the dissertation. Additional special purpose notation will be specified explicitly, within each respective section.

Table 1.1: Notation

Symbol	**Meaning**
$=$	"equals" (indirect equivalence)
$\equiv$	"coincides" (immediate equivalence)
$\triangleq$	"equals *by definition*"
$(\cdot)^{\boldsymbol{T}}$	Matrix transposition
$\lambda_{min}(\cdot)$	Minimum eigenvalue operator
$\lambda_{max}(\cdot)$	Maximum eigenvalue operator
$\sigma\{\cdot\}$	σ-algebra generated by $(\cdot)$ (a random element)
$\|\boldsymbol{x}\|_p$, $\boldsymbol{x} \in \mathbb{C}^n$	ℓ_p-norm of a vector; $\|\boldsymbol{x}\|_p \triangleq (\sum_{i=1}^n \|x_i\|^p)^{1/p}$, $\boldsymbol{x} \in \mathbb{C}^n$
$\|\mathbf{X}\|_2$, $\mathbf{X} \in \mathbb{C}^{n\times n}$	Spectral norm of a matrix; $\|\mathbf{X}\|_2 \triangleq \max_{\|\boldsymbol{x}\|_2 \equiv 1} \|\mathbf{X}\boldsymbol{x}\|_2$
$\|\mathbf{X}\|_F$, $\mathbf{X} \in \mathbb{C}^{n\times n}$	Frobenius norm of a matrix; $\|\mathbf{X}\|_2 \triangleq \sqrt{\sum_{i=1}^n \left\|\mathbf{X}_{ij}\right\|^2}$
$\mathbf{X} \succeq \mathbf{0}$, $\mathbf{X} \in \mathbb{C}^{n\times n}$	Positive demidefiniteness
$\mathbf{X} \succ \mathbf{0}$, $\mathbf{X} \in \mathbb{C}^{n\times n}$	Positive definiteness
$\mathbf{I}_{n\times n}$ or $\mathbf{I}_n$, both $\in \mathbb{C}^{n\times n}$	Finite dimensional identity operator of dimension n.
$\mathbb{R}_+$	$[0, \infty)$
$\mathbb{R}_{++}$	$(0, \infty)$
$\mathbb{N}^+$	$\{1, 2, \ldots\}$
$\mathbb{N}_n^+$, $n \in \mathbb{N}^+$	$\{1, 2, \ldots, n\}$
$\mathbb{N}_n$, $n \in \mathbb{N}^+$	$\{0\} \cup \mathbb{N}_n^+ \equiv \{0, 1, 2, \ldots, n\}$
$\{\mathcal{A}, \mathcal{B}\} \times \{\mathcal{C}, \mathcal{D}\}$ ($\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ are sets)	$\{\mathcal{A} \times \mathcal{C}, \mathcal{A} \times \mathcal{D}, \mathcal{B} \times \mathcal{C}, \mathcal{A} \times \mathcal{D}\}$ (overloading of the Cartesian Product)

Chapter 2

Grid-Based Nonlinear Stochastic Filtering

A novel, complete and rigorous convergence analysis of approximate nonlinear filtering in discrete time is presented. The ultimate focus of the analysis is grid-based approximate nonlinear filtering of Markov processes, observed in conditionally Gaussian noise. First, a general asymptotic result is established, applicable to general, possibly non-Markovian, non-recursive approximate estimators. Then, this result is exploited, providing strong asymptotic optimality guarantees for grid-based, *recursive* approximate filters, under various settings.

2.1 Asymptotically Optimal Discrete Time Nonlinear Filters From Stochastically Convergent State Process Approximations

2.1.1 Introduction

Nonlinear stochastic filtering refers to problems in which a stochastic process, usually called the state, is partially observed as a result of measuring another stochastic process, usually called the observations or measurements, and the objective is to estimate the state or some functional of it, based only on past and present observations. The nonlinearity is due to the general, possibly non Gaussian nature of the state and observations processes, as well as the fact that, in general, the state may be partially observed as a nonlinear functional of the observations. Usually, nonlinear state estimators are designed so as to optimize some performance criterion. Most commonly, this corresponds to the Minimum Mean Squared Error (MMSE), which is also adopted in this work.

A desirable feature of a nonlinear filter is *recursiveness* in time, as it greatly reduces

computational complexity and allows for real time estimation as new measurements become available. However, not all nonlinear filters possess this important property [11, 12]. Recursive nonlinear filters exist for some very special cases, such as those in which the transition model of the state process is linear (Gauss-Markov), or when the state is a Markov chain (discrete state space) [13–16]. In the absence of recursive filter representations, practical filtering schemes have been developed, which typically approximate the desired quantities of interest, either heuristically (e.g., Gaussian approximations [17, 18]) or in some more powerful, rigorous sense (e.g., Markov chain approximations [19, 20]).

In this work, we follow the latter research direction. Specifically, we consider a partially observable system in *discrete time*, comprised by a hidden, almost surely compactly bounded state process, observed through another, conditionally Gaussian measurement process. The mean and covariance matrix of the measurements both constitute nonlinear, time varying and state dependent functions, assumed to be known apriori. Employing a change of measure argument and using the original measurements, an approximate filtering operator can be defined, by replacing the "true" state process by an appropriate approximation. Our contribution is summarized in showing that if the approximation converges to the state either in probability or in the $\mathcal{C}$-weak sense (Section II.C), the resulting filtering operator converges to the true optimal nonlinear filter in a relatively strong and well defined sense; the convergence is compact in time and uniform in a measurable set of probability measure almost unity (Theorem 2.3). The aforementioned set is completely characterized in terms of a subset of the parameters of the filtering problem of interest. Consequently, our results provide a purely quantitative justification of Egoroff's theorem [21] for the problem at hand, which concerns the equivalence of almost sure convergence and almost uniform convergence of measurable functions.

To better motivate the reader, and in order to embed the problems to be considered in the framework of the dissertation, let us describe two problems that fit the scenario described above and can benefit from the contributions of this work, namely, those of *sequential channel state estimation* and *(sequential) spatiotemporal channel prediction.*

These problems will be considered in detail later in Chapter 3. Modern distributed networks usually consist of cooperating mobile sensors, each of them being capable of observing its communication channel (under a flat fading assumption), relative to a reference point in the space. In most practical scenarios, the dominant quantities characterizing the wireless links, such as the path loss exponent and the shadowing power, may be modeled as stochastic processes themselves. For instance, such behavior may be due to physical changes in the environment and also the inherent randomness of the communication medium itself. Then, the path loss exponent and the shadowing power can be collectively considered as the hidden state (suggestively called the channel state) of a partially observable system, where the channel gains measured at each sensor can be considered as the corresponding observations. In general, such observations are nonlinear functionals of the state. Assuming additionally that the channel state is a Markov process, the main results presented herein can essentially provide strong asymptotic guarantees for approximate sequential nonlinear channel state estimation and spatiotemporal channel prediction, enabling physical layer aware motion planning and stochastic control. For more details, the reader is referred to Chapter 3.

The idea of replacing the process of interest with some appropriate approximation is borrowed from [20]. However, [20] deals almost exclusively with continuous time stochastic systems and the results presented in there do not automatically extend to the discrete time system setting we are dealing with here. In fact, the continuous time counterparts of the discrete time stochastic processes considered here are considerably more general than the ones treated in [20]. More specifically, although some relatively general results are indeed provided for continuous time hidden processes, [20] is primarily focused on the standard hidden diffusion case, which constitutes a Markov process (and aiming to the development of recursive approximate filters), whereas, in our setting, the hidden process is initially assumed to be arbitrary (as long as it is confined to a compact set). Also, different from our formulation (see above), in [20], the covariance matrix of the observation process *does not* depend on the hidden state; the state affects only the mean of the observations. Further, the modes of stochastic convergence considered here are different compared to [20] (in fact, they are stronger), both regarding

convergence of approximations and convergence of approximate filters.

The results presented in this work provide a framework for analyzing a number of heuristic techniques for numerically approximating optimal nonlinear filters in discrete time, such as approximate grid based recursive approaches, known to perform well in a wide variety of applications [22, 23]. Additionally, our results do not refer exclusively to recursive nonlinear filters. The sufficient conditions which we provide for the convergence of approximate filtering operators are independent of the way a filter is realized (see Section 2.1.3). This is useful because, as highlighted in [24], no one prevents one from designing an efficient (approximate) nonlinear filter which is part recursive and part nonrecursive, or even possibly trying to combine the best of both worlds, and there are practical filters designed in this fashion [24].

2.1.2 Partially Observable System Model & Technical Preliminaries

In this section, we give a detailed description of the partially observable (or hidden) system model of interest and present our related technical assumptions on its components. Additionally, we present some essential background on the measure theoretic concept of change of probability measures and state some definitions and known results regarding specific modes of stochastic convergence, which will be employed in our subsequent theoretical developments.

2.1.2.1 Hidden Model: Definitions & Technical Assumptions

First, let us set the basic probabilistic framework, as well as precisely define the hidden system model considered throughout this work:

- All stochastic processes considered below are fundamentally generated on a common complete probability space (the base space), defined by a triplet $(\Omega, \mathscr{F}, \mathcal{P})$, at each time instant taking values in a measurable state space, consisting of some Euclidean subspace and the associated Borel σ-algebra on that subspace. For example, for each $t \in \mathbb{N}$, the state process $X_t \equiv X_t(\omega)$, where $\omega \in \Omega$, takes its values in the measurable state space $\left(\mathbb{R}^{M\times 1}, \mathscr{B}\left(\mathbb{R}^{M\times 1}\right)\right)$, where $\mathscr{B}\left(\mathbb{R}^{M\times 1}\right)$

constitutes the Borel σ-algebra of measurable subsets of $\mathbb{R}^{M\times 1}$.

- In this work, the evolution mechanism of state process X_t is assumed to be arbitrary. However, in order to avoid unnecessary technical complications, we assume that, for each $t \in \mathbb{N}$, the induced probability measure of X_t is absolutely continuous with respect to the Lebesgue measure on its respective state space. Then, by the Radon-Nikodym Theorem, it admits a density, unique up to sets of zero Lebesgue measure. Also, we will generically assume that for all $t \in \mathbb{N}$, $X_t \in \mathcal{Z}$ almost surely, where $\mathcal{Z}$ constitutes a compact strict subset of $\mathbb{R}^{M\times 1}$. In what follows, however, in order to lighten the presentation, we will assume that $M \equiv 1$. Nevertheless, all stated results hold with the same validity if $M > 1$ (See also Assumption 2.2 below).

- The state X_t is partially observed through the observation process

$$\mathbf{y}_t \triangleq \boldsymbol{\mu}_t(X_t) + \boldsymbol{\sigma}_t(X_t) + \boldsymbol{\xi}_t \in \mathbb{R}^{N\times 1}, \quad \forall t \in \mathbb{N}, \tag{2.1}$$

  where, *conditioned on* X_t and for each $t \in \mathbb{N}$, the sequence $\left\{\boldsymbol{\mu}_t : \mathcal{Z} \mapsto \mathbb{R}^{N\times 1}\right\}_{t\in\mathbb{N}}$ is known apriori, the process $\boldsymbol{\sigma}_t(X_t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t(X_t) \succ \mathbf{0})$ constitutes Gaussian noise, with the sequence $\{\boldsymbol{\Sigma}_t : \mathcal{Z} \mapsto \mathcal{D}_{\boldsymbol{\Sigma}}\}_{t\in\mathbb{N}}$, where $\mathcal{D}_{\boldsymbol{\Sigma}}$ is a bounded subset of $\mathbb{R}^{N\times N}$, also known apriori, and $\boldsymbol{\xi}_t \overset{i.i.d.}{\sim} \mathcal{N}\left(\mathbf{0}, \sigma_\xi^2 \mathbf{I}_{N\times N}\right)$.

As a pair, the state X_t and the observations process described by (2.1) define a very wide family of partially observable systems. In particular, any Hidden Markov Model (HMM) of *any* order, in which the respective Markov state process is almost surely confined in a compact subset of its respective Euclidean state space, is indeed a member of this family. More specifically, let us rewrite (2.1) in the canonical form

$$\mathbf{y}_t \equiv \boldsymbol{\mu}_t(X_t) + \sqrt{\mathbf{C}_t(X_t)}\boldsymbol{u}_t \in \mathbb{R}^{N\times 1}, \quad \forall t \in \mathbb{N}, \tag{2.2}$$

where $\boldsymbol{u}_t \equiv \boldsymbol{u}_t(\omega)$ constitutes a standard Gaussian white noise process and, for all $x \in \mathcal{Z}$, $\mathbf{C}_t(x) \triangleq \boldsymbol{\Sigma}_t(x) + \sigma_\xi^2 \mathbf{I}_{N\times N} \in \mathcal{D}_{\mathbf{C}}$, with $\mathcal{D}_{\mathbf{C}}$ bounded. Then, for a possibly nonstationary HMM of order m, assuming the existence of an explicit functional model for describing the temporal evolution of the state (being a Markov process of order m),

we get the system of *standardized* stochastic difference equations

$$\begin{aligned} X_t &\equiv f_t\left(\{X_{t-i}\}_{i\in\mathbb{N}_m^+}, \boldsymbol{W}_t\right) \in \mathcal{Z} \\ \mathbf{y}_t &\equiv \boldsymbol{\mu}_t\left(X_t\right) + \sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t \end{aligned}, \quad \forall t \in \mathbb{N}, \tag{2.3}$$

where, for each t, $f_t : \mathcal{Z}^m \times \mathcal{W} \overset{a.s.}{\mapsto} \mathcal{Z}$ (with $\mathcal{Z}^m \triangleq \times_{m \text{ times}} \mathcal{Z}$) constitutes a measurable nonlinear state transition mapping and $\boldsymbol{W}_t \equiv \boldsymbol{W}_t(\omega) \in \mathcal{W} \subseteq \mathbb{R}^{M_W \times 1}$, denotes a (discrete time) white noise process with state space $\mathcal{W}$. For a first order stationary HMM, the above system of equations reduces to

$$\begin{aligned} X_t &\equiv f\left(X_{t-1}, \boldsymbol{W}_t\right) \in \mathcal{Z} \\ \mathbf{y}_t &\equiv \boldsymbol{\mu}_t\left(X_t\right) + \sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t \end{aligned}, \quad \forall t \in \mathbb{N}, \tag{2.4}$$

which arguably constitutes the most typical partially observable system model encountered in both Signal Processing and Control, with plethora of important applications.

Let us also present some more specific assumptions, regarding the nature (boundedness, continuity and expansiveness) of the aforementioned sequences of functions.

Assumption 2.1. (Boundedness) For later reference, let

$$\lambda_{inf} \triangleq \inf_{t\in\mathbb{N}} \inf_{x\in\mathcal{Z}} \lambda_{min}\left(\mathbf{C}_t\left(x\right)\right), \tag{2.5}$$

$$\lambda_{sup} \triangleq \sup_{t\in\mathbb{N}} \sup_{x\in\mathcal{Z}} \lambda_{max}\left(\mathbf{C}_t\left(x\right)\right), \tag{2.6}$$

$$\mu_{sup} \triangleq \sup_{t\in\mathbb{N}} \sup_{x\in\mathcal{Z}} \left\|\boldsymbol{\mu}_t\left(x\right)\right\|_2, \tag{2.7}$$

where each quantity of the above is uniformly and finitely bounded for all $t \in \mathbb{N}$ and for all $x \in \mathcal{Z}$. If x is substituted by the stochastic process $X_t(\omega)$, then all the above definitions continue to hold in the essential sense. For technical reasons related to the bounding-from-above arguments presented in Section IV, containing the proof of the main result of this work, it is also assumed that $\lambda_{inf} > 1$, a requirement which can always be satisfied by appropriate normalization of the observations.

Assumption 2.2. (Continuity & Expansiveness) All members of the functional family $\left\{\boldsymbol{\mu}_t : \mathcal{Z} \mapsto \mathbb{R}^{N\times 1}\right\}_{t\in\mathbb{N}}$ are uniformly Lipschitz continuous, that is, there exists a

bounded constant $K_{\mu} \in \mathbb{R}_{+}$, such that, for all $t \in \mathbb{N}$,

$$\left\| \boldsymbol{\mu}_t (x) - \boldsymbol{\mu}_t (y) \right\|_2 \leq K_{\mu} \left| x - y \right|, \quad \forall (x, y) \in \mathcal{Z} \times \mathcal{Z}. \tag{2.8}$$

Additionally, all members of the functional family $\left\{ \boldsymbol{\Sigma}_t : \mathcal{Z} \mapsto \mathcal{D}_{\boldsymbol{\Sigma}} \subset \mathbb{R}^{N \times N} \right\}_{t \in \mathbb{N}}$ are *elementwise* uniformly Lipschitz continuous, that is, there exists some universal and bounded constant $K_{\boldsymbol{\Sigma}} \in \mathbb{R}_{+}$, such that, for all $t \in \mathbb{N}$ and for all $(i, j) \in \mathbb{N}_N^{+} \times \mathbb{N}_N^{+}$,

$$\left| \boldsymbol{\Sigma}_t^{ij} (x) - \boldsymbol{\Sigma}_t^{ij} (y) \right| \leq K_{\boldsymbol{\Sigma}} \left| x - y \right|, \quad \forall (x, y) \in \mathcal{Z} \times \mathcal{Z}. \tag{2.9}$$

If x is substituted by the stochastic process $X_t (\omega)$, then all the above statements are understood in the almost sure sense.

Remark 2.1. As we have already said, for simplicity, we assume that $\mathcal{Z} \subset \mathbb{R}$, that is, $M \equiv 1$. In any other case (when $M > 1$), we modify the Lipschitz assumptions stated above simply by replacing $|x - y|$ with $\|\mathbf{x} - \mathbf{y}\|_1$, that is, Lipschitz continuity is meant to be with respect to the ℓ_1 norm in the domain of the respective function. If this holds, everything that follows works also in $\mathbb{R}^{M>1}$, just with some added complexity in the proofs of the results. Also, because $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$ for any $\mathbf{x} \in \mathbb{R}^M$, the assumed Lipschitz continuity with respect to ℓ_1 norm can be replaced by Lipschitz continuity with respect to the ℓ_2 norm, since the latter implies the former, and again everything holds. Further, if $M > 1$, convergence in probability and $\mathcal{L}_1$ convergence of random vectors are both defined by replacing absolute values with the ℓ_1 norms of the vectors under consideration. ■

2.1.2.2 Conditional Expectations, Change of Measure & Filters

Before proceeding with the general formulation of our estimation problem and for later reference, let us define the complete natural filtrations of the processes X_t and $\mathbf{y}_t$ as

$$\{\mathscr{X}_t\}_{t \in \mathbb{N}} \triangleq \left\{ \sigma \left\{ \{X_i\}_{i \in \mathbb{N}_t} \right\} \right\}_{t \in \mathbb{N}} \quad \text{and} \tag{2.10}$$

$$\{\mathscr{Y}_t\}_{t \in \mathbb{N}} \triangleq \left\{ \sigma \left\{ \{\mathbf{y}_i\}_{i \in \mathbb{N}_t} \right\} \right\}_{t \in \mathbb{N}}, \tag{2.11}$$

respectively, and also the complete filtration generated by both X_t and $\mathbf{y}_t$ as

$$\{\mathscr{H}_t\}_{t \in \mathbb{N}} \triangleq \left\{ \sigma \left\{ \{X_i, \mathbf{y}_i\}_{i \in \mathbb{N}_t} \right\} \right\}_{t \in \mathbb{N}}. \tag{2.12}$$

In all the above, $\sigma\{Y\}$ denotes the σ-algebra generated by the random variable Y.

In this work, we adopt the MMSE as an optimality criterion. In this case, one would ideally like to discover a solution to the stochastic optimization problem

$$\begin{aligned} &\inf_{\widehat{X}_t} \quad \mathbb{E}\left\{\left\|X_t-\widehat{X}_t\right\|_2^2\right\} \\ &\text{subject to} \quad \mathbb{E}\left\{\widehat{X}_t\middle|\mathscr{Y}_t\right\}\equiv\widehat{X}_t \end{aligned}, \quad \forall t\in\mathbb{N}, \tag{2.13}$$

where the constraint is equivalent to confining the search for possible estimators $\widehat{X}_t$ to the subset of interest, that is, containing the ones which constitute $\mathscr{Y}_t$-measurable random variables. Of course, the solution to the program (2.13) coincides with the conditional expectation [25]

$$\mathbb{E}\{X_t|\mathscr{Y}_t\}\equiv\widehat{X}_t, \quad \forall t\in\mathbb{N}, \tag{2.14}$$

which, in the nonlinear filtering literature, is frequently called a *filter*. There is also an alternative and very useful way of reexpressing the filter process $\widehat{X}_t$, using the concept of *change of probability measures*, which will allow us to stochastically decouple the state and observations of our hidden system and then let us formulate precisely the approximation problem of interest in this work. Change of measure techniques have been extensively used in discrete time nonlinear filtering, mainly in order to discover recursive representations for various hidden Markov models [12,15,16,26]. In the following, we provide a brief introduction to these type of techniques (suited to our purposes) which is also intuitive, simple and technically accessible, including direct proofs of the required results.

Change of Probability Measure in Discrete Time:
Demystification & Useful Results

So far, all stochastic processes we have considered are defined on the base space $(\Omega,\mathscr{F},\mathcal{P})$. In fact, it is the structure of the probability measure $\mathcal{P}$ that is responsible for the coupling between the stochastic processes X_t and $\mathbf{y}_t$, being, for each $t\in\mathbb{N}$, measurable functions from $(\Omega,\mathscr{F})$ to $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ and $\left(\mathbb{R}^N,\mathscr{B}\left(\mathbb{R}^N\right)\right)$, respectively. Intuitively, the measure $\mathcal{P}$ constitutes our "reference measurement tool" for measuring

the events contained in the base σ-algebra $\mathscr{F}$, and any random variable serves as a "medium" or "channel" for observing these events.

As a result, some very natural questions arise from the above discussion. First, one could ask if and under what conditions it is possible to change the probability measure $\mathcal{P}$, which constitutes our *fixed* way of assigning probabilities to events, to another measure $\widetilde{\mathcal{P}}$ on the same measurable space $(\Omega, \mathscr{F})$, in a way such that there exists some sort of transformation connecting $\mathcal{P}$ and $\widetilde{\mathcal{P}}$. Second, if we can indeed make the transition from $\mathcal{P}$ to $\widetilde{\mathcal{P}}$, could we choose the latter probability measure in a way such that the processes X_t and $\mathbf{y}_t$ behave according to a prespecified statistical model? For instance, we could demand that, under $\widetilde{\mathcal{P}}$, X_t and $\mathbf{y}_t$ constitute independent stochastic processes. Third and most important, is it possible to derive an expression for the "original" filter $\widehat{X}_t \equiv \mathbb{E}_{\mathcal{P}}\{X_t | \mathscr{Y}_t\}$ under measure $\mathcal{P}$, using only (conditional) expectations under $\widetilde{\mathcal{P}}$ (denoted as $\mathbb{E}_{\widetilde{\mathcal{P}}}\{\cdot | \cdot\}$)?

The answers to all three questions stated above are affirmative under very mild assumptions and the key result in order to prove this assertion is the Radon-Nikodym Theorem [27]. However, assuming that the induced joint probability measure of the processes of interest is absolutely continuous with respect to the Lebesgue measure of the appropriate dimension, in the following we provide an answer to these questions, employing only elementary probability theory, avoiding the direct use of the Radon-Nikodym Theorem.

Theorem 2.1. (Conditional Bayes' Theorem for Densities) *Consider the (possibly vector) stochastic processes $X_t(\omega) \in \mathbb{R}^{N_t \times 1}$ and $Y_t(\omega) \in \mathbb{R}^{M_t \times 1}$, both defined on the same measurable space $(\Omega, \mathscr{F})$, for all $t \in \mathbb{N}$. Further, if $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ are two probability measures on $(\Omega, \mathscr{F})$, suppose that:*

- *Under both $\mathcal{P}$ and $\widetilde{\mathcal{P}}$, the process X_t is integrable.*
- *Under the base probability measure $\mathcal{P}$ (resp. $\widetilde{\mathcal{P}}$), the induced joint probability measure of $\left(\{X_i\}_{i\in\mathbb{N}_t}, \{Y_i\}_{i\in\mathbb{N}_t}\right)$ is absolutely continuous with respect to the Lebesgue measure of the appropriate dimension, implying the existence of a density f_t (resp.*

$\widetilde{f}_t$*), with*

$$f_t : \left(\underset{i \in \mathbb{N}_t}{\times} \mathbb{R}^{N_i \times 1} \right) \times \left(\underset{i \in \mathbb{N}_t}{\times} \mathbb{R}^{M_i \times 1} \right) \mapsto \mathbb{R}_+. \tag{2.15}$$

- *For each set of points, it is true that*

$$\widetilde{f}_t \left(\cdots \right) \equiv 0 \quad \Rightarrow \quad f_t \left(\cdots \right) \equiv 0, \tag{2.16}$$

or, equivalently, the support of f_t *is contained in the support of* $\widetilde{f}_t$*.*

Also, for all $t \in \mathbb{N}$*, define the Likelihood Ratio (LR) at* t *as the* $\{\mathscr{H}_t\}$*-adapted, nonnegative stochastic process*[1]

$$\Lambda_t \triangleq \frac{f_t \left(X_0, X_1, \ldots, X_t, Y_0, Y_1, \ldots, Y_t \right)}{\widetilde{f}_t \left(X_0, X_1, \ldots, X_t, Y_0, Y_1, \ldots, Y_t \right)}. \tag{2.17}$$

Then, it is true that

$$\widehat{X}_t \equiv \mathbb{E}_{\mathcal{P}} \left\{ X_t \middle| \mathscr{Y}_t \right\} \equiv \frac{\mathbb{E}_{\widetilde{\mathcal{P}}} \left\{ X_t \Lambda_t \middle| \mathscr{Y}_t \right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}} \left\{ \Lambda_t \middle| \mathscr{Y}_t \right\}}, \tag{2.18}$$

almost everywhere with respect to $\mathcal{P}$*.*

Proof of Theorem 2.1. See Section 2.1.6 (Appendix). ■

Remark 2.2. The $\{\mathscr{H}_t\}$-adapted LR process

$$\Lambda_t \equiv \Lambda_t \left(\mathcal{X}_t \triangleq \{X_i\}_{i \in \mathbb{N}_t}, \mathcal{Y}_t \triangleq \{Y_i\}_{i \in \mathbb{N}_t} \right), \quad t \in \mathbb{N}, \tag{2.19}$$

as defined in (2.17), actually coincides with the restriction of the Radon-Nikodym derivative of $\mathcal{P}$ with respect to $\widetilde{\mathcal{P}}$ to the filtration $\{\mathscr{H}_t\}_{t \in \mathbb{N}}$, that is,

$$\left. \frac{\mathrm{d}\mathcal{P}(\omega)}{\mathrm{d}\widetilde{\mathcal{P}}(\omega)} \right|_{\mathscr{H}_t} \equiv \Lambda_t \left(\mathcal{X}_t(\omega), \mathcal{Y}_t(\omega) \right), \quad \forall t \in \mathbb{N}, \tag{2.20}$$

a statement which, denoting the collections $\{x_i\}_{i \in \mathbb{N}_t}$ and $\{y_i\}_{i \in \mathbb{N}_t}$ as $\mathcal{x}_t$ and $\mathcal{y}_t$, respectively, is rigorously equivalent to

$$\begin{aligned} \mathcal{P}(\mathcal{F}) &\equiv \int_{\mathcal{F}} \Lambda_t \left(\mathcal{X}_t(\omega), \mathcal{Y}_t(\omega) \right) \mathrm{d}\widetilde{\mathcal{P}}(\omega) \\ &\equiv \int_{\mathcal{B}} \Lambda_t \left(\mathcal{x}_t, \mathcal{y}_t \right) \mathrm{d}^{2t} \widetilde{\mathcal{P}}_{(\mathcal{X}_t, \mathcal{Y}_t)} \left(\mathcal{x}_t, \mathcal{y}_t \right) \end{aligned}$$

---

[1] *With zero probability of confusion, we use* $\{\mathscr{Y}_t\}_{t \in \mathbb{N}}$ *and* $\{\mathscr{H}_t\}_{t \in \mathbb{N}}$ *to denote the complete filtrations generated by* Y_t *and* $\{X_t, Y_t\}$*.*

$$\equiv \mathcal{P}_{(\mathcal{X}_t,\mathcal{Y}_t)}(\mathcal{B}) \equiv \mathcal{P}\left((\mathcal{X}_t,\mathcal{Y}_t)\in\mathcal{B}\right), \tag{2.21}$$

$$\tag{2.22}$$

$$\forall \mathcal{F} \triangleq \{\omega\in\Omega\,|(\mathcal{X}_t(\omega),\mathcal{Y}_t(\omega))\in\mathcal{B}\}\in\mathscr{H}_t \quad \text{and} \tag{2.23}$$

$$\forall \mathcal{B}\in\left(\bigotimes_{i\in\mathbb{N}_t}\mathscr{B}\left(\mathbb{R}^{N_i\times 1}\right)\right)\otimes\left(\bigotimes_{i\in\mathbb{N}_t}\mathscr{B}\left(\mathbb{R}^{M_i\times 1}\right)\right), \quad \forall t\in\mathbb{N}, \tag{2.24}$$

respectively (in the above, "$\otimes$" denotes the product operator for σ-algebras). Of course, the existence and almost everywhere uniqueness of Λ_t are guaranteed by the Radon-Nikodym Theorem, provided that the base measure $\mathcal{P}$ is absolutely continuous with respect to $\widetilde{\mathcal{P}}$ on $\mathscr{H}_t$ ($\mathcal{P}\ll_{\mathscr{H}_t}\widetilde{\mathcal{P}}$). Further, for the case where there exist densities characterizing $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ (as in Theorem 1), demanding that $\mathcal{P}\ll_{\mathscr{H}_t}\widetilde{\mathcal{P}}$ is precisely equivalent to demanding that (2.16) is true and, again through the Radon-Nikodym Theorem, it can be easily shown that the derivative Λ_t actually coincides with the likelihood ratio process defined in (2.17), almost everywhere. ■

Now, let us apply Theorem 1 for the stochastic processes X_t and $\mathbf{y}_t$, comprising our partially observed system, as defined in Section II.A. In this respect, we present the following result.

Theorem 2.2. (Change of Measure for the Hidden System under Study) *Consider the hidden stochastic system of Section II.A on the usual base space $(\Omega,\mathscr{F},\mathcal{P})$, where $X_t\in\mathcal{Z}$ and $\mathbf{y}_t\in\mathbb{R}^{N\times 1}$, almost surely $\forall t\in\mathbb{N}$, constitute the hidden state process and the observation process, respectively. Then, there exists an alternative,* ***equivalent*** *to $\mathcal{P}$, base measure $\widetilde{\mathcal{P}}$ on $(\Omega,\mathscr{F})$, under which:*

- *The processes X_t and $\mathbf{y}_t$ are statistically independent.*
- *X_t constitutes a stochastic process with exactly the same dynamics as under $\mathcal{P}$.*
- *$\mathbf{y}_t$ constitutes a Gaussian vector white noise process with zero mean and covariance matrix equal to the identity.*

Additionally, the filter $\widehat{X}_t$ can be expressed as in (2.18), where the $\{\mathscr{H}_t\}$-adapted stochastic process $\Lambda_t, t\in\mathbb{N}$ is defined as in (2.25) (top of next page).

$$\begin{aligned}\Lambda_t &\triangleq \prod_{i\in\mathbb{N}_t}\lambda_i \\ &\triangleq \prod_{i\in\mathbb{N}_t}\frac{\exp\left(\frac{1}{2}\left\|\mathbf{y}_i\right\|_2^2-\frac{1}{2}\left(\mathbf{y}_i-\boldsymbol{\mu}_i\left(X_i\right)\right)^{\boldsymbol{T}}\left(\boldsymbol{\Sigma}_i\left(X_i\right)+\sigma_\xi^2\mathbf{I}_{N\times N}\right)^{-1}\left(\mathbf{y}_i-\boldsymbol{\mu}_i\left(X_i\right)\right)\right)}{\sqrt{\det\left(\boldsymbol{\Sigma}_i\left(X_i\right)+\sigma_\xi^2\mathbf{I}_{N\times N}\right)}} \\ &\equiv \frac{\exp\left(\frac{1}{2}\sum_{i\in\mathbb{N}_t}\left\|\mathbf{y}_i\right\|_2^2-\left(\mathbf{y}_i-\boldsymbol{\mu}_i\left(X_i\right)\right)^{\boldsymbol{T}}\left(\boldsymbol{\Sigma}_i\left(X_i\right)+\sigma_\xi^2\mathbf{I}_{N\times N}\right)^{-1}\left(\mathbf{y}_i-\boldsymbol{\mu}_i\left(X_i\right)\right)\right)}{\prod_{i\in\mathbb{N}_t}\sqrt{\det\left(\boldsymbol{\Sigma}_i\left(X_i\right)+\sigma_\xi^2\mathbf{I}_{N\times N}\right)}}\in\mathbb{R}_{++}\end{aligned} \tag{2.25}$$

Proof of Theorem 2. Additionally to the similar identifications made above (see (2.19)) and for later reference, let

$$\boldsymbol{\mathcal{Y}}_t\triangleq\{\mathbf{y}_i\}_{i\in\mathbb{N}_t}\quad\text{and}\quad \boldsymbol{y}_t\triangleq\{\boldsymbol{y}_i\}_{i\in\mathbb{N}_t}. \tag{2.26}$$

First, we construct the probability measure $\widetilde{\mathcal{P}}$, this way showing its existence. To accomplish this, define, for each $t\in\mathbb{N}$, a probability measure $\widetilde{\mathcal{P}}_{\mathcal{R}_t}$ on the measurable space $(\mathcal{R}_t,\mathscr{B}(\mathcal{R}_t))$, where

$$\mathcal{R}_t\triangleq\left(\underset{i\in\mathbb{N}_t}{\times}\mathbb{R}\right)\times\left(\underset{i\in\mathbb{N}_t}{\times}\mathbb{R}^{N\times 1}\right), \tag{2.27}$$

being absolutely continuous with respect to the Lebesgue measure on $(\mathcal{R}_t,\mathscr{B}(\mathcal{R}_t))$ and with density $\widetilde{f}_t:\mathcal{R}\mapsto\mathbb{R}_+$. Since, for each $t\in\mathbb{N}$, the processes $X_t(\omega)$ and $\mathbf{y}_t(\omega)$ are both, by definition, *fixed* and *measurable* functions from $(\Omega,\mathscr{H}_t)$ to $(\mathcal{R}_t,\mathscr{B}(\mathcal{R}_t))$, with[2]

$$\mathscr{H}_t\subseteq\mathscr{H}_\infty\triangleq\sigma\left\{\bigcup_{t\in\mathbb{N}}\mathscr{H}_t\right\}\subseteq\mathscr{F}, \tag{2.28}$$

measuring any $\mathcal{B}\in\mathscr{B}(\mathcal{R}_t)$ under $\widetilde{\mathcal{P}}_{\mathcal{R}_t}$ can be replaced by measuring the event (preimage)

$\{\omega\in\Omega|\,(\mathcal{X}_t,\boldsymbol{\mathcal{Y}}_t)\in\mathcal{B}\}\in\mathscr{H}_t$ under another measure, say $\widetilde{\mathcal{P}}$, defined collectively for all $t\in\mathbb{N}$ on the general measurable space $(\Omega,\mathscr{H}_\infty)$ as

$$\widetilde{\mathcal{P}}\left(\{\omega\in\Omega|\,(\mathcal{X}_t,\boldsymbol{\mathcal{Y}}_t)\in\mathcal{B}\}\right)\equiv\widetilde{\mathcal{P}}\left((\mathcal{X}_t,\boldsymbol{\mathcal{Y}}_t)\in\mathcal{B}\right)\triangleq\widetilde{\mathcal{P}}_{\mathcal{R}_t}(\mathcal{B}),\quad\forall\mathcal{B}\in\mathscr{B}(\mathcal{R}_t).$$

[2] $\mathscr{H}_\infty$ constitutes the *join*, that is, the *smallest* σ-algebra generated by the union of all $\mathscr{H}_t,\forall t\in\mathbb{N}$.

That is, the restriction of the probability measure $\widetilde{\mathcal{P}}$ to the σ-algebra $\mathscr{H}_\infty$ is *induced* by the probability measure $\widetilde{\mathcal{P}}_{\mathcal{R}_\infty}$ (also see Kolmogorov's Extension Theorem [12]). Further, in order to define the alternative base measure $\widetilde{\mathcal{P}}$ fully on $(\Omega, \mathscr{F})$, we have to extend its behavior on the remaining events which belong to the potentially finer σ-algebra $\mathscr{F}$ but are not included in $\mathscr{H}_\infty$. However, since we are interested in change of measure only for the augmented process $(\mathcal{X}_t, \boldsymbol{\mathcal{Y}}_t)$, these events are irrelevant to us. Therefore, $\widetilde{\mathcal{P}}$ can be defined arbitrarily on these events, as long as it remains a valid and consistent probability measure.

Now, to finalize the construction of the restriction of $\widetilde{\mathcal{P}}$ to $\mathscr{H}_t, \forall t \in \mathbb{N}$, we have to explicitly specify the density of $\widetilde{\mathcal{P}}_{\mathcal{R}_t}$, or, equivalently, of the joint density of the random variables $(\mathcal{X}_t, \boldsymbol{\mathcal{Y}}_t)$, $\widetilde{f}_t$, for all $t \in \mathbb{N}$. According to the statement of Theorem 2, we have to demand that

$$\begin{aligned}
\widetilde{f}_t\left(\mathcal{x}_t, \mathcal{y}_t\right) &\equiv \widetilde{f}_{\boldsymbol{\mathcal{y}}_t|\mathcal{X}_t}\left(\mathcal{y}_t \middle|\, \mathcal{x}_t\right) \widetilde{f}_{\mathcal{X}_t}\left(\mathcal{x}_t\right) \\
&= \widetilde{f}_{\boldsymbol{\mathcal{y}}_t}\left(\mathcal{y}_t\right) f_{\mathcal{X}_t}\left(\mathcal{x}_t\right) \\
&= \left(\prod_{i \in \mathbb{N}_t} \widetilde{f}_{\mathbf{y}_i}\left(\boldsymbol{y}_i\right)\right) f_{\mathcal{X}_t}\left(\mathcal{x}_t\right) \\
&= \left(\prod_{i \in \mathbb{N}_t} \frac{\exp\left(-\frac{\|\boldsymbol{y}_i\|_2^2}{2}\right)}{\sqrt{(2\pi)^N}}\right) f_{\mathcal{X}_t}\left(\mathcal{x}_t\right) \\
&\equiv \frac{\exp\left(-\frac{1}{2}\sum_{i \in \mathbb{N}_t} \|\boldsymbol{y}_i\|_2^2\right)}{\sqrt{(2\pi)^{N(t+1)}}} f_{\mathcal{X}_t}\left(\mathcal{x}_t\right). \qquad (2.29)
\end{aligned}$$

Next, by definition, we know that, under $\mathcal{P}$, the joint density of $(\mathcal{X}_t, \boldsymbol{\mathcal{Y}}_t)$ can be expressed as

$$\begin{aligned}
f_t\left(\mathcal{x}_t, \mathcal{y}_t\right) &\equiv f_{\boldsymbol{\mathcal{y}}_t|\mathcal{X}_t}\left(\mathcal{y}_t \middle|\, \mathcal{x}_t\right) f_{\mathcal{X}_t}\left(\mathcal{x}_t\right) \\
&\equiv \left(\prod_{i \in \mathbb{N}_t} f_{\mathbf{y}_i|X_i}\left(\boldsymbol{y}_i \middle|\, x_i\right)\right) f_{\mathcal{X}_t}\left(\mathcal{x}_t\right)
\end{aligned}$$

$$= \left(\prod_{i \in \mathbb{N}_t} \frac{\exp\left(\frac{\overline{\boldsymbol{y}}_i^{\boldsymbol{T}} \mathbf{C}_i^{-1} \overline{\boldsymbol{y}}_i}{-2} \right)}{\sqrt{\det\left(\mathbf{C}_i\right) (2\pi)^N}} \right) f_{\mathcal{X}_t}\left(\mathcal{X}_t\right)$$

$$\equiv \frac{\exp\left(\sum_{i \in \mathbb{N}_t} \frac{\overline{\boldsymbol{y}}_i^{\boldsymbol{T}} \mathbf{C}_i^{-1} \overline{\boldsymbol{y}}_i}{-2} \right)}{\left(\prod_{i \in \mathbb{N}_t} \sqrt{\det\left(\mathbf{C}_i\right)} \right) \sqrt{(2\pi)^{N(t+1)}}} f_{\mathcal{X}_t}\left(\mathcal{X}_t\right), \tag{2.30}$$

where, for all $t \in \mathbb{N}$,

$$\overline{\boldsymbol{y}}_t \equiv \overline{\boldsymbol{y}}_t\left(x_t\right) \triangleq \boldsymbol{y}_t - \boldsymbol{\mu}_t\left(x_t\right) \in \mathbb{R}^{N \times 1} \quad \text{and} \tag{2.31}$$

$$\mathbf{C}_t \equiv \mathbf{C}_t\left(x_t\right) \equiv \boldsymbol{\Sigma}_t\left(x_t\right) + \sigma_\xi^2 \mathbf{I}_{N \times N} \in \mathcal{D}_{\mathbf{C}}, \tag{2.32}$$

where $\mathcal{D}_{\mathbf{C}}$ constitutes a bounded subset of $\mathbb{R}^{N \times N}$. From (2.29) and (2.30), it is obvious that the sufficient condition (2.16) of Theorem 1 is satisfied (actually, in this case, we have an equivalence; as a result, the change of measure is an invertible transformation). Applying Theorem 1, (2.18) must be true by defining the $\{\mathscr{H}_t\}$-adapted stochastic process

$$\Lambda_t \equiv \Lambda_t\left(\mathcal{X}_t, \boldsymbol{\mathcal{Y}}_t\right) \triangleq \frac{f_t\left(\mathcal{X}_t, \boldsymbol{\mathcal{Y}}_t\right)}{\widetilde{f}_t\left(\mathcal{X}_t, \boldsymbol{\mathcal{Y}}_t\right)} \equiv \frac{f_{\boldsymbol{\mathcal{Y}}_t|\mathcal{X}_t}\left(\boldsymbol{\mathcal{Y}}_t|\,\mathcal{X}_t\right)}{\widetilde{f}_{\boldsymbol{\mathcal{Y}}_t|\mathcal{X}_t}\left(\boldsymbol{\mathcal{Y}}_t|\,\mathcal{X}_t\right)} \equiv \frac{f_{\boldsymbol{\mathcal{Y}}_t|\mathcal{X}_t}\left(\boldsymbol{\mathcal{Y}}_t|\,\mathcal{X}_t\right)}{\widetilde{f}_{\boldsymbol{\mathcal{Y}}_t}\left(\boldsymbol{\mathcal{Y}}_t\right)}$$

$$\equiv \frac{\exp\left(\sum_{i \in \mathbb{N}_t} \frac{\|\mathbf{y}_i\|_2^2 - \left(\overline{\mathbf{y}}_i\left(X_i\right)\right)^{\boldsymbol{T}} \left(\mathbf{C}_i\left(X_i\right)\right)^{-1} \overline{\mathbf{y}}_i\left(X_i\right)}{2} \right)}{\prod_{i \in \mathbb{N}_t} \sqrt{\det\left(\mathbf{C}_i\left(X_i\right)\right)}}, \tag{2.33}$$

or, alternatively,

$$\Lambda_t \equiv \prod_{i \in \mathbb{N}_t} \lambda_i \triangleq \prod_{i \in \mathbb{N}_t} \frac{\exp\left(\sum_{i \in \mathbb{N}_t} \frac{\|\mathbf{y}_i\|_2^2 - \left(\overline{\mathbf{y}}_i\left(X_i\right)\right)^{\boldsymbol{T}} \left(\mathbf{C}_i\left(X_i\right)\right)^{-1} \overline{\mathbf{y}}_i\left(X_i\right)}{2} \right)}{\sqrt{\det\left(\mathbf{C}_i\left(X_i\right)\right)}}, \tag{2.34}$$

therefore completing the proof. ■

2.1.2.3 Weak & $\mathcal{C}$-Weak Convergence of (Random) Probability Measures

In the analysis that will take place in Section IV, we will make use of the notions of weak and conditionally weak ($\mathcal{C}$-weak) convergence of sequences of probability measures.

Thus, let us define these notions of stochastic convergence consistently, suited at least for the purposes of our investigation.

Definition 2.1. (Weak Convergence [28]) Let $\mathcal{S}$ be an arbitrary metric space, let $\mathscr{S} \triangleq \mathscr{B}(\mathcal{S})$ be the associated Borel σ-algebra and consider a sequence of probability measures $\{\pi_n\}_{n\in\mathbb{N}}$ on $\mathscr{S}$. If π constitutes another "limit" probability measure on $\mathscr{S}$ such that

$$\begin{gathered}\lim_{n\to\infty} \pi_n(\mathcal{A}) = \pi(\mathcal{A}), \\ \forall \mathcal{A} \in \mathscr{S} \text{ such that } \pi(\partial\mathcal{A}) \equiv 0,\end{gathered} \tag{2.35}$$

where $\partial\mathcal{A}$ denotes the boundary set of the Borel set $\mathcal{A}$, then we say that the sequence $\{\pi_n\}_{n\in\mathbb{N}}$ converges to π *weakly* or *in the weak sense* and we equivalently write

$$\pi_n \xrightarrow[n\to\infty]{\mathcal{W}} \pi. \tag{2.36}$$

Of course, weak convergence of probability measures is equivalent to weak convergence or convergence in distribution, in case we are given sequences of $(S, \mathscr{S})$-valued random variables whose induced probability measures converge in the aforementioned sense.

Next, we present a definition for conditionally weak convergence of probability measures. To avoid possibly complicating technicalities, this definition is not presented in full generality. Rather, it is presented in an appropriately specialized form, which will be used later on, in the analysis that follows.

Definition 2.2. (Conditionally Weak Convergence) Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a base probability triplet and consider the measurable spaces $\left(\mathcal{S}_i, \mathscr{S}_i \triangleq \mathscr{B}(\mathcal{S}_i)\right), i = \{1, 2\}$, where $\mathcal{S}_1$ and $\mathcal{S}_2$ constitute a complete separable metric (Polish) space and an arbitrary metric space, respectively. Also, let $\{X_1^n : \Omega \to \mathcal{S}_1\}_{n\in\mathbb{N}}$ be a sequence of random variables, let $X_2 : \Omega \to \mathcal{S}_2$ be another random variable and consider the sequence of *(regular) induced conditional probability distributions (or measures)* $\mathcal{P}^n_{X_1^n|X_2} : \mathscr{S}_1 \times \Omega \to [0, 1]$, such that

$$\mathcal{P}^n_{X_1^n|X_2}(\mathcal{A}|\, X_2(\omega)) \equiv \mathcal{P}(X_1^n \in \mathcal{A}|\, \sigma\{X_2\}), \tag{2.37}$$

$\mathcal{P}-a.e.$, for any Borel set $\mathcal{A}\in\mathscr{S}_1$. If $X_1:\Omega\rightarrow\mathcal{S}_1$ constitutes a "limit" random variable, whose induced conditional measure $\mathcal{P}_{X_1|X_2}:\mathscr{S}_1\times\Omega\rightarrow[0,1]$ is such that

$$\begin{gathered}\lim_{n\rightarrow\infty}\mathcal{P}^n_{X_1^n|X_2}\left(\mathcal{A}|\,X_2\left(\omega\right)\right)=\mathcal{P}_{X_1|X_2}\left(\mathcal{A}|\,X_2\left(\omega\right)\right),\\ \forall\mathcal{A}\in\mathscr{S}_1\text{ such that }\pi\left(\partial\mathcal{A}\right)\equiv 0\text{ and }\mathcal{P}-a.e.,\end{gathered}\tag{2.38}$$

then we say that the sequence $\left\{\mathcal{P}^n_{X_1^n|X_2}\right\}_{n\in\mathbb{N}}$ converges to $\mathcal{P}_{X_1|X_2}$ *conditionally weakly ($\mathcal{C}$-weakly)* or *in the conditionally weak ($\mathcal{C}$-weak) sense* and we equivalently write

$$\mathcal{P}^n_{X_1^n|X_2}\left(\cdot|\,X_2\right)\underset{n\rightarrow\infty}{\overset{\mathcal{W}}{\longrightarrow}}\mathcal{P}_{X_1|X_2}\left(\cdot|\,X_2\right).\tag{2.39}$$

Remark 2.3. Actually, $\mathcal{C}$-weak convergence, as defined above, is strongly related to the more general concepts of *almost sure weak convergence* and random probability measures. For instance, the reader is referred to the related articles [29] and [30]. ■

Further, the following lemma characterizes weak convergence of probability measures (and random variables) [28].

Lemma 2.1. (Weak Convergence & Expectations) *Let $\mathcal{S}$ be an arbitrary metric space and let $\mathscr{S}\triangleq\mathscr{B}\left(\mathcal{S}\right)$. Suppose we are given a sequence of random variables $\{X^n\}_{n\in\mathbb{N}}$ and a "limit" X, all $(S,\mathscr{S})$-valued, but possibly defined on different base probability spaces, with $\{\mathcal{P}_{X^n}\}_{n\in\mathbb{N}}$ and $\mathcal{P}_X$ being their induced probability measures on $\mathscr{S}$, respectively. Then,*

$$X^n\underset{n\rightarrow\infty}{\overset{\mathcal{D}}{\longrightarrow}}X\Leftrightarrow\mathcal{P}_{X^n}\underset{n\rightarrow\infty}{\overset{\mathcal{W}}{\longrightarrow}}\mathcal{P}_X,\tag{2.40}$$

if and only if

$$\mathbb{E}\left\{f\left(X^n\right)\right\}\equiv\int_{\mathcal{S}}f\mathrm{d}\mathcal{P}_{X^n}\underset{n\rightarrow\infty}{\longrightarrow}\int_{\mathcal{S}}f\mathrm{d}\mathcal{P}_X\equiv\mathbb{E}\left\{f\left(X\right)\right\},\tag{2.41}$$

for all bounded, continuous functions $f:\mathcal{S}\rightarrow\mathbb{R}$.

Of course, if we replace weak convergence by $\mathcal{C}$-weak convergence, Lemma 2.1 continues to hold, but, in this case, (2.41) should be understood in the almost everywhere sense (see, for example, [30]). More specifically, under the generic notation of Definition 2.2 and under the appropriate assumptions according to Lemma 2.1, it will be true that

$$\mathbb{E}\left\{f\left(X_1^n\right)|\,X_2\right\}\left(\omega\right)\underset{n\rightarrow\infty}{\longrightarrow}\mathbb{E}\left\{f\left(X_1\right)|\,X_2\right\}\left(\omega\right),\tag{2.42}$$

for almost all $\omega \in \Omega$.

2.1.3 Problem Formulation & Statement of Main Results

In this section, we formulate the problem of interest, that is, in a nutshell, the problem of approximating a nonlinear MMSE filter by another (asymptotically optimal) filtering operator, defined by replacing the true process we would like to filter by an appropriate approximation. Although we do not deal with such a problem here, such an approximation would be chosen in order to yield a practically realizable approximate filtering scheme. We also present the main result of this work, establishing sufficient conditions for convergence of the respective approximate filters, in an indeed strong sense.

Let us start from the beginning. From Theorem 2.2, we know that

$$\mathbb{E}_{\mathcal{P}}\left\{ X_t \middle| \mathscr{Y}_t \right\} \equiv \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ X_t \Lambda_t \middle| \mathscr{Y}_t \right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ \Lambda_t \middle| \mathscr{Y}_t \right\}}, \quad \forall t \in \mathbb{N}, \tag{2.43}$$

where the RHS constitutes an alternative representation for the filter on the LHS, which constitutes the optimal in the MMSE sense estimator of the partially observed process X_t, given the available observations up to time t. If the numerical evaluation of either of the sides of (2.43) is difficult (either we are interested in a recursive realization of the filter or not), one could focus on the RHS, where the state and the observations constitute independent processes, and, keeping the same observations, replace X_t by another process $X_t^{\mathtt{A}}$, called the *approximation*, with *resolution* or *approximation parameter* $\mathtt{A} \in \mathbb{N}$ (for simplicity), also independent of the observations (with respect to $\widetilde{\mathcal{P}}$), for which the evaluation of the resulting "filter" might be easier. Under some appropriate, well defined sense, the approximation to the original process improves as $\mathtt{A} \to \infty$. This general idea of replacing the true state process with an approximation is employed in, for instance, [19, 20], and will be employed here, too.

At this point, a natural question arises: Why are we complicating things with change of measure arguments and not using $X_t^{\mathtt{A}}$ directly in the LHS of (2.43)? Indeed, using classical results such as the Dominated Convergence Theorem, one could prove at least pointwise convergence of the respective filter approximations. The main and most important issue with such an approach is that, in order for such a filter to be realizable

in any way, special attention must be paid to the choice of the approximation, regarding its stochastic dependence on the observations process. This is due to the original stochastic coupling between the state and the observations of the hidden system of interest. However, using change of measure, one can find an alternative representation of the filter process, where, under another probability measure, the state and observations are stochastically decoupled (independent). This makes the problem much easier, because the approximation can also be chosen to be independent of the observations. If we especially restrict our attention to recursive nonlinear filters, change of measure provides a rather versatile means for discovering recursive filter realizations. See, for example, the detailed treatment presented in [12].

Thus, concentrating on the RHS of (2.43), we can define an *approximate filtering operator* of the process X_t, also with resolution $\mathtt{A} \in \mathbb{N}$, as

$$\mathcal{E}^{\mathtt{A}}\left(X_t \middle| \mathscr{Y}_t\right) \triangleq \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ X_t^{\mathtt{A}} \Lambda_t^{\mathtt{A}} \middle| \mathscr{Y}_t \right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ \Lambda_t^{\mathtt{A}} \middle| \mathscr{Y}_t \right\}}, \quad \forall t \in \mathbb{N}. \tag{2.44}$$

Observe that the above quantity *is not a conditional expectation of* $X_t^{\mathtt{A}}$, because $X_t^{\mathtt{A}}$ does not follow the probability law of the true process of interest, X_t [20]. Of course, the question is if and under which sense,

$$\mathcal{E}^{\mathtt{A}}\left(X_t \middle| \mathscr{Y}_t\right) \xrightarrow[\mathtt{A}\to\infty]{?} \mathbb{E}_{\mathcal{P}}\left\{X_t \middle| \mathscr{Y}_t\right\}, \tag{2.45}$$

that is, if and in which sense our chosen approximate filtering operator is *asymptotically optimal*, as the resolution of the approximation increases. In other words, we are looking for a class of approximations, whose members approximate the process X_t well, in the sense that the resulting approximate filtering operators converge to the true filter as the resolution parameter increases, that is, as $\mathtt{A} \to \infty$, and under some appropriate notion of convergence. In this respect, below we formulate and prove the following theorem, which constitutes the main result of this work (recall the definition of $\mathcal{C}$-weak convergence given in Section II.C). In the following, $\mathbb{1}_{\mathcal{A}} : \mathbb{R} \to \{0, 1\}$ denotes the indicator of the set $\mathcal{A}$. Also, for any Borel set $\mathcal{A}$, $\mathbb{1}_{\mathcal{A}}(\cdot)$ constitutes a Dirac (atomic) probability measure. Equivalently, we write $\mathbb{1}_{\mathcal{A}}(\cdot) \equiv \delta_{(\cdot)}(\mathcal{A})$.

Theorem 2.3. (Convergence to the Optimal Filter) *Pick any natural $T < \infty$ and suppose either of the following:*

- *For all $t \in \mathbb{N}_T$, the sequence $\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}$ is marginally $\mathcal{C}$-weakly convergent to X_t, given X_t, that is,*

$$\mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}\left(\cdot\middle|\, X_t\right) \xrightarrow[\mathtt{A}\to\infty]{\mathcal{W}} \delta_{X_t}\left(\cdot\right), \quad \forall t \in \mathbb{N}_T. \tag{2.46}$$

- *For all $t \in \mathbb{N}_T$, the sequence $\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}$ is (marginally) convergent to X_t in probability, that is,*

$$X_t^{\mathtt{A}} \xrightarrow[\mathtt{A}\to\infty]{\mathcal{P}} X_t, \quad \forall t \in \mathbb{N}_T. \tag{2.47}$$

Then, there exists a measurable subset $\widehat{\Omega}_T \subseteq \Omega$ with $\mathcal{P}$-measure at least $1-(T+1)^{1-CN}\exp\left(-CN\right)$, such that

$$\lim_{\mathtt{A}\to\infty}\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left|\mathcal{E}^{\mathtt{A}}\left(X_t\middle|\,\mathscr{Y}_t\right) - \mathbb{E}_{\mathcal{P}}\left\{X_t\middle|\,\mathscr{Y}_t\right\}\right|(\omega) \equiv 0, \tag{2.48}$$

for any free, finite constant $C \geq 1$. In other words, the convergence of the respective approximate filtering operators is compact in $t \in \mathbb{N}$ and, with probability at least $1-(T+1)^{1-CN}\exp\left(-CN\right)$, uniform in ω.

Interestingly, as noted in the beginning of this section, the mode of convergence of the resulting approximate filtering operator is particularly strong. In fact, it is interesting that, for fixed T, the approximate filter $\mathcal{E}^{\mathtt{A}}\left(X_t\middle|\,\mathscr{Y}_t\right)$ converges to $\mathbb{E}_{\mathcal{P}}\left\{X_t\middle|\,\mathscr{Y}_t\right\}$ (uniformly) in a set that approaches the certain event, exponentially in N. That is, convergence to the optimal filter ***tends*** *to be in the uniformly almost everywhere sense, at an exponential rate (in N).* Consequently, it is revealed that the dimensionality of the observations process essentially stabilizes the behavior of the approximate filter, in a stochastic sense. Along the lines of the discussion presented above, it is clear that Theorem 2.3 provides a way of quantitatively justifying Egoroff's theorem [21], which bridges almost uniform convergence with almost sure convergence, however in an indeed abstract fashion.

Remark 2.4. The $\mathcal{C}$-weak convergence condition (2.46) is a rather strong one. In particular, as we show later in Lemma 2.8 (see Section IV), it implies $\mathcal{L}_1$ convergence, which

means that it also implies (marginal) convergence in probability (which constitutes the alternative sufficient condition of Theorem 2.3). In simple words, (2.46) resembles a situation where, at any time step, one is given or defines an approximation to the original process, in the sense that, *conditioned on the original process* at the same time step, the probability of being equal to the latter approaches unity. At this point, because $\mathcal{C}$-weak convergence is stronger than (and implies) convergence in probability, one could wonder why we presented both as alternative sufficient conditions for filter convergence in Theorem 2.3 (and also in Lemma 2.10 presented in Section IV). The reason is that, contrary to convergence in probability, condition (2.46) provides a nice *structural* criterion for *constructing* state process approximations in a natural way, which is also consistent with our intuition: If, at any time step, we could observe the value of true state process, then the respective value of the approximation at that same time step should be "sufficiently close" to the value of the state. Condition (2.46) expresses this intuitive idea and provides a version of the required sense of "closeness". ■

In order to demonstrate the applicability of Theorem 2.3, as well as demystify the $\mathcal{C}$-weak convergence condition (2.46), let us present a simple but illustrative example. The example refers to a class of approximate grid based filters, based on the so called marginal approximation [22], according to which the (compactly restricted) state process is fed into a uniform spatial quantizer of variable resolution. As we will see, this intuitively reasonable approximation idea constitutes a simple instance of the condition (2.46).

More specifically, assume that $X_t \in [a,b] \equiv \mathcal{Z}$, $\forall t \in \mathbb{N}$, almost surely. Let us discretize $\mathcal{Z}$ uniformly into $\mathtt{A}$ subintervals, of identical length, called cells. The l-th cell and its respective center are denoted as $\mathcal{Z}_{\mathtt{A}}^{l}$ and $x_{\mathtt{A}}^{l}, l \in \mathbb{N}_{\mathtt{A}}^{+}$. Then, letting $\mathcal{X}_{\mathtt{A}} \triangleq \left\{x_{\mathtt{A}}^{l}\right\}_{l\in\mathbb{N}_{L_S}^{+}}$, the *quantizer* $\mathcal{Q}_{\mathtt{A}} : (\mathcal{Z}, \mathscr{B}(\mathcal{Z})) \mapsto \left(\mathcal{X}_{\mathtt{A}}, 2^{\mathcal{X}_{\mathtt{A}}}\right)$ is defined as the bijective and measurable function which uniquely maps the l-th cell to the respective *reconstruction point* $x_{\mathtt{A}}^{l}$, $\forall l \in \mathbb{N}_{\mathtt{A}}^{+}$. That is, $\mathcal{Q}_{\mathtt{A}}(x) \triangleq x_{\mathtt{A}}^{l}$ if and only if $x \in \mathcal{Z}_{\mathtt{A}}^{l}$. Having defined the quantizer $\mathcal{Q}_{\mathtt{A}}(\cdot)$, the *Marginal Quantization* of the state is defined as [22]

$$X_t^{\mathtt{A}}(\omega) \triangleq \mathcal{Q}_{\mathtt{A}}(X_t(\omega)) \in \mathcal{X}_{\mathtt{A}},\ \forall t \in \mathbb{N},\ \mathcal{P}-a.s., \tag{2.49}$$

where $\mathtt{A} \in \mathbb{N}$ is identified as the approximation parameter. That is, X_t is approximated by its nearest neighbor on the cell grid. That is, the state is represented by a discrete set of reconstruction points, each one of them uniquely corresponding to a member of a partition of $\mathcal{Z}$.

By construction of marginal state approximations, it can be easily shown that (see Section 2.2)

$$X_t^{\mathtt{A}}(\omega) \overset{\mathcal{P}-a.s.}{\underset{\mathtt{A}\to\infty}{\longrightarrow}} X_t(\omega), \tag{2.50}$$

a fact that will be used in the following. Of course, almost sure convergence implies convergence in probability and, as we will see, $\mathcal{C}$-weak convergence as well. First, let us determine the conditional probability measure $\mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}(\mathrm{d}x|\, X_t)$. Since knowing X_t uniquely determines the value of $X_t^{\mathtt{A}}$, it must be true that

$$\begin{aligned}\mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}(\mathrm{d}x|\, X_t) &\equiv \mathcal{P}^{\mathtt{A}}_{\mathcal{Q}_{\mathtt{A}}(X_t)|X_t}(\mathrm{d}x|\, X_t)\\ &\equiv \delta_{\mathcal{Q}_{\mathtt{A}}(X_t)}(\mathrm{d}x), \quad \mathcal{P}-a.s..\end{aligned} \tag{2.51}$$

However, from Lemma 2.1, we know that weak convergence of measures is equivalent to showing that the expectations $\mathbb{E}\left\{ f\left(X_t^{\mathtt{A}}\right) \middle|\, X_t \right\}$ converge to $\mathbb{E}\{f(X_t)|\, X_t\} \equiv f(X_t)$, for all bounded and continuous $f(\cdot)$, almost everywhere. Indeed,

$$\begin{aligned}\mathbb{E}\left\{ f\left(X_t^{\mathtt{A}}\right) \middle|\, X_t \right\}(\omega) &\equiv \int_{\mathcal{Z}} f(x)\, \mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}(\mathrm{d}x|\, X_t(\omega))\\ &\equiv \int_{\mathcal{Z}} f(x)\, \delta_{\mathcal{Q}_{\mathtt{A}}(X_t(\omega))}(\mathrm{d}x)\\ &\equiv f(\mathcal{Q}_{\mathtt{A}}(X_t(\omega))) \overset{\mathcal{P}-a.s.}{\underset{\mathtt{A}\to\infty}{\longrightarrow}} f(X_t(\omega)),\end{aligned} \tag{2.52}$$

due to the continuity of $f(\cdot)$. Consequently, we have shown that

$$\mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}(\cdot|\, X_t) \equiv \delta_{\mathcal{Q}_{\mathtt{A}}(X_t)}(\cdot) \overset{\mathcal{W}}{\underset{\mathtt{A}\to\infty}{\longrightarrow}} \delta_{X_t}(\cdot), \tag{2.53}$$

fulfilling the first requirement of Theorem 2.3. This very simple example constitutes the basis for constructing more complicated and cleverly designed state approximations (for example, using stochastic quantizers). The challenge here is to come up with such approximations exhibiting nice properties, which would potentially lead to the development of effective approximate recursive or, in general, sequential filtering schemes,

well suited for dynamic inference in complex partially observable stochastic nonlinear systems. As far as grid based approximate recursive filtering is concerned, a relatively complete discussion of the problem is presented in Section 2.2 of this dissertation, where marginal state approximations are also treated in full generality.

An important and direct consequence of Theorem 2.3, also highlighted by the example presented above, is that, interestingly, the nature of the state process is completely irrelevant when one is interested in convergence of the respective approximate filters, in the respective sense of the aforementioned theorem. This fact has the following pleasing and intuitive interpretation: It implies that if any of the two conditions of Theorem 2.3 are satisfied, then we should forget about the internal stochastic structure of the state, and instead focus exclusively on the way the latter is being observed through time. That is, *we do not really care about what we partially observe, but how well we observe it; and if we observe it well, we can filter it well, too.* Essentially, the observations should constitute a stable functional of the state, of course in some well defined sense. In this work, this notion of stability is expressed precisely through Assumption 2.1 and 2.2, presented earlier in Section II.

Note, however, that the existence of a consistent approximate filter in the sense of Theorem 2.3 does not automatically imply that this filter will be efficiently implementable; usually, we would like such a filter to admit a recursive/sequential representation (or possibly a semirecursive one [24]). As it turns out, this can happen when the chosen state approximation admits a valid semimartingale type representation (in addition to satisfying one of the sufficient conditions of Theorem 2.3). For example, the case where the state is Markovian and the chosen state approximation is of the marginal type, discussed in the basic example presented above, is treated in detail in Section 2.2.

Remark 2.5. The filter representation (2.43) coincides with the respective expression employed in importance sampling [23, 31]. Since, under the alternative measure $\widetilde{\mathcal{P}}$, the observations and state constitute statistically independent processes, one can directly sample from the (joint) distribution of the state, fixing the observations to their respective value at each time t (of course, assuming that a relevant "sampling device" exists).

However, note that that due to the assumptions of Theorem 2.3, related at least to convergence in probability of the corresponding state approximations, the aforementioned result cannot be used directly in order to show convergence of importance sampling or related particle filtering techniques, which are directly related to empirical measures. The possible ways Theorem 2.3 can be utilized in order to provide asymptotic guarantees for particle filtering (using additional assumptions) constitutes an interesting open topic for further research. ■

The rest of our development is fully devoted in the detailed proof of Theorem 2.3.

2.1.4 Proof of Theorem 2.3

In order to facilitate the presentation, the proof is divided in a number of subsections.

2.1.4.1 Two Basic Lemmata, Linear Algebra - Oriented

Parts of the following useful results will be employed several times in the analysis that follows[3].

Lemma 2.2. *Consider arbitrary matrices* $\mathbf{A} \in \mathbb{C}^{N_1 \times M_1}$, $\mathbf{B} \in \mathbb{C}^{N_1 \times M_1}$, $\mathbf{X} \in \mathbb{C}^{M_2 \times N_2}$, $\mathbf{Y} \in \mathbb{C}^{M_2 \times N_2}$, *and let* $\|\cdot\|_{\mathfrak{M}}$ *be any matrix norm. Then, the following hold:*

- *If either*
  - $N_1 \equiv M_1 \equiv 1$, *or*
  - $N_1 \equiv N_2 \equiv M_1 \equiv M_2$ *and* $\|\cdot\|_{\mathfrak{M}}$ *is submultiplicative,*

  then

$$\|\mathbf{AX} - \mathbf{BY}\|_{\mathfrak{M}} \leq \|\mathbf{A}\|_{\mathfrak{M}} \|\mathbf{X} - \mathbf{Y}\|_{\mathfrak{M}} + \|\mathbf{Y}\|_{\mathfrak{M}} \|\mathbf{A} - \mathbf{B}\|_{\mathfrak{M}} . \tag{2.54}$$

- *If* $N_2 \equiv 1$, $M_1 \equiv M_2$ *and* $\|\cdot\|_{\mathfrak{M}}$ *constitutes any subordinate matrix norm to the* ℓ_p *vector norm,* $\|\cdot\|_p$, *then*

$$\|\mathbf{AX} - \mathbf{BY}\|_p \leq \|\mathbf{A}\|_{\mathfrak{M}} \|\mathbf{X} - \mathbf{Y}\|_p + \|\mathbf{Y}\|_p \|\mathbf{A} - \mathbf{B}\|_{\mathfrak{M}} . \tag{2.55}$$

[3] In this work, Lemma 2.3 presented in this subsection will be applied only for scalars (and where the metric considered coincides with the absolute value). However, the general version of the result (considering matrices and submultiplicative norms) is presented for the sake of generality.

Proof of Lemma 2.2. We prove the result only for the case where $N_1 \equiv N_2 \equiv M_1 \equiv M_2$ and $\|\cdot\|_{\mathfrak{M}}$ is submultiplicative. By definition of such a matrix norm,

$$\begin{aligned}\|\mathbf{AX}-\mathbf{BY}\|_{\mathfrak{M}} &\equiv \|\mathbf{AX}+\mathbf{AY}-\mathbf{AY}-\mathbf{BY}\|_{\mathfrak{M}}\\ &\equiv \|\mathbf{A}(\mathbf{X}-\mathbf{Y})+(\mathbf{A}-\mathbf{B})\mathbf{Y}\|_{\mathfrak{M}}\\ &\leq \|\mathbf{A}(\mathbf{X}-\mathbf{Y})\|_{\mathfrak{M}}+\|(\mathbf{A}-\mathbf{B})\mathbf{Y}\|_{\mathfrak{M}}\\ &\leq \|\mathbf{A}\|_{\mathfrak{M}}\|(\mathbf{X}-\mathbf{Y})\|_{\mathfrak{M}}+\|\mathbf{Y}\|_{\mathfrak{M}}\|(\mathbf{A}-\mathbf{B})\|_{\mathfrak{M}}\,, \end{aligned} \tag{2.56}$$

apparently completing the proof. The results for the other two cases considered in Lemma 2.2 can be readily shown following similar procedure. ■

Lemma 2.3. *Consider the collections of arbitrary, square matrices*

$$\left\{\mathbf{A}_i \in \mathbb{C}^{N\times N}\right\}_{i\in\mathbb{N}_n} \quad \textit{and} \quad \left\{\mathbf{B}_i \in \mathbb{C}^{N\times N}\right\}_{i\in\mathbb{N}_n}.$$

Then, for any submultiplicative matrix norm $\|\cdot\|_{\mathfrak{M}}$, it is true that

$$\left\|\prod_{i=0}^{n}\mathbf{A}_i-\prod_{i=0}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}} \leq \sum_{i=0}^{n}\left(\prod_{j=0}^{i-1}\|\mathbf{A}_j\|_{\mathfrak{M}}\right)\left(\prod_{j=i+1}^{n}\|\mathbf{B}_j\|_{\mathfrak{M}}\right)\|\mathbf{A}_i-\mathbf{B}_i\|_{\mathfrak{M}}\,. \tag{2.57}$$

Proof of Lemma 2.3. Applying Lemma 2.2 to the LHS of (2.57), we get

$$\begin{aligned}\left\|\prod_{i=0}^{n}\mathbf{A}_i-\prod_{i=0}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}} &\equiv \left\|\mathbf{A}_0\prod_{i=1}^{n}\mathbf{A}_i-\mathbf{B}_0\prod_{i=1}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}}\\ &\leq \|\mathbf{A}_0\|_{\mathfrak{M}}\left\|\prod_{i=1}^{n}\mathbf{A}_i-\prod_{i=1}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}}+\left\|\prod_{i=1}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}}\|\mathbf{A}_0-\mathbf{B}_0\|_{\mathfrak{M}}\,. \end{aligned} \tag{2.58}$$

The repeated application of Lemma 2.2 to the quantity multiplying $\|\mathbf{A}_0\|_{\mathfrak{M}}$ on the RHS of the expression above yields

$$\begin{aligned}\left\|\prod_{i=0}^{n}\mathbf{A}_i-\prod_{i=0}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}} &\leq \|\mathbf{A}_0\|_{\mathfrak{M}}\|\mathbf{A}_1\|_{\mathfrak{M}}\left\|\prod_{i=2}^{n}\mathbf{A}_i-\prod_{i=2}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}}\\ &\quad+\|\mathbf{A}_0\|_{\mathfrak{M}}\left\|\prod_{i=2}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}}\|\mathbf{A}_1-\mathbf{B}_1\|_{\mathfrak{M}}+\left\|\prod_{i=1}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}}\|\mathbf{A}_0-\mathbf{B}_0\|_{\mathfrak{M}}\,, \end{aligned} \tag{2.59}$$

where, the "temporal pattern" is apparent. Indeed, iterating (2.59) and proceeding inductively, we end up with the bound

$$\left\|\prod_{i=0}^{n}\mathbf{A}_i-\prod_{i=0}^{n}\mathbf{B}_i\right\|_{\mathfrak{M}} \leq \sum_{i=0}^{n}\left(\prod_{j=0}^{i-1}\|\mathbf{A}_j\|_{\mathfrak{M}}\right)\left\|\prod_{j=i+1}^{n}\mathbf{B}_j\right\|_{\mathfrak{M}}\|\mathbf{A}_i-\mathbf{B}_i\|_{\mathfrak{M}} \tag{2.60}$$

and the result readily follows invoking the submultiplicativeness of $\|\cdot\|_{\mathfrak{M}}$. ■

2.1.4.2 Preliminary Results

Here, we present and prove a number of preliminary results, which will help us towards the proof of an important lemma, which will be the key to showing the validity of Theorem 2.3.

First, under Assumption 2.2, stated in Section II.A, the following trivial lemmata hold.

Lemma 2.4. *Each member of the functional family* $\{\mathbf{\Sigma}_t : \mathcal{Z} \mapsto \mathcal{D}_{\mathbf{\Sigma}}\}_{t\in\mathbb{N}}$ *is Lipschitz continuous on* $\mathcal{Z}$*, in the Euclidean topology induced by the Frobenius norm. That is,* $\forall t \in \mathbb{N}$*,*

$$\left\|\mathbf{\Sigma}_t(x) - \mathbf{\Sigma}_t(y)\right\|_F \leq (NK_{\mathbf{\Sigma}})\,|x-y|\,, \tag{2.61}$$

$\forall\,(x,y) \in \mathcal{Z}\times\mathcal{Z}$*, for the same constant* $K_{\mathbf{\Sigma}} \in \mathbb{R}_+$*, as defined in Assumption 2.2. The same also holds for the family* $\{\mathbf{C}_t : \mathcal{Z} \mapsto \mathcal{D}_{\mathbf{C}}\}_{t\in\mathbb{N}}$*.*

Proof of Lemma 2.4. By definition of the Frobenius norm,

$$\begin{aligned}\left\|\mathbf{\Sigma}_t(x) - \mathbf{\Sigma}_t(y)\right\|_F &\equiv \sqrt{\sum_{(i,j)\in\mathbb{N}_N^+\times\mathbb{N}_N^+}\left(\mathbf{\Sigma}_t^{ij}(x) - \mathbf{\Sigma}_t^{ij}(y)\right)^2}\\ &\leq \sqrt{\sum_{(i,j)\in\mathbb{N}_N^+\times\mathbb{N}_N^+} K_{\mathbf{\Sigma}}^2\,|x-y|^2}\\ &\equiv \sqrt{N^2K_{\mathbf{\Sigma}}^2\,|x-y|^2}, \quad \forall t\in\mathbb{N}\end{aligned} \tag{2.62}$$

and our first claim follows. The second follows trivially if we recall the definition of each $\mathbf{C}_t(x)$. ■

Lemma 2.5. *For each member of the functional family* $\{\mathbf{C}_t : \mathcal{Z} \mapsto \mathcal{D}_{\mathbf{C}}\}_{t\in\mathbb{N}}$*, it is true that,* $\forall t \in \mathbb{N}$*,*

$$\left|\det\left(\mathbf{C}_t(x)\right) - \det\left(\mathbf{C}_t(y)\right)\right| \leq (NK_{DET})\left\|\mathbf{C}_t(x) - \mathbf{C}_t(y)\right\|_F, \tag{2.63}$$

$\forall\,(x,y) \in \mathcal{Z}\times\mathcal{Z}$*, for some bounded constant* $K_{DET} \equiv K_{DET}(N) \in \mathbb{R}_+$*, possibly dependent on* N *but independent of* t*.*

Proof of Lemma 2.5. As a consequence of the fact that the determinant of a matrix can be expressed as a polynomial function in N^2 variables (for example, see the Leibniz formula), it must be true that, $\forall t \in \mathbb{N}$,

$$\begin{aligned} \left|\det\left(\mathbf{C}_t\left(x\right)\right) - \det\left(\mathbf{C}_t\left(y\right)\right)\right| &\leq K_{DET} \sum_{(i,j)\in\mathbb{N}_N^+\times\mathbb{N}_N^+} \left|\mathbf{C}_t^{ij}\left(x\right) - \mathbf{C}_t^{ij}\left(y\right)\right| \\ &\equiv K_{DET}\left\|\mathbf{C}_t\left(x\right) - \mathbf{C}_t\left(y\right)\right\|_1, \end{aligned} \tag{2.64}$$

where the constant K_{DET} depends on *maximized* (using the fact that the domain $\mathcal{D}_{\mathbf{C}}$ is bounded) $(N-1)$-fold products of elements of $\mathbf{C}_t\left(x\right)$ and $\mathbf{C}_t\left(y\right)$, with respect to x (resp. y) and t. Consequently, although K_{DET} may depend on N, it certainly does not depend on t. Now, since the ℓ_1 entrywise norm of an $N \times N$ matrix corresponds to the norm of a vector with N^2 elements, we may further bound the right had side of the expression above by the Frobenius norm of $\mathbf{C}_t\left(x\right) - \mathbf{C}_t\left(y\right)$, yielding

$$\left|\det\left(\mathbf{C}_t\left(x\right)\right) - \det\left(\mathbf{C}_t\left(y\right)\right)\right| \leq NK_{DET}\left\|\mathbf{C}_t\left(x\right) - \mathbf{C}_t\left(y\right)\right\|_F, \tag{2.65}$$

which is what we were set to prove. ■

Remark 2.6. The fact that the constant K_{DET} may be a function of the dimension of the observation vector, N, does not constitute a significant problem throughout our analysis, simply because N is always considered a *finite and fixed* parameter of our problem. However, it is true that the (functional) way N appears in the various constants in our derived expressions can potentially affect speed of convergence and, for that reason, it constitutes an important analytical aspect. Therefore, throughout the analysis presented below, a great effort has been made in order to keep the dependence of our bounds on N within reasonable limits. ■

We also present another useful lemma, related to the expansiveness of each member of the functional family $\left\{\mathbf{C}_t^{-1} : \mathcal{Z} \mapsto \mathcal{D}_{\mathbf{C}^{-1}}\right\}_{t\in\mathbb{N}}$.

Lemma 2.6. *Each member of the functional family* $\left\{\mathbf{C}_t^{-1} : \mathcal{Z} \mapsto \mathcal{D}_{\mathbf{C}^{-1}}\right\}_{t\in\mathbb{N}}$ *is Lipschitz continuous on* $\mathcal{Z}$*, in the Euclidean topology induced by the Frobenius norm. That is,* $\forall t \in \mathbb{N}$*,*

$$\left\|\mathbf{C}_t^{-1}\left(x\right) - \mathbf{C}_t^{-1}\left(y\right)\right\|_F \leq K_{INV}\left|x - y\right|, \tag{2.66}$$

$\forall\,(x,y)\in\mathcal{Z}\times\mathcal{Z}$, *for some bounded constant* $K_{INV}\equiv K_{INV}\,(N)\in\mathbb{R}_{+}$, *possibly dependent on* N *but independent of* t.

Proof of Lemma 2.6. As a consequence of Laplace's formula for the determinant of a matrix and invoking Lemma 2.2, it is true that

$$\begin{aligned}
&\left\|\mathbf{C}_t^{-1}\,(x)-\mathbf{C}_t^{-1}\,(y)\right\|_F\\
&\equiv\left\|\frac{\mathrm{adj}\,(\mathbf{C}_t\,(x))}{\det\,(\mathbf{C}_t\,(x))}-\frac{\mathrm{adj}\,(\mathbf{C}_t\,(y))}{\det\,(\mathbf{C}_t\,(y))}\right\|_F\\
&\leq\frac{\|\mathrm{adj}\,(\mathbf{C}_t\,(x))-\mathrm{adj}\,(\mathbf{C}_t\,(y))\|_F}{\det\,(\mathbf{C}_t\,(x))}+\|\mathrm{adj}\,(\mathbf{C}_t\,(y))\|_F\,\frac{|\det\,(\mathbf{C}_t\,(x))-\det\,(\mathbf{C}_t\,(y))|}{\det\,(\mathbf{C}_t\,(x))\det\,(\mathbf{C}_t\,(y))},
\end{aligned}\tag{2.67}$$

where $\mathrm{adj}\,(\mathbf{A})$ denotes the adjugate of the square matrix $\mathbf{A}$. Since $\mathbf{C}_t\,(x)$ (resp. $\mathbf{C}_t\,(y)$) is a symmetric and positive definite matrix, so is its adjugate. Employing one more property regarding the eigenvalues of the adjugate [32] and the fact that $\lambda_{inf}>1$, we can write

$$\begin{aligned}
\|\mathrm{adj}\,(\mathbf{C}_t\,(y))\|_F&\leq\sqrt{N}\,\|\mathrm{adj}\,(\mathbf{C}_t\,(y))\|_2\\
&\equiv\sqrt{N}\lambda_{max}\,(\mathrm{adj}\,(\mathbf{C}_t\,(y)))\\
&\equiv\sqrt{N}\max_{i\in\mathbb{N}_N^+}\prod_{j\neq i}\lambda_j\,(\mathbf{C}_t\,(y))\\
&\leq\sqrt{N}\det\,(\mathbf{C}_t\,(y)),
\end{aligned}\tag{2.68}$$

and then (2.67) becomes

$$\begin{aligned}
\left\|\mathbf{C}_t^{-1}\,(x)-\mathbf{C}_t^{-1}\,(y)\right\|_F&\leq\frac{\|\mathrm{adj}\,(\mathbf{C}_t\,(x))-\mathrm{adj}\,(\mathbf{C}_t\,(y))\|_F}{\det\,(\mathbf{C}_t\,(x))}+\sqrt{N}\frac{|\det\,(\mathbf{C}_t\,(x))-\det\,(\mathbf{C}_t\,(y))|}{\det\,(\mathbf{C}_t\,(x))}\\
&\leq\frac{\|\mathrm{adj}\,(\mathbf{C}_t\,(x))-\mathrm{adj}\,(\mathbf{C}_t\,(y))\|_F}{\lambda_{inf}^N}+\frac{N^3K_{DET}K_{\mathbf{\Sigma}}}{\lambda_{inf}^N}\,|x-y|\,.
\end{aligned}\tag{2.69}$$

Next, the numerator of the first fraction from the left may be expressed as

$$\begin{aligned}
\|\mathrm{adj}\,(\mathbf{C}_t\,(x))-\mathrm{adj}\,(\mathbf{C}_t\,(y))\|_F&\equiv\sqrt{\sum_{(i,j)\in\mathbb{N}_N^+\times\mathbb{N}_N^+}\left(\mathrm{adj}\,(\mathbf{C}_t\,(x))^{ij}-\mathrm{adj}\,(\mathbf{C}_t\,(y))^{ij}\right)^2}\\
&\equiv\sqrt{\sum_{(i,j)\in\mathbb{N}_N^+\times\mathbb{N}_N^+}\left((-1)^{i+j}\left[\mathcal{M}_{ij}\,(\mathbf{C}_t\,(x))-\mathcal{M}_{ij}\,(\mathbf{C}_t\,(y))\right]\right)^2}
\end{aligned}$$

$$\equiv \sqrt{\sum_{(i,j)\in\mathbb{N}_N^+\times\mathbb{N}_N^+} \left(\mathcal{M}_{ij}\left(\mathbf{C}_t\left(x\right)\right)-\mathcal{M}_{ij}\left(\mathbf{C}_t\left(y\right)\right)\right)^2}, \quad (2.70)$$

where $\mathcal{M}_{ij}\left(\mathbf{C}_t\left(x\right)\right)$ denotes the (i,j)-th minor of $\mathbf{C}_t\left(x\right)$, which constitutes the *determinant* of the $(N-1)\times(N-1)$ matrix formulated by removing the i-th row and the j-th column of $\mathbf{C}_t\left(x\right)$. Consequently, from Lemma 2.5, there exists a constant K_{det}, possibly dependent on N, such that, $\forall t\in\mathbb{N}$,

$$\left\|\mathrm{adj}\left(\mathbf{C}_t\left(x\right)\right)-\mathrm{adj}\left(\mathbf{C}_t\left(y\right)\right)\right\|_F \leq \sqrt{\sum_{(i,j)\in\mathbb{N}_N^+\times\mathbb{N}_N^+} N^4K_{det}^2K_{\mathbf{\Sigma}}^2\left|x-y\right|^2}, \quad (2.71)$$

or, equivalently,

$$\left\|\mathrm{adj}\left(\mathbf{C}_t\left(x\right)\right)-\mathrm{adj}\left(\mathbf{C}_t\left(y\right)\right)\right\|_F \leq N^3K_{det}K_{\mathbf{\Sigma}}\left|x-y\right|, \quad (2.72)$$

$\forall\left(x,y\right)\in\mathcal{Z}\times\mathcal{Z}$. Therefore, combining with (2.69), we get

$$\begin{aligned}\left\|\mathbf{C}_t^{-1}\left(x\right)-\mathbf{C}_t^{-1}\left(y\right)\right\|_F &\leq \frac{N^3}{\lambda_{inf}^N}\left(K_{DET}+K_{det}\right)K_{\mathbf{\Sigma}}\left|x-y\right| \\ &\leq \frac{27\lambda_{inf}^{-3/\log\left(\lambda_{inf}\right)}}{\left(\log\left(\lambda_{inf}\right)\right)^3}\left(K_{DET}+K_{det}\right)K_{\mathbf{\Sigma}}\left|x-y\right| \\ &\triangleq K_{INV}\left|x-y\right|, \end{aligned} \quad (2.73)$$

and the proof is complete. ■

Next, we state the following simple probabilistic result, related to the expansiveness of the norm of the observation vector in a stochastic sense, under both base measures $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ considered (see Section 2.1.2.2).

Lemma 2.7. *Consider the random quadratic form*

$$Q_t\left(\omega\right)\triangleq\left\|\mathbf{y}_t\left(\omega\right)\right\|_2^2\equiv\left\|\overline{\mathbf{y}}_t\left(X_t\left(\omega\right)\right)+\boldsymbol{\mu}_t\left(X_t\left(\omega\right)\right)\right\|_2^2, \quad t\in\mathbb{N}. \quad (2.74)$$

Then, for any fixed $t\in\mathbb{N}$ and any freely chosen $C\geq 1$, there exists a bounded constant $\gamma>1$, such that the measurable set

$$\mathcal{T}_t\triangleq\left\{\omega\in\Omega\,\middle|\,\sup_{i\in\mathbb{N}_t}Q_i\left(\omega\right)<\gamma CN\left(1+\log\left(t+1\right)\right)\right\} \quad (2.75)$$

satisfies

$$\min\left\{\mathcal{P}\left(\mathcal{T}_t\right), \widetilde{\mathcal{P}}\left(\mathcal{T}_t\right)\right\} \geq 1 - \frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN-1}}, \tag{2.76}$$

that is, the sequence of quadratic forms $\left\{Q_i\left(\omega\right)\right\}_{i\in\mathbb{N}_t}$ *is uniformly bounded with very high probability under both base measures* $\mathcal{P}$ *and* $\widetilde{\mathcal{P}}$*.*

Proof of Lemma 2.7. First, it is true that

$$\begin{aligned}
\left\|\mathbf{y}_t\left(\omega\right)\right\|_2^2 &\equiv \left\|\overline{\mathbf{y}}_t\left(X_t\left(\omega\right)\right) + \boldsymbol{\mu}_t\left(X_t\left(\omega\right)\right)\right\|_2^2 \\
&\equiv \left\|\overline{\mathbf{y}}_t\left(X_t\left(\omega\right)\right)\right\|_2^2 + 2\mathbf{y}_t^{\boldsymbol{T}}\left(X_t\left(\omega\right)\right)\boldsymbol{\mu}_t\left(X_t\left(\omega\right)\right) + \left\|\boldsymbol{\mu}_t\left(X_t\left(\omega\right)\right)\right\|_2^2 \\
&\leq \left\|\overline{\mathbf{y}}_t\left(X_t\left(\omega\right)\right)\right\|_2^2 + 2\left\|\overline{\mathbf{y}}_t\left(X_t\left(\omega\right)\right)\right\|_2\mu_{sup} + \mu_{sup}^2.
\end{aligned} \tag{2.77}$$

Also, under $\mathcal{P}$, for each $t \in \mathbb{N}$, the random variable $\overline{\mathbf{y}}_t\left(X_t\right)$ constitutes an N-dimensional, conditionally (on X_t) Gaussian random variable with zero mean and covariance matrix $\mathbf{C}_t\left(X_t\right)$, that is

$$\overline{\mathbf{y}}_t\left|X_t\right. \sim \mathcal{N}\left(0, \mathbf{C}_t\left(X_t\right) \equiv \mathbf{C}_{\overline{\mathbf{y}}_t|X_t}\right). \tag{2.78}$$

Then, *if* X_t *is given*,

$$\overline{Q}_t\left(\omega\right) \triangleq \left\|\overline{\mathbf{y}}_t\left(X_t\left(\omega\right)\right)\right\|_2^2 \tag{2.79}$$

can be shown to admit the very useful alternative representation (for instance, see [33], pp. 89 - 90)

$$\overline{Q}_t \equiv \sum_{j\in\mathbb{N}_N^+} \lambda_j\left(\mathbf{C}_t\left(X_t\right)\right) U_j^2, \quad \forall t \in \mathbb{N}, \quad \text{with} \tag{2.80}$$

$$\left\{U_j\right\}_{j\in\mathbb{N}_N^+} \overset{i.i.d.}{\sim} \mathcal{N}\left(0,1\right). \tag{2.81}$$

From (2.80), one can readily observe that the statistical dependence of $\overline{Q}_t$ on X_t concentrates only on the eigenvalues of the covariance matrix $\mathbf{C}_t\left(X_t\right)$, for which we have already assumed the existence of a finite supremum explicitly (see Assumption 2.1). Consequently, *conditioning on the process* X_t, we can bound (2.80) as

$$\overline{Q}_t \leq \lambda_{sup} \sum_{j\in\mathbb{N}_N^+} U_j^2 \triangleq \lambda_{sup} U, \quad \text{with} \quad U \sim \chi^2\left(N\right), \tag{2.82}$$

almost everywhere and everywhere in time, where the RHS is independent of X_t. Next, from ([34], p. 1325), we know that for any chi squared random variable U with N

degrees of freedom,

$$\mathcal{P}\left(U \geq N + 2\sqrt{Nu} + 2u\right) \leq \exp\left(-u\right), \quad \forall u > 0. \tag{2.83}$$

Setting $u \equiv CN\left(1 + \log\left(t+1\right)\right)$ for any $C \geq 1$ and any $t \in \mathbb{N}$,

$$\mathcal{P}\left(U \geq N + 2N\sqrt{C\left(1 + \log\left(t+1\right)\right)} + 2CN\left(1 + \log\left(t+1\right)\right)\right) \leq \frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN}}. \tag{2.84}$$

a statement which equivalently means that, with probability at least $1 - \left(t+1\right)^{-CN}\exp\left(-CN\right)$,

$$U < N + 2N\sqrt{C\left(1 + \log\left(t+1\right)\right)} + 2CN\left(1 + \log\left(t+1\right)\right). \tag{2.85}$$

However, because the RHS of the above inequality is upper bounded by $5CN\left(1 + \log\left(t+1\right)\right)$,

$$\begin{aligned}\mathcal{P}\left(U < 5CN\left(1 + \log\left(t+1\right)\right)\right) & \\ \geq \mathcal{P}\left(U < N + 2N\sqrt{C\left(1 + \log\left(t+1\right)\right)} + 2CN\left(1 + \log\left(t+1\right)\right)\right) & \\ \geq 1 - \frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN}}. & \end{aligned} \tag{2.86}$$

Hence, $\forall i \in \mathbb{N}_t$,

$$\begin{aligned}\mathcal{P}\left(\overline{Q}_i \geq 5\lambda_{sup}CN\left(1 + \log\left(t+1\right)\right)\middle|\, X_i\right) &\leq \mathcal{P}\left(U \geq 5CN\left(1 + \log\left(t+1\right)\right)\right) \\ &\leq \frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN}},\end{aligned} \tag{2.87}$$

and, thus,

$$\begin{aligned}\mathcal{P}\left(\overline{Q}_i \geq 5\lambda_{sup}CN\left(1 + \log\left(t+1\right)\right)\right) &= \int \mathcal{P}\left(\overline{Q}_i \geq 5\lambda_{sup}CN\left(1 + \log\left(t+1\right)\right)\middle|\, X_i\right) \mathrm{d}\mathcal{P}_{X_i} \\ &\leq \frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN}} \int \mathrm{d}\mathcal{P}_{X_i} \equiv \frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN}}.\end{aligned} \tag{2.88}$$

However, we would like to produce a bound on the supremum of all the $Q_i, i \in \mathbb{N}_t$. Indeed, using the naive union bound,

$$\begin{aligned}\mathcal{P}\left(\bigcup_{i\in\mathbb{N}_t}\left\{\overline{Q}_i \geq 5\lambda_{sup}CN\left(1 + \log\left(t+1\right)\right)\right\}\right) &\leq \sum_{i\in\mathbb{N}_t}\mathcal{P}\left(\overline{Q}_i \geq \lambda_{sup}5CN\left(1 + \log\left(t+1\right)\right)\right) \\ &\leq \frac{\left(t+1\right)\exp\left(-CN\right)}{\left(t+1\right)^{CN}} \equiv \frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN-1}}\end{aligned}$$

or, equivalently,

$$\mathcal{P}\left(\sup_{i\in\mathbb{N}_t}\overline{Q}_i < 5\lambda_{sup}CN\left(1+\log\left(t+1\right)\right)\right) \equiv \mathcal{P}\left(\left\{\overline{Q}_i < 5\lambda_{sup}CN\left(1+\log\left(t+1\right)\right), \forall i\in\mathbb{N}_t\right\}\right)$$
$$\equiv \mathcal{P}\left(\bigcap_{i\in\mathbb{N}_t}\left\{\overline{Q}_i < 5\lambda_{sup}CN\left(1+\log\left(t+1\right)\right)\right\}\right)$$
$$\geq 1-\frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN-1}}, \tag{2.89}$$

holding true $\forall t\in\mathbb{N}$. Consequently, working in the same fashion as above, it is true that, with at least the same probability of success,

$$\sup_{i\in\mathbb{N}_t}Q_i\left(\omega\right) < 5\lambda_{sup}CN\left(1+\log\left(t+1\right)\right)+2\sqrt{5\lambda_{sup}CN\left(1+\log\left(t+1\right)\right)}\mu_{sup}+\mu_{sup}^2$$
$$< 5\lambda_{sup}\left(1+2\mu_{sup}+\mu_{sup}^2\right)CN\left(1+\log\left(t+1\right)\right) \tag{2.90}$$

or, setting $\gamma_1 \triangleq 5\lambda_{sup}\left(1+\mu_{sup}\right)^2 > 1$,

$$\sup_{i\in\mathbb{N}_t}Q_i\left(\omega\right) < \gamma_1 CN\left(1+\log\left(t+1\right)\right). \tag{2.91}$$

Now, under the alternative base measure $\widetilde{\mathcal{P}}$, $\mathbf{y}_t$ constitutes a Gaussian vector white noise process with zero mean and covariance matrix the identity, statistically independent of the process X_t (see Theorem 2). That is, for each t, the elements of $\mathbf{y}_t$ are themselves independent to each other. Thus, for all $t\in\mathbb{N}$ and for all $i\in\mathbb{N}_t$ and using similar arguments as the ones made above, it should be true that

$$\widetilde{\mathcal{P}}\left(Q_i < 5CN\left(1+\log\left(t+1\right)\right)\right) \geq 1-\frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN}} \tag{2.92}$$

and taking the union bound, we end up with the inequality

$$\widetilde{\mathcal{P}}\left(\sup_{i\in\mathbb{N}_t}Q_i < 5CN\left(1+\log\left(t+1\right)\right)\right) \geq 1-\frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN-1}}. \tag{2.93}$$

Defining $\gamma \triangleq \max\left\{\gamma_1, 5\right\} \equiv \gamma_1$, it must be true that, for all $t\in\mathbb{N}$,

$$\min\left\{\mathcal{P}\left(\sup_{i\in\mathbb{N}_t}Q_i < \gamma CN\left(1+\log\left(t+1\right)\right)\right), \widetilde{\mathcal{P}}\left(\sup_{i\in\mathbb{N}_t}Q_i < \gamma CN\left(1+\log\left(t+1\right)\right)\right)\right\}$$
$$\geq 1-\frac{\exp\left(-CN\right)}{\left(t+1\right)^{CN-1}}, \tag{2.94}$$

therefore completing the proof of the lemma. ■

Continuing our presentation of preliminary results towards the proof of Theorem (2.3) and leveraging the power of $\mathcal{C}$-weak convergence and Lemma 2.1, let us present the following lemma, connecting $\mathcal{C}$-weak convergence of random variables with convergence in the $\mathcal{L}_1$ sense.

Lemma 2.8. (From $\mathcal{C}$-Weak Convergence to Convergence in $\mathcal{L}_1$) *Consider the sequence of discrete time stochastic processes* $\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}$, *as well as a "limit" process* X_t,$t \in \mathbb{N}$, *all being* $\left(\mathbb{R}, \mathscr{S} \triangleq \mathscr{B}(\mathbb{R})\right)$*-valued and all defined on a common base space* $(\Omega, \mathscr{F}, \mathcal{P})$. *Further, suppose that all members of the collection* $\left\{\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}, X_t\right\}_{t\in\mathbb{N}}$ *are almost surely bounded in* $\mathcal{Z} \equiv [a, b]$ *(with* $-\infty < a < b < \infty$*) and that*

$$\mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}(\cdot|\, X_t) \xrightarrow[\mathtt{A}\to\infty]{\mathcal{W}} \delta_{X_t}(\cdot) \equiv \mathbb{1}_{(\cdot)}(X_t), \quad \forall t \in \mathbb{N}, \tag{2.95}$$

that is, the sequence $\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}$ *is marginally* $\mathcal{C}$*-weakly convergent to* X_t, *given* X_t, *for all* t. *Then, it is true that*

$$\mathbb{E}\left\{\left|X_t - X_t^{\mathtt{A}}\right|\right\} \xrightarrow[\mathtt{A}\to\infty]{} 0, \quad \forall t \in \mathbb{N}, \tag{2.96}$$

or, equivalently, $X_t^{\mathtt{A}} \xrightarrow[\mathtt{A}\to\infty]{\mathcal{L}_1} X_t$, *for all* t.

Proof of Lemma 2.8. Let all the hypotheses of Lemma 2.8 hold true. Then, we know that, $\forall t \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}(\mathcal{A}|\, X_t(\omega)) = \delta_{X_t(\omega)}(\mathcal{A}), \quad \mathcal{P}-a.e., \tag{2.97}$$

for all continuity Borel sets $\mathcal{A} \in \mathscr{S}$. Using the tower property, it is also true that

$$\mathbb{E}\left\{\left|X_t - X_t^{\mathtt{A}}\right|\right\} \equiv \mathbb{E}\left\{\mathbb{E}\left\{\left|X_t - X_t^{\mathtt{A}}\right| \middle| \sigma\{X_t\}\right\}\right\}. \tag{2.98}$$

Therefore, in order to show that $X_t^{\mathtt{A}} \xrightarrow[\mathtt{A}\to\infty]{\mathcal{L}_1} X_t$ for each $t \in \mathbb{N}$, it suffices to show that

$$\mathbb{E}\left\{\left|X_t - X_t^{\mathtt{A}}\right| \middle| \sigma\{X_t\}\right\}(\omega) \xrightarrow[\mathtt{A}\to\infty]{a.s.} 0, \quad \forall t \in \mathbb{N}. \tag{2.99}$$

Then, the Dominated Convergence Theorem would produce the desired result.

Of course, because all members of the collection $\left\{\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}, X_t\right\}_{t\in\mathbb{N}}$ are almost surely bounded in $\mathcal{Z}$, all members of the collection $\left\{\left\{\left|X_t - X_t^{\mathtt{A}}\right|\right\}_{\mathtt{A}\in\mathbb{N}}\right\}_{t\in\mathbb{N}}$ must be bounded almost surely in the compact set $\widehat{\mathcal{Z}} \triangleq [0, 2\delta] \subset \mathbb{R}$, where $\delta \triangleq \max\{|a|, |b|\}$.

Let us define the continuous and bounded function

$$f(x) \triangleq \begin{cases} x, & \text{if } x \in \widehat{\mathcal{Z}} \\ 2\delta, & \text{if } x > 2\delta \\ 0, & \text{if } x < 0 \end{cases} \cdot \tag{2.100}$$

Then, from Lemma 2.1 and using conditional probability measures it must be true that for each $t \in \mathbb{N}$, a version of the conditional expectation of interest is explicitly given by

$$\begin{aligned} \mathbb{E}\left\{ f\left(\left|X_t - X_t^{\mathtt{A}}\right|\right) \middle| \sigma\{X_t\} \right\}(\omega) \equiv \int f\left(|X_t(\omega) - x|\right) \mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}\left(\mathrm{d}x \,|X_t(\omega)\right) \underset{\mathtt{A}\to\infty}{\longrightarrow} \\ \underset{\mathtt{A}\to\infty}{\longrightarrow} \int f\left(|X_t(\omega) - x|\right) \delta_{X_t(\omega)}(\mathrm{d}x) \equiv 0, \ \mathcal{P} - a.e., \end{aligned} \tag{2.101}$$

since, for each $\omega \in \Omega$, $X_t(\omega)$ is constant. Further, by definition of f,

$$\begin{aligned} \mathbb{E}\left\{ f\left(\left|X_t - X_t^{\mathtt{A}}\right|\right) \middle| \sigma\{X_t\} \right\}(\omega) &\equiv \mathbb{E}\left\{ \left|X_t - X_t^{\mathtt{A}}\right| \mathbb{1}_{\left(|X_t - X_t^{\mathtt{A}}|\right)}\left(\widehat{\mathcal{Z}}\right) \middle| \sigma\{X_t\} \right\}(\omega) \\ &\equiv \mathbb{E}\left\{ \left|X_t - X_t^{\mathtt{A}}\right| \middle| \sigma\{X_t\} \right\}(\omega), \quad \mathcal{P} - a.e., \end{aligned} \tag{2.102}$$

and for all $t \in \mathbb{N}$, which means that

$$\mathbb{E}\left\{ \left|X_t - X_t^{\mathtt{A}}\right| \middle| \sigma\{X_t\} \right\}(\omega) \underset{\mathtt{A}\to\infty}{\overset{a.s.}{\longrightarrow}} 0, \quad \forall t \in \mathbb{N}. \tag{2.103}$$

Calling dominated convergence proves the result. ■

Additionally, the following useful (to us) result is also true. The proof, being elementary, is omitted.

Lemma 2.9. (Convergence of the Supremum) *Pick any natural $T < \infty$. If, under any circumstances,*

$$\mathbb{E}\left\{ \left|X_t - X_t^{\mathtt{A}}\right| \right\} \underset{\mathtt{A}\to\infty}{\longrightarrow} 0, \quad \forall t \in \mathbb{N}_T, \tag{2.104}$$

then

$$\sup_{t \in \mathbb{N}_T} \mathbb{E}\left\{ \left|X_t - X_t^{\mathtt{A}}\right| \right\} \underset{\mathtt{A}\to\infty}{\longrightarrow} 0. \tag{2.105}$$

2.1.4.3 The Key Lemma

We are now ready to present our key lemma, which will play an important role in establishing our main result (Theorem 2.3) later on. For proving this result, we make use of all the intermediate ones presented in the previous subsections.

Lemma 2.10. (Convergence of the Likelihoods) *Consider the stochastic process*

$$\widehat{\Lambda}_t \triangleq \frac{\exp\left(-\frac{1}{2}\sum_{i\in\mathbb{N}_t}\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i\right)\mathbf{C}_i^{-1}\left(X_i\right)\overline{\mathbf{y}}_i\left(X_i\right)\right)}{\prod_{i\in\mathbb{N}_t}\sqrt{\det\left(\mathbf{C}_i\left(X_i\right)\right)}} \triangleq \frac{\mathfrak{N}_t}{\mathfrak{D}_t}, \quad t\in\mathbb{N}. \tag{2.106}$$

Consider also the process $\widehat{\Lambda}_t^{\mathtt{A}} \triangleq \mathfrak{N}_t^{\mathtt{A}}/\mathfrak{D}_t^{\mathtt{A}}$*, defined exactly the same way as* $\widehat{\Lambda}_t$*, but replacing* X_i *with the approximation* $X_i^{\mathtt{A}}$*,* $\forall i\in\mathbb{N}_t$*. Further, pick any natural* $T<\infty$ *and suppose either of the following:*

- *For all* $t\in\mathbb{N}_T$*, the sequence* $\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}$ *is marginally* $\mathcal{C}$*-weakly convergent to* X_t*, given* X_t*, that is,*

$$\mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}\left(\cdot|\,X_t\right)\xrightarrow[\mathtt{A}\to\infty]{\mathcal{W}}\delta_{X_t}\left(\cdot\right), \quad \forall t\in\mathbb{N}_T. \tag{2.107}$$

- *For all* $t\in\mathbb{N}_T$*, the sequence* $\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}$ *is marginally convergent to* X_t *in probability, that is,*

$$X_t^{\mathtt{A}}\xrightarrow[\mathtt{A}\to\infty]{\mathcal{P}}X_t, \quad \forall t\in\mathbb{N}_T. \tag{2.108}$$

Then, there exists a measurable subset $\widehat{\Omega}_T\subseteq\Omega$*, such that*

$$\lim_{\mathtt{A}\to\infty}\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\,\mathscr{Y}_t\right\}\left(\omega\right)\equiv 0, \tag{2.109}$$

where the $\mathcal{P}$*,*$\widetilde{\mathcal{P}}$*-measures of* $\widehat{\Omega}_T$ *satisfy*

$$\min\left\{\mathcal{P}\left(\widehat{\Omega}_T\right),\widetilde{\mathcal{P}}\left(\widehat{\Omega}_T\right)\right\}\geq 1-\frac{\exp\left(-CN\right)}{\left(T+1\right)^{CN-1}}, \tag{2.110}$$

for any free but finite constant $C\geq 1$*.*

Proof of Lemma 2.10. From Lemma 2.2, it is true that

$$\begin{aligned}\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right| \equiv \left|\frac{\mathfrak{N}_t}{\mathfrak{D}_t}-\frac{\mathfrak{N}_t^{\mathtt{A}}}{\mathfrak{D}_t^{\mathtt{A}}}\right| &\leq \frac{\left|\mathfrak{N}_t-\mathfrak{N}_t^{\mathtt{A}}\right|}{\left|\mathfrak{D}_t^{\mathtt{A}}\right|}+\left|\mathfrak{N}_t\right|\left|\frac{1}{\mathfrak{D}_t}-\frac{1}{\mathfrak{D}_t^{\mathtt{A}}}\right| \\ &\leq \frac{\left|\mathfrak{N}_t-\mathfrak{N}_t^{\mathtt{A}}\right|}{\left|\mathfrak{D}_t^{\mathtt{A}}\right|}+\left|\frac{1}{\mathfrak{D}_t}-\frac{1}{\mathfrak{D}_t^{\mathtt{A}}}\right|.\end{aligned} \tag{2.111}$$

We first concentrate on the determinant part (second term) of the RHS of (2.111). Directly invoking Lemma 2.3, it will be true that

$$\begin{aligned}
&\left|\frac{1}{\mathfrak{D}_t}-\frac{1}{\mathfrak{D}_t^{\mathtt{A}}}\right| \\
&\equiv\left|\prod_{i=0}^{t}\frac{1}{\sqrt{\det\left(\mathbf{C}_i\left(X_i\right)\right)}}-\prod_{i=0}^{t}\frac{1}{\sqrt{\det\left(\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right)}}\right| \\
&\leq\sum_{i=0}^{t}\left(\prod_{j=0}^{i-1}\frac{1}{\sqrt{\det\left(\mathbf{C}_j\left(X_j\right)\right)}}\right)\left(\prod_{j=i+1}^{t}\frac{1}{\sqrt{\det\left(\mathbf{C}_j\left(X_j^{\mathtt{A}}\right)\right)}}\right) \\
&\qquad\qquad\times\frac{\left|\sqrt{\det\left(\mathbf{C}_i\left(X_i\right)\right)}-\sqrt{\det\left(\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right)}\right|}{\sqrt{\det\left(\mathbf{C}_i\left(X_i\right)\right)\det\left(\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right)}} \\
&=\sum_{i=0}^{t}\left(\prod_{j=0}^{i-1}\frac{1}{\sqrt{\prod_{n=1}^{N}\lambda_n\left(\mathbf{C}_j\left(X_j\right)\right)}}\right)\left(\prod_{j=i+1}^{t}\frac{1}{\sqrt{\prod_{n=1}^{N}\lambda_n\left(\mathbf{C}_j\left(X_j^{\mathtt{A}}\right)\right)}}\right) \\
&\qquad\qquad\times\frac{\left|\sqrt{\det\left(\mathbf{C}_i\left(X_i\right)\right)}-\sqrt{\det\left(\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right)}\right|}{\sqrt{\prod_{n=1}^{N}\lambda_n\left(\mathbf{C}_j\left(X_j\right)\right)\lambda_n\left(\mathbf{C}_j\left(X_j^{\mathtt{A}}\right)\right)}} \\
&\leq\sum_{i=0}^{t}\frac{1}{2\lambda_{inf}^{Ni/2}\lambda_{inf}^{N(t-i)/2}}\frac{\left|\det\left(\mathbf{C}_i\left(X_i\right)\right)-\det\left(\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right)\right|}{\lambda_{inf}^{N}} \\
&\equiv\frac{1}{2\sqrt{\lambda_{inf}^{N(t+2)}}}\sum_{i=0}^{t}\left|\det\left(\mathbf{C}_i\left(X_i\right)\right)-\det\left(\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right)\right|. \qquad (2.112)
\end{aligned}$$

From Lemma 2.5, we can bound the RHS of the above expression as

$$\begin{aligned}
\left|\frac{1}{\mathfrak{D}_t}-\frac{1}{\mathfrak{D}_t^{\mathtt{A}}}\right| &\leq\frac{NK_{DET}}{2\sqrt{\lambda_{inf}^{N(t+2)}}}\sum_{i=0}^{t}\left\|\mathbf{C}_i\left(X_i\right)-\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right\|_F \\
&\equiv\frac{NK_{DET}}{2\sqrt{\lambda_{inf}^{N(t+2)}}}\sum_{i=0}^{t}\left\|\mathbf{\Sigma}_i\left(X_i\right)-\mathbf{\Sigma}_i\left(X_i^{\mathtt{A}}\right)\right\|_F. \qquad (2.113)
\end{aligned}$$

And from Lemma 2.4, (2.113) becomes

$$\left|\frac{1}{\mathfrak{D}_t}-\frac{1}{\mathfrak{D}_t^{\mathtt{A}}}\right|\leq\frac{N^2K_{DET}K_{\boldsymbol{\Sigma}}}{2\sqrt{\lambda_{inf}^{N(t+2)}}}\sum_{i=0}^{t}\left|X_i-X_i^{\mathtt{A}}\right|. \tag{2.114}$$

We now turn our attention to the "difference of exponentials" part (first term) of the RHS of (2.111). First, we know that

$$\prod_{i=0}^{t}\det\left(\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right)\geq\prod_{i=0}^{t}\prod_{j=1}^{N}\lambda_{inf}\equiv\lambda_{inf}^{N(t+1)}, \tag{2.115}$$

yielding the inequality

$$\frac{\left|\mathfrak{N}_t-\mathfrak{N}_t^{\mathtt{A}}\right|}{\left|\mathfrak{D}_t^{\mathtt{A}}\right|}\leq\frac{\left|\mathfrak{N}_t-\mathfrak{N}_t^{\mathtt{A}}\right|}{\sqrt{\lambda_{inf}^{N(t+1)}}}, \tag{2.116}$$

where $\lambda_{inf}>1$ (see Assumption 2.1). Next, making use of the inequality [20]

$$\left|\exp\left(\alpha\right)-\exp\left(\beta\right)\right|\leq\left|\alpha-\beta\right|\left(\exp\left(\alpha\right)+\exp\left(\beta\right)\right), \tag{2.117}$$

$\forall\left(\alpha,\beta\right)\in\mathbb{R}^2$, the absolute difference on the numerator of (2.116) can be upper bounded as

$$\begin{aligned}\left|\mathfrak{N}_t-\mathfrak{N}_t^{\mathtt{A}}\right|&\leq\frac{1}{2}\left|\sum_{i=0}^{t}\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i\right)\mathbf{C}_i^{-1}\left(X_i\right)\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i^{\mathtt{A}}\right)\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right|\left(\mathfrak{N}_t+\mathfrak{N}_t^{\mathtt{A}}\right)\\&\leq\sum_{i=0}^{t}\left|\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i\right)\mathbf{C}_i^{-1}\left(X_i\right)\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i^{\mathtt{A}}\right)\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right|.\end{aligned} \tag{2.118}$$

Concentrating on each member of the series above in the last line of (2.118) and calling Lemma 2.2, it is true that

$$\begin{aligned}&\left|\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i\right)\left[\mathbf{C}_i^{-1}\left(X_i\right)\overline{\mathbf{y}}_i\left(X_i\right)\right]-\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i^{\mathtt{A}}\right)\left[\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right]\right|\\&\qquad\leq\left\|\overline{\mathbf{y}}_i\left(X_i\right)\right\|_2\left\|\mathbf{C}_i^{-1}\left(X_i\right)\overline{\mathbf{y}}_i\left(X_i\right)-\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2+\\&\qquad\qquad+\left\|\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\left\|\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2.\end{aligned} \tag{2.119}$$

Calling Lemma 2.2 again for the term multiplying the quantity $\left\|\overline{\mathbf{y}}_i\left(X_i\right)\right\|_2$ in the RHS of the above expression, we arrive at the inequalities

$$\begin{aligned}&\left|\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i\right)\mathbf{C}_i^{-1}\left(X_i\right)\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i^{\mathtt{A}}\right)\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right|\\&\qquad\leq\left\|\overline{\mathbf{y}}_i\left(X_i\right)\right\|_2\left\|\mathbf{C}_i^{-1}\left(X_i\right)\right\|_2\left\|\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\end{aligned}$$

$$
\begin{aligned}
&+\left\|\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\left\|\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\right\|_2\left\|\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\\
&\qquad+\left\|\overline{\mathbf{y}}_i\left(X_i\right)\right\|_2\left\|\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\left\|\mathbf{C}_i^{-1}\left(X_i\right)-\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\right\|_2\\
\leq\ &\frac{\left\|\overline{\mathbf{y}}_i\left(X_i\right)\right\|_2}{\lambda_{inf}}\left\|\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2+\frac{\left\|\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2}{\lambda_{inf}}\left\|\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\\
&\qquad+\left\|\overline{\mathbf{y}}_i\left(X_i\right)\right\|_2\left\|\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\left\|\mathbf{C}_i^{-1}\left(X_i\right)-\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\right\|_2, \qquad (2.120)
\end{aligned}
$$

or, equivalently,

$$
\begin{aligned}
&\left|\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i\right)\mathbf{C}_i^{-1}\left(X_i\right)\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i^{\mathtt{A}}\right)\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right|\\
&\qquad\leq\frac{\left\|\overline{\mathbf{y}}_i\left(X_i\right)\right\|_2+\left\|\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2}{\lambda_{inf}}\left\|\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2+\\
&\qquad\qquad+\left\|\overline{\mathbf{y}}_i\left(X_i\right)\right\|_2\left\|\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\left\|\mathbf{C}_i^{-1}\left(X_i\right)-\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\right\|_2. \qquad (2.121)
\end{aligned}
$$

Now, recalling Assumption 2.2, the definition of $\overline{\mathbf{y}}_i\left(X_i\right)$ (resp. for $X_i^{\mathtt{A}}$) and invoking Lemma 2.6, it must be true that

$$
\begin{aligned}
&\left|\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i\right)\mathbf{C}_i^{-1}\left(X_i\right)\overline{\mathbf{y}}_i\left(X_i\right)-\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i^{\mathtt{A}}\right)\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right|\\
&\quad\leq\frac{\left\|\mathbf{y}_i-\boldsymbol{\mu}_i\left(X_i\right)\right\|_2+\left\|\mathbf{y}_i-\boldsymbol{\mu}_i\left(X_i^{\mathtt{A}}\right)\right\|_2}{\lambda_{inf}}\left\|\boldsymbol{\mu}_i\left(X_i\right)-\boldsymbol{\mu}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\\
&\qquad\qquad+\left\|\mathbf{y}_i-\boldsymbol{\mu}_i\left(X_i\right)\right\|_2\left\|\mathbf{y}_i-\boldsymbol{\mu}_i\left(X_i^{\mathtt{A}}\right)\right\|_2\left\|\mathbf{C}_i^{-1}\left(X_i\right)-\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\right\|_2\\
&\quad\leq\left(K_\mu\frac{2\left(\left\|\mathbf{y}_i\right\|_2+\mu_{sup}\right)}{\lambda_{inf}}+K_{INV}\left(\left\|\mathbf{y}_i\right\|_2+\mu_{sup}\right)^2\right)\left|X_i-X_i^{\mathtt{A}}\right|\\
&\quad\triangleq\Theta\left(\mathbf{y}_i\right)\left|X_i-X_i^{\mathtt{A}}\right|. \qquad (2.122)
\end{aligned}
$$

Using the above inequality, the RHS of (2.116) can be further bounded from above as

$$
\frac{\left|\mathfrak{N}_t-\mathfrak{N}_t^{\mathtt{A}}\right|}{\left|\mathfrak{D}_t^{\mathtt{A}}\right|}\leq\frac{\sup_{i\in\mathbb{N}_t}\Theta\left(\mathbf{y}_i\right)}{\sqrt{\lambda_{inf}^{N(t+1)}}}\sum_{i=0}^{t}\left|X_i-X_i^{\mathtt{A}}\right|. \qquad (2.123)
$$

Therefore, we can bound the RHS of (2.111) as

$$
\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\leq\left(\frac{\sup_{i\in\mathbb{N}_t}\Theta\left(\mathbf{y}_i\right)}{\sqrt{\lambda_{inf}^{N(t+1)}}}+\frac{N^2K_{DET}K_{\boldsymbol{\Sigma}}}{2\sqrt{\lambda_{inf}^{N(t+2)}}}\right)\sum_{i=0}^{t}\left|X_i-X_i^{\mathtt{A}}\right|. \qquad (2.124)
$$

Taking conditional expectations on both sides of (2.124), observing that the quantity $\sup_{i\in\mathbb{N}_t}\Theta\left(\mathbf{y}_i\right)$ constitutes a $\{\mathscr{Y}_t\}$-adapted process and recalling that under the base

measure $\widetilde{\mathcal{P}}$ (see Theorem 2), the processes $\mathbf{y}_t$ and X_t (resp. $X_t^{\mathtt{A}}$) are statistically independent, we can write

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}\leq\left(\frac{\sup_{i\in\mathbb{N}_t}\Theta\left(\mathbf{y}_i\right)}{\sqrt{\lambda_{inf}^{N(t+1)}}}+\frac{N^2K_{DET}K_{\mathbf{\Sigma}}}{2\sqrt{\lambda_{inf}^{N(t+2)}}}\right)\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\sum_{i=0}^{t}\left|X_i-X_i^{\mathtt{A}}\right|\right\},\quad(2.125)$$

$\widetilde{\mathcal{P}}-a.e.$, and, because $\mathcal{P}\ll_{\mathscr{H}_t}\widetilde{\mathcal{P}}$, $\mathcal{P}-a.e.$ as well. From the last inequality, we can readily observe that in order to be able to talk about any kind of uniform convergence regarding the RHS, it is vital to ensure that the random variable $\sup_{i\in\mathbb{N}_t}\Theta\left(\mathbf{y}_i\right)$ is bounded from above. However, because the support of $\|\mathbf{y}_i\|_2$ is infinite, it is impossible to bound $\sup_{i\in\mathbb{N}_t}\Theta\left(\mathbf{y}_i\right)$ in the almost sure sense. Nevertheless, Lemma 2.7 immediately implies that there exists a measurable subset $\widehat{\Omega}_\tau\subseteq\Omega$ with

$$\min\left\{\mathcal{P}\left(\widehat{\Omega}_\tau\right),\widetilde{\mathcal{P}}\left(\widehat{\Omega}_\tau\right)\right\}\geq1-\frac{\exp\left(-CN\right)}{\left(\tau+1\right)^{CN-1}}\quad(2.126)$$

such that, for all $\omega\in\widehat{\Omega}_\tau$,

$$\sup_{i\in\mathbb{N}_\tau}\left\|\mathbf{y}_i\left(\omega\right)\right\|_2^2\equiv\sup_{i\in\mathbb{N}_\tau}\left\|\mathbf{y}_i\right\|_2^2<\gamma CN\left(1+\log\left(1+\tau\right)\right),\quad(2.127)$$

for some fixed constant $\gamma>1$, for any $C\geq1$ and for any *fixed* $\tau\in\mathbb{N}$. Choosing $\tau\equiv T<\infty$, it is true that

$$\begin{aligned}\sup_{i\in\mathbb{N}_t}\Theta\left(\mathbf{y}_i\right)&\leq\sup_{i\in\mathbb{N}_T}\Theta\left(\mathbf{y}_i\right)\\&\leq\sup_{i\in\mathbb{N}_T}\left[K_\mu\frac{2\left(\left\|\mathbf{y}_i\right\|_2+\mu_{sup}\right)}{\lambda_{inf}}+K_{INV}\left(\left\|\mathbf{y}_i\right\|_2+\mu_{sup}\right)^2\right]\\&<\left(K_\mu\frac{2\widetilde{\gamma}}{\lambda_{inf}}+K_{INV}\widetilde{\gamma}^2\right)CN\left(1+\log\left(1+T\right)\right)\\&\triangleq K_oCN\left(1+\log\left(1+T\right)\right),\quad\forall t\in\mathbb{N}_T,\end{aligned}\quad(2.128)$$

where $\widetilde{\gamma}\triangleq\sqrt{\gamma}+\mu_{sup}$. Therefore, it will be true that

$$\begin{aligned}&\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}\\&\quad\leq\left(\frac{K_oCN\left(1+\log\left(1+T\right)\right)}{\sqrt{\lambda_{inf}^{N(t+1)}}}+\frac{N^2K_{DET}K_{\mathbf{\Sigma}}}{2\sqrt{\lambda_{inf}^{N(t+2)}}}\right)\sum_{i=0}^{t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_i-X_i^{\mathtt{A}}\right|\right\},\end{aligned}\quad(2.129)$$

for all $t\in\mathbb{N}_T$, with probability at least

$$1-\frac{\exp\left(-CN\right)}{\left(T+1\right)^{CN-1}},$$

under either $\mathcal{P}$ or $\widetilde{\mathcal{P}}$. Further,

$$\sum_{i=0}^{t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_i-X_i^{\mathtt{A}}\right|\right\}\leq(t+1)\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_\tau-X_\tau^{\mathtt{A}}\right|\right\}. \tag{2.130}$$

Then, with the same probability of success,

$$\begin{aligned}
&\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}\\
&\leq\left(\frac{K_oCN\left(1+\log\left(1+T\right)\right)(t+1)}{\sqrt{\lambda_{inf}^{N(t+1)}}}+\frac{K_{DET}K_{\mathbf{\Sigma}}N^2\left(t+1\right)}{2\sqrt{\lambda_{inf}^{N(t+2)}}}\right)\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_\tau-X_\tau^{\mathtt{A}}\right|\right\}\\
&\leq\left(\frac{K_oCN\left(1+\log\left(1+T\right)\right)(T+1)}{\lambda_{inf}^{N/2}}+\frac{K_{DET}K_{\mathbf{\Sigma}}N^2\left(T+1\right)}{2\lambda_{inf}^{N}}\right)\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_\tau-X_\tau^{\mathtt{A}}\right|\right\}\\
&\triangleq K_G\left(T\right)\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_\tau-X_\tau^{\mathtt{A}}\right|\right\},\quad\forall t\in\mathbb{N}_T,
\end{aligned}\tag{2.131}$$

where $K_G\left(T\right)\equiv\mathcal{O}\left(T\log\left(T\right)\right)$. *Alternatively*, upper bounding the functions comprised by the quantities t,N,λ_{inf} in the second and third lines of the expressions above as (note that, obviously, $t+1\geq1,\forall t\in\mathbb{R}_+$)

$$\frac{N\left(t+1\right)}{\sqrt{\lambda_{inf}^{N(t+1)}}}\leq\max_{t\in\mathbb{R}_+}\frac{N\left(t+1\right)^2}{\sqrt{\lambda_{inf}^{N(t+1)}}}\quad\text{and} \tag{2.132}$$

$$\frac{N^2\left(t+1\right)}{\sqrt{\lambda_{inf}^{N(t+2)}}}\leq\max_{t\in\mathbb{R}_+}\frac{N^2\left(t+1\right)^2}{\sqrt{\lambda_{inf}^{N(t+2)}}}, \tag{2.133}$$

respectively, we can also define

$$K_G\left(T\right)\triangleq\frac{16K_oC\left(1+\log\left(1+T\right)\right)\lambda_{inf}^{-2/\log\left(\lambda_{inf}\right)}}{N\left(\log\left(\lambda_{inf}\right)\right)^2}+\frac{8K_{DET}K_{\mathbf{\Sigma}}\lambda_{inf}^{-N}\lambda_{inf}^{-2/\log\left(\lambda_{inf}\right)}}{\left(\log\left(\lambda_{inf}\right)\right)^2}, \tag{2.134}$$

where, in this case, $K_G\left(T\right)\equiv\mathcal{O}\left(\log\left(T\right)\right)$. Note, however, that although its dependence on T is logarithmic, $K_G\left(T\right)$ may still be large due to the inability to compensate for the size of K_o. In any case, for all $\omega\in\widehat{\Omega}_T$,

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}(\omega)\leq K_G\left(T\right)\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_\tau-X_\tau^{\mathtt{A}}\right|\right\},\quad\forall t\in\mathbb{N}_T, \tag{2.135}$$

Therefore, we get

$$\sup_{\omega\in\widehat{\Omega}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}(\omega)\leq K_G\left(T\right)\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_\tau-X_\tau^{\mathtt{A}}\right|\right\},\quad\forall t\in\mathbb{N}_T \tag{2.136}$$

and further taking the supremum over $t \in \mathbb{N}_T$ on both sides, it must be true that

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}(\omega)\leq K_G(T)\sup_{t\in\mathbb{N}_T}\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_\tau-X_\tau^{\mathtt{A}}\right|\right\}$$
$$\equiv K_G(T)\sup_{t\in\mathbb{N}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_t-X_t^{\mathtt{A}}\right|\right\}. \tag{2.137}$$

Finally, if either

$$\mathcal{P}^{\mathtt{A}}_{X_t^{\mathtt{A}}|X_t}(\cdot|X_t)\underset{\mathtt{A}\to\infty}{\overset{\mathcal{W}}{\longrightarrow}}\delta_{X_t}(\cdot)\equiv\mathbb{1}_{(\cdot)}(X_t),\quad\forall t\in\mathbb{N}_T, \tag{2.138}$$

or

$$X_t^{\mathtt{A}}\underset{\mathtt{A}\to\infty}{\overset{\mathcal{P}}{\longrightarrow}}X_t,\quad\forall t\in\mathbb{N}_T \tag{2.139}$$

and given that since the members of $\left\{X_t^{\mathtt{A}}\right\}_{\mathtt{A}\in\mathbb{N}}$ are almost surely bounded in $\mathcal{Z}$, the aforementioned sequence is also uniformly integrable for all $t\in\mathbb{N}$, it must be true that (see Lemma 2.8)

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_t-X_t^{\mathtt{A}}\right|\right\}\underset{\mathtt{A}\to\infty}{\longrightarrow}0,\quad\forall t\in\mathbb{N}_T. \tag{2.140}$$

Then, Lemma 2.9 implies that

$$\sup_{t\in\mathbb{N}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_t-X_t^{\mathtt{A}}\right|\right\}\underset{\mathtt{A}\to\infty}{\longrightarrow}0, \tag{2.141}$$

which in turn implies the existence of the limit on the LHS of (2.137). QED. ■

2.1.4.4 Finishing the Proof of Theorem 2.3

Considering the absolute difference of the RHSs of (2.44) and (2.43), it is true that (see Lemma 2.2)

$$\left|\frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\{X_t\Lambda_t|\mathscr{Y}_t\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\{\Lambda_t|\mathscr{Y}_t\}}-\frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t^{\mathtt{A}}\Lambda_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}\right|\equiv\left|\frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t\widehat{\Lambda}_t\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t\middle|\mathscr{Y}_t\right\}}-\frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t^{\mathtt{A}}\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}\right|, \tag{2.142}$$

due to the fact that the increasing stochastic process

$$\prod_{i\in\mathbb{N}_t}\exp\left(\frac{1}{2}\|\mathbf{y}_i\|_2^2\right)\equiv\exp\left(\frac{1}{2}\sum_{i\in\mathbb{N}_t}\|\mathbf{y}_i\|_2^2\right) \tag{2.143}$$

is $\{\mathscr{Y}_t\}$-adapted. Then, we can write

$$\left|\frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t\widehat{\Lambda}_t\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t\middle|\mathscr{Y}_t\right\}}-\frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t^{\mathtt{A}}\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}\right|$$

$$\equiv \frac{\left|\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t\widehat{\Lambda}_t\middle|\mathscr{Y}_t\right\}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}-\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t^{\mathtt{A}}\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t\middle|\mathscr{Y}_t\right\}\right|}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t\middle|\mathscr{Y}_t\right\}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}$$
$$\leq \frac{\left|\mathbb{E}_{\mathcal{P}}\left\{X_t\middle|\mathscr{Y}_t\right\}\right|\left|\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}\right|}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}+\frac{\left|\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t\widehat{\Lambda}_t-X_t^{\mathtt{A}}\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}\right|}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}},\quad \widetilde{\mathcal{P}},\mathcal{P}-a.e..$$
(2.144)

Let us first focus on the difference on the numerator of the second ratio of the RHS of (2.144). Recalling that $\delta\equiv\max\{|a|,|b|\}$, we can then write

$$\begin{aligned}\left|\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t\widehat{\Lambda}_t-X_t^{\mathtt{A}}\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}\right| &\leq \mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_t\widehat{\Lambda}_t-X_t^{\mathtt{A}}\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}\\ &\leq \mathbb{E}_{\widetilde{\mathcal{P}}}\left\{|X_t|\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|+\left|\widehat{\Lambda}_t^{\mathtt{A}}\right|\left|X_t-X_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}\\ &\leq \delta\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}+\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_t-X_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}\\ &\equiv \delta\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}+\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_t-X_t^{\mathtt{A}}\right|\right\},\end{aligned}$$
(2.145)

On the other hand, for the denominator for (2.144), it is true that

$$\begin{aligned}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\} &\equiv \mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\frac{\exp\left(-\frac{1}{2}\sum_{i\in\mathbb{N}_t}\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i^{\mathtt{A}}\right)\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right)}{\prod_{i\in\mathbb{N}_t}\sqrt{\det\left(\mathbf{C}_i\left(X_i^{\mathtt{A}}\right)\right)}}\middle|\mathscr{Y}_t\right\}\\ &\geq \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\exp\left(-\frac{1}{2}\sum_{i\in\mathbb{N}_t}\overline{\mathbf{y}}_i^{\boldsymbol{T}}\left(X_i^{\mathtt{A}}\right)\mathbf{C}_i^{-1}\left(X_i^{\mathtt{A}}\right)\overline{\mathbf{y}}_i\left(X_i^{\mathtt{A}}\right)\right)\middle|\mathscr{Y}_t\right\}}{\sqrt{\lambda_{sup}^{N(t+1)}}}\\ &\geq \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\exp\left(-\frac{1}{2\lambda_{inf}}\sum_{i\in\mathbb{N}_t}\left(\|\mathbf{y}_i\|_2+\mu_{sup}\right)^2\right)\middle|\mathscr{Y}_t\right\}}{\sqrt{\lambda_{sup}^{N(t+1)}}}\\ &\equiv \frac{\exp\left(-\frac{1}{2\lambda_{inf}}\sum_{i\in\mathbb{N}_t}\left(\|\mathbf{y}_i\|_2+\mu_{sup}\right)^2\right)}{\sqrt{\lambda_{sup}^{N(t+1)}}},\quad \widetilde{\mathcal{P}},\mathcal{P}-a.e.,\end{aligned}$$
(2.146)

since the process $\sum_{i\in\mathbb{N}_t}\left(\|\mathbf{y}_i\|_2+\mu_{sup}\right)^2$ is $\{\mathscr{Y}_t\}$-adapted. Now, from Lemma 2.7, we know that

$$\sup_{i\in\mathbb{N}_t}\|\mathbf{y}_i\|_2^2\leq\sup_{i\in\mathbb{N}_T}\|\mathbf{y}_i\|_2^2<\gamma CN\left(1+\log\left(T+1\right)\right),\quad\forall t\in\mathbb{N}_T,$$
(2.147)

where the last inequality holds with probability at least $1-(T+1)^{1-CN}\exp(-CN)$, under both base measures $\mathcal{P}$ and $\widetilde{\mathcal{P}}$, for any finite constant $C \geq 1$. Therefore, it can be trivially shown that

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\} \geq \frac{\exp\left(-\frac{\left(\sqrt{\gamma CN\left(1+\log\left(T+1\right)\right)}+\mu_{sup}\right)^2(T+1)}{2\lambda_{inf}}\right)}{\sqrt{\lambda_{sup}^{N(t+1)}}} > 0, \quad (2.148)$$

for all $t \in \mathbb{N}_T$, implying that

$$\inf_{t\in\mathbb{N}_T}\inf_{\omega\in\widehat{\Omega}_T}\inf_{\mathtt{A}\in\mathbb{N}}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}(\omega) > 0, \quad (2.149)$$

where $\widehat{\Omega}_T$ coincides with the event

$$\left\{\omega\in\Omega\middle|\sup_{i\in\mathbb{N}_t}\|\mathbf{y}_i\|_2^2 < \gamma CN\left(1+\log\left(T+1\right)\right), \forall t\in\mathbb{N}_T\right\}$$

with $\mathcal{P}$,$\widetilde{\mathcal{P}}$-measure at least $1-(T+1)^{1-CN}\exp(-CN)$. Of course, the existence of $\widehat{\Omega}_T$ follows from Lemma 2.7. Putting it altogether, (2.144) becomes (recall that the base measures $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ are equivalent)

$$\begin{aligned}&\left|\mathbb{E}_{\mathcal{P}}\left\{X_t\middle|\mathscr{Y}_t\right\}-\mathcal{E}^{\mathtt{A}}\left(X_t\middle|\mathscr{Y}_t\right)\right| \\ &\qquad\leq \frac{2\delta\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}+\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_t-X_t^{\mathtt{A}}\right|\right\}}{\inf_{\mathtt{A}\in\mathbb{N}}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}}, \quad \mathcal{P},\widetilde{\mathcal{P}}-a.e., \quad (2.150)\end{aligned}$$

where $\delta \equiv \max\left\{|a|,|b|\right\}$. Taking the supremum both with respect to $\omega \in \widehat{\Omega}_T$ and $t \in \mathbb{N}_T$ on both sides, we get

$$\begin{aligned}&\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left|\mathbb{E}_{\mathcal{P}}\left\{X_t\middle|\mathscr{Y}_t\right\}(\omega)-\mathcal{E}^{\mathtt{A}}\left(X_t\middle|\mathscr{Y}_t\right)(\omega)\right| \\ &\qquad\leq \frac{\sup_{t\in\mathbb{N}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|X_t-X_t^{\mathtt{A}}\right|\right\}+2\delta\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left|\widehat{\Lambda}_t-\widehat{\Lambda}_t^{\mathtt{A}}\right|\middle|\mathscr{Y}_t\right\}(\omega)}{\inf_{t\in\mathbb{N}_T}\inf_{\omega\in\widehat{\Omega}_T}\inf_{\mathtt{A}\in\mathbb{N}}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_t^{\mathtt{A}}\middle|\mathscr{Y}_t\right\}(\omega)}, \quad (2.151)\end{aligned}$$

with

$$\min\left\{\mathcal{P}\left(\widehat{\Omega}_T\right),\widetilde{\mathcal{P}}\left(\widehat{\Omega}_T\right)\right\} \geq 1-\frac{\exp\left(-CN\right)}{\left(T+1\right)^{CN-1}}. \quad (2.152)$$

Finally, calling Lemma 2.10 and Lemma 2.9, and since the denominators of the fractions appearing in (2.151) are nonzero, its RHS tends to zero as $\mathtt{A} \to \infty$, under the respective hypotheses. Consequently, the LHS will also converge, therefore completing the proof of the Theorem 2.3. ■

2.1.5 Conclusion

In this work, we have provided sufficient conditions for convergence of approximate, asymptotically optimal nonlinear filtering operators, for a general class of hidden stochastic processes, observed in a conditionally Gaussian noisy environment. In particular, employing a common change of measure argument, we have shown that using the same measurements, but replacing the "true" state by an approximation process, which converges to the former either in probability or in the $\mathcal{C}$-weak sense, one can define an approximate filtering operator, which converges to the optimal filter compactly in time and uniformly in an event occurring with probability nearly 1, at the same time constituting a purely quantitative justification of Egoroff's theorem for the problem of interest. The results presented in this work essentially provide a framework for analyzing the convergence properties of various classes of approximate nonlinear filters (either recursive or nonrecursive), such as existing grid based approaches, which are known to perform well in various applications.

2.1.6 Appendix: Proof of Theorem 2.1

The proof is astonishingly simple. Let the hypotheses of the statement of Theorem 1 hold true. To avoid useless notational congestion, let us also make the identifications

$$\mathcal{X}_t \triangleq \{X_i\}_{i\in\mathbb{N}_t} \quad \text{and} \quad \mathcal{Y}_t \triangleq \{Y_i\}_{i\in\mathbb{N}_t}\,. \tag{2.153}$$

Now, by definition of the conditional expectation operator and since we have assumed the existence of densities, it is true that

$$\begin{aligned}\widehat{X}_t &\equiv \mathbb{E}_{\mathcal{P}}\left\{ X_t \middle| Y_0, Y_1, \ldots, Y_t\right\} \\ &= \int x_t f_{X_t|\mathcal{Y}_t}\left(x_t \middle| \mathcal{Y}_t\right) \mathrm{d}x_t\end{aligned}$$

$$
\begin{aligned}
&= \frac{\int x_t f_{(X_t,\mathcal{Y}_t)}\left(x_t,\mathcal{Y}_t\right)\mathrm{d}x_t}{f_{\mathcal{Y}_t}\left(\mathcal{Y}_t\right)} \\
&\equiv \frac{\int x_t f_t\left(x_0,x_1,\ldots x_t,\mathcal{Y}_t\right)\prod_{i=0}^{t}\mathrm{d}x_i}{\int f_t\left(x_0,x_1,\ldots x_t,\mathcal{Y}_t\right)\prod_{i=0}^{t}\mathrm{d}x_i} \\
&\equiv \frac{\int x_t \lambda_t \widetilde{f}_t\left(\{x_i\}_{i\in\mathbb{N}_t},\mathcal{Y}_t\right)\prod_{i=0}^{t}\mathrm{d}x_i}{\int \lambda_t \widetilde{f}_t\left(\{x_i\}_{i\in\mathbb{N}_t},\mathcal{Y}_t\right)\prod_{i=0}^{t}\mathrm{d}x_i},
\end{aligned}
\tag{2.154}
$$

where

$$
\begin{aligned}
\lambda_t &\triangleq \frac{f_t\left(x_0,x_1,\ldots x_t,Y_0\left(\omega\right),Y_1\left(\omega\right),\ldots,Y_t\left(\omega\right)\right)}{\widetilde{f}_t\left(x_0,x_1,\ldots x_t,Y_0\left(\omega\right),Y_1\left(\omega\right),\ldots,Y_t\left(\omega\right)\right)} \\
&\equiv \lambda_t\left(\{x_i\}_{i\in\mathbb{N}_t},\mathcal{Y}_t\left(\omega\right)\right)\in\mathbb{R}_+, \quad \forall\omega\in\Omega,
\end{aligned}
\tag{2.155}
$$

constitutes a "half ordinary function - half random variable" likelihood ratio and where the condition (2.16) ensures its boundedness. Of course, although the likelihood ratio can be indeterminate when both densities are zero, the respective points do not contribute in the computation of the relevant integrals presented above, because these belong to measurable sets corresponding to events of measure zero. From (2.154) and by definition of the conditional density of $\{X_i\}_{i\in\mathbb{N}_t}$ given $\{Y_i\}_{i\in\mathbb{N}_t}$, we immediately get

$$
\begin{aligned}
\widehat{X}_t &\equiv \mathbb{E}_{\mathcal{P}}\left\{X_t\middle|\mathscr{Y}_t\right\} \\
&\equiv \frac{\int x_t \lambda_t \widetilde{f}_{\mathcal{X}_t|\mathcal{Y}_t}\left(\{x_i\}_{i\in\mathbb{N}_t}\middle|\mathcal{Y}_t\right)\prod_{i=0}^{t}\mathrm{d}x_i}{\int \lambda_t \widetilde{f}_{\mathcal{X}_t|\mathcal{Y}_t}\left(\{x_i\}_{i\in\mathbb{N}_t}\middle|\mathcal{Y}_t\right)\prod_{i=0}^{t}\mathrm{d}x_i} \\
&= \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t\Lambda_t\middle|Y_0,Y_1,\ldots,Y_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t\middle|Y_0,Y_1,\ldots,Y_t\right\}} \\
&\equiv \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t\Lambda_t\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t\middle|\mathscr{Y}_t\right\}},
\end{aligned}
\tag{2.156}
$$

which constitutes what we were initially set to show. ■

2.2 Grid Based Nonlinear Filtering Revisited: Recursive Estimation & Asymptotic Optimality

2.2.1 Introduction

It is well known that except for a few special cases [13–16, 24], general nonlinear filters of partially observable Markov processes (or Hidden Markov Models (HMMs)) do not admit finite dimensional (recursive) representations [11, 12]. Nonlinear filtering problems, though, arise naturally in a wide variety of important applications, including target tracking [35,36], localization and robotics [37,38], mathematical finance [39] and channel prediction in wireless sensor networks (see Section 3.2), just to name a few. Adopting the Minimum Mean Square Error (MMSE) as the standard optimality criterion, in most cases, the nonlinear filtering problem results in a dynamical system in the infinite dimensional space of measures, making the need for robust approximate solutions imperative.

Approximate nonlinear filtering methods can be primarily categorized into two major groups [40]: *local* and *global*. Local methods include the celebrated extended Kalman filter [**?**], the unscented Kalman filter [**?**], Gaussian approximations [18], cubature Kalman filters [**?**] and quadrature Kalman filters [**?**]. These methods are mainly based on the local "assumed form of the conditional density" approach, which dates back to the 1960's [17]. Local methods are characterized by relatively small computational complexity, making them applicable in relatively higher dimensional systems. However, they are strictly suboptimal and, thus, they at most constitute efficient heuristics, but *without explicit theoretical guarantees*. On the other hand, global methods, which include grid based approaches (relying on proper quantizations of the state space of the state process [19, 20, 22]) and Monte Carlo approaches (particle filters and related methods [23]), provide approximations to the *whole posterior measure of the state*. Global methods possess very powerful asymptotic optimality properties, providing explicit theoretical guarantees and predictable performance. For that reason, they are very important both in theory and practice, either as solutions, or as benchmarks for the evaluation of suboptimal techniques. The main common disadvantage of global

methods is their high computational complexity as the dimensionality of the underlying model increases. This is true both for grid based and particle filtering techniques [31, 41–44].

In this work, we focus on *grid based approximate filtering of Markov processes observed in conditionally Gaussian noise, constructed by exploiting uniform quantizations of the state*. Two types of state quantizations are considered: the *Markovian* and the *marginal* ones (see [22] and/or Section 2.2.3). Based on existing results [12, 22, 40], one can derive grid based, recursive nonlinear filtering schemes, exploitting the properties of the aforementioned types of state approximations. The novelty of our work lies in the development of an original convergence analysis of those schemes, under generic assumptions on the expansiveness of the observations (see Section 2.2.2). Our contributions can be summarized as follows:

1) For marginal state quantizations, we propose the notion of *conditional regularity of Markov kernels* (Definition 2.4), which is an easily verifiable condition for guaranteeing strong asymptotic consistency of the resulting grid based filter. Conditional regularity is a simple and relaxed condition, in contrast to more complicated and potentially stronger conditions found in the literature, such as the Lipschitz assumption imposed on the stochastic kernel(s) of the underlying process in [22].

2) Under certain conditions, we show that all grid based filters considered here converge to the true optimal nonlinear filter in a strong and controllable sense (Theorems 2.6 and 2.7). In particular, the convergence is compact in time and uniform in a measurable set occurring with probability almost 1; this event is completely characterized in terms of the filtering horizon and the dimensionality of the observations.

3) We show that all our results can be easily extended in order to support filters of *functionals* of the state and recursive, grid based approximate prediction (Theorem 2.8). More specifically, we show that grid based filters are asymptotically optimal as long as the state functional is bounded and continuous; this is a typical assumption (see also [12,20,45]). Of course, this latter assumption is in addition to and independent from any other condition (e.g., conditional regularity) imposed on the structure of the partially observable system under consideration. As it will be seen in Chapter 3, this simple

property has been proven particularly useful, in the context of channel estimation in wireless sensor networks. The assumption of a bounded and continuous state functional is more relaxed as compared to the respective bounded and Lipschitz assumption found in [22].

Another novel aspect of our contribution is that our original theoretical development is based more on linear-algebraic arguments and less on measure theoretic ones, making the presentation shorter, clearer and easy to follow.

Relation to the Literature

In this work, conditional regularity is presented as a relaxed sufficient condition for asymptotic consistency of discrete time grid based filters, employing marginal state quantizations. Another set of conditions ensuring asymptotic convergence of state approximations to optimal nonlinear filters are the Kushner's local consistency conditions (see, example, [19, 20]). These refer to Markov chain approximations for continuous time Gaussian diffusion processes and the related standard nonlinear filtering problem.

It is important to stress that, as it can be verified in Section IV, the constraints which conditional regularity imposes on the stochastic kernel of the hidden Markov process under consideration are *general and do not require* the assumption of any specific class of hidden models. In this sense, conditional regularity is a *nonparametric condition* for ensuring convergence to the optimal nonlinear filter. For example, hidden Markov processes driven by strictly non-Gaussian noise are equally supported as their Gaussian counterparts, provided the same conditions are satisfied, as suggested by conditional regularity (see Section IV). Consequently, it is clear that conditional regularity advocated in this work is different in nature than Kushner's local consistency conditions [19,20]. In fact, putting the differences between continuous and discrete time aside, conditional regularity is more general as well.

Convergence of discrete time approximate nonlinear filters (not necessarily recursive) was previously studied in Section 2.1. No special properties of the state were assumed, such as the Markov property; it was only assumed that the state is almost surely compactly supported. In this work, the results of Section 2.1 provide the tools for

showing asymptotic optimality of grid based, *recursive* approximate estimators. Further, our results will be leveraged later in Chapter 3, showing asymptotic consistency of sequential spatiotemporal estimators/predictors of the magnitude of the wireless channel over a geographical region, as well as its variance. The estimation is based on limited channel observations, obtained by a small number of sensors.

2.2.2 System Model & Problem Formulation

2.2.2.1 System Model & Technical Assumptions

All stochastic processes defined below are defined on a common complete probability space (the base space), defined by a triplet $(\Omega, \mathscr{F}, \mathcal{P})$. Also, for a set $\mathcal{A}$, $\mathscr{B}(\mathcal{A})$ denotes the respective Borel σ-algebra.

Let $X_t \in \mathbb{R}^{M\times 1}$ be Markov with *known dynamics* (stochastic kernel)[4]

$$\mathcal{K}_t : \mathscr{B}\left(\mathbb{R}^{M\times 1}\right) \times \mathbb{R}^{M\times 1} \mapsto [0,1], \quad t \in \mathbb{N}, \tag{2.157}$$

which, together with an initial probability measure $\mathcal{P}_{X_{-1}}$ on $\left(\mathbb{R}^{M\times 1}, \mathscr{B}\left(\mathbb{R}^{M\times 1}\right)\right)$, completely describe its stochastic behavior. Generically, the state is assumed to be compactly supported in $\mathbb{R}^{M\times 1}$, that is, for all $t \in \{-1\} \cup \mathbb{N}, X_t \in \mathcal{Z} \subset \mathbb{R}^{M\times 1}$, $\mathcal{P} - a.s.$. We may also alternatively assume the existence of an explicit state transition model describing the temporal evolution of the state, as

$$X_t \triangleq f_t(X_{t-1}, W_t) \in \mathcal{Z}, \quad \forall t \in \mathbb{N}, \tag{2.158}$$

where, for each t, $f_t : \mathcal{Z} \times \mathcal{W} \overset{a.s.}{\mapsto} \mathcal{Z}$ constitutes a measurable nonlinear state transition mapping with somewhat "favorable" analytical behavior (see below) and $W_t \equiv W_t(\omega) \in \mathcal{W} \subseteq \mathbb{R}^{M_W\times 1}$, for $t \in \mathbb{N}$, $\omega \in \Omega$, denotes a white noise process with state space $\mathcal{W}$. The recursion defined in (2.158) is initiated by choosing $X_{-1} \sim \mathcal{P}_{X_{-1}}$, independently of W_t.

The state X_t is partially observed through the *conditionally Gaussian* process

$$\mathbb{R}^{N\times 1} \ni \mathbf{y}_t | X_t \overset{i.i.d.}{\sim} \mathcal{N}\left(\boldsymbol{\mu}_t(X_t), \boldsymbol{\Sigma}_t(X_t) + \sigma^2_{\boldsymbol{\Sigma}}\mathbf{I}_N\right), \tag{2.159}$$

[4]Hereafter, we employ the usual notation $\mathcal{K}_t(\mathcal{A}|X_{t-1} \equiv \boldsymbol{x}) \equiv \mathcal{K}_t(\mathcal{A}|\boldsymbol{x})$, for $\mathcal{A}$ Borel.

$\sigma_{\mathbf{\Sigma}} \geq 0$, with conditional means and variances known apriori, for all $t \in \mathbb{N}$. Additionally, we assume that $\mathbf{\Sigma}_t(X_t) \succ \mathbf{0}$, with $\mathbf{\Sigma}_t : \mathcal{Z} \mapsto \mathcal{D}_{\mathbf{\Sigma}}$, for all $t \in \mathbb{N}$, where $\mathcal{D}_{\mathbf{\Sigma}} \subset \mathbb{R}^{N \times N}$ is bounded. The observations (2.159) can also be rewritten in the canonical form $\mathbf{y}_t \equiv \boldsymbol{\mu}_t(X_t) + \sqrt{\mathbf{C}_t(X_t)}\boldsymbol{u}_t$, for all $t \in \mathbb{N}$, where $\boldsymbol{u}_t \equiv \boldsymbol{u}_t(\omega)$ constitutes a standard Gaussian white noise process and, for all $\boldsymbol{x} \in \mathcal{Z}$, $\mathbf{C}_t(\boldsymbol{x}) \triangleq \mathbf{\Sigma}_t(\boldsymbol{x}) + \sigma_{\mathbf{\Sigma}}^2 \mathbf{I}_N$. The process $\boldsymbol{u}_t$ is assumed to be mutually independent of X_{-1}, and of the innovations W_t, in case $X_t \equiv f_t(X_{t-1}, W_t)$.

The class of partially observable systems described above is very wide, containing all (first order) Hidden Markov Models (HMMs) with compactly supported state processes and conditionally Gaussian measurements. Hereafter, without loss of generality and in order to facilitate the presentation, we will assume stationarity of state transitions, dropping the subscript "t" in the respective stochastic kernels and/or transition mappings. However, we should mention that all subsequent results hold true also for the nonstationary case, if one assumes that any condition hereafter imposed on the mechanism generating X_t holds for all $t \in \mathbb{N}$, that is, for all different "modes" of the state process. As in Section 2.1, Assumptions 2.1 and 2.2 are hereafter considered to be in power.

Remark 2.7. In certain applications, conditional Gaussianity of the observations given the state may not be a valid modeling assumption. However, such a structural assumption not only allows for analytical tractability when it holds, but also provides important insights related to the performance of the respective approximate filter, even if the conditional distribution of the observations is not Gaussian, provided it is "sufficiently smooth and unimodal". ■

2.2.2.2 Prior Results & Problem Formulation

Before proceeding and for later reference, let us define the complete natural filtrations generated by the processes X_t and $\mathbf{y}_t$ as $\{\mathscr{X}_t\}_{t \in \mathbb{N} \cup \{-1\}}$ and $\{\mathscr{Y}_t\}_{t \in \mathbb{N}}$, respectively.

Adopting the MMSE as an optimality criterion for inferring the hidden process X_t on the basis of the observations, one would ideally like to discover an efficient way for evaluating the conditional expectation or *filter* of the state, given the available

information encoded in $\mathscr{Y}_t$, sequentially in time. Unfortunately, except for some very special cases, [13–16], it is well known that the optimal nonlinear filter does not admit an explicit finite dimensional representation [11, 12].

As a result, one must resort to properly designed approximations to the general nonlinear filtering problem, leading to well behaved, finite dimensional, approximate filtering schemes. Such schemes are typically derived by approximating the desired quantities of interest either heuristically (see, e.g. [17, 18]), or in some more powerful, rigorous sense, (see, e.g., Markov chain approximations [19, 20, 22], or particle filtering techniques [23, 45]). In this work, we follow the latter direction and propose a novel, rigorous development of grid based approximate filtering, focusing on the class of partially observable systems described in Section 2.2.2.A. For this, we exploit the general asymptotic results presented in Section 2.1.

As in Section 2.1, our analysis is based on the well known representation of the optimal filter, employing the simple concept (at least in discrete time) of *change of probability measures* (see, e.g., [12, 15, 16, 46]). Let $\mathbb{E}_{\mathcal{P}}\{X_t|\,\mathscr{Y}_t\}$ denote the filter of X_t given $\mathscr{Y}_t$, under the base measure $\mathcal{P}$. Then, there exists another (hypothetical) probability measure $\widetilde{\mathcal{P}}$ [12], *equivalent to* $\mathcal{P}$, such that

$$\mathbb{E}_{\mathcal{P}}\{X_t|\,\mathscr{Y}_t\} \equiv \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\{X_t\Lambda_t|\,\mathscr{Y}_t\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\{\Lambda_t|\,\mathscr{Y}_t\}}, \tag{2.160}$$

where $\Lambda_t \triangleq \prod_{i\in\mathbb{N}_t} \mathsf{L}_i(X_i, \mathbf{y}_i)$ and $\mathsf{L}_t(X_t, \mathbf{y}_t) \triangleq \left(\sqrt{2\pi}\right)^N \mathcal{N}(\mathbf{y}_t; \boldsymbol{\mu}_t(X_t), \mathbf{C}_t(X_t))$, for all $t \in \mathbb{N}$, with $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \mathbf{C})$ denoting the multivariate Gaussian density as a function of $\boldsymbol{x}$, with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. Here, we also define $\Lambda_{-1} \equiv 1$. The most important part is that, under $\widetilde{\mathcal{P}}$, the processes X_t (including the initial value X_{-1}) and $\mathbf{y}_t$ are mutually statistically independent, with X_t being the same as under the original measure and $\mathbf{y}_t$ being a Gaussian vector white noise process with zero mean and covariance matrix the identity. As one might guess, the measure $\widetilde{\mathcal{P}}$ is more convenient to work with. It is worth mentioning (see Section 2.1) that the *Feynman-Kac formula* (2.160) is true regardless of the nature of the state X_t, that is, it holds even if X_t is not Markov. In fact, the machinery of change of measures can be applied to any nonlinear filtering problem and is not tied to the particular filtering formulations

considered in this work [12].

Let us now replace X_t in the RHS of (2.160) with another process $X_t^{L_S}$, called the *approximation*, with *resolution* or *approximation parameter* $L_S \in \mathbb{N}$ (conventionally), also independent of the observations under $\widetilde{\mathcal{P}}$, for which the evaluation of the resulting "filter" might be easier. Then, we can define the *approximate filter* of the state X_t

$$\mathcal{E}^{L_S}\left(X_t \middle| \mathscr{Y}_t\right) \triangleq \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t^{L_S}\Lambda_t^{L_S}\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t^{L_S}\middle|\mathscr{Y}_t\right\}}, \quad \forall t \in \mathbb{N}. \tag{2.161}$$

It was shown in Section 2.1 that, under certain conditions, this approximate filter is strongly asymptotically consistent, as follows.

Theorem 2.4. (Convergence to the Optimal Filter) *Pick any natural $T < \infty$ and suppose either of the following:*

- *For all $t \in \mathbb{N}_T$, the sequence $\left\{X_t^{L_S}\right\}_{L_S\in\mathbb{N}}$ is marginally $\mathcal{C}$-weakly convergent to X_t, given X_t, that is,*

$$\mathcal{P}^{L_S}_{X_t^{L_S}\big|X_t}\left(\cdot\middle|X_t\right) \underset{L_S\to\infty}{\overset{\mathcal{W}}{\longrightarrow}} \delta_{X_t}\left(\cdot\right), \quad \forall t \in \mathbb{N}_T. \tag{2.162}$$

- *For all $t \in \mathbb{N}_T$, the sequence $\left\{X_t^{L_S}\right\}_{L_S\in\mathbb{N}}$ is (marginally) convergent to X_t in probability, that is,*

$$X_t^{L_S} \underset{L_S\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} X_t, \quad \forall t \in \mathbb{N}_T. \tag{2.163}$$

Then, there exists a measurable subset $\widehat{\Omega}_T \subseteq \Omega$ with $\mathcal{P}$-measure at least $1-(T+1)^{1-CN}\exp\left(-CN\right)$, such that

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left\|\mathcal{E}^{L_S}\left(X_t\middle|\mathscr{Y}_t\right)-\mathbb{E}_{\mathcal{P}}\left\{X_t\middle|\mathscr{Y}_t\right\}\right\|_1(\omega) \underset{L_S\to\infty}{\longrightarrow} 0, \tag{2.164}$$

for any free, finite constant $C \geq 1$. In other words, the convergence of the respective approximate filtering operators is compact in $t \in \mathbb{N}$ and, with probability at least $1-(T+1)^{1-CN}\exp\left(-CN\right)$, uniform in ω.

Theorem 2.4 constitutes the full-blown version of Theorem 2.3 of Section 2.1.3, and is presented here for completeness. Let us also recall some basic facts from our earlier development, presented in Section 2.1.

Remark 2.8. It should be mentioned that Theorem 2.4 holds for *any process* X_t, *Markov or not*, as long as X_t is almost surely compactly supported. ■

Remark 2.9. The mode of filter convergence reported in Theorem 2.4 is particularly strong. It implies that *inside any fixed finite time interval and among almost all possible paths* of the observations process, the approximation error between the true and approximate filters is finitely bounded and converges to zero, as the grid resolution increases, resulting in a practically appealing asymptotic property. This mode of convergence constitutes, in a sense, a practically useful, quantitative justification of Egorov's Theorem [21], which abstractly relates almost uniform convergence with almost sure convergence of measurable functions. Further, it is important to mention that, for fixed T, convergence to the optimal filter tends to be in the uniformly almost everywhere sense, *at an exponential rate* with respect to the dimensionality of the observations, N. This shows that, in a sense, the dimensionality of the observations *stochastically stabilizes* the approximate filtering process. ■

Remark 2.10. Observe that the adopted approach concerning construction of the approximate filter of X_t, the approximation $X_t^{L_S}$ is naturally constructed under the base measure $\widetilde{\mathcal{P}}$, satisfying the constraint of being independent of the observations, $\mathbf{y}_t$. However, it is easy to see that if, for each t in the horizon of interest, $X_t^{L_S}$ is $\{\mathscr{X}_t\}$-adapted, then it may be defined under the original base measure $\mathcal{P}$ without any complication; under $\widetilde{\mathcal{P}}$, X_t (and, thus, $X_t^{L_S}$) is independent of $\mathbf{y}_t$ by construction. In greater generality, $X_t^{L_S}$ may be constructed under $\mathcal{P}$, as long as it can be somehow guaranteed to follow the same distribution *and* be independent of $\mathbf{y}_t$ under $\widetilde{\mathcal{P}}$. As we shall see below, this is not always obvious or true; if fact, it is strongly dependent on the information (encoded in the appropriate σ-algebra) exploited in order to define the process $X_t^{L_S}$, as well as the particular choice of the alternative measure $\widetilde{\mathcal{P}}$. ■

2.2.3 Uniform State Quantizations

Although Theorem 2.3 presented above provides the required conditions for convergence of the respective approximate filter, it does not specify any specific class of processes to be used as the required approximations. In order to satisfy either of the conditions

of Theorem 2.3, $X_t^{L_S}$ must be strongly dependent on X_t. For example, if the approximation is merely weakly convergent to the original state process (as, for instance, in particle filtering techniques), the conditions of Theorem 2.3 will not be fulfilled. In this work, the state X_t is approximated by another closely related process with discrete state space, constituting a uniformly quantized approximation of the original one.

Similarly to [22], we will consider two types of state approximations: *Marginal Quantizations* and *Markovian Quantizations*. Specifically, in the following, we study pathwise properties of the aforementioned state approximations. Nevertheless, and as in every meaningful filtering formulation, neither the state nor its approximations need to be known or constructed by the user. Only the (conditional) laws of the approximations need to be known. To this end, let us state a general definition of a quantizer.

Definition 2.3. (Quantizers) Consider a compact subset $\mathcal{A} \subset \mathbb{R}^N$, a partition $\Pi \triangleq \{\mathcal{A}_i\}_{i\in\mathbb{N}_L^+}$ of $\mathcal{A}$ and let $\mathcal{B} \triangleq \left\{\{b_i\}_{i\in\mathbb{N}_L^+}\right\}$ be a discrete set consisting of *distinct reconstruction points*, with $b_i \in \mathbb{R}^M, \forall i \in \mathbb{N}_L^+$. Then, an *L-level Euclidean Quantizer* is any bounded and measurable function $\mathcal{Q}_L : (\mathcal{A}, \mathscr{B}(\mathcal{A})) \mapsto \left(\mathcal{B}, 2^{\mathcal{B}}\right)$, defined by assigning all $x \in \mathcal{A}_i \in \Pi, i \in \mathbb{N}_L^+$ to a unique $b_j \in \mathcal{B}, j \in \mathbb{N}_L^+$, such that the mapping between the elements of Π and $\mathcal{B}$ is one to one and onto (a bijection).

2.2.3.1 Uniformly Quantizing $\mathcal{Z}$

For simplicity and without any loss of generality, suppose that $\mathcal{Z} \equiv [a,b]^M$ (for $a \in \mathbb{R}$ and $b \in \mathbb{R}$ with obviously $a < b$), representing the compact set of support of the state X_t. Also, consider a uniform L-set partition of the interval $[a,b]$, $\Pi_L \triangleq \{\mathcal{Z}_l\}_{l\in\mathbb{N}_{L-1}}$ and, additionally, let $\Pi_{L_S} \triangleq \underset{M \text{ times}}{\times} \Pi_L$ be the overloaded Cartesian product of M copies of the partitions defined above, with cardinality $L_S \triangleq L^M$. As usual, our reconstruction points will be chosen as the center of masses of the hyperrectangles comprising the hyperpartition Π_{L_S}, denoted as $\boldsymbol{x}_{L_S}^{\{l_m\}_{m\in\mathbb{N}_M^+}} \equiv \boldsymbol{x}_{L_S}^{\{l_m\}}$, where $l_m \in \mathbb{N}_{L-1}$. According to some predefined ordering, we make the identification $\boldsymbol{x}_{L_S}^{\{l_m\}} \equiv \boldsymbol{x}_{L_S}^{l}$, $l \in \mathbb{N}_{L_S}^+$. Further, let $\mathcal{X}_{L_S} \triangleq \left\{\boldsymbol{x}_{L_S}^1, \boldsymbol{x}_{L_S}^2, \ldots, \boldsymbol{x}_{L_S}^{L_S}\right\}$ and define the quantizer $\mathcal{Q}_{L_S} : (\mathcal{Z}, \mathscr{B}(\mathcal{Z})) \mapsto$

$\left(\mathcal{X}_{L_S}, 2^{\mathcal{X}_{L_S}}\right)$, where

$$\begin{aligned} &\mathcal{Q}_{L_S}(\boldsymbol{x}) \triangleq \boldsymbol{x}_{L_S}^{\{l_m\}} \equiv \boldsymbol{x}_{L_S}^{l} \in \mathcal{X}_{L_S} \\ &\text{iff} \quad \boldsymbol{x} \in \underset{m \in \mathbb{N}_M^+}{\times} \mathcal{Z}_{l_m} \triangleq \mathcal{Z}_{L_S}^{l} \in \Pi_{L_S} \end{aligned} . \tag{2.165}$$

Given the definitions stated above, the following simple and basic result is true. The proof, being elementary, is omitted.

Lemma 2.11. (Uniform Convergence of Quantized Values) *It is true that*

$$\lim_{L_S \to \infty} \sup_{\boldsymbol{x} \in \mathcal{Z}} \left\| \mathcal{Q}_{L_S}(\boldsymbol{x}) - \boldsymbol{x} \right\|_1 \equiv 0, \tag{2.166}$$

that is, $\mathcal{Q}_{L_S}(\boldsymbol{x})$ converges as $L_S \to \infty$, uniformly in $\boldsymbol{x}$.

Remark 2.11. We should mention here that Lemma 2.11, as well as all the results to be presented below hold equally well when the support of X_t is different in each dimension, or when different quantization resolutions are chosen in each dimension, just by adding additional complexity to the respective arguments. ■

2.2.3.2 Marginal Quantization

The first class of state process approximations of interest is that of marginal state quantizations, according to which X_t is approximated by its nearest neighbor

$$X_t^{L_S}(\omega) \triangleq \mathcal{Q}_{L_S}(X_t(\omega)) \in \mathcal{X}_{L_S}, \quad \forall t \in \{-1\} \cup \mathbb{N}, \tag{2.167}$$

$\mathcal{P} - a.s.$, where $L_S \in \mathbb{N}$ is identified as the approximation parameter. Next, we present another simple but important lemma, concerning the behavior of the quantized stochastic process $X_t^{L_S}(\omega)$, as L_S gets large. Again, the proof is relatively simple, and it is omitted.

Lemma 2.12. (Uniform Convergence of Marginal State Quantizations) *For $X_t(\omega) \in \mathcal{Z}$, for all $t \in \mathbb{N}$, almost surely, it is true that*

$$\lim_{L_S \to \infty} \sup_{t \in \mathbb{N}} \operatorname*{ess\,sup}_{\omega \in \Omega} \left\| X_t^{L_S}(\omega) - X_t(\omega) \right\|_1 \equiv 0, \tag{2.168}$$

that is, $X_t^{L_S}(\omega)$ converges as $L_S \to \infty$, uniformly in t and uniformly $\mathcal{P}$-almost everywhere in ω.

Remark 2.12. One drawback of marginal approximations is that they *do not* possess the Markov property any more. This fact introduces considerable complications in the development of recursive estimators, as shown later in Section 2.2.4. However, marginal approximations are practically appealing, because they do not require explicit knowledge of the stochastic kernel describing the transitions of X_t. ■

Remark 2.13. Note that the implications of Lemma 2.12 continue to be true under the base measure $\widetilde{\mathcal{P}}$. This is true because $X_t^{L_S}$ is $\{\mathscr{X}_t\}$-adapted, and also due to the fact that the "local" probability spaces $(\Omega, \mathscr{X}_\infty, \mathcal{P})$ and $\left(\Omega, \mathscr{X}_\infty, \widetilde{\mathcal{P}}\right)$ are completely identical. Here, $\mathscr{X}_\infty \triangleq \sigma\left\{\bigcup_{t\in\mathbb{N}\cup\{-1\}} \mathscr{X}_t\right\}$ constitutes the join of the filtration $\{\mathscr{X}_t\}_{t\in\mathbb{N}\cup\{-1\}}$. In other words, the restrictions of $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ on $\mathscr{X}_\infty$ -the collection of events ever to be generated by X_t- coincide; that is, $\mathcal{P}|_{\mathscr{X}_\infty} \equiv \widetilde{\mathcal{P}}\Big|_{\mathscr{X}_\infty}$. ■

2.2.3.3 Markovian Quantization

The second class of approximations considered is that of Markovian quantizations of the state. In this case, we assume explicit knowledge of a transition mapping, modeling the temporal evolution of X_t. In particular, we assume a recursion as in (2.158), where the process W_t acts as the driving noise of the state X_t and constitutes an intrinsic characteristic of it. Then, the Markovian quantization of X_t is defined as

$$X_t^{L_S} \triangleq \mathcal{Q}_{L_S}\left(f\left(X_{t-1}^{L_S}, W_t\right)\right) \in \mathcal{X}_{L_S},\ \forall t \in \mathbb{N}, \tag{2.169}$$

with $X_{-1}^{L_S} \equiv \mathcal{Q}_{L_S}(X_{-1}) \in \mathcal{X}_{L_S}$, $\mathcal{P}-a.s.$, and which satisfies the Markov property trivially; since $\mathcal{X}_{L_S}$ is finite, it constitutes a (time-homogeneous) *finite state space Markov Chain*. A scheme for generating $X_t^{L_S}$ is shown in Fig. 2.1.

At this point, it is very important to observe that, whereas X_t is guaranteed to be Markov with the same dynamics and independent of $\mathbf{y}_t$ under $\widetilde{\mathcal{P}}$, we cannot immediately say the same for the Markovian approximation $X_t^{L_S}$. The reason is that $X_t^{L_S}$ is measurable with respect to the filtration generated by the initial condition X_{-1} and the innovations process W_t and not with respect to $\{\mathscr{X}_t\}_{t\in\mathbb{N}\cup\{-1\}}$. Without any additional considerations, W_t may very well be partially correlated relative to $\mathbf{y}_t$ and/or

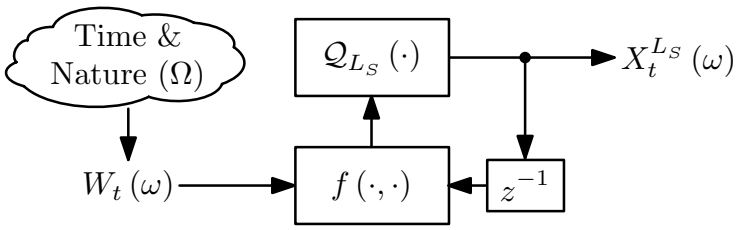


Figure 2.1: Block representation of Markovian quantization. As noted in the cloud, "Nature" here refers to the sample space Ω of the base triplet $(\Omega, \mathscr{F}, \mathcal{P})$.

X_{-1}, and/or even non white itself! Nevertheless, $\widetilde{\mathcal{P}}$ may be chosen such that W_t indeed satisfies the aforementioned properties under question, as the following result suggests.

Lemma 2.13. (Choice of $\widetilde{\mathcal{P}}$) *Without any other modification, the base measure $\widetilde{\mathcal{P}}$ may be chosen such that the initial condition X_{-1} and the innovations process W_t follow the same distributions as under $\mathcal{P}$* ***and*** *are all mutually independent relative to the observations, $\mathbf{y}_t$.*

Proof of Lemma 2.13. See Section 2.2.7.6 (Appendix F). ■

Lemma 2.13 essentially implies that Markovian quantizations may be constructed and analyzed either under $\mathcal{P}$ or $\widetilde{\mathcal{P}}$, interchangeably. Also adapt Remark 2.13 to this case.

Under the assumption of a transition mapping, every possible path of $X_t(\omega)$ is completely determined by fixing $X_{-1}(\omega)$ and $W_t(\omega)$ at any particular realization, for each $\omega \in \Omega$. As in the case of marginal quantizations, the goal of the Markovian quantization is the *pathwise* approximation of X_t by $X_t^{L_S}$, for almost all realizations of the white noise process W_t and initial value X_{-1}. In practice, however, as noted in the beginning of this section, knowledge of W_t is of course not required by the user. What is required by the user is the transition matrix of the Markov chain $X_t^{L_S}$, which could be obtained via, for instance, simulation (also see Section IV).

For analytical tractability, we will impose the following reasonable regularity assumption on the expansiveness of the transition mapping $f(\cdot,\cdot)$:

Assumption 2.3. (Expansiveness of Transition Mappings) For all $\boldsymbol{y} \in \mathcal{W}$, $f : \mathcal{Z} \times \mathcal{W} \mapsto \mathcal{Z}$ is *Lipschitz continuous* in $\boldsymbol{x} \in \mathcal{Z}$, that is, possibly dependent on each

$\boldsymbol{y}$, there exists a non-negative, bounded constant $K(\boldsymbol{y})$, where $\sup_{\boldsymbol{y}\in\mathcal{W}} K(\boldsymbol{y})$ exists and is finite, such that

$$\|f(\boldsymbol{x}_1,\boldsymbol{y}) - f(\boldsymbol{x}_2,\boldsymbol{y})\|_1 \leq K(\boldsymbol{y})\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_1, \tag{2.170}$$

$\forall(\boldsymbol{x}_1,\boldsymbol{x}_2) \in \mathcal{Z}\times\mathcal{Z}$. If, additionally, $\sup_{\boldsymbol{y}\in\mathcal{W}} K(\boldsymbol{y}) < 1$, then $f(\cdot,\cdot)$ will be referred to as *uniformly contractive*.

Employing Assumption 2.3, the next result presented below characterizes the convergence of the Markovian state approximation $X_t^{L_S}$ to the true process X_t, as the quantization of the state space $\mathcal{Z}$ gets finer and under appropriate conditions.

Lemma 2.14. (Uniform Convergence of Markovian State Quantizations) *Suppose that the transition mapping $f:\mathcal{Z}\times\mathcal{W}\mapsto\mathcal{Z}$ of the Markov process $X_t(\omega)$ is Lipschitz, almost surely and for all $t\in\mathbb{N}$. Also, consider the approximating Markov process $X_t^{L_S}(\omega)$, as defined in (2.169). Then,*

$$\lim_{L_S\to\infty} \operatorname*{ess\,sup}_{\omega\in\Omega} \left\| X_t^{L_S}(\omega) - X_t(\omega)\right\|_1 \equiv 0, \quad \forall t\in\mathbb{N}, \tag{2.171}$$

that is, $X_t^{L_S}(\omega)$ converges as $L_S\to\infty$, in the pointwise sense in t and uniformly almost everywhere in ω. If, additionally, $f(\cdot,\cdot)$ is uniformly contractive, almost surely and for all $t\in\mathbb{N}$, then it is true that

$$\lim_{L_S\to\infty} \sup_{t\in\mathbb{N}} \operatorname*{ess\,sup}_{\omega\in\Omega} \left\| X_t^{L_S}(\omega) - X_t(\omega)\right\|_1 \equiv 0, \tag{2.172}$$

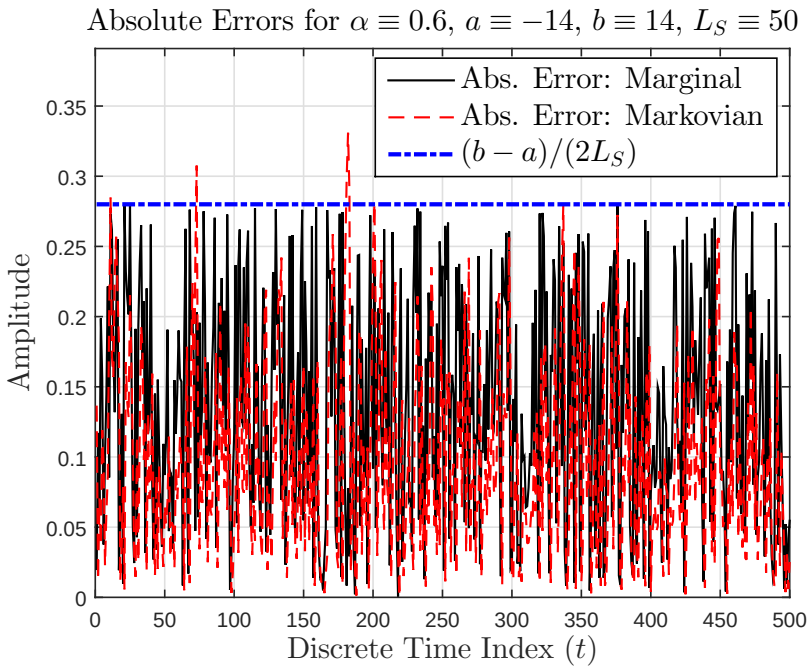
that is, the convergence is additionally uniform in t.

Proof of Lemma 2.14. See Section 2.2.7.1 (Appendix A). ■

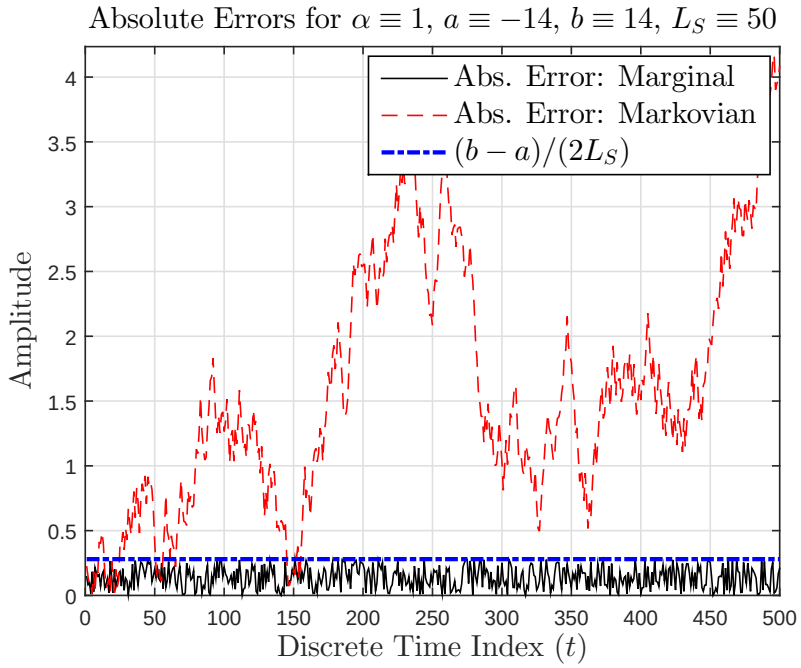
Especially concerning temporally uniform convergence of the quantization schemes under consideration, and to highlight its great practical importance, it would be useful to illustrate the implications of Lemmata 2.12 and 2.14 by means of the following simple numerical example.

Example 2.1. Let X_t be a scalar, first order autoregressive process ($AR(1)$), defined via the linear stochastic difference equation

$$X_t \triangleq \alpha X_{t-1} + W_t, \quad \forall t\in\mathbb{N}, \tag{2.173}$$



(a)



(b)

Figure 2.2: Absolute errors between each of the quantized versions of the $AR(1)$ process of our example, and the true process itself, respectively, for (a) $\alpha \equiv 0.6$ (stable process) and (b) $\alpha \equiv 1$ (a random walk).

where $W_t \overset{i.i.d}{\sim} \mathcal{N}(0,1), \forall t \in \mathbb{N}$. In our example, the parameter $\alpha \in [-1,1]$ is known apriori and controls the stability of the process, with the case where $\alpha \equiv 1$ corresponding to a Gaussian random walk. Of course, it is true that the state space of the process defined by (2.173) is the whole $\mathbb{R}$, which means that, strictly speaking, there are no finite a and b such that $X_t \in [a,b] \equiv \mathcal{Z}, \forall t \in \mathbb{N}$, with probability 1. However, it is true that for sufficiently large but finite a and b, there exists a "large" measurable set of possible outcomes for which X_t, being a Gaussian process, indeed belongs to $\mathcal{Z}$ with very high probability. Whenever this happens, we should be able to verify Lemmata 2.12 and 2.14 directly.

Additionally, it is trivial to verify that the linear transition function in (2.173) is always a contraction, with Lipschitz constant $K \equiv |\alpha|$, whenever the $AR(1)$ process of interest is stable, that is, whenever $|\alpha| < 1$.

Fig. 2.2(a) and 2.2(b) show the absolute errors between two $AR(1)$ processes and their quantized versions according to Lemmata 2.12 and 2.14, for $\alpha \equiv 0.6$ and $\alpha \equiv 1$, respectively. From the figure, one can readily observe that the marginal quantization of X_t always converges to X_t uniformly in time, regardless of the particular value of α, experimentally validating Lemma 2.12. On the other hand, it is obvious that when the transition function of our system is not a contraction (Lemma 2.14), uniform convergence of the respective Markovian quantization to the true state X_t cannot be guaranteed. Of course, we have not proved any additional necessity regarding our sufficiency assumption related to the contractiveness of the transition mapping of the process of interest, meaning that there might exist processes which do not fulfill this requirement and still converge uniformly. However, for uniform contractions, the convergence will always be uniform *whenever the process X_t is bounded in $\mathcal{Z}$.* ■

2.2.4 Recursive Estimation & Asymptotic Optimality

It is indeed easy to show that when used as candidate state approximations for defining approximate filtering operators in the fashion of Section 2.2.2.B, both the marginal and Markovian quantization schemes presented in Sections 2.2.3.B and 2.2.3.C, respectively, converge to the optimal nonlinear filter of the state X_t. Convergence is in the sense of

Theorem 2.3 presented in Section 2.2.2.B, *corroborating asymptotic optimality under a unified convergence criterion.*

Specifically, under the respective (and usual) assumptions, Lemmata 2.12 and 2.14 presented above imply that both the marginal and Markovian approximations converge to the true state X_t at least in the almost sure sense, for all $t \in \mathbb{N}$. Therefore, both will also converge to the true state in probability, satisfying the second sufficient condition of Theorem 2.3. The following result is true. Its proof, being apparent, is omitted.

Theorem 2.5. (Convergence of Approximate Filters) *Pick any natural $T < \infty$ and let the process $X_t^{L_S}$ represent either the marginal or the Markovian approximation of the state X_t. Then, under the respective assumptions implied by Lemmata 2.12 and 2.14, the approximate filter $\mathcal{E}^{L_S}\left(X_t \middle| \mathscr{Y}_t\right)$ converges to the true nonlinear filter $\mathbb{E}_{\mathcal{P}}\left\{X_t \middle| \mathscr{Y}_t\right\}$, in the sense of Theorem 2.3.*

Although Theorem 2.5 shows asymptotic consistency of the marginal and Markovian approximate filters in a strong sense, it does not imply the existence of any finite dimensional scheme for actually realizing these estimators. This is the purpose of the next subsections. In particular, we develop recursive representations for the asymptotically optimal (as $L_S \to \infty$) filter $\mathcal{E}^{L_S}\left(X_t \middle| \mathscr{Y}_t\right)$, as defined previously in (2.44).

For later reference, let us define the *bijective* mapping (a trivial quantizer) $\mathcal{Q}_{L_S}^e : \left(\mathcal{X}_{L_S}, 2^{\mathcal{X}_{L_S}}\right) \mapsto \left(\mathcal{V}_{L_S}, 2^{\mathcal{V}_{L_S}}\right)$, where the set $\mathcal{V}_{L_S} \triangleq \left\{\mathbf{e}_1^{L_S}, \ldots, \mathbf{e}_{L_S}^{L_S}\right\}$ contains the complete standard basis in $\mathbb{R}^{L_S \times 1}$. Since $\boldsymbol{x}_{L_S}^l$ is bijectively mapped to $\mathbf{e}_l^{L_S}$ for all $l \in \mathbb{N}_{L_S}^+$, we can write $\boldsymbol{x}_{L_S}^l \equiv \mathbf{X}\mathbf{e}_l^{L_S}$, where $\mathbf{X} \triangleq \left[\boldsymbol{x}_{L_S}^1 \, \boldsymbol{x}_{L_S}^2 \, \ldots \, \boldsymbol{x}_{L_S}^{L_S}\right] \in \mathbb{R}^{M \times L_S}$ constitutes the respective *reconstruction* matrix. From this discussion, it is obvious that

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_t^{L_S}\Lambda_t^{L_S}\middle|\mathscr{Y}_t\right\} \equiv \mathbf{X}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}_{L_S}^e\left(X_t^{L_S}\right)\Lambda_t^{L_S}\middle|\mathscr{Y}_t\right\}, \tag{2.174}$$

leading to the expression

$$\mathcal{E}^{L_S}\left(X_t \middle| \mathscr{Y}_t\right) \equiv \frac{\mathbf{X}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}_{L_S}^e\left(X_t^{L_S}\right)\Lambda_t^{L_S}\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t^{L_S}\middle|\mathscr{Y}_t\right\}}, \tag{2.175}$$

for all $t \in \mathbb{N}$, regardless of the type of state quantization employed. We additionally define the *likelihood* matrix

$$\mathbf{\Lambda}_t \triangleq \operatorname{diag}\left(\mathsf{L}_t\left(\boldsymbol{x}_{L_S}^1, \mathbf{y}_t\right) \, \ldots \, \mathsf{L}_t\left(\boldsymbol{x}_{L_S}^{L_S}, \mathbf{y}_t\right)\right) \in \mathbb{R}^{L_S \times L_S}. \tag{2.176}$$

Also to be subsequently used, given the quantization type, define the column stochastic matrix $\boldsymbol{P} \in [0,1]^{L_S \times L_S}$ as

$$\boldsymbol{P}(i,j) \triangleq \mathcal{P}\left(X_t^{L_S} \equiv \boldsymbol{x}_{L_S}^i \middle| X_{t-1}^{L_S} \equiv \boldsymbol{x}_{L_S}^j \right), \tag{2.177}$$

for all $(i,j) \in \mathbb{N}_{L_S}^+ \times \mathbb{N}_{L_S}^+$.

At this point, it will be important to note that the transition matrix $\boldsymbol{P}$ defined in (2.177) is implicitly assumed to be time invariant, regardless of the state approximation employed. Under the system model established in Section 2.2.2.A (assuming temporal homogeneity for the original Markov process X_t), this is unconditionally true when one considers Markovian state quantizations, simply because the resulting approximating process $X_t^{L_S}$ constitutes a Markov chain with finite state space, as stated earlier in Section 2.2.3.C. On the other hand, the situation is quite different when one considers marginal quantizations of the state. In that case, the conditional probabilities

$$\mathcal{P}\left(X_t^{L_S} \equiv \boldsymbol{x}_{L_S}^i \middle| X_{t-1}^{L_S} \equiv \boldsymbol{x}_{L_S}^j \right) \equiv \mathcal{P}\left(X_t \in \mathcal{Z}_{L_S}^i \middle| X_{t-1} \in \mathcal{Z}_{L_S}^j \right), \tag{2.178}$$

which would correspond to the (i,j)-th element of the resulting transition matrix, are, in general, *not time invariant any more, even if the original Markov process is time homogeneous*. Nevertheless, assuming the existence of at least one *invariant measure* (a stationary distribution) for the Markov process X_t, *also chosen as its initial distribution*, the aforementioned probabilities are indeed time invariant. This is a very common and reasonable assumption employed in practice, especially when tracking stationary signals. For notational and intuitional simplicity, and in order to present a unified treatment of all the approximate filters considered in this work, the aforementioned assumption will also be adopted in the analysis that follows.

2.2.4.1 Markovian Quantization

We start with the case of Markovian quantizations, since it is easier and more straightforward. Here, the development of the respective approximate filter is based on the fact that $X_t^{L_S}$ constitutes a Markov chain. Actually, this fact is the only requirement for the existence of a recursive realization of the filter, with Lemma 3 providing a sufficient

condition, ensuring asymptotic optimality. The resulting recursive scheme is summarized in the following result. The proof is omitted, since it involves standard arguments in nonlinear filtering, similar to the ones employed in the derivation of the filtering recursions for a partially observed Markov chain with finite state space [12, 15, 47], as previously mentioned.

Theorem 2.6. (The Markovian Filter) *Consider the Markovian state approximation* $X_t^{L_S}$ *and define* $\boldsymbol{E}_t \triangleq \mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ \mathcal{Q}_{L_S}^e\left(X_t^{L_S}\right)\Lambda_t^{L_S}\middle|\,\mathscr{Y}_t\right\} \in \mathbb{R}^{L_S\times 1}$, *for all* $t\in\mathbb{N}$. *Then, under the appropriate assumptions (Lipschitz property of Lemma 2.14), the asymptotically optimal in* L_S *approximate grid based filter* $\mathcal{E}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)$ *can be expressed as*

$$\mathcal{E}^{L_S}\left(X_t|\,\mathscr{Y}_t\right) \equiv \frac{\mathbf{X}\boldsymbol{E}_t}{\|\boldsymbol{E}_t\|_1}, \quad \forall t\in\mathbb{N}, \tag{2.179}$$

where the process $\boldsymbol{E}_t$ *satisfies the linear recursion*

$$\boldsymbol{E}_t \equiv \boldsymbol{\Lambda}_t\boldsymbol{P}\boldsymbol{E}_{t-1}, \quad \forall t\in\mathbb{N}. \tag{2.180}$$

The filter is initialized setting $\boldsymbol{E}_{-1} \triangleq \mathbb{E}_{\mathcal{P}}\left\{\mathcal{Q}_{L_S}^e\left(X_{-1}^{L_S}\right)\right\}$.

Remark 2.14. It is worth mentioning that, although *formally similar* to, the approximate filter introduced in Theorem 2.6 *does not* refer to a Markov chain with finite state space, because the observations process utilized in the filtering iterations corresponds to that of the *real* partially observable system under consideration. The quantity $\mathcal{E}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)$ *does not constitute a conditional expectation* of the Markov chain associated with $\boldsymbol{P}$, because the latter process does not follow the probability law of the true state process X_t. ■

Remark 2.15. In fact, $\boldsymbol{E}_t$ may be interpreted as a vector encoding an *unnormalized* point mass function, which, *roughly speaking*, expresses *the belief of the quantized state*, given the observations up to and including time t. Normalization by $\|\boldsymbol{E}_t\|_1$ corresponds precisely to a point mass function. ■

Remark 2.16. For the benefit of the reader, we should mention that the Markovian filter considered above essentially coincides with the approximate grid based filter reported in ([23], Section IV.B), although the construction of the two filters is different: the

former is constructed via a Markovian quantization of the state, whereas the latter [23] is based on a "quasi-marginal" approach (compare with (2.178)). Nevertheless, given our assumptions on the HMM under consideration, both formulations result in exactly the same transition matrix. Therefore, the optimality properties of the Markovian filter are indeed inherited by the grid based filter described in [23]. ■

2.2.4.2 Marginal Quantization

We now move on to the case of marginal quantizations. In order to be able to come up with a simple, Markov chain based, recursive filtering scheme, as in the case of Markovian quantizations previously treated, it turns out that a further assumption is required, this time concerning the stochastic kernel of the Markov process X_t. But before embarking on the relevant analysis, let us present some essential definitions.

First, for any process X_t, we will say that a sequence of functions $\{f_n(\cdot)\}_n$ is $\mathcal{P}_{X_t}$–UI, if $\{f_n(\cdot)\}_n$ is Uniformly Integrable with respect to the pushforward measure induced by X_t, $\mathcal{P}_{X_t}$, where $t \in \mathbb{N} \cup \{-1\}$, i.e.,

$$\lim_{K\to\infty} \sup_n \int_{\{|f_n(\boldsymbol{x})|>K\}} |f_n(\boldsymbol{x})| \, \mathcal{P}_{X_t}(\mathrm{d}\boldsymbol{x}) \equiv 0. \tag{2.181}$$

Second, given L_S, recall from Section 2.2.3.A that the set Π_{L_S} contains as members all quantization regions of $\mathcal{Z}$, $\mathcal{Z}_{L_S}^j$, $j \in \mathbb{N}_{L_S}^+$. Then, given the stochastic kernel $\mathcal{K}(\cdot|\cdot)$ associated with the time invariant transitions of X_t and for each $L_S \in \mathbb{N}^+$, we define the *cumulative kernel*

$$\begin{aligned}
\mathcal{K}\big(\mathcal{A}|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\big) &\triangleq \frac{\displaystyle\int_{\mathcal{Z}_{L_S}(\boldsymbol{x})} \mathcal{K}(\mathcal{A}|\,\boldsymbol{\theta})\,\mathcal{P}_{X_{t-1}}(\mathrm{d}\boldsymbol{\theta})}{\mathcal{P}\left(X_{t-1}\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right)} \\
&\equiv \frac{\mathbb{E}\left\{\mathcal{K}(\mathcal{A}|\,X_{t-1})\,\mathbb{1}_{\left\{X_{t-1}\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right\}}\right\}}{\mathbb{E}\left\{\mathbb{1}_{\left\{X_{t-1}\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right\}}\right\}} \\
&\equiv \mathbb{E}\left\{\mathcal{K}(\mathcal{A}|\,X_{t-1})|\,X_{t-1}\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right\},
\end{aligned} \tag{2.182}$$

for all Borel $\mathcal{A} \in \mathscr{B}\left(\mathbb{R}^{M\times 1}\right)$ and all $\boldsymbol{x} \in \mathcal{Z}$, where $\mathcal{Z}_{L_S}(\boldsymbol{x}) \in \Pi_{L_S}$ denotes the unique quantization region, which includes $\boldsymbol{x}$. Note that if $\boldsymbol{x}$ is substituted by $X_{t-1}(\omega)$,

the resulting quantity $\mathcal{Z}_{L_S}(X_{t-1}(\omega))$ constitutes an $\mathscr{X}_t$-predictable *set-valued random element*. Now, if, for any $\boldsymbol{x} \in \mathcal{Z}$, $\mathcal{K}(\cdot|\,\boldsymbol{x})$ admits a stochastic kernel density $\kappa : \mathbb{R}^{M\times 1} \times \mathbb{R}^{M\times 1} \mapsto \mathbb{R}_+$, suggestively denoted as $\kappa(\boldsymbol{y}|\,\boldsymbol{x})$, we define, in exactly the same fashion as above, the *cumulative kernel density*

$$\kappa\left(\boldsymbol{y}|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right) \triangleq \mathbb{E}\left\{\kappa(\boldsymbol{y}|\,X_{t-1})|\,X_{t-1}\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right\}, \tag{2.183}$$

for all $\boldsymbol{y} \in \mathbb{R}^{M\times 1}$. The fact that $\kappa\left(\cdot|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right)$ is indeed a Radon-Nikodym derivative of $\mathcal{K}\left(\cdot|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right)$ readily follows by definition of the latter and Fubini's Theorem.

Remark 2.17. Observe that, although integration is with respect to $\mathcal{P}_{X_{t-1}}$ on the RHS of (2.182), $\mathcal{K}\left(\cdot|\in\mathcal{Z}_{L_S}(\cdot)\right)$ is time invariant. This is due to stationarity of X_t, as assumed in the beginning of Section 2.2.4, implying time invariance of the marginal measure $\mathcal{P}_{X_t}$, for all $t \in \mathbb{N}\cup\{-1\}$. Additionally, for each $\boldsymbol{x} \in \mathcal{Z}$, when $\mathcal{A}$ is restricted to Π_{L_S}, $\mathcal{K}\left(\mathcal{A}|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right)$ corresponds to an entry of the (time invariant) matrix $\boldsymbol{P}$, also defined earlier. In the general case, where the aforementioned cumulative kernel is time varying, all subsequent analysis continues to be valid, just by adding additional notational complexity. ■

In respect to the relevant assumption required on $\mathcal{K}(\cdot|\,\cdot)$, as asserted above, let us now present the following definition.

Definition 2.4. (Cumulative Conditional Regularity of Markov Kernels) Consider the kernel $\mathcal{K}(\cdot|\,\cdot)$, associated with X_t, for all $t \in \mathbb{N}$. We say that $\mathcal{K}(\cdot|\,\cdot)$ is *Conditionally Regular of Type I (CRT I)*, if, for $\mathcal{P}_{X_t} \equiv \mathcal{P}_{X_{-1}}$-almost all $\boldsymbol{x}$, there exists a $\mathcal{P}_{X_{-1}} - UI$ sequence $\left\{\delta_n^I(\cdot) \geq 0\right\}_{n\in\mathbb{N}^+}$ with $\delta_n^I(\cdot) \xrightarrow[n\to\infty]{a.e.} 0$, such that

$$\sup_{\mathcal{A}\in\Pi_{L_S}} \left|\mathcal{K}(\mathcal{A}|\,\boldsymbol{x}) - \mathcal{K}\left(\mathcal{A}|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right)\right| \leq \frac{\delta_{L_S}^I(\boldsymbol{x})}{L_S}. \tag{2.184}$$

If, further, for $\mathcal{P}_{X_{-1}}$-almost all $\boldsymbol{x}$, the measure $\mathcal{K}(\cdot|\,\boldsymbol{x})$ admits a density $\kappa(\cdot|\,\boldsymbol{x})$, and if there exists another $\mathcal{P}_{X_{-1}} - UI$ sequence $\left\{\delta_n^{II}(\cdot) \geq 0\right\}_{n\in\mathbb{N}^+}$ with $\delta_n^{II}(\cdot) \xrightarrow[n\to\infty]{a.e.} 0$, such that

$$\operatorname*{ess\,sup}_{\boldsymbol{y}\in\mathbb{R}^{M\times 1}} \left|\kappa(\boldsymbol{y}|\,\boldsymbol{x}) - \kappa\left(\boldsymbol{y}|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right)\right| \leq \delta_{L_S}^{II}(\boldsymbol{x}), \tag{2.185}$$

$\mathcal{K}(\cdot|\,\cdot)$ is called *Conditionally Regular of Type II (CRT II)*. In any case, X_t will also be called conditionally regular.

A consequence of conditional regularity is the following Martingale Difference (MD) [11, 12] type representation of the marginally quantized process $\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)$.

Lemma 2.15. (Semirecursive MD-type Representation of Marginal Quantizations) *Assume that the state process X_t is conditionally regular. Then, the quantized process $\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)$ admits the representation*

$$\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right) \equiv \boldsymbol{P}\mathcal{Q}^e_{L_S}\left(X_{t-1}^{L_S}\right) + \boldsymbol{\mathcal{M}}_t^e + \boldsymbol{\varepsilon}_t^{L_S}, \tag{2.186}$$

where, under the base measure $\widetilde{\mathcal{P}}$,$\boldsymbol{\mathcal{M}}_t^e \in \mathbb{R}^{L_S\times 1}$ constitutes an $\mathscr{X}_t$-MD process and $\boldsymbol{\varepsilon}_t^{L_S} \in \mathbb{R}^{L_S\times 1}$ constitutes a $\{\mathscr{X}_t\}$-predictable process, such that

- *if X_t is CRT I, then*

$$\left\|\boldsymbol{\varepsilon}_t^{L_S}\right\|_1 \leq \delta^I_{L_S}\left(X_{t-1}\right) \underset{L_S\to\infty}{\longrightarrow} 0,\ \widetilde{\mathcal{P}}-a.s., \tag{2.187}$$

- *whereas, if X_t is CRT II, then*

$$\left\|\boldsymbol{\varepsilon}_t^{L_S}\right\|_1 \leq |b-a|^M \delta^{II}_{L_S}\left(X_{t-1}\right) \underset{L_S\to\infty}{\longrightarrow} 0,\ \widetilde{\mathcal{P}}-a.s., \tag{2.188}$$

everywhere in time.

Proof of Lemma 2.15. See Section 2.2.7.2 (Appendix B). ■

Now, consider an auxiliary Markov chain $Z_t^{L_S} \in \mathcal{V}_{L_S}$, with $\boldsymbol{P}$ (defined as in (2.177)) as its transition matrix and with initial distribution to be specified. Of course, $Z_t^{L_S}$ can be represented as $Z_t^{L_S} \equiv \boldsymbol{P}Z_{t-1}^{L_S} + \widetilde{\boldsymbol{\mathcal{M}}}_t^e$, where $\widetilde{\boldsymbol{\mathcal{M}}}_t^e \in \mathbb{R}^{L_S\times 1}$ constitutes a $\mathscr{Z}_t$-MD process, with $\{\mathscr{Z}_t\}_{t\in\mathbb{N}}$ being the complete natural filtration generated by $Z_t^{L_S}$.

Due to the existence of the "bias" process $\boldsymbol{\varepsilon}_t^{L_S}$ in the martingale difference representation of $\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)$ (see Lemma 2.15), the direct derivation of a filtering recursion for this process is difficult. However, it turns out that the approximate filter involving the marginal state quantization $X_t^{L_S}$, $\mathcal{E}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)$, can be further approximated by the also approximate filter

$$\widetilde{\mathcal{E}}^{L_S}\left(X_t|\,\mathscr{Y}_t\right) \triangleq \frac{\mathbf{X}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{Z_t^{L_S}\Lambda_t^{Z,L_S}\middle|\,\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t^{Z,L_S}\middle|\,\mathscr{Y}_t\right\}}, \tag{2.189}$$

for all $t \in \mathbb{N}$, where the functional Λ_t^{Z,L_S} is defined exactly like $\Lambda_t^{L_S}$, but replacing $X_t^{L_S}$ with $Z_t^{L_S}$. This latter filter indeed admits the recursive representation proposed in Theorem 2.6 (with $\boldsymbol{P}$ defined as in (2.177), reflecting the choice of a marginal state approximation).

Consequently, if we are interested in the asymptotic behavior of the approximation error between $\widetilde{\mathcal{E}}^{L_S}(X_t|\mathscr{Y}_t)$ and the original nonlinear filter $\mathbb{E}_{\mathcal{P}}\{X_t|\mathscr{Y}_t\}$, we can write

$$\begin{aligned}&\left\|\mathbb{E}_{\mathcal{P}}\{X_t|\mathscr{Y}_t\} - \widetilde{\mathcal{E}}^{L_S}(X_t|\mathscr{Y}_t)\right\|_1 \\ &\qquad \leq \left\|\mathbb{E}_{\mathcal{P}}\{X_t|\mathscr{Y}_t\} - \mathcal{E}^{L_S}(X_t|\mathscr{Y}_t)\right\|_1 + \left\|\mathcal{E}^{L_S}(X_t|\mathscr{Y}_t) - \widetilde{\mathcal{E}}^{L_S}(X_t|\mathscr{Y}_t)\right\|_1. \quad (2.190)\end{aligned}$$

However, from Theorem 2.5, we know that, under the respective conditions,

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left\|\mathcal{E}^{L_S}(X_t|\mathscr{Y}_t) - \mathbb{E}_{\mathcal{P}}\{X_t|\mathscr{Y}_t\}\right\|_1 \underset{L_S\to\infty}{\longrightarrow} 0. \quad (2.191)$$

Therefore, if we show that error between $\mathcal{E}^{L_S}(X_t|\mathscr{Y}_t)$ and $\widetilde{\mathcal{E}}^{L_S}(X_t|\mathscr{Y}_t)$ vanishes in the above sense, then, $\widetilde{\mathcal{E}}^{L_S}(X_t|\mathscr{Y}_t)$ will converge to $\mathbb{E}_{\mathcal{P}}\{X_t|\mathscr{Y}_t\}$, also in the same sense. It turns out that if X_t is conditionally regular, the aforementioned desired statement always holds, as follows.

Lemma 2.16. (Convergence of Approximate Filters) *For any natural $T < \infty$, suppose that the state process X_t is conditionally regular and that the initial measure of the chain $Z_t^{L_S}$ is chosen such that*

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{Z_{-1}^{L_S}\right\} \equiv \mathbb{E}_{\mathcal{P}}\left\{\mathcal{Q}_{L_S}^{e}\left(X_{-1}^{L_S}\right)\right\}. \quad (2.192)$$

Then, for the same measurable subset $\widehat{\Omega}_T \subseteq \Omega$ of Theorem 2.3, it is true that

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left\|\mathcal{E}^{L_S}(X_t|\mathscr{Y}_t) - \widetilde{\mathcal{E}}^{L_S}(X_t|\mathscr{Y}_t)\right\|_1 \underset{L_S\to\infty}{\longrightarrow} 0. \quad (2.193)$$

Additionally, under the same setting, it follows that

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left\|\widetilde{\mathcal{E}}^{L_S}(X_t|\mathscr{Y}_t) - \mathbb{E}_{\mathcal{P}}\{X_t|\mathscr{Y}_t\}\right\|_1 \underset{L_S\to\infty}{\longrightarrow} 0. \quad (2.194)$$

Proof of Lemma 2.16. See Section 2.2.7.3 (Appendix C). ■

Finally, the next theorem establishes precisely the form of the recursive grid based filter, employing the marginal quantization of the state.

Theorem 2.7. (The Marginal Filter) *Consider the marginal state approximation* $X_t^{L_S}$ *and suppose that the state process* X_t *is conditionally regular. Then, for each* $t \in \mathbb{N}$*, the asymptotically optimal in* L_S *approximate filtering operator* $\widetilde{\mathcal{E}}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)$ *can be recursively expressed exactly as in Theorem 2.6, with initial conditions as in Lemma 2.16 and transition matrix* $\boldsymbol{P}$ *defined as in (2.177).*

Remark 2.18. **(Weak Conditional Regularity)** All the derivations presented above are still valid if, in the definition of conditional regularity (Definition 2), one replaces almost everywhere convergence of the sequences $\left\{\delta_n^I\left(\cdot\right)\right\}_n$ and $\left\{\delta_n^{II}\left(\cdot\right)\right\}_n$ with *convergence in probability*. This is due to the fact that uniform integrability plus convergence in measure are necessary and sufficient conditions for showing convergence in $\mathcal{L}_1$ (for finite measure spaces). Consequently, if we focus on, for instance, CRT I (CRT II is similar), it is easy to see that in order to ensure asymptotic consistency of the marginal approximate filter in the sense of Theorem 2.7, it suffices that, for $\mathcal{A} \in \Pi_{L_S}$ and for any $\epsilon > 0$,

$$\mathcal{P}_{X_{t-1}}\left(\sup_{\mathcal{A}}\left|\mathcal{K}(\mathcal{A}|\boldsymbol{x})-\mathcal{K}\big(\mathcal{A}|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\big)\right|>\frac{\epsilon}{L_S}\right)\underset{L_S\to\infty}{\longrightarrow}0, \tag{2.195}$$

for all $t \in \mathbb{N}$ (in general), given the stochastic kernel $\mathcal{K}\left(\cdot|\,\cdot\right)$ and for the desired choice of the quantizer $\mathcal{Q}_{L_S}\left(\cdot\right)$. Here, the $\mathcal{P}_{X_t}-UI$ sequence $\left\{\delta_n^I\left(\cdot\right)\right\}_n$ is identified as

$$\delta_{L_S}^I\left(\boldsymbol{x}\right) \equiv \sup_{\mathcal{A}\in\Pi_{L_S}} L_S\left|\mathcal{K}\left(\mathcal{A}|\,\boldsymbol{x}\right)-\mathcal{K}\left(\mathcal{A}|\in\mathcal{Z}_{L_S}\left(\boldsymbol{x}\right)\right)\right|, \tag{2.196}$$

for all $L_S \in \mathbb{N}^+$ and for almost all $\boldsymbol{x} \in \mathbb{R}^{M\times 1}$. In other words, it is required that, for any $\epsilon > 0$,

$$\sup_{\mathcal{A}\in\Pi_{L_S}}\left|\mathcal{K}\left(\mathcal{A}|\,X_{t-1}\right)-\mathcal{K}\left(\mathcal{A}|\in\mathcal{Z}_{L_S}\left(X_{t-1}\right)\right)\right|\leq\frac{\epsilon}{L_S}, \tag{2.197}$$

with probability at least $1-\pi_{t-1}(\epsilon, L_S)$, for all $t \in \mathbb{N}$ (in general), where, for each t, $\{\pi_{t-1}(\epsilon,n)\}_{n\in\mathbb{N}^+}$ constitutes a sequence vanishing at infinity. This is a considerably weaker form of conditional regularity, as stated in Definition 2. ■

2.2.4.3 Extensions: State Functionals & Approximate Prediction

All the results presented so far can be extended as follows. First, if $\left\{\boldsymbol{\phi}_t : \mathbb{R}^{M\times 1} \mapsto \mathbb{R}^{M_{\phi_t}\times 1}\right\}_{t\in\mathbb{N}}$ is a family of bounded and continuous functions, it is easy

to show that every relevant theorem presented so far is still true if one replaces X_t by $\boldsymbol{\phi}_t(X_t)$ in the respective formulations of the approximate filters discussed. This is made possible by observing that (2.18) still holds if we replace X_t by $\boldsymbol{\phi}_t(X_t)$, by invoking the Continuous Mapping Theorem and using the boundedness of $\boldsymbol{\phi}_t(X_t)$, instead of the boundedness of X_t, whenever required.

Second, exploiting very similar arguments as in the previous sections, it is possible to derive asymptotically optimal ρ-step state predictors, where $\rho > 0$ denotes the desired (and finite) prediction horizon. In particular, under the usual assumptions [12], it is easy to show that, as in the filtering case, the optimal nonlinear temporal predictor $\mathbb{E}_{\mathcal{P}}\left\{X_{t+\rho}\middle|\mathscr{Y}_t\right\}$ can be expressed through the Feynman-Kac type of formula

$$\mathbb{E}_{\mathcal{P}}\left\{X_{t+\rho}\middle|\mathscr{Y}_t\right\} \equiv \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{X_{t+\rho}\Lambda_t\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t\middle|\mathscr{Y}_t\right\}}, \quad \forall t \in \mathbb{N}. \tag{2.198}$$

Therefore, in analogy to (2.175), it is reasonable to consider grid based approximations of the form

$$\mathcal{E}^{L_S}\left(X_{t+\rho}\middle|\mathscr{Y}_t\right) \equiv \frac{\mathbf{X}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}^e_{L_S}\left(X^{L_S}_{t+\rho}\right)\Lambda^{L_S}_t\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda^{L_S}_t\middle|\mathscr{Y}_t\right\}}, \tag{2.199}$$

for all $t \in \mathbb{N}$. Focusing on marginal state quantizations (the Markovian case is similar, albeit easier), then, exploiting Lemma 2.15 and using induction, it is easy to show that

$$\mathcal{Q}^e_{L_S}\left(X^{L_S}_{t+\rho}\right) \equiv \boldsymbol{P}^{\rho}\mathcal{Q}^e_{L_S}\left(X^{L_S}_t\right) + \sum_{i=1}^{\rho}\boldsymbol{P}^{\rho-i}\boldsymbol{\mathcal{M}}^e_{t+i} + \sum_{i=1}^{\rho}\boldsymbol{P}^{\rho-i}\boldsymbol{\varepsilon}^{L_S}_{t+i}, \quad \forall t \in \mathbb{N}. \tag{2.200}$$

Thus, using simple properties of MD sequences, it follows that the numerator of the fraction on the RHS of (2.199) can be decomposed as

$$\begin{aligned}&\mathcal{E}^{L_S}\left(X_{t+\rho}\middle|\mathscr{Y}_t\right)\\ &\quad\equiv \frac{\mathbf{X}\boldsymbol{P}^{\rho}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}^e_{L_S}\left(X^{L_S}_{t+\rho}\right)\Lambda^{L_S}_t\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda^{L_S}_t\middle|\mathscr{Y}_t\right\}} + \frac{\mathbf{X}\sum_{i=1}^{\rho}\boldsymbol{P}^{\rho-i}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\boldsymbol{\varepsilon}^{L_S}_{t+i}\Lambda^{L_S}_t\middle|\mathscr{Y}_t\right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda^{L_S}_t\middle|\mathscr{Y}_t\right\}}.\end{aligned} \tag{2.201}$$

The first term on the RHS of (2.201) is analyzed exactly as in the proof of Lemma 2.16. For the second term, it is true that

$$\left\|\mathbf{X}\sum_{i=1}^{\rho}\boldsymbol{P}^{\rho-i}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\boldsymbol{\varepsilon}^{L_S}_{t+i}\Lambda^{L_S}_t\middle|\mathscr{Y}_t\right\}\right\|_1 \leq M\gamma\sqrt{\lambda_{inf}^{-N(t+1)}}\sum_{i=1}^{\rho}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left\|\boldsymbol{\varepsilon}^{L_S}_{t+i}\right\|_1\right\}, \tag{2.202}$$

which can be treated as an extra error term, also in the fashion of Lemma 2.16.

Putting it altogether (state functionals plus prediction), the following general theorem holds, covering every aspect of the investigation presented in this work.

Theorem 2.8. (Grid Based Filtering/Prediction & Functionals of the State) *For any deterministic functional family* $\left\{\boldsymbol{\phi}_t : \mathbb{R}^{M\times 1} \mapsto \mathbb{R}^{M_{\phi_t}\times 1}\right\}_{t\in\mathbb{N}}$ *with bounded and continuous members and any finite prediction horizon* $\rho \geq 0$*, the strictly optimal filter and* ρ*-step predictor of the* ***transformed process*** $\boldsymbol{\phi}_t(X_t)$ *can be approximated as*

$$\mathcal{E}^{L_S}\left(\boldsymbol{\phi}_{t+\rho}\left(X_{t+\rho}\right)\middle|\,\mathscr{Y}_t\right) \triangleq \boldsymbol{\Phi}_{t+\rho}\frac{\boldsymbol{P}^{\rho}\boldsymbol{E}_t}{\|\boldsymbol{E}_t\|_1} \in \mathbb{R}^{M_{\phi_t}\times 1}, \tag{2.203}$$

for all $t \in \mathbb{N}$*, where the process* $E_t \in \mathbb{R}^{L_S\times 1}$ *can be recursively evaluated as in Theorem 2.6,* $\boldsymbol{P}$ *is defined according to the chosen state quantization and*

$$\boldsymbol{\Phi}_{t+\rho} \triangleq \left[\boldsymbol{\phi}_{t+\rho}\left(\boldsymbol{x}_{L_S}^{1}\right) \ \ldots \ \boldsymbol{\phi}_{t+\rho}\left(\boldsymbol{x}_{L_S}^{L_S}\right)\right] \in \mathbb{R}^{M_{\phi_t}\times L_S}. \tag{2.204}$$

Additionally, under the appropriate assumptions (see Lemma 2.14 and Lemma 2.16, respectively) the approximate filter is asymptotically optimal. in the sense of Theorem 2.3.

Remark 2.19. As Theorem 2.8 clearly states, for each choice of state functionals and any finite prediction horizon, convergence of the respective approximate grid based filters is in the sense of Theorem 2.3. This implies the existence of an exceptional measurable set of measure almost unity, inside of which convergence is in the uniform sense. It is important to emphasize that this exceptional event, $\widehat{\Omega}_T$, as well as its measure, are *independent of the particular choice of both the bounded family* $\{\boldsymbol{\phi}_t\}_t$ *and the prediction horizon* ρ. This fact can be easily verified by a quick detour of the proof of Theorem 2.3. In particular, for any fixed choice of T, $\widehat{\Omega}_T$ characterizes exclusively the growth of the observations $\mathbf{y}_t$, which are the same regardless of filtering, prediction, or any functional imposed on the state. Therefore, *stochastically uniform (in* $\widehat{\Omega}_T$*) convergence of one estimator implies stochastically uniform convergence of any other estimator*, within any class of estimators, constructed employing any uniformly bounded and continuous class of functionals of the state and finite prediction horizons. ■

2.2.4.4 Filter Performance

The uncertainty of a filtering estimator can be quantified via its *posterior* quadratic deviation from the true state, at each time t. This information is encoded into the posterior covariance matrix

$$\mathbb{V}\{X_t|\,\mathscr{Y}_t\} \equiv \mathbb{E}\left\{X_t X_t^{\boldsymbol{T}}\middle|\,\mathscr{Y}_t\right\} - \mathbb{E}\{X_t|\,\mathscr{Y}_t\}(\mathbb{E}\{X_t|\,\mathscr{Y}_t\})^{\boldsymbol{T}}, \tag{2.205}$$

for all $t \in \mathbb{N}$. Next, in a general setting, we consider asymptotically consistent approximations of $\mathbb{V}\left\{\boldsymbol{\phi}_{t+\rho}\left(X_{t+\rho}\right)\middle|\,\mathscr{Y}_t\right\}$, which, at the same time, admit finite dimensional representations. In the following, $\|\cdot\|_1^{\mathsf{E}}$ denotes the *entrywise* ℓ_1-norm for matrices, which upper bounds both the ℓ_1-operator-induced and the Frobenius norms.

Theorem 2.9. (Posterior Covariance Recursions) *Under the same setting as in Theorem 2.8, the posterior covariance matrix of the optimal filter of the transformed process $\boldsymbol{\phi}_{t+\rho}\left(X_{t+\rho}\right)$ can be approximated as*

$$\mathcal{V}^{L_S}\left(\boldsymbol{\phi}_{t+\rho}\left(X_{t+\rho}\right)\middle|\,\mathscr{Y}_t\right) \triangleq \boldsymbol{\Phi}_{t+\rho}\left[\operatorname{diag}\left(\frac{\boldsymbol{P}^{\rho}\boldsymbol{E}_t}{\|\boldsymbol{E}_t\|_1}\right) - \frac{\boldsymbol{P}^{\rho}\boldsymbol{E}_t}{\|\boldsymbol{E}_t\|_1}\left(\frac{\boldsymbol{P}^{\rho}\boldsymbol{E}_t}{\|\boldsymbol{E}_t\|_1}\right)^{\boldsymbol{T}}\right]\boldsymbol{\Phi}_{t+\rho}^{\boldsymbol{T}}, \tag{2.206}$$

for all $t \in \mathbb{N}$. Under the appropriate assumptions (Lemma 2.14/2.16), the approximate estimator is asymptotically optimal in the sense of Theorem 2.3.

Proof of Theorem 2.9. See Section 2.2.7.4 (Appendix D). ■

2.2.5 Analytical Examples & Some Simulations

This section is centered around a discussion about the practical applicability of the grid based filters under consideration, mainly in regard to filter implementation, as well as the sufficient conditions for asymptotic optimality presented and analyzed in Section 2.2.4. In what follows, we consider a class of 1-dimensional (for simplicity), common and rather practically important *additive Nonlinear AutoRegressions (NARs)*, where X_t evolves according to the stochastic difference equation

$$X_t \equiv h\left(X_{t-1}\right) + W_t, \quad \forall t \in \mathbb{N}, \tag{2.207}$$

$X_{-1} \sim \mathcal{P}_{X_{-1}}$, where $h\left(\cdot\right)$ constitutes a uniformly bounded and at least continuous nonlinear functional and W_t is a white noise process with known measure. To ensure

that the state is bounded, we will assume that the white noise W_t follows, for each $t \in \mathbb{N}$, a zero location (and mean), *truncated Gaussian distribution* in $[-\alpha, \alpha]$, with scale σ and with density

$$f_W(x) \triangleq \frac{\varphi(x/\sigma)}{2\sigma\Phi(\alpha/\sigma) - \sigma} \mathbb{1}_{[-\alpha,\alpha]}(x), \ \forall x \in \mathbb{R}, \tag{2.208}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ denote the standard Gaussian density and cumulative distribution functions, respectively. Under these considerations, if $\sup_{x\in\mathbb{R}} |h(x)| \equiv B$, then $|X_t| \leq B + \alpha$ and, thus, $\mathcal{Z}$ is identified as the set $[a, b \equiv -a]$, with $b \triangleq B + \alpha$.

2.2.5.1 Markovian Filter

In this case, the respective approximation of the state process is given by the quantized stochastic difference equation

$$X_t^{L_S} \triangleq \mathcal{Q}_{L_S}\left(h\left(X_{t-1}^{L_S}\right) + W_t\right), \quad \forall t \in \mathbb{N}, \tag{2.209}$$

initialized as $X_{-1}^{L_S} \equiv \mathcal{Q}_{L_S}(X_{-1})$, with probability 1. In order to guarantee asymptotic optimality of the respective approximate filter described in Theorem 2.6, the original process X_t is required to at least satisfy the basic Lipschitz condition of Assumption 2.3. Indeed, *if we merely assume that* $h(\cdot)$ *is additionally Lipschitz* with constant $L_h > 0$ (that is, regardless of the stochastic character of W_t, in general), then the function

$$f(x, y) \triangleq h(x) + y, \quad (x, y) \in [-B, B] \times [-\alpha, \alpha] \tag{2.210}$$

is also Lipschitz with respect to x (for all y), with constant L_h as well. Therefore, under the mild Lipschitz assumption for $h(\cdot)$, we have shown that the resulting Markovian filter will indeed be asymptotically consistent. In practice, we expect that a smaller constant L_h would result in better performance of the approximate filter, with best results if $h(\cdot)$ constitutes a contraction, which makes $f(\cdot, y)$ uniformly contractive in y. The above is indeed true, since filtering is essentially implemented via a stochastic difference equation itself, and, in general, any discretized approximation to this difference equation is subject to error accumulation.

Of course, in order for the Markovian filter to be realizable, both the transition matrix $\boldsymbol{P}$ and the initial value $\boldsymbol{E}_{-1}$ have to be determined. In all cases, under our

assumptions, $\boldsymbol{P}$ (and obviously $\boldsymbol{E}_{-1}$) may be determined during an *offline training phase*, and stored in memory. A brute force way for estimating $\boldsymbol{P}$ is to simulate X_t (recall that the stochastic description of the transitions of X_t is known apriori). Then, $\boldsymbol{P}$ can be empirically estimated using the Strong Law of Large Numbers (SLLN). The aforementioned procedure results in excellent performance in practice (see Section 3.2). Exactly the same idea may be employed in order to estimate $\boldsymbol{E}_{-1}$, given the initial measure of X_t. Note that the above described empirical method for the estimation of $\boldsymbol{P}$ and $\boldsymbol{E}_{-1}$ does not assume a specific model describing the temporal evolution of X_t, or any particular choice of state quantization. Thus, it is generally applicable.

However, for the specific (though general) class of systems discussed above, we may also present an analytical construction for $\boldsymbol{P}$ (and $\boldsymbol{E}_{-1}$, assuming $\mathcal{P}_{X_{-1}}$ is known), resulting in compact, closed form expressions. Indeed, by definition of $X_t^{L_S}$, $\boldsymbol{P}(i,j)$ and each $\mathcal{Z}_{L_S}^i$, whose center is $\boldsymbol{x}_{L_S}^i$, we get

$$\begin{aligned}\boldsymbol{P}(i,j) &\equiv \mathcal{P}\left(h\left(X_{t-1}^{L_S}\right)+W_t \in \mathcal{Z}_{L_S}^i \,\middle|\, X_{t-1}^{L_S} \equiv \boldsymbol{x}_{L_S}^j\right) \\ &= \int_{\mathcal{Z}_{L_S}^i} f_W\Big(x - h\left(\boldsymbol{x}_{L_S}^j\right)\Big)\mathrm{d}x, \end{aligned} \tag{2.211}$$

which, based on (2.207), can be written in closed form as

$$\boldsymbol{P}(i,j) \equiv \frac{\Phi\left(\dfrac{p_{L_S}^{ij}(\alpha,B)}{\sigma}\right) - \Phi\left(\dfrac{q_{L_S}^{ij}(\alpha,B)}{\sigma}\right)}{2\sigma\Phi\left(\alpha/\sigma\right)-\sigma} \mathbb{1}_{(-\infty,p)}(q)\,, \tag{2.212}$$

for all $(i,j) \in \mathbb{N}_{L_S}^+ \times \mathbb{N}_{L_S}^+$, where

$$p_{L_S}^{ij}(\alpha,B) \triangleq \min\left\{\alpha, \boldsymbol{x}_{L_S}^i - h\left(\boldsymbol{x}_{L_S}^j\right) + \frac{B+\alpha}{L_S}\right\} \text{ and} \tag{2.213}$$

$$q_{L_S}^{ij}(\alpha,B) \triangleq \max\left\{-\alpha, \boldsymbol{x}_{L_S}^i - h\left(\boldsymbol{x}_{L_S}^j\right) - \frac{B+\alpha}{L_S}\right\}. \tag{2.214}$$

Consequently, via (2.212), one may obtain the whole matrix $\boldsymbol{P}$ for any set of parameters σ, α, B and for any resolution L_S. As far as the initial value $\boldsymbol{E}_{-1}$ is concerned, assuming that the initial measure of X_t, $\mathcal{P}_{X_{-1}}$, is known and recalling that the mapping $\mathcal{Q}_{L_S}^e(\cdot)$ is bijective, it will be true that

$$\boldsymbol{E}_{-1} \equiv \sum_{j \in \mathbb{N}_{L_S}^+} \mathbf{e}_j^{L_S} \mathcal{P}\left(X_{-1}^{L_S} \equiv \boldsymbol{x}_{L_S}^j\right), \tag{2.215}$$

where $\mathcal{P}\left(X_{-1}^{L_S} \equiv \boldsymbol{x}_{L_S}^{j}\right) \equiv \int_{\mathcal{Z}_{L_S}^{j}} \mathcal{P}_{X_{-1}}(\mathrm{d}x)$, for all $j \in \mathbb{N}_{L_S}^{+}$. Thus, $\boldsymbol{E}_{-1}$ can be evaluated in closed form, as long as the aforementioned integrals can be analytically computed.

2.2.5.2 Marginal Filter

Marginal filters are, in general, slightly more complicated. However, at least in theory, they are provably more powerful than Markovian filters, as the following result suggests.

Theorem 2.10. (Additive NARs are *Almost* CRT II) *Let $X_t \in \mathbb{R}$ evolve as in (2.207), with $X_{-1} \sim \mathcal{P}_{X_{-1}}$, and where*

- *$h(\cdot)$ is continuous and uniformly bounded by $B > 0$.*
- *W_t follows the truncated Gaussian law in $[-\alpha, \alpha]$, $\alpha > 0$, with scale zero and location $\sigma > 0$.*

*Then, for any quantizer $\mathcal{Q}_{L_S}(\cdot)$ and any initial measure $\mathcal{P}_{X_{-1}}$, X_t is **almost** conditionally regular, in the sense that*

$$\operatorname*{ess\,sup}_{y \in \mathbb{R}} \left| \kappa(y|\, x) - \kappa_t\big(y| \in \mathcal{Z}_{L_S}(x)\big) \right| \leq \delta_{L_S}^{II}(x) + f_W(\alpha), \tag{2.216}$$

for some uniformly bounded, time invariant, nonnegative sequence $\left\{\delta_n^{II}(\cdot)\right\}_{n \in \mathbb{N}^+}$, converging to zero $\mathcal{P}_{X_t}$-almost everywhere, for all $t \in \{-1\} \cup \mathbb{N}$.

Proof of Theorem 2.10. See Section 2.2.7.5 (Appendix E). ■

As Theorem 2.10 suggests, regardless of the respective initial measures and without any additional assumptions on the nature of $h(\cdot)$, except for continuity, the truncated Gaussian NARs under consideration are *almost* conditionally regular of type II, in the sense that the relevant condition on the respective stochastic kernel is modified by adding *the drift* $f_W(\alpha)$. In general, this drift parameter might cause error accumulation during the implementation of the marginal filter. On the other hand though, it is true that for any fixed scale parameter σ, $f_W(\alpha) \equiv \mathcal{O}\Big(\exp\left(-\alpha^2\right)\Big)$. Thus, for sufficiently large α, $f_W(\alpha)$ will not essentially affect filter performance.

Nevertheless, technically, this drift error can vanish, if one considers a white noise W_t following a distribution admitting a finitely supported and essentially Lipschitz in $[-\alpha, \alpha]$ density, *taking zero values at* $\pm\alpha$. This is possible by observing that the proof to Theorem 2.10 in fact works for such densities, without significant modifications. Then, $f_W(\pm\alpha) \equiv 0$ and, hence, the resulting NAR will be CRT II. Such densities exist and are, in fact, popular; examples are the Logit-Normal and the Raised Cosine densities, which constitute nice truncated approximations to the Gaussian density, or more interesting choices, such as the Beta and Kumarasawmy densities.

Regarding the implementation of the marginal filter, unlike the Markovian case, closed forms for the elements of $\boldsymbol{P}_{(t)}$ are very difficult to obtain, because they explicitly depend on the marginal measures of X_t, for each t, as (2.182) suggests. Even if $\mathcal{P}_{X_{-1}}$ is an invariant measure, implying that the transition matrix is time invariant, the closed form determination of $\boldsymbol{P}(i,j)$ requires proper choice of $\mathcal{P}_{X_{-1}}$, which, in most cases, cannot be made by the user. Therefore, in most cases, $\boldsymbol{P}_{(t)}$ has to be computed via, for instance, simulation, and employing the SLLN. As restated above, this simple technique gives excellent empirical results. Also, assuming knowledge of the initial measure $\mathcal{P}_{X_{-1}}$, $\boldsymbol{E}_{-1}$ is again given by (2.215).

In order to demonstrate the applicability of the marginal filter, as well as empirically evaluate the training-by-simulation technique advocated above, below we present some additional experimental results (note that the following also holds for the Markovian filter, under the appropriate assumptions). As we shall see, these results will also confirm some aspects of the particular mode of convergence advocated in Theorem 2.3. Specifically, consider an additive NAR of the form discussed above, where $h(x) \equiv \tanh(1.3x) \in (-1,1)$, that is, $B \equiv 1$, and where $\alpha \equiv 1$ and $\sigma \equiv 0.3$. Additionally, the resulting state process X_t is observed via the nonlinear functional $\mathbf{y}_t \equiv [X_t]^3 \mathbf{1}_N + \boldsymbol{w}_t$ ($\mathbf{1}_n$ being the n-by-1 all-ones vector), where $\boldsymbol{w}_t \overset{i.i.d}{\sim} \mathcal{N}\left(0, \sigma_{\boldsymbol{w}}^2 \mathbf{I}_N\right)$, $\sigma_{\boldsymbol{w}}^2 \equiv 2$, for all $t \in \mathbb{N}$. In order to stress test the marginal approximation approach, we set $\mathcal{P}_{X_{-1}} \equiv \mathcal{U}[-2,2]$ and we arbitrarily assume stationarity of X_t, regardless of $\mathcal{P}_{X_{-1}}$ being an invariant measure or not. This is a common tactic in practice. Under this setting, $\boldsymbol{E}_{-1} \equiv L_S^{-1} \mathbf{1}_{L_S}$, whereas a *single* $\boldsymbol{P}$ is estimated *offline* from $3 \cdot 10^5$ samples of a *single* simulated version of X_t.

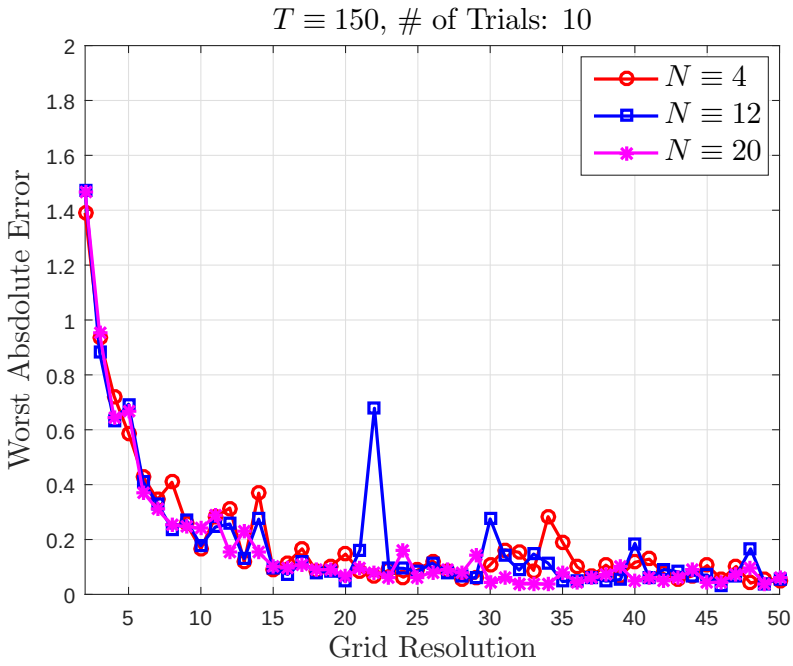


Figure 2.3: Marginal filter: worst error with respect to filter resolution (L_S), over 10 trials and for different values of N.

As Theorem 2.3 suggests, one should be interested in the approximation error between the approximate and exact filters of X_t. However, the exact nonlinear filter of X_t is impossible to compute in a reasonable manner; besides, this is the motive for developing approximate filters. For that reason, we will further approximate the approximation error by replacing the optimal filter of X_t by a *particle filter* (an also approximate *global* method), but employing a very high number of particles. The resampling step of the particle filter is implemented using *systematic resampling*, known to minimize Monte Carlo (MC) variation [23]. In our simulations, 5000 particles are employed in each filtering iteration.

In the above fashion, Fig. 2.3 shows, for each filtering resolution L_S, ranging from 2 to 50, the *worst* absolute approximation error, chosen amongst 10 realizations (MC trials) of the (approximate) filtering process, where the filtering horizon was chosen as $T \equiv 150$ time steps. The error process depicted in Fig. 2.3 provides a good approximation to the exact uniform approximation error of (2.164) in Theorem 2.3.

From the figure, we observe that convergence of the worst approximation error is confirmed; for all values of N, a clear *strictly decreasing error trend* is identified, as

L_S increases. This roughly justifies Theorem 2.3. What is more, at least for the 10 realizations collected for each combination of N and L_S, the decay of the approximation error is *superstable, for all values of* N. This indicates that, in practice, the realizations of the approximate filtering process, which will ever be observed by the user, will be such that convergence to the optimal filter is indeed uniform, and *almost monotonic* (the "outliers" present at $L_S \equiv 22, 30, 34$ are most probably due to the use of a particle filter -a randomized estimator- for emulating the true filter of X_t). In the language of Theorem 2.3, it will "always" be the case that $\omega \in \widehat{\Omega}_T$ (an event occurring with high probability). This in turn implies that, although general, Theorem 2.3 might be somewhat looser than reality for "good" hidden model setups. Finally, another practically significant detail, which is revealed via Fig. 2.3, and seems to be a common feature of grid based methods, is that the uniform error bound of the approximate filters *does not increase as a function of* N. Note that this fact cannot be verified via Theorem 2.3.

2.2.6 Conclusion

We have presented a comprehensive treatment of grid based approximate nonlinear filtering of discrete time Markov processes observed in conditionally Gaussian noise, relying on Markovian and marginal approximations of the state. For the Markovian case, it has been shown that the resulting approximate filter is strongly asymptotically optimal as long as the transition mapping of the state is Lipschitz. For the marginal case, the novel concept of conditional regularity was proposed as a sufficient condition for ensuring asymptotic optimality. Conditional regularity is proven to be potentially more relaxed, compared to the state of the art in grid based filtering, revealing the potential strength of the grid based approach, and also justifying its good performance in applications. For both state approximation cases, convergence to the optimal filter has been proven to be in a strong sense, i.e., compact in time and uniform in a fully characterized event occurring almost certainly. Additionally, typical but important extensions of our results were discussed and justified. The whole theoretical development

was based on a novel methodological scheme, especially for marginal state approximations. This focused more on the use of linear-algebraic techniques and less on measure theoretic arguments, making the presentation more tangible and easier to grasp. In our companion work (see Section 3.2), the results presented herein have been successfully exploited, providing theoretical guarantees in the context of channel estimation in mobile wireless sensor networks.

2.2.7 Appendices

2.2.7.1 Appendix A: Proof of Lemma 2.14

Consider the event $\mathcal{E} \triangleq \{\omega \in \Omega \,|X_t(\omega) \in \mathcal{Z},\ \forall t \in \mathbb{N}\}$ of unity probability measure, that is, with $\mathcal{P}(\mathcal{E}) \equiv 1$. Of course, by our assumptions so far, $\mathcal{P}(\mathcal{E}^c) \equiv 0$, with

$$\mathcal{E}^c \triangleq \{\omega \in \Omega \,|X_t(\omega) \notin \mathcal{Z},\ \text{for some } t \in \mathbb{N}\} \tag{2.217}$$

being an "impossible" measurable set. Then, for $\omega \in \mathcal{E}$, we have $X_t(\omega) \in \mathcal{Z}$ for all $t \in \mathbb{N}$ and we may rewrite (2.169) as

$$X_t^{L_S}(\omega) = f\left(X_{t-1}^{L_S}(\omega), W_t(\omega)\right) + \varepsilon_t^{L_S}(\omega), \tag{2.218}$$

for some bounded process $\varepsilon_t^{L_S}(\omega)$. By Assumption 2.3,

$$\left\|X_t^{L_S}(\omega) - X_t(\omega)\right\|_1 \leq K(W_t(\omega))\left\|X_{t-1}^{L_S}(\omega) - X_{t-1}(\omega)\right\|_1 + \left\|\varepsilon_t^{L_S}(\omega)\right\|_1, \tag{2.219}$$

for all $t \in \mathbb{N}$. By construction of the quantizer $\mathcal{Q}_{L_S}(\cdot)$, it is easy to show that, for all $\omega \in \mathcal{E}$, $\left\|\varepsilon_t^{L_S}(\omega)\right\|_1 \leq M|b-a|/2L_S$, for all $t \in \mathbb{N}$. Then, iterating the right hand side of (2.219) and using induction, it can be easily shown that

$$\begin{aligned}&\left\|X_t^{L_S}(\omega) - X_t(\omega)\right\|_1 \\ &\leq \left(\prod_{i=0}^{t} K(W_i(\omega))\right)\left\|X_{-1}^{L_S}(\omega) - X_{-1}(\omega)\right\|_1 + \frac{M|b-a|}{2L_S}\left(1 + \sum_{j=1}^{t}\prod_{i=j}^{t} K(W_i(\omega))\right),\end{aligned} \tag{2.220}$$

where $X_{-1}^{L_S}(\omega)$ and $X_{-1}(\omega)$ constitute the initial values of the processes $X_t^{L_S}(\omega)$ and $X_t(\omega)$, respectively. Let us focus on the second term on the RHS of (2.220). Since,

by assumption, the respective Lipschitz constants are bounded with respect to the supremum norm in $\mathcal{E}$ and for all $t \in \mathbb{N}$, it holds that

$$\sum_{j=1}^{t}\prod_{i=j}^{t} K\left(W_i\left(\omega\right)\right) \leq \sup_{\omega\in\mathcal{E}} \sum_{j=1}^{t}\prod_{i=j}^{t} K\left(W_i\left(\omega\right)\right) \triangleq \sum_{j=1}^{t}\prod_{i=j}^{t} K\left(W_i^*\right). \tag{2.221}$$

Note, however, that the supremum of (2.221) in $t \in \mathbb{N}$ indeed might not be finite. Likewise, regarding the first term on the RHS of (2.220), we have

$$\prod_{i=0}^{t} K(W_i\left(\omega\right)) \leq \sup_{\omega\in\mathcal{E}} \prod_{i=0}^{t} K(W_i\left(\omega\right)) \triangleq \prod_{i=0}^{t} K\big(W_i^\star\big). \tag{2.222}$$

As a result, assuming only Lipschitz continuity of $f\left(\cdot,\cdot\right)$ and recalling that $X_{-1}^{L_S} \equiv \mathcal{Q}_{L_S}\left(X_{-1}\right)$, taking the supremum on both sides (2.220) yields

$$\begin{aligned}\sup_{\omega\in\mathcal{E}} \left\| X_t^{L_S}\left(\omega\right) - X_t\left(\omega\right)\right\|_1 &\equiv \operatorname*{ess\,sup}_{\omega\in\Omega} \left\| X_t^{L_S}\left(\omega\right) - X_t\left(\omega\right)\right\|_1 \\ &\leq \frac{M\left|b-a\right|}{2L_S}\left(1+\sum_{j=1}^{t}\prod_{i=j}^{t} K\big(W_i^*\big)+\prod_{i=0}^{t} K\big(W_i^\star\big)\right) \underset{L_S\to\infty}{\longrightarrow} 0, \end{aligned} \tag{2.223}$$

where the convergence rate may depend on each finite t, therefore only guaranteeing convergence of $X_t^{L_S}\left(\omega\right)$ in the pointwise sense in t and uniformly almost everywhere in ω. Now, if $f(\cdot,\cdot)$ is uniformly contractive for all $\omega \in \mathcal{E}$, and for all $t \in \mathbb{N}$, then it will be true that $K\left(W_t\left(\omega\right)\right) \in [0,1)$, surely in $\mathcal{E}$ and everywhere in time as well. Consequently, focusing on the second term on the RHS of (2.220), it should be true that

$$\begin{aligned}1+\sum_{j=1}^{t}\prod_{i=j}^{t} K\left(W_i\left(\omega\right)\right) &\leq 1+\sum_{j=1}^{t}\prod_{i=j}^{t} \sup_{l\in\mathbb{N}} \sup_{\omega\in\mathcal{E}} K\left(W_l\left(\omega\right)\right) \\ &\triangleq 1+\sum_{j=1}^{t}\prod_{i=j}^{t} K_* \equiv \sum_{j=0}^{t} K_*^j = \frac{1-K_*^{t+1}}{1-K_*} \leq \frac{1}{1-K_*}, \end{aligned} \tag{2.224}$$

where $K_* \in [0,1)$ constitutes a global "Lipschitz constant" for $f\left(\cdot,\cdot\right)$ in $\mathcal{E}$ and for all $t \in \mathbb{N}$. The situation is of course similar for the simpler first term on the RHS of (2.220). As a result, we readily get that

$$\sup_{t\in\mathbb{N}} \operatorname*{ess\,sup}_{\omega\in\Omega} \left\| X_t^{L_S}(\omega) - X_t(\omega)\right\|_1 \leq \frac{M\left|b-a\right|}{2L_S}\frac{2-K_*}{1-K_*}, \tag{2.225}$$

where the RHS vanishes as $L_S \to \infty$, thus proving the second part of the lemma. ■

2.2.7.2 Appendix B: Proof of Lemma 2.15

Since the mapping $\mathcal{Q}^e_{L_S}(\cdot)$ is bijective and using the Markov property of X_t, it is true that

$$\begin{aligned}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)\middle|\,\mathscr{X}_{t-1}\right\} &\equiv \mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)\middle|\,X_{t-1}\right\}\\ &\equiv \sum_{j\in\mathbb{N}^+_{L_S}} \mathbf{e}_j^{L_S}\widetilde{\mathcal{P}}\left(X_t\in\mathcal{Z}^j_{L_S}\middle|\,X_{t-1}\right).\end{aligned} \tag{2.226}$$

First, let us consider the case where X_t is CRT II. Then, assuming the existence of a stochastic kernel density, it follows that there is a nonnegative sequence $\left\{\delta^{II}_{L_S}(\cdot)\right\}_{L_S\in\mathbb{N}^+}$ converging almost everywhere to 0 as $L_S\to\infty$, such that, for all $\boldsymbol{y}\in\mathbb{R}^{M\times 1}$, $\kappa(\boldsymbol{y}|\,\boldsymbol{x})\leq \delta^{II}_{L_S}(\boldsymbol{x})+\kappa\left(\boldsymbol{y}|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right)$. Thus, for each particular choice of $(\boldsymbol{y},\boldsymbol{x})$, there exists a process $\varepsilon_{L_S}(\boldsymbol{y},\boldsymbol{x})\in\left[-\delta^{II}_{L_S}(\boldsymbol{x}),\delta^{II}_{L_S}(\boldsymbol{x})\right]$, such that

$$\kappa(\boldsymbol{y}|\,\boldsymbol{x})\equiv\varepsilon_{L_S}(\boldsymbol{y},\boldsymbol{x})+\kappa\left(\boldsymbol{y}|\in\mathcal{Z}_{L_S}(\boldsymbol{x})\right). \tag{2.227}$$

Consequently, (2.226) can be expressed as

$$\begin{aligned}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)\middle|\,\mathscr{X}_{t-1}\right\} &= \sum_{j\in\mathbb{N}^+_{L_S}}\mathbf{e}_j^{L_S}\int_{\mathcal{Z}^j_{L_S}}\kappa(\boldsymbol{x}_t|\,X_{t-1}(\omega))\,\mathrm{d}\boldsymbol{x}_t\\ &= \sum_{j\in\mathbb{N}^+_{L_S}}\mathbf{e}_j^{L_S}\int_{\mathcal{Z}^j_{L_S}}\kappa\left(\boldsymbol{x}_t|\in\mathcal{Z}_{L_S}(X_{t-1}(\omega))\right)\mathrm{d}\boldsymbol{x}_t+\boldsymbol{\varepsilon}_t^{L_S}\\ &\equiv \sum_{j\in\mathbb{N}^+_{L_S}}\mathbf{e}_j^{L_S}\widetilde{\mathcal{P}}\left(X_t^{L_S}\equiv\boldsymbol{x}^j_{L_S}\middle|\,X_{t-1}^{L_S}\right)+\boldsymbol{\varepsilon}_t^{L_S},\end{aligned} \tag{2.228}$$

where the $\{\mathscr{X}_t\}$-predictable error process $\boldsymbol{\varepsilon}_t^{L_S}\in\mathbb{R}^{L_S\times 1}$ is defined as

$$\boldsymbol{\varepsilon}_t^{L_S}\triangleq\left[\left\{\int_{\mathcal{Z}^j_{L_S}}\varepsilon_{L_S}(\boldsymbol{x}_t,X_{t-1})\,\mathrm{d}\boldsymbol{x}_t\right\}_{j\in\mathbb{N}^+_{L_S}}\right]^{\boldsymbol{T}}. \tag{2.229}$$

Then, since the state space of $\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)$ is finite with cardinality L_S, we can write

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)\middle|\,\mathscr{X}_{t-1}\right\}\equiv\boldsymbol{P}\mathcal{Q}^e_{L_S}\left(X_{t-1}^{L_S}\right)+\boldsymbol{\varepsilon}_t^{L_S}, \tag{2.230}$$

or, equivalently,

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}^e_{L_S}\left(X_t^{L_S}\right)-\boldsymbol{P}\mathcal{Q}^e_{L_S}\left(X_{t-1}^{L_S}\right)-\boldsymbol{\varepsilon}_t^{L_S}\middle|\,\mathscr{X}_{t-1}\right\}\triangleq\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\boldsymbol{\mathcal{M}}^e_t|\,\mathscr{X}_{t-1}\right\}\equiv 0. \tag{2.231}$$

As far as the quantity $\left\|\boldsymbol{\varepsilon}_t^{L_S}\right\|_1$ is concerned, it is true that

$$\begin{aligned}\left\|\boldsymbol{\varepsilon}_t^{L_S}\right\|_1 &\leq \sum_{j\in\mathbb{N}_{L_S}^+}\int_{\mathcal{Z}_{L_S}^j}\left|\varepsilon_{L_S}\left(\boldsymbol{x}_t, X_{t-1}\right)\right|\mathrm{d}\boldsymbol{x}_t\\ &\leq \sum_{j\in\mathbb{N}_{L_S}^+}\int_{\mathcal{Z}_{L_S}^j}\delta_{L_S}^{II}\left(X_{t-1}\right)\mathrm{d}\boldsymbol{x}_t\\ &= |b-a|^M\,\delta_{L_S}^{II}\left(X_{t-1}\right)\underset{L_S\to\infty}{\longrightarrow}0,\quad \widetilde{\mathcal{P}}-a.s.\end{aligned}\tag{2.232}$$

and for all $t\in\mathbb{N}$.

For the case where X_t constitutes a CRT I process, the situation is similar. Specifically, (2.226) can be expressed as

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}_{L_S}^e\left(X_t^{L_S}\right)\middle|\,\mathscr{X}_{t-1}\right\}\equiv\sum_{j\in\mathbb{N}_{L_S}^+}\mathbf{e}_j^{L_S}\widetilde{\mathcal{P}}\left(X_t^{L_S}\equiv\boldsymbol{x}_{L_S}^j\middle|\,X_{t-1}^{L_S}\right)+\boldsymbol{\varepsilon}_t^{L_S},\tag{2.233}$$

where the process $\boldsymbol{\varepsilon}_t^{L_S}\in\mathbb{R}^{L_S\times 1}$ is defined similarly to the previous case as

$$\boldsymbol{\varepsilon}_t^{L_S}\triangleq\left[\varepsilon_{L_S}\left(\mathcal{Z}_{L_S}^1, X_{t-1}\right)\ \ldots\ \varepsilon_{L_S}\left(\mathcal{Z}_{L_S}^{L_S}, X_{t-1}\right)\right]^{\boldsymbol{T}},\tag{2.234}$$

with

$$\begin{aligned}\left\|\boldsymbol{\varepsilon}_t^{L_S}\right\|_1 &\equiv \sum_{j\in\mathbb{N}_{L_S}^+}\left|\varepsilon_{L_S}\left(\mathcal{Z}_{L_S}^j, X_{t-1}\right)\right|\\ &\leq \sum_{j\in\mathbb{N}_{L_S}^+}\frac{\delta_{L_S}^I\left(X_{t-1}\right)}{L_S}\equiv\delta_{L_S}^I\left(X_{t-1}\right)\underset{L_S\to\infty}{\longrightarrow}0,\end{aligned}\tag{2.235}$$

$\widetilde{\mathcal{P}}-a.s.$ and for all $t\in\mathbb{N}$. The proof is complete. ■

2.2.7.3 Appendix C: Proof of Lemma 2.16

Let us first recall some identifications. First, it can be easily shown that

$$\begin{aligned}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t^{L_S}\middle|\,\mathscr{Y}_t\right\} &\equiv \left\|\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathcal{Q}_{L_S}^e\left(X_t^{L_S}\right)\Lambda_t^{L_S}\middle|\,\mathscr{Y}_t\right\}\right\|_1\\ &\triangleq \left\|\boldsymbol{E}_t^X\right\|_1,\quad\text{and}\end{aligned}\tag{2.236}$$

$$\begin{aligned}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t^{Z,L_S}\middle|\,\mathscr{Y}_t\right\} &\equiv \left\|\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{Z_t^{L_S}\Lambda_t^{Z,L_S}\middle|\,\mathscr{Y}_t\right\}\right\|_1\\ &\triangleq \left\|\boldsymbol{E}_t^Z\right\|_1.\end{aligned}\tag{2.237}$$

Then, we can write[5]

$$\left\|\mathcal{E}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)-\widetilde{\mathcal{E}}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)\right\|_1 \leq \|\mathbf{X}\|_1 \frac{\left|\left\|\boldsymbol{E}_t^Z\right\|_1-\left\|\boldsymbol{E}_t^X\right\|_1\right|+\left\|\boldsymbol{E}_t^X-\boldsymbol{E}_t^Z\right\|_1}{\left\|\boldsymbol{E}_t^X\right\|_1}. \quad (2.238)$$

Since $\|\mathbf{X}\|_1 \equiv M\max\{|a|,|b|\} \triangleq M\gamma$ and using the reverse triangle inequality, we get

$$\left\|\mathcal{E}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)-\widetilde{\mathcal{E}}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)\right\|_1 \leq 2M\gamma\frac{\left\|\boldsymbol{E}_t^X-\boldsymbol{E}_t^Z\right\|_1}{\left\|\boldsymbol{E}_t^X\right\|_1}. \quad (2.239)$$

Since $Z_t^{L_S}$ is a Markov chain, it can be readily shown that $\boldsymbol{E}_t^Z$ satisfies the linear recursion $\boldsymbol{E}_t^Z = \boldsymbol{\Lambda}_t\boldsymbol{P}\boldsymbol{E}_{t-1}^Z$, for all $t \in \mathbb{N}$ (also see Theorem 2.6). Similarly, using the martingale difference type representation given in Lemma 2.15, it easy to show that $\boldsymbol{E}_t^X$ satisfies another recursion of the form

$$\boldsymbol{E}_t^X = \boldsymbol{\Lambda}_t\boldsymbol{P}\boldsymbol{E}_{t-1}^X + \boldsymbol{\Lambda}_t\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\boldsymbol{\varepsilon}_t^{L_S}\Lambda_{t-1}^{L_S}\middle|\,\mathscr{Y}_{t-1}\right\},\ \forall t\in\mathbb{N}. \quad (2.240)$$

Then, by induction, the error process $\boldsymbol{E}_t^Z - \boldsymbol{E}_t^X$ satisfies

$$\begin{aligned}\boldsymbol{E}_t^Z - \boldsymbol{E}_t^X = &\left(\prod_{i\in\mathbb{N}_t}\left(\boldsymbol{\Lambda}_{t-i}\boldsymbol{P}\right)\right)\left(\boldsymbol{E}_{-1}^Z-\boldsymbol{E}_{-1}^X\right)\\ &-\sum_{j\in\mathbb{N}_t}\left(\prod_{i=0}^{j-1}\left(\boldsymbol{\Lambda}_{t-i}\boldsymbol{P}\right)\right)\boldsymbol{\Lambda}_{t-j}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\boldsymbol{\varepsilon}_{t-j}^{L_S}\Lambda_{t-j-1}^{L_S}\middle|\,\mathscr{Y}_{t-j-1}\right\}, \quad (2.241)\end{aligned}$$

for all $t \in \mathbb{N}$. Setting

$$\boldsymbol{E}_{-1}^Z \equiv \mathbb{E}_{\widetilde{\mathcal{P}}}\left\{Z_{-1}^{L_S}\right\} \equiv \mathbb{E}_{\mathcal{P}}\left\{\mathcal{Q}_{L_S}^e\left(X_{-1}^{L_S}\right)\right\} \equiv \boldsymbol{E}_{-1}^X \quad (2.242)$$

and taking the ℓ_1-norm of $\boldsymbol{E}_t^Z - \boldsymbol{E}_t^X$, it is true that

$$\begin{aligned}\left\|\boldsymbol{E}_t^Z-\boldsymbol{E}_t^X\right\|_1 &\leq \sum_{\tau\in\mathbb{N}_t}\left(\prod_{i=0}^{t-\tau-1}\left\|\widehat{\boldsymbol{\Lambda}}_{t-i}\right\|_1\right)\left\|\widehat{\boldsymbol{\Lambda}}_\tau\right\|_1\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\widehat{\Lambda}_{\tau-1}^{L_S}\left\|\boldsymbol{\varepsilon}_\tau^{L_S}\right\|_1\middle|\,\mathscr{Y}_{\tau-1}\right\}\\ &\leq \sum_{\tau\in\mathbb{N}_t}\sqrt{\lambda_{inf}^{-N(t-\tau+1)}}\sqrt{\lambda_{inf}^{-N\tau}}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left\|\boldsymbol{\varepsilon}_\tau^{L_S}\right\|_1\middle|\,\mathscr{Y}_{\tau-1}\right\}\\ &\equiv \sqrt{\lambda_{inf}^{-N(t+1)}}\sum_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\left\|\boldsymbol{\varepsilon}_\tau^{L_S}\right\|_1\right\},\quad \widetilde{\mathcal{P}}-a.e., \quad (2.243)\end{aligned}$$

[5]Here, $\|\mathbf{A}\|_1$ denotes the operator norm induced by the ℓ_1-vector norm.

since the process $\boldsymbol{\varepsilon}_t^{L_S}$ is $\{\mathscr{X}_t\}$-predictable and, under $\widetilde{\mathcal{P}}$, the processes X_t and $\mathbf{y}_t$ are statistically independent.

Now, assuming, for example, that X_t is CRT II (the case where X_t is CRT I is similar), we get

$$\begin{aligned}\left\|\boldsymbol{E}_t^Z-\boldsymbol{E}_t^X\right\|_1 &\leq \sqrt{\lambda_{inf}^{-N(t+1)}}\,|b-a|^M \sum_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\delta_{L_S}^{II}(X_{\tau-1})\right\}\\ &\leq \frac{|b-a|^M}{N\log\left(\lambda_{inf}\right)}\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\delta_{L_S}^{II}(X_{\tau-1})\right\},\end{aligned} \tag{2.244}$$

$\widetilde{\mathcal{P}}-a.e.$ (and, since $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ are equivalent, $\mathcal{P}-a.e.$ as well) and for all $t\in\mathbb{N}$. Regarding the denominator on the RHS of (2.239), in the last part of the proof of Theorem 2.3, the authors have shown that, in general, for any fixed $T<\infty$,

$$\inf_{t\in\mathbb{N}_T}\inf_{\omega\in\widehat{\Omega}_T}\inf_{L_S\in\mathbb{N}}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\Lambda_t^{L_S}\,\middle|\,\mathscr{Y}_t\right\}(\omega)>0, \tag{2.245}$$

where $\widehat{\Omega}_T\subseteq\Omega$ constitutes exactly the same measurable set of Theorem 2.4, occurring with $\mathcal{P}$-probability at least $1-\,(T+1)^{1-CN}\exp\left(-CN\right)$, for $C\geq 1$. Thus, (2.239) becomes

$$\left\|\mathcal{E}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)-\widetilde{\mathcal{E}}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)\right\|_1 \leq \frac{2M\gamma|b-a|^M\sup_{\tau\in\mathbb{N}_t}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\delta_{L_S}^{II}\left(X_{\tau-1}\right)\right\}}{N\log\left(\lambda_{inf}\right)\inf_{L_S\in\mathbb{N}}\left\|\boldsymbol{E}_t^X\right\|_1}, \tag{2.246}$$

$\mathcal{P}-a.e.$, and taking the supremum both with respect to $\omega\in\widehat{\Omega}_T$ and $t\in\mathbb{N}_T$ on both sides, we get

$$\begin{aligned}&\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left\|\mathcal{E}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)-\widetilde{\mathcal{E}}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)\right\|_1(\omega)\\ &\qquad\leq \frac{2M\gamma\,|b-a|^M\sup_{\tau\in\mathbb{N}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\delta_{L_S}^{II}\left(X_{\tau-1}\right)\right\}}{N\log\left(\lambda_{inf}\right)\inf_{t\in\mathbb{N}_T}\inf_{\omega\in\widehat{\Omega}_T}\inf_{L_S\in\mathbb{N}}\left\|\boldsymbol{E}_t^X\right\|_1(\omega)}.\end{aligned} \tag{2.247}$$

Since the sequence $\left\{\delta_{L_S}^{II}(\cdot)\right\}_{L_S}$ is $\mathcal{P}_{X_{-1}}\equiv\mathcal{P}_{X_t}-UI$, it is trivial that the sequence $\left\{\delta_{L_S}^{II}\left(X_{t-1}(\cdot)\right)\right\}_{L_S}$ is uniformly integrable, for all $t\in\mathbb{N}_T$. Then, because $\delta_{L_S}^{II}\left(X_{t-1}(\cdot)\right)\xrightarrow[L_S\to\infty]{a.e.}0$ (with respect to $\widetilde{\mathcal{P}}$), Vitali's Convergence Theorem implies that $\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\delta_{L_S}^{II}\left(X_{t-1}\right)\right\}\underset{L_S\to\infty}{\longrightarrow}0$, for all $t\in\mathbb{N}_T$, which in turn implies that $\sup_{\tau\in\mathbb{N}_T}\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\delta_{L_S}^{II}\left(X_{\tau-1}\right)\right\}\underset{L_S\to\infty}{\longrightarrow}0$. Thus, the RHS of (2.247) converges, and so does its LHS as well. ■

2.2.7.4 Appendix D: Proof of Theorem 2.9

For simplicity and clarity in the exposition, we consider the standard case where $\boldsymbol{\phi}_t(X_t) \equiv X_t$, for all $t \in \mathbb{N}$ and $\rho \equiv 1$. Starting with the definitions, since $\mathbb{V}\{X_t|\,\mathscr{Y}_t\}$ is given by (2.54), for all $t \in \mathbb{N}$, it is reasonable to define the grid based "filter"

$$\mathcal{V}^{L_S}(X_t|\,\mathscr{Y}_t) \triangleq \mathcal{E}^{L_S}\Big(X_t X_t^{\boldsymbol{T}}\Big|\,\mathscr{Y}_t\Big) - \mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t)\Big(\mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t)\Big)^{\boldsymbol{T}}, \tag{2.248}$$

for all $t \in \mathbb{N}$, where $\mathcal{E}^{L_S}\Big(X_t X_t^{\boldsymbol{T}}\Big|\,\mathscr{Y}_t\Big)$ constitutes an *entrywise* operator on the matrix $X_t X_t^{\boldsymbol{T}} \in \mathbb{R}^{M\times M}$, defined as

$$\mathcal{E}^{L_S}\Big(X_t X_t^{\boldsymbol{T}}\Big|\,\mathscr{Y}_t\Big)(i,j) \triangleq \frac{1}{\|\boldsymbol{E}_t\|_1}\sum_{l\in\mathbb{N}_{L_S}^+} \boldsymbol{x}_{L_S}^l(i)\,\boldsymbol{x}_{L_S}^l(j)\,\boldsymbol{E}_t(l) \tag{2.249}$$

$$\triangleq \frac{1}{\|\boldsymbol{E}_t\|_1}\sum_{l\in\mathbb{N}_{L_S}^+} \boldsymbol{\phi}^{ij}\Big(\boldsymbol{x}_{L_S}^l\Big)\,\boldsymbol{E}_t(l) \triangleq \boldsymbol{\Phi}^{ij}\frac{\boldsymbol{E}_t}{\|\boldsymbol{E}_t\|_1}, \tag{2.250}$$

for all $(i,j) \in \mathbb{N}_M^+ \times \mathbb{N}_M^+$. In the above, the function(al) $\boldsymbol{\phi}^{ij} : \mathbb{R}^{L_S\times 1} \to \mathbb{R}$ is obviously bounded as continuous. Then, making use of the triangle inequality, the entrywise ℓ_1-norm of $\mathcal{V}^{L_S}\{X_t|\,\mathscr{Y}_t\} - \mathbb{V}\{X_t|\,\mathscr{Y}_t\}$ may be bounded from above by the sum of the entrywise ℓ_1-norms of the differences between the first (Difference 1) and the second (Difference 2) terms on the RHSs of (2.54) and (2.248), respectively. For Difference 1,

$$\Big\|\mathcal{E}^{L_S}\Big(X_t X_t^{\boldsymbol{T}}\Big|\,\mathscr{Y}_t\Big) - \mathbb{E}\Big\{X_t X_t^{\boldsymbol{T}}\Big|\,\mathscr{Y}_t\Big\}\Big\|_1^{\mathsf{E}} \leq M^2 \sup_{(i,j)\in\mathbb{N}_M^+\times\mathbb{N}_M^+}\left|\mathbb{E}\Big\{\boldsymbol{\phi}^{ij}(X_t)\Big|\,\mathscr{Y}_t\Big\} - \boldsymbol{\Phi}^{ij}\frac{\boldsymbol{E}_t}{\|\boldsymbol{E}_t\|_1}\right|, \tag{2.251}$$

for all $t \in \mathbb{N}$, where we have exploited the definitions above and which means that Difference 1 converges to zero as $L_S \to \infty$, in the sense of Theorem 2.8, for any fixed natural $T < \infty$ and for the same measurable set $\widehat{\Omega}_T$ of Theorem 2.8 (also see Remark 2.19). For Difference 2, it is easy to show that

$$\begin{aligned}\Big\|\mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t)\Big(\mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t)\Big)^{\boldsymbol{T}} - \mathbb{E}\{X_t|\,\mathscr{Y}_t\}(\mathbb{E}\{X_t|\,\mathscr{Y}_t\})^{\boldsymbol{T}}\Big\|_1^{\mathsf{E}} \\ \leq \Big(\Big\|\mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t)\Big\|_1 + \|\mathbb{E}\{X_t|\,\mathscr{Y}_t\}\|_1\Big)\Big\|\mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t) - \mathbb{E}\{X_t|\,\mathscr{Y}_t\}\Big\|_1 \\ \leq 2M\gamma\Big\|\mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t) - \mathbb{E}\{X_t|\,\mathscr{Y}_t\}\Big\|_1,\end{aligned} \tag{2.252}$$

for all $t \in \mathbb{N}$, where we recall that $\gamma \equiv \max\{|a|, |b|\}$. Again, Difference 2 converges to zero as $L_S \to \infty$, exactly in the same sense as Difference 1 above. Consequently, putting it altogether, we have shown that

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left\|\mathcal{V}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)-\mathbb{V}\left\{X_t|\,\mathscr{Y}_t\right\}\right\|_1^{\mathsf{E}}\underset{L_S\to\infty}{\longrightarrow}0, \tag{2.253}$$

proving asymptotic consistency of the approximate estimator.

Now, in order to show that $\mathcal{V}^{L_S}\left(X_t|\,\mathscr{Y}_t\right)$ indeed has the form advocated in Theorem 2.9, it suffices to observe that (2.249) in fact coincides with the (i,j)-th element of the matrix

$$\mathbf{X}\mathrm{diag}\left(\frac{\boldsymbol{E}_t}{\|\boldsymbol{E}_t\|_1}\right)\mathbf{X}^{\boldsymbol{T}}\equiv\mathcal{E}^{L_S}\left(X_tX_t^{\boldsymbol{T}}\Big|\,\mathscr{Y}_t\right), \tag{2.254}$$

for all $t \in \mathbb{N}$. The proof is now complete. ■

2.2.7.5 Appendix E: Proof of Theorem 2.10

$$\begin{aligned}
&\operatorname*{ess\,sup}_{y\in\mathbb{R}}\left|\kappa(y|\,x)-\kappa_t\big(y|\in\mathcal{Z}_{L_S}(x)\big)\right|\\
&\equiv\operatorname*{ess\,sup}_{y\in\mathbb{R}}\left|\kappa(y|\,x)-\frac{\displaystyle\int_{\mathcal{Z}_{L_S}(x)}\kappa(y|\,\theta)\,\mathcal{P}_{X_{t-1}}(\mathrm{d}\theta)}{\mathcal{P}\left(X_{t-1}\in\mathcal{Z}_{L_S}\left(x\right)\right)}\right|\equiv\operatorname*{ess\,sup}_{y\in\mathbb{R}}\left|\frac{\displaystyle\int_{\mathcal{Z}_{L_S}(x)}\kappa(y|\,x)-\kappa(y|\,\theta)\,\mathcal{P}_{X_{t-1}}(\mathrm{d}\theta)}{\mathcal{P}\left(X_{t-1}\in\mathcal{Z}_{L_S}\left(x\right)\right)}\right|\\
&\leq\operatorname*{ess\,sup}_{y,\theta\in\mathcal{Z}_{L_S}(x)}\left|\kappa\left(y|\,x\right)-\kappa\left(y|\,\theta\right)\right|\equiv\operatorname*{ess\,sup}_{y,\theta\in\mathcal{Z}_{L_S}(x)}\left|f_W\left(y-h\left(x\right)\right)-f_W\left(y-h\left(\theta\right)\right)\right|\\
&\leq\operatorname*{ess\,sup}_{y,\theta\in\mathcal{Z}_{L_S}(x)}\frac{\left|\varphi\left(\dfrac{y-h\left(x\right)}{\sigma}\right)\mathbb{1}_{[-\alpha,\alpha]}\left(y-h\left(x\right)\right)-\varphi\left(\dfrac{y-h\left(\theta\right)}{\sigma}\right)\mathbb{1}_{[-\alpha,\alpha]}\left(y-h\left(\theta\right)\right)\right|}{2\sigma\Phi\left(\alpha/\sigma\right)-\sigma},\\
&\leq\operatorname*{ess\,sup}_{y,\theta\in\mathcal{Z}_{L_S}(x)}\left[\frac{\min\left\{\varphi\left(\dfrac{y-h\left(x\right)}{\sigma}\right),\varphi\left(\dfrac{y-h\left(\theta\right)}{\sigma}\right)\right\}\left|\mathbb{1}_{[-\alpha,\alpha]}(y-h\left(x\right))-\mathbb{1}_{[-\alpha,\alpha]}(y-h\left(\theta\right))\right|}{2\sigma\Phi\left(\alpha/\sigma\right)-\sigma}\right.\\
&\qquad\qquad\left.+\frac{\left|\varphi\left(\dfrac{y-h\left(x\right)}{\sigma}\right)-\varphi\left(\dfrac{y-h\left(\theta\right)}{\sigma}\right)\right|}{2\sigma\Phi\left(\alpha/\sigma\right)-\sigma}\right]
\end{aligned} \tag{2.255}$$

By Definition 2.4 and the additive model under consideration, it is obvious that we are interested in CRT II, which, for the case of an arbitrary initial measure $\mathcal{P}_{X_{-1}}$, is

equivalent to the strengthened *global* demand that

$$\operatorname*{ess\,sup}_{\boldsymbol{y}\in\mathbb{R}^{M\times 1}}\left|\kappa\left(\boldsymbol{y}\right|\boldsymbol{x}\right)-\kappa_{t}\left(\boldsymbol{y}\right|\in\mathcal{Z}_{L_{S}}\left(\boldsymbol{x}\right)\right)\left|\leq\delta_{L_{S},t}^{II}\left(\boldsymbol{x}\right),\right. \tag{2.256}$$

being true $\mathcal{P}_{X_t}-a.e.$, for some $\mathcal{P}_{X_t}-UI$, nonnegative sequence $\left\{\delta_{n,t}^{II}\left(\cdot\right)\right\}_{n\in\mathbb{N}^{+}}$, with $\delta_{n,t}^{II}\left(\cdot\right)\underset{n\to\infty}{\longrightarrow}0$, $\mathcal{P}_{X_t}-a.e$, ***for all*** $t\in\{-1\}\cup\mathbb{N}_{T}$, for some desired $T\in[0,\infty]$. Of course, $\kappa_{t}\left(\cdot|\in\mathcal{Z}_{L_{S}}\left(\cdot\right)\right)$ is defined exactly as in (2.183), but with an explicit subscript "t", indicating possible temporal variability.

Then, in regard to the additive NAR under consideration and using the respective definitions, it is true that (see (2.255))

$$\begin{aligned}\operatorname*{ess\,sup}_{y\in\mathbb{R}}\left|\kappa(y|\,x)-\kappa_{t}\left(y\right|\in\mathcal{Z}_{L_{S}}(x)\right)\left|\right.&\leq\operatorname*{ess\,sup}_{y\in\mathbb{R},\theta\in\mathcal{Z}_{L_{S}}(x)}\frac{\varphi\left(\frac{\alpha}{\sigma}\right)+\left|\varphi\left(\frac{y-h\left(x\right)}{\sigma}\right)-\varphi\left(\frac{y-h\left(\theta\right)}{\sigma}\right)\right|}{2\sigma\Phi\left(\alpha/\sigma\right)-\sigma}\\&\equiv f_{W}\left(\alpha\right)+\operatorname*{ess\,sup}_{y\in\mathbb{R},\theta\in\mathcal{Z}_{L_{S}}(x)}\frac{\left|\varphi\left(\frac{y-h\left(x\right)}{\sigma}\right)-\varphi\left(\frac{y-h\left(\theta\right)}{\sigma}\right)\right|}{2\sigma\Phi\left(\alpha/\sigma\right)-\sigma}\\&\leq f_{W}\left(\alpha\right)+\frac{\sup\limits_{\theta\in\mathcal{Z}_{L_{S}}(x)}\left|h\left(x\right)-h\left(\theta\right)\right|}{\left(2\sigma^{2}\Phi\left(\alpha/\sigma\right)-\sigma^{2}\right)\sqrt{2e\pi}},\ \mathcal{P}_{X_t}-a.e.,\end{aligned} \tag{2.257}$$

for all $t\in\{-1\}\cup\mathbb{N}_{T}$. From (2.257), it is almost obvious that $\sup_{\theta\in\mathcal{Z}_{L_{S}}(x)}\left|h\left(x\right)-h\left(\theta\right)\right|$ vanishes as $L_S\to\infty$. Indeed, for each fixed x, by definition of $\mathcal{Z}_{L_{S}}\left(x\right)$, it follows that

$$\begin{aligned}\sup_{\theta\in\mathcal{Z}_{L_{S}}(x)}\left|h\left(x\right)-h\left(\theta\right)\right|&\equiv\sup_{\left|\theta-\mathcal{Q}_{L_{S}}(x)\right|\leq\frac{B+\alpha}{L_{S}}}\left|h\left(x\right)-h\left(\theta\right)\right|\\&\equiv\left|h\left(x\right)-h\left(\theta_{L_{S}}^{*}\left(x\right)\right)\right|,\end{aligned} \tag{2.258}$$

where $\theta_{L_{S}}^{*}\left(x\right)\underset{L_{S}\to\infty}{\longrightarrow}x$, $\mathcal{P}_{X_t}-a.e.$. Thus, due to the continuity of $h\left(\cdot\right)$,

$$\sup_{\theta\in\mathcal{Z}_{L_{S}}(x)}\left|h\left(x\right)-h\left(\theta\right)\right|\underset{L_{S}\to\infty}{\longrightarrow}0, \tag{2.259}$$

$\mathcal{P}_{X_t}-a.e.$, for all $t\in\{-1\}\cup\mathbb{N}_{T}$. Now, note that $\sup_{\theta\in\mathcal{Z}_{L_{S}}(x)}\left|h\left(x\right)-h\left(\theta\right)\right|\leq 2B$, set

$$\delta_{L_{S}}^{II}\left(x\right)\triangleq\frac{\sup\limits_{\theta\in\mathcal{Z}_{L_{S}}(x)}\left|h\left(x\right)-h\left(\theta\right)\right|}{\left(2\sigma^{2}\Phi\left(\alpha/\sigma\right)-\sigma^{2}\right)\sqrt{2e\pi}} \tag{2.260}$$

and choose $T\equiv\infty$. The proof is complete. ■

2.2.7.6 Appendix F: Proof of Lemma 2.13

This is a technical proof and requires a deeper appeal to the theoretics of change of probability measures. Until now, we have made use of the so called ***reverse*** [12] change of measure formula

$$\mathbb{E}_{\mathcal{P}}\left\{ X_t \middle|\, \mathscr{Y}_t \right\} \equiv \frac{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ X_t \Lambda_t \middle|\, \mathscr{Y}_t \right\}}{\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ \Lambda_t \middle|\, \mathscr{Y}_t \right\}}, \quad \forall t \in \mathbb{N}. \tag{2.261}$$

Formula (2.261) is characterized as reverse, simply because it provides a representation for the conditional expectation of X_t under the original base measure $\mathcal{P}$ via operations performed exclusively under another auxiliary, hypothetical base measure $\widetilde{\mathcal{P}}$. In full generality, the likelihood ratio process Λ_t on the RHS of (2.261) may be expressed as

$$\begin{aligned}
\Lambda_t &\equiv \frac{\prod\limits_{i\in\mathbb{N}_t} \exp\left(\frac{1}{2}\left\|\mathbf{y}_i\right\|_2^2 - \frac{1}{2}\left(\mathbf{y}_i - \boldsymbol{\mu}_i\left(X_i\right)\right)^{\boldsymbol{T}} \left(\boldsymbol{\Sigma}_i\left(X_i\right) + \sigma_\xi^2 \mathbf{I}_{N\times N} \right)^{-1} \left(\mathbf{y}_i - \boldsymbol{\mu}_i\left(X_i\right)\right) \right)}{\prod\limits_{i\in\mathbb{N}_t} \sqrt{\det\left(\boldsymbol{\Sigma}_i\left(X_i\right) + \sigma_\xi^2 \mathbf{I}_{N\times N} \right)}} \\
&\equiv \prod_{i\in\mathbb{N}_t} \frac{\sqrt{\left(2\pi\right)^N}}{\exp\left(-\frac{1}{2}\left\|\mathbf{y}_i\right\|_2^2 \right)} \frac{\exp\left(-\frac{1}{2}\left(\mathbf{y}_i - \boldsymbol{\mu}_i\left(X_i\right)\right)^{\boldsymbol{T}} \left(\boldsymbol{\Sigma}_i\left(X_i\right) + \sigma_\xi^2 \mathbf{I}_{N\times N} \right)^{-1} \left(\mathbf{y}_i - \boldsymbol{\mu}_i\left(X_i\right)\right) \right)}{\sqrt{\left(2\pi\right)^N} \sqrt{\det\left(\boldsymbol{\Sigma}_i\left(X_i\right) + \sigma_\xi^2 \mathbf{I}_{N\times N} \right)}} \\
&\equiv \prod_{i\in\mathbb{N}_t} \frac{\mathcal{N}\left(\mathbf{y}_i; \boldsymbol{\mu}_i\left(X_i\right), \mathbf{C}_i\left(X_i\right)\right)}{\mathcal{N}\left(\mathbf{y}_i; \mathbf{0}, \mathbf{I}\right)} \\
&\triangleq \prod_{i\in\mathbb{N}_t} \mathsf{L}_i\left(X_i, \mathbf{y}_i\right) \in \mathbb{R}_{++},
\end{aligned} \tag{2.262}$$

for all $t \in \mathbb{N}$. Also, $\Lambda_{-1} \equiv 1$. Note that we have slightly overloaded the definition of the Λ_i's and L_i's, compared to (2.18). But this is fine, since the term $\exp\left(-\left\|\mathbf{y}_t\right\|_2^2 / 2 \right)$ is $\{\mathscr{Y}_t\}$-adapted. Here, Λ_t, *as defined in (2.262)*, is interpreted precisely as the restriction of the Radon-Nikodym derivative $\mathrm{d}\mathcal{P}/\mathrm{d}\widetilde{\mathcal{P}}$ on the filtration $\{\mathscr{H}_t\}_{t\in\mathbb{N}}$, generated by both X_t (including X_{-1} in $\mathscr{H}_0$) and $\mathbf{y}_t$. That is,

$$\left.\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\widetilde{\mathcal{P}}}\right|_{\mathscr{H}_t} \equiv \Lambda_t, \quad \forall t \in \mathbb{N} \cup \{-1\} \quad \text{with} \tag{2.263}$$

$$1 \equiv \Lambda_{-1}. \tag{2.264}$$

Observe that, for at $t \in \mathbb{N}$, $\mathscr{Y}_t \subset \mathscr{H}_t$ and, thus, (2.261) is a valid expression. In other words, the Radon-Nikodym Theorem is applied accordingly on the measurable space $(\Omega, \mathscr{H}_t)$, for each $t \in \mathbb{N}$.

However, because the base measures $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ are *equivalent on* $\mathscr{H}_t$ (that is, the one is absolutely continuous with respect to the other), it is possible, in exactly the same fashion as above, to "start" under $\mathcal{P}$ and express conditional expectations under $\widetilde{\mathcal{P}}$ via a ***forward*** change of measure formula. In particular, it is true that

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ X_t \middle| \mathscr{Y}_t \right\} \equiv \frac{\mathbb{E}_{\mathcal{P}}\left\{ X_t \Lambda_t^{-1} \middle| \mathscr{Y}_t \right\}}{\mathbb{E}_{\mathcal{P}}\left\{ \Lambda_t^{-1} \middle| \mathscr{Y}_t \right\}}, \quad \forall t \in \mathbb{N} \tag{2.265}$$

where, as it is natural, this time we have

$$\left.\frac{\mathrm{d}\widetilde{\mathcal{P}}}{\mathrm{d}\mathcal{P}}\right|_{\mathscr{H}_t} \equiv \Lambda_t^{-1}, \quad \forall t \in \mathbb{N} \cup \{-1\} \quad \text{with} \tag{2.266}$$

$$1 \equiv \Lambda_{-1}^{-1}. \tag{2.267}$$

From the above, one may realize that the "mechanics" of the change of measure procedures (forward and reverse), at least in discrete time, are very well structured and much simpler than they may initially seem to be at a first glance. In more generality, it is true that if $\mathscr{C}_t$ is a sub σ-algebra of $\mathscr{H}_t$ and for a $\{\mathscr{H}_t\}$-adapted process H_t [**?**, 12, 15],

$$\mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ H_t \middle| \mathscr{C}_t \right\} \equiv \frac{\mathbb{E}_{\mathcal{P}}\left\{ H_t \Lambda_t^{-1} \middle| \mathscr{C}_t \right\}}{\mathbb{E}_{\mathcal{P}}\left\{ \Lambda_t^{-1} \middle| \mathscr{C}_t \right\}}, \quad \forall t \in \mathbb{N}. \tag{2.268}$$

And, of course, we can even evaluate (conditional) probabilities under $\widetilde{\mathcal{P}}$ as

$$\widetilde{\mathcal{P}}\left(H_t \in \mathcal{A} \middle| \mathscr{C}_t \right) \equiv \mathbb{E}_{\widetilde{\mathcal{P}}}\left\{ \mathbb{1}_{\{H_t \in \mathcal{A}\}} \middle| \mathscr{C}_t \right\} \equiv \frac{\mathbb{E}_{\mathcal{P}}\left\{ \mathbb{1}_{\{H_t \in \mathcal{A}\}} \Lambda_t^{-1} \middle| \mathscr{C}_t \right\}}{\mathbb{E}_{\mathcal{P}}\left\{ \Lambda_t^{-1} \middle| \mathscr{C}_t \right\}}, \quad \forall t \in \mathbb{N}, \tag{2.269}$$

for any Borel set $\mathcal{A}$.

Now, consider the process $X_t \equiv f\left(X_{t-1}, W_t\right), t \in \mathbb{N}$. As assumed throughout this work, X_t is Markov under $\mathcal{P}$, with W_t being a white noise (i.i.d.) innovations process. Also, under $\widetilde{\mathcal{P}}$, X_t is again Markov with exactly the same dynamics, *but independent of* $\mathbf{y}_t$. However, at this point nothing is known regarding the nature of W_t (distribution, whiteness) and how it is related to X_{-1} and $\mathbf{y}_t$. The proof of the remarkable fact that, without any other modification, $\widetilde{\mathcal{P}}$ may be chosen such that W_t indeed satisfies the aforementioned properties under question, follows.

Without changing the respective Radon-Nikodym derivatives for either the forward or reverse change of measure formulas presented above, let us *enlarge* the measurable

space for which the change of measure procedure is valid, by defining $\{\mathscr{H}_t\}_{t\in\mathbb{N}}$ to be the joint filtration generated by, $\mathbf{y}_t$, the initial condition X_{-1} and the innovations process W_t (Why enlarged?). Our goal in the following will be to show the following, regarding the base measure $\widetilde{\mathcal{P}}$, defined, for each $t \in \mathbb{N}$, on the enlarged measurable space $(\Omega, \mathscr{H}_t)$:

1. First, we will show that, under $\widetilde{\mathcal{P}}$, the observations process $\mathbf{y}_t$ is mutually independent of both X_{-1} and W_t and therefore also independent of the state X_t.

2. Second, we will show that, under $\widetilde{\mathcal{P}}$, W_t is white and identically distributed as as under $\mathcal{P}$ (in addition to it being independent of $\mathbf{y}_t$ from (1)).

3. Third, we will show that, under $\widetilde{\mathcal{P}}$, X_t is Markov with the same dynamics as under $\mathcal{P}$ (in addition to it being independent of $\mathbf{y}_t$ from (1)).

In order to embark on the rigorous proof of the above, define, for each $t \in \mathbb{N}$, the auxiliary σ-algebra $\mathscr{H}_t^-$, generated by $\{\mathbf{y}_i\}_{i\in\mathbb{N}_{t-1}}$, X_{-1} and $\{W_i\}_{i\in\mathbb{N}_t}$.

1. For any $\alpha \in \mathbb{R}^{N\times 1}$, it is true that (the "$\leq$" operator is interpreted in the elementwise sense)

$$\begin{aligned}\widetilde{\mathcal{P}}\left(\mathbf{y}_t \leq \alpha \,\middle|\, \mathscr{H}_t^-\right) &\equiv \mathbb{E}_{\widetilde{\mathcal{P}}}\left\{\mathbb{1}_{\{\mathbf{y}_t\leq\alpha\}}\middle|\,\mathscr{H}_t^-\right\}\\ &\equiv \frac{\mathbb{E}_{\mathcal{P}}\left\{\mathbb{1}_{\{\mathbf{y}_t\leq\alpha\}}\Lambda_t^{-1}\middle|\,\mathscr{H}_t^-\right\}}{\mathbb{E}_{\mathcal{P}}\left\{\Lambda_t^{-1}\middle|\,\mathscr{H}_t^-\right\}}\\ &= \frac{\mathbb{E}_{\mathcal{P}}\left\{\mathbb{1}_{\{\mathbf{y}_t\leq\alpha\}}\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_t^-\right\}}{\mathbb{E}_{\mathcal{P}}\left\{\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_t^-\right\}}, \quad \forall t\in\mathbb{N}. \end{aligned} \tag{2.270}$$

Let us consider the denominator $\mathbb{E}_{\mathcal{P}}\left\{\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_t^-\right\}$. We have

$$\begin{aligned}\mathbb{E}_{\mathcal{P}}\left\{\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_t^-\right\} &\equiv \mathbb{E}_{\mathcal{P}}\left\{\frac{\mathcal{N}\left(\mathbf{y}_t;\mathbf{0},\mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)}\middle|\,\mathscr{H}_t^-\right\}\\ &= \mathbb{E}_{\mathcal{P}}\left\{\frac{\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)}{\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)}\middle|\,\mathscr{H}_t^-\right\}, \end{aligned} \tag{2.271}$$

and given the facts that knowledge of X_{-1} and $\{W_i\}_{i\in\mathbb{N}_t}$ completely determines $\{X_i\}_{i\in\mathbb{N}_t}$ and that the observations are conditionally independent given the states $\{X_i\}_{i\in\mathbb{N}_t}$, we get, for every $t \in \mathbb{N}$,

$$\mathbb{E}_{\mathcal{P}}\left\{\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_t^-\right\} = \int \frac{\mathcal{N}\left(\mathbf{y}_t;\mathbf{0},\mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)}\mathcal{N}\left(\mathbf{y}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)\mathrm{d}\mathbf{y}_t \equiv 1. \tag{2.272}$$

Likewise, concerning the numerator $\mathbb{E}_{\mathcal{P}}\left\{\mathbb{1}_{\{\mathbf{y}_t\leq\alpha\}}\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_t^-\right\}$, it is true that

$$\begin{aligned}\mathbb{E}_{\mathcal{P}}\left\{\mathbb{1}_{\{\mathbf{y}_t\leq\alpha\}}\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_t^-\right\}&\equiv\mathbb{E}_{\mathcal{P}}\left\{\mathbb{1}_{\{\mathbf{y}_t\leq\alpha\}}\frac{\mathcal{N}\left(\mathbf{y}_t;\mathbf{0},\mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)}\middle|\,\mathscr{H}_t^-\right\}\\&=\int\frac{\mathbb{1}_{\{\mathbf{y}_t\leq\alpha\}}\mathcal{N}\left(\mathbf{y}_t;\mathbf{0},\mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)}\mathcal{N}\left(\mathbf{y}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)\mathrm{d}\mathbf{y}_t\\&\equiv\int\mathbb{1}_{\{\mathbf{y}_t\leq\alpha\}}\mathcal{N}\left(\mathbf{y}_t;\mathbf{0},\mathbf{I}\right)\mathrm{d}\mathbf{y}_t,\end{aligned}\tag{2.273}$$

or, equivalently,

$$\widetilde{\mathcal{P}}\left(\mathbf{y}_t\leq\alpha\middle|\,\mathscr{H}_t^-\right)\equiv\widetilde{\mathcal{P}}\left(\mathbf{y}_t\leq\alpha\right),\quad\forall t\in\mathbb{N}\tag{2.274}$$

and for any $\alpha\in\mathbb{R}^{N\times1}$. Therefore, $\mathbf{y}_t$ is white standard normal under $\widetilde{\mathcal{P}}$ and, additionally, mutually independent of X_{-1} and W_t and, therefore, mutually independent of X_t, too.

2. Similarly, concerning the innovations process W_t, for any $\alpha\in\mathbb{R}^{M_W\times1}$, it is true that

$$\widetilde{\mathcal{P}}\left(W_t\leq\alpha\,|\mathscr{H}_{t-1}\right)\equiv\frac{\mathbb{E}_{\mathcal{P}}\left\{\mathbb{1}_{\{W_t\leq\alpha\}}\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_{t-1}\right\}}{\mathbb{E}_{\mathcal{P}}\left\{\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_{t-1}\right\}},\quad\forall t\in\mathbb{N}.\tag{2.275}$$

In this case, for the denominator, we again have

$$\mathbb{E}_{\mathcal{P}}\left\{\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_{t-1}\right\}\equiv\mathbb{E}_{\mathcal{P}}\left\{\frac{\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)}{\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)}\middle|\,\mathscr{H}_{t-1}\right\},\tag{2.276}$$

but because $X_t\equiv f\left(X_{t-1},W_t\right)$, knowledge of X_{-1} and $\{W_i\}_{i\in\mathbb{N}_{t-1}}$ completely determines $\{X_i\}_{i\in\mathbb{N}_{t-1}}$, the processes W_t and $\boldsymbol{u}_t$ are mutually independent and since the random variable W_t is independent of $\{\mathbf{y}_i\}_{i\in\mathbb{N}_{t-1}^+}$, we get

$$\begin{aligned}&\mathbb{E}_{\mathcal{P}}\left\{\mathsf{L}_t^{-1}\middle|\,\mathscr{H}_{t-1}\right\}\\&=\int_{W_t}\int_{\boldsymbol{u}_t}\frac{\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)}{\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)}\mathcal{N}\left(\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)\mathrm{d}\boldsymbol{u}_t\mathcal{P}_{W_t}\left(\mathrm{d}W_t\right)\\&=\int_{W_t}\int_{\boldsymbol{u}_t}\sqrt{\det\left(\mathbf{C}_t\left(X_t\right)\right)}\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)\mathrm{d}\boldsymbol{u}_t\mathcal{P}_{W_t}\left(\mathrm{d}W_t\right)\\&\equiv\int_{W_t}\int_{\boldsymbol{u}_t}\det\left(\sqrt{\mathbf{C}_t\left(X_t\right)}\right)\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)\mathrm{d}\boldsymbol{u}_t\mathcal{P}_{W_t}\left(\mathrm{d}W_t\right)\\&=\int_{W_t}\int_{\boldsymbol{u}_t}\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)\mathrm{d}\left[\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t\right]\mathcal{P}_{W_t}\left(\mathrm{d}W_t\right)\\&\equiv\int_{W_t}\mathcal{P}_{W_t}\left(\mathrm{d}W_t\right)\end{aligned}$$

$$\equiv 1. \tag{2.277}$$

Likewise, the numerator can be expanded as

$$\begin{aligned}
&\mathbb{E}_{\mathcal{P}}\left\{\mathbb{1}_{\{W_t\leq\alpha\}}\mathsf{L}_t^{-1}\middle|\mathscr{H}_{t-1}\right\}\\
&\quad\equiv\int_{W_t}\mathbb{1}_{\{W_t\leq\alpha\}}\int_{\boldsymbol{u}_t}\frac{\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)\mathcal{N}\left(\boldsymbol{u}_t;\mathbf{0},\mathbf{I}\right)}{\mathcal{N}\left(\boldsymbol{\mu}_t\left(X_t\right)+\sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t;\boldsymbol{\mu}_t\left(X_t\right),\mathbf{C}_t\left(X_t\right)\right)}\mathrm{d}\boldsymbol{u}_t\mathcal{P}_{W_t}\left(\mathrm{d}W_t\right)\\
&\quad\equiv\int_{W_t}\mathbb{1}_{\{W_t\leq\alpha\}}\mathcal{P}_{W_t}\left(\mathrm{d}W_t\right),
\end{aligned}\tag{2.278}$$

or, equivalently,

$$\widetilde{\mathcal{P}}\left(W_t\leq\alpha\,|\mathscr{H}_{t-1}\right)\equiv\widetilde{\mathcal{P}}\left(W_t\leq\alpha\right)\equiv\mathcal{P}\left(W_t\leq\alpha\right),\quad\forall t\in\mathbb{N}\tag{2.279}$$

and for any $\alpha\in\mathbb{R}^{M_W\times 1}$. Therefore, W_t is white under $\widetilde{\mathcal{P}}$, in addition to it being independent of $\mathbf{y}_t$ and with the same distribution as under $\mathcal{P}$.

3. It suffices to show that, under $\widetilde{\mathcal{P}}$, the initial condition X_{-1} has the same distribution as under $\mathcal{P}$. If this is true, then, given all the above facts, under $\widetilde{\mathcal{P}}$, the process $X_t\equiv f\left(X_{t-1},W_t\right),t\in\mathbb{N}$ is Markov with the same dynamics as under $\mathcal{P}$. Indeed, for any $\alpha\in\mathbb{R}^{M\times 1}$, it is trivially true that

$$\begin{aligned}
\widetilde{\mathcal{P}}\left(X_{-1}\leq\alpha\right)&\equiv\widetilde{\mathcal{P}}\left(X_{-1}\leq\alpha\,|\{\emptyset,\Omega\}\right)\\
&=\frac{\mathbb{E}_{\mathcal{P}}\left\{\mathbb{1}_{\{X_{-1}\leq\alpha\}}\Lambda_t^{-1}\right\}}{\mathbb{E}_{\mathcal{P}}\left\{\Lambda_t^{-1}\right\}},\quad\forall t\in\mathbb{N}\cup\{-1\}.
\end{aligned}\tag{2.280}$$

Simply, choose $t\equiv -1$. QED. ■

Chapter 3

Space-Time Dynamic Wireless Channel Modeling

In addition to the temporal variation of the wireless medium, recently, considerable interest has been expressed concerning its spatial variation as well. Such knowledge are beneficial in emerging areas such as mobile beamforming [48] spatially controlled communications (which is the concentration of the dissertation, as well), mobility enhanced physical layer security [49–51], communication-aware motion and path planning, network routing, connectivity maintenance and physical layer based dynamic coverage [52–54]. In all these cases, *dynamic spatiotemporal channel estimation/tracking and prediction* becomes an essential part of mobility control, since it would provide valuable physical layer related information (channel maps), which is necessary for dynamic decision making and stochastic control.

Two distinct models describing the spatiotemporal interactions of the communication channel, the latter seen holistically as a stochastic space-time field, are presented below. The first model is based on parametric spatiotemporal Gaussian random field theory. The second model is based on a hierarchical approach to channel modeling. Additionally, for the latter model, *non-trivial* statistical inference is discussed, under an approximate grid-based nonlinear filtering framework, exploiting the results presented previously in Chapter 2.

The two models are presented under the assumption of a set of possibly mobile, single antenna nodes in the space, communicating with some spatially fixed, reference base station, as in Section 1.1. Both models are based on the and widely acceptable description of the respective communication links as a multiplicative stochastic system, consisting of three components; path loss, shadowing (large scale fading), and multipath (small scale) fading.

3.1 Parametric Spatiotemporal Gaussian Channel Modeling

This section introduces a general *parametric* stochastic model for describing the spatiotemporal evolution of the wireless channel. For the benefit of the reader, a more intuitive justification of this general model is also provided. Additionally, some extensions to the model are briefly discussed, highlighting its versatility, along with some technical considerations, which will be of importance later, for analyzing the theoretical consistency of the subsequently proposed techniques.

3.1.1 Large Scale Gaussian Channel Modeling in the dB Domain

At each space-time point $(\mathbf{p},t) \in \mathcal{S} \times \mathbb{N}_{N_T}^{+}$, the source-relay channel field may be decomposed as the product of three space-time varying components [55], as

$$f\left(\mathbf{p},t\right) \equiv \underbrace{f^{PL}(\mathbf{p})}_{\text{path loss}} \underbrace{f^{SH}(\mathbf{p},t)}_{\text{shadowing}} \underbrace{f^{MF}(\mathbf{p},t)}_{\text{fading}} e^{\mathfrak{J}\frac{2\pi\|\mathbf{p}-\mathbf{p}_S\|_2}{\lambda}}, \tag{3.1}$$

where $\mathfrak{J} \triangleq \sqrt{-1}$ denotes the imaginary unit, $\lambda > 0$ denotes the wavelength employed for the communication, and:

1. $f^{PL}\left(\mathbf{p}\right) \triangleq \|\mathbf{p}-\mathbf{p}_S\|_2^{-\ell/2}$ is the *path loss field*, a deterministic quantity, with $\ell > 0$ being the path loss exponent.

2. $f^{SH}\left(\mathbf{p},t\right) \in \mathbb{R}$ is the *shadowing field*, whose square is, for each $(\mathbf{p},t) \in \mathcal{S} \times \mathbb{N}_{N_T}^{+}$, a base-10 log-normal random variable with zero location.

3. $f^{MF}\left(\mathbf{p},t\right) \in \mathbb{C}$ constitutes the *multipath fading field*, a stationary process with known statistics.

The same decomposition holds in direct correspondence for the relay-destination channel field, $g\left(\mathbf{p},t\right)$. Additionally, if "$\perp\!\!\!\perp$" means "is statistically independent of", it is assumed that [56]

$$\left[f^{MF}\left(\mathbf{p},t\right) g^{MF}\left(\mathbf{p},t\right)\right] \perp\!\!\!\perp \left[f^{SH}\left(\mathbf{p},t\right) g^{SH}\left(\mathbf{p},t\right)\right] \quad \text{and} \tag{3.2}$$

$$f^{MF}\left(\mathbf{p},t\right) \perp\!\!\!\perp g^{MF}\left(\mathbf{p},t\right). \tag{3.3}$$

In particular, if the phase of $f^{MF}(\mathbf{p},t)$ is denoted as $\phi_f(\mathbf{p},t) \in [-\pi,\pi]$, is further assumed that

$$\left|f^{MF}(\mathbf{p},t)\right| \perp\!\!\!\perp \phi_f(\mathbf{p},t), \tag{3.4}$$

and the same for for $g^{MF}(\mathbf{p},t)$. It also follows that

$$\left[\left|f^{MF}(\mathbf{p},t)\right| \left|g^{MF}(\mathbf{p},t)\right|\right] \perp\!\!\!\perp \left[f^{SH}(\mathbf{p},t)\, g^{SH}(\mathbf{p},t)\right]. \tag{3.5}$$

We are interested in the magnitudes of both fields $f(\mathbf{p},t)$ and $g(\mathbf{p},t)$. Instead of working with the multiplicative model described by (3.1), it is much preferable to work in logarithmic scale. We may define the *log-scale magnitude field*

$$F(\mathbf{p},t) \triangleq \alpha_S(\mathbf{p})\,\ell + \sigma_S(\mathbf{p},t) + \xi_S(\mathbf{p},t), \tag{3.6}$$

where we define

$$-\alpha_S(\mathbf{p}) \triangleq 10\log_{10}\left(\|\mathbf{p}-\mathbf{p}_S\|_2\right), \tag{3.7}$$

$$\sigma_S(\mathbf{p},t) \triangleq 10\log_{10}\left(f^{SH}(\mathbf{p},t)\right)^2 \quad \text{and} \tag{3.8}$$

$$\xi_S(\mathbf{p},t) \triangleq 10\log_{10}\left|f^{MF}(\mathbf{p},t)\right|^2 - \rho, \quad \text{with} \tag{3.9}$$

$$\rho \triangleq \mathbb{E}\left\{10\log_{10}\left|f^{MF}(\mathbf{p},t)\right|^2\right\}, \tag{3.10}$$

for all $(\mathbf{p},t) \in \mathcal{S}\times\mathbb{N}^+_{N_T}$. It is then trivial to show that the magnitude of $f(\mathbf{p},t)$ may be reconstructed via the *bijective* formula

$$|f(\mathbf{p},t)| \equiv 10^{\rho/20}\exp\left(\frac{\log(10)}{20}F(\mathbf{p},t)\right), \tag{3.11}$$

for all $(\mathbf{p},t) \in \mathcal{S}\times\mathbb{N}^+_{N_T}$, a "trick" that will prove very useful in the next section. Regarding $g(\mathbf{p},t)$, the log-scale field $G(\mathbf{p},t)$ is defined in the same fashion, but replacing the subscript "S" by "D".

For each relay $i \in \mathbb{N}^+_R$, let us define the respective log-scale channel magnitude processes $F_i(t) \triangleq F(\mathbf{p}_i(t),t)$ and $G_i(t) \triangleq G(\mathbf{p}_i(t),t)$, for all $t \in \mathbb{N}^+_{N_T}$. Of course, we may stack all the $F_i(t)$'s defined in (3.6), resulting in the vector additive model

$$\boldsymbol{F}(t) \triangleq \boldsymbol{\alpha}_S(\mathbf{p}(t))\,\ell + \boldsymbol{\sigma}_S(t) + \boldsymbol{\xi}_S(t) \in \mathbb{R}^{R\times 1}, \tag{3.12}$$

where $\boldsymbol{\alpha}_S(t)$, $\boldsymbol{\sigma}_S(t)$ and $\boldsymbol{\xi}_S(t)$ are defined accordingly. We can also define $\boldsymbol{G}(t) \triangleq \boldsymbol{\alpha}_D(\mathbf{p}(t))\,\ell + \boldsymbol{\sigma}_D(t) + \boldsymbol{\xi}_D(t) \in \mathbb{R}^{R\times 1}$, with each quantity in direct correspondence with (3.12). We may also define, in the same manner, the log-scale shadowing and multipath fading processes $\sigma^i_{S(D)}(t) \triangleq \sigma_{S(D)}(\mathbf{p}_i(t), t)$ and $\xi^i_{S(D)}(t) \triangleq \xi_{S(D)}(\mathbf{p}_i(t), t)$, for all $t \in \mathbb{N}^+_{N_T}$, respectively.

Next, let us focus on the spatiotemporal dynamics of $\{|f_i(t)|\}_i$ and $\{|g_i(t)|\}_i$, which are modeled through those of the shadowing components of $\{F_i(t)\}_i$ and $\{G_i(t)\}_i$. It is assumed that, for any N_T and any *deterministic* ensemble of positions of the relays in $\mathbb{N}^+_{N_T}$, say $\{\mathbf{p}(t)\}_{t\in\mathbb{N}^+_{N_T}}$, the random vector

$$\left[\boldsymbol{F}^{\boldsymbol{T}}(1)\ \boldsymbol{G}^{\boldsymbol{T}}(1)\ \ldots\ \boldsymbol{F}^{\boldsymbol{T}}(N_T)\ \boldsymbol{G}^{\boldsymbol{T}}(N_T)\right]^{\boldsymbol{T}} \in \mathbb{R}^{2RN_T\times 1} \tag{3.13}$$

is *jointly Gaussian* with known means and known covariance matrix. More specifically, on a per node basis, we let $\xi^i_{S(D)}(t) \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma^2_\xi\right)$ and $\sigma^i_{S(D)}(t) \overset{i.d.}{\sim} \mathcal{N}\left(0, \eta^2\right)$, for all $t \in \mathbb{N}^+_{N_T}$ and $i \in \mathbb{N}^+_R$ [56, 57]. In particular, extending Gudmundson's model [58] in a straightforward way, we propose defining the spatiotemporal correlations of the shadowing part of the channel as

$$\mathbb{E}\left\{\sigma^i_S(k)\,\sigma^j_S(l)\right\} \triangleq \eta^2 \exp\left(-\frac{\left\|\mathbf{p}_i(k) - \mathbf{p}_j(l)\right\|_2}{\beta} - \frac{|k-l|}{\gamma}\right), \tag{3.14}$$

and correspondingly for $\left\{\sigma^i_D(t)\right\}_{i\in\mathbb{N}^+_R}$, and additionally,

$$\mathbb{E}\left\{\sigma^i_S(k)\,\sigma^j_D(l)\right\} \triangleq \mathbb{E}\left\{\sigma^i_S(k)\,\sigma^j_S(l)\right\} \exp\left(-\frac{\|\mathbf{p}_S - \mathbf{p}_D\|_2}{\delta}\right), \tag{3.15}$$

for all $(i,j) \in \mathbb{N}^+_R \times \mathbb{N}^+_R$ and for all $(k,l) \in \mathbb{N}^+_{N_T} \times \mathbb{N}^+_{N_T}$. In the above, $\eta^2 > 0$ and $\beta > 0$ are called the *shadowing power* and the *correlation distance*, respectively [58]. In this fashion, we will call $\gamma > 0$ and $\delta > 0$ the *correlation time* and the *BS (Base Station) correlation*, respectively. For later reference, let us define the (cross)covariance matrices

$$\boldsymbol{\Sigma}_{SD}(k,l) \triangleq \mathbb{E}\left\{\boldsymbol{\sigma}_S(k)\,\boldsymbol{\sigma}^{\boldsymbol{T}}_D(l)\right\} + \mathbb{1}_{\{S\equiv D\}}\mathbb{1}_{\{k\equiv l\}}\sigma^2_\xi \mathbf{I}_R \in \mathbb{S}^R, \tag{3.16}$$

as well as

$$\boldsymbol{\Sigma}(k,l) \triangleq \begin{bmatrix} \boldsymbol{\Sigma}_{SS}(k,l) & \boldsymbol{\Sigma}_{SD}(k,l) \\ \boldsymbol{\Sigma}_{SD}(k,l) & \boldsymbol{\Sigma}_{DD}(k,l) \end{bmatrix} \in \mathbb{S}^{2R}, \tag{3.17}$$

for all $(k,l) \in \mathbb{N}_{N_T}^{+} \times \mathbb{N}_{N_T}^{+}$. Using these definitions, the covariance matrix of the joint distribution describing (3.13) can be readily expressed as

$$\mathbf{\Sigma} \triangleq \begin{bmatrix} \mathbf{\Sigma}(1,1) & \mathbf{\Sigma}(1,2) & \dots & \mathbf{\Sigma}(1,N_T) \\ \mathbf{\Sigma}(2,1) & \mathbf{\Sigma}(2,2) & \dots & \mathbf{\Sigma}(2,N_T) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}(N_T,1) & \mathbf{\Sigma}(N_T,2) & \cdots & \mathbf{\Sigma}(N_T,N_T) \end{bmatrix} \in \mathbb{S}^{2RN_T}. \tag{3.18}$$

Of course, in order for $\mathbf{\Sigma}$ to be a valid covariance matrix, it must be at least positive semidefinite, that is, in $\mathbb{S}_{+}^{2RN_T}$. If fact, for nearly all cases of interest, $\mathbf{\Sigma}$ is guaranteed to be strictly positive definite (or in $\mathbb{S}_{++}^{2RN_T}$), as the following result suggests.

Lemma 3.1. (Positive (Semi)Definiteness of $\mathbf{\Sigma}$) *For all possible deterministic trajectories of the relays on $\mathcal{S}^R \times \mathbb{N}_{N_T}^{+}$, it is true that $\mathbf{\Sigma} \in \mathbb{S}_{++}^{2RN_T}$, as long as $\sigma_\xi^2 \neq 0$. Otherwise, $\mathbf{\Sigma} \in \mathbb{S}_{+}^{2RN_T}$. In other words, as long as multipath (small-scale) fading is present in the channel response, the joint Gaussian distribution of the channel vector in (3.13) is guaranteed to be nonsingular.*

Proof of Lemma 3.1. See Section 3.1.4.1 (Appendix A). ■

3.1.2 Model Justification

As already mentioned, the spatial dependence among the source-relay and relay-destination channel magnitudes (due to shadowing) is described via Gudmundson's model [58] (position related component in (3.14)), which has been very popular in the literature and also experimentally verified [56, 58, 59]. Second, the Laplacian type of temporal dependence among the same groups of channel magnitudes also constitutes a reasonable choice, in the sense that channel magnitudes are expected to be significantly correlated only for small time lags, whereas, for larger time lags, such dependence should decay at a fast rate. For an experimental justification of the adopted model, see, for instance, [60]. Of course, one could use any other positive (semi)definite kernel, multiplying Gudmundson's spatial correlation exponential kernel, without changing the statement and proof of Lemma 3.1. Third, the incorporation of the spherical/isotropic BS correlation term in our proposed general model (in (3.15)) can be justified by the

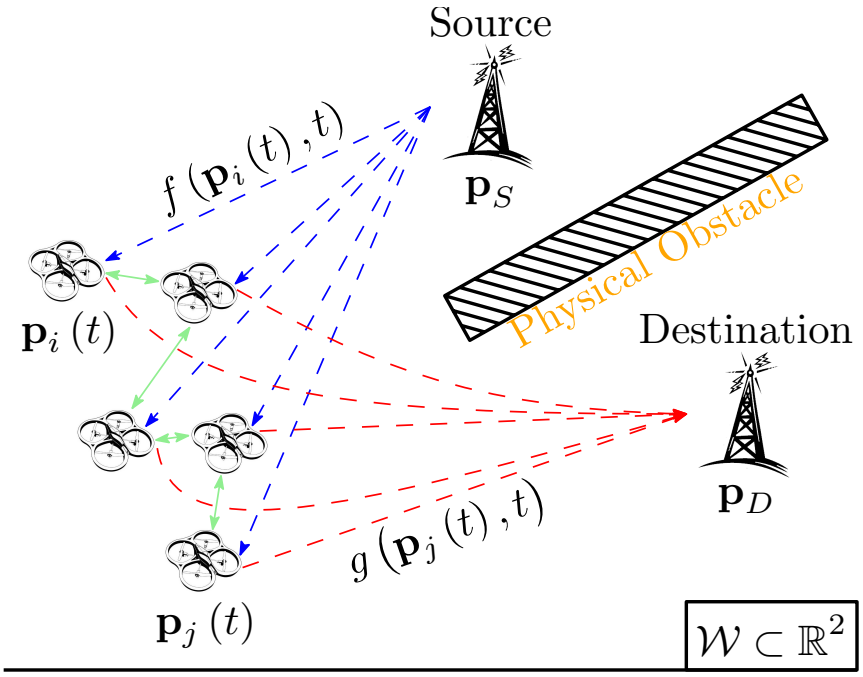


Figure 3.1: A case where source-relay and relay-destination links are likely to be correlated.

the existence of important cases where the source and destination might be close to each other and yet no direct link may exist between them, as least as far as relay beamforming is concerned, which is the main ultimate focus of this work. See, for instance, Fig. 3.1, where a "large" physical obstacle makes the direct communication between the source and the destination impossible. Then, relay beamforming can be exploited in order to enable efficient communication between the source and the destination, making intelligent use of the available resources, in order to improve or maintain a certain QoS in the network. In such cases, however, it is very likely that the shadowing parts of the source-relay and relay-destination links will be spatially and/or temporally correlated among each other, since shadowing is very much affected by the spatial characteristics of the terrain, which, in such cases, is common for both beamforming phases (recall that an AF policy is considered). Of course, by taking the BS station correlation $\delta \to 0$, one recovers the generic/trivial case where the source-relay and relay-destination links are *mutually independent*.

3.1.3 Extensions & Some Technical Considerations

It should be also mentioned that our general description of the wireless channel as a spatiotemporal Gaussian field, does not limit the covariance matrix $\mathbf{\Sigma}$ to be formed as in (3.18); other choices for $\mathbf{\Sigma}$ will work fine in our subsequent developments, as long as, *for each fixed* $t \in \mathbb{N}_{N_T}^+$, some mild conditions on the *spatial interactions* of the fields $\sigma_{S(D)}(\mathbf{p}, t)$ and $\xi_{S(D)}(\mathbf{p}, t)$, are satisfied. In what follows, we consider only the source-relay fields $\sigma_S(\mathbf{p}, t)$ and $\xi_S(\mathbf{p}, t)$. The same arguments hold for the relay-destination fields $\sigma_D(\mathbf{p}, t)$ and $\xi_D(\mathbf{p}, t)$, in direct correspondence.

Fix $t \in \mathbb{N}_{N_T}^+$. Recall that, so far, we have defined the statistical behavior of both $\sigma_S(\mathbf{p}, t)$ and $\xi_S(\mathbf{p}, t)$ *only on a per-node basis*. However, since the spatiotemporal statistical model introduced in Section 3.1.1 is assumed to be valid for any possible trajectory of the relays in $\mathcal{S}^R \times \mathbb{N}_{N_T}^+$, each relay is allowed to be anywhere in $\mathcal{S}$, at each time slot t. This statistical construction *induces* the statistical structure (the *laws*) of both fields $\sigma_S(\mathbf{p}, t)$ and $\xi_S(\mathbf{p}, t)$ on $\mathcal{S}$.

As far as $\sigma_S(\mathbf{p}, t)$ is concerned, it is straightforward to see that it constitutes a Gaussian process with zero mean, and a continuous and *isotropic* covariance kernel $\mathbf{\Sigma}_\sigma : \mathbb{R}^2 \to \mathbb{R}$, defined as

$$\mathbf{\Sigma}_\sigma(\boldsymbol{\tau}) \triangleq \eta^2 \exp\left(-\frac{\|\boldsymbol{\tau}\|_2}{\beta}\right), \tag{3.19}$$

where $\boldsymbol{\tau} \triangleq \mathbf{p} - \mathbf{q} \geq 0$, for all $(\mathbf{p}, \mathbf{q}) \in \mathcal{S}^2$, which agrees with the model introduced in (3.14), for $k \equiv l$ (Gudmundson's model). Thus, $\sigma_S(\mathbf{p}, t)$ is a well defined random field.

However, this is not the case with $\xi_S(\mathbf{p}, t)$. Under no additional restrictions, $\xi_S(\mathbf{p}, t)$ and $\xi_S(\mathbf{q}, t)$ are implicitly assumed to be independent for all $(\mathbf{p}, \mathbf{q}) \in \mathcal{S}^2$, such that $\mathbf{p} \neq \mathbf{q}$. Thus, we are led to consider $\xi_S(\mathbf{p}, t)$ as a zero-mean white process in continuous space. However, it is well known that such a process is technically problematic in a measure theoretic framework. Nevertheless, we may observe that it is *not* actually essential to characterize the covariance structure of $\xi_S(\mathbf{p}, t)$ *for all* $(\mathbf{p}, \mathbf{q}) \in \mathcal{S}^2$, with $\mathbf{p} \neq \mathbf{q}$. This is due to the fact that, at each time slot $t \in \mathbb{N}_{N_T}^+$, it is *physically impossible* for any two relays to be arbitrarily close to each other. We may thus make the following simple assumption on the positions of the network nodes/sensors/relays, at each time

slot $t \in \mathbb{N}_{N_T}^{+}$.

Assumption 3.1. (Relay Separation) *There exists an $\varepsilon_{MF} > 0$, such that, for all $t \in \mathbb{N}_{N_T}^{+}$ and any ensemble of relay positions at time slot t, $\{\mathbf{p}_i(t)\}_{i \in \mathbb{N}_R^{+}}$, it is true that*

$$\inf_{\substack{(i,j) \in \mathbb{N}_R^{+} \times \mathbb{N}_R^{+} \\ \text{with } i \neq j}} \left\| \mathbf{p}_i(t) - \mathbf{p}_j(t) \right\|_2 > \varepsilon_{MF}. \tag{3.20}$$

Assumption 3.1 simply states that, at each $t \in \mathbb{N}_{N_T}^{+}$, all relays are at least ε_{MF} distance units apart from each other. If this constraint is satisfied, then, without any loss of generality, we may define $\xi_S(\mathbf{p}, t)$ as a Gaussian field with zero mean, and with *any* continuous, isotropic (say) covariance kernel $\mathbf{\Sigma}_\xi : \mathbb{R}^2 \to \mathbb{R}$, which satisfies

$$\mathbf{\Sigma}_\xi(\boldsymbol{\tau}) \triangleq \begin{cases} \sigma_\xi^2, & \text{if } \boldsymbol{\tau} \equiv \mathbf{0} \\ 0, & \text{if } \|\boldsymbol{\tau}\|_2 \geq \varepsilon_{MF} \end{cases}, \tag{3.21}$$

and is arbitrarily defined otherwise. A simple example is the spherical, compactly supported kernel with width ε_{MF}, defined as [61]

$$\frac{\mathbf{\Sigma}_o(\boldsymbol{\tau})}{\sigma_\xi^2} \triangleq \begin{cases} 1 - \frac{3}{2}\frac{\|\boldsymbol{\tau}\|_2}{\varepsilon_{MF}} + \frac{1}{2}\left(\frac{\|\boldsymbol{\tau}\|_2}{\varepsilon_{MF}}\right)^3, & \text{if } \|\boldsymbol{\tau}\|_2 < \varepsilon_{MF} \\ 0, & \text{if } \|\boldsymbol{\tau}\|_2 \geq \varepsilon_{MF} \end{cases}. \tag{3.22}$$

Of course, across (discrete) time slots, $\xi_S(\mathbf{p}, t)$ inherits whiteness without any technical issue.

We should stress that the above assumptions are made for technical reasons and will be transparent in the subsequent analysis, as long as the mild constraint (3.20) is satisfied; from the perspective of the relays, all evaluations of $\xi_S(\mathbf{p}, t)$, at each time slot, will be independent to each other. And, of course, ε_{MF} may be chosen small enough, such that (3.20) is satisfied virtually always, assuming that the relays are sufficiently far apart from each other, and/or that, at each time slot t, their new positions are relatively close to their old positions, at time slot $t-1$.

Based on the explicit statistical description of $\sigma_S(\mathbf{p}, t)$ and $\xi_S(\mathbf{p}, t)$ presented above, we now additionally demand that both are spatial fields with *(everywhere) continuous sample paths*. Equivalently, we demand that, for every $\omega \in \Omega$, $\sigma_S(\omega, \mathbf{p}, t) \in \mathsf{C}(\mathcal{S})$

and $\xi_S(\omega, \mathbf{p}, t) \in \mathsf{C}(\mathcal{S})$, where $\mathsf{C}(\mathcal{A})$ denotes the set of continuous functions on some qualifying set $\mathcal{A}$. Sample path continuity of stationary Gaussian fields may be guaranteed under mild conditions on the respective lag-dependent covariance kernel, as the following result suggests, however in a, slightly weaker, *almost everywhere* sense.

Theorem 3.1. (*a.e.*-**Continuity of Gaussian Fields [62–64])** *Let $X(\boldsymbol{s})$, $\boldsymbol{s} \in \mathbb{R}^N$, be a real-valued, zero-mean, stationary Gaussian random field with a continuous covariance kernel $\boldsymbol{\Sigma}_X : \mathbb{R}^N \to \mathbb{R}$. Suppose that there exist constants $0 < c < +\infty$ and $\varepsilon, \zeta > 0$, such that*

$$1 - \frac{\boldsymbol{\Sigma}_X(\boldsymbol{\tau})}{\boldsymbol{\Sigma}_X(\mathbf{0})} \leq \frac{c}{\left|\log\left(\|\boldsymbol{\tau}\|_2\right)\right|^{1+\varepsilon}}, \tag{3.23}$$

for all $\boldsymbol{\tau} \in \left\{ \boldsymbol{x} \in \mathbb{R}^N \middle| \|\boldsymbol{x}\|_2 < \zeta \right\}$. Then, $X(\boldsymbol{s})$ is $\mathcal{P}$-almost everywhere sample path continuous, or, equivalently, $\mathcal{P} - a.e.$-continuous, on every compact subset $\mathcal{K} \subset \mathbb{R}^N$ and, therefore, on $\mathbb{R}^N$ itself. Additionally, $X(\boldsymbol{s})$ is bounded, $\mathcal{P}$-almost everywhere, as well.

Utilizing Theorem 3.1 and generically assuming that $\boldsymbol{\Sigma}_\xi \triangleq \boldsymbol{\Sigma}_o$, it is possible to show that both fields $\sigma_S(\mathbf{p}, t)$ and $\xi_S(\mathbf{p}, t)$ satisfy the respective conditions and thus, that both fields are *a.e.*-continuous on $\mathcal{S}$. For $\sigma_S(\mathbf{p}, t)$, the reader is referred to ([63], Example 2.2). Of course, instead of $\boldsymbol{\Sigma}_\sigma$, any other kernel may be considered, as long as the condition Theorem 3.1 is satisfied.

As far as $\xi_S(\mathbf{p}, t)$ is concerned, let us choose $\varepsilon \equiv 1$ and $\zeta \equiv 1$. We thus need to show that, for every $\tau \triangleq \|\boldsymbol{\tau}\|_2 \in [0, 1)$, it holds that

$$1 - \frac{\boldsymbol{\Sigma}_o(\boldsymbol{\tau})}{\boldsymbol{\Sigma}_o(\mathbf{0})} \leq \frac{c}{\left(\log\left(\|\boldsymbol{\tau}\|_2\right)\right)^2}, \tag{3.24}$$

or, equivalently,

$$1 - \left(1 - \frac{3}{2}\frac{\tau}{\varepsilon_{MF}} + \frac{1}{2}\left(\frac{\tau}{\varepsilon_{MF}}\right)^3\right) \mathbb{1}_{\{\tau < \varepsilon_{MF}\}} \leq \frac{c}{\left(\log(\tau)\right)^2}, \tag{3.25}$$

for some finite, positive constant c. We first consider the case where $1 > \tau \geq \varepsilon_{MF} > 0$ (whenever $\varepsilon_{MF} < 1$, of course). We then have

$$1 \leq \frac{\left(\log(\varepsilon_{MF})\right)^2}{\left(\log(\tau)\right)^2} \triangleq \frac{c_1}{\left(\log(\tau)\right)^2}, \tag{3.26}$$

easily verifying the condition required by Theorem 3.1. Now, when $0 \leq \tau < \min\{\varepsilon_{MF}, 1\}$, it is easy to see that there exists a finite $c_2 > 0$, such that

$$\tau \leq \frac{c_2}{(\log(\tau))^2}. \tag{3.27}$$

If $\tau \equiv 0$, then the inequality above holds for any choice of c_2. If $\tau > 0$, define a function $h : (0,1) \to \mathbb{R}_+$, as

$$h(\tau) \triangleq \tau (\log(\tau))^2. \tag{3.28}$$

By a simple first derivative test, it follows that

$$\begin{aligned} h(\tau) &\leq \max_{\tau \in (0,1)} h(\tau) \\ &\equiv h(\exp(-2)) \\ &\equiv 4\exp(-2), \quad \forall \tau \in (0,1). \end{aligned} \tag{3.29}$$

Consequently, (3.27) is (loosely) satisfied for all $\tau \in [0, \min\{\varepsilon_{MF}, 1\}) \subseteq (0,1)$, by choosing $c_2 \equiv 4\exp(-2)$. Now, observe that

$$\frac{3}{2}\frac{\tau}{\varepsilon_{MF}} - \frac{1}{2}\left(\frac{\tau}{\varepsilon_{MF}}\right)^3 < \frac{3}{2}\frac{\tau}{\varepsilon_{MF}} \leq \frac{3c_2}{2\varepsilon_{MF}(\log(\tau))^2}. \tag{3.30}$$

Finally, simply choose

$$\begin{aligned} c &\equiv \max\left\{c_1, \frac{3c_2}{2\varepsilon_{MF}}\right\} \\ &\equiv \max\left\{(\log(\varepsilon_{MF}))^2, \frac{6\exp(-2)}{\varepsilon_{MF}}\right\} < +\infty, \end{aligned} \tag{3.31}$$

which immediately implies (3.25). Therefore, we have shown that, if we choose $\mathbf{\Sigma}_\xi \equiv \mathbf{\Sigma}_o$, then, for any fixed, but *arbitrarily small* $\varepsilon_{MF} > 0$, the spatial field $\xi_S(\mathbf{p}, t)$ will also be almost everywhere sample path continuous.

Observe that, via the analysis above, sample path continuity of the involved fields can be ascertained, but only in the only almost everywhere sense. Nevertheless, it easy to show that there always exist everywhere sample path continuous fields $\widetilde{\sigma}_S(\mathbf{p}, t)$ and $\widetilde{\xi}_S(\mathbf{p}, t)$, which are *indistinguishable* from $\sigma_S(\mathbf{p}, t)$ and $\xi_S(\mathbf{p}, t)$, respectively [65]. Therefore, there is absolutely no loss of generality if we take both $\sigma_S(\mathbf{p}, t)$ and $\xi_S(\mathbf{p}, t)$ to be sample path continuous, *everywhere* in Ω, and we will do so, hereafter.

Sample path continuity of all fields $\sigma_{S(D)}(\mathbf{p},t)$ and $\xi_{S(D)}(\mathbf{p},t)$ will be essential in Chapter 4, where we rigorously discuss optimality of our proposed relay motion control framework, with special focus on the relay beamforming problem.

We close this section by discussing, in some more detail, the temporal properties of the *evaluations* of the fields $\sigma_S(\mathbf{p},t)$ and $\sigma_D(\mathbf{p},t)$ at any *deterministic* set of N (say) positions $\{\mathbf{p}_i \in \mathcal{S}\}_{i\in\mathbb{N}_N^+}$, *same* across all N_T time slots. This results in the zero-mean, stationary temporal Gaussian process

$$\boldsymbol{C}(t) \triangleq \left[\{\sigma_S(\mathbf{p}_i,t)\}_{i\in\mathbb{N}_N^+}\ \{\sigma_D(\mathbf{p}_i,t)\}_{i\in\mathbb{N}_N^+}\right]^{\boldsymbol{T}} \in \mathbb{R}^{2N\times 1}, \quad t\in\mathbb{N}_{N_T}^+, \tag{3.32}$$

with *matrix covariance kernel* $\boldsymbol{\Sigma}_{\boldsymbol{C}}:\mathbb{Z}\to\mathbb{S}_+^{2N}$, defined, under the specific spatiotemporal model considered, as

$$\boldsymbol{\Sigma}_{\boldsymbol{C}}(\nu) \triangleq \exp\left(-\frac{|\nu|}{\gamma}\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}} \in \mathbb{S}_+^{2N}, \tag{3.33}$$

where $\nu \triangleq t-s$, for all $(t,s)\in\mathbb{N}_{N_T}^+\times\mathbb{N}_{N_T}^+$,

$$\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}} \triangleq \begin{bmatrix} 1 & \kappa \\ \kappa & 1 \end{bmatrix} \circledcirc \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{C}} \in \mathbb{S}_+^{2N}, \tag{3.34}$$

$$\kappa \triangleq \exp\left(-\frac{\|\mathbf{p}_S-\mathbf{p}_D\|_2}{\delta}\right) < 1, \tag{3.35}$$

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{C}}(i,j) \triangleq \boldsymbol{\Sigma}_{\sigma}(\mathbf{p}_i-\mathbf{p}_j), \quad \forall\,(i,j)\in\mathbb{N}_N^+\times\mathbb{N}_N^+, \tag{3.36}$$

and with "$\circledcirc$" denoting the operator of the Kronecker product. Then, the following result is true.

Theorem 3.2. ($\boldsymbol{C}(t)$ is Markov) *For any deterministic, time invariant set of points $\{\mathbf{p}_i\in\mathcal{S}\}_{i\in\mathbb{N}_N^+}$, the vector process $\boldsymbol{C}(t)\in\mathbb{R}^{2N\times 1}$, $t\in\mathbb{N}_{N_T}^+$, as defined in (3.32)-(3.36), may be represented as a stable order-1 vector autoregression, satisfying the linear stochastic difference equation*

$$\boldsymbol{X}(t) \equiv \varphi\boldsymbol{X}(t-1)+\boldsymbol{W}(t), \quad t\in\mathbb{N}_{N_T}^+, \tag{3.37}$$

where

$$\varphi \triangleq \exp(-1/\gamma) < 1, \tag{3.38}$$

$$\boldsymbol{X}(0) \sim \mathcal{N}\left(\mathbf{0},\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}\right) \quad \textit{and} \tag{3.39}$$

$$\boldsymbol{W}(t) \overset{i.i.d.}{\sim} \mathcal{N}\left(\mathbf{0}, \left(1-\varphi^2\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}\right), \quad \forall t \in \mathbb{N}_{N_T}^{+}. \tag{3.40}$$

In particular, $\boldsymbol{C}(t)$ *is Markov.*

Proof of Theorem 3.2. The proof is a standard exercise in time series; see Section 3.1.4.2 (Appendix B). ■

From a practical point of view, Theorem 3.2 is extremely valuable. Specifically, the Markovian representation of $\boldsymbol{C}(t)$ may be employed in order to *efficiently simulate* the spatiotemporal paths of the communication channel on any finite, but arbitrarily fine grid. This is important, since it allows detailed *numerical evaluation* of all methods developed in this work. Theorem 3.2 also reveals that the channel model we have considered actually agrees with experimental results presented in, for instance, [60, 66], which show that autoregressive processes constitute an adequate model for stochastically describing temporal correlations among wireless communication links.

Remark 3.1. Unfortunately, to the best of our knowledge, the channel process along a specific relay trajectory, presented in Section 3.1.1, where the positions of the relays are allowed to vary across time slots is no longer stationary and may not be shown to satisfy the Markov Property. Therefore, in our analysis presented hereafter, we regard the aforementioned process as a general, nonstationary Gaussian process. All inference results presented below are based on this generic representation. ■

Remark 3.2. For simplicity, all motion control problems in this work are formulated on the plane (some subset of $\mathbb{R}^2$). This means that any motion of the relays of the network along the third dimension of the space is indifferent to our channel model. Nevertheless, under appropriate (based on the requirements discussed above) assumptions concerning 3D wireless channel modeling, all subsequent arguments would hold in exactly the same fashion when fully unconstrained motion in $\mathbb{R}^3$ is assumed to affect the quality of the wireless channel. ■

3.1.4 Appendices

3.1.4.1 Appendix A: Proof of Lemma 3.1

In the following, we will rely on an *incremental* construction of $\mathbf{\Sigma}$. Initially, consider the matrix

$$\widetilde{\mathbf{\Sigma}} \triangleq \begin{bmatrix} \widetilde{\mathbf{\Sigma}}(1,1) & \widetilde{\mathbf{\Sigma}}(1,2) & \dots & \widetilde{\mathbf{\Sigma}}(1,N_T) \\ \widetilde{\mathbf{\Sigma}}(2,1) & \widetilde{\mathbf{\Sigma}}(2,2) & \dots & \widetilde{\mathbf{\Sigma}}(2,N_T) \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\mathbf{\Sigma}}(N_T,1) & \widetilde{\mathbf{\Sigma}}(N_T,2) & \cdots & \widetilde{\mathbf{\Sigma}}(N_T,N_T) \end{bmatrix} \in \mathbb{S}^{RN_T}, \tag{3.41}$$

where, for each combination $(k,l) \in \mathbb{N}_{N_T}^{+} \times \mathbb{N}_{N_T}^{+}$, $\widetilde{\mathbf{\Sigma}}(k,l) \in \mathbb{S}^{R}$, with

$$\begin{aligned} \widetilde{\mathbf{\Sigma}}(k,l)(i,j) &\triangleq \widetilde{\mathbf{\Sigma}}\left(\mathbf{p}_i(k), \mathbf{p}_j(l)\right) \\ &\triangleq \eta^2 \exp\left(-\frac{\left\|\mathbf{p}_i(k) - \mathbf{p}_j(l)\right\|_2}{\beta}\right), \end{aligned} \tag{3.42}$$

for all $(i,j) \in \mathbb{N}_{R}^{+} \times \mathbb{N}_{R}^{+}$. By construction, $\widetilde{\mathbf{\Sigma}}$ is positive semidefinite, because the well known *exponential kernel* $\widetilde{\mathbf{\Sigma}} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_{++}$ defined above is positive (semi)definite.

Next, define the positive definite matrix

$$\mathbf{K} \triangleq \begin{bmatrix} 1 & \kappa \\ \kappa & 1 \end{bmatrix}, \quad \text{with} \tag{3.43}$$

$$\kappa \triangleq \exp\left(-\frac{\|\mathbf{p}_S - \mathbf{p}_D\|_2}{\delta}\right) < 1 \tag{3.44}$$

and consider the *Tracy-Singh* type of product of $\mathbf{K}$ and $\widetilde{\mathbf{\Sigma}}$

$$\widetilde{\mathbf{\Sigma}}_{\mathbf{K}} \triangleq \mathbf{K} \circ \widetilde{\mathbf{\Sigma}} \triangleq \begin{bmatrix} \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(1,1) & \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(1,2) & \dots & \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(1,N_T) \\ \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(2,1) & \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(2,2) & \dots & \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(2,N_T) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(N_T,1) & \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(N_T,2) & \cdots & \mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(N_T,N_T) \end{bmatrix} \in \mathbb{S}^{2RN_T}, \tag{3.45}$$

where "$\circledcirc$" denotes the operator of the Kronecker product. Then, for each $(k,l) \in \mathbb{N}_{N_T}^{+} \times \mathbb{N}_{N_T}^{+}$, we have

$$\mathbf{K} \circledcirc \widetilde{\mathbf{\Sigma}}(k,l) \equiv \begin{bmatrix} \widetilde{\mathbf{\Sigma}}(k,l) & \kappa\widetilde{\mathbf{\Sigma}}(k,l) \\ \kappa\widetilde{\mathbf{\Sigma}}(k,l) & \widetilde{\mathbf{\Sigma}}(k,l) \end{bmatrix} \in \mathbb{S}^{2R}. \tag{3.46}$$

It is easy to show that $\widetilde{\mathbf{\Sigma}}_{\mathbf{K}}$ is positive semidefinite, that is, in $\mathbb{S}_{+}^{2RN_T}$. First, via a simple inductive argument, it can be shown that, for compatible matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$,

$$(\mathbf{AB}) \circ (\mathbf{CD}) \equiv (\mathbf{A} \circ \mathbf{C}) (\mathbf{B} \circ \mathbf{D}) . \tag{3.47}$$

Also, for compatible $\mathbf{A}, \mathbf{B}$, it is true that $(\mathbf{A} \circ \mathbf{B})^{\boldsymbol{T}} \equiv \mathbf{A}^{\boldsymbol{T}} \circ \mathbf{B}^{\boldsymbol{T}}$. Since $\mathbf{K}$ and $\widetilde{\mathbf{\Sigma}}$ are symmetric, consider their spectral decompositions $\mathbf{K} \equiv \mathbf{U}_{\mathbf{K}} \mathbf{\Lambda}_{\mathbf{K}} \mathbf{U}_{\mathbf{K}}^{\boldsymbol{T}}$ and $\widetilde{\mathbf{\Sigma}} \equiv \mathbf{U}_{\widetilde{\mathbf{\Sigma}}} \mathbf{\Lambda}_{\widetilde{\mathbf{\Sigma}}} \mathbf{U}_{\widetilde{\mathbf{\Sigma}}}^{\boldsymbol{T}}$. Given the identities stated above, we may write

$$\begin{aligned} \widetilde{\mathbf{\Sigma}}_{\mathbf{K}} \equiv \mathbf{K} \circ \widetilde{\mathbf{\Sigma}} &\equiv \left(\mathbf{U}_{\mathbf{K}} \mathbf{\Lambda}_{\mathbf{K}} \mathbf{U}_{\mathbf{K}}^{\boldsymbol{T}} \right) \circ \left(\mathbf{U}_{\widetilde{\mathbf{\Sigma}}} \mathbf{\Lambda}_{\widetilde{\mathbf{\Sigma}}} \mathbf{U}_{\widetilde{\mathbf{\Sigma}}}^{\boldsymbol{T}} \right) \\ &\equiv \left(\mathbf{U}_{\mathbf{K}} \circ \mathbf{U}_{\widetilde{\mathbf{\Sigma}}} \right) \left(\mathbf{\Lambda}_{\mathbf{K}} \circ \mathbf{\Lambda}_{\widetilde{\mathbf{\Sigma}}} \right) \left(\mathbf{U}_{\mathbf{K}}^{\boldsymbol{T}} \circ \mathbf{U}_{\widetilde{\mathbf{\Sigma}}}^{\boldsymbol{T}} \right) \\ &\equiv \left(\mathbf{U}_{\mathbf{K}} \circ \mathbf{U}_{\widetilde{\mathbf{\Sigma}}} \right) \left(\mathbf{\Lambda}_{\mathbf{K}} \circ \mathbf{\Lambda}_{\widetilde{\mathbf{\Sigma}}} \right) \left(\mathbf{U}_{\mathbf{K}} \circ \mathbf{U}_{\widetilde{\mathbf{\Sigma}}} \right)^{\boldsymbol{T}} , \end{aligned} \tag{3.48}$$

where $\left(\mathbf{U}_{\mathbf{K}} \circ \mathbf{U}_{\widetilde{\mathbf{\Sigma}}} \right) \left(\mathbf{U}_{\mathbf{K}}^{\boldsymbol{T}} \circ \mathbf{U}_{\widetilde{\mathbf{\Sigma}}}^{\boldsymbol{T}} \right) \equiv \left(\mathbf{U}_{\mathbf{K}} \mathbf{U}_{\mathbf{K}}^{\boldsymbol{T}} \right) \circ \left(\mathbf{U}_{\widetilde{\mathbf{\Sigma}}} \mathbf{U}_{\widetilde{\mathbf{\Sigma}}}^{\boldsymbol{T}} \right) \equiv \mathbf{I}_2 \circ \mathbf{I}_{RN_T} \equiv \mathbf{I}_{2RN_T}$, and where the matrix $\mathbf{\Lambda}_{\mathbf{K}} \circ \mathbf{\Lambda}_{\widetilde{\mathbf{\Sigma}}}$ is easily shown to be diagonal and with nonnegative elements. Thus, since (3.48) constitutes a valid spectral decomposition for $\widetilde{\mathbf{\Sigma}}_{\mathbf{K}}$, it follows that $\widetilde{\mathbf{\Sigma}}_{\mathbf{K}} \in \mathbb{S}_{+}^{2RN_T}$.

As a last step, let $\mathbf{E} \in \mathbb{S}^{N_T}$, such that

$$\mathbf{E}(k,l) \triangleq \exp\left(-\frac{|k-l|}{\gamma} \right), \tag{3.49}$$

for all $(k,l) \in \mathbb{N}_{N_T}^{+} \times \mathbb{N}_{N_T}^{+}$. Again, $\mathbf{E}$ is positive semidefinite, because the well known *Laplacian kernel* is positive (semi)definite. Consider the matrix

$$\widetilde{\mathbf{\Sigma}}_{\mathbf{E}} \triangleq (\mathbf{E} \circledcirc \mathbf{1}_{2R \times 2R}) \odot \widetilde{\mathbf{\Sigma}}_{\mathbf{K}} \in \mathbb{S}^{2RN_T}, \tag{3.50}$$

where "$\odot$" denotes the operator of the Schur-Hadamard product. Of course, since the matrix $\mathbf{1}_{2R \times 2R}$ is rank-1 and positive semidefinite, $\mathbf{E} \circledcirc \mathbf{1}_{2R \times 2R}$ will be positive semidefinite as well. Consequently, by the Schur Product Theorem, $\widetilde{\mathbf{\Sigma}}_{\mathbf{E}}$ will also be positive semidefinite. Finally, observe that

$$\mathbf{\Sigma} \equiv \widetilde{\mathbf{\Sigma}}_{\mathbf{E}} + \sigma_{\xi}^2 \mathbf{I}_{2RN_T}, \tag{3.51}$$

from where it follows that $\mathbf{\Sigma} \in \mathbb{S}_{++}^{2RN_T}$, whenever $\sigma_{\xi}^2 \neq 0$. Our claims follow. ■

3.1.4.2 Appendix B: Proof of Theorem 3.2

Obviously, the vector process $\boldsymbol{X}(t)$ is Gaussian with mean zero. This is straightforward to show. Therefore, what remains is, simply, to verify that the covariance structure of $\boldsymbol{X}(t)$ is the same as that of $\boldsymbol{C}(t)$, that is, we need to show that

$$\mathbb{E}\left\{\boldsymbol{X}(s)\,\boldsymbol{X}^{\boldsymbol{T}}(t)\right\} \equiv \mathbb{E}\left\{\boldsymbol{C}(s)\,\boldsymbol{C}^{\boldsymbol{T}}(t)\right\}, \tag{3.52}$$

for all $(s,t) \in \mathbb{N}_{N_T}^{+} \times \mathbb{N}_{N_T}^{+}$.

First, consider the case where $s \equiv t$. Then, we have

$$\begin{aligned}\mathbb{E}\left\{\boldsymbol{X}(s)\,\boldsymbol{X}(t)\right\} &\equiv \mathbb{E}\left\{\boldsymbol{X}(t)\,\boldsymbol{X}^{\boldsymbol{T}}(t)\right\} \\ &= \varphi^2\mathbb{E}\left\{\boldsymbol{X}(t-1)\,\boldsymbol{X}^{\boldsymbol{T}}(t-1)\right\} + \left(1-\varphi^2\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}.\end{aligned} \tag{3.53}$$

Observe, though, that, similarly to the scalar order-1 autoregressive model, the quantity

$$\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}} \equiv \mathbb{E}\left\{\boldsymbol{X}(0)\,\boldsymbol{X}^{\boldsymbol{T}}(0)\right\} \tag{3.54}$$

is a fixed point of the previously stated recursion for $\mathbb{E}\left\{\boldsymbol{X}(t)\,\boldsymbol{X}^{\boldsymbol{T}}(t)\right\}$. Therefore, it is true that

$$\mathbb{E}\left\{\boldsymbol{X}(t)\,\boldsymbol{X}^{\boldsymbol{T}}(t)\right\} \equiv \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}} \equiv \boldsymbol{\Sigma}_{\boldsymbol{C}}(0) \equiv \mathbb{E}\left\{\boldsymbol{C}(t)\,\boldsymbol{C}^{\boldsymbol{T}}(t)\right\}, \tag{3.55}$$

which the desired result.

Now, consider the case where $s < t$. Then, it may be easily shown that

$$\mathbb{E}\left\{\boldsymbol{X}(s)\,\boldsymbol{X}(t)\right\} \equiv \varphi^2\mathbb{E}\left\{\boldsymbol{X}(s-1)\,\boldsymbol{X}^{\boldsymbol{T}}(t-1)\right\} + \varphi\mathbb{E}\left\{\boldsymbol{W}(s)\,\boldsymbol{X}^{\boldsymbol{T}}(t-1)\right\}. \tag{3.56}$$

Let us consider the second term on the RHS of (3.56). Expanding the recursion, we may write

$$\begin{aligned}\varphi\mathbb{E}\left\{\boldsymbol{W}(s)\,\boldsymbol{X}^{\boldsymbol{T}}(t-1)\right\} &\equiv \varphi\mathbb{E}\left\{\boldsymbol{W}(s)\left(\varphi\boldsymbol{X}^{\boldsymbol{T}}(t-2) + \boldsymbol{W}^{\boldsymbol{T}}(t-1)\right)\right\} \\ &\equiv \varphi^2\mathbb{E}\left\{\boldsymbol{W}(s)\,\boldsymbol{X}^{\boldsymbol{T}}(t-2)\right\} \\ &\vdots \\ &\equiv \varphi^{t-s}\mathbb{E}\left\{\boldsymbol{W}(s)\,\boldsymbol{W}^{\boldsymbol{T}}(s)\right\} \\ &\equiv \varphi^{t-s}\left(1-\varphi^2\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}.\end{aligned} \tag{3.57}$$

We observe that this term depends only on the lag $t-s$. Thus, it is true that

$$\begin{aligned}\mathbb{E}\left\{\boldsymbol{X}\left(s\right)\boldsymbol{X}\left(t\right)\right\} &\equiv \varphi^{2}\mathbb{E}\left\{\boldsymbol{X}\left(s-1\right)\boldsymbol{X}^{\boldsymbol{T}}\left(t-1\right)\right\}+\varphi^{t-s}\left(1-\varphi^{2}\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}\\ &= \varphi^{2\cdot 2}\mathbb{E}\left\{\boldsymbol{X}\left(s-2\right)\boldsymbol{X}^{\boldsymbol{T}}\left(t-2\right)\right\}+\varphi^{t-s}\left(1-\varphi^{2}\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}\left(1+\varphi^{2}\right)\\ &\vdots\\ &\equiv \varphi^{2s}\mathbb{E}\left\{\boldsymbol{X}\left(0\right)\boldsymbol{X}^{\boldsymbol{T}}\left(t-s\right)\right\}+\varphi^{t-s}\left(1-\varphi^{2}\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}\sum_{i\in\mathbb{N}_{s-1}}\left(\varphi^{2}\right)^{s-1}\\ &= \varphi^{2s}\mathbb{E}\left\{\boldsymbol{X}\left(0\right)\boldsymbol{X}^{\boldsymbol{T}}\left(t-s\right)\right\}+\varphi^{t-s}\left(1-\varphi^{2s}\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}. \qquad (3.58)\end{aligned}$$

Further, we may further expand $\mathbb{E}\left\{\boldsymbol{X}\left(0\right)\boldsymbol{X}^{\boldsymbol{T}}\left(t-s\right)\right\}$ in similar fashion as above, to get that

$$\mathbb{E}\left\{\boldsymbol{X}\left(0\right)\boldsymbol{X}^{\boldsymbol{T}}\left(t-s\right)\right\}\equiv\varphi^{t-s}\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}. \qquad (3.59)$$

Exactly the same arguments may be made for the symmetric case where $t<s$. Therefore, it follows that

$$\begin{aligned}\mathbb{E}\left\{\boldsymbol{X}\left(s\right)\boldsymbol{X}\left(t\right)\right\} &\equiv \varphi^{|t-s|}\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}\\ &\equiv \exp\left(-\frac{|t-s|}{\gamma}\right)\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{C}}\\ &\equiv \boldsymbol{\Sigma}_{\boldsymbol{C}}\left(t-s\right) \qquad (3.60)\end{aligned}$$

for all $(s,t)\in\mathbb{N}_{N_T}^{+}\times\mathbb{N}_{N_T}^{+}$, and we are done. ■

3.2 Hierarchical Spatiotemporal Channel Modeling & Tracking

3.2.1 Introduction

The first basic approach to joint hierarchical spatiotemporal channel (specifically shadowing) tracking and prediction was presented in [67,68], where the use of Channel Gain (CG) maps was advocated as an advantageous alternative to Power Spectral Density (PSD) maps for cooperative spectrum sensing in the context of cognitive radios. Although analytically appealing, the state space model considered in [67,68] for describing the evolution of the wireless channel is rather restrictive; both the temporal dependencies of the shadowing field and its spatial interactions are characterized by purely linear

relationships, focusing mainly on modeling the spatiotemporal variations of the *trend* of the field.

In this work, the descriptive channel parameters (e.g., the path loss exponent, the shadowing power, etc.), referred to here as the *channel state*, are assumed to be temporally varying. Specifically, the whole channel state constitutes a Markov process, with known, but potentially *non stationary, nonlinear* and/or *non Gaussian* transition model. Then, the spatiotemporal evolution of the channel is modeled as a two layer stochastic system, or, in more specific terms, as a Hidden Markov Model (HMM) [12].

Our main contributions are summarized in the following. **1)** Recognizing the intractability of state estimation in partially observable nonlinear systems, we show that grid based approximate nonlinear filtering is meaningfully applicable to the channel state tracking and spatiotemporal channel prediction problems of interest. Due to the relatively small dimension of the channel state, grid based methods constitute excellent approximation candidates for the problems at hand. Exploiting filtered estimates of the channel state, a recursive spatiotemporal predictor of the channel gains (magnitudes) is developed (Theorem 3.4), providing *real time sequential estimates* for the respective CG map, for each sensor in the network. Relevant recursive estimators of the (conditional) variance of the channel gain predictions are also developed, providing the user with an additional measure of estimate uncertainty (Theorem 3.5). **2)** Leveraging the fundamental theory developed in Section 2.2, we provide conditions, under which the proposed channel state tracker, spatiotemporal predictor and conditional variance estimator, briefly described above, are *asymptotically optimal*, in a common, strong sense (Theorems 3.4 and 3.5), providing a unified convergence criterion for all three sequential estimators.

The proposed statistical model describing the joint spatiotemporal behavior of the channel is inspired by [56], where the channel state was assumed to be constant through time and space, therefore constituting a set of fixed but unknown parameters. Moreover, the proposed formulation is more general than [67], since it can deal with complex variations in the channel characteristics, other than linear variations in the shadowing

trend. However, in our state space description of the channel, spatial statistical dependencies are present only in the observations process, whereas in [67], the trend of the shadowing component of the channel, constituting the hidden state, respectively, is *jointly* spatiotemporally colored. Also, here, we will consider the *detrended* problem, which, as stated, is similar to the one treated in [56] (in a non Bayesian framework). This is a good approximation of reality [52–54, 56, 59]. A complete channel model, combining both a non zero spatiotemporally varying shadowing trend in the fashion of [67, 68] with the temporally varying channel parameters advocated here, would result in a non trivial problem in nonlinear estimation and constitutes a subject of future research.

*Note on **notation**: In this work, notation is generic and is not in direct agreement with Section 1.1. This is because the respective channel modeling problem is considered in generic terms. However, it should be easy for the reader to make the necessary connections. Additionally, the reader might observe that some quantities, defined previously in Section 3.1, are refined under another framework. This is intentional and is made for the sake of clarity in the exposition. Essentially, the meaning of those quantities will depend on the particular channel model considered.*

3.2.2 System Model & Problem Formulation

Here, for simplicity, we consider the wireless network illustrated in Fig. 3.2. This network may be seen as either the source-relay or relay-destination *half* of the beamforming network of Section 1.1. As before, the environment is assumed to be a compact planar region $\mathcal{S} \subset \mathbb{R}^2$, where there exists a fixed, stationary antenna at a reference position $\mathbf{p}_{ref} \in \mathcal{S}$, capable of at least information broadcasting. There also exist a set of N single antenna sensors, possibly mobile and located at $\mathbf{p}_i \equiv \mathbf{p}_i(t) \in \mathcal{S}, i \in \mathbb{N}_N^+$, monitoring the channel relative to the reference antenna. These sensors may be a subset of the total nodes in the network and are responsible for the respective channel estimation tasks. The sensors can cooperate, and further, can either communicate with a fusion center (in a centralized setting), or exchange basic messages amongst each other (in a decentralized/infrastructureless scenario) using a low rate dedicated channel. Concerning

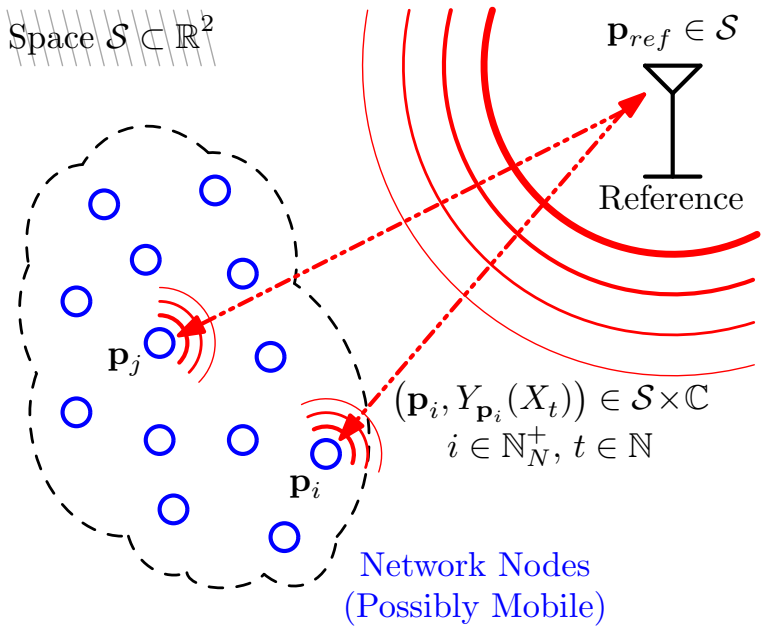


Figure 3.2: The wireless network of interest.

channel modeling, we adopt a flat fading model between each node and the reference antenna. It is additionally assumed that channel reciprocity holds and that all network nodes can perfectly observe their individual channel realizations (e.g. magnitudes and potentially phases) relative to the reference antenna [56]. The channels are modeled as spatially and temporally correlated, discrete time random processes (spatiotemporal random fields), sharing the same channel environment, at least as far as the underlying characteristics of the communication medium are concerned.

As already mentioned, the channel state encompasses statistics of the communication medium, and is here modeled as a multidimensional discrete time stochastic process, evolving in time according to a known statistical model. The channel state is assumed to be hidden from the network nodes; the nodes can observe their respective channel realizations, but they cannot directly observe the characteristics of the mechanism that generates these realizations.

3.2.2.1 Hierarchical Description of the Wireless Channel

Let $X_t \equiv X_t(\omega) \subset \mathbb{R}^{M\times 1}$, $t \in \mathbb{N}$, $\omega \in \Omega$ denote the hidden channel state. As in Section 3.1, *and employing generic notation*, the relative to the reference antenna complex

channel process at each network node $i \in \mathbb{N}_N^+$ can be decomposed as [55]

$$Y_i\left(\mathbf{p}_i\left(t\right), X_t\right) \equiv Y_{\mathbf{p}_i}\left(X_t\right) \equiv \underbrace{Y_{\mathbf{p}_i}^{PL}\left(X_t\right)}_{\text{path loss}} \underbrace{Y_i^{SH}\left(X_t\right)}_{\text{shadowing}} \underbrace{Y_{\mathbf{p}_i}^{MF}\left(t\right)}_{\text{fading}} \exp\left(\mathfrak{J}\frac{2\pi d_i\left(t\right)}{\lambda}\right), \quad (3.61)$$

where $\mathfrak{J} \triangleq \sqrt{-1}$, $\lambda > 0$ denotes the wavelength employed for the communication, and where: **1)** $Y_{\mathbf{p}_i}^{PL}\left(X_t\right) \in \mathbb{R}$ denotes path loss, defined as

$$Y_{\mathbf{p}_i}^{PL}\left(X_t\right) \triangleq \left\|\mathbf{p}_i\left(t\right) - \mathbf{p}_{ref}\right\|_2^{-\ell(X_t)/2} \triangleq \left(d_i\left(t\right)\right)^{-\ell(X_t)/2}, \quad (3.62)$$

where $\ell\left(X_t\right) > 0$ is the state dependent path loss exponent, which is the same for all network nodes. **2)** $Y_i^{SH}\left(X_t\right) \in \mathbb{R}$ denotes the shadowing part of the channel model. Its square, *conditionally on* X_t, constitutes a base-10 log-normal random variable with zero location and scale *depending on* X_t. **3)** $Y_{\mathbf{p}_i}^{MF}\left(t\right) \in \mathbb{C}$ represents multipath fading, which, for simplicity, is assumed to be a *spatiotemporally* white[1], strictly stationary process with fully known statistical description, not associated with X_t, therefore being an unpredictable complex "observation noise". Making the substitution $Y_i\left(\mathbf{p}_i\left(t\right), X_t\right) \leftarrow \exp\left(-\mathfrak{J}2\pi d_i\left(t\right)/\lambda\right) Y_i\left(\mathbf{p}_i\left(t\right), X_t\right)$, we can define the observation of node i in logarithmic scale as

$$\begin{aligned} y_t^i \triangleq 10\log_{10}\left|Y_{\mathbf{p}_i}\left(X_t\right)\right|^2 - 10\mathbb{E}\left\{\log_{10}\left|Y_{\mathbf{p}_i}^{MF}\left(t\right)\right|^2\right\} \\ = -10\ell\left(X_t\right)\log_{10}\left(d_i\left(t\right)\right) + 10\log_{10}\left(Y_i^{SH}\left(X_t\right)\right)^2 \\ + \overline{10\log_{10}\left|Y_{\mathbf{p}_i}^{MF}\left(t\right)\right|^2} \triangleq \alpha_t^i \ell\left(X_t\right) + \sigma_t^i\left(X_t\right) + \xi_t^i, \end{aligned} \quad (3.63)$$

where $\overline{(\cdot)}$ denotes the zero mean version of a random variable. We should emphasize here that by "measurement" or "observation" we refer to the *predictable* component of the channel, which is described in terms of the channel magnitude.

Remark 3.3. In the above, we have assumed that the path loss exponent is spatially invariant over $\mathcal{S}$. This assumption is essential for our subsequent derivations. It should hold in sufficiently small, spatially homogeneous environments, in a *statistical* sense.

[1]See [56] and references therein for arguing about the validity of this assumption. Also, throughout this work, the samples of a discrete time white stochastic process are understood to be independent.

For instance, the path loss exponent is spatially invariant when changes in the spatial characteristics of the environment have a common global effect on the large scale characteristics of the wireless channel. Also see [54, 56] and the references therein. ■

3.2.2.2 Modeling SpatioTemporal Correlations of the Observations

In similar fashion as in [52–54, 56, 59], where relevant experimental verification was also presented, the following further assumptions are made[2]: $\xi_t^i \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_\xi^2\right), \forall i \in \mathbb{N}_N^+$ and $\forall t \in \mathbb{N}$ [57]. Second, conditioned on X_t, $\sigma_t^i\left(X_t\right) \overset{i.d.}{\sim} \mathcal{N}\left(0, \eta^2\left(X_t\right)\right), \forall i \in \mathbb{N}_N^+$. This stems from the fact that $\left(Y_i^{SH}\left(X_t\right)\right)^2$ is (base-10) log-normally distributed. Additionally, it is assumed that the members of the set $\left\{\sigma_t^i\left(X_t\right)\right\}_{i \in \mathbb{N}_N^+}$ constitute *jointly normal, spatially correlated* random variables with conditional on X_t autocorrelation kernel (covariance matrix) [56, 58]

$$\mathbf{\Sigma}_t\left(\boldsymbol{\theta}\left(X_t\right)\right)(i,j) \triangleq \theta_1\left(X_t\right) \exp\left(-\frac{d_{ij}\left(t\right)}{\theta_2\left(X_t\right)}\right), \tag{3.64}$$

for all $(i,j) \in \mathbb{N}_N^+ \times \mathbb{N}_N^+$, where $d_{ij}\left(t\right) \triangleq \left\|\mathbf{p}_i\left(t\right) - \mathbf{p}_j\left(t\right)\right\|_2 \in \mathbb{R}_+$ and $\boldsymbol{\theta}\left(X_t\right) \triangleq \left[\theta_1\left(X_t\right)\ \theta_2\left(X_t\right)\right]^{\boldsymbol{T}}$ with $\theta_1\left(X_t\right) \equiv \eta^2\left(X_t\right)$. As in Section 3.1, this is the Gudmundson's model, where, in this case, the first parameter, $\theta_1\left(X_t\right)$, is the *shadowing power*, controls the variance of the shadowing part of the channel, whereas the second, $\theta_2\left(X_t\right)$, is the *correlation distance*, controls the decay rate of the spatial correlation between the channels for each pair of network nodes. Note that instead of (3.64), any other kernel can be assumed, as long as it satisfies some mild continuity properties (see below).

In order to completely define an overall observation process for all nodes in the network, we may stack the N individual channel processes of (3.63), yielding

$$\mathbf{y}_t \equiv \boldsymbol{\alpha}_t \ell\left(X_t\right) + \boldsymbol{\sigma}_t\left(X_t\right) + \boldsymbol{\xi}_t, \quad \forall t \in \mathbb{N}, \tag{3.65}$$

where $\mathbf{y}_t \in \mathbb{R}^{N \times 1}$, $\boldsymbol{\alpha}_t \in \mathbb{R}^{N \times 1}$, $\boldsymbol{\sigma}_t\left(X_t\right) \in \mathbb{R}^{N \times 1}$ and $\boldsymbol{\xi}_t \in \mathbb{R}^{N \times 1}$ are defined accordingly. The observation process (3.65) can also be rewritten in the canonical form $\mathbf{y}_t \equiv \boldsymbol{\alpha}_t \ell\left(X_t\right) + \sqrt{\mathbf{C}_t\left(X_t\right)}\boldsymbol{u}_t, \forall t \in \mathbb{N}$, where $\boldsymbol{u}_t \equiv \boldsymbol{u}_t\left(\omega\right)$ constitutes a standard Gaussian white noise process and $\mathbf{C}_t\left(X_t\right) \triangleq \mathbf{\Sigma}_t\left(\boldsymbol{\theta}\left(X_t\right)\right) + \sigma_\xi^2 \mathbf{I}_{N \times N} \in \mathcal{D}_{\mathbf{C}}$, with $\mathcal{D}_{\mathbf{C}}$ obviously bounded.

[2] In what follows, "i.d." means "identically distributed" and "i.i.d." means "independent and i.d.".

3.2.2.3 Modeling Temporal Correlations of the State

Let us now concentrate more on the channel state process $X_t \in \mathbb{R}^{M\times 1}$. In this work, we will assume that X_t constitutes a Markov process with *known but nonlinear and (possibly) nonstationary dynamics*, described by a stochastic kernel $\mathcal{K}_t : \mathscr{B}\left(\mathbb{R}^{M\times 1}\right) \times \mathbb{R}^{M\times 1} \mapsto [0,1], t \in \mathbb{N}$ ($\mathscr{B}(\mathcal{A})$ denotes the Borel σ-algebra generated by the set $\mathcal{A}$). Also, we will assume that the state is compactly supported, that is, $\forall t \in \mathbb{N}, X_t \in [a,b]^M \triangleq \mathcal{Z} \subset \mathbb{R}^{M\times 1}$, almost surely. Depending on the available information, instead of using stochastic kernels, we may alternatively assume that $X_t \triangleq f_t(X_{t-1}, W_t) \in \mathcal{Z}, \forall t \in \mathbb{N}$, where, for each t, $f_t : \mathcal{Z} \times \mathcal{W} \overset{a.s.}{\mapsto} \mathcal{Z}$ constitutes a state transition mapping and $W_t \equiv W_t(\omega) \in \mathcal{W} \subseteq \mathbb{R}^{M_W\times 1}$, for $t \in \mathbb{N}$, $\omega \in \Omega$, denotes a white noise process with known measure and state space $\mathcal{W}$.

From now on, without loss of generality, we will drop the subscript "t" from both the stochastic kernels and transition mappings governing X_t, therefore assuming stationarity of the state. In the case of nonstationary nonlinear dynamics, all subsequent conditions on the stochastic mechanism generating X_t (see Section 3.2.3) must hold *for all* $t \in \mathbb{N}$ (that is, for each "mode" of the state X_t). Further, for mathematical simplicity, we will assume that $\ell(X_t) \equiv X_t(1) \in \mathbb{R}$ and $\boldsymbol{\theta}(X_t) \equiv [X_t(2)\ X_t(3)]^{\boldsymbol{T}} \in \mathbb{R}^{M\times 1}$, that is, $M \equiv 3$. From the previous discussion, it follows that the partially observable system defined above can be described by

$$\begin{cases} X_t \,|X_{t-1} \sim \mathcal{K}(X_t \in \mathrm{d}\boldsymbol{x} \,|X_{t-1}) \quad \text{or} \\ X_t \equiv f(X_{t-1}, W_t) \qquad\qquad\qquad\qquad , \ \forall t \in \mathbb{N}, \\ \mathbf{y}_t \equiv \boldsymbol{A}_t X_t + \boldsymbol{\sigma}_t(X_t) + \boldsymbol{\xi}_t \end{cases} \tag{3.66}$$

where $\boldsymbol{A}_t \triangleq \left[\boldsymbol{\alpha}_t\ \mathbf{0}_{N\times(M-1)}\right] \in \mathbb{R}^{N\times M}$.

Remark 3.4. The assumption that the channel state is almost surely compactly supported in $\mathcal{Z}$ is not restrictive. There is no constraint on how large $\mathcal{Z}$ is; it just needs to be compact. In fact, our approach will still be valid as long as the state process lies in some compact set, at least with very high probability. Here, the aforementioned compactness assumption is made mainly for analytical tractability. ■

3.2.2.4 Technical Assumptions

In addition to the above and in favor of supporting our analytical arguments presented in subsequent sections, we make the following mild assumptions on the functional structure of the observation process of (3.66). First, throughout this work, we assume that

$$\lambda_{inf} \equiv \inf_{t\in\mathbb{N}} \inf_{\boldsymbol{x}\in\mathcal{Z}} \lambda_{min}\left(\mathbf{C}_t\left(\boldsymbol{x}\right)\right) > 1, \tag{3.67}$$

a requirement which can always be satisfied by appropriate normalization of the observations. Second, regarding modeling spatial correlations due to shadowing, any choice of autocorrelation kernel is permitted, as long as the resulting covariance matrix $\mathbf{\Sigma}_t : \mathcal{Z} \mapsto \mathcal{D}_{\mathbf{\Sigma}} \subset \mathbb{R}^{N\times N}$ is *elementwise* Lipschitz continuous on $\mathcal{Z}$ with respect to the ℓ_1-norm, uniformly in $\mathbb{N}$ (Assumption 2.2 in Section 2.1). Note that the isotropic autocorrelation previously defined by (3.64) can be easily verified to satisfy the aforementioned Lipschitz assumption, considering the compactness of the state vector.

3.2.2.5 Precise Definition of Problems of Interest

Let us now define the problems of interest in this work in a mathematically precise way. Hereafter, *strict optimality* will be meant to be in the *Minimum Mean Square Sense (MMSE)*. Also, in the following, the natural filtration generated by the causal observation process $\mathbf{y}_t$ is defined as the sequence $\{\mathscr{Y}_t\}_{t\in\mathbb{N}}$, where $\mathscr{Y}_t \triangleq \sigma\left\{\{\mathbf{y}_i\}_{i\in\mathbb{N}_t}\right\}$, for all $t \in \mathbb{N}$.

Problem 3.1. (Sequential Channel State Tracking (SCST)) *Develop a sequential, finite dimensional scheme for (approximately) evaluating the optimal filter or ρ-step predictor of the channel state X_t on the basis of the available channel magnitude observations up to time t, given by*

$$\widehat{X}_{t+\rho} \triangleq \mathbb{E}\left\{ X_{t+\rho} \middle| \mathscr{Y}_t \right\}, \quad \forall t \in \mathbb{N}, \tag{3.68}$$

where $\rho \geq 0$ constitutes the ***prediction horizon****. The computational complexity of the sequential scheme may not grow as more observations become available.*

Problem 3.2. (Sequential Spatiotemporal Channel Prediction (SSCP)) *Develop a sequential, finite dimensional scheme for (approximately) evaluating the optimal*

spatiotemporal predictor of the channel magnitude at position $\mathbf{q} \in \mathbb{R}^2$ *and time* $t+\rho$ *(*$\rho \geq 0$ *is the prediction horizon) given the available channel magnitude observations up to time* t*, expressed as*

$$\widehat{y}_{t+\rho}(\mathbf{q}) \triangleq \mathbb{E}\left\{ y_{t+\rho}(\mathbf{q}) \middle| \mathscr{Y}_t \right\}, \quad \forall t \in \mathbb{N}. \tag{3.69}$$

Again, the computational complexity of the sequential scheme may not grow as more observations become available.

It should be emphasized that the main ingredients rendering the Bayesian solution of the SCST and SSCP problems possible are the assumption of spatial invariance of the path loss exponent in $\mathcal{S}$, as well as modeling the log-shadowing part of the wireless channel as a conditionally Gaussian stochastic process. As a result, the channel measurements at the sensors can be stochastically described by the HMM defined through (3.66), enabling the use of the theory of nonlinear filtering for deriving effective recursive algorithmic schemes allowing temporal tracking/prediction of the channel state, as well as spatiotemporal prediction of the channel itself at unobserved locations.

As we will see later in Section 3.2.4, the SSCP problem can be solved sequentially using the respective sequential solution of the SCST problem. However, unfortunately, it is well known that, except for some very special cases [13–16], the respective nonlinear filtering and prediction problems do not admit any known sequential (in particular, recursive) representation [11,12]. Therefore, in order to solve the SCST problem defined above, one typically has to rely on carefully designed and robust approximations to the problem of nonlinear filtering of Markov processes in discrete time, focusing on the class of systems (HMMs) described by (3.66). This is exactly where the theory developed in Chapter 2 may be exploited, providing theoretical guarantees for correctness and stability.

3.2.3 Grid-Based Approximate Filtering: Preliminaries

In the following, we present some preliminary results in asymptotically optimal, approximate recursive filtering of Markov processes. These results were previously developed in Section 2.2, and are presented here for immediate reference and completeness.

3.2.3.1 Uniform State Quantizations

In the previous section, we have assumed that $X_t \in \mathcal{Z} \equiv [a,b]^M, \forall t \in \mathbb{N}, a.s.$, where, geometrically, $\mathcal{Z}$ constitutes an M-hypercube, representing the compact set of support of the state X_t. In the fashion of Section 2.2, let us discretize $\mathcal{Z}$ into $L_S \triangleq L^M$ hypercubic M-dimensional cells of identical volume (each dimension is partitioned in to L intervals). The center of mass of the l-th cell is denoted as $\boldsymbol{x}_{L_S}^l, l \in \mathbb{N}_{L_S}^+$. Then, letting $\mathcal{X}_{L_S} \triangleq \left\{\boldsymbol{x}_{L_S}^l\right\}_{l\in\mathbb{N}_{L_S}^+}$, the *quantizer* $\mathcal{Q}_{L_S} : (\mathcal{Z}, \mathscr{B}(\mathcal{Z})) \mapsto \left(\mathcal{X}_{L_S}, 2^{\mathcal{X}_{L_S}}\right)$ is defined as the bijective and measurable function which uniquely maps the l-th cell to the respective *reconstruction point* $\boldsymbol{x}_{L_S}^l$, $\forall l \in \mathbb{N}_{L_S}^+$, according to some predefined ordering. That is, $\mathcal{Q}_{L_S}(\boldsymbol{x}) \triangleq \boldsymbol{x}_{L_S}^l$ if and only if $\boldsymbol{x}$ belongs to the respective cell. Given $\mathcal{Q}_{L_S}(\cdot)$, we consider the following discrete approximations of the process X_t:

- The *Markovian Quantization* of the state, defined as

$$\widetilde{X}_t^{L_S} \triangleq \mathcal{Q}_{L_S}\left(f\left(\widetilde{X}_{t-1}^{L_S}, W_t\right)\right) \in \mathcal{X}_{L_S}, \quad \forall t \in \mathbb{N}. \tag{3.70}$$

  where we have assumed explicitly apriori knowledge of a transition mapping, and

- The *Marginal Quantization* of the state, defined as

$$\overline{X}_t^{L_S} \triangleq \mathcal{Q}_{L_S}(X_t) \in \mathcal{X}_{L_S}, \quad \forall t \in \mathbb{N}. \tag{3.71}$$

Additionally, for later reference, define the column stochastic matrices $\widetilde{\boldsymbol{P}} \in [0,1]^{L_S \times L_S}$ and $\overline{\boldsymbol{P}} \in [0,1]^{L_S \times L_S}$ as

$$\widetilde{\boldsymbol{P}}(i,j) \triangleq \mathcal{P}\left(\widetilde{X}_t^{L_S} \equiv \boldsymbol{x}_{L_S}^i \middle| \widetilde{X}_{t-1}^{L_S} \equiv \boldsymbol{x}_{L_S}^j\right) \quad \text{and} \tag{3.72}$$

$$\overline{\boldsymbol{P}}(i,j) \triangleq \mathcal{P}\left(\overline{X}_t^{L_S} \equiv \boldsymbol{x}_{L_S}^i \middle| \overline{X}_{t-1}^{L_S} \equiv \boldsymbol{x}_{L_S}^j\right), \tag{3.73}$$

for all $(i,j) \in \mathbb{N}_{L_S}^+ \times \mathbb{N}_{L_S}^+$, obviously related to the Markovian and marginal state quantizations, respectively. Due to its structure, $\widetilde{\boldsymbol{P}}$ can at least be constructed simulating $\widetilde{X}_t^{L_S}$. From the Law of Large Numbers, the entries of $\widetilde{\boldsymbol{P}}$ can be estimated with arbitrary precision from a sufficiently large number of realizations of $\widetilde{X}_t^{L_S}, t \in \mathbb{N}_T$, for some $T < \infty$. Similarly, $\overline{\boldsymbol{P}}$ can be estimated also with arbitrary precision from multiple realizations of $\overline{X}_t^{L_S}$. Note, however, that in this case, it is possible to obtain $\overline{\boldsymbol{P}}$ only using

$$\lambda_t\left(\boldsymbol{x}_{L_S}^j\right) \triangleq \frac{\exp\left(-\frac{1}{2}\left(\mathbf{y}_t-\boldsymbol{A}_t\boldsymbol{x}_{L_S}^j\right)^{\boldsymbol{T}}\left(\boldsymbol{\Sigma}_t\left(\boldsymbol{x}_{L_S}^j\right)+\sigma_\xi^2\mathbf{I}_{N\times N}\right)^{-1}\left(\mathbf{y}_t-\boldsymbol{A}_t\boldsymbol{x}_{L_S}^j\right)\right)}{\sqrt{\det\left(\boldsymbol{\Sigma}_t\left(\boldsymbol{x}_{L_S}^j\right)+\sigma_\xi^2\mathbf{I}_{N\times N}\right)}} \tag{3.74}$$
$$> 0, \quad \forall j \in \mathbb{N}_{L_S}^+$$

available realizations of the state, *without actually knowing either the stochastic kernel or the transition mapping of* X_t (if such exists). For example, this could be made possible in sufficiently controlled physical experiments, specially designed for system identification, where the state X_t would be a fully observable stochastic process. In all subsequent analytical arguments, we will assume perfect knowledge of either $\widetilde{\boldsymbol{P}}$ or $\overline{\boldsymbol{P}}$, depending on the type of state quantization employed. The case where the aforementioned transition matrix is imperfectly known (as it actually happens in reality) or not (initially) known at all raises interesting questions with practical significance, however out of the scope of this particular work.

3.2.3.2 Asymptotically Optimal Recursive Estimators

Leveraging the ensemble of results presented in Section 2.2, we may carefully formulate the following fundamental result, which will be the basis for all subsequent analysis. In the following, $\mathcal{Q}_{L_S}^e : \left(\mathcal{X}_{L_S}, 2^{\mathcal{X}_{L_S}}\right) \mapsto \left(\mathcal{B}_{L_S}, 2^{\mathcal{B}_{L_S}}\right)$ constitutes a fixed bijective mapping between the sets $\mathcal{X}_{L_S}$ and $\mathcal{B}_{L_S} \triangleq \left\{\mathbf{e}_l^{L_S}\right\}_{l\in\mathbb{N}_{L_S}^+}$, where the latter contains as elements the complete standard basis in $\mathbb{R}^{L_S\times 1}$.

Theorem 3.3. (Approximate Filtering of Markov Processes) *Define the* ***reconstruction*** *and* ***likelihood*** *matrices*

$$\mathbf{X} \triangleq \left[\boldsymbol{x}_{L_S}^1\, \boldsymbol{x}_{L_S}^2\, \ldots\, \boldsymbol{x}_{L_S}^{L_S}\right] \in \mathbb{R}^{M\times L_S} \quad \textit{and} \tag{3.75}$$

$$\boldsymbol{\Lambda}_t \triangleq \operatorname{diag}\left(\lambda_t\left(\boldsymbol{x}_{L_S}^1\right) \ldots \lambda_t\left(\boldsymbol{x}_{L_S}^{L_S}\right)\right) \in \mathbb{R}^{L_S\times L_S}, \tag{3.76}$$

respectively, where, for all $t \in \mathbb{N}$, $\lambda_t\left(\boldsymbol{x}_{L_S}^j\right)$ *is given by (3.74) (top of page). Then, for any deterministic functional family* $\left\{\boldsymbol{\phi}_t : \mathbb{R}^{M\times 1} \mapsto \mathbb{R}^{M_{\phi_t}\times 1}\right\}_{t\in\mathbb{N}}$ *with bounded and*

continuous members, and any finite prediction horizon $\rho \geq 0$*, the optimal filter and* ρ*-step predictor of the* ***transformed process*** $\boldsymbol{\phi}_t(X_t)$ *may be approximated as*

$$\mathcal{E}^{L_S}\left(\boldsymbol{\phi}_{t+\rho}\left(X_{t+\rho}\right)\middle|\,\mathscr{Y}_t\right) \triangleq \boldsymbol{\Phi}_{t+\rho}\frac{\boldsymbol{P}^{\rho}E_t}{\|E_t\|_1} \in \mathbb{R}^{M_{\phi_t}\times 1}, \tag{3.77}$$

for all $t \in \mathbb{N}$*, where the process* $E_t \in \mathbb{R}^{L_S\times 1}$ *on the RHS of (3.77) satisfies the simple linear recursion*

$$E_t \equiv \boldsymbol{\Lambda}_t \boldsymbol{P} E_{t-1}, \quad \forall t \in \mathbb{N}, \tag{3.78}$$

with $\boldsymbol{P} \equiv \widetilde{\boldsymbol{P}}\left(\overline{\boldsymbol{P}}\right)$ *being the transition matrix when the Markovian (marginal) quantization is employed, and*

$$\boldsymbol{\Phi}_{t+\rho} \triangleq \left[\boldsymbol{\phi}_{t+\rho}\left(\boldsymbol{x}_{L_S}^{1}\right) \ \ldots \ \boldsymbol{\phi}_{t+\rho}\left(\boldsymbol{x}_{L_S}^{L_S}\right)\right] \in \mathbb{R}^{M_{\phi_t}\times L_S}. \tag{3.79}$$

The filter is initialized accordingly setting $E_{-1} \equiv \mathbb{E}\left\{\mathcal{Q}_{L_S}^{e}\left(\widetilde{X}_{-1}^{L_S}\right)\right\}$ *or* $\mathbb{E}\left\{\mathcal{Q}_{L_S}^{e}\left(\overline{X}_{-1}^{L_S}\right)\right\}$*. Additionally, pick any natural* $T < \infty$ *and suppose either of the following:*

- *The Markovian quantization is employed, whose initial value coincides with that of* X_t*, and* $f : \mathcal{Z}\times\mathcal{W} \overset{a.s.}{\mapsto} \mathcal{Z}$ *is Lipschitz in* $\mathcal{Z}$*, for every element of* $\mathcal{W}$*.*
- *The marginal quantization is employed and* X_t *is conditionally regular.*

Then, for any finite prediction horizon $\rho \geq 0$*, there exists a measurable subset* $\widehat{\Omega}_T \subseteq \Omega$ *with* $\mathcal{P}$*-measure at least* $1-(T+1)^{1-CN}\exp(-CN)$*, such that*

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left\|\mathcal{E}^{L_S}\left(\boldsymbol{\phi}_{t+\rho}\left(X_{t+\rho}\right)\middle|\,\mathscr{Y}_t\right) - \mathbb{E}\left\{\boldsymbol{\phi}_{t+\rho}\left(X_{t+\rho}\right)\middle|\,\mathscr{Y}_t\right\}\right\|_1 \underset{L_S\to\infty}{\longrightarrow} 0, \tag{3.80}$$

for any free, finite constant $C \geq 1$*.*

We observe that the two quantization strategies of interest (Markovian and marginal) both guarantee the same filter performance, in the sense that the respective approximate filters converge to the optimal MMSE state estimator under the same criteria. However, the assumptions on the internal nonlinear dynamics of the hidden state are quite different for the two types of state quantization considered and require different level of apriori knowledge about the structure of the hidden system under consideration (e.g., system description via a transition mapping or a stochastic kernel).

3.2.4 SCST & SSCP in Mobile Wireless Networks

In this section, we present the main results of this work. In a nutshell, we propose two theoretically consistent sequential algorithms for approximately solving the SCST and SSCP problems, defined previously in Section 3.2.2.5, both derived as applications of Theorem 3.3, presented in Section 3.2.3. Non trivial, real time recursive estimators of the variance of the proposed spatiotemporal channel predictors are also developed.

3.2.4.1 SCST

At this point, it is apparent that Theorem 3.3 in fact directly provides us with an effective approximate and recursive estimator for the channel state X_t. Simply, choose $\boldsymbol{\phi}_t$ as the identity, for all $t \in \mathbb{N}$ (recall that the channel state is assumed to be almost surely compactly supported). Therefore, Theorem 3.3 immediately solves the SCST problem, since the resulting filtering/prediction scheme is sequential and, as new channel measurements become available, its computational complexity is fixed, due to time invariance of the type of numerical operations required for each filter update.

3.2.4.2 SSCP

Defining the natural filtration generated by both the state X_t and the observations $\mathbf{y}_t$ as $\{\mathscr{H}_t\}_{t\in\mathbb{N}}$, $\mathscr{H}_t \triangleq \sigma\left\{\{X_t, \mathbf{y}_i\}_{i\in\mathbb{N}_t}\right\}$ and using the tower property of expectations, it is true that $\widehat{y}_t(\mathbf{q}) \equiv \mathbb{E}\left\{\mathbb{E}\left\{y_t(\mathbf{q})\middle|\mathscr{H}_t\right\}\middle|\mathscr{Y}_t\right\}$, for all $t \in \mathbb{N}$. Let us define the quantities

$$\alpha^{\mathbf{q}} \triangleq -10\log_{10}\left(\left\|\mathbf{q}-\mathbf{p}_{ref}\right\|_2\right), \tag{3.81}$$

$$\sigma^{\mathbf{q}}(X_t)\big|\, X_t \overset{def}{\sim} \mathcal{N}\left(0, \eta^2(X_t)\right) \quad \text{and} \tag{3.82}$$

$$\xi^{\mathbf{q}} \overset{def}{\sim} \mathcal{N}\left(0, \sigma_\xi^2\right), \tag{3.83}$$

where, also by definition,

$$\mathbb{E}\left\{\begin{bmatrix}\boldsymbol{\sigma}_t(X_t)\\ \sigma^{\mathbf{q}}(X_t)\end{bmatrix}\begin{bmatrix}\boldsymbol{\sigma}_t(X_t)\\ \sigma^{\mathbf{q}}(X_t)\end{bmatrix}^{\boldsymbol{T}}\middle| X_t\right\} \triangleq \begin{bmatrix}\boldsymbol{\Sigma}_t(X_t) & \boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\\ \left(\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\right)^{\boldsymbol{T}} & \eta^2(X_t)\end{bmatrix}, \tag{3.84}$$

with each element of $\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t) \in \mathbb{R}^{N\times 1}$ given by

$$\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)(j) \triangleq \theta_1(X_t)\exp\left(-\frac{\left\|\mathbf{q}-\mathbf{p}_j(t)\right\|_2}{\theta_2(X_t)}\right), \tag{3.85}$$

for all $j \in \mathbb{N}_N^+$. Then, it must be true that

$$y_t(\mathbf{q}) \equiv A^{\mathbf{q}} X_t + \sigma^{\mathbf{q}}(X_t) + \xi^{\mathbf{q}}, \tag{3.86}$$

where $A^{\mathbf{q}} \triangleq \left[\alpha^{\mathbf{q}} \mathbf{0}_{1\times(M-1)}\right] \in \mathbb{R}^{1\times M}$, since $y_t(\mathbf{q})$ can be equivalently considered as an additional observation, measured by an imaginary sensor at position $\mathbf{q}$, which of course was not used for state estimation in the SCST problem treated above. Using well known properties of jointly Gaussian random vectors [25],

$$\begin{aligned}\mathbb{E}\left\{y_t(\mathbf{q})\middle|\mathscr{H}_t\right\} &= A^{\mathbf{q}} X_t + \left(\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t)\left(\mathbf{y}_t - \boldsymbol{A}_t X_t\right) \\ &\triangleq \phi_t(X_t, \mathbf{y}_t).\end{aligned} \tag{3.87}$$

As a result, $\widehat{y}_t(\mathbf{q})$ can be expressed as

$$\widehat{y}_t(\mathbf{q}) \equiv \mathbb{E}\left\{\phi_t(X_t, \mathbf{y}_t)\middle|\mathscr{Y}_t\right\}, \tag{3.88}$$

that is, the SSCP problem coincides with the problem of sequentially evaluating the optimal nonlinear filter of the functional $\phi_t(\cdot,\cdot)$. However, note that Theorem 3.3 is not directly applicable for developing an approximate recursive filter for the MMSE optimal predictor (3.88), because the functional $\phi_t(\cdot,\cdot)$ *depends, except for the state, on the observations process* (sensor measurements) as well. However, exploiting the linearity of $\phi_t(\cdot,\cdot)$ on $\mathbf{y}_t$, the following result is true, which constitutes a generalization of Theorem 3.3, and provides a closed form approximate solution to the SSCP problem, at the same time enjoying asymptotic optimality in the sense of Theorem (3.3).

Theorem 3.4. (Approximate Solution to the SSCP Problem) *The optimal spatiotemporal predictor of the channel magnitude at an arbitrary position* $\mathbf{q} \in \mathbb{R}^2$*,* $\widehat{y}_{t+\rho}(\mathbf{q})$*, can be approximated as*

$$\mathcal{E}^{L_S}\left(y_{t+\rho}(\mathbf{q})\middle|\mathscr{Y}_t\right) \triangleq \begin{cases} \left\langle \boldsymbol{\phi}_t(\mathbf{y}_t), \dfrac{E_t}{\|E_t\|_1} \right\rangle, & \rho \equiv 0 \\ A^{\mathbf{q}} \mathbf{X} \dfrac{\boldsymbol{P}^{\rho} E_t}{\|E_t\|_1}, & \rho \geq 1 \end{cases}, \tag{3.89}$$

for all $t \in \mathbb{N}$*, where the process* $E_t \in \mathbb{R}^{L_S\times 1}$ *can be recursively evaluated as in Theorem 3.3 and where the stochastic process* $\boldsymbol{\phi}_t(\mathbf{y}_t) \in \mathbb{R}^{L_S\times 1}$ *is defined as*

$$\boldsymbol{\phi}_t(\mathbf{y}_t) \triangleq \left[\phi_t\left(\boldsymbol{x}_{L_S}^1, \mathbf{y}_t\right) \ \ldots \ \phi_t\left(\boldsymbol{x}_{L_S}^{L_S}, \mathbf{y}_t\right)\right]^{\boldsymbol{T}}, \tag{3.90}$$

with $\phi_t : \mathbb{R}^{M\times 1} \times \mathbb{R}^{N\times 1} \mapsto \mathbb{R}$ *defined as in (3.87). Additionally, under the same conditions as in Theorem 3.3, it is true that*

$$\sup_{t\in\mathbb{N}_T} \sup_{\omega\in\widehat{\Omega}_T} \left| \mathcal{E}^{L_S}\left(y_{t+\rho}\left(\mathbf{q}\right) \middle| \mathscr{Y}_t \right) - \widehat{y}_{t+\rho}\left(\mathbf{q}\right) \right| \underset{L_S\to\infty}{\longrightarrow} 0. \tag{3.91}$$

Proof of Theorem 3.4. See Section 3.2.7.1 (Appendix A). ■

Theorem 3.4 presented above provides asymptotically optimal estimators for the approximate evaluation of the conditional expectation of $y_{t+\rho}\left(\mathbf{q}\right)$, for all $\rho \geq 0$ and $\mathbf{q} \in \mathbb{R}^2$. In a practical setting, though, the availability of a measure of uncertainty of the point estimate $\widehat{y}_{t+\rho}\left(\mathbf{q}\right)$ would also be of interest. Next, we provide approximate estimators for the conditional variance of $y_{t+\rho}\left(\mathbf{q}\right)$, given the available information up to and including time t, $\mathscr{Y}_t$. In the following, let $B \triangleq \left[0\,1\,\mathbf{0}_{1\times(M-2)}\right] \in \mathbb{R}^{1\times M}$.

Theorem 3.5. (Approximate Filters for the Conditional Variance of Channel Predictions) *The optimal predictor of the variance of* $y_{t+\rho}\left(\mathbf{q}\right)$*, conditional on the filtration* $\{\mathscr{Y}_t\}_{t\in\mathbb{N}}$*, at an arbitrary position* $\mathbf{q} \in \mathbb{R}^2$*,* $\mathbb{V}_{\mathscr{Y}_t}\left\{y_{t+\rho}\left(\mathbf{q}\right)\right\}$*, can be approximated as follows:*

If $\rho \equiv 0$*,*

$$\begin{aligned}\mathcal{V}_{\mathscr{Y}_t}^{L_S}\left(y_t\left(\mathbf{q}\right)\right) \triangleq \sigma_\xi^2 &+ \left\langle \boldsymbol{\phi}_t^{\mathcal{V}}\left(\mathbf{y}_t\right), \frac{E_t}{\left\|E_t\right\|_1} \right\rangle \\ &- \left(\mathcal{E}^{L_S}\left(y_t\left(\mathbf{q}\right)\middle|\mathscr{Y}_t\right)\right)^2,\end{aligned} \tag{3.92}$$

whereas, if $\rho > 0$*,*

$$\begin{aligned}\mathcal{V}_{\mathscr{Y}_t}^{L_S}\left(y_{t+\rho}\left(\mathbf{q}\right)\right) \triangleq \sigma_\xi^2 &+ \left\langle \boldsymbol{\phi}_{t+\rho}^{\mathcal{V}}, \frac{\boldsymbol{P}^\rho E_t}{\left\|E_t\right\|_1} \right\rangle \\ &- \left(\mathcal{E}^{L_S}\left(y_{t+\rho}\left(\mathbf{q}\right)\middle|\mathscr{Y}_t\right)\right)^2,\end{aligned} \tag{3.93}$$

for all $t \in \mathbb{N}$*. In the above, the process* $E_t \in \mathbb{R}^{L_S\times 1}$ *can be recursively evaluated as in Theorem 3.3, the process* $\boldsymbol{\phi}_t\left(\mathbf{y}_t\right) \in \mathbb{R}^{L_S\times 1}$ *is defined as*

$$\boldsymbol{\phi}_t^{\mathcal{V}}\left(\mathbf{y}_t\right) \triangleq \left[\phi_t^{\mathcal{V}}\left(\boldsymbol{x}_{L_S}^1, \mathbf{y}_t\right) \ldots \phi_t^{\mathcal{V}}\left(\boldsymbol{x}_{L_S}^{L_S}, \mathbf{y}_t\right)\right]^{\boldsymbol{T}}, \tag{3.94}$$

with $\phi_t^{\mathcal{V}}\left(\boldsymbol{x}_{L_S}^j, \mathbf{y}_t\right) \in \mathbb{R}$ *given by (3.97) and (3.98) (bottom of next page), and*

$$\boldsymbol{\phi}_{t+\rho}^{\mathcal{V}} \triangleq \left[\phi_{t+\rho}^{\mathcal{V}}\left(\boldsymbol{x}_{L_S}^1\right) \ldots \phi_{t+\rho}^{\mathcal{V}}\left(\boldsymbol{x}_{L_S}^{L_S}\right)\right]^{\boldsymbol{T}} \in \mathbb{R}^{L_S\times 1}, \tag{3.95}$$

with $\phi_{t+\rho}^{\mathcal{V}}\left(\boldsymbol{x}_{L_S}^j\right) \triangleq \left(A^{\mathbf{q}}\boldsymbol{x}_{L_S}^j\right)^2 + B\boldsymbol{x}_{L_S}^j, j \in \mathbb{N}_{L_S}^+$. *Additionally, under the usual circumstances, it is true that*

$$\sup_{\substack{t\in\mathbb{N}_T \\ \omega\in\widehat{\Omega}_T}} \left|\mathcal{V}_{\mathscr{Y}_t}^{L_S}\left(y_{t+\rho}\left(\mathbf{q}\right)\right) - \mathbb{V}_{\mathscr{Y}_t}\left\{y_{t+\rho}\left(\mathbf{q}\right)\right\}\right| \underset{L_S\to\infty}{\longrightarrow} 0. \tag{3.96}$$

Proof of Theorem 3.5. See Section 3.2.7.2 (Appendix B). ■

3.2.4.3 Computational Complexity: A Fair Comparison

A careful inspection of the filtering schemes proposed for the solution of the SCST and SSCP problems, respectively, reveals that, in the worst case, the computational complexity of both algorithms scales as $\mathcal{O}\left(L_S^2 + L_S N^3\right)$. The two algorithms can also be combined into one with the same computational requirements. The cubic term is due to the inversion and the determinant calculation of the involved covariance matrices and it is computationally bearable, at least for a relatively small number of sensors. However, note that when the sensors are stationary or when their trajectories are fixed and known apriori, these operations may be completely bypassed by precomputing the required set of matrices and storing them in memory. Then, the computational complexity reduces to $\mathcal{O}\left(L_S^2 + L_S N^2\right)$.

Treating N as a constant, the complexity of both algorithms considered scales as $\mathcal{O}\left(L_S^2\right)$. In the particular problems we are interested in here, though, the state dimension M is relatively low, almost always between 2 and 5 (here, it is assumed that $M \equiv 3$ at most), which makes grid based filters practically feasible. Additionally, as it has already been shown in [56] in a non Bayesian framework, the sensitivity of the quality of spatial channel prediction on the estimation error of the shadowing power and decorrelation distance of the channel is indeed very weak, making it possible to potentially consider a lower quantization resolution for the aforementioned quantities without significant compromise in terms of the prediction quality. As a result, grid

---

$$\phi_t^{\mathcal{V}}\left(\boldsymbol{x}_{L_S}^j, \mathbf{y}_t\right) \triangleq \phi_t^2\left(\boldsymbol{x}_{L_S}^j, \mathbf{y}_t\right) + \psi_{t,\mathbf{q}}^{\boldsymbol{o}}\left(\boldsymbol{x}_{L_S}^j\right) + B\boldsymbol{x}_{L_S}^j \in \mathbb{R}, \tag{3.97}$$

$$\psi_{t,\mathbf{q}}^{\boldsymbol{o}}\left(\boldsymbol{x}_{L_S}^j\right) \triangleq \left(\boldsymbol{\sigma}_t^{\mathbf{q}}\left(\boldsymbol{x}_{L_S}^j\right)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}\left(\boldsymbol{x}_{L_S}^j\right) \boldsymbol{\sigma}_t^{\mathbf{q}}\left(\boldsymbol{x}_{L_S}^j\right) \in \mathbb{R}, \quad j \in \mathbb{N}_{L_S}^+. \tag{3.98}$$

based approximate filters are indeed adequate for the problems of interest, taking advantage of their strong asymptotic properties.

Naturally, particle filters [23,69] would constitute the rivals of our grid based filtering approach. Particle filters exhibit a computational complexity of $\mathcal{O}(L_S)$, with L_S being the number of particles [23]. That is, their complexity is one order of magnitude smaller compared to the complexity of grid based filters. Note, though, that in grid based filtering, the one and only computational operation incurring a complexity of $\mathcal{O}\left(L_S^2\right)$ is the Matrix-Vector (MV) operation $\boldsymbol{P}E_{t-1}$. In fact, in the numerical simulations conducted in Section 3.2.5, it was revealed that, at least for the problems of interest, the required operations on the involved covariance matrices are far more computationally intensive than the aforementioned MV multiplication. These operations would be also required in any particle filter implementation as well.

Continuing the comparison with particle filters, another issue of major importance is filter behavior with respect to the curse of dimensionality. Particle filters do suffer from the curse. In general, their expected approximation error scales exponentially with respect to the dimensionality of the HMM under consideration [41–43, 70]. Of course, it can be easily seen that grid based filters also suffer from similar drawbacks, a fact that strengthens the common belief that, at least in the context of nonlinear filtering, the curse of dimensionality constitutes a ubiquitous phenomenon.

3.2.4.4 Advantages of Grid Based Filters over Particle Filters

When applied to lower dimensional hidden systems, which are of interest in this work (see previous section), grid based filtering possesses some definite advantages over particle filtering techniques. From the theoretical point of view, as stated in Theorems 3.3, 3.4 and 3.5, the convergence of the grid based approximate filters proposed in this work for effectively solving the SCST and SSCP problems is compact in time and, most importantly, uniform in a *purely characterized* set of almost full probability measure, which explicitly depends on T and N. In favor of the particle filtering approach, *Egoroff's Theorem* [21] states that almost sure convergence implies almost uniform convergence, that is, there exists a measurable set of arbitrarily small measure, such that

the convergence is uniform in the complement of this set (almost uniform convergence). However, Egoroff's Theorem constitutes a purely abstract result, being of small practical importance, providing no practically useful theoretical guarantee. On the other hand, for the proposed grid based filters, Egoroff's Theorem is purely quantitatively justified. In fact, for fixed T, convergence to *both* the MMSE optimal nonlinear filter and channel gain map tracker (channel spatiotemporal predictor) occurs *uniformly* in a *common* set that approximates the certain event, *at an exponential rate in* N. As a result, the dimensionality of the observations process *stochastically stabilizes* the proposed approximate filter. This is practically important because it shows that the proposed estimators will behave better with a larger number of channel measurements, favoring networks with a large number of sensors. Our results provide a way of theoretically justifying the aforementioned intuitively expected behavior of the proposed estimators. To the best of our knowledge, such type of results do not exist in the case of particle filters. Also see for a thorough theoretical comparison of grid based filters with particle filters, showing that, at least theoretically, a larger class of hidden processes can be supported by the former, compared to the latter.

From the practical point of view, grid based approximate filters exhibit two important advantages over particle filters, as follows. First, at least for lower dimensional systems, grid based filters require a smaller number of quantization cells when compared to the number of particles required in order to attain the same level of performance. See the comparative survey [44], where this favorable behavior of grid based filters is clearly demonstrated. Of course, smaller number of cells are practically important, since, this way, the actual computational resources required for the implementation of the filtering process are greatly reduced, also justifying the somewhat higher computational complexity of grid based filtering schemes. Second, grid based approximate filters can be implemented in a lot easier and definitely more robust manner, compared to particle filtering schemes. One reason for this is that grid based filters constitute *truly recursive* and *deterministic* estimators, without the need of any random sampling operation, which can be a complicated procedure, especially for hidden systems with

complex internal nonlinear dynamics. Additionally, it is well known [23, 45] that particle filters suffer from some inherent defects, such as the problem of properly choosing the importance density and the phenomenon of particle degeneracy, potentially causing performance degradation. On the other hand, by nature, grid based filters require no fine tuning at all and their performance is determined almost exclusively by the grid resolution, in order to ensure satisfactory filtering performance.

3.2.5 Numerical Simulations

The effectiveness of the proposed estimation schemes will be validated through a number of synthetic experiments. Specifically, we consider $N \equiv 30$ sensors randomly scattered on a fine square grid, in the region $\mathcal{S} \equiv [0, 40]^2$ (in $m \times m$). The position of the reference antenna is fixed at $\mathbf{p}_{ref} \equiv [25\,10]^{\boldsymbol{T}}$. The variance of the multipath fading is set at $\sigma_\xi^2 \equiv 2$ and, as far as shadowing is concerned, we assume that the correlation distance is known and equal to $10\,m$. As a result, in this example, the channel state is two dimensional, with $X_t(1) \equiv \mu(X_t)$ and $X_t(2) \equiv \eta^2(X_t) \equiv \boldsymbol{\theta}(X_t)$. Each component of the channel state evolves according to

$$X_t(1) \equiv \tanh(\gamma(X_{t-1}(1) - 2)) + W_t + 2, \quad \text{and} \tag{3.99}$$

$$X_t(2) \equiv 0.3\,|\tanh(\sin(\gamma X_{t-1}(2)\, W_t) + \\ + X_{t-1}(2)\, W_t) + W_t| + 25, \quad \forall t \in \mathbb{N}, \tag{3.100}$$

for some arbitrary but known initial conditions, where $\gamma \equiv 1.6$ and $W_t \equiv \text{clip}_{[-1,1]}(G_t)$, $G_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, with $\text{clip}_{[-1,1]}(\cdot)$, denoting the hard limiter operation into $[-1, 1]$. Note the strong coupling between the two equations, driven by the same noise realizations. The above equations attempt to model a situation where the path loss exponent is somewhat slowly varying between 0 and 4, whereas the shadowing power is rapidly varying between 25 and 25.6. The state $X_t \equiv [X_t(1)\ X_t(2)]^{\boldsymbol{T}}$ was uniformly quantized into $L_S \equiv 30^2$ cells (that is, $L \equiv 30$). Concerning state quantization, the marginal type was employed, where the transition matrix $\boldsymbol{P} \equiv \overline{\boldsymbol{P}}$ was simulated from 10^5 realizations of the Markov process under consideration. More specifically, each entry of the matrix, corresponding to a conditional probability, was estimated by counting how many

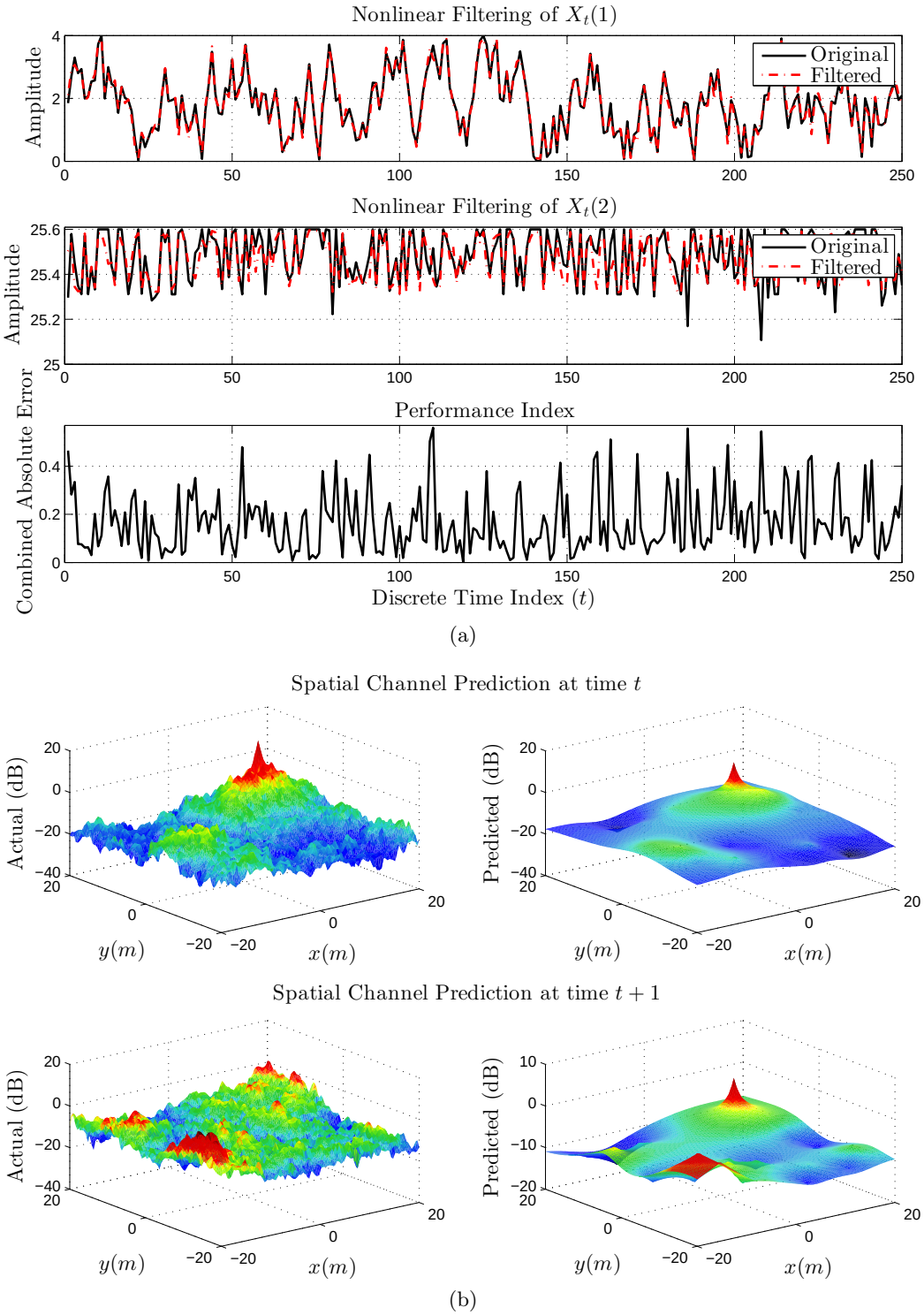


Figure 3.3: (a) Demonstration of channel state tracking for 250 time steps. The estimates are produced from the observations of 30 randomly scattered sensors in the square region $[-20\,m, 20\,m]^2$. (b) Spatial prediction and temporal tracking of the channel combined. The spatial grid consists of 3600 points and the results are obtained using just 30 spatial measurements.

times the respective event happens given the respective condition, how many times the respective condition happens and, then, dividing the former with the latter. For simplicity, we focus on the case where $\rho \equiv 0$, that is, we consider the problems of *temporal filtering of the channel state* and *spatial prediction of the channel*, both being instances of the SCST and SSCP problems, respectively.

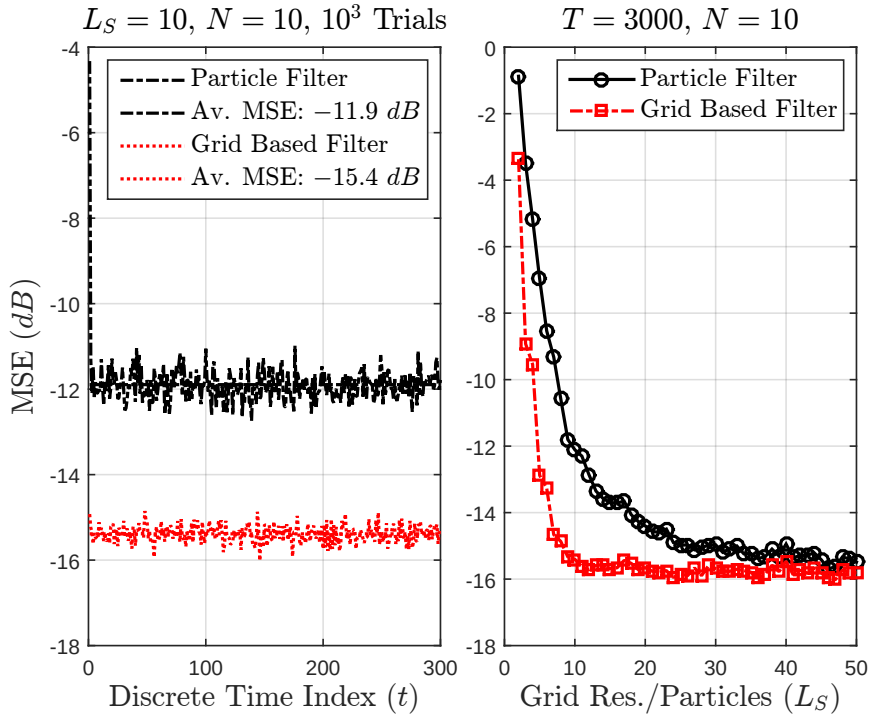
Fig. 3.3a demonstrates the channel state tracking (temporal filtering of the state) for 250 time steps, according to the experimental setting stated above. As illustrated in the figure, the quality of the estimates is very good, considering the nonlinearity present both in the state process and the observations at each sensor in the network. It also apparent that the produced estimation process behaves in a stable manner, as time increases.

The filter of the channel state can subsequently be used for predicting the channel magnitude at unexplored positions. This is illustrated in Fig. 3.3b. The random field used for modeling the spatial channel process was generated using a spatial grid of 3600 points and the respective predicted values were obtained from just $N \equiv 30$ randomly scattered spatial channel measurements in the region of interest. From the figure, it can be seen that the quality of the predicted process is very good, especially considering the fact that the channel is reconstructed using only 0.83 % of the total number of grid points in the region of interest. Of course, the quality of the spatial prediction improves as the number of spatial measurements (and therefore nodes/sensors) increases.

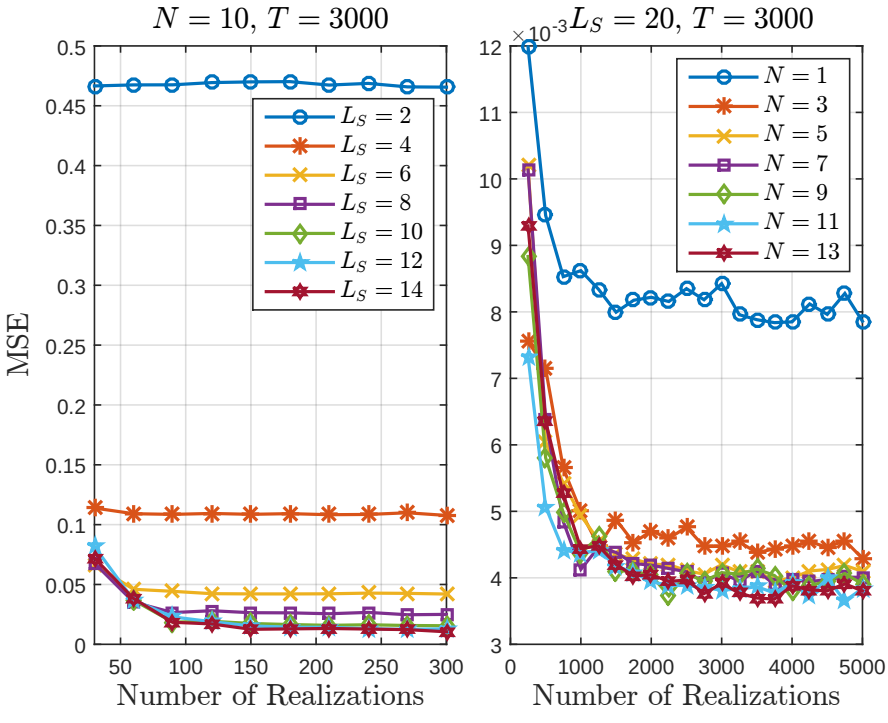
The practical advantages of the grid based filtering approach discussed in this work against particle filtering techniques will also be experimentally evaluated, confirming the discussion presented in Section 3.2.4.4. Focusing only on hidden state estimation (in our context, SCST), let us additionally *fix* the shadowing power as $\eta^2 \equiv 25$, set $\sigma_\xi^2 \equiv 4.1$ and assume that the time varying path loss coefficient $\mu(X_t) \equiv X_t$ evolves according to the *non-Gaussian* model

$$X_t \equiv \tanh\left(\gamma\left(X_{t-1} - 2\right) W_t\right) + W_t + 2, \quad \forall t \in \mathbb{N}, \tag{3.101}$$

with the rest of parameters same as before. This system simulates a particularly noisy environment for state estimation, which is typical in wireless channel realizations. Fig.



(a)



(b)

Figure 3.4: (a) MSE comparison between particle and grid based filters. Left: Monte Carlo (MC) estimate of the MSE for $T \equiv 300$ time instants (10^3 MC trials). Right: MSE as a function of grid resolution/number of particles. (b) MSE as a function of training realizations: Left: keeping N fixed and varying L_S. Right: keeping L_S fixed and varying N.

3.4a shows two different comparisons of the achieved MSE between the grid based estimator described herein and a standard SIR (Sequential Importance Resampling) particle filter with systematic resampling in the resampling step, known to minimize the Monte Carlo (MC) variation [23]. Resampling is performed at each iteration of the particle filter. Specifically, Fig. 3.4a (left) shows an MC estimate of the MSE for $T \equiv 300$ time instants of the operation of the two filtering schemes, where $L_S \equiv 10$, $N \equiv 10$ and the total number of MC trials averaged in order to produce the presented results is 1000. From the figure, it is apparent that the grid based filter performs considerably better than the particle filter. The average MSE for the grid based filter more than $3dB$ lower than that for the particle filter, which translates into more than twice as large quadratic error for the latter. What is more, the fluctuations of the estimates of the MSE around the average is quite larger for the particle filter compared to the grid based one, translating in larger variance of the corresponding state estimates.

Fig. 3.4a (right) shows the MSE of the two filtering schemes as a function of the grid resolution/number of particles, L_S. N is the same as before and the respective estimates of the MSE were computed by time averaging of $T \equiv 3000$ filtering estimates. Again, grid based filtering performs strictly better than particle filtering, since all MSE estimates for the former are strictly smaller than the respective estimates for the latter, for the same L_S. In particular, we observe that grid based filtering with a grid resolution of 10 to 20 performs better than particle filtering utilizing 50 particles, or more. Overall, Fig. 3.4a shows that grid based filters perform better especially in the low-resolution-high-noise regime, which is important regarding the applications of interest in this work.

Finally, Fig. 3.4b provides an experimental assessment of the discussed grid based filters with respect to model mismatch. As noted earlier in this section, the transition matrix $\boldsymbol{P}$ is produced using the Law of Large Numbers, by simulation. This is performed as an offline training phase, before the actual operation of the approximate filter. As a result, it is important to at least experimentally verify the behavior of the filter as a function of the number of state realizations employed in the training phase for estimating $\boldsymbol{P}$. This is what Fig. 3.4b depicts, where the hidden system is assumed to

be the same as before (see (3.101)), but where $\eta^2 \equiv 1$ and $\sigma_\xi^2 \equiv 1$. These choices are made in order to reduce the error floor of the filtering process, making the errors due to model mismatch more pronounced.

More specifically, Fig. 3.4b (left) shows the MSE, computed by time averaging as above, as a function of training realizations, for different values of L_S, keeping N fixed at 10. From the figure, it is apparent that the sensitivity to model mismatch is very well behaved, with around 200 training realizations being sufficient in order for the approximate filter to reach its error floor. The situation is similar in Fig. 3.4b (right), where $L_S \equiv 20$ and N is varying. However, in this case, a larger number of realizations (~ 3000) is required in order to achieve the respective MSE floors. This is due to the larger grid resolution ($L_S \equiv 20$). Nevertheless, in any case, the sensitivity to model mismatch shows a definitely regular behavior, and a relatively small number of (channel) state realizations is required in order for the respective grid based filters to be able to produce consistent estimates.

3.2.6 Conclusion

A nonlinear filtering framework was proposed for addressing the fundamental problems of sequential channel state tracking and spatiotemporal channel prediction in mobile wireless sensor networks. First, we formulated the channel observations at each sensor as a partially observable nonlinear system with temporally varying state and spatiotemporally varying observations. Then, a grid based approximate filtering scheme was employed for accurately tracking the temporal variation of the channel state, based on which we proposed a recursive spatiotemporal channel gain predictor, providing real time sequential CG map estimation at each sensor in the network. Non trivial, approximate recursive estimators of the variance of the CG map predictions were also proposed. Further, we showed that all three estimators are asymptotically optimal, in the sense that they converge to the respective optimal MMSE estimators/predictors, in a technically strong sense. In addition to these theoretical results, numerical simulations were presented, validating the practical effectiveness of the proposed approach and increasing the user's confidence for practical consideration in real world wireless

networks.

3.2.7 Appendices

3.2.7.1 Appendix A: Proof of Theorem 3.4

Let us first consider the filtering case, that is, the one where $\rho \equiv 0$. Substituting (3.87) to (3.88) and defining

$$\boldsymbol{\chi}^1_{t,\mathbf{q}}(X_t) \triangleq \left(\left(\boldsymbol{\sigma}^{\mathbf{q}}_t(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t)\right)^{\boldsymbol{T}} \in \mathbb{R}^{N\times 1} \quad \text{and} \tag{3.102}$$

$$\chi^2_{t,\mathbf{q}}(X_t) \triangleq \left(\boldsymbol{\sigma}^{\mathbf{q}}_t(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t)\, \boldsymbol{A}_t X_t \in \mathbb{R},\ t \in \mathbb{N}, \tag{3.103}$$

we can write

$$\widehat{y}_t(\mathbf{q}) \equiv A^{\mathbf{q}}\mathbb{E}\{X_t|\,\mathscr{Y}_t\} + \left(\mathbb{E}\left\{\boldsymbol{\chi}^1_{t,\mathbf{q}}(X_t)\middle|\,\mathscr{Y}_t\right\}\right)^{\boldsymbol{T}}\mathbf{y}_t - \mathbb{E}\left\{\chi^2_{t,\mathbf{q}}(X_t)\middle|\,\mathscr{Y}_t\right\}. \tag{3.104}$$

Then, for all $t \in \mathbb{N}$, define the approximate operator

$$\begin{aligned}\mathcal{E}^{L_S}(y_t(\mathbf{q})|\,\mathscr{Y}_t) \triangleq\ & A^{\mathbf{q}}\mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t)\\ &+ \left(\mathcal{E}^{L_S}\left(\boldsymbol{\chi}^1_{t,\mathbf{q}}(X_t)\middle|\,\mathscr{Y}_t\right)\right)^{\boldsymbol{T}}\mathbf{y}_t - \mathcal{E}^{L_S}\left(\chi^2_{t,\mathbf{q}}(X_t)\middle|\,\mathscr{Y}_t\right).\end{aligned} \tag{3.105}$$

Using the triangle inequality, the Cauchy-Schwarz Inequality and the fact that the ℓ_2 norm of a vector is upper bounded by its ℓ_1 norm, it is true that

$$\begin{aligned}&\left|\mathcal{E}^{L_S}(y_t(\mathbf{q})|\,\mathscr{Y}_t) - \widehat{y}_t(\mathbf{q})\right|\\ &\quad\le |\alpha^{\mathbf{q}}|\left\|\left(\mathcal{E}^{L_S}(X_t|\,\mathscr{Y}_t) - \mathbb{E}\{X_t|\,\mathscr{Y}_t\}\right)\right\|_1\\ &\qquad + \|\mathbf{y}_t\|_2\left\|\mathcal{E}^{L_S}\left(\boldsymbol{\chi}^1_{t,\mathbf{q}}(X_t)\middle|\,\mathscr{Y}_t\right) - \mathbb{E}\left\{\boldsymbol{\chi}^1_{t,\mathbf{q}}(X_t)\middle|\,\mathscr{Y}_t\right\}\right\|_1\\ &\qquad + \left|\mathcal{E}^{L_S}\left(\chi^2_{t,\mathbf{q}}(X_t)\middle|\,\mathscr{Y}_t\right) - \mathbb{E}\left\{\chi^2_{t,\mathbf{q}}(X_t)\middle|\,\mathscr{Y}_t\right\}\right|.\end{aligned} \tag{3.106}$$

Now, from Lemma 2.7, it follows that for any natural $T < \infty$, there exists a bounded $\gamma > 1$, such that $\sup_{t\in\mathbb{N}_T}\|\mathbf{y}_t(\omega)\|_2 < \sqrt{\gamma CN(1+\log(T+1))}$, for all $\omega \in \widehat{\Omega}_T \subseteq \Omega$, with measure at least $1-(T+1)^{1-CN}\exp(-CN)$, exactly as in Theorem 3.3. Therefore, invoking Theorems 3.3 and 2.8, it readily follows that, under the respective conditions,

$$\lim_{L_S\to\infty}\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left|\mathcal{E}^{L_S}(y_t(\mathbf{q})|\,\mathscr{Y}_t) - \widehat{y}_t(\mathbf{q})\right| \equiv 0, \tag{3.107}$$

showing the second part of the theorem, when $\rho \equiv 0$. For the first part, observe that $\mathcal{E}^{L_S}\left(y_t\left(\mathbf{q}\right)\middle|\,\mathscr{Y}_t\right)$ can be explicitly expressed as (see Theorems 3.3 and 2.8)

$$\mathcal{E}^{L_S}\left(y_t\left(\mathbf{q}\right)\middle|\,\mathscr{Y}_t\right) \equiv \left\langle \left(A^{\mathbf{q}}\mathbf{X}\right)^{\boldsymbol{T}} + \left(\boldsymbol{\Psi}^1_{t,\mathbf{q}}\right)^{\boldsymbol{T}}\mathbf{y}_t + \boldsymbol{\chi}^2_{t,\mathbf{q}}, \frac{E_t}{\left\|E_t\right\|_1}\right\rangle, \tag{3.108}$$

where

$$\boldsymbol{\Psi}^1_{t,\mathbf{q}} \triangleq \left[\boldsymbol{\chi}^1_{t,\mathbf{q}}\left(\boldsymbol{x}^1_{L_S}\right) \cdots \boldsymbol{\chi}^1_{t,\mathbf{q}}\left(\boldsymbol{x}^{L_S}_{L_S}\right)\right] \in \mathbb{R}^{N\times L_S}, \tag{3.109}$$

$$\left(\boldsymbol{\psi}^2_{t,\mathbf{q}}\right)^{\boldsymbol{T}} \triangleq \left[\chi^2_{t,\mathbf{q}}\left(\boldsymbol{x}^1_{L_S}\right) \cdots \chi^2_{t,\mathbf{q}}\left(\boldsymbol{x}^{L_S}_{L_S}\right)\right] \in \mathbb{R}^{1\times L_S} \tag{3.110}$$

and which, after simple algebra, can be easily shown to coincide with the vector process $\boldsymbol{\phi}_t\left(\mathbf{y}_t\right)$ of Theorem 3.4.

In the prediction case ($\rho \geq 1$), the procedure is slightly different. Let us define $\left\{\mathscr{Q}^{+\rho}_t\right\}_{t\in\mathbb{N}}$, $\mathscr{Q}^{+\rho}_t \triangleq \sigma\left\{\left\{X_{i+\rho}, \mathbf{y}_i, y_i\left(\mathbf{q}\right)\right\}_{i\in\mathbb{N}_t}\right\}$, the complete filtration generated by $X_{t+\rho}, \mathbf{y}_t$ and $y_t\left(\mathbf{q}\right)$. Also, note that, for all $\rho \geq 1$, the augmented observation vector process (and therefore each one of its elements) $\mathbf{y}^{aug}_{t+\rho} \triangleq \left[\mathbf{y}_{t+\rho}\, y_{t+\rho}\left(\mathbf{q}\right)\right]^{\boldsymbol{T}} \in \mathbb{R}^{(N+1)\times 1}$ is *conditionally independent of* $\mathbf{y}^{aug}_t, \mathbf{y}^{aug}_{t-1}, \ldots$, *given the state at time* $t+\rho$, $X_{t+\rho}$. Thus, using the tower property, it is true that

$$\begin{aligned}
\widehat{y}_{t+\rho}\left(\mathbf{q}\right) &\equiv \mathbb{E}\left\{\mathbb{E}\left\{y_{t+\rho}\left(\mathbf{q}\right)\middle|\,\mathscr{Q}^{+\rho}_t\right\}\middle|\,\mathscr{Y}_t\right\} \\
&\equiv \mathbb{E}\left\{\mathbb{E}\left\{A^{\mathbf{q}}X_{t+\rho} + \sigma^{\mathbf{q}}\left(X_{t+\rho}\right) + \xi^{\mathbf{q}}\middle|\,X_{t+\rho}\right\}\middle|\,\mathscr{Y}_t\right\} \\
&\equiv A^{\mathbf{q}}\mathbb{E}\left\{X_{t+\rho}\middle|\,\mathscr{Y}_t\right\}.
\end{aligned} \tag{3.111}$$

Consequently, defining the approximate spatiotemporal predictor $\mathcal{E}^{L_S}\left(y_{t+\rho}\left(\mathbf{q}\right)\middle|\,\mathscr{Y}_t\right) \triangleq A^{\mathbf{q}}\mathcal{E}^{L_S}\left(X_{t+\rho}\middle|\,\mathscr{Y}_t\right)$, substituting $\mathcal{E}^{L_S}\left(X_{t+\rho}\middle|\,\mathscr{Y}_t\right)$ from Theorem 3.3 and following a very similar convergence analysis to the filtering case treated above, the respective results present in the statement of Theorem 3.4 follow. The proof is complete. ■

3.2.7.2 Appendix B: Proof of Theorem 3.5

Fix $\mathbf{q} \in \mathbb{R}^2$. If $\rho \equiv 0$, the conditional variance of $y_t\left(\mathbf{q}\right)$ given $\mathscr{Y}_t$ is given by

$$\mathbb{V}_{\mathscr{Y}_t}\left\{y_t\left(\mathbf{q}\right)\right\} = \mathbb{E}\left\{y^2_t\left(\mathbf{q}\right)\middle|\,\mathscr{Y}_t\right\} - \widehat{y}^2_t\left(\mathbf{q}\right), \quad \forall t \in \mathbb{N}. \tag{3.115}$$

Also, from (3.87), we have $\mathbb{E}\left\{ y_t^2(\mathbf{q}) \middle| \mathscr{H}_t \right\} = \mathbb{V}_{\mathscr{H}_t}\{y_t(\mathbf{q})\} + \phi_t^2(X_t, \mathbf{y}_t)$. Using the tower property in (3.115), we get

$$\mathbb{V}_{\mathscr{Y}_t}\{y_t(\mathbf{q})\} \equiv \mathbb{E}\left\{ \mathbb{V}_{\mathscr{H}_t}\{y_t(\mathbf{q})\} \middle| \mathscr{Y}_t \right\} + \mathbb{E}\left\{ \phi_t^2(X_t, \mathbf{y}_t) \middle| \mathscr{Y}_t \right\} - \widehat{y}_t^2(\mathbf{q}), \quad \forall t \in \mathbb{N}, \tag{3.116}$$

what is otherwise called the Law of Total Variance. From (3.116), we observe that $\mathbb{V}_{\mathscr{Y}_t}\{y_t(\mathbf{q})\}$ can be split as a nonlinear combination of three estimators.

Using well known properties of jointly Gaussian random vectors and the definition of $\psi_{t,\mathbf{q}}^{\boldsymbol{o}}(\cdot)$ in (3.98), it is true that [25] $\mathbb{V}_{\mathscr{H}_t}\{y_t(\mathbf{q})\} = \sigma_\xi^2 + \eta^2(X_t) + \psi_{t,\mathbf{q}}^{\boldsymbol{o}}(X_t)$, and, therefore,

$$\mathbb{E}\left\{ \mathbb{V}_{\mathscr{H}_t}\{y_t(\mathbf{q})\} \middle| \mathscr{Y}_t \right\} \equiv \sigma_\xi^2 + \mathbb{E}\left\{ BX_t + \psi_{t,\mathbf{q}}^{\boldsymbol{o}}(X_t) \middle| \mathscr{Y}_t \right\}, \tag{3.117}$$

for all $t \in \mathbb{N}$, with $B \equiv \left[0\,1\,\mathbf{0}_{1\times(M-2)}\right] \in \mathbb{R}^{1\times M}$. For the second term on the RHS of (3.116), from (3.87) and after quite some algebra, we arrive at the expression

$$\phi_t^2(X_t, \mathbf{y}_t) = \mathbf{y}_t^{\boldsymbol{T}} \boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t)\, \mathbf{y}_t + \boldsymbol{\psi}_{t,\mathbf{q}}^{\alpha}(X_t)\, \mathbf{y}_t + \psi_{t,\mathbf{q}}^{\beta}(X_t), \tag{3.118}$$

holding true for all $t \in \mathbb{N}$, where $\boldsymbol{\psi}_{t,\mathbf{q}}^{\alpha}(\cdot)$, $\psi_{t,\mathbf{q}}^{\beta}(\cdot)$ and $\boldsymbol{\Gamma}_t^{\mathbf{q}}(\cdot)$ are defined as in (3.112), (3.113) and (3.114), respectively. Consequently, $\mathbb{V}_{\mathscr{Y}_t}\{y_t(\mathbf{q})\}$ can be expressed as

$$\begin{aligned}\mathbb{V}_{\mathscr{Y}_t}\{y_t(\mathbf{q})\} \equiv \sigma_\xi^2 &+ \mathbf{y}_t^{\boldsymbol{T}} \mathbb{E}\left\{ \boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t) \middle| \mathscr{Y}_t \right\} \mathbf{y}_t \\ &+ \mathbb{E}\left\{ \boldsymbol{\psi}_{t,\mathbf{q}}^{\alpha}(X_t) \middle| \mathscr{Y}_t \right\} \mathbf{y}_t - \widehat{y}_t^2(\mathbf{q}) \\ &+ \mathbb{E}\left\{ BX_t + \psi_{t,\mathbf{q}}^{\boldsymbol{o}}(X_t) + \psi_{t,\mathbf{q}}^{\beta}(X_t) \middle| \mathscr{Y}_t \right\}, \quad \forall t \in \mathbb{N}. \end{aligned} \tag{3.119}$$

Now, for all $t \in \mathbb{N}$, define the approximate filtering operator

$$\begin{aligned}\mathcal{V}_{\mathscr{Y}_t}^{L_S}(y_t(\mathbf{q})) \triangleq \sigma_\xi^2 &+ \mathbf{y}_t^{\boldsymbol{T}} \mathcal{E}^{L_S}\left(\boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t) \middle| \mathscr{Y}_t \right) \mathbf{y}_t \\ &+ \mathcal{E}^{L_S}\left(\boldsymbol{\psi}_{t,\mathbf{q}}^{\alpha}(X_t) \middle| \mathscr{Y}_t \right) \mathbf{y}_t - \left(\mathcal{E}^{L_S}\left(y_t(\mathbf{q}) \middle| \mathscr{Y}_t \right) \right)^2 \end{aligned}$$

$$\boldsymbol{\psi}_{t,\mathbf{q}}^{\alpha}(X_t) \triangleq 2A^{\mathbf{q}} X_t \left(\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t) - 2\left(\boldsymbol{A}_t X_t\right)^{\boldsymbol{T}} \boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t) \in \mathbb{R}^{1\times N} \tag{3.112}$$

$$\begin{aligned}\psi_{t,\mathbf{q}}^{\beta}(X_t) \triangleq A^{\mathbf{q}} X_t &\left(A^{\mathbf{q}} X_t - 2\left(\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t)\, \boldsymbol{A}_t X_t \right) \\ &+ \left(\boldsymbol{A}_t X_t\right)^{\boldsymbol{T}} \boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t)\, \boldsymbol{A}_t X_t \in \mathbb{R} \end{aligned} \tag{3.113}$$

$$\boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t) \triangleq \left(\left(\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t) \right)^{\boldsymbol{T}} \times \left(\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t) \in \mathbb{R}^{N\times N}, t \in \mathbb{N}. \tag{3.114}$$

$$+\,\mathcal{E}^{L_S}\left(BX_t+\psi^{\boldsymbol{o}}_{t,\mathbf{q}}\left(X_t\right)+\psi^{\beta}_{t,\mathbf{q}}\left(X_t\right)\middle|\,\mathscr{Y}_t\right), \quad (3.120)$$

where $\mathcal{E}^{L_S}\left(\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right)\in\mathbb{R}^{N\times N}$ is defined as

$$\mathcal{E}^{L_S}\left(\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right)\triangleq\frac{1}{\|E_t\|_1}\sum_{j\in\mathbb{N}^+_{L_S}}\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(\boldsymbol{x}^j_{L_S}\right)E_t\left(j\right). \quad (3.121)$$

Using the triangle inequality as in the proof of Theorem 3.4, the error $\left|\mathcal{V}^{L_S}_{\mathscr{Y}_t}\left(y_t\left(\mathbf{q}\right)\right)-\mathbb{V}_{\mathscr{Y}_t}\left\{y_t\left(\mathbf{q}\right)\right\}\right|$ is bounded from above by the sum of the errors between each pair of true and approximate estimators, respectively (excluding σ^2_ξ). Then, because of the boundedness and continuity of the functionals associated with the third and fifth terms on the RHS of (3.120), these will converge to the respective MMSE optimal estimators, in the sense of Theorem 3.3 (also see Appendix A). Thus, we only need to focus on the second and fourth terms on the RHS of (3.120).

For the second term, the Cauchy-Schwarz Inequality and the consistency of the spectral norm imply that

$$\begin{aligned}&\left|\mathbf{y}^{\boldsymbol{T}}_t\mathcal{E}^{L_S}\left(\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right)\mathbf{y}_t-\mathbf{y}^{\boldsymbol{T}}_t\mathbb{E}\left\{\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right\}\mathbf{y}_t\right|\\&\qquad\leq\|\mathbf{y}_t\|^2_2\left\|\mathcal{E}^{L_S}\left(\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right)-\mathbb{E}\left\{\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right\}\right\|_2. \quad (3.122)\end{aligned}$$

As in the proof of Theorem 3.4 presented above, we know that $\sup_{t\in\mathbb{N}_T}\|\mathbf{y}_t\left(\omega\right)\|^2_2<\gamma CN\left(1+\log\left(T+1\right)\right)$, for all $\omega\in\widehat{\Omega}_T\subseteq\Omega$, with $\mathcal{P}\left(\widehat{\Omega}_T\right)\geq 1-(T+1)^{1-CN}\exp\left(-CN\right)$. As a result, it suffices to show that

$$\left\|\mathcal{E}^{L_S}\left(\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right)-\mathbb{E}\left\{\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right\}\right\|_2\underset{L_S\to\infty}{\longrightarrow}0, \quad (3.123)$$

uniformly in $\widehat{\Omega}_T$ and for all $t\in\mathbb{N}_T$. Of course, it is true that $\mathbb{E}\left\{\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right\}\left(i,j\right)\equiv\mathbb{E}\left\{\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\left(i,j\right)\middle|\,\mathscr{Y}_t\right\}$, for all $(i,j)\in\mathbb{N}^+_N\times\mathbb{N}^+_N$, that is, the conditional expectation of $\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)$ given $\mathscr{Y}_t$ is the matrix formed by the respective conditional expectations of each entry. But observe that the approximate estimator $\mathcal{E}^{L_S}\left(\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right)$ in (3.121) also satisfies the same property. As a result, the error matrix $\mathcal{E}^{L_S}\left(\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right)-\mathbb{E}\left\{\boldsymbol{\Gamma}^{\mathbf{q}}_t\left(X_t\right)\middle|\,\mathscr{Y}_t\right\}\triangleq\boldsymbol{R}^{\mathscr{Y}_t}_{t,\mathbf{q}}$ contains as elements the errors of the corresponding entries. Additionally, we can easily show that

$$\left\|\boldsymbol{R}^{\mathscr{Y}_t}_{t,\mathbf{q}}\right\|_2\leq\left\|\boldsymbol{R}^{\mathscr{Y}_t}_{t,\mathbf{q}}\right\|_F\leq N\sup_{(i,j)\in\mathbb{N}^+_N\times\mathbb{N}^+_N}\left|\boldsymbol{R}^{\mathscr{Y}_t}_{t,\mathbf{q}}\left(i,j\right)\right|. \quad (3.124)$$

Now, all entries of $\boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t)$ are of the form

$$\boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t)(i,j) \equiv \left(\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t)(i) \times \left(\boldsymbol{\sigma}_t^{\mathbf{q}}(X_t)\right)^{\boldsymbol{T}} \mathbf{C}_t^{-1}(X_t)(j), \tag{3.125}$$

for all (i,j), from where it follows that each $\boldsymbol{\Gamma}_t^{\mathbf{q}}(X_t)(i,j)$ constitutes a bounded and continuous functional of the (almost surely compactly supported) state X_t, for all $t \in \mathbb{N}_T$. Therefore, the conditions of Theorem 3.3 are satisfied, from which it follows that $\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left\|\boldsymbol{R}_{t,\mathbf{q}}^{\mathscr{Y}_t}\right\|_2 \underset{L_S\to\infty}{\longrightarrow} 0$, which is what we were set to show.

Regarding the fourth term on the RHS of (3.120), it is true that

$$\left|\left(\mathcal{E}^{L_S}\left(y_t(\mathbf{q})|\,\mathscr{Y}_t\right)\right)^2 - \widehat{y}_t^2(\mathbf{q})\right| \equiv \left|\mathcal{E}^{L_S}\left(y_t(\mathbf{q})|\,\mathscr{Y}_t\right) - \widehat{y}_t(\mathbf{q})\right|\left|\mathcal{E}^{L_S}\left(y_t(\mathbf{q})|\,\mathscr{Y}_t\right) + \widehat{y}_t(\mathbf{q})\right|. \tag{3.126}$$

Since, from Theorem 3.4, we know that the error $\left|\mathcal{E}^{L_S}\left(y_t(\mathbf{q})|\,\mathscr{Y}_t\right) - \widehat{y}_t(\mathbf{q})\right|$ converges to zero in the sense of Theorem 3.3, it will suffice to show that the quantity $\left|\mathcal{E}^{L_S}\left(y_t(\mathbf{q})|\,\mathscr{Y}_t\right) + \widehat{y}_t(\mathbf{q})\right|$ is uniformly bounded in $\widehat{\Omega}_T$ and for all $t \in \mathbb{N}_T$. Indeed, the triangle inequality implies that

$$\left|\mathcal{E}^{L_S}\left(y_t(\mathbf{q})|\,\mathscr{Y}_t\right) + \widehat{y}_t(\mathbf{q})\right| \leq \left|\mathcal{E}^{L_S}\left(y_t(\mathbf{q})|\,\mathscr{Y}_t\right)\right| + \left|\mathbb{E}\left\{\phi_t(X_t,\mathbf{y}_t)|\,\mathscr{Y}_t\right\}\right|. \tag{3.127}$$

Again from the Cauchy-Schwarz Inequality, it can be easily shown that

$$\left|\mathbb{E}\left\{\phi_t(X_t,\mathbf{y}_t)|\,\mathscr{Y}_t\right\}\right| \leq C_{t,\mathbf{q}}^1 + C_{t,\mathbf{q}}^2\left\|\mathbf{y}_t\right\|_2, \tag{3.128}$$

where

$$C_{t,\mathbf{q}}^1 \triangleq \sup_{\boldsymbol{x}\in\mathcal{Z}}\left|A^{\mathbf{q}}\boldsymbol{x} + \left(\boldsymbol{\sigma}_t^{\mathbf{q}}(\boldsymbol{x})\right)^{\boldsymbol{T}}\mathbf{C}_t^{-1}(\boldsymbol{x})\,\boldsymbol{A}_t\boldsymbol{x}\right| \quad \text{and} \tag{3.129}$$

$$C_{t,\mathbf{q}}^2 \triangleq \sup_{\boldsymbol{x}\in\mathcal{Z}}\left\|\left(\boldsymbol{\sigma}_t^{\mathbf{q}}(\boldsymbol{x})\right)^{\boldsymbol{T}}\mathbf{C}_t^{-1}(\boldsymbol{x})\right\|_2, \tag{3.130}$$

implying that $\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left|\mathbb{E}\left\{\phi_t(X_t,\mathbf{y}_t)|\,\mathscr{Y}_t\right\}\right| \leq C_{\mathbf{q}}$, with

$$C_{\mathbf{q}} \triangleq \sup_{t\in\mathbb{N}_T} C_{t,\mathbf{q}}^1 + \left(\sqrt{\gamma C N\left(1+\log(T+1)\right)}\right)\sup_{t\in\mathbb{N}_T} C_{t,\mathbf{q}}^2, \tag{3.131}$$

where the RHS is always finite, by assumption. For the first term on the RHS of (3.127), by definition of $\mathcal{E}^{L_S}\left(y_t(\mathbf{q})|\,\mathscr{Y}_t\right)$ and using Hölder's Inequality, we have (see Theorem

3.4)

$$\left|\mathcal{E}^{L_S}\left(y_t\left(\mathbf{q}\right)|\,\mathscr{Y}_t\right)\right| \leq \sup_{j\in\mathbb{N}_{L_S}^{+}}\left|\phi_t\left(\boldsymbol{x}_{L_S}^{j},\mathbf{y}_t\right)\right|$$
$$\leq C_{t,\mathbf{q}}^{1}+C_{t,\mathbf{q}}^{2}\left\|\mathbf{y}_t\right\|_2, \tag{3.132}$$

as above, yielding the same uniform bound for $\left|\mathcal{E}^{L_S}\left(y_t\left(\mathbf{q}\right)|\,\mathscr{Y}_t\right)\right|$. The above lead directly to the inequality

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left|\mathcal{E}^{L_S}\left(y_t\left(\mathbf{q}\right)|\,\mathscr{Y}_t\right)+\widehat{y}_t\left(\mathbf{q}\right)\right|\leq 2C_{\mathbf{q}}, \tag{3.133}$$

showing uniform boundedness in $\widehat{\Omega}_T$ and $\mathbb{N}_T$. Therefore,

$$\sup_{t\in\mathbb{N}_T}\sup_{\omega\in\widehat{\Omega}_T}\left|\mathcal{V}_{\mathscr{Y}_t}^{L_S}\left(y_t\left(\mathbf{q}\right)\right)-\mathbb{V}_{\mathscr{Y}_t}\left\{y_t\left(\mathbf{q}\right)\right\}\right|\underset{L_S\to\infty}{\longrightarrow}0. \tag{3.134}$$

The explicit form of $\mathcal{V}_{\mathscr{Y}_t}^{L_S}\left(y_t\left(\mathbf{q}\right)\right)$ presented in Theorem 3.5 can be determined after some algebraic manipulations of (3.120).

The case where $\rho>0$ is similar, albeit simpler. Specifically, the conditional variance of $y_{t+\rho}\left(\mathbf{q}\right)$ given $\mathscr{Y}_t$ is given by $\mathbb{V}_{\mathscr{Y}_t}\left\{y_{t+\rho}\left(\mathbf{q}\right)\right\}\triangleq\mathbb{E}\left\{y_{t+\rho}^{2}\left(\mathbf{q}\right)\middle|\,\mathscr{Y}_t\right\}-\widehat{y}_{t+\rho}^{2}\left(\mathbf{q}\right)$, for all $t\in\mathbb{N}$. Also, recall the definition of the filtration $\left\{\mathscr{Q}_t^{+\rho}\right\}_{t\in\mathbb{N}}$ in the proof of Theorem 3.4. Then, the law of total variance reads

$$\mathbb{V}_{\mathscr{Y}_t}\left\{y_{t+\rho}\left(\mathbf{q}\right)\right\}$$
$$\equiv\mathbb{E}\left\{\mathbb{V}_{\mathscr{Q}_t^{+\rho}}\left\{y_{t+\rho}\left(\mathbf{q}\right)\right\}\middle|\,\mathscr{Y}_t\right\}+\mathbb{E}\left\{\left(A^{\mathbf{q}}X_{t+\rho}\right)^2\middle|\,\mathscr{Y}_t\right\}-\widehat{y}_{t+\rho}^{2}\left(\mathbf{q}\right),\quad\forall t\in\mathbb{N}. \tag{3.135}$$

Exploiting the fact that $y_{t+\rho}\left(\mathbf{q}\right)$ is conditionally independent of $\mathbf{y}_t^{aug},\mathbf{y}_{t-1}^{aug},\ldots,$ *given the state* at time $t+\rho$, $X_{t+\rho}$ and using the tower property, it can be shown that $\mathbb{V}_{\mathscr{Q}_t^{+\rho}}\left\{y_{t+\rho}\left(\mathbf{q}\right)\right\}=BX_{t+\rho}+\sigma_{\xi}^{2}$, yielding

$$\mathbb{V}_{\mathscr{Y}_t}\left\{y_{t+\rho}\left(\mathbf{q}\right)\right\}$$
$$\equiv\sigma_{\xi}^{2}+\mathbb{E}\left\{BX_{t+\rho}\middle|\,\mathscr{Y}_t\right\}+\mathbb{E}\left\{\left(A^{\mathbf{q}}X_{t+\rho}\right)^2\middle|\,\mathscr{Y}_t\right\}-\widehat{y}_{t+\rho}^{2}\left(\mathbf{q}\right),\quad\forall t\in\mathbb{N}. \tag{3.136}$$

Let us define the approximate predictor

$$\mathcal{V}_{\mathscr{Y}_t}^{L_S}\left(y_{t+\rho}\left(\mathbf{q}\right)\right)$$

$$\triangleq \sigma_{\xi}^{2} + \mathcal{E}^{L_S}\left(\left(A^{\mathbf{q}}X_{t+\rho}\right)^2 + BX_{t+\rho}\middle|\,\mathscr{Y}_t\right) - \left(\mathcal{E}^{L_S}\left(y_{t+\rho}\left(\mathbf{q}\right)\middle|\,\mathscr{Y}_t\right)\right)^2, \quad \forall t \in \mathbb{N}. \quad (3.137)$$

Since $\left(A^{\mathbf{q}}X_{t+\rho}\right)^2 + BX_{t+\rho}$ is functionally independent of the observations, Theorem 3.3 can be applied to arrive at the estimator of Theorem 3.5. Further, the convergence analysis of the approximate predictor is very similar to the case where $\rho \equiv 0$, treated in detail above, and it is omitted. QED. ■

3.2.8 A Note on Practical Applicability

The validation and verification of the proposed model based approach to spatiotemporal wireless channel estimation/prediction, as well as its adaptation, on the basis of real experimental data, is a challenging task. It constitutes an important topic for further investigation in our research agenda.

In particular, we are currently in the process of developing an efficient meta-method, termed as Markovian Channel Profiling (MCP), which will ultimately solve the respective parameter estimation problem, enabling the use of the channel map tracking methods described above. Additionally, efforts are currently made for the complete experimental validation of the proposed channel modeling approach, using real world wireless channel measurements, coming from inexpensive sensing equipment, such as consumer oriented cell phones.

Chapter 4

Spatially Controlled Relay Beamforming: 2-Stage Optimal Policies

4.1 Introduction

Distributed, networked communication systems, such as relay beamforming networks [1, 2, 6, 7, 9, 10, 71] (e.g., Amplify & Forward (AF)) are typically designed without explicitly considering how the positions of the networking nodes might affect the quality of the communication. Optimum physical placement of assisting networking nodes, which could potentially improve the quality of the communication, does not constitute a clear network design aspect. However, in most practical settings in physical layer communications, the Channel State Information (CSI) observed by each networking node, per channel use, although (modeled as) random, it is both spatially and temporally correlated. It is, therefore, reasonable to ask if and how system performance could be improved by controlling the positions of certain network nodes, based on causal side (CSI) information, and exploiting the spatiotemporal dependencies of the wireless medium.

Recently, autonomous node mobility has been proposed as an effective means to further enhance performance in various distributed network settings. In [72], optimal transmit AF beamforming has been combined with potential field based relay mobility control in multiuser cooperative networks, in order to minimize relay transmit power, while meeting certain Quality-of-Service (QoS) constraints. In [50], in the framework of information theoretic physical layer security, decentralized jammer motion control has been jointly combined with noise nulling and cooperative jamming, maximizing the network secrecy rate. In [51], optimal relay positioning has been studied in systems where multiple relays deliver information to a destination, in the presence of an eavesdropper,

with a goal of maximizing or achieving a target level of ergodic secrecy.

In the complementary context of communication aware (comm-aware) robotics, node mobility has been exploited in distributed robotic networks, in order to enhance system performance, in terms of maintaining reliable, in-network communication connectivity [73–76], and optimizing network energy management [77]. Networked node motion control has also been exploited in special purpose applications, such as networked robotic surveillance [78] and target tracking [79].

In [50,51,72], the links among the nodes of the network (or the related statistics) are assumed to be available in the form of static channel maps, during the whole motion of the jammers/relays. However, this is an oversimplifying assumption in scenarios where the channels change significantly in time and space [56, 67, 68].

In this paper, we try to overcome this major limitation, and we consider the problem of optimally and dynamically updating relay positions in one source/destination relay beamforming networks, *in a dynamic channel environment*. Different from [50, 51, 72], we model the wireless channel as a *spatiotemporal stochastic field*; this approach may be seen as a versatile extension of a realistic, commonly employed "log-normal" channel model [56]. We then propose a 2-*stage stochastic programming* problem formulation, optimally specifying the positions of the relays *at each time slot*, such that the *Signal-to-Interference+Noise Ratio (SINR)* or *QoS* at the destination, at the same time slot, is maximized *on average*, based on *causal CSI*, and subject to a *total power constraint at the relays*. At each time slot, the relays not only beamform to the destination, but also optimally, *predictively* decide their positions at the next time slot, based on their experience (causal actions and channel observations). This novel, cyber-physical system approach to relay beamforming is termed as *Spatially Controlled Relay Beamforming*.

Exploiting the assumed stochastic channel structure, it is first shown that the proposed optimal motion control problem is equivalent to a set of simpler, two dimensional subproblems, which can be solved in a *distributed fashion*, one at each relay, *without the need for intermediate exchange of messages* among the relays. However, each the objectives of the aforementioned subproblems involves the evaluation of a conditional expectation of a well defined ratio of almost surely positive random variables, which is

impossible to perform analytically, calling for the development of easily implementable approximations to each of the original problems. Two such heuristics are considered. The first is based on the so-called *Method of Statistical Differentials* [80], whereas the second constitutes a brute force approach, based on the *multidimensional Gauss-Hermite Quadrature Rule*, a readily available routine for numerical integration. In both cases, the original problem objective is replaced by the respective approximation, which, in both cases, is shown to be easily computed via simple, closed form expressions. The computational complexity of both approaches is also discussed and characterized. Subsequently, we present an important result, along with the respective detailed technical development, characterizing the performance of the proposed system, *across time slots* (Theorems 4.4 and 4.5). In a nutshell, this result states that, although our *problem objective is itself myopic* at each time slot, the *expected network QoS exhibits an increasing trend across time slots* (in other words, the expected QoS increases in time, within a small positive slack), under optimal decision making at the relays. Lastly, we present representative numerical simulations, experimentally confirming both the efficacy and feasibility of the proposed approach, as well as the validity of our theoretical predictions.

During exposition of the proposed spatially controlled relay beamforming system, we concurrently develop and utilize a rigorous discussion concerning the optimality of our approach, and with interesting results (Section 4.5.1 / Appendix B). Clearly, our problem formulation is challenging; it involves a *variational* stochastic optimization problem, where, at each time slot, the decision variable, *a function* of the so far available useful information in the system (also called a *policy*, or a *decision rule*), *constitutes itself the spatial coordinates, from which every network relay will observe the underlying spatiotemporal channel field, at the next time slot.* In other words, our formulation requires solving an *optimal spatial field sampling problem, in a dynamic fashion.* Such a problem raises certain fundamental questions, not only related to our proposed spatially controlled beamforming formulation, but also to a large class of variational stochastic programs of similar structure.

In this respect, our contributions are partially driven by assuming an underlying

complete base probability space of otherwise *arbitrary structure*, generating all random phenomena considered in this work. Under this general setting, we explicitly identify sufficient conditions, which guarantee the validity of the so-called *substitution rule for conditional expectations*, specialized to such expectations of random spatial (in general) fields/functions *with an also random spatial parameter*, relative to some σ-algebra, which makes the latter parameter measurable (fixed) (Definition 4.6 & Theorem 4.6). General validity of the substitution rule, without imposing additional, special conditions, traces back to the existence of regular conditional distributions, defined *directly* on the sample space of the underlying base probability space. Such regular conditional distributions cannot be guaranteed to exist, unless the sample space has nice topological properties, for instance, if it is Polish [81]. In the context of our spatially controlled beamforming application, such structural requirements on the sample space, which, by assumption, is conceived as a model of "nature", and generates the spatiotemporal channel field sampled by the relays, are simply not reasonable. Considering this, our first contribution is to show that it is possible to guarantee the validity of the form of the substitution rule under consideration by imposing conditions on the topological structure of the involved random field, rather than that of the sample space (a part of its domain). This results in a rather generally applicable problem setting (Theorem 4.6).

In this work, the validity of the substitution rule is ascertained by imposing simple continuity assumptions on the random functions involved, which, in some cases, might be considered somewhat restrictive. Nevertheless, those assumptions can be significantly weakened, guaranteeing the validity of the substitution rule for vastly discontinuous random functions, including, for instance, cases with random discontinuities, or random jumps. The development of this extended analysis, though, is out of the scope of this paper, and will be presented elsewhere.

The validity of the substitution rule is vitally important in the treatment of a wide class of variational stochastic programs, including that involved in the proposed spatially controlled beamforming approach. In particular, leveraging the power of the substitution rule, we develop a version of the so-called *Fundamental Lemma of Stochastic*

Control (FLSC) [82–87] (Lemma 4.2), which provides sufficient conditions that permit interchange of integration (expectation) and max/minimization in general variational (stochastic) programming settings. The FLSC allows the initial variational problem to be *exchanged* by a related, though *pointwise* (ordinary) optimization problem, thus efficiently reducing the search over a class of functions (initial problem) to searching over constants, which is, of course, a standard and much more handleable optimization setting. In slightly different ways, the FLSC is evidently utilized in relevant optimality analysis both in Stochastic Programming [84, 85], and in Dynamic Programming & Stochastic Optimal Control [82, 83, 86, 87].

A very general version of the FLSC is given in ([84], Theorem 14.60), where unconstrained variational optimization of integrals of extended real-valued *random lower semicontinuous functions* [85], or, by another name, *normal integrands* [84], with respect to a general σ-finite measure, is considered. Our version of the FLSC may be considered a useful variation of Theorem 14.60 in [84], and considers *constrained* variational optimization problems involving integrals of random functions, but with respect to some base *probability* measure (that is, expectations). In our result, via the tower property of expectations, the role of the normal integrand in ([84], Theorem 14.60) is played by the conditional expectation of the random function considered, relative to a σ-algebra, which makes the respective decision variable of the problem (a function(al)) measurable. Assuming a base probability space of arbitrary structure, this argument is justified by assuming validity of the substitution rule, which, in turn, is ascertained under our previously developed sufficient conditions. Different from ([84], Theorem 14.60), in our version of the FLSC, apart from natural Borel measurability requirements, *no continuity assumptions are directly imposed* on the structure on either the random function, or the respective conditional expectation. In this respect, our result extends ([84], Theorem 14.60), and is of independent interest.

On the other hand, from the strongly related perspective of Stochastic Optimal Control, our version of the FLSC may be considered as the basic building block for further development of Bellman Equation-type, Dynamic Programming solutions [86, 87], under a strictly Borel measurability framework, sufficient for our purposes. Quite

differently though, in our formulation, the respective cost (at each stage of the problem) is itself a random function (a spatial field), whose domain is the Cartesian product of a base space of arbitrary topology, with another, nicely behaved Borel space, instead of the usual Cartesian product of two Borel spaces (the spaces of state and controls), as in the standard dynamic programming setting [86, 87]. Essentially, our formulation is "one step back" as compared to the basic dynamic programming model of [86, 87], in the sense that the cost considered herein refers *directly* back to the base space. As a result, different treatment of the problem is required; essentially, the validity of the substitution rule for our cost function bypasses the requirement for existence of conditional distributions, and exploits potential nice properties of the respective conditional cost (in our case, joint *Borel* measurability).

Finally, *emphasizing on our particular problem formulation*, our functional assumptions, which guarantee the validity of the substitution rule, combined with the FLSC, result in a total of six sufficient conditions, under which strict optimality via problem exchangeability is guaranteed (conditions **C1-C6** in Lemma 4.3). Those conditions are subsequently shown to be satisfied specifically for the spatially controlled beamforming problem under consideration (verification Theorem 4.1), ensuring strict optimality of a solution obtained by exploiting problem exchangeability.

4.2 Spatially Controlled Relay Beamforming

As mentioned above, the beamforming objective adopted will be *maximization of the Signal-to-Interference+Noise Ratio (SINR) at the destination (measuring network QoS)*, under a total power budget at the relays, as in Section. For the single-source single-destination setting considered herein, the aforementioned beamforming problem admits a closed form solution, a fact which will be important in deriving optimal relay motion control policies, in a tractable fashion. But first, let us present the general scheduling schema of the proposed mobile beamforming system, as well as some technical preliminaries on stochastic programming and optimal control, which will be used repeatedly in the analysis to follow.

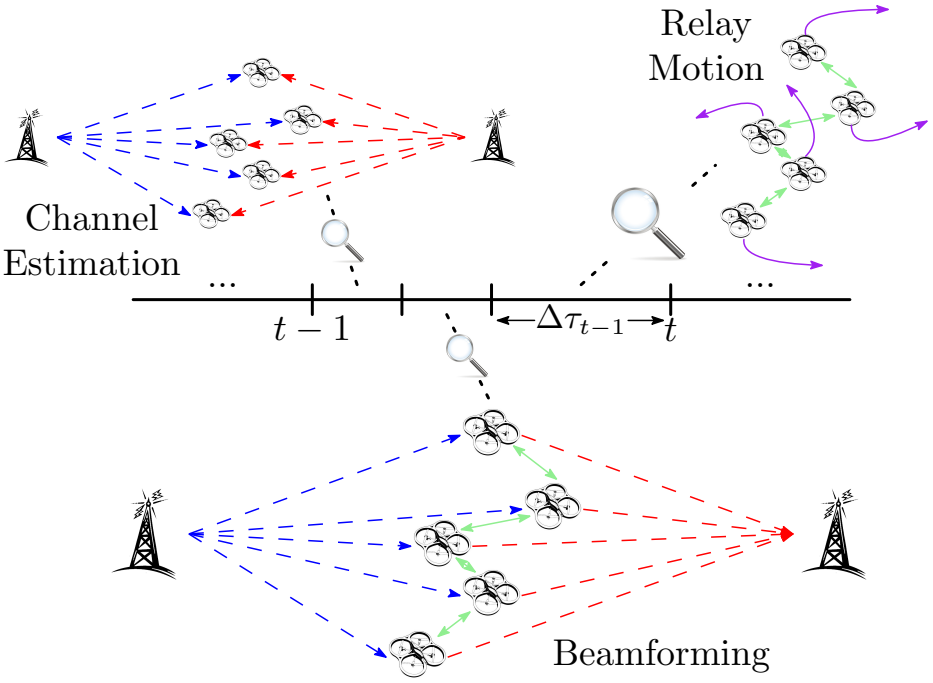


Figure 4.1: Proposed TDMA-like joint scheduling protocol for communications and controls.

4.2.1 Joint Scheduling of Communications & Controls

At each time slot $t \in \mathbb{N}^{+}_{N_T}$ and assuming the same carrier for all communication tasks, we employ a basic joint communication/decision making TDMA-like protocol, as follows:

1. The source broadcasts a pilot signal to the relays, which then estimate their respective channels relative to the source.

2. The same procedure is carried out for the channels relative to the destination.

3. Then, based on the estimated CSI, the relays beamform in AF mode (assume perfect CSI estimation).

4. Based on the CSI received *so far*, strategic decision making is implemented, motion controllers of the relays are determined and relays are steered to their updated positions.

The above sequence of actions is repeated for all N_T time slots, corresponding to the total operational horizon of the system. This simple scheduling protocol is graphically depicted in Fig. 4.1.

Concerning relay kinematics, it is assumed that the relays obey the differential equation

$$\dot{\mathbf{p}}\left(\tau\right) \equiv \mathbf{u}\left(\tau\right), \quad \forall \tau \in \left[0, T\right], \tag{4.1}$$

where $\mathbf{u} \triangleq \left[\mathbf{u}_1 \ \ldots \ \mathbf{u}_R\right]^{\boldsymbol{T}} \in \mathcal{S}^R$, with $\mathbf{u}_i : \left[0, T\right] \to \mathcal{S}$ being the motion controller of relay $i \in \mathbb{N}_R^+$. Apparently, relay motion is in continuous time. However, assuming the relays may move *only after their controls have been determined and up to the start of the next time slot*, we can write

$$\mathbf{p}\left(t\right) \equiv \mathbf{p}\left(t-1\right) + \int_{\Delta\tau_{t-1}} \mathbf{u}_{t-1}\left(\tau\right) \mathrm{d}\tau, \quad \forall t \in \mathbb{N}_{N_T}^2, \tag{4.2}$$

with $\mathbf{p}\left(1\right) \equiv \mathbf{p}_{init}$, and where $\Delta\tau_t \subset \mathbb{R}$ and $\mathbf{u}_t : \Delta\tau_t \to \mathcal{S}^R$ denote the time interval that the relays are allowed to move in and the respective relay controller, in each time slot $t \in \mathbb{N}_{N_T-1}^+$. It holds that $\mathbf{u}\left(\tau\right) \equiv \sum_{t \in \mathbb{N}_{N_T-1}^+} \mathbf{u}_t\left(\tau\right) \mathbb{1}_{\Delta\tau_t}\left(\tau\right)$, *where τ belongs in the first $N_T - 1$* time slots. Of course, at each time slot t, $\Delta\tau_t$ must be sufficiently small such that the temporal correlations of the CSI at adjacent time slots are sufficiently strong. These correlations are controlled by the correlation time parameter γ, which can be a function of the slot width. Therefore, the velocity of the relays must be of the order of $\left(\Delta\tau_t\right)^{-1}$. In this work, though, we assume that the relays are not explicitly resource constrained, in terms of their motion.

Now, regarding the form of the relay motion controllers $\mathbf{u}_{t-1}\left(\tau\right), \tau \in \Delta\tau_{t-1}$, *given a goal position vector at time slot t*, $\mathbf{p}^o\left(t\right)$, it suffices to fix a path in $\mathcal{S}^R$, such that the points $\mathbf{p}^o\left(t\right)$ and $\mathbf{p}\left(t-1\right)$ are connected in at most time $\Delta\tau_t$. A generic choice for such a path is the straight line[1] connecting $\mathbf{p}_i^o\left(t\right)$ and $\mathbf{p}_i\left(t-1\right)$, for all $i \in \mathbb{N}_R^+$. Therefore, we may choose the relay controllers at time slot $t-1 \in \mathbb{N}_{N_T-1}^+$ as

$$\mathbf{u}_{t-1}^o\left(\tau\right) \triangleq \frac{1}{\Delta\tau_{t-1}} \left(\mathbf{p}^o\left(t\right) - \mathbf{p}\left(t-1\right)\right), \quad \forall \tau \in \Delta\tau_{t-1}. \tag{4.3}$$

As a result, any motion control problem considered hereafter can now be formulated in terms of specifying the goal relay positions at the next time slot, given their positions at the current time slot (and the observed CSI).

---

[1]Caution is needed here, due to the possibility of physical collisions among relays themselves, or among relays and other physical obstacles in the workspace, $\mathcal{S}$. Nevertheless, for simplicity, we assume that either such events never occur, or that, if they do, there exists some transparent collision avoidance mechanism implemented at each relay, which is out of our direct control.

In the following, let $\mathscr{C}(\mathcal{T}_t)$ denote the set of channel gains observed by the relays, *along the paths of their point trajectories* $\mathcal{T}_t \triangleq \{\mathbf{p}(1) \dots \mathbf{p}(t)\}$, $t \in \mathbb{N}^+_{N_T}$. Then, $\mathcal{T}_t$ may be recursively updated as $\mathcal{T}_t \equiv \mathcal{T}_{t-1} \cup \{\mathbf{p}(t)\}$, for all $t \in \mathbb{N}^+_{N_T}$, with $\mathcal{T}_0 \triangleq \varnothing$. In a technically precise sense, $\{\mathscr{C}(\mathcal{T}_t)\}_{t \in \mathbb{N}^+_{N_T}}$ will also denote the filtration generated by the CSI observed at the relays, *along* $\mathcal{T}_t$, interchangeably. In other words, in case the trajectories of the relays are themselves random, then $\mathscr{C}(\mathcal{T}_t)$ denotes the σ-algebra generated by *both* the CSI observed up to and including time slot t *and* $\mathbf{p}(1) \dots \mathbf{p}(t)$, for all $t \in \mathbb{N}^+_{N_T}$. Additionally, we define $\mathscr{C}(\mathcal{T}_0) \equiv \mathscr{C}(\{\varnothing\})$ as $\mathscr{C}(\mathcal{T}_0) \triangleq \{\varnothing, \Omega\}$, that is, as the trivial σ-algebra, and we may occasionally refer to time $t \equiv 0$, as a *dummy time slot*, by convention.

4.2.2 2-Stage Stochastic Optimization of Beamforming Weights and Relay Positions: Base Formulation & Methodology

At each time slot $t \in \mathbb{N}^+_{N_T}$, given the current CSI encoded in $\mathscr{C}(\mathcal{T}_t)$, we are interested in determining $\boldsymbol{w}^o(t) \triangleq [w_1(t)\ w_2(t) \dots w_R(t)]^{\boldsymbol{T}}$, as an optimal solution to a beamforming optimization problem, as a functional of $\mathscr{C}(\mathcal{T}_t)$. Let the *optimal value* (say infimum) of this problem be the process $V_t \equiv V(\mathbf{p}(t), t)$, a functional of the CSI encoded in $\mathscr{C}(\mathcal{T}_t)$, depending on the positions of the relays at time slot t.

Suppose that, *at time slot* $t-1$, an oracle reveals $\mathscr{C}(\mathcal{T}_t \equiv \mathcal{T}_{t-1} \cup \{\mathbf{p}(t)\})$, which also determines the channels corresponding to the new positions of the relays at the next time slot t. Then, we could further consider optimizing V_t with respect to $\mathbf{p}(t)$, representing the new position of the relays. But note that, $\mathscr{C}(\mathcal{T}_t)$ is not physically observable and in the absence of the oracle, optimizing V_t with respect to $\mathbf{p}(t)$ is impossible, since given $\mathscr{C}(\mathcal{T}_{t-1})$, the channels at any position of the relays are nontrivial random variables. However, it is reasonable to search for the best decision on the positions of the relays at time slot t, as a functional of the available information encoded in $\mathscr{C}(\mathcal{T}_{t-1})$, such that V_t is optimized *on average*. This procedure may be formally formulated as a 2-*stage*

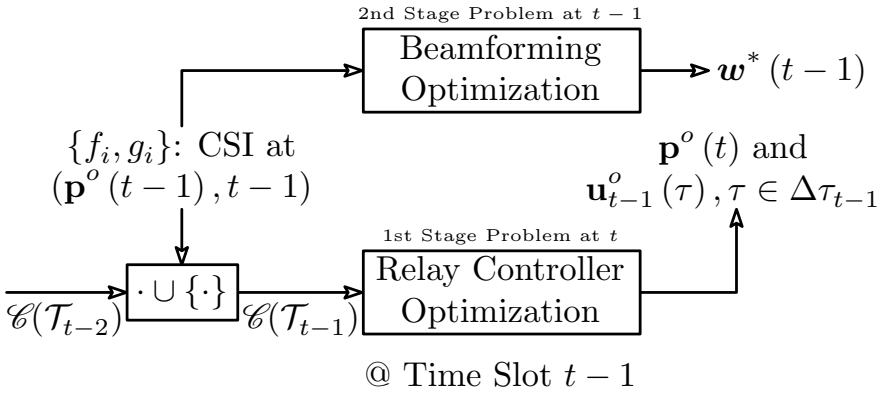


Figure 4.2: 2-Stage optimization of beamforming weights and spatial relay controllers. The variables $\boldsymbol{w}^o(t-1)$, $\mathbf{u}^o_{t-1}$ and $\mathbf{p}^o(t)$ denote the optimal beamforming weights and relay controllers at time slot $t-1$, and the optimal relay positions at time slot t, respectively.

stochastic program [85],

$$\begin{aligned} \underset{\mathbf{p}(t)}{\text{minimize}} \quad & \mathbb{E}\left\{V\left(\mathbf{p}\left(t\right), t\right)\right\} \\ \text{subject to} \quad & \mathbf{p}\left(t\right) \equiv \mathcal{M}\left(\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right) \in \mathcal{C}\left(\mathbf{p}^o\left(t-1\right)\right), \ , \\ & \text{for some } \mathcal{M} : \mathbb{R}^{2R(t-1)} \to \mathbb{R}^{2R} \end{aligned} \tag{4.4}$$

to be solved at each $t-1 \in \mathbb{N}^+_{N_T-1}$, where $\mathcal{C} : \mathbb{R}^{2R} \rightrightarrows \mathbb{R}^{2R}$ is a multifunction, with $\mathcal{C}\left(\mathbf{p}^o\left(t-1\right)\right) \subseteq \mathcal{S}^R$ representing a physically feasible spatial neighborhood around the point $\mathbf{p}^o\left(t-1\right) \in \mathcal{S}^R$, the decision vector *selected* at time $t-2 \in \mathbb{N}_{N_T-2}$ (recall that $t \equiv 0$ denotes a dummy time slot). Note that, in general, the decision *selected* at $t-2$, $\mathbf{p}^o\left(t-1\right)$, may not be an optimal decision for the respective problem solved at $t-2$ and implemented at $t-1$. To distinguish $\mathbf{p}^o\left(t-1\right)$ from an optimal decision at $t-2$, the latter will be denoted as $\mathbf{p}^*\left(t-1\right)$, for all $t \in \mathbb{N}^2_{N_T}$. Also note that, in order for (4.4) to be well defined, important technical issues, such as measurability of V_t and existence of its expectation at least for each feasible decision $\mathbf{p}\left(t\right)$, should be precisely resolved. Problem (4.5), together with the respective beamforming problem with optimal value V_t (which will focus on shortly) are referred to as the *first-stage problem* and the *second-stage problem*, respectively [85]. Hereafter, aligned with the literature, any feasible choice for the decision variable $\mathbf{p}\left(t\right)$ in (4.5), will be interchangeably called an *(admissible) policy*. A generic block representation of the proposed 2-stage stochastic programming approach is depicted in Fig. 4.2.

Mainly due to the arbitrary structure of the function $\mathcal{M}$, (4.4) is too general to consider, within a reasonable analytical framework. Thus, let us slightly constrain the decision set of (4.4) to include *only measurable decisions*, resulting in the formulation

$$\begin{aligned} \underset{\mathbf{p}(t)}{\text{minimize}} \quad & \mathbb{E}\left\{V\left(\mathbf{p}\left(t\right),t\right)\right\} \\ \text{subject to} \quad & \mathbf{p}\left(t\right) \equiv \mathcal{M}\left(\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right) \in \mathcal{C}\left(\mathbf{p}^{o}\left(t-1\right)\right), \\ & \mathcal{M}^{-1}\left(\mathcal{A}\right) \in \mathscr{B}\left(\mathbb{R}^{2R}\right), \forall \mathcal{A} \in \mathscr{B}\left(\mathbb{R}^{2R(t-1)}\right) \end{aligned} \quad , \tag{4.5}$$

provided, of course, that the stochastic program (4.5) is well defined. The second constraint in (4.5) is equivalent to $\mathcal{M}$ being Borel measurable, instead of being any arbitrary function, as in (4.4).

Provided its well definiteness, the stochastic program (4.5) is difficult to solve, most importantly because of its *variational character*; the decision variable $\mathbf{p}\left(t\right)$ is constrained to be a functional of the CSI observed up to and including time $t-1$. A very powerful tool, which will enable us to both make (4.5) meaningful and overcome the aforementioned difficulty, is the *Fundamental Lemma of Stochastic Control* [82–87], which in fact refers to a family of technical results related to the *interchangeability* of integration (expectation) and minimization in general stochastic programming. Under the framework of the Fundamental Lemma, in Appendix A, we present a detailed discussion, best suited for the purposes of this work, which is related to the important technical issues, arising when one wishes to meaningfully define and tractably simplify "hard", variational problems of the form of (4.5).

In particular, Lemma 4.3, presented in Section 4.5.1.4 (Appendix A), identifies six sufficient technical conditions (conditions **C1-C6**, see statement of Lemma 4.3), under which the variational problem (4.5) is *exchangeable* by the structurally simpler, *pointwise* optimization problem

$$\begin{aligned} \underset{\mathbf{p}(t)}{\text{minimize}} \quad & \mathbb{E}\left\{V\left(\mathbf{p}\left(t\right),t\right)|\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\} \\ \text{subject to} \quad & \mathbf{p}\left(t\right) \in \mathcal{C}\left(\mathbf{p}^{o}\left(t-1\right)\right) \end{aligned} \quad , \tag{4.6}$$

to be solved at each $t-1 \in \mathbb{N}_{N_T-1}^{+}$. Observe that, in (4.6), the decision variable $\mathbf{p}\left(t\right)$ is constant, as opposed to (4.5), where the decision variable $\mathbf{p}\left(t\right)$ is itself a functional of the observed information at time slot $t-1$, that is, a policy. Provided that CSI

$\mathscr{C}(\mathcal{T}_{t-1})$ and $\mathbf{p}^o(t-1)$ are known and that the involved conditional expectation can be somehow evaluated, (4.6) constitutes an ordinary, nonlinear optimization problem.

If Lemma 4.3 is in power, *exchangeability* of (4.5) by (4.6) is understood in the sense that the optimal value of (4.5), which is a number, coincides with the *expectation* of optimal value of (4.6), which turns out to be a measurable function of $\mathscr{C}(\mathcal{T}_{t-1})$. In other words, minimization is *interchangeable* with integration, in the sense that

$$\inf_{\mathbf{p}(t)\in\mathcal{D}_t} \mathbb{E}\left\{V\left(\mathbf{p}\left(t\right),t\right)\right\} \equiv \mathbb{E}\left\{\inf_{\mathbf{p}(t)\in\mathcal{C}\left(\mathbf{p}^o(t-1)\right)} \mathbb{E}\left\{V\left(\mathbf{p}\left(t\right),t\right)|\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}\right\}, \tag{4.7}$$

for all $t \in \mathbb{N}_{N_T}^2$, where $\mathcal{D}_t$ denotes the set of feasible decisions for (4.5). What is more, under the aforementioned technical conditions of Lemma 4.3, exchangeability implies that, if there exists an admissible policy of (4.5), say $\mathbf{p}^*(t)$, which solves (4.6), then $\mathbf{p}^*(t)$ is also optimal for (4.5). Additionally, Lemma 4.3 implies existence of at least one optimal solution to (4.6), which is simultaneously feasible and, thus, optimal, for the original stochastic program (4.5). If, further, (4.6) features a unique optimal solution, say $\mathbf{p}^*(t)$, then $\mathbf{p}^*(t)$ must be an optimal solution to (4.5).

In the next subsection, we will specify the optimal value of the second-stage subproblem, V_t, for each time $t \in \mathbb{N}_{N_T}^+$. That is, we will consider a fixed criterion for implementing relay beamforming (recourse actions) at each t, *after* the predictive decisions on the positions of the relays have been made (*at time* $t-1$) and the relays have moved to their new positions, implying that the CSI at time at time t has been revealed. Of course, one of the involved challenges will be to explicitly show that Conditions **C1-C6** are satisfied for each case considered, so that we can focus on solving the ordinary nonlinear optimization problem (4.6), instead of the much more difficult variational problem (4.5). The other challenge we will face is actually solving (4.6).

Remark 4.1. It would be important to note that the pointwise problem (4.5) admits a reasonable and intuitive interpretation: At each time slot $t-1$, instead of (deterministically) optimizing V_t with respect to $\mathbf{p}(t)$ in $\mathcal{C}(\mathbf{p}^o(t-1))$, which is, of course, impossible, one considers optimizing *a projection of* $V(\mathbf{p},t)$, $\mathbf{p} \in \mathcal{S}^R$ *onto the space of all measurable functionals of* $\mathscr{C}(\mathcal{T}_{t-1})$, which corresponds to the information observed by the relays, up to $t-1$. Provided that, for every $\mathbf{p} \in \mathcal{S}^R$, $V(\mathbf{p},t)$ is in the Hilbert

space of square-integrable, real-valued functions relative to $\mathcal{P}$, $\mathcal{L}_2(\Omega, \mathscr{F}, \mathcal{P}; \mathbb{R})$, it is then reasonable to consider orthogonal projections, that is, the Minimum Mean Square Error (MMSE) estimate, or, more accurately, prediction of $V(\mathbf{p}, t)$ given $\mathscr{C}(\mathcal{T}_{t-1})$. This, of course, coincides with the conditional expectation $\mathbb{E}\{V(\mathbf{p}, t) | \mathscr{C}(\mathcal{T}_{t-1})\}$. One then optimizes the *random utility* $\mathbb{E}\{V(\mathbf{p}, t) | \mathscr{C}(\mathcal{T}_{t-1})\}$, with respect to $\mathbf{p}$ in the random set $\mathcal{C}(\mathbf{p}^o(t-1))$, as in (4.6).

Although there is nothing technically wrong with actually starting with (4.6) as our *initial* problem formulation, and essentially bypassing the technical difficulties of (4.5), the fact that the objective of (4.6) depends on $\mathscr{C}(\mathcal{T}_{t-1})$ does not render it a useful optimality criterion. This is because the objective of (4.6) quantifies the performance of a *single* decision, *only conditioned* on $\mathscr{C}(\mathcal{T}_{t-1})$, *despite* the fact that an optimal solution to (4.6) (provided it exists) constitutes itself a functional of $\mathscr{C}(\mathcal{T}_{t-1})$. In other words, the objective of (4.6) *does not* quantify the performance of *a policy (a decision rule)*; in order to do that, any reasonable performance criterion should assign *a number* to each policy, ranking its quality, and *not* a function depending on $\mathscr{C}(\mathcal{T}_{t-1})$. The expected utility $\mathbb{E}\{V_t\}$ of the variational problem (4.5) constitutes a suitable such criterion. And by the Fundamental Lemma, (4.5) may be indeed reduced to (4.6), which can thus be regarded as a proxy for solving the former.

There are two main reasons justifying our interest in policies, rather than individual decisions. First, one should be interested in the *long-term behavior* of the beamforming (in our case) system, in the sense that it should be possible to assess system performance if the system is used repeatedly over time, e.g., periodically (every hour, day) or on demand. For example, consider a beamforming system (the "experiment"), which operates for N_T time slots and dependently *restarts* its operation at time slots $kN_T + 1$, for k in some subset of $\mathbb{N}^+$. This might be practically essential for maintaining system stability over time, saving on resources, etc. It is then clear that merely quantifying the performance of individual decisions is meaningless, from an operational point of view; simply, the random utility approach quantifies performance *only along a specific path of the observed information*, $\mathscr{C}(\mathcal{T}_{t-1})$, for $t \in \mathbb{N}^+_{N_T}$. This issue is more profound when channel observations taking specific values correspond to events of zero measure (this

is actually the case with the Gaussian channel model introduced in Section **??**). On the contrary, it is of interest to jointly quantify system performance when decisions are made for different outcomes of the sample space Ω. This immediately results in the need for quantifying the performance of different policies (decision rules), and this is only possible by considering variational optimization problems, such as (4.5).

Additionally, because decisions are made *in stages*, it is of great interest to consider how the system performs *across time slots*, or, in other words, to discover *temporal trends* in performance, if such trends exist. In particular, for the beamforming problem considered in this work, we will be able to theoretically characterize system behavior under both suboptimal and optimal decision making, in the average (expected) sense (see Section 4.2.4), *across all time slots*; this is impossible to do for each possible outcome of the sample space, individually, when the random utility approach is considered.

The second main reason for considering the variational program (4.5) as our main objective, instead of (4.6), is practical, and extremely important from an engineering point of view. The expected utility approach assigns, at each time slot, a number to each policy, quantifying its quality. Simulating repeatedly the system and invoking the Law of Large Numbers, one may obtain excellent estimates of the expected performance of the system, quantified by the chosen utility. Therefore, the systematic experimental assessment of a particular sequence of policies (one for each time slot) is readily possible. Apparently, such experimental validation approach is impossible to perform by adopting the random (conditional) utility approach, since the performance of the system will be quantified via a real valued (in general) random quantity. ■

Remark 4.2. The stochastic programming methodology presented in this subsection is very general and can support lots of choices in regard to the structure of the second-stage subproblem, V_t. As shown in the discussion developed in Appendix A, the key to showing the validity of the Fundamental Lemma is the set of conditions **C1-C6**. If these are satisfied, it is then possible to convert the original, variational problem into a pointwise one, while strictly preserving optimality. ■

4.2.3 SINR Maximization at the Destination

The basic and fundamentally important beamforming criterion considered in this work is that of enhancing network QoS, or, in other words, maximizing the respective SINR at the destination, subject to a total power budget at the relays. At each time slot $t \in \mathbb{N}^{+}_{N_T}$, *given* CSI encoded in $\mathscr{C}(\mathcal{T}_t)$ and with $\boldsymbol{w}(t) \triangleq [w_1(t) \ldots w_R(t)]^{\boldsymbol{T}}$, this may be achieved as in Section 1.2.1, by formulating the constrained optimization problem [2, 9]

$$\begin{aligned} \underset{\boldsymbol{w}(t)}{\text{maximize}} \quad & \frac{\mathbb{E}\{P_S(t)|\mathscr{C}(\mathcal{T}_t)\}}{\mathbb{E}\{P_{I+N}(t)|\mathscr{C}(\mathcal{T}_t)\}} \\ \text{subject to} \quad & \mathbb{E}\{P_R(t)|\mathscr{C}(\mathcal{T}_t)\} \leq P_c \end{aligned}, \tag{4.8}$$

where $P_c > 0$ denotes the total available relay transmission power. It was shown earlier in Section 1.2.1, that the optimal value of (4.8) may be expressed *analytically* as [2]

$$\begin{aligned} V_t &\equiv \sum_{i \in \mathbb{N}^{+}_{R}} \frac{P_c P_0 \left|f\left(\mathbf{p}_i(t), t\right)\right|^2 \left|g\left(\mathbf{p}_i(t), t\right)\right|^2}{P_0 \sigma_D^2 \left|f\left(\mathbf{p}_i(t), t\right)\right|^2 + P_c \sigma^2 \left|g\left(\mathbf{p}_i(t), t\right)\right|^2 + \sigma^2 \sigma_D^2} \\ &\triangleq \sum_{i \in \mathbb{N}^{+}_{R}} V_I\left(\mathbf{p}_i(t), t\right), \quad \forall t \in \mathbb{N}^{+}_{N_T}. \end{aligned} \tag{4.9}$$

Adopting the 2-stage stochastic optimization framework presented and discussed in Section 4.2.2, we are now interested, at each time slot $t-1 \in \mathbb{N}^{+}_{N_T-1}$, in the program

$$\begin{aligned} \underset{\mathbf{p}(t)}{\text{maximize}} \quad & \mathbb{E}\left\{\sum_{i \in \mathbb{N}^{+}_{R}} V_I\left(\mathbf{p}_i(t), t\right)\right\} \\ \text{subject to} \quad & \mathbf{p}(t) \equiv \mathcal{M}\left(\mathscr{C}(\mathcal{T}_{t-1})\right) \in \mathcal{C}\left(\mathbf{p}^o(t-1)\right), \\ & \mathcal{M}^{-1}(\mathcal{A}) \in \mathscr{B}\left(\mathbb{R}^{2R}\right), \forall \mathcal{A} \in \mathscr{B}\left(\mathbb{R}^{2R(t-1)}\right) \end{aligned}, \tag{4.10}$$

where $\mathbf{p}^o(1) \in \mathcal{S}^R$ is a known constant, representing the initial positions of the relays. But in order to be able to formulate (4.10) in a well defined manner fully and and simplify it by exploitting the Fundamental Lemma, we have to explicitly verify Conditions **C1-C6** of Lemma 4.3 in Section 4.5.1.4 of Appendix A. To this end, let us present a definition.

Definition 4.1. (Translated Multifunctions) Given $\mathcal{H} \subset \mathbb{R}^N$, $\mathcal{A} \subseteq \mathbb{R}^N$ and any

fixed $\boldsymbol{h} \in \mathcal{H}$, $\mathcal{D} : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is called the $(\mathcal{H}, \boldsymbol{h})$-translated multifunction in $\mathcal{A}$, if and only if $\mathcal{D}(\boldsymbol{y}) \triangleq \{\boldsymbol{x} \in \mathcal{A} | \, \boldsymbol{x} - \boldsymbol{y} \in \mathcal{H}\}$, for all $\boldsymbol{y} \in \mathcal{A} - \boldsymbol{h} \triangleq \left\{ \boldsymbol{x} \in \mathbb{R}^N \middle| \, \boldsymbol{x} + \boldsymbol{h} \in \mathcal{A} \right\}$.

Note that translated multifunctions, in the sense of Definition 4.1, are always unique and non empty, whenever $\boldsymbol{y} \in \mathcal{A} - \boldsymbol{h}$. We also observe that, if $\boldsymbol{y} \notin \mathcal{A} - \boldsymbol{h}$, $\mathcal{D}(\boldsymbol{y})$ is undefined; in fact, outside $\mathcal{A} - \boldsymbol{h}$, $\mathcal{D}$ may be defined arbitrarily, and this will be irrelevant in our analysis. The following assumption on the structure of the compact-valued multifunction $\mathcal{C} : \mathbb{R}^{2R} \rightrightarrows \mathbb{R}^{2R}$ is adopted hereafter, and for the rest of this work.

Assumption 4.1. ($\mathcal{C}$ is Translated) *Given any arbitrary compact set $\mathbf{0} \in \mathcal{G} \subset \mathcal{S}^R$, $\mathcal{C}$ constitutes the corresponding $(\mathcal{G}, \mathbf{0})$-translated, compact-valued multifunction in $\mathcal{S}^R$.*

Then, the following important result is true.

Theorem 4.1. (Verification Theorem / SINR Maximization) *Suppose that, at time slot $t - 1 \in \mathbb{N}^+_{N_T - 1}$, the selected decision at $t - 2$, $\mathbf{p}^o(t-1) \equiv \mathbf{p}^o(\omega, t-1)$, is measurable relative to $\mathscr{C}(\mathcal{T}_{t-2})$. Then, the stochastic program (4.10) satisfies conditions* **C1-C6** *and the Fundamental Lemma applies (see Appendix A, Section 4.5.1.4, Lemma 4.3). Additionally, as long as the pointwise program*

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \sum_{i \in \mathbb{N}^+_R} \mathbb{E}\left\{ V_I(\mathbf{p}_i, t) | \, \mathscr{C}(\mathcal{T}_{t-1}) \right\} \\ \text{subject to} \quad & \mathbf{p} \in \mathcal{C}(\mathbf{p}^o(t-1)) \end{aligned} \tag{4.11}$$

has a unique maximizer $\mathbf{p}^*(t)$, and $\mathbf{p}^o(t) \equiv \mathbf{p}^*(t)$, then $\mathbf{p}^o(t)$ is $\mathscr{C}(\mathcal{T}_{t-1})$-measurable and the condition of the theorem is automatically satisfied at time slot t.

Proof of Theorem 4.1. See Appendix B. ■

As Theorem 4.1 suggests, in order for conditions **C1-C6** to be simultaneously satisfied for all $t \in \mathbb{N}^2_{N_T}$, it is sufficient that the program (4.11) has a *unique* optimal solution, for each t. Although, it general, such requirement might not be particularly appealing, for the problems of interest in this work, the event where (4.11) does not

have a unique optimizer is extremely rare, almost never occurring in practice. Nevertheless, uniqueness of the optimal solution to (4.11) does not constitute a necessary condition for $\mathscr{C}(\mathcal{T}_{t-1})$-measurability of the optimal decision at time slot $t-1$. For instance, $\mathbf{p}^*(t)$ will always be $\mathscr{C}(\mathcal{T}_{t-1})$-measurable when the compact-valued, closed multifunction $\mathcal{C}:\mathbb{R}^{2R}\rightrightarrows\mathbb{R}^{2R}$ is additionally *finite-valued*, and $\mathbf{p}^o(t)\equiv\mathbf{p}^*(t)$. This choice for $\mathcal{C}$ is particularly useful for practical implementations. In any case, as long as conditions **C1-C6** are guaranteed to be satisfied, we may focus exclusively on the pointwise program (4.11), whose expected optimal value, via the Fundamental Lemma, coincides with the optimal value of the original problem (4.10).

By definition, we readily observe that the problem (4.11) is separable. In fact, given that, *for each $t\in\mathbb{N}^+_{N_T-1}$, decisions taken and CSI collected so far are available to all relays*, (4.11) can be solved in a *completely distributed fashion at the relays*, with the i-th relay being responsible for solving the program

$$\begin{aligned}\underset{\mathbf{p}}{\text{maximize}}\quad & \mathbb{E}\left\{V_I(\mathbf{p},t)\middle|\mathscr{C}(\mathcal{T}_{t-1})\right\}\\ \text{subject to}\quad & \mathbf{p}\in\mathcal{C}_i(\mathbf{p}^o(t-1))\end{aligned}, \tag{4.12}$$

at each $t-1\in\mathbb{N}^+_{N_T-1}$, where $\mathcal{C}_i:\mathbb{R}^2\rightrightarrows\mathbb{R}^2$ denotes the corresponding part of $\mathcal{C}$, for each $i\in\mathbb{N}^+_R$. Note that no local exchange of intermediate results is required among relays; given the available information, each relay independently solves its own subproblem. It is also evident that apart from the obvious difference in the feasible set, the optimization problems at each of the relays are identical. The problem, however, with (4.12) is that its objective involves the evaluation of a conditional expectation of a well defined ratio of almost surely positive random variables, which is *impossible to perform analytically.* For this reason, it is imperative to resort to the development of well behaved approximations to (4.12), which, at the same time, would facilitate implementation. In the following, we present two such heuristic approaches.

4.2.3.1 Approximation by the Method of Statistical Differentials

The first idea we are going to explore is that of approximating the objective of (4.12) by truncated Taylor expansions. Observe that V_I can be equivalently expressed as

$$\begin{aligned} V_I(\mathbf{p},t) &\equiv \frac{1}{\frac{\sigma_D^2}{P_c}|g(\mathbf{p},t)|^{-2}+\frac{\sigma^2}{P_0}|f(\mathbf{p},t)|^{-2}+\frac{\sigma^2\sigma_D^2}{P_cP_0}|f(\mathbf{p},t)|^{-2}|g(\mathbf{p},t)|^{-2}} \\ &\triangleq \frac{1}{V_{II}(\mathbf{p},t)}, \end{aligned} \tag{4.13}$$

for all $(\mathbf{p},t) \in \mathcal{S}\times\mathbb{N}_{N_T}^+$. Then, for $t \in \mathbb{N}_{N_T}^2$, we may locally approximate $\mathbb{E}\{V_I(\mathbf{p},t)|\mathscr{C}(\mathcal{T}_{t-1})\}$ around the point $\mathbb{E}\{V_{II}(\mathbf{p},t)|\mathscr{C}(\mathcal{T}_{t-1})\}$ (see Section 3.14.2 in [80]; also known as the *Method of Statistical Differentials*) via a first order Taylor expansion as

$$\mathbb{E}\{V_I(\mathbf{p},t)|\mathscr{C}(\mathcal{T}_{t-1})\} \approx \frac{1}{\mathbb{E}\{V_{II}(\mathbf{p},t)|\mathscr{C}(\mathcal{T}_{t-1})\}}, \tag{4.14}$$

or via a second order Taylor expansion as

$$\mathbb{E}\{V_I(\mathbf{p},t)|\mathscr{C}(\mathcal{T}_{t-1})\} \approx \frac{\mathbb{E}\left\{(V_{II}(\mathbf{p},t))^2\middle|\mathscr{C}(\mathcal{T}_{t-1})\right\}}{\left(\mathbb{E}\{V_{II}(\mathbf{p},t)|\mathscr{C}(\mathcal{T}_{t-1})\}\right)^3}, \tag{4.15}$$

where it is straightforward to show that the square on the numerator can be expanded as

$$\begin{aligned} (V_{II}(\mathbf{p},t))^2 \equiv{}& \left(\frac{\sigma^2\sigma_D^2}{P_cP_0}\right)^2|f(\mathbf{p},t)|^{-4}|g(\mathbf{p},t)|^{-4}+2\frac{\sigma^2\sigma_D^2}{P_cP_0}|f(\mathbf{p},t)|^{-2}|g(\mathbf{p},t)|^{-2} \\ &+2\left(\frac{\sigma^2}{P_0}\right)^2\frac{\sigma_D^2}{P_c}|f(\mathbf{p},t)|^{-4}|g(\mathbf{p},t)|^{-2}+2\frac{\sigma^2}{P_0}\left(\frac{\sigma_D^2}{P_c}\right)^2|f(\mathbf{p},t)|^{-2}|g(\mathbf{p},t)|^{-4} \\ &+\left(\frac{\sigma^2}{P_0}\right)^2|f(\mathbf{p},t)|^{-4}+\left(\frac{\sigma_D^2}{P_c}\right)^2|g(\mathbf{p},t)|^{-4}. \end{aligned} \tag{4.16}$$

The approximate formula (4.15) may be in fact computed in closed form at any point $\mathbf{p}\in\mathcal{S}$, thanks to the following technical, but simple, result.

Lemma 4.1. (Big Expectations) *Under the wireless channel model introduced in Section* **??**, *it is true that, at any* $\mathbf{p}\in\mathcal{S}$,

$$[F(\mathbf{p},t)\ G(\mathbf{p},t)]^{\boldsymbol{T}}\Big|\mathscr{C}(\mathcal{T}_{t-1}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p}),\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})\right), \tag{4.17}$$

for all $t \in \mathbb{N}_{N_T}^2$, *and where we define*

$$\boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p}) \triangleq \left[\mu_{t|t-1}^{F}(\mathbf{p})\ \mu_{t|t-1}^{G}(\mathbf{p})\right]^{\boldsymbol{T}}, \tag{4.18}$$

$$\mu_{t|t-1}^{F}(\mathbf{p}) \triangleq \alpha_S(\mathbf{p})\,\ell + \boldsymbol{c}_{1:t-1}^{F}(\mathbf{p})\,\boldsymbol{\Sigma}_{1:t-1}^{-1}\left(\boldsymbol{m}_{1:t-1}-\boldsymbol{\mu}_{1:t-1}\right) \in \mathbb{R}, \tag{4.19}$$

$$\mu_{t|t-1}^{G}(\mathbf{p}) \triangleq \alpha_D(\mathbf{p})\,\ell + \boldsymbol{c}_{1:t-1}^{G}(\mathbf{p})\,\boldsymbol{\Sigma}_{1:t-1}^{-1}\left(\boldsymbol{m}_{1:t-1}-\boldsymbol{\mu}_{1:t-1}\right) \in \mathbb{R} \quad \textit{and} \tag{4.20}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p}) \triangleq &\begin{bmatrix} \eta^2+\sigma_\xi^2 & \eta^2 e^{-\frac{\|\mathbf{p}_S-\mathbf{p}_D\|_2}{\delta}} \\ \eta^2 e^{-\frac{\|\mathbf{p}_S-\mathbf{p}_D\|_2}{\delta}} & \eta^2+\sigma_\xi^2 \end{bmatrix} \\ &- \begin{bmatrix} \boldsymbol{c}_{1:t-1}^{F}(\mathbf{p}) \\ \boldsymbol{c}_{1:t-1}^{G}(\mathbf{p}) \end{bmatrix} \boldsymbol{\Sigma}_{1:t-1}^{-1} \begin{bmatrix} \boldsymbol{c}_{1:t-1}^{F}(\mathbf{p}) \\ \boldsymbol{c}_{1:t-1}^{G}(\mathbf{p}) \end{bmatrix}^{\boldsymbol{T}} \in \mathbb{S}_{++}^2,\end{aligned} \tag{4.21}$$

with

$$\boldsymbol{m}_{1:t-1} \triangleq \left[\boldsymbol{F}^{\boldsymbol{T}}(1)\ \boldsymbol{G}^{\boldsymbol{T}}(1)\ \ldots\ \boldsymbol{F}^{\boldsymbol{T}}(t-1)\ \boldsymbol{G}^{\boldsymbol{T}}(t-1)\right]^{\boldsymbol{T}} \in \mathbb{R}^{2R(t-1)\times 1}, \tag{4.22}$$

$$\boldsymbol{\mu}_{1:t-1} \triangleq [\boldsymbol{\alpha}_S(\mathbf{p}(1))\ \boldsymbol{\alpha}_D(\mathbf{p}(1))\ \ldots\ \boldsymbol{\alpha}_S(\mathbf{p}(t-1))\ \boldsymbol{\alpha}_D(\mathbf{p}(t-1))]^{\boldsymbol{T}}\ell \in \mathbb{R}^{2R(t-1)\times 1}, \tag{4.23}$$

$$\boldsymbol{c}_{1:t-1}^{F}(\mathbf{p}) \triangleq \left[\boldsymbol{c}_1^{F}(\mathbf{p})\ \ldots\ \boldsymbol{c}_{t-1}^{F}(\mathbf{p})\right] \in \mathbb{R}^{1\times 2R(t-1)}, \tag{4.24}$$

$$\boldsymbol{c}_{1:t-1}^{G}(\mathbf{p}) \triangleq \left[\boldsymbol{c}_1^{G}(\mathbf{p})\ \ldots\ \boldsymbol{c}_{t-1}^{G}(\mathbf{p})\right] \in \mathbb{R}^{1\times 2R(t-1)}, \tag{4.25}$$

$$\boldsymbol{c}_k^{F}(\mathbf{p}) \triangleq \left[\left\{\mathbb{E}\left\{\sigma_S(\mathbf{p},t)\,\sigma_S^j(k)\right\}\right\}_{j\in\mathbb{N}_R^+}\ \left\{\mathbb{E}\left\{\sigma_S(\mathbf{p},t)\,\sigma_D^j(k)\right\}\right\}_{j\in\mathbb{N}_R^+}\right], \forall k \in \mathbb{N}_{t-1}^+ \tag{4.26}$$

$$\boldsymbol{c}_k^{G}(\mathbf{p}) \triangleq \left[\left\{\mathbb{E}\left\{\sigma_D(\mathbf{p},t)\,\sigma_S^j(k)\right\}\right\}_{j\in\mathbb{N}_R^+}\ \left\{\mathbb{E}\left\{\sigma_D(\mathbf{p},t)\,\sigma_D^j(k)\right\}\right\}_{j\in\mathbb{N}_R^+}\right], \forall k \in \mathbb{N}_{t-1}^+ \textit{ and} \tag{4.27}$$

$$\boldsymbol{\Sigma}_{1:t-1} \triangleq \begin{bmatrix} \boldsymbol{\Sigma}(1,1) & \cdots & \boldsymbol{\Sigma}(1,t-1) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}(t-1,1) & \cdots & \boldsymbol{\Sigma}(t-1,t-1) \end{bmatrix} \in \mathbb{S}_{++}^{2R(t-1)}, \tag{4.28}$$

for all $(\mathbf{p},t) \in \mathcal{S}\times\mathbb{N}_{N_T}^2$. *Further, for any choice of* $(m,n)\in\mathbb{Z}\times\mathbb{Z}$, *the conditional correlation of the fields* $|f(\mathbf{p},t)|^m$ *and* $|g(\mathbf{p},t)|^n$ *relative to* $\mathscr{C}(\mathcal{T}_{t-1})$ *may be expressed in closed form as*

$$\begin{aligned}&\mathbb{E}\left\{|f(\mathbf{p},t)|^m\,|g(\mathbf{p},t)|^n\,|\,\mathscr{C}(\mathcal{T}_{t-1})\right\} \\ &\equiv 10^{(m+n)\rho/20}\exp\left(\frac{\log(10)}{20}\begin{bmatrix} m \\ n \end{bmatrix}^{\boldsymbol{T}} \boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p}) + \left(\frac{\log(10)}{20}\right)^2 \begin{bmatrix} m \\ n \end{bmatrix}^{\boldsymbol{T}} \boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p}) \begin{bmatrix} m \\ n \end{bmatrix}\right),\end{aligned} \tag{4.29}$$

at any $\mathbf{p}\in\mathcal{S}$ *and for all* $t\in\mathbb{N}_{N_T}^2$.

Proof of Lemma 4.1. See Appendix B. ■

Since, by exploitting Lemma 4.1 and (4.16), formula (4.15) can be evaluated without any particular difficulty, we now propose the replacement of the original pointwise problem of interest, (4.12), with either of the heuristics

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \frac{1}{\mathbb{E}\left\{V_{II}\left(\mathbf{p},t\right)\middle|\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}} \\ \text{subject to} \quad & \mathbf{p}\in\mathcal{C}_i\left(\mathbf{p}^o\left(t-1\right)\right) \end{aligned} \tag{4.30}$$

and

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \frac{\mathbb{E}\left\{\left(V_{II}\left(\mathbf{p},t\right)\right)^2\middle|\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}}{\left(\mathbb{E}\left\{V_{II}\left(\mathbf{p},t\right)\middle|\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}\right)^3} \\ \text{subject to} \quad & \mathbf{p}\in\mathcal{C}_i\left(\mathbf{p}^o\left(t-1\right)\right) \end{aligned}, \tag{4.31}$$

to be solved at relay $i\in\mathbb{N}_R^+$, at each time $t-1\in\mathbb{N}_{N_T-1}^+$, depending on the order of approximation employed, respectively. Observe that Jensen's Inequality directly implies that the objective of (4.30) is always *lower than or equal* that of (4.31) and that of the original program (4.12), as well. As a result, (4.30) is also a *lower bound relaxation* to (4.12). Both approximations are technically well behaved, though, as made precise by the next theorem.

Theorem 4.2. (Behavior of Approximation Chains I / SINR Maximization) *Both heuristics (4.30) and (4.31) each feature at least one measurable maximizer. Therefore, provided that any of the two heuristics is solved at each time slot* $t-1\in\mathbb{N}_{N_T-1}^+$*, that the selected one features a unique maximizer,* $\widetilde{\mathbf{p}}^*\left(t\right)$*, and that* $\widetilde{\mathbf{p}}^*\left(t\right)\equiv\mathbf{p}^o\left(t\right)$*, for all* $t\in\mathbb{N}_{N_T}^2$*, the produced decision chain is measurable and condition* **C2** *is satisfied at all times.*

Proof of Theorem 4.2. See Appendix B. ■

Theorem 4.2 implies that, at each time slot $t\in\mathbb{N}_{N_T-1}^+$ and under the respective conditions, the chosen heuristic constitutes a well defined approximation to the original problem, (4.12) and, in turn, to (4.10), in the sense that all conditions **C1-C4** are satisfied.

At this point, it will be important to note that, for *each* $\mathbf{p} \in \mathcal{S}$, computation of the conditional mean and covariance in (4.17) of Lemma 4.1 require execution of matrix operations, which are of *expanding dimension in* $t \in \mathbb{N}_{N_T}^2$; observe that, for instance, the covariance matrix $\mathbf{\Sigma}_{1:t-1}$ is of size $2R\,(t-1)$, which is increasing in $t \in \mathbb{N}_{N_T}^2$. Fortunately, however, the increase is linear in t. Additionally, the reader may readily observe that the inversion of the covariance matrix $\mathbf{\Sigma}_{1:t-1}$ constitutes the computationally dominant operation in the long formulas of Lemma 4.1. The computational complexity of this matrix inversion, which takes place at each time slot $t-1 \in \mathbb{N}_{N_T-1}^+$, is, in general, of the order of $\mathcal{O}\left(R^3 t^3\right)$ elementary operations. Fortunately though, we may exploit the Matrix Inversion Lemma, in order to reduce the computational complexity of the aforementioned matrix inversion to the order of $\mathcal{O}\left(R^3 t^2\right)$. Indeed, by construction, $\mathbf{\Sigma}_{1:t-1}$ may be expressed as

$$\mathbf{\Sigma}_{1:t-1} \equiv \begin{bmatrix} \mathbf{\Sigma}_{1:t-2} & \mathbf{\Sigma}_{1:t-2}^{c} \\ \left(\mathbf{\Sigma}_{1:t-2}^{c}\right)^{\boldsymbol{T}} & \mathbf{\Sigma}\,(t-1,t-1) \end{bmatrix}, \tag{4.32}$$

where

$$\mathbf{\Sigma}_{1:t-2}^{c} \triangleq \left[\mathbf{\Sigma}\,(1,t-1)\ \ldots\ \mathbf{\Sigma}\,(t-2,t-1)\right]^{\boldsymbol{T}} \in \mathbb{R}^{2R(t-2)\times 2R}. \tag{4.33}$$

Invoking the Matrix Inversion Lemma, we obtain the *recursive* expression

$$\mathbf{\Sigma}_{1:t-1}^{-1} = \begin{bmatrix} \mathbf{\Sigma}_{1:t-2}^{-1} + \mathbf{\Sigma}_{1:t-2}^{-1}\mathbf{\Sigma}_{1:t-2}^{c}\mathbf{S}_{t-1}^{-1}\left(\mathbf{\Sigma}_{1:t-2}^{c}\right)^{\boldsymbol{T}}\mathbf{\Sigma}_{1:t-2}^{-1} & -\mathbf{\Sigma}_{1:t-2}^{-1}\mathbf{\Sigma}_{1:t-2}^{c}\mathbf{S}_{t-1}^{-1} \\ -\mathbf{S}_{t-1}^{-1}\left(\mathbf{\Sigma}_{1:t-2}^{c}\right)^{\boldsymbol{T}}\mathbf{\Sigma}_{1:t-2}^{-1} & \mathbf{S}_{t-1}^{-1} \end{bmatrix}, \quad \text{with} \tag{4.34}$$

$$\mathbf{S}_{t-1} \triangleq \mathbf{\Sigma}\,(t-1,t-1) - \left(\mathbf{\Sigma}_{1:t-2}^{c}\right)^{\boldsymbol{T}}\mathbf{\Sigma}_{1:t-2}^{-1}\mathbf{\Sigma}_{1:t-2}^{c} \in \mathbb{S}_{++}^{2R}, \tag{4.35}$$

where $\mathbf{S}_{t-1}$ is the respective Schur complement. From (4.34) and (4.35), it can be easily verified that the most computationally demanding operation involved is $\mathbf{\Sigma}_{1:t-2}^{-1}\mathbf{\Sigma}_{1:t-2}^{c}$, of order $\mathcal{O}\left(R^3 t^2\right)$. Since the inversion of $\mathbf{S}_{t-1}$ is of the order of $\mathcal{O}\left(R^3\right)$, we arrive at a total complexity of $\mathcal{O}\left(R^3 t^2\right)$ elementary operations of the recursive scheme presented above, and implemented at each time slot $t-1$. The achieved reduction in complexity is important. In most scenarios, R, the number of relays, will be relatively small and fixed for the whole operation of the system, whereas t, the time slot index, might generally

take large values, since it is common for the operational horizon of the system, N_T, to be large. Additionally, the reader may readily observe that the aforementioned covariance matrix is independent of the position at which the channel is predicted, $\mathbf{p}$. As a result, its inversion may be performed just once in each time slot, for all evaluations of the mean and covariance of the Gaussian density in (4.17), for all different choices of $\mathbf{p}$ on a fixed grid (say). Consequently, if the total number of such evaluations is $P \in \mathbb{N}^+$, and recalling that the complexity for a matrix-vector multiplication is quadratic in the dimension of the quantities involved, then, *at worst*, the total computational complexity for channel prediction is of the order of $\mathcal{O}\left(PR^2t^2 + R^3t^2\right)$, *at each* $t-1 \in \mathbb{N}^+_{N_T-1}$. This means that a potential actual computational system would have to be able to execute matrix operations with complexity at most of the order of $\mathcal{O}\left(PR^2N_T^2 + R^3N_T^2\right)$, which constitutes the worst case complexity, *over all* N_T *time slots*. The analysis above characterizes the complexity for solving either of the heuristics (4.30) and (4.31), if the feasible set $\mathcal{C}_i$ is assumed to be finite, for all $i \in \mathbb{N}^+_R$. Of course, if the quantity RN_T is considered a fixed constant, implying that computation of the mean and covariance in (4.17) is considered the result of a black box with fixed (worst) execution time and with input $\mathbf{p}$, then, at each $t-1 \in \mathbb{N}^+_{N_T-1}$, the total computational complexity for channel prediction is of the order of $\mathcal{O}(P)$ function evaluations, that is, linear in P.

4.2.3.2 Brute Force

The second approach to the solution of (4.12), considered in this section, is based on the fact that the objective of the aforementioned program can be evaluated rather efficiently, relying on the *multidimensional Gauss-Hermite Quadrature Rule* [88], which constitutes a readily available routine for numerical integration. It is particularly effective for computing expectations of complicated functions of Gaussian random variables [89]. This is indeed the case here, as shown below.

Leveraging Lemma 4.1 and as it can also be seen in the proof of Theorem 4.1 (condition **C6**), the objective of (4.12) can be equivalently represented, for all $t \in \mathbb{N}^2_{N_T}$,

via a Lebesgue integral as

$$\mathbb{E}\left\{V_I\left(\mathbf{p},t\right)|\,\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}=\int_{\mathbb{R}^2} r\left(\boldsymbol{x}\right)\mathcal{N}\left(\boldsymbol{x};\boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p}),\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})\right)\mathrm{d}\boldsymbol{x}, \tag{4.36}$$

for any choice of $\mathbf{p}\in\mathcal{S}$, where $\mathcal{N}\left(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma}\right):\mathbb{R}^2\rightarrow\mathbb{R}_{++}$ denotes the bivariate Gaussian density, with mean $\boldsymbol{\mu}\in\mathbb{R}^{2\times1}$ and covariance $\boldsymbol{\Sigma}\in\mathbb{S}_{+}^{2\times2}$, and the function $r:\mathbb{R}^2\rightarrow\mathbb{R}_{++}$ is defined exploitting the trick (3.11) as

$$r\left(\boldsymbol{x}\right)\equiv r\left(x_1,x_2\right)\triangleq\frac{P_cP_0 10^{2\rho/10}\left[\exp\left(x_1+x_2\right)\right]^{\frac{\log(10)}{10}}}{P_0\sigma_D^2\left[\exp\left(x_1\right)\right]^{\frac{\log(10)}{10}}+P_c\sigma^2\left[\exp\left(x_2\right)\right]^{\frac{\log(10)}{10}}+10^{-\rho/10}\sigma^2\sigma_D^2}, \tag{4.37}$$

for all $\boldsymbol{x}\equiv(x_1,x_2)\in\mathbb{R}^2$. Exploitting the Lebesgue integral representation (4.36), it can be easily shown that the conditional expectation may be closely approximated by the double summation formula (see Section IV in [89])

$$\mathbb{E}\left\{V_I\left(\mathbf{p},t\right)|\,\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}\approx\sum_{l_1\in\mathbb{N}_M^+}\varpi_{l_1}\sum_{l_2\in\mathbb{N}_M^+}\varpi_{l_2}r\left(\sqrt{\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})}\boldsymbol{q}_{(l_1,l_2)}+\boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p})\right), \tag{4.38}$$

where $M\in\mathbb{N}^+$ denotes the quadrature resolution, $\boldsymbol{q}_{(l_1,l_2)}\triangleq\left[q_{l_1}\,q_{l_2}\right]^{\boldsymbol{T}}\in\mathbb{R}^{2\times1}$ denotes the (l_1,l_2)-th quadrature point and $\boldsymbol{\varpi}_{(l_1,l_2)}\triangleq\left[\varpi_{l_1}\,\varpi_{l_2}\right]^{\boldsymbol{T}}\in\mathbb{R}^{2\times1}$ denotes respective weighting coefficient, for all $(l_1,l_2)\in\mathbb{N}_M^+\times\mathbb{N}_M^+$. Both sets of quadrature points and weighting coefficients are automatically selected *apriori and independently in each dimension*, via the following simple procedure [89,90]. Let us define a matrix $\boldsymbol{J}\in\mathbb{R}^{M\times M}$, such that

$$\boldsymbol{J}\left(i,j\right)\triangleq\begin{cases}\sqrt{\dfrac{\min\left\{i,j\right\}}{2}}, & |j-i|\equiv1\\ 0, & \text{otherwise}\end{cases},\quad\forall\left(i,j\right)\in\mathbb{N}_M^+\times\mathbb{N}_M^+. \tag{4.39}$$

That is, $\boldsymbol{J}$ constitutes a hollow, tridiagonal, symmetric matrix. Let the sets $\left\{\lambda_i\left(\boldsymbol{J}\right)\in\mathbb{R}\right\}_{i\in\mathbb{N}_M^+}$ and $\left\{\boldsymbol{v}_i\left(\boldsymbol{J}\right)\in\mathbb{R}^{M\times1}\right\}_{i\in\mathbb{N}_M^+}$ contain the eigenvalues and *normalized* eigenvectors of $\boldsymbol{J}$, respectively. Then, simply, quadrature points and the respective weighting coefficients are selected independently in each dimension $j\in\{1,2\}$ as

$$q_{l_j}\equiv\sqrt{2}\lambda_{l_j}\left(\boldsymbol{J}\right)\quad\text{and} \tag{4.40}$$

$$\varpi_{l_j}\equiv\left(\boldsymbol{v}_{l_j}\left(\boldsymbol{J}\right)\left(1\right)\right)^2,\quad\forall l_j\in\mathbb{N}_M^+. \tag{4.41}$$

In (4.41), $\boldsymbol{v}_{l_j}(\boldsymbol{J})(1)$ denotes the first entry of the involved vector.

Under the above considerations, in this subsection, we propose, for a sufficiently large number of quadrature points M, the replacement of the original pointwise problem (4.12) with the heuristic

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \sum_{(l_1,l_2)\in\mathbb{N}_M^+\times\mathbb{N}_M^+} \varpi_{l_1}\varpi_{l_2} r\left(\sqrt{\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})}\boldsymbol{q}_{(l_1,l_2)}+\boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p})\right), \\ \text{subject to} \quad & \mathbf{p}\in\mathcal{C}_i\left(\mathbf{p}^o\left(t-1\right)\right) \end{aligned} \tag{4.42}$$

to be solved at relay $i\in\mathbb{N}_R^+$, at each time $t-1\in\mathbb{N}_{N_T-1}^+$. As in Section 4.2.3.1 above, the following result is in power, concerning the technical consistency of the decision chain produced by considering the approximate program (4.42), for all $t\in\mathbb{N}_{N_T}^2$. Proof is omitted, as it is essentially identical to that of Theorem 4.2.

Theorem 4.3. (Behavior of Approximation Chains II / SINR Maximization) *Consider the the heuristic (4.42). Then, under the same circumstances, all conclusions of Theorem 4.2 hold true.*

Since the computations in (4.40) and (4.41) do not depend on $\mathbf{p}$ or the information collected so far, encoded in $\mathscr{C}\left(\mathcal{T}_{t-1}\right)$, for $t\in\mathbb{N}_{N_T}^2$, quadrature points and the respective weights can be determined offline and stored in memory. Therefore, the computational burden of (4.38) concentrates solely on the computation of an inner product, whose computational complexity is of the order of $\mathcal{O}\left(M^2\right)$, as well as a total of M^2 evaluations of $r\left(\sqrt{\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})}\boldsymbol{q}_{(l_1,l_2)}+\boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p})\right)$, for *each* value of $\mathbf{p}$. Excluding temporarily the computational burden of $\boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p})$ and $\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})$, each of the latter evaluations is of fixed complexity, since each involves elementary operations among matrices and vectors in $\mathbb{R}^{2\times 2}$ and $\mathbb{R}^{2\times 1}$, respectively and, additionally, the involved matrix square root can be evaluated in closed form, via the formula [91]

$$\sqrt{\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})}\equiv\frac{\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})+\sqrt{\det\left(\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})\right)}\mathbf{I}_2}{\sqrt{\operatorname{tr}\left(\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})\right)+2\sqrt{\det\left(\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})\right)}}}\in\mathbb{S}_+^{2\times 2}, \tag{4.43}$$

where we have taken into account that $\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})$ is always a (conditional) covariance matrix and, thus, (conditionally) positive semidefinite. As a result and considering

the last paragraph of Section 4.2.3.1, if (4.38) is evaluated on a finite grid of possible locations, say $P \in \mathbb{N}^{+}$, then, at each $t-1 \in \mathbb{N}_{N_T-1}^{+}$, the total computational complexity of the Gauss-Hermite Quadrature Rule outlined above is of the order of $\mathcal{O}\left(PM^2 + PR^2t^2 + R^3t^2\right)$ elementary operations / function evaluations. This will be the total, worst case computational complexity for solving (4.42), if the feasible set $\mathcal{C}_i$ is assumed to be finite, for all $i \in \mathbb{N}_R^{+}$. As noted above, a finite feasible set greatly simplifies implementation, since a trial-and-error approach may be employed for solving the respective optimization problem. If M is considered a fixed constant (e.g., $M \equiv 10^3$), and the same holds for $Rt \leq RN_T$, then, in each time slot, the total complexity of the Gauss-Hermite Quadrature Rule is of the order of $\mathcal{O}(P)$ evaluations of (4.38), that is, linear in P. In that case, the whole numerical integration routine is considered a black box of fixed computational load, which, in each time slot, takes $\mathbf{p}$ as its input. Observe that, whenever $M \approx RN_T$, the worst case complexity of the brute force method, described in this subsection, *over all N_T time slots*, is essentially the same as that of the Taylor approximation method, presented earlier in Section 4.2.3.1.

4.2.4 Theoretical Guarantees: Network QoS Increases Across Time Slots

The proposed relay position selection approach presented in Section 4.2.3 enjoys a very important and useful feature, initially observed via numerical simulations: Although a 2-stage stochastic programming procedure is utilized *independently* at each time slot for determining optimal relay positioning and beamforming weights at the next time slot, *the average network QoS (that is, the achieved SINR) actually increases, as a function of time (the time slot).* Then, it was somewhat surprising to discover that, additionally, this behavior of the achieved SINR can be predicted *theoretically*, in an indeed elegant manner and, as it will be clear below, under mild and reasonable assumptions on the structure of the spatially controlled beamforming problem under consideration. But first, it would be necessary to introduce the following definition.

Definition 4.2. (L.MD.G Fields) On $(\Omega, \mathscr{F}, \mathcal{P})$, an integrable stochastic field Ξ :

$\Omega \times \mathbb{R}^N \times \mathbb{N} \to \mathbb{R}$ is said to be a *Linear Martingale Difference (MD) Generator, relative to a filtration* $\{\mathscr{H}_t \subseteq \mathscr{F}\}_{t\in\mathbb{N}}$, and *with scaling factor* $\mu \in \mathbb{R}$, or, equivalently, $\mathbf{L.MD.G}\diamondsuit(\mathscr{H}_t, \mu)$, if and only if, for each $t \in \mathbb{N}^+$, there exists a measurable set $\Omega_t \subseteq \Omega$, with $\mathcal{P}(\Omega_t) \equiv 1$, such that, for every $\boldsymbol{x} \in \mathbb{R}^N$, it is true that

$$\mathbb{E}\{\Xi(\boldsymbol{x}, t)|\,\mathscr{H}_{t-1}\}(\omega) \equiv \mu\mathbb{E}\{\Xi(\boldsymbol{x}, t-1)|\,\mathscr{H}_{t-1}\}(\omega), \tag{4.44}$$

for all $\omega \in \Omega_t$.

Remark 4.3. A fine detail in the definition of a $\mathbf{L.MD.G}\diamondsuit(\mathscr{H}_t, \mu)$ field is that, for each $t \in \mathbb{N}$, the event Ω_t *does not depend* on the choice of point $\boldsymbol{x} \in \mathbb{R}^N$. Nevertheless, even if the event where (4.44) is satisfied is indeed dependent on the particular $\boldsymbol{x} \in \mathbb{R}^N$, let us denote it as $\Omega_{\boldsymbol{x},t}$, we may leverage the fact that conditional expectations are unique almost everywhere, and *arbitrarily define*

$$\mathbb{E}\{\Xi(\boldsymbol{x}, t)|\,\mathscr{H}_{t-1}\}(\omega) \triangleq \mu\mathbb{E}\{\Xi(\boldsymbol{x}, t-1)|\,\mathscr{H}_{t-1}\}(\omega), \tag{4.45}$$

for all $\omega \in \Omega^c_{\boldsymbol{x},t}$, where $\mathcal{P}\left(\Omega^c_{\boldsymbol{x},t}\right) \equiv 0$. That is, we modify *both, or either of* the random elements $\mathbb{E}\{\Xi(\boldsymbol{x}, t-1)|\,\mathscr{H}_{t-1}\}$ and $\mathbb{E}\{\Xi(\boldsymbol{x}, t)|\,\mathscr{H}_{t-1}\}$, on the null set $\Omega^c_{\boldsymbol{x},t}$, such that (4.44) is satisfied. Then, it may be easily verified that both such modifications result in valid versions of the conditional expectations of $\Xi(\boldsymbol{x}, t-1)$ and $\Xi(\boldsymbol{x}, t)$ relative to $\mathscr{H}_{t-1}$, respectively and satisfy property (4.44), *everywhere with respect to* $\omega \in \Omega$.

In Definition 4.2, invariance of Ω_t with respect to $\boldsymbol{x} \in \mathbb{R}^N$, in conjunction with the power of the substitution rule for conditional expectations (Section 4.5.1.1), will allow the development of strong conditional arguments, *when* $\boldsymbol{x}$ *is replaced by a random element, measurable relative to* $\mathscr{H}_{t-1}$. ■

Remark 4.4. There are lots of examples of **L.MD.G** stochastic fields, satisfying the technical properties of Definition 4.2. For completeness, let us present two such examples. Employing generic notation, consider an integrable real-valued stochastic field $Y(\boldsymbol{x}, t)$, $\boldsymbol{x} \in \mathbb{R}^N$, $t \in \mathbb{N}$. Let the natural filtration associated with $Y(\boldsymbol{x}, t)$ be $\{\mathscr{Y}_t\}_{t\in\mathbb{N}}$, with $\mathscr{Y}_t \triangleq \sigma\left\{Y(\boldsymbol{x}, t), \boldsymbol{x} \in \mathbb{R}^N\right\}$, for all $t \in \mathbb{N}$. Also, consider another, for simplicity temporal, integrable real-valued process $W(t), t \in \mathbb{N}$. Suppose, further, that $Y(\boldsymbol{x}, t)$ is a martingale with respect to $t \in \mathbb{N}$ (relative to $\{\mathscr{Y}_t\}_{t\in\mathbb{N}}$), and that $W(t)$ is a zero mean

process, independent of $Y(\boldsymbol{x},t)$. In particular, we assume that, for every $t\in\mathbb{N}^{+}$, there exist events $\Omega_t^Y\subseteq\Omega$ and $\Omega_t^W\subseteq\Omega$, satisfying $\mathcal{P}\left(\Omega_t^Y\right)\equiv 1$ and $\mathcal{P}\left(\Omega_t^W\right)\equiv 1$, such that, for all $\boldsymbol{x}\in\mathbb{R}^N$,

$$\mathbb{E}\left\{Y(\boldsymbol{x},t)\middle|\,\mathscr{Y}_{t-1}\right\}(\omega)\equiv Y(\omega,\boldsymbol{x},t-1)\quad\text{and}\tag{4.46}$$

$$\mathbb{E}\left\{W(t)\middle|\,\mathscr{Y}_{t-1}\right\}(\omega)\equiv 0,\tag{4.47}$$

for all $\omega\in\Omega_t^Y\bigcap\Omega_t^W$, where, apparently, $\mathcal{P}\left(\Omega_t^Y\bigcap\Omega_t^W\right)\equiv 1$.

Our first, probably most basic example of a **L.MD.G** field is simply the martingale $Y(\boldsymbol{x},t)$ itself. Of course, in order to verify this statement, we need to show that it satisfies the technical requirements of Definition 4.2, relative to a given filtration; in particular, let us choose $\{\mathscr{Y}_t\}_{t\in\mathbb{N}}$ to be that filtration. Then, for every $(\boldsymbol{x},t)\in\mathbb{R}^N\times\mathbb{N}^{+}$, it is trivial to see that

$$\mathbb{E}\left\{Y(\boldsymbol{x},t)\middle|\,\mathscr{Y}_{t-1}\right\}(\omega)\equiv Y(\omega,\boldsymbol{x},t-1)\equiv\mathbb{E}\left\{Y(\boldsymbol{x},t-1)\middle|\,\mathscr{Y}_{t-1}\right\}(\omega),\tag{4.48}$$

for all $\omega\in\Omega_t^Y$, where $Y(\boldsymbol{x},t-1)$ *is chosen as our version of* $\mathbb{E}\left\{Y(\boldsymbol{x},t-1)\middle|\,\mathscr{Y}_{t-1}\right\}$, *everywhere* in Ω. As a result, the martingale $Y(\boldsymbol{x},t)$ is itself a $\mathbf{L.MD.G}\diamondsuit(\mathscr{Y}_t,1)$, as expected.

The second, somewhat more interesting example of a **L.MD.G** field is defined as

$$X(\boldsymbol{x},t)\triangleq\varrho Y(\boldsymbol{x},t)+W(t),\tag{4.49}$$

for all $(\boldsymbol{x},t)\in\mathbb{R}^N\times\mathbb{N}$, where, say, $0<\varrho\leq 1$. In order to verify the technical requirements of Definition 4.2, let us again choose $\{\mathscr{Y}_t\}_{t\in\mathbb{N}}$ as our filtration. Then, for every $(\boldsymbol{x},t)\in\mathbb{R}^N\times\mathbb{N}^{+}$, there exists a measurable set $\Omega_{\boldsymbol{x},t}^{Y,W}\subseteq\Omega$, with $\mathcal{P}\left(\Omega_{\boldsymbol{x},t}^{Y,W}\right)\equiv 1$, such that, for all $\omega\in\Omega_{\boldsymbol{x},t}^{Y,W}$,

$$\begin{aligned}\mathbb{E}\left\{X(\boldsymbol{x},t)\middle|\,\mathscr{Y}_{t-1}\right\}(\omega)&\equiv\varrho Y(\omega,\boldsymbol{x},t-1)+\mathbb{E}\left\{W(t)\right\}\\&\equiv\varrho Y(\omega,\boldsymbol{x},t-1).\end{aligned}\tag{4.50}$$

Therefore, we may choose our version for $\mathbb{E}\left\{X(\boldsymbol{x},t)\middle|\,\mathscr{Y}_{t-1}\right\}$ as

$$\mathbb{E}\left\{X(\boldsymbol{x},t)\middle|\,\mathscr{Y}_{t-1}\right\}(\omega)\equiv\varrho Y(\omega,\boldsymbol{x},t-1),\quad\forall\omega\in\Omega.\tag{4.51}$$

In exactly the same fashion, we may choose, for every $(\boldsymbol{x}, t) \in \mathbb{R}^N \times \mathbb{N}^+$,

$$\mathbb{E}\left\{ X\left(\boldsymbol{x}, t-1\right) \middle| \mathscr{Y}_{t-1} \right\}(\omega) \equiv \varrho Y\left(\omega, \boldsymbol{x}, t-1\right), \quad \forall \omega \in \Omega. \tag{4.52}$$

Consequently, for every $(\boldsymbol{x}, t) \in \mathbb{R}^N \times \mathbb{N}^+$, it will be true that

$$\mathbb{E}\left\{ X\left(\boldsymbol{x}, t\right) \middle| \mathscr{Y}_{t-1} \right\}(\omega) \equiv \varrho Y\left(\omega, \boldsymbol{x}, t-1\right) \equiv \mathbb{E}\left\{ X\left(\boldsymbol{x}, t-1\right) \middle| \mathscr{Y}_{t-1} \right\}(\omega), \tag{4.53}$$

for all $\omega \in \Omega$, showing that the field $X\left(\boldsymbol{x}, t\right)$ is also $\mathbf{L.MD.G}\diamondsuit\left(\mathscr{Y}_t, 1\right)$. ■

Leveraging the notion of a **L.MD.G** field, the following result may be proven, characterizing the temporal (in discrete time) evolution of the objective of myopic stochastic programs of the form of (4.5). In order to introduce the result, let us define the family $\left\{ \mathscr{P}_t^{\uparrow} \right\}_{t \in \mathbb{N}_{N_T}^+}$, with $\mathscr{P}_t^{\uparrow}$ being the *limit σ-algebra generated by all admissible policies at time slot t,* defined as

$$\mathscr{P}_t^{\uparrow} \triangleq \sigma\left\{ \bigcup_{\mathbf{p}(t) \in \mathcal{D}_t} \sigma\left\{ \mathbf{p}\left(t\right) \right\} \right\} \subseteq \mathscr{C}\left(\mathcal{T}_{t-1}\right), \quad \forall t \in \mathbb{N}_{N_T}^+, \tag{4.54}$$

with $\mathscr{P}_1^{\uparrow}$ being the trivial σ-algebra; recall that $\mathbf{p}\left(1\right) \in \mathcal{S}^R$ is assumed to be a constant. Also, for every $t \in \mathbb{N}_{N_T}^+$, let us define the class

$$\overline{\mathcal{D}}_t \equiv \left\{ \mathbf{p} : \Omega \to \mathcal{S}^R \middle| \mathbf{p}^{-1}\left(\mathcal{A}\right) \in \mathscr{P}_t^{\uparrow}, \text{ for all } \mathcal{A} \in \mathscr{B}\left(\mathcal{S}^R\right) \right\}. \tag{4.55}$$

The result now follows.

Theorem 4.4. (L.MD.G Objectives Increase over Time) *Consider, for each $t \in \mathbb{N}_{N_T}^2$, the maximization version of the 2-stage stochastic program (4.5), for some choice of the second-stage optimal value $V\left(\mathbf{p}, t\right)$, $\mathbf{p} \in \mathcal{S}^R$, $t \in \mathbb{N}_{N_T}^2$. Suppose that conditions* **C1-C6** *are satisfied at all times and let $\mathbf{p}^*\left(t\right)$ denote an optimal solution to (4.10), decided at $t-1 \in \mathbb{N}_{N_T-1}^+$. Suppose, further, that, for every $t \in \mathbb{N}_{N_T}^+$,*

- *$V\left(\mathbf{p}, t\right)$ is $\mathbf{L.MD.G}\diamondsuit\left(\mathscr{H}_t, \mu\right)$, for a filtration $\left\{ \mathscr{H}_t \supseteq \mathscr{P}_t^{\uparrow} \right\}_{t \in \mathbb{N}_{N_T}^+}$ and some $\mu \in \mathbb{R}$, and that*
- *$V\left(\cdot, \cdot, t\right)$ is both $\boldsymbol{SP}\diamondsuit\mathfrak{C}_{\mathscr{H}_t}$ and $\boldsymbol{SP}\diamondsuit\mathfrak{C}^I_{\mathscr{H}_{t-1}}$, with $\overline{\mathcal{D}}_t \subseteq \mathfrak{C}_{\mathscr{H}_t} \subseteq \mathfrak{I}_{\mathscr{H}_t}$ (Remark 4.8 / Section 4.5.1.1).*

Then, for any admissible policy $\mathbf{p}^o(t-1)$*, it is true that*

$$\mu\mathbb{E}\left\{V\left(\mathbf{p}^o(t-1),t-1\right)\right\} \equiv \mathbb{E}\left\{V\left(\mathbf{p}^o(t-1),t\right)\right\} \quad \textit{and} \tag{4.56}$$

$$\mu\mathbb{E}\left\{V\left(\mathbf{p}^*(t-1),t-1\right)\right\} \leq \mathbb{E}\left\{V\left(\mathbf{p}^*(t),t\right)\right\}, \quad \forall t \in \mathbb{N}^2_{N_T}. \tag{4.57}$$

In particular, if $\mu \equiv 1$*, the objective:* • *does not decrease by not updating the decision variable, and* • *is nondecreasing over time, under optimal decision making.*

Proof of Theorem 4.4. See Appendix B. ■

Remark 4.5. When the stochastic program under study is *separable*, that is, when the objective is of the form

$$V(\mathbf{p}(t),t) \equiv \sum_{i\in\mathbb{N}^+_M} V_i(\mathbf{p}_i(t),t) \tag{4.58}$$

(and the respective constraints of the problem decoupled), then, in order to reach the conclusions of Theorem 4.4 for V, it *suffices* for Theorem 4.4 to hold *individually* for each V_i, $i \in \mathbb{N}^+_M$. This is true, for instance, for the spatially controlled beamforming problem (4.10). ■

We may now return to the beamforming problem under consideration, namely (4.10). By Remark 4.5 and Theorem 4.4, it would suffice if we could show that the field $V(\mathbf{p},t)$ is a linear MD generator, relative to a properly chosen filtration. Unfortunately, though, it does not seem to be the case; the statistical structure of $V(\mathbf{p},t)$ does not match that of a linear MD generator *exactly*, relative to any reasonably chosen filtration. Nevertheless, under the channel model of Section **??**, it is indeed possible to show that $V(\mathbf{p},t)$ is *approximately* $\mathbf{L.MD.G}\diamondsuit(\mathscr{C}(\mathcal{T}_{t-1}),1)$, a fact that explains, in an elegant manner, why our proposed spatially controlled beamforming framework is expected to work so well, both under optimal and suboptimal decision making.

To that $V(\mathbf{p},t)$ is *approximately* $\mathbf{L.MD.G}\diamondsuit(\mathscr{C}(\mathcal{T}_{t-1}),1)$, simply consider projecting $V(\mathbf{p},t-1)$ onto $\mathscr{C}(\mathcal{T}_{t-2})$, via the conditional expectation $\mathbb{E}\{V(\mathbf{p},t-1)|\mathscr{C}(\mathcal{T}_{t-2})\}$. Of course, and based on what we have seen so far, $\mathbb{E}\{V(\mathbf{p},t-1)|\mathscr{C}(\mathcal{T}_{t-2})\}$ can be written as a Lebesgue integral of $V(\mathbf{p},t-1)$ expressed in terms of the vector field $[F(\mathbf{p},t-1)\ G(\mathbf{p},t-1)]^{\boldsymbol{T}}$, times its conditional density relative to $\mathscr{C}(\mathcal{T}_{t-2})$. It then

easy to see that this conditional density will be, of course, Gaussian, and will be of exactly the same form as the conditional density of $[F(\mathbf{p},t)\ G(\mathbf{p},t)]^{\boldsymbol{T}}$ relative to $\mathscr{C}(\mathcal{T}_{t-1})$, as presented in Lemma 4.1, but with t replaced by $t-1$. Likewise, $\mathbb{E}\{V(\mathbf{p},t)|\,\mathscr{C}(\mathcal{T}_{t-2})\}$ is of the same form as $\mathbb{E}\{V(\mathbf{p},t-1)|\,\mathscr{C}(\mathcal{T}_{t-2})\}$, but with all terms

$$\exp\left(-\frac{1}{\gamma}\right),\exp\left(-\frac{2}{\gamma}\right),\ldots,\exp\left(-\frac{t-2}{\gamma}\right) \tag{4.59}$$

simply replaced by

$$\exp\left(-\frac{2}{\gamma}\right),\exp\left(-\frac{3}{\gamma}\right),\ldots,\exp\left(-\frac{t-1}{\gamma}\right), \tag{4.60}$$

for all $t\in\mathbb{N}_{N_T}^3$. Of course, if $t\equiv 2$, we have

$$\begin{aligned}\mathbb{E}\{V(\mathbf{p},2)|\,\mathscr{C}(\mathcal{T}_0)\} &\equiv \mathbb{E}\{V(\mathbf{p},2)\}\\ &\equiv \mathbb{E}\{V(\mathbf{p},1)\} \equiv \mathbb{E}\{V(\mathbf{p},1)|\,\mathscr{C}(\mathcal{T}_0)\}.\end{aligned} \tag{4.61}$$

Now, for γ *sufficiently large*, we may approximately write

$$\exp\left(-\frac{x+1}{\gamma}\right)\approx\exp\left(-\frac{x}{\gamma}\right),\quad \forall x>1, \tag{4.62}$$

and, therefore, due to continuity, it should be true that

$$\mathbb{E}\{V(\mathbf{p},t)|\,\mathscr{C}(\mathcal{T}_{t-2})\}\approx\mathbb{E}\{V(\mathbf{p},t-1)|\,\mathscr{C}(\mathcal{T}_{t-2})\}, \tag{4.63}$$

for all $t\in\mathbb{N}_{N_T}^2$ (and everywhere with respect to $\omega\in\Omega$). As a result, we have shown that, *at least approximately*, $V(\mathbf{p},t)$ is $\mathbf{L.MD.G}\diamondsuit(\mathscr{C}(\mathcal{T}_{t-1}),1)$. We may then invoke Theorem 4.4 in an approximate manner, leading to the following important result. Hereafter, for $x\in\mathbb{R}$ and $y\in\mathbb{R}$, $x\lesssim y$ will imply that x *is approximately smaller or equal than* y, in the sense that $x\leq y+\varepsilon$, where $\varepsilon>0$ is some small slack.

Theorem 4.5. (QoS Increases over Time Slots) *Consider the separable stochastic program (4.10). For γ sufficiently large, and for any admissible policy $\mathbf{p}^o(t-1)$, it is true that*

$$\mathbb{E}\{V_I(\mathbf{p}_i^o(t-1),t-1)\}\approx\mathbb{E}\{V_I(\mathbf{p}_i^o(t-1),t)\}, \tag{4.64}$$

$$\mathbb{E}\left\{V_I\big(\mathbf{p}_i^*(t-1),t-1\big)\right\}\lesssim\mathbb{E}\left\{V_I\big(\mathbf{p}_i^*(t),t\big)\right\},\quad \forall i\in\mathbb{N}_R^+ \tag{4.65}$$

$$\mathbb{E}\left\{V(\mathbf{p}^{o}(t-1),t-1)\right\} \approx \mathbb{E}\left\{V(\mathbf{p}^{o}(t-1),t)\right\} \quad \textit{and} \tag{4.66}$$

$$\mathbb{E}\left\{V\big(\mathbf{p}^{*}(t-1),t-1\big)\right\} \lesssim \mathbb{E}\left\{V\big(\mathbf{p}^{*}(t),t\big)\right\}, \tag{4.67}$$

for all $t \in \mathbb{N}_{N_T}^2$. *In other words,* ***approximately****, the average network QoS:* • *does not decrease by not updating the positions of the relays and* • *is nondecreasing across time slots, under (per relay) optimal decision making.*

Theorem 4.5 is very important from a practical point of view, and has the following additional implications. Roughly speaking, under the conditions of Theorem 4.5, that is, if the temporal interactions of the channel are sufficiently strong, the average network QoS is not (approximately) expected to, at least *abruptly,* decrease if one or more relays stop moving at some point. Such event might indeed happen in an actual autonomous network, possibly due to power limitations, or a failure in the motion mechanisms of some network nodes. In the same framework, Theorem 4.5 implies that the relays which continue moving contribute (approximately) positively to increasing the average network QoS, across time slots. Such behavior of the proposed spatially controlled beamforming system may be also confirmed numerically, as discussed in Section 4.3. For the record, and as it will be also shown in Section 4.3, relatively small values for the correlation time γ, such as $\gamma \equiv 5$, are sufficient in order to practically observe the nice system behavior promised by Theorem 4.5. This fact makes the proposed spatially controlled beamforming system attractive in terms of practical feasibility, and shows that such an approach could actually enhance system performance in a well-behaved, real world situation.

4.3 Numerical Simulations & Experimental Validation

In this section, we present synthetic numerical simulations, which essentially confirm that the proposed approach, previously presented in Section 4.2, actually works, and results in relay motion control policies, which yield improved beamforming performance. All synthetic experiments were conducted on an imaginary square terrain of dimensions 30×30 squared units of length, with $\mathcal{W} \equiv [0, 30]^2$, uniformly divided into $30 \times 30 \equiv 900$ square regions. The locations of the source and destination are fixed as $\mathbf{p}_S \equiv [15\,0]^{\boldsymbol{T}}$ and

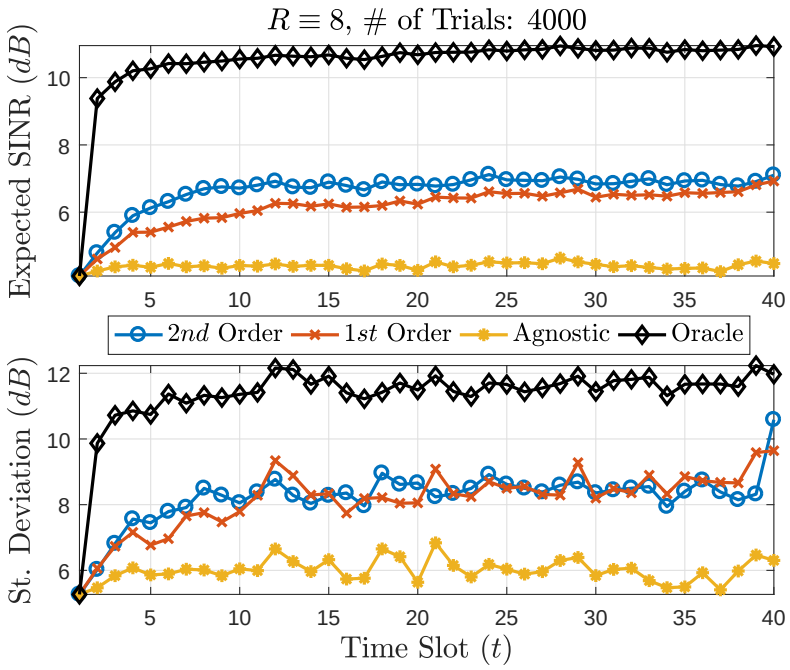


Figure 4.3: Comparison of the proposed strategic relay planning schemes, versus an agnostic, randomized relay motion policy.

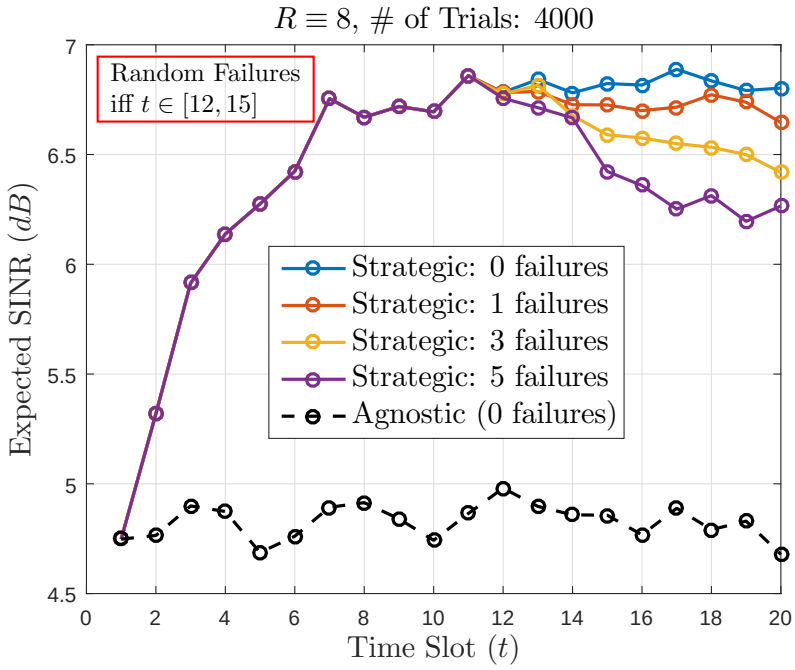
$\mathbf{p}_D \equiv [15\,30]^{\boldsymbol{T}}$. The beamforming temporal horizon is chosen as $T \equiv 40$ and the number of relays is fixed at $R \equiv 8$. The wavelength is chosen as $\lambda \equiv 0.125$, corresponding to a carrier frequency of $2.4\,GHz$. The various parameters of the assumed channel model are set as $\ell \equiv 3$, $\rho \equiv 20$, $\sigma_\xi^2 \equiv 20$, $\eta^2 \equiv 50$, $\beta \equiv 10$, $\gamma \equiv 5$ and $\delta \equiv 1$. The variances of the reception noises at the relays and the destination are fixed as $\sigma^2 \equiv \sigma_D^2 \equiv 1$. Lastly, both the transmission power of the source and the *total* transmission power budget of the relays are chosen as $P \equiv P_c \equiv 25$ ($\approx 14dB$) units of power.

The relays are restricted to the rectangular region $\mathcal{S} \equiv [0, 30] \times [12, 18]$. Further, at each time instant, each of the relays is allowed to move inside a 9-region area, centered at each current position, thus defining its closed set of feasible directions $\mathcal{C}_i$, for each relay $i \in \mathbb{N}_R^+$. Basic collision and out-of-bounds control was also considered and implemented.

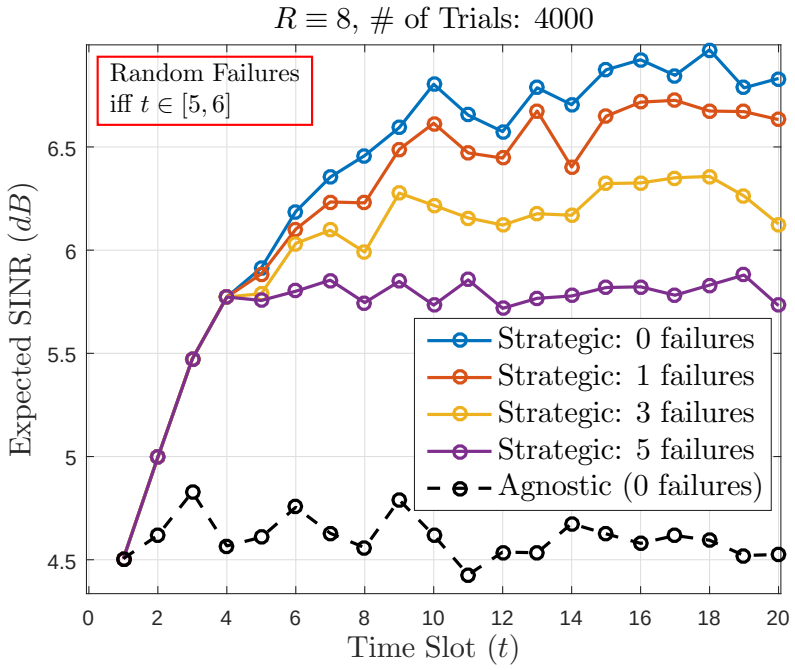
In order to assess the effectiveness of our proposed approach, we compare both heuristics (4.30) and (4.31) against the case where an *agnostic*, *purely randomized* relay control policy is adopted; in this case, at each time slot, each relay moves randomly to a new available position, without taking previously observed CSI into consideration. For simplicity, we do not consider the brute force method presented earlier in Section

4.2.3.2. For reference, we also consider the performance of an oracle control policy at the relays, where, at each time slot $t-1 \in \mathbb{N}^{+}_{N_T-1}$, relay $i \in \mathbb{N}^{+}_{R}$ updates its position by *noncausally* looking into the future and choosing the position $\mathbf{p}_i$, which maximizes directly the quantity $V_I(\mathbf{p}_i, t)$, over $\mathcal{C}_i(\mathbf{p}_i(t-1))$. Of course, the comparison of all controlled systems is made under exactly the same communication environment.

Fig. 4.3 shows the expectation and standard deviation of the achieved QoS for all controlled systems, approximated by executing 4000 trials of the whole experiment. As seen in the figure, there is a clear advantage in exploiting strategically designed relay motion control. Whereas the agnostic system maintains an average SINR of about $4\,dB$ at all times, the system based on the proposed $2nd$ order heuristic (4.31) is clearly superior, exhibiting an increasing trend in the achieved SINR, with a gap starting from about $0.5\,dB$ at time slot $t \equiv 2$, up to $3\,dB$ at time slots $t \equiv 10, 11, \ldots, 40$. The $1st$ order heuristic (4.30) comes second, with always slightly lower average SINR, and which also exhibits a similar increasing trend as the $2nd$ order heuristic (4.31). Additionally, it seems to converge to the performance achieved by (4.31), across time slots. The existence of an increasing trend in the achieved average network QoS has already been predicted by Theorem 4.4 *for a strictly optimal policy*, and our experiments confirm this behavior for both heuristics (4.30) and (4.31), as well. This shows that both heuristics constitute excellent approximations to the original problem (4.12). Consequently, it is both theoretically and experimentally verified that, although the proposed stochastic programming formulation is essentially myopic, the resulting system performance is not, and this is dependent on the fact that the channel exhibits non trivial temporal statistical interactions. We should also comment on the standard deviation of all systems, which, from Fig. 4.3, seems somewhat high, relative to the range of the respective average SINR. This is *exclusively* due to the wild variations of the channel, which, in turn, are due to the effects of shadowing and multipath fading; it is *not* due to the adopted beamforming technique. This is reasonable, since, when the channel is not *actually* in deep fade at time t (an event which might happen with positive probability), the relays, at time $t-1$, are predictively steered to locations, which, most probably, incur higher network QoS. As clearly shown in Fig. 4.3, for all systems under study,

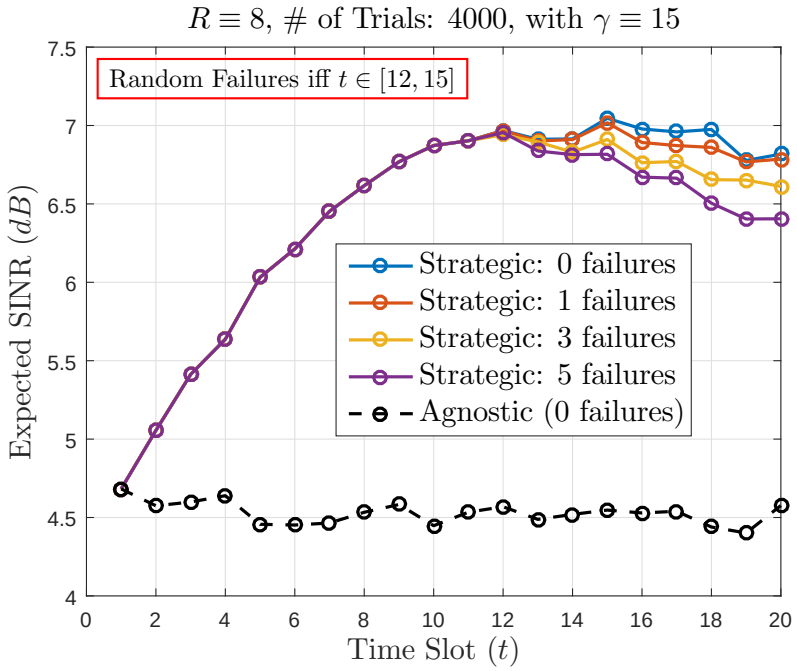


(a)

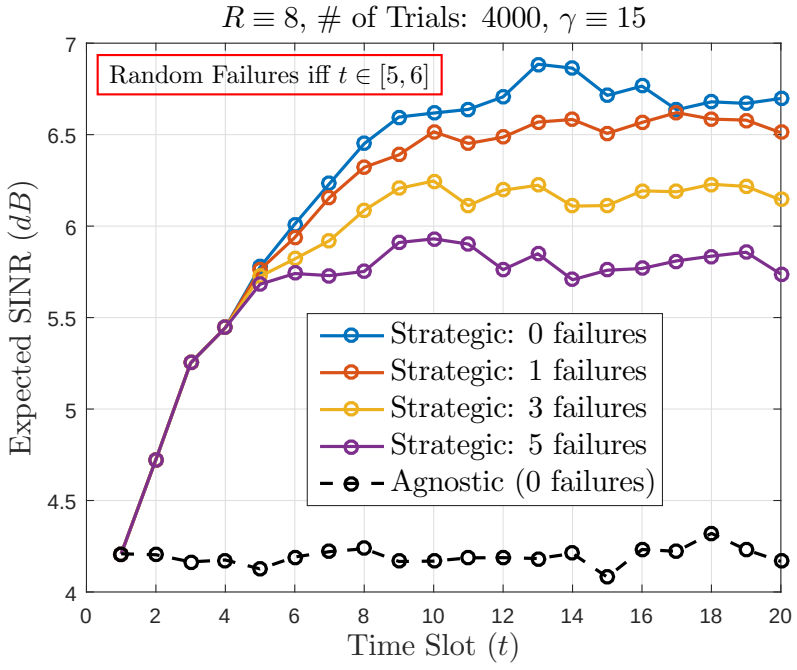


(b)

Figure 4.4: Performance of the proposed spatially controlled system, at the presence of motion failures.



(a)



(b)

Figure 4.5: Performance of the proposed spatially controlled system, at the presence of motion failures.

including that implementing the oracle policy, an increase in system performance also implies a proportional increase in the respective standard deviation.

Next, we experimentally evaluate the performance of the system at the presence of *random motion failures* in the network. Hereafter, we work with the $2nd$ order heuristic (4.31), and set $T \equiv 20$. Random motion failures are modeled by choosing, at each trial, a *random sample* of a fixed number of relays and a *random time* when the failures occur, that is, at each time, the selected relays just stop moving; they continue to beamform staying still, at the position each of them visited last. Two cases are considered; in the first case, motion failures happen if and only if $t \in [12, 15]$ (Figs. 4.4a and 4.5a), whereas, in the second case, $t \in [5, 6]$ (Figs. 4.4b and 4.5b). In both cases, zero, one, three and five relays (chosen at random, at each trial) stop moving. Two cases for γ are considered, $\gamma \equiv 5$ (Figs. 4.4a and 4.4b) and $\gamma \equiv 15$ (Figs. 4.5a and 4.5b).

Again, the results presented in Fig. 4.4 pleasingly confirm our predictions implied by Theorem 4.4 (note, however, that Theorem 4.4 does not support *randomized* motion failures; on the other hand, our simulations are such in order to stress test the proposed system in more adverse motion failure cases). In particular, Fig. 4.4a clearly demonstrates that a larger number of motion failures induces a proportional, relatively (depending on γ) slight decrease in performance; this decrease, though, is smoothly evolving, and is not abrupt. This behavior is more pronounced in Fig. 4.5a, where the correlation time parameter γ has been increased to 15 (recall that, in Theorem 4.4, γ is assumed to be sufficiently large). We readily observe that, in this case, over the same horizon, the operation of the system is smoother, and decrease in performance, as well as its slope, are significantly smaller than those in Fig. 4.4a, for all cases of motion failures. Now, in Figs. 4.4b and 4.5b, when motion failures happen early, well before the network QoS converges to its maximal value, we observe that, although some relays might stop moving at some point, the achieved expected network QoS continues exhibiting its usual increasing trend. Of course, the performance of the system converges values strictly proportional to the number of failures in each of the cases considered. This means that the relays which continue moving contribute positively to increasing network QoS. This has been indeed predicted by Theorem 4.4, as well.

4.4 Conclusions

We have considered the problem of enhancing QoS in time slotted relay beamforming networks with one source/destination, via stochastic relay motion control. Modeling the wireless channel as a spatiotemporal stochastic field, we proposed a novel 2-stage stochastic programming formulation for predictively specifying relay positions, such that the future expected network QoS is maximized, based on causal CSI and under a total relay power constraint. We have shown that this problem can be effectively approximated by a set of simple, two dimensional subproblems, which can be distributively solved, one at each relay. System optimality was tediously analyzed under a rigorous mathematical framework, and our analysis resulted in the development of an extended version of the Fundamental Lemma of Stochastic Control, which constitutes a result of independent interest, as well. We have additionally provided strong theoretical guarantees, characterizing the performance of the proposed system, and showing that the average QoS achieved improves over time. Our simulations confirmed the success of the proposed approach, which results in relay motion control policies yielding significant performance improvement, when compared to agnostic, randomized relay motion.

4.5 Appendices

4.5.1 Appendix A: Measurability & The Fundamental Lemma of Stochastic Control

In the following, aligned with the purposes of this paper, a detailed discussion is presented, which is related to important technical issues, arising towards the analysis and simplification of variational problems of the form of (4.5).

At this point, it would be necessary to introduce some important concepts. Let us first introduce the useful class of *Carathéodory functions* [85, 92][2].

[2]Instead of working with the class of Carathéodory functions, we could also consider the more general class of *random lower semicontinuous functions* [85], which includes the former. However, this might lead to overgeneralization and, thus, we prefer not to do so; the class of Carathéodory functions will be perfectly sufficient for our purposes.

Definition 4.3. (Carathéodory Function) On $(\Omega, \mathscr{F}, \mathcal{P})$, the mapping $H : \Omega \times \mathbb{R}^N \to \overline{\mathbb{R}}$ is called Carathéodory, if and only if $H(\cdot, \boldsymbol{x})$ is $\mathscr{F}$-measurable for all $\boldsymbol{x} \in \mathbb{R}^N$ and $H(\omega, \cdot)$ is continuous for all $\omega \in \Omega$.

In the analysis that follows, we will exploit the notion of *measurability* for closed-valued multifunctions.

Definition 4.4. (Measurable Multifunctions [84, 85]**)** On the measurable space $(\Omega, \mathscr{F})$, a closed-valued multifunction $\mathcal{X} : \Omega \rightrightarrows \mathbb{R}^N$ is $\mathscr{F}$-measurable if and only if, for all closed $\mathcal{A} \subseteq \mathbb{R}^N$, the preimage

$$\mathcal{X}^{-1}(\mathcal{A}) \triangleq \left\{ \omega \in \Omega \,\middle|\, \mathcal{X}(\omega) \bigcap \mathcal{A} \neq \varnothing \right\} \tag{4.68}$$

is in $\mathscr{F}$. If $\mathscr{F}$ constitutes a Borel σ-algebra, generated by a topology on Ω, then an $\mathscr{F}$-measurable $\mathcal{X}$ will be equivalently called *Borel measurable*.

We will also make use of the concept of a *closed multifunction* (Remark 28 in [85], p. 365), whose definition is also presented below, restricted to the case of Euclidean spaces, of interest in this work.

Definition 4.5. (Closed Multifunction) A closed-valued multifunction $\mathcal{X} : \mathbb{R}^M \rightrightarrows \mathbb{R}^N$ (a function from $\mathbb{R}^M$ to closed sets in $\mathbb{R}^N$) is closed if and only if, for all sequences $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ and $\{\boldsymbol{y}_k\}_{k\in\mathbb{N}}$, such that $\boldsymbol{x}_k \underset{k\to\infty}{\longrightarrow} \boldsymbol{x}$, $\boldsymbol{y}_k \underset{k\to\infty}{\longrightarrow} \boldsymbol{y}$ and $\boldsymbol{x}_k \in \mathcal{X}(\boldsymbol{y}_k)$, for all $k \in \mathbb{N}$, it is true that $\boldsymbol{x} \in \mathcal{X}(\boldsymbol{y})$.

4.5.1.1 Random Functions & The Substitution Rule for Conditional Expectations

Given a random function $g(\omega, \boldsymbol{x})$, a sub σ-algebra $\mathscr{Y}$, another $\mathscr{Y}$-measurable random element X, and as long as $\mathbb{E}\{g(\cdot, \boldsymbol{x}) | \mathscr{Y}\}$ exists for all $\boldsymbol{x}$ in the range of X, we would also need to make extensive use of the *substitution rule*

$$\begin{aligned} \mathbb{E}\{g(\cdot, X) | \mathscr{Y}\}(\omega) &\equiv \mathbb{E}\{g(\cdot, X(\omega)) | \mathscr{Y}\}(\omega) \\ &\equiv \mathbb{E}\{g(\cdot, \boldsymbol{x}) | \mathscr{Y}\}(\omega)|_{\boldsymbol{x}\equiv X(\omega)}, \quad \mathcal{P}-a.e., \end{aligned} \tag{4.69}$$

which would allow us to evaluate conditional expectations, by essentially fixing the quantities that are constant relative to the information we are conditioning on, carry out the evaluation, and then let those quantities vary in ω again. Although the substitution rule is a concept readily taken for granted when conditional expectations of Borel measurable functions of random elements (say, from products of Euclidean spaces to $\mathbb{R}$) are considered, it does not hold, in general, for arbitrary random functions. As far as our general formulation is concerned, it is necessary to consider random functions, whose domain is a product of a well behaved space (such as $\mathbb{R}^N$) and the sample space, Ω, whose structure is assumed to be and should be arbitrary, at least in regard to the applications of interest in this work.

One common way to ascertain the validity of the substitution rule is by exploitting the representation of conditional expectations via integrals with respect to the relevant regular conditional distributions, whenever the latter exist. But because of the arbitrary structure of the base space $(\Omega, \mathscr{F}, \mathcal{P})$, regular conditional distributions defined on points in the sample space Ω cannot be guaranteed to exist and, therefore, the substitution rule may fail to hold. However, as we will see, the substitution rule will be very important for establishing the Fundamental Lemma. Therefore, we may choose to impose it as a property on the structures of g and/or X instead, as well as establish sufficient conditions for this property to hold. The relevant definition follows.

Definition 4.6. (Substitution Property (*SP*)) On $(\Omega, \mathscr{F}, \mathcal{P})$, consider a random element $Y : \Omega \to \mathbb{R}^M$, the associated sub σ-algebra $\mathscr{Y} \triangleq \sigma\{Y\} \subseteq \mathscr{F}$, and a random function $g : \Omega \times \mathbb{R}^N \to \mathbb{R}$, such that $\mathbb{E}\{g(\cdot, \boldsymbol{x})\}$ exists for all $\boldsymbol{x} \in \mathbb{R}^N$. Let $\mathfrak{C}_{\mathscr{Y}}$ be any functional class, such that[3]

$$\mathfrak{C}_{\mathscr{Y}} \subseteq \mathfrak{I}_{\mathscr{Y}} \triangleq \left\{ X : \Omega \to \mathbb{R}^N \,\middle|\, \begin{array}{c} X^{-1}(\mathcal{A}) \in \mathscr{Y}, \text{ for all } \mathcal{A} \in \mathscr{B}\left(\mathbb{R}^N\right) \\ \mathbb{E}\{g(\cdot, X)\} \text{ exists} \end{array} \right\}. \tag{4.70}$$

We say that g *possesses the Substitution Property within* $\mathfrak{C}_{\mathscr{Y}}$, or, equivalently, that g *is* $\boldsymbol{SP}\diamondsuit\mathfrak{C}_{\mathscr{Y}}$, if and only if there exists a jointly Borel measurable function $h : \mathbb{R}^M \times \mathbb{R}^N \to$

[3]Hereafter, statements of type "$\mathbb{E}\{g(\cdot, X)\}$ exists" will *implicitly* imply that $g(\cdot, X)$ is an $\mathscr{F}$-measurable function.

$\overline{\mathbb{R}}$, with $h\left(Y\left(\omega\right),\boldsymbol{x}\right)\equiv\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)|\,\mathscr{Y}\right\}\left(\omega\right)$, everywhere in $\left(\omega,\boldsymbol{x}\right)\in\Omega\times\mathbb{R}^{N}$, such that, for any $X\in\mathfrak{C}_{\mathscr{Y}}$, it is true that

$$\mathbb{E}\left\{g\left(\cdot,X\right)|\,\mathscr{Y}\right\}\left(\omega\right)\equiv h\left(Y\left(\omega\right),X\left(\omega\right)\right),\tag{4.71}$$

almost everywhere in $\omega\in\Omega$ with respect to $\mathcal{P}$.

Remark 4.6. Observe that, in Definition 4.6, h is required to be the same for all $X\in\mathfrak{C}_{\mathscr{Y}}$. That is, h should be determined only by the structure of g, relative to $\mathscr{Y}$, regardless of the specific X within $\mathfrak{C}_{\mathscr{Y}}$, considered each time. On the other hand, it is also important to note that the set of unity measure, where (4.71) is valid, *might indeed be dependent on the particular* X. ■

Remark 4.7. Another detail of Definition 4.6 is that, because $\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)\right\}$ is assumed to exist for all $\boldsymbol{x}\in\mathbb{R}^{N}$, $\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)|\,\mathscr{Y}\right\}$ also exists and, as an extended $\mathscr{Y}$-measurable random variable, for every $\boldsymbol{x}\in\mathbb{R}^{N}$, there exists a Borel measurable function $h_{\boldsymbol{x}}:\mathbb{R}^{M}\rightarrow\overline{\mathbb{R}}$, such that

$$h_{\boldsymbol{x}}\left(Y\left(\omega\right)\right)\equiv\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)|\,\mathscr{Y}\right\}\left(\omega\right),\quad\forall\omega\in\Omega.\tag{4.72}$$

One may then readily define a function $h:\mathbb{R}^{M}\times\mathbb{R}^{N}\rightarrow\overline{\mathbb{R}}$, such that $h\left(Y\left(\omega\right),\boldsymbol{x}\right)\equiv\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)|\,\mathscr{Y}\right\}\left(\omega\right)$, uniformly ***for all points*** ω, ***of the sample space*** Ω. This is an extremely important fact, in regard to the analysis that follows. Observe, however, that, in general, h will be Borel measurable *only in its first argument*; h is not guaranteed to be measurable in $\boldsymbol{x}\in\mathbb{R}^{N}$, for each $Y\in\mathbb{R}^{M}$, let alone jointly measurable in both its arguments. ■

Remark 4.8. **(Generalized *SP*)** Definition 4.6 may be reformulated in a more general setting. In particular, $\mathscr{Y}$ may be assumed to be any arbitrary sub σ-algebra of $\mathscr{F}$, but with the subtle difference that, in such case, one would instead directly demand that the random function $h:\Omega\times\mathbb{R}^{N}\rightarrow\overline{\mathbb{R}}$, with $h\left(\omega,\boldsymbol{x}\right)\equiv\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)|\,\mathscr{Y}\right\}\left(\omega\right)$, everywhere in $\left(\omega,\boldsymbol{x}\right)\in\Omega\times\mathbb{R}^{N}$, is jointly $\mathscr{Y}\otimes\mathscr{B}\left(\mathbb{R}^{N}\right)$-measurable and such that, for any $X\in\mathfrak{C}_{\mathscr{Y}}$ (with $\mathfrak{C}_{\mathscr{Y}}$ defined accordingly), it is true that

$$\mathbb{E}\left\{g\left(\cdot,X\right)|\,\mathscr{Y}\right\}\left(\omega\right)\equiv h\left(\omega,X\left(\omega\right)\right),\quad\mathcal{P}-a.e..\tag{4.73}$$

Although such a generalized definition of the substitution property is certainly less enlightening, it is still useful. Specifically, this version of $\boldsymbol{SP}$ is explicitly used in the statement and proof of Theorem 4.4, presented in Section 4.2.4. ■

Keeping $(\Omega, \mathscr{F}, \mathcal{P})$ of arbitrary structure, we will be interested in the set of g's which are $\boldsymbol{SP}\diamondsuit\mathfrak{I}_{\mathscr{Y}}$. The next result provides a large class of such random functions, which is sufficient for our purposes.

Theorem 4.6. (Sufficient Conditions for the $\boldsymbol{SP}\diamondsuit\mathfrak{I}_{\mathscr{Y}}$) *On $(\Omega, \mathscr{F}, \mathcal{P})$, consider a random element $Y : \Omega \to \mathbb{R}^M$, the associated sub σ-algebra $\mathscr{Y} \triangleq \sigma\{Y\} \subseteq \mathscr{F}$, and a random function $g : \Omega \times \mathbb{R}^N \to \mathbb{R}$. Suppose that:*

- *g is dominated by a $\mathcal{P}$-integrable function; that is,*

$$\exists \psi \in \mathcal{L}_1(\Omega, \mathscr{F}, \mathcal{P}; \mathbb{R}), \text{ such that } \sup_{\boldsymbol{x} \in \mathbb{R}^N} |g(\omega, \boldsymbol{x})| \leq \psi(\omega), \quad \forall \omega \in \Omega, \tag{4.74}$$

- *g is Carathéodory on $\Omega \times \mathbb{R}^N$, and that*
- *the extended real valued function $\mathbb{E}\{g(\cdot, \boldsymbol{x}) | \mathscr{Y}\}$ is Carathéodory on $\Omega \times \mathbb{R}^N$.*

Then, g is $\boldsymbol{SP}\diamondsuit\mathfrak{I}_{\mathscr{Y}}$.

Proof of Theorem 4.6. Under the setting of the theorem, consider any $\mathscr{Y}$-measurable random element $X : \Omega \to \mathbb{R}^N$, for which $\mathbb{E}\{g(\cdot, X)\}$ exists. Then, $\mathbb{E}\{g(\cdot, X) | \mathscr{Y}\}$ exists. Also, by domination of g by ψ, for all $\boldsymbol{x} \in \mathbb{R}^N$, $\mathbb{E}\{g(\cdot, \boldsymbol{x}) | \mathscr{Y}\}$ exists and constitutes a $\mathcal{P}$-integrable, $\mathscr{Y}$-measurable random variable. By Remark 4.7, we know that

$$\mathbb{E}\{g(\cdot, \boldsymbol{x}) | \mathscr{Y}\}(\omega) \equiv h(Y(\omega), \boldsymbol{x}), \quad \forall (\omega, \boldsymbol{x}) \in \Omega \times \mathbb{R}^N, \tag{4.75}$$

where $h : \mathbb{R}^M \times \mathbb{R}^N \to \overline{\mathbb{R}}$ is Borel measurable *in its first argument.* However, since $\mathbb{E}\{g(\cdot, \boldsymbol{x}) | \mathscr{Y}\}(\omega)$
$\equiv h(Y(\omega), \boldsymbol{x})$ is Carathéodory on $\Omega \times \mathbb{R}^N$, h is Carathéodory on $\mathbb{R}^M \times \mathbb{R}^N$, as well. Thus, h will be jointly $\mathscr{B}\left(\mathbb{R}^M\right) \otimes \mathscr{B}\left(\mathbb{R}^N\right)$-measurable (Lemma 4.51 in [92], along with the fact that $\overline{\mathbb{R}}$ is metrizable).

We claim that, actually, h is such that

$$\mathbb{E}\left\{g\left(\cdot,X\right)|\,\mathscr{Y}\right\}\equiv h\left(Y,X\right),\quad \mathcal{P}-a.e.. \tag{4.76}$$

Employing a common technique, the result will be proven in steps, starting from indicators and building up to arbitrary measurable functions, as far as X is concerned. Before embarking with the core of the proof, note that, for any $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\mathbb{R}^N$ and any $\mathcal{A}\in\mathscr{F}$, the sum $g\left(\cdot,\boldsymbol{x}_1\right)\mathbb{1}_{\mathcal{A}}+g\left(\cdot,\boldsymbol{x}_2\right)\mathbb{1}_{\mathcal{A}^c}$ is always well defined, and $\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_1\right)\mathbb{1}_{\mathcal{A}}\right\}$ and $\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_2\right)\mathbb{1}_{\mathcal{A}^c}\right\}$ both exist and are finite by domination. This implies that $\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_1\right)\mathbb{1}_{\mathcal{A}}\right\}+\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_2\right)\mathbb{1}_{\mathcal{A}^c}\right\}$ is always well-defined, which in turn implies the validity of the additivity properties (Theorem 1.6.3 and Theorem 5.5.2 in [27])

$$\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_1\right)\mathbb{1}_{\mathcal{A}}+g\left(\cdot,\boldsymbol{x}_2\right)\mathbb{1}_{\mathcal{A}^c}\right\}\equiv\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_1\right)\mathbb{1}_{\mathcal{A}}\right\}+\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_2\right)\mathbb{1}_{\mathcal{A}^c}\right\}\in\mathbb{R},\text{ and} \tag{4.77}$$

$$\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_1\right)\mathbb{1}_{\mathcal{A}}+g\left(\cdot,\boldsymbol{x}_2\right)\mathbb{1}_{\mathcal{A}^c}|\,\mathscr{Y}\right\}\equiv\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_1\right)\mathbb{1}_{\mathcal{A}}|\,\mathscr{Y}\right\}+\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}_2\right)\mathbb{1}_{\mathcal{A}^c}|\,\mathscr{Y}\right\}, \tag{4.78}$$

$\mathcal{P}-a.e.$. Hence, under our setting, any such manipulation is technically justified.

Suppose first that $X\left(\omega\right)\equiv\widetilde{\boldsymbol{x}}\mathbb{1}_{\mathcal{A}}\left(\omega\right)$, for some $\widetilde{\boldsymbol{x}}\in\mathbb{R}^N$ and some $\mathcal{A}\in\mathscr{Y}$. Then, by ([27], Theorem 5.5.11 & Comment 5.5.12), it is true that

$$\begin{aligned}
\mathbb{E}\left\{g\left(\cdot,X\right)|\,\mathscr{Y}\right\}&\equiv\mathbb{E}\left\{g\left(\cdot,\widetilde{\boldsymbol{x}}\right)\mathbb{1}_{\mathcal{A}}+g\left(\cdot,\mathbf{0}\right)\mathbb{1}_{\mathcal{A}^c}|\,\mathscr{Y}\right\}\\
&\equiv\mathbb{E}\left\{g\left(\cdot,\widetilde{\boldsymbol{x}}\right)\mathbb{1}_{\mathcal{A}}|\,\mathscr{Y}\right\}+\mathbb{E}\left\{g\left(\cdot,\mathbf{0}\right)\mathbb{1}_{\mathcal{A}^c}|\,\mathscr{Y}\right\}\\
&\equiv\mathbb{1}_{\mathcal{A}}\mathbb{E}\left\{g\left(\cdot,\widetilde{\boldsymbol{x}}\right)|\,\mathscr{Y}\right\}+\mathbb{1}_{\mathcal{A}^c}\mathbb{E}\left\{g\left(\cdot,\mathbf{0}\right)|\,\mathscr{Y}\right\}\\
&\equiv\mathbb{1}_{\mathcal{A}}h\left(Y,\widetilde{\boldsymbol{x}}\right)+\mathbb{1}_{\mathcal{A}^c}h\left(Y,\mathbf{0}\right)\\
&\equiv h\left(Y,\widetilde{\boldsymbol{x}}\mathbb{1}_{\mathcal{A}}\right)\\
&\equiv h\left(Y,X\right),\quad\mathcal{P}-a.e.,
\end{aligned} \tag{4.79}$$

proving the claim for indicators.

Consider now simple functions of the form

$$X\left(\omega\right)\equiv\sum_{i\in\mathbb{N}_I^+}\widetilde{\boldsymbol{x}}_i\mathbb{1}_{\mathcal{A}_i}\left(\omega\right), \tag{4.80}$$

where $\widetilde{\boldsymbol{x}}_i\in\mathbb{R}^N$, $\mathcal{A}_i\in\mathscr{Y}$, for all $i\in\mathbb{N}_I^+$, with $\mathcal{A}_i\bigcap\mathcal{A}_j\equiv\varnothing$, for $i\neq j$ and $\bigcup_{i\in\mathbb{N}_I^+}\mathcal{A}_i\equiv\Omega$.

Then, we again have

$$\begin{aligned}
\mathbb{E}\left\{ g\left(\cdot, X\right) \middle| \mathscr{Y} \right\} &\equiv \mathbb{E}\left\{ \sum_{i \in \mathbb{N}_I^+} g\left(\cdot, \widetilde{\boldsymbol{x}}_i\right) \mathbb{1}_{\mathcal{A}_i} \middle| \mathscr{Y} \right\} \\
&\equiv \sum_{i \in \mathbb{N}_I^+} \mathbb{E}\left\{ g\left(\cdot, \widetilde{\boldsymbol{x}}_i\right) \mathbb{1}_{\mathcal{A}_i} \middle| \mathscr{Y} \right\} \\
&\equiv \sum_{i \in \mathbb{N}_I^+} \mathbb{1}_{\mathcal{A}_i} \mathbb{E}\left\{ g\left(\cdot, \widetilde{\boldsymbol{x}}_i\right) \middle| \mathscr{Y} \right\} \\
&\equiv \sum_{i \in \mathbb{N}_I^+} \mathbb{1}_{\mathcal{A}_i} h\left(Y, \widetilde{\boldsymbol{x}}_i\right) \\
&\equiv h\left(Y, \sum_{i \in \mathbb{N}_I^+} \widetilde{\boldsymbol{x}}_i \mathbb{1}_{\mathcal{A}_i} \right) \\
&\equiv h\left(Y, X\right), \quad \mathcal{P} - a.e., \qquad (4.81)
\end{aligned}$$

and the proved is claimed for simple functions.

To show that our claims are true for any arbitrary random function g, we take advantage of the continuity of both h and g in $\boldsymbol{x}$. First, we know that h is Carathéodory, which means that, for every $\omega \in \Omega$, if any sequence $\left\{ \boldsymbol{x}_n \in \mathbb{R}^N \right\}_{n \in \mathbb{N}}$ is such that $\boldsymbol{x}_n \underset{n \to \infty}{\longrightarrow} \boldsymbol{x}$ (for arbitrary $\boldsymbol{x} \in \mathbb{R}^N$), it is true that

$$h\left(Y\left(\omega\right), \boldsymbol{x}_n\right) \equiv \mathbb{E}\left\{ g\left(\cdot, \boldsymbol{x}_n\right) \middle| \mathscr{Y} \right\}\left(\omega\right) \underset{n \to \infty}{\longrightarrow} \mathbb{E}\left\{ g\left(\cdot, \boldsymbol{x}\right) \middle| \mathscr{Y} \right\}\left(\omega\right) \equiv h\left(Y\left(\omega\right), \boldsymbol{x}\right). \qquad (4.82)$$

Second, we know that g is Carathéodory as well, also implying that, for every $\omega \in \Omega$, if any sequence $\left\{ \boldsymbol{x}_n \in \mathbb{R}^N \right\}_{n \in \mathbb{N}}$ is such that $\boldsymbol{x}_n \underset{n \to \infty}{\longrightarrow} \boldsymbol{x}$, it is true that

$$g\left(\omega, \boldsymbol{x}_n\right) \underset{n \to \infty}{\longrightarrow} g\left(\omega, \boldsymbol{x}\right). \qquad (4.83)$$

Next, let $\left\{ X_n : \Omega \to \mathbb{R}^N \right\}_{n \in \mathbb{N}}$ be a sequence of simple Borel functions, such that, for all $\omega \in \Omega$,

$$X_n\left(\omega\right) \underset{n \to \infty}{\longrightarrow} X\left(\omega\right). \qquad (4.84)$$

Note that such a sequence always exists (see Theorem 1.5.5 (b) in [27]). Consequently, for each $\omega \in \Omega$, we may write (note that g is $\mathscr{F} \otimes \mathscr{B}\left(\mathbb{R}^N\right)$-measurable; see ([92], Lemma 4.51))

$$g\left(\omega, X_n\left(\omega\right)\right) \underset{n \to \infty}{\longrightarrow} g\left(\omega, X\left(\omega\right)\right), \qquad (4.85)$$

that is, the sequence $\{g(\cdot, X_n)\}_{n\in\mathbb{N}}$ converges to $g(\cdot, X)$, everywhere in Ω.

Now, let us try to apply the Dominated Convergence Theorem for conditional expectations (Theorem 5.5.5 in [27]) to the aforementioned sequence of functions. Of course, we have to show that all members of the sequence $\{g(\cdot, X_n)\}_{n\in\mathbb{N}}$ are dominated by another integrable function, uniformly in $n \in \mathbb{N}$. By assumption, there exists an integrable function $\psi : \Omega \to \mathbb{R}$, such that

$$|g(\omega, \boldsymbol{x})| \leq \psi(\omega), \quad \forall (\omega, \boldsymbol{x}) \in \Omega \times \mathbb{R}^N. \tag{4.86}$$

In particular, it must also be true that

$$|g(\omega, X_n(\omega))| \leq \psi(\omega), \quad \forall (\omega, n) \in \Omega \times \mathbb{N}, \tag{4.87}$$

verifying the domination requirement. Thus, Dominated Convergence implies the existence of an event $\Omega_{\Pi_1} \subseteq \Omega$, with $\mathcal{P}(\Omega_{\Pi_1}) \equiv 1$, such that, for all $\omega \in \Omega_{\Pi_1}$,

$$\mathbb{E}\{g(\cdot, X_n)|\mathscr{Y}\}(\omega) \underset{n\to\infty}{\longrightarrow} \mathbb{E}\{g(\cdot, X)|\mathscr{Y}\}(\omega). \tag{4.88}$$

Also, for every $\omega \in \Omega \bigcap \Omega_{\Pi_1} \equiv \Omega_{\Pi_1}$, (4.82) yields

$$h(Y(\omega), X_n(\omega)) \underset{n\to\infty}{\longrightarrow} h(Y(\omega), X(\omega)). \tag{4.89}$$

However, by what we have shown above, because the sequence $\{X_n\}_{n\in\mathbb{N}}$ consists of simple functions, then, for every $n \in \mathbb{N}$, there exists $\Omega_{\Pi^n} \subseteq \Omega$, with $\mathcal{P}(\Omega_{\Pi^n}) \equiv 1$, such that, for all $\omega \in \Omega_{\Pi^n}$,

$$\mathbb{E}\{g(\cdot, X_n)|\mathscr{Y}\}(\omega) \equiv h(Y(\omega), X_n(\omega)). \tag{4.90}$$

Since $\mathbb{N}$ is countable, there exists a "global" event $\Omega_{\Pi_2} \subseteq \Omega$, with $\mathcal{P}(\Omega_{\Pi_2}) \equiv 1$, such that, for all $\omega \in \Omega_{\Pi_2}$,

$$\mathbb{E}\{g(\cdot, X_n)|\mathscr{Y}\}(\omega) \equiv h(Y(\omega), X_n(\omega)), \quad \forall n \in \mathbb{N}. \tag{4.91}$$

Now define the event $\Omega_{\Pi_3} \triangleq \Omega_{\Pi_1} \bigcap \Omega_{\Pi_2}$. Of course, $\mathcal{P}(\Omega_{\Pi_3}) \equiv 1$. Then, for every $\omega \in \Omega_{\Pi_3}$, (4.88), (4.89) and (4.91) all hold simultaneously. Therefore, for every $\omega \in \Omega_{\Pi_3}$, it is true that (say)

$$h(Y(\omega), X_n(\omega)) \underset{n\to\infty}{\longrightarrow} \mathbb{E}\{g(\cdot, X)|\mathscr{Y}\}(\omega) \quad \text{and} \tag{4.92}$$

$$h\left(Y\left(\omega\right),X_{n}\left(\omega\right)\right)\underset{n\rightarrow\infty}{\longrightarrow}h\left(Y\left(\omega\right),X\left(\omega\right)\right), \tag{4.93}$$

which immediately yields

$$\mathbb{E}\left\{g\left(\cdot,X\right)|\,\mathscr{Y}\right\}\left(\omega\right)\equiv h\left(Y\left(\omega\right),X\left(\omega\right)\right),\quad\mathcal{P}-a.e., \tag{4.94}$$

showing that g is $\boldsymbol{SP}\diamondsuit\mathfrak{I}_{\mathscr{Y}}$. ■

Remark 4.9. Note that the assumptions of Theorem 4.6 can be significantly weakened, guaranteeing the validity of the substitution rule for vastly discontinuous random functions, including, for instance, cases with random discontinuities, or random jumps. This extended analysis, though, is out of the scope of the paper and will be presented elsewhere. ■

4.5.1.2 A Base Form of the Lemma

We will first state a base, very versatile version of the Fundamental Lemma, treating a general class of problems, which includes the particular stochastic problem of interest, (4.5), as a subcase.

Lemma 4.2. (Fundamental Lemma / Base Version) *On $(\Omega,\mathscr{F},\mathcal{P})$, consider a random element $Y:\Omega\rightarrow\mathbb{R}^{M}$, the sub σ-algebra $\mathscr{Y}\triangleq\sigma\left\{Y\right\}\subseteq\mathscr{F}$, a random function $g:\Omega\times\mathbb{R}^{N}\rightarrow\mathbb{R}$, such that $\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)\right\}$ exists for all $\boldsymbol{x}\in\mathbb{R}^{N}$, a Borel measurable closed-valued multifunction $\mathcal{X}:\mathbb{R}^{N}\rightrightarrows\mathbb{R}^{N}$, with $\mathrm{dom}\left(\mathcal{X}\right)\equiv\mathbb{R}^{N}$, as well as another $\mathscr{Y}$-measurable random element $Z_{Y}:\Omega\rightarrow\mathbb{R}^{N}$, with $Z_{Y}\left(\omega\right)\equiv\mathcal{Z}\left(Y\left(\omega\right)\right)$, for all $\omega\in\Omega$, for some Borel $\mathcal{Z}:\mathbb{R}^{M}\rightarrow\mathbb{R}^{N}$. Consider also the decision set*

$$\mathcal{F}_{\mathcal{X}\left(Z_{Y}\right)}^{\mathscr{Y}}\triangleq\left\{X:\Omega\rightarrow\mathbb{R}^{N}\left|\begin{array}{l}X\left(\omega\right)\in\mathcal{X}\left(Z_{Y}\left(\omega\right)\right),\ a.e.\ in\ \omega\in\Omega\\X^{-1}\left(\mathcal{A}\right)\in\mathscr{Y},\ for\ all\ \mathcal{A}\in\mathscr{B}\left(\mathbb{R}^{N}\right)\end{array}\right.\right\}, \tag{4.95}$$

containing all $\mathscr{Y}$-measurable selections of $\mathcal{X}\left(Z_{Y}\right)$. Then, $\mathcal{F}_{\mathcal{X}\left(Z_{Y}\right)}^{\mathscr{Y}}$ is nonempty. Suppose that:

- *$\mathbb{E}\left\{g\left(\cdot,X\right)\right\}$ exists for all $X\in\mathcal{F}_{\mathcal{X}\left(Z_{Y}\right)}^{\mathscr{Y}}$, with $\inf_{X\in\mathcal{F}_{\mathcal{X}\left(Z_{Y}\right)}^{\mathscr{Y}}}\mathbb{E}\left\{g\left(\cdot,X\right)\right\}<+\infty$, and that*

- g *is* $\boldsymbol{SP}\diamondsuit\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$.

Then, if $\overline{\mathscr{Y}}$ *denotes the completion of* $\mathscr{Y}$ *relative to the restriction* $\mathcal{P}|_{\mathscr{Y}}$*, then the optimal value function* $\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)|\,\mathscr{Y}\right\}\triangleq\vartheta$ *is* $\overline{\mathscr{Y}}$*-measurable and it is true that*

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{g\left(\cdot,X\right)\right\}\equiv\mathbb{E}\left\{\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)|\,\mathscr{Y}\right\}\right\}\equiv\mathbb{E}\left\{\vartheta\right\}. \tag{4.96}$$

In other words, variational minimization over $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ *is exchangeable by pointwise (over constants) minimization over the random multifunction* $\mathcal{X}\left(Z_Y\right)$*, relative to* $\mathscr{Y}$*.*

Remark 4.10. Note that, in the statement of Theorem 4.7, assuming that the infimum of $\mathbb{E}\left\{g\left(\cdot,X\right)\right\}$ over $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ is less than $+\infty$ is *equivalent* to assuming the existence of an X in $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, such that $\mathbb{E}\left\{g\left(\cdot,X\right)\right\}$ is less than $+\infty$. ■

Before embarking with the proof of Lemma 4.2, it would be necessary to state an old, fundamental selection theorem, due to Mackey [93].

Theorem 4.7. (Borel Measurable Selections [93]**)** *Let* $\left(\mathcal{S}_1,\mathscr{B}\left(\mathcal{S}_1\right)\right)$ *and* $\left(\mathcal{S}_2,\mathscr{B}\left(\mathcal{S}_2\right)\right)$ *be Borel spaces and let* $\left(\mathcal{S}_2,\mathscr{B}\left(\mathcal{S}_2\right)\right)$ *be standard. Let* $\mu:\mathscr{B}\left(\mathcal{S}_1\right)\rightarrow[0,\infty]$ *be a standard measure on* $\left(\mathcal{S}_1,\mathscr{B}\left(\mathcal{S}_1\right)\right)$*. Suppose that* $\mathcal{A}\in\mathscr{B}\left(\mathcal{S}_1\right)\otimes\mathscr{B}\left(\mathcal{S}_2\right)$*, such that, for each* $y\in\mathcal{S}_1$*, there exists* $x_y\in\mathcal{S}_2$*, so that* $\left(y,x_y\right)\in\mathcal{A}$*. Then, there exists a Borel subset* $\mathcal{O}\in\mathscr{B}\left(\mathcal{S}_1\right)$ *with* $\mu\left(\mathcal{O}\right)\equiv0$*, as well as a Borel measurable function* $\phi:\mathcal{S}_1\rightarrow\mathcal{S}_2$*, such that* $\left(y,\phi\left(y\right)\right)\in\mathcal{A}$*, for all* $y\in\mathcal{S}_1\setminus\mathcal{O}$*.*

Remark 4.11. Theorem 4.7 refers to the concepts of a *Borel space,* a *standard Borel space* and a *standard measure*. These are employed as structural assumptions, in order for the conclusions of the theorem to hold true. In this paper, except for the base probability space $(\Omega,\mathscr{F},\mathcal{P})$, whose structure may be arbitrary, all other spaces and measures considered will satisfy those assumptions by default. We thus choose not to present the respective definitions; instead, the interested reader is referred to the original article, [93]. ■

We are now ready to prove Lemma 4.2, as follows.

Proof of Lemma 4.2. As usual with such results, the proof will rely on showing a double sided inequality [82–84, 86, 87, 94]. There is one major difficulty, though, in the

optimization setting considered, because all infima may be potentially unattainable, within the respective decision sets. However, it is immediately evident that, because g is assumed to be $\boldsymbol{SP}\diamondsuit\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, and via a simple application of the tower property, it will suffice to show that

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}} \mathbb{E}\left\{h\left(Y,X\right)\right\} \equiv \mathbb{E}\left\{\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)} h\left(Y,\boldsymbol{x}\right)\right\}. \tag{4.97}$$

This is because it is true that, for any $\mathscr{Y}$-measurable selection of $\mathcal{X}\left(Z_Y\right)$, say $X:\Omega\rightarrow\mathbb{R}^N$, for which $\mathbb{E}\left\{g\left(\cdot,X\right)\right\}$ exists,

$$\mathbb{E}\left\{g\left(\cdot,X\right)\middle|\mathscr{Y}\right\}\left(\omega\right) \equiv h\left(Y\left(\omega\right),\boldsymbol{x}\right)\big|_{\boldsymbol{x}=X(\omega)}, \quad \forall\omega\in\Omega_{\Pi_X}, \tag{4.98}$$

where the event $\Omega_{\Pi_X}\in\mathscr{F}$ is such that $\mathcal{P}\left(\Omega_{\Pi_X}\right)\equiv 1$ and h is jointly Borel, satisfying

$$\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)\middle|\mathscr{Y}\right\} \equiv h\left(Y\left(\omega\right),\boldsymbol{x}\right), \tag{4.99}$$

everywhere in $(\omega,\boldsymbol{x})\in\Omega\times\mathbb{R}^N$.

For the sake of clarity in the exposition, we will break the proof into a number of discrete subsections, providing a tractable roadmap to the final result.

Step 1. $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ *is nonempty.*

It suffices to show that there exists at least one $\mathscr{Y}$-measurable selection of $\mathcal{X}\left(Z_Y\right)$, that is, a $\mathscr{Y}$-measurable random variable, say $X:\Omega\rightarrow\mathbb{R}^N$, such that $X\left(\omega\right)\in\mathcal{X}\left(Z_Y\left(\omega\right)\right)$, for all ω in the domain of $\mathcal{X}\left(Z_Y\right)$.

We first show that the composite multifunction $\mathcal{X}\left(Z_Y\left(\cdot\right)\right):\Omega\rightrightarrows\mathbb{R}^N$ is $\mathscr{Y}$-measurable. Recall from Definition 4.4 that it suffices to show that

$$\mathcal{X}Z_Y^{-1}\left(\mathcal{A}\right) \triangleq \left\{\omega\in\Omega\,\middle|\,\mathcal{X}\left(Z_Y\left(\omega\right)\right)\bigcap\mathcal{A}\neq\varnothing\right\}\in\mathscr{Y}, \tag{4.100}$$

for every closed $\mathcal{A}\subseteq\mathbb{R}^N$. Since the closed-valued multifunction $\mathcal{X}$ is Borel measurable, it is true that $\mathcal{X}^{-1}\left(\mathcal{A}\right)\in\mathscr{B}\left(\mathbb{R}^N\right)$, for all closed $\mathcal{A}\subseteq\mathbb{R}^N$. We also know that Z_Y is $\mathscr{Y}$-measurable, or that $Z_Y^{-1}\left(\mathcal{B}\right)\in\mathscr{Y}$, for all $\mathcal{B}\in\mathscr{B}\left(\mathbb{R}^N\right)$. Setting $\mathcal{B}\equiv\mathcal{X}^{-1}\left(\mathcal{A}\right)\in\mathscr{B}\left(\mathbb{R}^N\right)$, for any arbitrary closed $\mathcal{A}\subseteq\mathbb{R}^N$, it is true that

$$\mathscr{Y}\ni Z_Y^{-1}\left(\mathcal{X}^{-1}\left(\mathcal{A}\right)\right) \equiv \left\{\omega\in\Omega\,\middle|\,Z_Y\left(\omega\right)\in\mathcal{X}^{-1}\left(\mathcal{A}\right)\right\}$$

$$\equiv \left\{\omega \in \Omega \,\middle|\, \mathcal{X}\left(Z_Y\left(\omega\right)\right) \bigcap \mathcal{A} \neq \varnothing \right\}$$
$$\equiv \mathcal{X} Z_Y^{-1}\left(\mathcal{A}\right), \tag{4.101}$$

and, thus, the composition $\mathcal{X}\left(Z_Y\left(\cdot\right)\right)$ is $\mathscr{Y}$-measurable, or, in other words, measurable on the measurable (sub)space $(\Omega, \mathscr{Y})$.

Now, since the closed-valued multifunction $\mathcal{X}\left(Z_Y\right)$ is measurable on $(\Omega, \mathscr{Y})$, it admits a *Castaing Representation* (Theorem 14.5 in [84] & Theorem 7.34 in [85]). Therefore, there exists at least one $\mathscr{Y}$-measurable selection of $\mathcal{X}\left(Z_Y\right)$, which means that $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ contains at least one element. ★

Step 2. *ϑ is $\overline{\mathscr{Y}}$-measurable.*

To show the validity of this statement, we first demonstrate that, for any chosen $h : \mathbb{R}^M \times \mathbb{R}^N \to \overline{\mathbb{R}}$, as in Definition 4.6, the function $\xi : \mathbb{R}^M \to \overline{\mathbb{R}}$, defined as

$$\xi\left(\boldsymbol{y}\right) \triangleq \inf_{\boldsymbol{x} \in \mathcal{X}(\mathcal{Z}(\boldsymbol{y}))} h\left(\boldsymbol{y}, \boldsymbol{x}\right), \quad \forall \boldsymbol{y} \in \mathbb{R}^M, \tag{4.102}$$

is measurable relative to $\overline{\mathscr{B}}\left(\mathbb{R}^M\right)$, the completion of $\mathscr{B}\left(\mathbb{R}^M\right)$ relative to the pushforward $\mathcal{P}_Y$. This follows easily from the following facts. First, the *graph* of the measurable multifunction $\mathcal{X}\left(\mathcal{Z}\left(\cdot\right)\right)$ is itself measurable and in $\mathscr{B}\left(\mathbb{R}^M\right) \otimes \mathscr{B}\left(\mathbb{R}^N\right)$ (Theorem 14.8 in [84]), and, therefore, *analytic* (Appendix A.2 in [86]). Second, h is jointly Borel measurable and, therefore, a *lower semianalytic* function (Appendix A.2 in [86]). As a result, ([87], Proposition 7.47) implies that ξ is also lower semianalytic, and, consequently, *universally measurable* (Appendix A.2 in [86]). Being universally measurable, ξ is also measurable relative to $\overline{\mathscr{B}}\left(\mathbb{R}^M\right)$, thus proving our claim. We also rely on the definitions of both $\overline{\mathscr{Y}}$ and $\overline{\mathscr{B}}\left(\mathbb{R}^M\right)$, stated as (Theorem 1.9 in [95])

$$\mathcal{B} \in \overline{\mathscr{Y}} \iff \mathcal{B} \equiv \mathcal{C} \bigcup \mathcal{D} \,\middle|\, \mathcal{C} \in \mathscr{Y} \text{ and } \mathcal{D} \subseteq \mathcal{O} \in \mathscr{Y}, \text{ with } \mathcal{P}|_{\mathscr{Y}}\left(\mathcal{O}\right) \equiv 0 \tag{4.103}$$
$$\mathcal{B} \in \overline{\mathscr{B}}\left(\mathbb{R}^M\right) \iff \mathcal{B} \equiv \mathcal{C} \bigcup \mathcal{D} \,\middle|\, \mathcal{C} \in \mathscr{B}\left(\mathbb{R}^M\right) \text{ and } \mathcal{D} \subseteq \mathcal{O} \in \mathscr{B}\left(\mathbb{R}^M\right), \text{ with } \mathcal{P}_Y\left(\mathcal{O}\right) \equiv 0. \tag{4.104}$$

Now, specifically, to show that ϑ is measurable relative to $\overline{\mathscr{Y}}$, it suffices to show

that, for every Borel $\mathcal{A} \in \mathscr{B}\left(\overline{\mathbb{R}}\right)$,

$$\vartheta^{-1}(\mathcal{A}) \triangleq \{\omega \in \Omega | \, \vartheta(\omega) \in \mathcal{A}\} \in \overline{\mathscr{Y}}. \tag{4.105}$$

Recall, that, by definition of ξ, it is true that $\xi(Y(\omega)) \equiv \vartheta(\omega)$, for all $\omega \in \Omega$. Then, for every $\mathcal{A} \in \mathscr{B}\left(\overline{\mathbb{R}}\right)$, we may write

$$\begin{aligned} \vartheta^{-1}(\mathcal{A}) &\equiv \xi Y^{-1}(\mathcal{A}) \\ &\equiv \{\omega \in \Omega | \, \xi(Y(\omega)) \in \mathcal{A}\} \\ &\equiv \left\{\omega \in \Omega | \, Y(\omega) \in \xi^{-1}(\mathcal{A})\right\} \\ &\triangleq Y^{-1}\left(\xi^{-1}(\mathcal{A})\right). \end{aligned} \tag{4.106}$$

But $\xi^{-1}(\mathcal{A}) \in \overline{\mathscr{B}}\left(\mathbb{R}^M\right)$, which, by (4.104), equivalently means that $\xi^{-1}(\mathcal{A}) \equiv \mathcal{G}_{\mathcal{A}} \bigcup \mathcal{H}_{\mathcal{A}}$, for some $\mathcal{G}_{\mathcal{A}} \in \mathscr{B}\left(\mathbb{R}^M\right)$ and some $\mathcal{H}_{\mathcal{A}} \subseteq \mathcal{E}_{\mathcal{A}} \in \mathscr{B}\left(\mathbb{R}^M\right)$, with $\mathcal{P}_Y(\mathcal{E}_{\mathcal{A}}) \equiv 0$. Thus, we may further express any $\mathcal{A}$-preimage of ϑ as

$$\begin{aligned} \vartheta^{-1}(\mathcal{A}) &\equiv Y^{-1}\left(\mathcal{G}_{\mathcal{A}} \bigcup \mathcal{H}_{\mathcal{A}}\right) \\ &\equiv Y^{-1}(\mathcal{G}_{\mathcal{A}}) \bigcup Y^{-1}(\mathcal{H}_{\mathcal{A}}). \end{aligned} \tag{4.107}$$

Now, because $\mathcal{G}_{\mathcal{A}}$ is Borel and Y is a random element, it is true that $Y^{-1}(\mathcal{G}_{\mathcal{A}}) \in \mathscr{Y}$. On the other hand, $\mathcal{H}_{\mathcal{A}} \subseteq \mathcal{E}_{\mathcal{A}}$, which implies that $Y^{-1}(\mathcal{H}_{\mathcal{A}}) \subseteq Y^{-1}(\mathcal{E}_{\mathcal{A}})$, where

$$\mathcal{P}|_{\mathscr{Y}}\left(Y^{-1}(\mathcal{E}_{\mathcal{A}})\right) \equiv \mathcal{P}_Y(\mathcal{E}_{\mathcal{A}}) \equiv 0. \tag{4.108}$$

Therefore, we have shown that, for every $\mathcal{A} \in \mathscr{B}\left(\overline{\mathbb{R}}\right)$, $\vartheta^{-1}(\mathcal{A})$ may always be written as a union of an element in $\mathscr{Y}$ and some subset of a $\mathcal{P}|_{\mathscr{Y}}$-null set, also in $\mathscr{Y}$. Enough said. ★

Step 3. *For every $X \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, it is true that $h(Y, X) \geq \inf_{\boldsymbol{x} \in \mathcal{X}(Z_Y)} h(Y, \boldsymbol{x}) \equiv \vartheta$.*

For each $\omega \in \Omega$ (which also determines Y), we may write

$$\begin{aligned} \vartheta(\omega) &\equiv \inf_{\boldsymbol{x} \in \mathcal{X}(\mathcal{Z}(Y(\omega)))} h(Y(\omega), \boldsymbol{x}) \\ &\equiv \inf_{\mathcal{M}(Y(\omega)) \in \mathcal{X}(Z_Y(\omega))} h(Y(\omega), \mathcal{M}(Y(\omega))), \end{aligned} \tag{4.109}$$

where $\mathcal{M}:\mathbb{R}^M\to\mathbb{R}^N$ is of arbitrary nature. Therefore, ϑ may be equivalently regarded as the result of infimizing h over the set of all, *measurable or not*, functionals of Y, which are also selections of $\mathcal{X}(Z_Y)$. This set, of course, includes $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$. Now, choose an $X\equiv\mathcal{M}_X(Y)\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, as above, for some Borel measurable $\mathcal{M}_X:\mathbb{R}^M\to\mathbb{R}^N$. Then, it must be true that

$$\vartheta(\omega)\leq h\left(Y(\omega),\mathcal{M}_X(Y(\omega))\right)\equiv h\left(Y(\omega),X(\omega)\right), \tag{4.110}$$

everywhere in $\omega\in\Omega$. ★

Step 4. *It is also true that*

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{h(Y,X)\right\}\geq\mathbb{E}\left\{\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}h(Y,\boldsymbol{x})\right\}. \tag{4.111}$$

From **Step 3**, we know that, for every $X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, we have

$$h(Y,X)\geq\vartheta. \tag{4.112}$$

At this point, we exploit measurability of ϑ, proved in **Step 2**. Since, by assumption,

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{h(Y,X)\right\}\equiv\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{g(\cdot,X)\right\}<+\infty, \tag{4.113}$$

it follows that there exists $X_F\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, such that $\mathbb{E}\left\{h(Y,X_F)\right\}<+\infty$ (recall that the integral $\mathbb{E}\left\{g(\cdot,X_F)\right\}$ exists anyway, also by assumption). Since (4.112) holds for every $X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, it also holds for $X_F\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ and, consequently, the integral of ϑ exists, with $\mathbb{E}\left\{\vartheta\right\}<+\infty$. Then, we may take expectations on both sides of (4.112) (Theorem 1.5.9 (b) in [27]), yielding

$$\mathbb{E}\left\{h(Y,X)\right\}\geq\mathbb{E}\left\{\vartheta\right\},\quad\forall X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}. \tag{4.114}$$

Infimizing additionally both sides over $X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, we obtain the desired inequality.

We may also observe that, if $\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{h(Y,X)\right\}\equiv-\infty$, then

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{h(Y,X)\right\}\equiv\mathbb{E}\left\{\vartheta\right\}\equiv-\infty, \tag{4.115}$$

and the conclusion of Lemma 4.2 holds immediately. Therefore, in the following, we may assume that $\inf_{X \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}} \mathbb{E}\left\{h\left(Y, X\right)\right\} > -\infty$. ★

Step 5. *For every $\varepsilon > 0$, $n \in \mathbb{N}$ and every $\boldsymbol{y} \in \mathbb{R}^M$, there exists $\boldsymbol{x} \equiv \boldsymbol{x}_{\boldsymbol{y}} \in \mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)$, such that*

$$h\left(\boldsymbol{y}, \boldsymbol{x}_{\boldsymbol{y}}\right) \leq \max\left\{\xi\left(\boldsymbol{y}\right), -n\right\} + \varepsilon. \tag{4.116}$$

This simple fact may be shown by contradiction; replacing the universal with existential quantifiers and vice versa in the above statement, suppose that there exists $\varepsilon > 0$, $n \in \mathbb{N}$, and $\boldsymbol{y} \in \mathbb{R}^M$ such that, for all $\boldsymbol{x} \in \mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)$, $h\left(\boldsymbol{y}, \boldsymbol{x}\right) > \max\left\{\xi\left(\boldsymbol{y}\right), -n\right\} + \varepsilon$. There are two cases: **1)** $\xi\left(\boldsymbol{y}\right) > -\infty$. In this case, $\max\left\{\xi\left(\boldsymbol{y}\right), -n\right\} \geq \xi\left(\boldsymbol{y}\right)$, which would imply that, for all $\boldsymbol{x} \in \mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)$,

$$h\left(\boldsymbol{y}, \boldsymbol{x}\right) > \xi\left(\boldsymbol{y}\right) + \varepsilon, \tag{4.117}$$

contradicting the fact that $\xi\left(\boldsymbol{y}\right)$ is the infimum (the greatest lower bound) of $h\left(\boldsymbol{y}, \boldsymbol{x}\right)$ over $\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)$, since $\varepsilon > 0$. **2)** $\xi\left(\boldsymbol{y}\right) \equiv -\infty$. Here, $\max\left\{\xi\left(\boldsymbol{y}\right), -n\right\} \equiv -n$, and, for all $\boldsymbol{x} \in \mathcal{X}_{\mathcal{Z}}\left(\boldsymbol{y}\right)$, we would write

$$h\left(\boldsymbol{y}, \boldsymbol{x}\right) > -n + \varepsilon \in \mathbb{R}, \tag{4.118}$$

which, again, contradicts the fact that $-\infty \equiv \xi\left(\boldsymbol{y}\right)$ is the infimum of $h\left(\boldsymbol{y}, \boldsymbol{x}\right)$ over $\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)$. Therefore, in both cases, we are led to a contradiction, implying that the statement preceding and including (4.116) is true. The idea of using the maximum operator, so that $\xi\left(\boldsymbol{y}\right)$ may be allowed to take the value $-\infty$, is credited to and borrowed from ([84], proof of Theorem 14.60). ★

Step 6. *There exists a Borel measurable function $\widetilde{\xi} : \mathbb{R}^M \to \overline{\mathbb{R}}$, such that*

$$\widetilde{\xi}\left(\boldsymbol{y}\right) \equiv \xi\left(\boldsymbol{y}\right), \quad \forall \boldsymbol{y} \in \overline{\mathcal{R}}_{\xi} \supseteq \mathcal{R}_{\xi}, \tag{4.119}$$

where $\mathcal{R}_{\xi} \in \mathscr{B}\left(\mathbb{R}^M\right)$ is such that $\mathcal{P}_Y\left(\mathcal{R}_{\xi}\right) \equiv 1$, and $\overline{\mathcal{R}}_{\xi} \in \overline{\mathscr{B}}\left(\mathbb{R}^M\right)$ is such that $\overline{\mathcal{P}}_Y\left(\overline{\mathcal{R}}_{\xi}\right) \equiv 1$, where $\overline{\mathcal{P}}_Y$ denotes the completion of the pushforward $\mathcal{P}_Y$.

From ([95], Proposition 2.12), we know that, since ξ is $\overline{\mathscr{B}}\left(\mathbb{R}^M\right)$-measurable, there exists a $\mathscr{B}\left(\mathbb{R}^M\right)$-measurable function $\widetilde{\xi}:\mathbb{R}^M\rightarrow\overline{\mathbb{R}}$, such that

$$\widetilde{\xi}\left(\boldsymbol{y}\right)\equiv\xi\left(\boldsymbol{y}\right),\quad\forall\boldsymbol{y}\in\overline{\mathcal{R}}_{\xi},\tag{4.120}$$

where $\overline{\mathcal{R}}_{\xi}$ is an event in $\overline{\mathscr{B}}\left(\mathbb{R}^M\right)$, such that $\overline{\mathcal{P}}_Y\left(\overline{\mathcal{R}}_{\xi}\right)\equiv 1$. However, from **Step 2** (see (4.104)), we know that $\overline{\mathcal{R}}_{\xi}\equiv\mathcal{R}_{\xi}\bigcup\overline{\mathcal{R}}_{\xi}^{E}$, where $\mathcal{R}_{\xi}\in\mathscr{B}\left(\mathbb{R}^M\right)$ and $\overline{\mathcal{P}}_Y\left(\overline{\mathcal{R}}_{\xi}^{E}\right)\equiv 0$. Then, it may be easily shown that $\overline{\mathcal{P}}_Y\left(\overline{\mathcal{R}}_{\xi}\right)\equiv\overline{\mathcal{P}}_Y\left(\mathcal{R}_{\xi}\right)\equiv 1$ and, since $\overline{\mathcal{P}}_Y$ and $\mathcal{P}_Y$ agree on the elements of $\mathscr{B}\left(\mathbb{R}^M\right)$, $\mathcal{P}_Y\left(\mathcal{R}_{\xi}\right)\equiv 1$, as well. ★

Step 7. *There exists a $(\mathcal{P},\varepsilon,n)$-optimal selector $X_n^{\varepsilon}\in\mathcal{F}_{\mathcal{X}(Z_Y)}^{\mathscr{Y}}$: For every $\varepsilon>0$ and for every $n\in\mathbb{N}$, there exists $X_n^{\varepsilon}\in\mathcal{F}_{\mathcal{X}(Z_Y)}^{\mathscr{Y}}$, such that*

$$h\left(Y,X_n^{\varepsilon}\right)\leq\max\left\{\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}h\left(Y,\boldsymbol{x}\right),-n\right\}+\varepsilon,\quad\mathcal{P}-a.e..\tag{4.121}$$

This is the most crucial property of the problem that needs to be established, in order to reach to the final conclusions of Lemma 4.2. In this step, we make use of Theorem 4.7. Because Theorem 4.7 works on Borel spaces, in the following, it will be necessary to work directly on the state space of the random element Y, equipped with its Borel σ-algebra, and the pushforward $\mathcal{P}_Y$. In the following, we will also make use of the results proved in **Step 5** and **Step 6**.

Recall the definition of ξ in the statement of Lemma 4.2. We may readily show that the multifunction $\mathcal{X}\left(\mathcal{Z}\left(\cdot\right)\right)$ is $\mathscr{B}\left(\mathbb{R}^M\right)$-measurable. This may be shown in exactly the same way as in **Step 1**, exploitting the hypotheses that the multifunction $\mathcal{X}$ and the function $\mathcal{Z}$ are both Borel measurable. Borel measurability of $\mathcal{X}\left(\mathcal{Z}\left(\cdot\right)\right)$ will be exploited shortly.

Compare the result of **Step 5** with what we would like to prove here; the statement preceding and including (4.116) is not enough for our purposes; what we would actually like is to be able to generate a *selector*, that is, a function of $\boldsymbol{y}$ such that (4.116) would hold at least almost everywhere with respect to $\mathcal{P}_Y$. This is why we need Theorem 4.7. The idea of using Theorem 4.7 into this context is credited to and borrowed from [96].

From **Step 2** and **Step 6**, we know that ξ is $\overline{\mathscr{B}}\left(\mathbb{R}^M\right)$-measurable and that there exists a Borel measurable function $\widetilde{\xi}:\mathbb{R}^M\to\overline{\mathbb{R}}$, such that $\widetilde{\xi}\left(\boldsymbol{y}\right)\equiv\xi\left(\boldsymbol{y}\right)$, everywhere in $\boldsymbol{y}\in\mathcal{R}_\xi$, where $\mathcal{R}_\xi\in\mathscr{B}\left(\mathbb{R}^M\right)$ is such that $\mathcal{P}_Y\left(\mathcal{R}_\xi\right)\equiv\overline{\mathcal{P}}_Y\left(\mathcal{R}_\xi\right)\equiv 1$. Then, it follows that

$$\widetilde{\xi}\left(\boldsymbol{y}\right)\equiv\inf_{\boldsymbol{x}\in\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)}h\left(\boldsymbol{y},\boldsymbol{x}\right), \tag{4.122}$$

for all $\boldsymbol{y}\in\mathcal{R}_\xi$.

Define, for brevity, $\mathcal{X}_{\mathcal{Z}}\left(\boldsymbol{y}\right)\triangleq\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)$, for all $\boldsymbol{y}\in\mathbb{R}^M$. Towards the application of Theorem 4.7, fix any $\varepsilon>0$ and any $n\in\mathbb{N}$ and consider the set

$$\Pi_{\mathcal{X}_{\mathcal{Z}}}^{\varepsilon,n}\equiv\left\{\left.\begin{array}{c}\left(\boldsymbol{y},\boldsymbol{x}\right)\in\mathbb{R}^M\times\mathbb{R}^N\end{array}\right|\begin{array}{cc}\boldsymbol{x}\in\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right) & \\ h\left(\boldsymbol{y},\boldsymbol{x}\right)\leq\max\left\{\widetilde{\xi}\left(\boldsymbol{y}\right),-n\right\}+\varepsilon & ,\quad\text{if }\boldsymbol{y}\in\mathcal{R}_\xi \\ \boldsymbol{x}\in\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right), & \text{if }\boldsymbol{y}\in\mathcal{R}_\xi^c\end{array}\right\}. \tag{4.123}$$

We will show that Π_ε^n constitutes a measurable set in $\mathscr{B}\left(\mathbb{R}^M\right)\otimes\mathscr{B}\left(\mathbb{R}^N\right)$. Observe that $\Pi_{\mathcal{X}_{\mathcal{Z}}}^{\varepsilon,n}\equiv\Pi_{\mathcal{X}_{\mathcal{Z}}}\bigcap\left(\Pi^{\varepsilon,n}\bigcup\Pi_{rem}\right)$, where we define

$$\Pi_{\mathcal{X}_{\mathcal{Z}}}\triangleq\left\{\left.\left(\boldsymbol{y},\boldsymbol{x}\right)\in\mathbb{R}^M\times\mathbb{R}^N\right|\boldsymbol{x}\in\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)\right\}, \tag{4.124}$$

$$\Pi^{\varepsilon,n}\triangleq\left\{\left.\left(\boldsymbol{y},\boldsymbol{x}\right)\in\mathbb{R}^M\times\mathbb{R}^N\right|\boldsymbol{y}\in\mathcal{R}_\xi,h\left(\boldsymbol{y},\boldsymbol{x}\right)\leq\max\left\{\widetilde{\xi}\left(\boldsymbol{y}\right),-n\right\}+\varepsilon\right\}\text{ and} \tag{4.125}$$

$$\Pi_{rem}\triangleq\left\{\left.\left(\boldsymbol{y},\boldsymbol{x}\right)\in\mathbb{R}^M\times\mathbb{R}^N\right|\boldsymbol{y}\in\mathcal{R}_\xi^c\right\}. \tag{4.126}$$

Clearly, it suffices to show that both $\Pi_{\mathcal{X}_{\mathcal{Z}}}$ and $\Pi^{\varepsilon,n}$ are in $\mathscr{B}\left(\mathbb{R}^M\right)\otimes\mathscr{B}\left(\mathbb{R}^N\right)$. First, the set $\Pi_{\mathcal{X}_{\mathcal{Z}}}$ is the *graph* of the multifunction $\mathcal{X}_{\mathcal{Z}}$, and, because $\mathcal{X}_{\mathcal{Z}}$ is measurable, it follows from ([84], Theorem 14.8) that $\Pi_{\mathcal{X}_{\mathcal{Z}}}\in\mathscr{B}\left(\mathbb{R}^M\right)\otimes\mathscr{B}\left(\mathbb{R}^N\right)$. Second, because g is $\boldsymbol{SP}\diamondsuit\mathcal{F}_{\mathcal{X}\left(Z_Y\right)}^{\mathscr{Y}}$, h is jointly Borel measurable. Additionally, $\mathcal{R}_\xi$ is Borel and $\widetilde{\xi}$ is Borel as well. Consequently, $\Pi^{\varepsilon,n}$ can be written as the intersection of two measurable sets, implying that it is in $\mathscr{B}\left(\mathbb{R}^M\right)\otimes\mathscr{B}\left(\mathbb{R}^N\right)$, as well. And third, $\Pi_{rem}\in\mathscr{B}\left(\mathbb{R}^M\right)\otimes\mathscr{B}\left(\mathbb{R}^N\right)$, since $\mathcal{R}_\xi^c$ is Borel, as a complement of a Borel set. Therefore, $\Pi_{\mathcal{X}_{\mathcal{Z}}}^{\varepsilon,n}\in\mathscr{B}\left(\mathbb{R}^M\right)\otimes\mathscr{B}\left(\mathbb{R}^N\right)$.

Now, we have to verify the selection property, set as a requirement in the statement of Theorem 4.7. Indeed, for every $\boldsymbol{y}\in\mathcal{R}_\xi$, there exists $\boldsymbol{x}_{\boldsymbol{y}}\in\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)$, such that (4.116) holds, where $\xi\left(\boldsymbol{y}\right)\equiv\widetilde{\xi}\left(\boldsymbol{y}\right)$ (see **Step 6** and above), while, for every $\boldsymbol{y}\in\mathcal{R}_\xi^c$, any $\boldsymbol{x}_{\boldsymbol{y}}\in\mathcal{X}\left(\mathcal{Z}\left(\boldsymbol{y}\right)\right)$ will do. Thus, for every $\boldsymbol{y}\in\mathbb{R}^M$, there exists $\boldsymbol{x}_{\boldsymbol{y}}\in\mathbb{R}^N$, such that $\left(\boldsymbol{y},\boldsymbol{x}_{\boldsymbol{y}}\right)\in\Pi_{\mathcal{X}_{\mathcal{Z}}}^{\varepsilon,n}$. As a result, Theorem 4.7 applies and implies that there exists a Borel

subset $\mathcal{R}^{c}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}$ of $\mathcal{P}_Y$-measure 0, as well as a Borel measurable selector $\mathsf{S}^{\varepsilon}_{n}:\mathbb{R}^{M}\rightarrow\mathbb{R}^{N}$, such that, $(\boldsymbol{y},\mathsf{S}^{\varepsilon}_{n}(\boldsymbol{y}))\in\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}$, for all $\boldsymbol{y}\in\mathcal{R}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}$. In other words, the Borel measurable selector $\mathsf{S}^{\varepsilon}_{n}$ is such that

$$\begin{aligned}\mathsf{S}^{\varepsilon}_{n}(\boldsymbol{y})&\in\mathcal{X}(\mathcal{Z}(\boldsymbol{y}))\quad\text{and}\\ h(\boldsymbol{y},\mathsf{S}^{\varepsilon}_{n}(\boldsymbol{y}))&\leq\max\left\{\widetilde{\xi}(\boldsymbol{y}),-n\right\}+\varepsilon,\qquad\forall\boldsymbol{y}\in\mathcal{R}_{\xi}\bigcap\mathcal{R}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}\triangleq\mathcal{R}^{\xi}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}},\end{aligned}\tag{4.127}$$

where, of course, $\mathcal{P}_Y\left(\mathcal{R}_{\xi}\bigcap\mathcal{R}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}\right)\equiv 1$. Additionally, (4.127) must be true at $\boldsymbol{y}=Y(\omega)$, as long as ω is such that the values of Y are restricted to $\mathcal{R}^{\xi}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}$. Equivalently, we demand that

$$\omega\in\left\{\omega\in\Omega|\,Y(\omega)\in\mathcal{R}^{\xi}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}\right\}\equiv Y^{-1}\left(\mathcal{R}^{\xi}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}\right)\triangleq\Omega^{\varepsilon}_{n}.\tag{4.128}$$

But $\mathcal{R}^{\xi}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}\in\mathscr{B}\left(\mathbb{R}^{M}\right)$ and Y is a random element and, hence, a measurable function for Ω to $\mathbb{R}^{M}$. This means that $\Omega^{\varepsilon}_{n}\in\mathscr{Y}$ and we are allowed to write

$$\begin{aligned}\mathcal{P}(\Omega^{\varepsilon}_{n})&\equiv\int_{Y^{-1}\left(\mathcal{R}^{\xi}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}\right)}\mathcal{P}(\mathrm{d}\omega)\\ &=\int_{\left\{\boldsymbol{y}\in\mathbb{R}^{M}\left|\boldsymbol{y}\in\mathcal{R}^{\xi}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}\right.\right\}}\mathcal{P}_Y(\mathrm{d}\boldsymbol{y})\\ &\equiv\mathcal{P}_Y\left(\mathcal{R}^{\xi}_{\Pi^{\varepsilon,n}_{\mathcal{X}_{\mathcal{Z}}}}\right)\equiv 1.\end{aligned}\tag{4.129}$$

Therefore, we may pull (4.127) back to the base space, and restate it as

$$\begin{aligned}\mathsf{S}^{\varepsilon}_{n}(Y(\omega))&\in\mathcal{X}(Z_Y(\omega))\quad\text{and}\\ h(Y(\omega),\mathsf{S}^{\varepsilon}_{n}(Y(\omega)))&\leq\max\{\xi(Y(\omega)),-n\}+\varepsilon,\qquad\forall\omega\in\Omega^{\varepsilon}_{n},\end{aligned}\tag{4.130}$$

where $\Omega^{\varepsilon}_{n}\subseteq\Omega$ is an event, such that $\mathcal{P}(\Omega^{\varepsilon}_{n})\equiv 1$. Then, by construction, $\mathsf{S}^{\varepsilon}_{n}(Y)\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$. As a result, for any choice of $\varepsilon>0$ and $n\in\mathbb{N}$, the selector $X^{\varepsilon}_{n}\triangleq\mathsf{S}^{\varepsilon}_{n}(Y)\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ is such that

$$\mathbb{E}\{g(\cdot,X^{\varepsilon}_{n})|\,\mathscr{Y}\}(\omega)\leq\max\{\vartheta(\omega),-n\}+\varepsilon,\quad\mathcal{P}-a.e..\tag{4.131}$$

We are done. ★

Step 8. *It is true that*

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\{h(Y,X)\}\leq\mathbb{E}\left\{\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}h(Y,\boldsymbol{x})\right\}.\tag{4.132}$$

Define the sequence of random variables $\left\{\varpi_n : \Omega \to \overline{\mathbb{R}}\right\}_{n\in\mathbb{N}}$ as (see the RHS of (4.121))

$$\varpi_n(\omega) \triangleq \max\{\vartheta(\omega), -n\}, \quad \forall(\omega, n) \in \Omega \times \mathbb{N}. \tag{4.133}$$

Also, recall that $\mathbb{E}\{\vartheta\} < +\infty$. Additionally, observe that

$$\varpi_n(\omega) \leq \max\{\vartheta(\omega), 0\} \geq 0, \quad \forall(\omega, n) \in \Omega \times \mathbb{N}, \tag{4.134}$$

where it is easy to show that $\mathbb{E}\{\max\{\vartheta, 0\}\} < +\infty$. Thus, all members of $\{\varpi_n\}_{n\in\mathbb{N}}$ are bounded by an integrable random variable, everywhere in ω and uniformly in n, whereas it is trivial that, for every $\omega \in \Omega$, $\varpi_n(\omega) \underset{n\to\infty}{\searrow} \vartheta(\omega)$.

Consider now the result of **Step 7**, where we showed that, for every $\varepsilon > 0$ and for every $n \in \mathbb{N}$, there exists a selector $X_n^{\varepsilon} \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, such that

$$h(Y, X_n^{\varepsilon}) \leq \varpi_n + \varepsilon, \quad \mathcal{P} - a.e.. \tag{4.135}$$

We can then take expectations on both sides (note that all involved integrals exist), to obtain

$$\mathbb{E}\{h(Y, X_n^{\varepsilon})\} \leq \mathbb{E}\{\varpi_n\} + \varepsilon. \tag{4.136}$$

Since $X_n^{\varepsilon} \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, it also follows that

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}} \mathbb{E}\{h(Y, X)\} \leq \mathbb{E}\{\varpi_n\} + \varepsilon. \tag{4.137}$$

It is also easy to see that ϖ_n fulfills the requirements of the Extended Monotone Convergence Theorem ([27], Theorem 1.6.7 (b)). Therefore, we may pass to the limit on both sides of (4.137) as $n \to \infty$, yielding

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}} \mathbb{E}\{h(Y, X)\} \leq \mathbb{E}\{\vartheta\} + \varepsilon. \tag{4.138}$$

But $\varepsilon > 0$ is arbitrary. ★

Finally, just combine the statements of **Step 4** and **Step 8**, and the result follows, completing the proof of Lemma 4.2. ■

Remark 4.12. Obviously, Lemma 4.2 holds also for maximization problems as well, by defining $g \equiv -f$, for some random function $f : \Omega \times \mathbb{R}^N \to \mathbb{R}$, under the corresponding setting and assumptions. Note that, in this case, we have to assume that $\sup_{X \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}} \mathbb{E}\{f(\cdot, X)\} > -\infty$. ■

Remark 4.13. Lemma 4.2 may be considered a useful variation of Theorem 14.60 in [84], in the following sense. First, it is specialized for conditional expectations of random functions, which are additionally $\boldsymbol{SP}\diamondsuit\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, in the context of stochastic control. The latter property allows these conditional expectations to be expressed as (Borel) random functions themselves. This is in contrast to ([84], Theorem 14.60), where it is assumed that the random function, whose role is played by the respective conditional expectation in Lemma 4.2, is somehow provided apriori. Second, Lemma 4.2 extends ([84], Theorem 14.60), in the sense that the decision set $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ confines any solution to the respective optimization problem to be a $\mathscr{Y}$-measurable selection of a closed-valued measurable multifunction, while at the same time, apart from natural (and important) measurability requirements, no continuity assumptions are imposed on the structure of the random function induced by the respective conditional expectation; only the validity of the substitution property is required. In ([84], Theorem 14.60), on the other hand, it is respectively assumed that the involved random function constitutes a normal integrand, or, in other words, that it is random lower semicontinuous. ■

Remark 4.14. In Lemma 4.2, variational optimization is performed over some subset of functions *measurable relative to* $\mathscr{Y} \equiv \sigma\{Y\}$, where Y is some given random element. Although we do not pursue such an approach here, it would most probably be possible to develop a more general version of Lemma 4.2, where the decision set would be appropriately extended to include $\overline{\mathscr{Y}}$-measurable random elements, as well. In such case, the definition of the substitution property could be extended under the framework of lower semianalytic functions and universal measurability, and would allow the development of arguments showing existence of everywhere ε-optimal and potentially everywhere optimal policies (decisions), in the spirit of [86, 87]. ■

4.5.1.3 Guaranteeing the Existence of Measurable Optimal Controls

Although Lemma 4.2 constitutes a very useful result, which enables the simplification of a stochastic variational problem, by essentially replacing it by an at least structurally simpler, pointwise optimization problem, it does not provide insight on the existence of a common optimal solution, *within the respective decision sets*.

On the one hand, it is easy to observe that, similarly to ([84], Theorem 14.60), if there exists an optimal selection $X^* \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, such that

$$X^* \in \underset{\boldsymbol{x}\in\mathcal{X}(Z_Y)}{\arg\min}\, \mathbb{E}\left\{ g\left(\cdot, \boldsymbol{x}\right) \middle| \mathscr{Y} \right\} \neq \varnothing, \quad \mathcal{P}-a.e., \tag{4.139}$$

and Lemma 4.2 applies, then, exploitting the fact that g is $\boldsymbol{SP}\diamondsuit\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, we may write

$$\begin{aligned} \inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}} \mathbb{E}\left\{g\left(\cdot, X\right)\right\} &\equiv \mathbb{E}\left\{\vartheta\right\} \equiv \mathbb{E}\left\{\xi\left(Y\right)\right\} \\ &= \mathbb{E}\left\{h\left(Y, X^*\right)\right\} \\ &= \mathbb{E}\left\{\mathbb{E}\left\{ g\left(\cdot, X^*\right) \middle| \mathscr{Y} \right\}\right\} \\ &\equiv \mathbb{E}\left\{g\left(\cdot, X^*\right)\right\}, \end{aligned} \tag{4.140}$$

implying that the infimum of $\mathbb{E}\left\{g\left(\cdot, X\right)\right\}$ over $X \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ is attained by X^*; therefore, X^* is also an optimal solution to the respective variational problem. *Conversely*, if X^* attains the infimum of $\mathbb{E}\left\{g\left(\cdot, X\right)\right\}$ over $X \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ *and* the infimum is greater than $-\infty$, then both $\mathbb{E}\left\{g\left(\cdot, X^*\right)\right\}$ and $\mathbb{E}\left\{\vartheta\right\}$ are finite, which also implies that $\mathbb{E}\left\{ g\left(\cdot, X^*\right) \middle| \mathscr{Y} \right\}$ and ϑ are finite $\mathcal{P}-a.e.$. As a result, and recalling **Step 3** in the proof of Lemma 4.2, we have

$$\mathbb{E}\left\{\mathbb{E}\left\{ g\left(\cdot, X^*\right) \middle| \mathscr{Y} \right\} - \vartheta\right\} \equiv 0 \quad \text{and} \tag{4.141}$$

$$\mathbb{E}\left\{ g\left(\cdot, X^*\right) \middle| \mathscr{Y} \right\} - \vartheta \geq 0, \quad \mathcal{P}-a.e.. \tag{4.142}$$

and, consequently, $\vartheta \equiv \mathbb{E}\left\{ g\left(\cdot, X^*\right) \middle| \mathscr{Y} \right\}$, $\mathcal{P}-a.e.$.

Unfortunately, it is not possible to guarantee existence of such an $X^* \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, in general. However, at least for the purposes of this paper, it is both reasonable and desirable to demand the existence of an optimal solution X^*, satisfying (**??**) (in our spatially controlled beamforming problem, we *need* to make a *feasible* decision on the

position of the relays at the next time slot). Additionally, such an optimal solution, if it exists, will not be available in closed form, and, consequently, it will be impossible to verify measurability directly. Therefore, we have to be able to show both existence and measurability of X^* *indirectly*, and specifically, by imposing constraints on the structure of the stochastic optimization problem under consideration. One way to do this, *emphasizing on our spatially controlled beamforming problem formulation*, is to restrict our attention to pointwise optimization problems involving Carathéodory objectives, over closed-valued multifunctions, which are additionally *closed* -see Definition 4.5.

Focusing on Carathéodory functions is not particularly restrictive, since it is already clear that, in order to guarantee the validity of the substitution rule (the ***SP*** Property), similar continuity assumptions would have to be imposed on both random functions g and $\mathbb{E}\{g(\cdot,\cdot)|\mathscr{Y}\} \equiv h$, as Theorem 4.6 suggests. At the same time, restricting our attention to optimizing Carathéodory functions over measurable multifunctions, measurability of optimal values and optimal decisions is preserved, as the next theorem suggests.

Theorem 4.8. (Measurability under Partial Minimization) *On the base subspace* $(\Omega, \mathscr{Y}, \mathcal{P}|_{\mathscr{Y}})$*, where* $\mathscr{Y} \subseteq \mathscr{F}$*, let the random function* $H : \Omega \times \mathbb{R}^N \to \overline{\mathbb{R}}$ *be Carathéodory, and consider another random element* $Z : \Omega \to \mathbb{R}^N$*, as well as any compact-valued multifunction* $\mathcal{X} : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$*, with* $\mathrm{dom}(\mathcal{X}) \equiv \mathbb{R}^N$*, which is also* ***closed****. Additionally, define* $H^* : \Omega \to \overline{\mathbb{R}}$ *as the optimal value to the optimization problem*

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & H(\omega, \boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}(Z(\omega)) \end{array}, \quad \forall \omega \in \Omega. \tag{4.143}$$

Then, H^* *is* $\mathscr{Y}$*-measurable and attained for at least one* $\mathscr{Y}$*-measurable minimizer* $X^* : \Omega \to \mathbb{R}^N$*. If the minimizer* X^* *is unique, then it has to be* $\mathscr{Y}$*-measurable.*

Proof of Theorem 4.8. From ([85], pp. 365 - 367 and/or [84], Example 14.32 & Theorem 14.37), we may immediately deduce that H^* is $\mathscr{Y}$-measurable and attained for at least one $\mathscr{Y}$-measurable minimizer X^*, as long as the compact (therefore closed, as well)-valued multifunction $\mathcal{X}(Z(\cdot)) : \Omega \rightrightarrows \mathbb{R}^N$ is measurable relative to $\mathscr{Y}$. In order to show that the composition $\mathcal{X}(Z(\cdot))$ is $\mathscr{Y}$-measurable, we use the assumption that the

compact-valued multifunction $\mathcal{X} : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is closed and, therefore, Borel measurable (Remark 28 in [85], p. 365). Then, $\mathscr{Y}$-measurability of $\mathcal{X}(Z(\cdot))$ follows by the same arguments as in **Step 1**, in the proof of Lemma 4.2. ■

Remark 4.15. It would be important to mention that if one replaces $\mathbb{R}^N$ with any compact (say) subset $\mathcal{H} \subset \mathbb{R}^N$ in the statement of Theorem 4.8, then the result continues to hold as is. No modification is necessary. In our spatially controlled beamforming problem, this compact set $\mathcal{H}$ is specifically identified either with the hypercubic region $\mathcal{S}^R$, or with some compact subset of it. ■

4.5.1.4 Fusion & Derivation of Conditions C1-C6

Finally, combining Theorem 4.6, Lemma 4.2, Theorem 4.8 and Corollary **??**, we may directly formulate the following constrained version of the Fundamental Lemma, which is of central importance regarding the special class of stochastic problems considered in this work and, in particular, (4.5).

Lemma 4.3. (Fundamental Lemma / Fused Version) *On* $(\Omega, \mathscr{F}, \mathcal{P})$*, consider a random element* $Y : \Omega \to \mathbb{R}^M$*, the sub* σ*-algebra* $\mathscr{Y} \triangleq \sigma\{Y\} \subseteq \mathscr{F}$*, a random function* $g : \Omega \times \mathbb{R}^N \to \mathbb{R}$*, such that* $\mathbb{E}\{g(\cdot, \boldsymbol{x})\}$ *exists for all* $\boldsymbol{x} \in \mathbb{R}^N$*, a multifunction* $\mathcal{X} : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$*, with* $\mathrm{dom}(\mathcal{X}) \equiv \mathbb{R}^N$ *and , as well as another function* $Z_Y : \Omega \to \mathbb{R}^N$*. Assume that:*

C1. $\mathcal{X}$ *is compact-valued and closed, and that*

C2. Z_Y *is a* $\mathscr{Y}$*-measurable random element.*

Consider also the nonempty decision set $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$*. Additionally, suppose that:*

C3. $\mathbb{E}\{g(\cdot, X)\}$ *exists for all* $X \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$*, with* $\inf_{X \in \mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}} \mathbb{E}\{g(\cdot, X)\} < +\infty$*,*

C4. g *is dominated by a* $\mathcal{P}$*-integrable function, uniformly in* $\boldsymbol{x} \in \mathbb{R}^N$*,*

C5. g *is Carathéodory on* $\Omega \times \mathbb{R}^N$*, and that*

C6. $\mathbb{E}\{g(\cdot, \boldsymbol{x}) | \mathscr{Y}\} \equiv h(Y, \boldsymbol{x})$ *is Carathéodory on* $\Omega \times \mathbb{R}^N$*.*

Then, the optimal value function $\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)\middle|\mathscr{Y}\right\}\triangleq\vartheta$ *is* $\mathscr{Y}$*-measurable, and it is true that*

$$\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{g\left(\cdot,X\right)\right\}\equiv\mathbb{E}\left\{g\left(\cdot,X^*\right)\right\}\equiv\mathbb{E}\left\{\vartheta\right\},\tag{4.144}$$

for at least one $X^*\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$ *, such that* $X^*\left(\omega\right)\in\arg\min_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)\middle|\mathscr{Y}\right\}\left(\omega\right)$*, everywhere in* $\omega\in\Omega$*. If there is only one minimizer attaining* ϑ*, then it has to be* $\mathscr{Y}$*-measurable.*

Proof of Lemma 4.3. We just carefully combine Theorem 4.6, Lemma 4.2 and Theorem 4.8. First, if conditions **C4-C6** are satisfied, then, from Theorem 4.6, it follows that g is $\boldsymbol{SP}\diamondsuit\mathfrak{I}_{\mathscr{Y}}$. Then, since $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}\subseteq\mathfrak{I}_{\mathscr{Y}}$, g is $\boldsymbol{SP}\diamondsuit\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, as well. Consequently, with **C1-C3** being true, all assumptions of Lemma 4.2 are satisfied, and the first equivalence of (4.144) from the left is true. Additionally, from Theorem 4.8, it easily follows that the optimal value ϑ is $\mathscr{Y}$-measurable, attained by an at least one $\mathscr{Y}$-measurable X^*, which, of course, constitutes a selection of $\mathcal{X}\left(\mathcal{Z}\left(Y\right)\right)\equiv\mathcal{X}\left(Z_Y\right)$, or, equivalently, $X^*\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$. Then, because g is $\boldsymbol{SP}\diamondsuit\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, we may write

$$\begin{aligned}\vartheta&\equiv\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}\mathbb{E}\left\{g\left(\cdot,\boldsymbol{x}\right)\middle|\mathscr{Y}\right\}\\&\equiv\inf_{\boldsymbol{x}\in\mathcal{X}(Z_Y)}h\left(Y,\boldsymbol{x}\right)\\&\equiv h\left(Y,X^*\right)\\&\equiv\mathbb{E}\left\{g\left(\cdot,X^*\right)\middle|\mathscr{Y}\right\},\quad\mathcal{P}-a.e.,\end{aligned}\tag{4.145}$$

which yields the equivalence $\mathbb{E}\left\{\vartheta\right\}\equiv\mathbb{E}\left\{g\left(\cdot,X^*\right)\right\}$. The proof is complete. ■

Remark 4.16. Note that, because, in Lemma 4.3, $X^*\left(\omega\right)\in\mathcal{X}\left(Z_Y\left(\omega\right)\right)$, *everywhere* in $\omega\in\Omega$, it is true that X^* is actually a minimizer of the slightly more constrained problem of infimizing $\mathbb{E}\left\{g\left(\cdot,X\right)\right\}$ over the set of *precisely* all $\mathscr{Y}$-measurable selections of $\mathcal{X}\left(Z_Y\right)$. Denoting this decision set as $\mathcal{F}^{\mathscr{Y},E}_{\mathcal{X}(Z_Y)}\subseteq\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$, the aforementioned statement is true since, simply,

$$\inf_{X\in\mathcal{F}^{\mathscr{Y},E}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{g\left(\cdot,X\right)\right\}\geq\inf_{X\in\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}}\mathbb{E}\left\{g\left(\cdot,X\right)\right\}\equiv\mathbb{E}\left\{g\left(\cdot,X^*\right)\right\}\tag{4.146}$$

$$\Longrightarrow \inf_{X\in\mathcal{F}^{\mathscr{Y},E}_{\mathcal{X}(Z_Y)}} \mathbb{E}\left\{g\left(\cdot,X\right)\right\} \equiv \mathbb{E}\left\{g\left(\cdot,X^*\right)\right\}. \tag{4.147}$$

where we have used the fact that $X^* \in \mathcal{F}^{\mathscr{Y},E}_{\mathcal{X}(Z_Y)}$. This type of decision set is considered, for simplicity, in (4.5), which corresponds to the original formulation of the spatially controlled beamforming problem. ■

Lemma 4.3 is of major importance, as it directly provides us with conditions **C1-C6**, which, being relatively easily verifiable, at least for our spatially controlled beamforming setting, ensure strict theoretical consistency of the methods developed in this paper. At this point, our discussion about the Fundamental Lemma has been concluded. ■

4.5.2 Appendix B: Proofs / Section 4.2

4.5.2.1 Proof of Theorem 4.1

Since, in the following, we are going to verify conditions **C1-C6** of Lemma 4.3 in Section 4.5.1.4 (Appendix A) for the 2-stage problem (4.10), it will be useful to first match it to the setting of Lemma 4.3, term-by-term. Table 4.1 shows how the components of (4.10) are matched to the respective components of the optimization problem considered in Lemma 4.3. For the rest of the proof, we consider this variable matching automatic.

Keep $t \in \mathbb{N}^2_{N_T}$ *fixed*. As in the statement of Theorem 4.1, suppose that, at time slot $t-1 \in \mathbb{N}^+_{N_T-1}$, $\mathbf{p}^o\left(t-1\right) \equiv \mathbf{p}^o\left(\omega, t-1\right)$ is measurable relative to $\mathscr{C}\left(\mathcal{T}_{t-2}\right)$. Then, condition **C2** is automatically verified. ★

Next, let us verify **C1**. For this, we will simply show directly that closed-valued translated multifunctions, in the sense of Definition 4.1, are also closed. Given two *closed* sets $\mathcal{H} \subset \mathbb{R}^N$, $\mathcal{A} \subseteq \mathbb{R}^N$ and a fixed reference $\boldsymbol{h} \in \mathcal{H}$, let $\mathcal{D} : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be $(\mathcal{H}, \boldsymbol{h})$-translated in $\mathcal{A}$ and consider any two arbitrary sequences

$$\left\{\boldsymbol{x}_k \in \mathcal{A}\right\}_{k\in\mathbb{N}} \quad \text{and} \quad \left\{\boldsymbol{y}_k \in \mathcal{A}-\boldsymbol{h}\right\}_{k\in\mathbb{N}}, \tag{4.148}$$

such that $\boldsymbol{x}_k \underset{k\to\infty}{\longrightarrow} \boldsymbol{x}$, $\boldsymbol{y}_k \underset{k\to\infty}{\longrightarrow} \boldsymbol{y}$ and $\boldsymbol{x}_k \in \mathcal{D}\left(\boldsymbol{y}_k\right)$, for all $k \in \mathbb{N}$. By Definition 4.1, $\boldsymbol{x}_k \in \mathcal{D}\left(\boldsymbol{y}_k\right)$ if and only if $\boldsymbol{x}_k - \boldsymbol{y}_k \in \mathcal{H}$, for all $k \in \mathbb{N}$. But $\boldsymbol{x}_k - \boldsymbol{y}_k \underset{k\to\infty}{\longrightarrow} \boldsymbol{x} - \boldsymbol{y}$ and $\mathcal{H}$ is closed. Therefore, it is true that $\boldsymbol{x} - \boldsymbol{y} \in \mathcal{H}$, as well, showing that $\mathcal{D}$ is closed. By

Problem of Lemma 4.3	2-Stage Problem (4.10)
Random element $Y : \Omega \to \mathbb{R}^M$	*All* relay positions *and* channel observations, up to (current) time slot $t-1$
σ-Algebra $\mathscr{Y} \triangleq \sigma\{Y\}$	σ-Algebra $\mathscr{C}(\mathcal{T}_{t-1})$, jointly generated by the above random vector
Random Function $g : \Omega \times \mathbb{R}^N \to \mathbb{R}$	Optimal value of the second-stage problem, $V(\cdot,\cdot,t-1) : \Omega \times \mathcal{S}^R \to \mathbb{R}_{++}$
Multifunction $\mathcal{X} : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$, with $\mathrm{dom}(\mathcal{X}) \equiv \mathbb{R}^N$	Spatially feasible motion region $\mathcal{C} : \mathcal{S}^R \rightrightarrows \mathcal{S}^R$, with $\mathrm{dom}(\mathcal{C}) \equiv \mathcal{S}^R$
Function $Z_Y : \Omega \to \mathbb{R}^N$	*Selected* motion policy at time slot $t-2$, $\mathbf{p}^o(\cdot, t-1) : \Omega \to \mathcal{S}^R$
Decision set $\mathcal{F}^{\mathscr{Y}}_{\mathcal{X}(Z_Y)}$	Decision set $\mathcal{D}_t$ (precisely matched with $\mathcal{F}^{\mathscr{Y},E}_{\mathcal{X}(Z_Y)}$)

Table 4.1: Variable matching for (4.10) and the respective problem considered in Lemma 4.3.

Assumption 4.1, $\mathcal{C} : \mathbb{R}^{2R} \rightrightarrows \mathbb{R}^{2R}$ is the $(\mathcal{G}, \mathbf{0})$-translated multifunction in $\mathcal{S}^R$, for some compact and, hence, closed, $\mathcal{G} \subset \mathcal{S}^R$. Consequently, the restriction of $\mathcal{C}$ in $\mathcal{S}^R$ is closed and **C3** is verified. ★

Condition **C5** is also easily verified; it suffices to show that both functions $|f(\cdot,\cdot,t)|^2$ and $|g(\cdot,\cdot,t)|^2$ are Carathéodory on $\Omega \times \mathcal{S}$, or, in other words, that the fields $|f(\mathbf{p},t)|^2$ and $|g(\mathbf{p},t)|^2$ are everywhere sample path continuous. Indeed, if this holds, $V_I(\cdot,\cdot,t)$ will be Carathéodory, as a continuous functional of $|f(\cdot,\cdot,t)|^2$ and $|g(\cdot,\cdot,t)|^2$, and since

$$V\left(\left[\mathbf{p}_1^{\boldsymbol{T}} \ \ldots \ \mathbf{p}_R^{\boldsymbol{T}}\right]^{\boldsymbol{T}}, t\right) \equiv \sum_{i \in \mathbb{N}_R^+} V_I(\mathbf{p}_i, t), \tag{4.149}$$

it readily follows that $V(\cdot,\cdot,t)$ is Carathéodory on $\Omega \times \mathcal{S}^R$. In order to show (everywhere) sample path continuity of $|f(\mathbf{p},t)|^2$ (respectively $|g(\mathbf{p},t)|^2$) on $\mathcal{S}$, we may utilize (3.11). As a result, sample path continuity of $|f(\mathbf{p},t)|^2$ is equivalent to sample path continuity of

$$F(\mathbf{p},t) \equiv \alpha_S(\mathbf{p})\,\ell + \sigma_S(\mathbf{p},t) + \xi_S(\mathbf{p},t), \quad \forall \mathbf{p} \in \mathcal{S}. \tag{4.150}$$

Of course, α_S is a continuous function of $\mathbf{p}$. As long as the fields $\sigma_S(\mathbf{p},t)$ and $\xi_S(\mathbf{p},t)$ are concerned, these are also sample path continuous; see Section 3.1.3. Enough said. ★

We continue with **C3**. Since we already know that $V(\cdot,\cdot,t)$ is Carathéodory, it follows from ([92], Lemma 4.51) that $V(\cdot,\cdot,t)$ is also jointly measurable relative to

$\mathscr{F} \otimes \mathscr{B}\left(\mathcal{S}^R\right)$. Next, let $\mathbf{p}(t) \equiv \mathbf{p}(\omega, t) \in \mathcal{S}^R$ be *any* random element, measurable with respect to $\mathscr{C}(\mathcal{T}_{t-1})$ and, thus, $\mathscr{F}$, too. Then, from ([92], Lemma 4.49), we know that the pair $(\mathbf{p}(t,\omega),(t,\omega))$ is also $\mathscr{F}$-measurable. Consequently, $|V(\cdot,\mathbf{p}(\cdot,t),t)|^2$ must be $\mathscr{F}$-measurable, as a composition of measurable functions. Additionally, $V(\cdot,\cdot,t)$ is, by definition, nonnegative. Thus, its expectation exists (Corollary 1.6.4 in [27]), and we are done. ★

Conditions **C4** and **C6** need slightly more work, in order to be established. To verify **C4**, we have to show existence of a function in $\mathcal{L}_1(\Omega, \mathscr{F}, \mathcal{P}; \mathbb{R})$, which dominates $V(\cdot,\cdot,t)$, *uniformly* in $\mathbf{p} \in \mathcal{S}^R$. Everywhere in Ω, again using (3.11), and with $c \triangleq \log(10)/10$ for brevity, we may write

$$
\begin{aligned}
V\left(\left[\mathbf{p}_1^{\boldsymbol{T}} \ \ldots \ \mathbf{p}_R^{\boldsymbol{T}}\right]^{\boldsymbol{T}}, t\right) &\equiv \sum_{i \in \mathbb{N}_R^+} \frac{P_c P_0 \left|f(\mathbf{p}_i,t)\right|^2 \left|g(\mathbf{p}_i,t)\right|^2}{P_0 \sigma_D^2 \left|f(\mathbf{p}_i,t)\right|^2 + P_c \sigma^2 \left|g(\mathbf{p}_i,t)\right|^2 + \sigma^2 \sigma_D^2} \\
&\leq \frac{P_0}{\sigma^2} \sum_{i \in \mathbb{N}_R^+} \left|f(\mathbf{p}_i,t)\right|^2 \\
&\leq \frac{P_0}{\sigma^2} \sum_{i \in \mathbb{N}_R^+} \sup_{\mathbf{p}_i \in \mathcal{S}} \left|f(\mathbf{p}_i,t)\right|^2 \\
&\equiv \frac{10^{\rho/10} P_0 R}{\sigma^2} \sup_{\mathbf{p} \in \mathcal{S}} \exp\left(cF(\mathbf{p},t)\right) \\
&\equiv \frac{10^{\rho/10} P_0 R}{\sigma^2} \exp\left(c \sup_{\mathbf{p} \in \mathcal{S}} F(\mathbf{p},t)\right) \\
&\equiv \frac{10^{\rho/10} P_0 R}{\sigma^2} \exp\left(c \sup_{\mathbf{p} \in \mathcal{S}} \alpha_S(\mathbf{p})\, \ell + \sigma_S(\mathbf{p},t) + \xi_S(\mathbf{p},t)\right) \\
&\triangleq \frac{10^{\rho/10} P_0 R}{\sigma^2} \exp\left(c \sup_{\mathbf{p} \in \mathcal{S}} \alpha_S(\mathbf{p})\, \ell + \chi_S(\mathbf{p},t)\right) \\
&\leq \frac{10^{\rho/10} P_0 R}{\sigma^2} \exp\left(c\ell \sup_{\mathbf{p} \in \mathcal{S}} \alpha_S(\mathbf{p})\right) \exp\left(c \sup_{\mathbf{p} \in \mathcal{S}} \chi_S(\mathbf{p},t)\right) \\
&\triangleq \varphi(\omega,t) > 0, \quad \forall \omega \in \Omega. \qquad (4.151)
\end{aligned}
$$

Due to the fact that α_S is continuous in $\mathbf{p} \in \mathcal{S}$ and that $\mathcal{S}$ is compact, the Extreme Value Theorem implies that the deterministic term $\sup_{\mathbf{p} \in \mathcal{S}} \alpha_S(\mathbf{p})$ is finite. Consequently, it suffices to show that

$$
\mathbb{E}\left\{\exp\left(c \sup_{\mathbf{p} \in \mathcal{S}} \chi_S(\mathbf{p},t)\right)\right\} < +\infty, \qquad (4.152)
$$

provided, of course, that the expectation is meaningfully defined. For this to happen, it suffices that the function $\sup_{\mathbf{p}\in\mathcal{S}} \chi_S(\mathbf{p},t)$ is a well defined random variable. Since both $\sigma_S(\mathbf{p},t)$ and $\xi_S(\mathbf{p},t)$ are sample path continuous, it follows that the sum field $\sigma_S(\mathbf{p},t)+\xi_S(\mathbf{p},t)$ is sample path continuous. It is then relatively easy to see that $\sup_{\mathbf{p}\in\mathcal{S}} \chi_S(\mathbf{p},t)$ is a measurable function. See, for instance, Corollary **??**, or [64]. Additionally, the Extreme Value Theorem again implies that $\sup_{\mathbf{p}\in\mathcal{S}} \chi_S(\mathbf{p},t)$ is finite *everywhere* on Ω, which in turn means that the field $\chi_S(\mathbf{p},t)$ is *at least* almost everywhere bounded on the compact set $\mathcal{S}$.

Now, in order to prove that (4.152) is indeed true, we will invoke a well known result from the theory of concentration of measure, the *Borell-TIS Inequality*, which now follows.

Theorem 4.9. (Borell-TIS Inequality [64]) *Let $X(\boldsymbol{s})$, $\boldsymbol{s}\in\mathbb{R}^N$, be a real-valued, zero-mean, Gaussian random field, $\mathcal{P}$-almost everywhere bounded on a compact subset $\mathcal{K}\subset\mathbb{R}^N$. Then, it is true that*

$$\mathbb{E}\left\{\sup_{\boldsymbol{s}\in\mathcal{K}} X(\boldsymbol{s})\right\} < +\infty \quad \textit{and} \tag{4.153}$$

$$\mathcal{P}\left(\sup_{\boldsymbol{s}\in\mathcal{K}} X(\boldsymbol{s}) - \mathbb{E}\left\{\sup_{\boldsymbol{s}\in\mathcal{K}} X(\boldsymbol{s})\right\} > u\right) \leq \exp\left(-\frac{u^2}{2\sup\limits_{\boldsymbol{s}\in\mathcal{K}}\mathbb{E}\left\{X^2(\boldsymbol{s})\right\}}\right), \tag{4.154}$$

for all $u>0$.

As highlighted in ([64], page 50), an immediate consequence of the Borell-TIS Inequality is that, under the setting of Theorem 4.9, we may further assert that

$$\mathcal{P}\left(\sup_{\boldsymbol{s}\in\mathcal{K}} X(\boldsymbol{s}) > u\right) \leq \exp\left(-\frac{\left(u-\mathbb{E}\left\{\sup\limits_{\boldsymbol{s}\in\mathcal{K}} X(\boldsymbol{s})\right\}\right)^2}{2\sup\limits_{\boldsymbol{s}\in\mathcal{K}}\mathbb{E}\left\{X^2(\boldsymbol{s})\right\}}\right), \tag{4.155}$$

for all $u > \mathbb{E}\{\sup_{\boldsymbol{s}\in\mathcal{K}} X(\boldsymbol{s})\}$.

To show (4.152), we exploit the Borell-TIS Inequality and follow a procedure similar to ([64], Theorem 2.1.2). First, from the discussion above, we readily see that the field $\chi_S(\mathbf{p},t)$ does satisfy the assumptions Theorem 4.9. Also, because $\chi_S(\mathbf{p},t)$ is the sum of two independent fields, it is true that

$$\mathbb{E}\left\{\chi_S^2(\mathbf{p},t)\right\} \equiv \eta^2 + \sigma_\xi^2. \tag{4.156}$$

As a result, Theorem 4.9 implies that $\mathbb{E}\left\{\sup_{\boldsymbol{s}\in\mathcal{K}} X(\boldsymbol{s})\right\}$ is finite and we may safely write

$$\begin{aligned}\mathbb{E}\left\{\exp\left(c\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right)\right\} &\equiv \int_0^\infty \mathcal{P}\left(\exp\left(c\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right)>x\right)\mathrm{d}x\\ &\equiv \int_0^\infty \mathcal{P}\left(\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)>\frac{\log(x)}{c}\right)\mathrm{d}x. \end{aligned} \tag{4.157}$$

In order to exploit (4.155), it must hold that

$$\frac{\log(x)}{c}>\mathbb{E}\left\{\sup_{\boldsymbol{s}\in\mathcal{K}}X(\boldsymbol{s})\right\}\Leftrightarrow x>\exp\left(c\mathbb{E}\left\{\sup_{\boldsymbol{s}\in\mathcal{K}}X(\boldsymbol{s})\right\}\right)>0. \tag{4.158}$$

Therefore, we may break (4.157) into two parts and bound from above, namely,

$$\begin{aligned}&\mathbb{E}\left\{\exp\left(c\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right)\right\}\\ &\quad\equiv\int_0^{\exp\left(c\mathbb{E}\left\{\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right\}\right)}\mathcal{P}\left(\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)>\frac{\log(x)}{c}\right)\mathrm{d}x\\ &\qquad+\int_{\exp\left(c\mathbb{E}\left\{\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right\}\right)}^\infty\mathcal{P}\left(\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)>\frac{\log(x)}{c}\right)\mathrm{d}x\\ &\quad\le\int_0^{\exp\left(c\mathbb{E}\left\{\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right\}\right)}\mathrm{d}x\\ &\qquad+\int_{\exp\left(c\mathbb{E}\left\{\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right\}\right)}^\infty\exp\left(-\frac{\left(\frac{\log(x)}{c}-\mathbb{E}\left\{\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right\}\right)^2}{2\left(\eta^2+\sigma_\xi^2\right)}\right)\mathrm{d}x\\ &\quad\le\exp\left(c\mathbb{E}\left\{\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right\}\right)\\ &\qquad+c\int_{\mathbb{E}\left\{\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right\}}^\infty\exp(cu)\exp\left(-\frac{\left(u-\mathbb{E}\left\{\sup_{\mathbf{p}\in\mathcal{S}}\chi_S(\mathbf{p},t)\right\}\right)^2}{2\left(\eta^2+\sigma_\xi^2\right)}\right)\mathrm{d}u. \end{aligned} \tag{4.159}$$

Since both terms on the RHS of (4.159) are finite, (4.152) is indeed satisfied. Consequently, it is true that

$$\mathbb{E}\{\varphi(\cdot,t)\}<+\infty\Leftrightarrow\varphi(\cdot,t)\in\mathcal{L}_1(\Omega,\mathscr{F},\mathcal{P};\mathbb{R}). \tag{4.160}$$

Enough said; **C4** is now verified. ★

Moving on to **C6**, the goal here is to show that, for each fixed $t \in \mathbb{N}_{N_T}^2$, the well defined random function $H : \Omega \times \mathcal{S}^R \to \overline{\mathbb{R}}$, defined as

$$H(\omega, \mathbf{p}) \triangleq \mathbb{E}\left\{V(\mathbf{p}, t) \,|\mathscr{C}(\mathcal{T}_{t-1})\right\}(\omega), \tag{4.161}$$

is Carathéodory. Observe, though, that we may write

$$H(\omega, \mathbf{p}) \equiv \sum_{i \in \mathbb{N}_R^+} H_I(\omega, \mathbf{p}_i), \tag{4.162}$$

where the random function $H_I : \Omega \times \mathcal{S} \to \mathbb{R}$ is defined as

$$H_I(\omega, \mathbf{p}) \triangleq \mathbb{E}\left\{ \frac{P_c P_0 |f(\mathbf{p}, t)|^2 |g(\mathbf{p}, t)|^2}{P_0 \sigma_D^2 |f(\mathbf{p}, t)|^2 + P_c \sigma^2 |g(\mathbf{p}, t)|^2 + \sigma^2 \sigma_D^2} \middle| \mathscr{C}(\mathcal{T}_{t-1}) \right\}(\omega). \tag{4.163}$$

Because a finite sum of Carathéodory functions (in this case, in different variables) is obviously Carathéodory, it suffices to show that H_I is Carathéodory.

First, it is easy to see that $H_I(\cdot, \mathbf{p})$ constitutes a well defined conditional expectation of a nonnegative random variable, for all $\mathbf{p} \in \mathcal{S}$. Therefore, what remains is to show that $H_I(\omega, \cdot)$ is continuous on $\mathcal{S}$, everywhere with respect to $\omega \in \Omega$. For this, we will rely on the sequential definition of continuity and the explicit representation of H_I as an integral with respect to the Lebesgue measure, which exploits the form of the projective system of finite dimensional distributions of $|f(\mathbf{p}, t)|^2$ and $|g(\mathbf{p}, t)|^2$. In particular, because of the trick (3.11), it is easy to show that H_I can be equivalently expressed as the Lebesgue integral

$$H_I(\omega, \mathbf{p}) = \int_{\mathbb{R}^2} r(\boldsymbol{x}) \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_2(\omega, \mathbf{p}), \boldsymbol{\Sigma}_2(\omega, \mathbf{p})) \, \mathrm{d}\boldsymbol{x}, \tag{4.164}$$

where the continuous function $r : \mathbb{R}^2 \to \mathbb{R}_{++}$ is defined as

$$r(\boldsymbol{x}) \equiv r(x_1, x_2) \triangleq \frac{P_c P_0 10^{\rho/10} \left[\exp(x_1 + x_2)\right]^{\frac{\log(10)}{10}}}{P_0 \sigma_D^2 \left[\exp(x_1)\right]^{\frac{\log(10)}{10}} + P_c \sigma^2 \left[\exp(x_2)\right]^{\frac{\log(10)}{10}} + 10^{-\rho/10} \sigma^2 \sigma_D^2}, \tag{4.165}$$

for all $\boldsymbol{x} \equiv (x_1, x_2) \in \mathbb{R}^2$, and $\mathcal{N} : \mathbb{R}^2 \times \mathcal{S} \times \Omega \to \mathbb{R}_{++}$, corresponds to the *jointly Gaussian* conditional density of $F(\mathbf{p}, t)$ and $G(\mathbf{p}, t)$, relative to $\mathscr{C}(\mathcal{T}_{t-1})$, with mean $\boldsymbol{\mu}_2 : (\omega, \mathbf{p}) \to \mathbb{R}^{2\times 1}$ and covariance $\boldsymbol{\Sigma}_2 : (\omega, \mathbf{p}) \to \mathbb{S}_{++}^{2\times 2}$ explicitly depending on ω and $\mathbf{p}$ as

$$\boldsymbol{\mu}_2(\omega, \mathbf{p}) \equiv \boldsymbol{\mu}_2(\mathscr{C}(\mathcal{T}_{t-1})(\omega); \mathbf{p}) \quad \text{and} \tag{4.166}$$

$$\boldsymbol{\Sigma}_2(\omega,\mathbf{p}) \equiv \boldsymbol{\Sigma}_2\left(\mathscr{C}\left(\mathcal{T}_{t-1}\right)(\omega);\mathbf{p}\right), \quad \forall(\omega,\mathbf{p}) \in \Omega\times\mathcal{S}. \tag{4.167}$$

Via a simple change of variables, we may reexpress $H_I(\omega,\mathbf{p})$ as

$$H_I(\omega,\mathbf{p}) \equiv \int_{\mathbb{R}^2} r\left(\boldsymbol{x}+\boldsymbol{\mu}_2(\omega,\mathbf{p})\right)\mathcal{N}\left(\boldsymbol{x};\mathbf{0},\boldsymbol{\Sigma}_2(\omega,\mathbf{p})\right)\mathrm{d}\boldsymbol{x}. \tag{4.168}$$

It is straightforward to verify that both $\boldsymbol{\mu}_2(\omega,\cdot)$ and $\boldsymbol{\Sigma}_2(\omega,\cdot)$ are continuous functions in $\mathbf{p}\in\mathcal{S}$, for all $\omega\in\Omega$. This is due to the fact that all functions involving $\mathbf{p}$ in the wireless channel model introduced in Section **??** are trivially continuous in this variable. Equivalently, we may assert that the whole integrand $r\left(\boldsymbol{x}+\boldsymbol{\mu}_2(\omega,\cdot)\right)\mathcal{N}\left(\boldsymbol{x};\mathbf{0},\boldsymbol{\Sigma}_2(\omega,\cdot)\right)$ is a continuous function, for all pairs $(\omega,\boldsymbol{x})\in\Omega\times\mathbb{R}^2$. Next, fix $\omega\in\Omega$, and for *arbitrary* $\mathbf{p}\in\mathcal{S}$, consider *any* sequence $\{\mathbf{p}_k\in\mathcal{S}\}_{k\in\mathbb{N}}$, such that $\mathbf{p}_k \underset{k\to\infty}{\longrightarrow} \mathbf{p}$. Then, $H_I(\omega,\cdot)$ is continuous if and only if $H_I(\omega,\mathbf{p}_k) \underset{k\to\infty}{\longrightarrow} H_I(\omega,\mathbf{p})$. We will show this via a simple application of the Dominated Convergence Theorem. Emphasizing the dependence on $\mathbf{p}$ as a superscript for the sake of clarity, we can write

$$\begin{aligned}
&r\left(\boldsymbol{x}+\boldsymbol{\mu}_2^{\mathbf{p}}\right)\mathcal{N}\left(\boldsymbol{x};\mathbf{0},\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)\\
&\quad\leq r\left(\boldsymbol{x}+\boldsymbol{\mu}_2^{\mathbf{p}}\right)\mathcal{N}\left(\boldsymbol{x};\mathbf{0},\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)\\
&\quad\equiv r\left(\boldsymbol{x}+\boldsymbol{\mu}_2^{\mathbf{p}}\right)\frac{\exp\left(-\frac{1}{2}\boldsymbol{x}^{\boldsymbol{T}}\left[\boldsymbol{\Sigma}_2^{\mathbf{p}}\right]^{-1}\boldsymbol{x}\right)}{2\pi\sqrt{\det\left(\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)}}\\
&\quad\leq r\left(\boldsymbol{x}+\boldsymbol{\mu}_2^{\mathbf{p}}\right)\frac{\exp\left(-\frac{1}{2}\lambda_{min}\left(\left[\boldsymbol{\Sigma}_2^{\mathbf{p}}\right]^{-1}\right)\|\boldsymbol{x}\|_2^2\right)}{2\pi\sqrt{\det\left(\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)}}\\
&\quad\equiv r\left(\boldsymbol{x}+\boldsymbol{\mu}_2^{\mathbf{p}}\right)\frac{\exp\left(-\frac{\|\boldsymbol{x}\|_2^2}{2\lambda_{max}\left(\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)}\right)}{2\pi\sqrt{\det\left(\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)}}\\
&\quad\leq \frac{P_0 10^{\rho/10}}{\sigma^2}\left[\exp\left(x_1+\boldsymbol{\mu}_2^{\mathbf{p}}(1)\right)\right]^{\frac{\log(10)}{10}}\frac{\exp\left(-\frac{\|\boldsymbol{x}\|_2^2}{2\lambda_{max}\left(\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)}\right)}{2\pi\sqrt{\det\left(\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)}}\\
&\quad\leq \frac{P_0 10^{\rho/10}}{\sigma^2}\left[\exp\left(x_1+\sup_{\mathbf{p}\in\mathcal{S}}\boldsymbol{\mu}_2^{\mathbf{p}}(1)\right)\right]^{\frac{\log(10)}{10}}\frac{\exp\left(-\frac{\|\boldsymbol{x}\|_2^2}{2\sup\limits_{\mathbf{p}\in\mathcal{S}}\lambda_{max}\left(\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)}\right)}{2\pi\sqrt{\inf\limits_{\mathbf{p}\in\mathcal{S}}\det\left(\boldsymbol{\Sigma}_2^{\mathbf{p}}\right)}}
\end{aligned}$$

$$\triangleq \frac{P_0 10^{\rho/10}}{\sigma^2}\left[\exp(x_1+p_1)\right]^{\frac{\log(10)}{10}}\frac{\exp\left(-\frac{\|\boldsymbol{x}\|_2^2}{2p_2}\right)}{2\pi\sqrt{p_3}}$$

$$\triangleq \psi(\omega,\boldsymbol{x}), \tag{4.169}$$

where, due to the continuity of $\boldsymbol{\mu}_2(\omega,\cdot)$ and $\boldsymbol{\Sigma}_2(\omega,\cdot)$, the continuity of the maximum eigenvalue and determinant operators, the fact that $\mathcal{S}$ is compact, and the power of the Extreme Value Theorem, all extrema involved are finite and, of course, independent of $\mathbf{p}$. It is now easy to verify that the RHS of (4.169) is integrable. Indeed, by Fubini's Theorem (Theorem 2.6.4 in [27])

$$\begin{aligned}
&\int_{\mathbb{R}^2}\psi(\omega,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&=\frac{P_0 10^{\rho/10}}{\sigma^2}\frac{\exp\left(\frac{\log(10)}{10}p_1\right)}{\sqrt{p_3}}\int_{\mathbb{R}^2}\exp\left(\frac{\log(10)}{10}x_1\right)\frac{1}{2\pi}\exp\left(-\frac{\|\boldsymbol{x}\|_2^2}{2p_2}\right)\mathrm{d}\boldsymbol{x}\\
&\equiv\frac{P_0 10^{\rho/10}}{\sigma^2}\frac{\exp\left(\frac{\log(10)}{10}p_1\right)p_2}{\sqrt{p_3}}\int_{\mathbb{R}^2}\exp\left(\frac{\log(10)}{10}x_1\right)\frac{1}{2\pi p_2}\exp\left(-\frac{\|\boldsymbol{x}\|_2^2}{2p_2}\right)\mathrm{d}\boldsymbol{x}\\
&=\frac{P_0 10^{\rho/10}}{\sigma^2}\frac{\exp\left(\frac{\log(10)}{10}p_1\right)p_2}{\sqrt{p_3}}\int_{\mathbb{R}}\exp\left(\frac{\log(10)}{10}x_1\right)\frac{1}{\sqrt{2\pi p_2}}\exp\left(-\frac{x_1^2}{2p_2}\right)\mathrm{d}x_1\\
&=\frac{P_0 10^{\rho/10}}{\sigma^2}\frac{\exp\left(\frac{\log(10)}{10}p_1(\omega)\right)p_2(\omega)}{\sqrt{p_3(\omega)}}\exp\left(\frac{p_2(\omega)}{2}\left(\frac{\log(10)}{10}\right)^2\right)<+\infty,
\end{aligned}\tag{4.170}$$

for $\omega\in\Omega$. That is,

$$\psi(\omega,\cdot)\in\mathcal{L}_1\left(\mathbb{R}^2,\mathscr{B}\left(\mathbb{R}^2\right),\mathcal{L};\mathbb{R}\right),\quad \omega\in\Omega, \tag{4.171}$$

where $\mathcal{L}$ denotes the Lebesgue measure. We can now call Dominated Convergence; since, for each $\boldsymbol{x}\in\mathbb{R}^2$ (and each $\omega\in\Omega$),

$$r(\boldsymbol{x}+\boldsymbol{\mu}_2(\omega,\mathbf{p}_k))\,\mathcal{N}(\boldsymbol{x};\mathbf{0},\boldsymbol{\Sigma}_2(\omega,\mathbf{p}_k))\underset{k\to\infty}{\longrightarrow} r(\boldsymbol{x}+\boldsymbol{\mu}_2(\omega,\mathbf{p}))\,\mathcal{N}(\boldsymbol{x};\mathbf{0},\boldsymbol{\Sigma}_2(\omega,\mathbf{p})) \tag{4.172}$$

and all members of this sequence are dominated by the integrable function $\psi(\omega,\cdot)$, it is true that

$$H_I(\omega,\mathbf{p}_k)\equiv\int_{\mathbb{R}^2} r(\boldsymbol{x}+\boldsymbol{\mu}_2(\omega,\mathbf{p}_k))\,\mathcal{N}(\boldsymbol{x};\mathbf{0},\boldsymbol{\Sigma}_2(\omega,\mathbf{p}_k))\,\mathrm{d}\boldsymbol{x}$$

$$\underset{k\to\infty}{\longrightarrow} \int_{\mathbb{R}^2} r\left(\boldsymbol{x}+\boldsymbol{\mu}_2\left(\omega,\mathbf{p}\right)\right)\mathcal{N}\left(\boldsymbol{x};\mathbf{0},\boldsymbol{\Sigma}_2\left(\omega,\mathbf{p}\right)\right)\mathrm{d}\boldsymbol{x} \equiv H_I\left(\omega,\mathbf{p}\right). \quad (4.173)$$

But $\{\mathbf{p}_k\}_{k\in\mathbb{N}}$ and $\mathbf{p}$ are arbitrary, showing that $H_I\left(\omega,\cdot\right)$ is continuous, for each fixed $\omega\in\Omega$. Hence, H_I is Carathéodory on $\Omega\times\mathcal{S}$. ★

The proof to the second part of Theorem 4.1 follows easily by direct application of the Fundamental Lemma (Lemma 4.3; also see Table 4.1) ■

4.5.2.2 Proof of Lemma 4.1

In the notation of the statement of the lemma, the joint conditional distribution of $\left[F\left(\mathbf{p},t\right)\ G\left(\mathbf{p},t\right)\right]^{\boldsymbol{T}}$ relative to the σ-algebra $\mathscr{C}\left(\mathcal{T}_{t-1}\right)$ can be readily shown to be Gaussian with mean $\boldsymbol{\mu}_{t|t-1}^{F,G}(\mathbf{p})$ and covariance $\boldsymbol{\Sigma}_{t|t-1}^{F,G}(\mathbf{p})$, for all $(\mathbf{p},t)\in\mathcal{S}\times\mathbb{N}_{N_T}^2$. This is due to the fact that we have implicitly assumed that the channel fields $F\left(\mathbf{p},t\right)$ and $G\left(\mathbf{p},t\right)$ are jointly Gaussian. It is then a typical exercise (possibly somewhat tedious though) to show that the functions $\boldsymbol{\mu}_{t|t-1}^{F,G}$ and $\boldsymbol{\Sigma}_{t|t-1}^{F,G}$ are of the form asserted in the statement of the lemma. Regarding the proof for (4.29), observe that we can write

$$\begin{aligned}
&\mathbb{E}\left\{\left|f\left(\mathbf{p},t\right)\right|^m\left|g\left(\mathbf{p},t\right)\right|^n\middle|\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}\\
&\quad\equiv 10^{(m+n)\rho/20}\mathbb{E}\left\{\exp\left(\frac{\log\left(10\right)}{20}\left(mF\left(\mathbf{p},t\right)+nG\left(\mathbf{p},t\right)\right)\right)\middle|\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}\\
&\quad\equiv 10^{(m+n)\rho/20}\mathbb{E}\left\{\exp\left(\frac{\log\left(10\right)}{20}\left[m\,n\right]\left[F\left(\mathbf{p},t\right)\ G\left(\mathbf{p},t\right)\right]^{\boldsymbol{T}}\right)\middle|\mathscr{C}\left(\mathcal{T}_{t-1}\right)\right\}, \qquad (4.174)
\end{aligned}$$

with the conditional expectation on the RHS being nothing else than the conditional moment generating function of the conditionally jointly Gaussian random vector $\left[F\left(\mathbf{p},t\right)\ G\left(\mathbf{p},t\right)\right]^{\boldsymbol{T}}$ at each $\mathbf{p}$ and t, evaluated at the point $\left(\log\left(10\right)/20\right)\left[m\,n\right]^{\boldsymbol{T}}$, for any choice of $(m,n)\in\mathbb{Z}\times\mathbb{Z}$. Recalling the special form of the moment generating function for Gaussian random vectors, the result readily follows. ■

4.5.2.3 Proof of Theorem 4.2

It will suffice to show that both objectives of (4.30) and (4.31) are Carathéodory in $\Omega\times\mathcal{S}$. But this statement may be easily shown by analytically expressing both (4.30) and (4.31) using Lemma 4.1. Now, since both objectives of (4.30) and (4.31) are Carathéodory, we may invoke Theorem 4.8 and Corollary **??** (Appendix A), in an

inductive fashion, for each $t \in \mathbb{N}_{N_T}^2$, guaranteeing the existence of at least one $\mathscr{C}(\mathcal{T}_{t-1})$-measurable decision for either (4.30), or (4.31), say $\widetilde{\mathbf{p}}^*(t)$, which solves the optimization problem considered, for all $\omega \in \Omega$. Proceeding inductively gives the result. ■

4.5.2.4 Proof of Theorem 4.4

By assumption, $V(\mathbf{p}, t)$ is $\mathbf{L}.\mathbf{MD}.\mathbf{G}\diamondsuit(\mathscr{H}_t, \mu)$, implying, for every $t \in \mathbb{N}_{N_T}^+$, the existence of an event $\Omega_t \subseteq \Omega$, satisfying $\mathcal{P}(\Omega_t) \equiv 1$, such that, for every $\mathbf{p} \in \mathcal{S}^R$,

$$\mu\mathbb{E}\{V(\mathbf{p}, t-1)|\mathscr{H}_{t-1}\}(\omega) \equiv \mathbb{E}\{V(\mathbf{p}, t)|\mathscr{H}_{t-1}\}(\omega), \quad \forall \omega \in \Omega_t. \tag{4.175}$$

Fix $t \in \mathbb{N}_{N_T}^2$. Consider any admissible policy $\mathbf{p}^o(t)$ at t, implemented at t and decided at $t-1 \in \mathbb{N}_{N_T-1}^+$. By our assumptions, $V(\cdot, \cdot, t)$ is $\boldsymbol{SP}\diamondsuit\mathfrak{C}_{\mathscr{H}_t}$. Additionally, because $\mathbf{p}^o(t)$ is admissible, it will be measurable relative to the limit σ-algebra $\mathscr{P}_t^{\uparrow}$ and, hence, measurable relative to $\mathscr{H}_t$. Thus, there exists an event $\Omega_t^{\mathbf{p}^o} \subseteq \Omega$, with $\mathcal{P}\left(\Omega_t^{\mathbf{p}^o}\right) \equiv 1$, such that, for every $\omega \in \Omega_t^{\mathbf{p}^o}$,

$$\begin{aligned}\mathbb{E}\{V(\mathbf{p}^o(t), t)|\mathscr{H}_t\}(\omega) &\equiv \mathbb{E}\{V(\mathbf{p}, t)|\mathscr{H}_t\}(\omega)|_{\mathbf{p}=\mathbf{p}^o(\omega,t)} \\ &\equiv h_t(\omega, \mathbf{p}^o(\omega, t)),\end{aligned} \tag{4.176}$$

where the extended real-valued random function $h_t : \Omega \times \mathcal{S}^R \to \overline{\mathbb{R}}$ is jointly $\mathscr{H}_t \otimes \mathscr{B}\left(\mathcal{S}^R\right)$-measurable, with $h_t(\omega, \mathbf{p}) \equiv \mathbb{E}\{V_I(\mathbf{p}, t)|\mathscr{H}_t\}(\omega)$, everywhere in $(\omega, \mathbf{p}) \in \Omega \times \mathcal{S}^R$.

Also by our assumptions, $V(\cdot, \cdot, t)$ is $\boldsymbol{SP}\diamondsuit\mathfrak{C}_{\mathscr{H}_{t-1}}$, as well. Similarly to the arguments made above, if $\mathbf{p}^o(t)$ is assumed to be measurable relative to the limit σ-algebra $\mathscr{P}_{t-1}^{\uparrow}$, or, in other words, admissible at $t-1$, then it will also be measurable relative to $\mathscr{H}_{t-1}$. Therefore, there exists an event $\Omega_{t^-}^{\mathbf{p}^o} \subseteq \Omega$, with $\mathcal{P}\left(\Omega_{t^-}^{\mathbf{p}^o}\right) \equiv 1$, such that, for every $\omega \in \Omega_{t^-}^{\mathbf{p}^o}$,

$$\begin{aligned}\mathbb{E}\{V(\mathbf{p}^o(t), t)|\mathscr{H}_{t-1}\}(\omega) &\equiv \mathbb{E}\{V(\mathbf{p}, t)|\mathscr{H}_{t-1}\}(\omega)|_{\mathbf{p}=\mathbf{p}^o(\omega,t)} \\ &\equiv h_{t^-}(\omega, \mathbf{p}^o(\omega, t)),\end{aligned} \tag{4.177}$$

where the random function $h_{t^-} : \Omega \times \mathcal{S}^R \to \overline{\mathbb{R}}$ is jointly $\mathscr{H}_{t-1} \otimes \mathscr{B}\left(\mathcal{S}^R\right)$-measurable, with $h_{t^-}(\omega, \mathbf{p}) \equiv \mathbb{E}\{V(\mathbf{p}, t)|\mathscr{H}_{t-1}\}(\omega)$, everywhere in $(\omega, \mathbf{p}) \in \Omega \times \mathcal{S}^R$. Note that, by

construction, $\mathbf{p}^o(t)$ will also be admissible at time t and, therefore, measurable relative to and $\mathscr{H}_t$, as well.

Now, we combine the arguments made above. Keep $t \in \mathbb{N}^2_{N_T}$ fixed. At time slot $t-2 \in \mathbb{N}_{N_T-2}$, let $\mathbf{p}^o(t-1) \equiv \mathbf{p}^o(\omega, t-1)$ be a $\mathscr{C}(\mathcal{T}_{t-2})$-measurable admissible policy (recall that **C1-C6** are satisfied by assumption; also recall that, if $t \equiv 2$, $\mathscr{C}(\mathcal{T}_{t-2}) \equiv \mathscr{C}(\mathcal{T}_0)$ is the trivial σ-algebra). At the *next* time slot $t-1 \in \mathbb{N}^+_{N_T-1}$, let us choose $\mathbf{p}^o(t) \equiv \mathbf{p}^o(\omega, t-1)$; in this case, $\mathbf{p}^o(t)$ will also be $\mathscr{C}(\mathcal{T}_{t-2})$-measurable and result in the *same final position for the relays at time slot* $t \in \mathbb{N}^2_{N_T}$. As a result, the relays just stay still. Under these circumstances, at time slot $t-1 \in \mathbb{N}^+_{N_T-1}$, the expected network QoS will be $\mathbb{E}\{V(\mathbf{p}^o(t-1), t-1)\}$, whereas, at the next time slot $t \in \mathbb{N}^2_{N_T}$, it will be $\mathbb{E}\{V(\mathbf{p}^o(t-1), t)\}$. Exploiting (4.175), we may write

$$\mu h_{t-1}(\omega, \mathbf{p}) \equiv h_{t^-}(\omega, \mathbf{p}), \quad \forall (\omega, \mathbf{p}) \in \Omega_t \bigcap \Omega^{\mathbf{p}^o}_{t-1} \bigcap \Omega^{\mathbf{p}^o}_{t^-} \times \mathcal{S}^R, \tag{4.178}$$

where, obviously, $\mathcal{P}\left(\Omega_t \bigcap \Omega^{\mathbf{p}^o}_{t-1} \bigcap \Omega^{\mathbf{p}^o}_{t^-}\right) \equiv 1$. Consequently, it will be true that

$$\mu h_{t-1}(\omega, \mathbf{p}^o(\omega, t-1)) \equiv h_{t^-}(\omega, \mathbf{p}^o(\omega, t-1)), \quad \forall \omega \in \Omega_t \bigcap \Omega^{\mathbf{p}^o}_{t-1} \bigcap \Omega^{\mathbf{p}^o}_{t^-}. \tag{4.179}$$

From (4.176) and (4.177), it is also true that

$$\mu\mathbb{E}\{V(\mathbf{p}^o(t-1), t-1) | \mathscr{H}_{t-1}\}(\omega) \equiv \mathbb{E}\{V(\mathbf{p}^o(t-1), t) | \mathscr{H}_{t-1}\}(\omega), \tag{4.180}$$

almost everywhere with respect to $\mathcal{P}$. This, of course, implies that

$$\mu\mathbb{E}\{V(\mathbf{p}^o(t-1), t-1)\} \equiv \mathbb{E}\{V(\mathbf{p}^o(t-1), t)\}, \tag{4.181}$$

and for all $t \in \mathbb{N}^2_{N_T}$, since t was arbitrary.

Since (4.181) holds for all admissible policies decided at time slot $t-2 \in \mathbb{N}_{N_T-2}$, it will also hold for the respective optimal policy, that is,

$$\mu\mathbb{E}\left\{V\left(\mathbf{p}^*(t-1), t-1\right)\right\} \equiv \mathbb{E}\left\{V\left(\mathbf{p}^*(t-1), t\right)\right\}, \quad \forall t \in \mathbb{N}^2_{N_T}. \tag{4.182}$$

Next, as discussed above, the choice $\mathbf{p}^o(t) \equiv \mathbf{p}^*(\omega, t-1)$ constitutes an admissible policy decided at time slot $t-1 \in \mathbb{N}^+_{N_T-1}$; it suffices to see that $\mathbf{p}^*(\omega, t-1) \in \mathcal{C}\left(\mathbf{p}^*(\omega, t-1)\right)$, by definition of our initial 2-stage problem, because "staying still"

is always a feasible decision for the relays. Consequently, because the optimal policy $\mathbf{p}^*(t)$ results in the highest network QoS, *among all admissible policies*, it will be true that

$$\mu\mathbb{E}\left\{V\left(\mathbf{p}^*(t-1),t-1\right)\right\} \leq \mathbb{E}\left\{V\left(\mathbf{p}^*(t),t\right)\right\}, \quad \forall t \in \mathbb{N}_{N_T}^2, \tag{4.183}$$

completing the proof of Theorem 4.4. ■

Chapter 5

Conclusions & Future Research Directions

This dissertation is the first *systematic* piece of work on spatially controlled communications and, in particular, on spatially controlled relay beamforming, which is, admittedly, a novel and highly interdisciplinary topic. The design of such systems is far from trivial. We have demonstrated the importance of realistic wireless channel modeling, as well as the need for utilizing and advancing theory and technical background from the vast area of Operations Research, such as stochastic programming, which is clearly non-traditional in the typical wireless communications and signal processing literature. In a nutshell, and for the first time, we were able to propose an effective, reasonably complicated spatially controlled relay beamforming system, suitable for realistic space and time varying communication channel environments. The performance of the proposed system was analyzed and predicted under a rigorous theoretical framework. Our numerical simulations, in different stages of this work provided experimental evidence which confirmed our theoretical predictions, and showed great potential of the proposed system. In other words, our work is not only theoretically correct, but is expected to result in significant (non-marginal) performance gains in practice.

Our research in the context of spatially controlled communications, also stimulated novel theoretical research, with significant intellectual merit. In particular, our need to characterize the performance of the hierarchical channel model of Chapter 3, led to the development of new insightful results related to the convergence properties and asymptotic consistency of approximate nonlinear filters, both in Markov and non-Markov cases. Additionally, our extensive work in the theoretical analysis of the proposed spatially controlled beamforming system led to the development, among other results, of a new, generalized version of the Fundamental Lemma of Stochastic Control, which is

of independent theoretical and practical interest.

In conclusion, we would also like to provide a few directions for future research, initiated by the research developed herein:

- As the reader might have noticed, although, in Chapter 3, two wireless channel models have been presented and analyzed, whereas in Chapter 4, the proposed spatially controlled beamforming system is designed assuming only one of them, namely, that assuming the channel is a parametric spatiotemporal Gaussian field. Therefore, an immediate direction for future research is to reformulate and extend the system proposed in Chapter 4 under a hierarchical channel modeling assumption (the second model presented in Chapter 3), which, due to to time constraints, has not been pursued in the dissertation. Such a channel model, although slightly more complicated, would provide more numerically efficient solution (due to Markovianity), and potentially increased flexibility.

- In the same context, real-world, experimental validation of the HMM-based, hierarchical channel model of Chapter 3 will be highly valuable, since it would provide evidence that such a model is practically relevant and realistic, therefore justifying its consideration in the design of a spatially controlled communication system. Additionally, the development of efficient methods for parameter estimation, which have not been pursued herein, are of high practical importance. At this point, this is a subject of current, active research.

- Another directly relevant research direction is the complete solution of the power minimization beamforming problem of Section 1.2.2. Although the power minimization problem was pursued at a preliminary stage in [97], its solution is far from complete and requires further developments.

- In Chapter 4, we proposed relay motion control schemes, based, however, on a myopic optimization approach, where the relays decide their future positions, by optimizing an one-step-ahead conditional cost. As we saw, in this case, it was also possible to characterize the performance of the system across time-slots, in an effective and elegant manner. What about multi-stage formulations of the

problem? Finite horizon or infinite horizon (stationary) solutions? In particular, would such an approach result in better *per-slot* performance than the myopic approach? And, would a *per-slot* characterization of system performance be possible?

- In the same spirit, what about controlling risk? In the numerical simulations provided, we have seen that, although the average performance of the system improves (average SINR), its variance also increases. Therefore, in a practical setting, it might be reasonable to search for solutions which also control the variability of the optimal system performance, in order to enhance system robustness, in a stochastic sense. There is increased interest in *risk-averse* stochastic programming over the last two decades, mainly in the Operations Research community, which may be leveraged and further developed in order to provide such robust solutions.

- Another possible research direction can be the consideration of other, potentially simpler models for describing the space time evolution of the communication channel. For instance, on such possibility could be the adoption of "more discrete" channel models of, say, the Gilbert-Elliot, "good", "bad", "better" type. Such an approach would be very different from what we have considered so far, and would require effective channel characterization, in terms of ensuring a certain performance level of the communication network under consideration, and then utilizing such knowledge for performing stochastic decision making. The main benefit of such an approach would be a much lower expected computational complexity, as well as the potential for identifying optimal resource allocation policies in closed form (for example, of the threshold type).

- Finally, there are numerous extensions of the idea of spatially controlled relay beamforming to physical layer communication tasks other than relay beamforming, such as Physical Layer Security and Privacy, Spectrum Sensing, Radar and Target Detection & Tracking, Multiuser/MIMO networks, Energy Harvesting and Green Communications, to name a few.

In this dissertation, we have provided the first systematic framework to spatially controlled communications, focusing on a basic beamforming scenario. We have shown that this idea is feasible, and that it provides significant performance gains. We thus hope that our research will initiate interest in the context of spatially controlled communications, and that this dissertation is just the beginning.

References

[1] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," *Information Theory, IEEE Transactions on*, vol. 55, no. 6, pp. 2499–2517, June 2009.

[2] G. Zheng, Kai-Kit Wong, A. Paulraj, and B. Ottersten, "Collaborative-Relay Beamforming with Perfect CSI: Optimum and Distributed Implementation," *Signal Processing Letters, IEEE*, vol. 16, no. 4, pp. 257–260, April 2009.

[3] L. Dong, A.P. Petropulu, and H.V. Poor, "Weighted cross-layer cooperative beamforming for wireless networks," *Signal Processing, IEEE Transactions on*, vol. 57, no. 8, pp. 3240–3252, Aug 2009.

[4] E. Koyuncu, Y. Jing, and H. Jafarkhani, "Distributed beamforming in wireless relay networks with quantized feedback," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 8, pp. 1429–1439, October 2008.

[5] S. Fazeli-Dehkordy, S. ShahbazPanahi, and S. Gazor, "Multiple peer-to-peer communications using a network of relays," *Signal Processing, IEEE Transactions on*, vol. 57, no. 8, pp. 3053–3062, Aug 2009.

[6] J. Li, A.P. Petropulu, and H.V. Poor, "Cooperative transmission for relay networks based on second-order statistics of channel state information," *Signal Processing, IEEE Transactions on*, vol. 59, no. 3, pp. 1280–1291, March 2011.

[7] Y. Liu and A.P. Petropulu, "On the sumrate of amplify-and-forward relay networks with multiple source-destination pairs," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 11, pp. 3732–3742, November 2011.

[8] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *Information Theory, IEEE Transactions on*, vol. 51, no. 9, pp. 3037–3063, Sept 2005.

[9] V. Havary-Nassab, S. ShahbazPanahi, A. Grami, and Zhi-Quan Luo, "Distributed Beamforming for Relay Networks based on Second-Order Statistics of the Channel State Information," *Signal Processing, IEEE Transactions on*, vol. 56, no. 9, pp. 4306–4316, Sept 2008.

[10] V. Havary-Nassab, S. ShahbazPanahi, and A. Grami, "Optimal distributed beamforming for two-way relay networks," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1238–1250, March 2010.

[11] A. Segall, "Stochastic processes in estimation theory," *Information Theory, IEEE Transactions on*, vol. 22, no. 3, pp. 275–286, May 1976.

[12] Robert J Elliott, Lakhdar Aggoun, and John B Moore, *Hidden Markov Models*, Springer, 1994.

[13] A. Segall, "Recursive estimation from discrete-time point processes," *Information Theory, IEEE Transactions on*, vol. 22, no. 4, pp. 422–431, Jul 1976.

[14] S.I. Marcus, "Optimal nonlinear estimation for a class of discrete-time stochastic systems," *Automatic Control, IEEE Transactions on*, vol. 24, no. 2, pp. 297–302, Apr 1979.

[15] Robert J Elliott, "Exact adaptive filters for markov chains observed in gaussian noise," *Automatica*, vol. 30, no. 9, pp. 1399–1408, 1994.

[16] Robert J Elliott and Hailiang Yang, "How to count and guess well: Discrete adaptive filters," *Applied Mathematics and Optimization*, vol. 30, no. 1, pp. 51–78, 1994.

[17] Harold J Kushner, "Approximations to optimal nonlinear filters," *Automatic Control, IEEE Transactions on*, vol. 12, no. 5, pp. 546–556, 1967.

[18] K. Ito and Kaiqi Xiong, "Gaussian filters for nonlinear filtering problems," *Automatic Control, IEEE Transactions on*, vol. 45, no. 5, pp. 910–927, May 2000.

[19] Harold J Kushner and Paul Dupuis, *Numerical methods for stochastic control problems in continuous time*, vol. 24, Springer, 2001.

[20] H. J. Kushner, "Numerical aproximation to optimal nonlinear filters," *http://www.dam.brown.edu/lcds/publications/documents/Kusher_Pub001.pdf*, 2008.

[21] Leonard F Richardson, *Measure and integration: a concise introduction to real analysis*, John Wiley & Sons, 2009.

[22] Gilles Pagès, Huyên Pham, et al., "Optimal quantization methods for nonlinear filtering with discrete-time observations," *Bernoulli*, vol. 11, no. 5, pp. 893–932, 2005.

[23] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, 2002.

[24] F. Daum, "Nonlinear filters: Beyond the Kalman Filter," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, pp. 57 – 69, Aug. 2005.

[25] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 4th edition, 2002.

[26] R.J. Elliott and V. Krishnamurthy, "New finite-dimensional filters for parameter estimation of discrete-time linear gaussian models," *Automatic Control, IEEE Transactions on*, vol. 44, no. 5, pp. 938–951, May 1999.

[27] Robert B. Ash and Catherine Doleans-Dade, *Probability and Measure Theory*, Academic Press, 2000.

[28] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, 2009.

[29] P. Berti, L. Pratelli, and P. Rigo, "Almost sure weak convergence of random probability measures," *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 78, no. 2, pp. 91 – 97, April 2006.

[30] R. Grubel and Z. Kabluchko, "A functional central limit theorem for branching random walks, almost sure weak convergence, and applications to random trees," *http://arxiv.org/pdf/1410.0469.pdf*, 2014.

[31] P. Rebeschini and R. van Handel, "Can local particle filters beat the curse of dimensionality?," *http://arxiv.org/abs/1301.6585*, 2013.

[32] Y. Sharifi, "Eigenvalues of adjugate matrix," *http://ysharifi.wordpress.com/2011/01/22/eigenvalues-of-adjugate-matrix*, January 2011.

[33] A. M. Mathai and S. B. Provost, *Quadratic forms in random variables: theory and applications*, M. Dekker New York, 1992.

[34] B Laurent and P Massart, "Adaptive estimation of a quadratic functional by model selection," *Annals of Statistics*, vol. 28, no. 5, pp. 1302 – 1338, 2000.

[35] A. Farina, B. Ristic, and D. Benvenuti, "Tracking a ballistic target: comparison of several nonlinear filters," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 38, no. 3, pp. 854–867, Jul 2002.

[36] Xiao-Rong Li and Vesselin P. Jilkov, "A survey of maneuvering target tracking: approximation techniques for nonlinear filtering," in *Proc. SPIE*, 2004, vol. 5428, pp. 537–550.

[37] S.I. Roumeliotis and George A. Bekey, "Bayesian estimation and kalman filtering: a unified framework for mobile robot localization," in *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, 2000, vol. 3, pp. 2985–2992 vol.3.

[38] Alexander S. Volkov, "Accuracy bounds of non-gaussian bayesian tracking in a {NLOS} environment," *Signal Processing*, vol. 108, no. 0, pp. 498 – 508, 2015.

[39] Dan Crisan and Boris Rozovskii, *The Oxford handbook of nonlinear filtering*, Oxford University Press, 2011.

[40] Zhe Chen, "Bayesian filtering: From kalman filters to particle filters, and beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[41] Thomas Bengtsson, Peter Bickel, Bo Li, et al., "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems," *Probability and Statistics: Essays in Honor of David A. Freedman*, vol. 2, pp. 316–334, 2008.

[42] Paul Bui Quang, Christian Musso, and Francois Le Gland, "An insight into the issue of dimensionality in particle filtering," in *Information Fusion (FUSION), 2010 13th Conference on*. IEEE, 2010, pp. 1–8.

[43] Patrick Rebeschini, *Nonlinear Filtering in High Dimension*, Ph.D. thesis, Princeton University, 2014.

[44] Afef Sellami, "Comparative survey on nonlinear filtering methods: the quantization and the particle filtering approaches," *Journal of Statistical Computation and Simulation*, vol. 78, no. 2, pp. 93–113, 2008.

[45] Dan Crisan and Arnaud Doucet, "A survey of convergence results on particle filtering methods for practitioners," *Signal Processing, IEEE Transactions on*, vol. 50, no. 3, pp. 736–746, 2002.

[46] Robert J Elliott, Francois Dufour, and W Paul Malcolm, "State and mode estimation for discrete-time jump markov systems," *SIAM journal on control and optimization*, vol. 44, no. 3, pp. 1081–1104, 2005.

[47] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, Springer Verlag, New York, 2005.

[48] N. Chatzipanagiotis, Y. Liu, A. Petropulu, and M. M. Zavlanos, "Controlling groups of mobile beamformers," in *IEEE Conference on Decision and Control*, Hawaii, 2012.

[49] Ke Ma, Y. Zhang, and W. Trappe, "Managing the mobility of a mobile sensor network using network dynamics," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 19, no. 1, pp. 106–120, Jan 2008.

[50] D. S. Kalogerias, N. Chatzipanagiotis, M. M. Zavlanos, and A. P. Petropulu, "Mobile jammers for secrecy rate maximization in cooperative networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 2901–2905.

[51] D. S. Kalogerias and A. P. Petropulu, "Mobi-cliques for improving ergodic secrecy in fading wiretap channels under power constraints," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 1578–1591.

[52] A. Ghaffarkhah and Y. Mostofi, "Communication-aware motion planning in mobile networks," *Automatic Control, IEEE Transactions on*, vol. 56, no. 10, pp. 2478–2485, Oct 2011.

[53] Yuan Yan and Y. Mostofi, "Robotic router formation in realistic communication environments," *Robotics, IEEE Transactions on*, vol. 28, no. 4, pp. 810–827, Aug 2012.

[54] Alireza Ghaffarkhah and Yasamin Mostofi, "Dynamic networked coverage of time-varying environments in the presence of fading communication channels," *ACM Transactions on Sensor Networks (TOSN)*, vol. 10, no. 3, pp. 45, 2014.

[55] A. Goldsmith, *Wireless Communications*, Cambridge university press, 2005.

[56] M. Malmirchegini and Y. Mostofi, "On the Spatial Predictability of Communication Channels," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 3, pp. 964–978, March 2012.

[57] S. L. Cotton and W. G. Scanlon, "Higher Order Statistics for Lognormal Small-Scale Fading in Mobile Radio Channels," *Antennas and Wireless Propagation Letters, IEEE*, vol. 6, pp. 540–543, 2007.

[58] M. Gudmundson, "Correlation Model for Shadow Fading in Mobile Radio Systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145–2146, Nov 1991.

[59] Alejandro Gonzalez-Ruiz, Alireza Ghaffarkhah, and Yasamin Mostofi, "A Comprehensive Overview and Characterization of Wireless Channels for Networked Robotic and Control Systems," *Journal of Robotics*, vol. 2011, 2012.

[60] C. Oestges, N. Czink, B. Bandemer, P. Castiglione, F. Kaltenberger, and A. J. Paulraj, "Experimental characterization and modeling of outdoor-to-indoor and indoor-to-indoor distributed channels," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2253–2265, Jun 2010.

[61] M. G. Genton, "Classes of Kernels for Machine Learning: A Statistics Perspective," *Journal of Machine Learning Research*, vol. 2, no. Dec, pp. 299–312, 2001.

[62] R. J. Adler, *The Geometry of Random Fields*, vol. 62, Siam, 2010.

[63] P. Abrahamsen, *A Review of Gaussian Random Fields and Correlation Functions*, Norsk Regnesentral/Norwegian Computing Center, 1997.

[64] R. J. Adler and J. E. Taylor, *Random Fields & Geometry*, Springer Science & Business Media, 2009.

[65] Lakhdar Aggoun and Robert J Elliott, *Measure theory and filtering: Introduction and applications*, vol. 15, Cambridge University Press, 2004.

[66] A.O. Kaya, L.J. Greenstein, and W. Trappe, "Characterizing indoor wireless channels via ray tracing combined with stochastic modeling," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 8, pp. 4165–4175, August 2009.

[67] Seung-Jun Kim, E. Dall'Anese, and G.B. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged kalman filtering," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 1, pp. 24–36, Feb 2011.

[68] E. Dall'Anese, Seung-Jun Kim, and G.B. Giannakis, "Channel gain map tracking via distributed kriging," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 3, pp. 1205–1211, March 2011.

[69] Neil J Gordon, David J Salmond, and Adrian FM Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*. IET, 1993, vol. 140, pp. 107–113.

[70] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Aerospace Conference, 2003, Proceedings IEEE*, March 2003, vol. 4, pp. 1979–1993.

[71] Y. Liu and A.P. Petropulu, "Relay selection and scaling law in destination assisted physical layer secrecy systems," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, Aug 2012, pp. 381–384.

[72] N. Chatzipanagiotis, Yupeng Liu, A. Petropulu, and M.M. Zavlanos, "Controlling groups of mobile beamformers," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, Dec 2012, pp. 1984–1989.

[73] J. Fink, A. Ribeiro, and V. Kumar, "Robust control of mobility and communications in autonomous robot teams," *IEEE Access*, vol. 1, pp. 290–309, 2013.

[74] Y. Yan and Y. Mostofi, "Co-optimization of communication and motion planning of a robotic operation under resource constraints and in fading environments," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1562–1572, April 2013.

[75] J. Fink, A. Ribeiro, and V. Kumar, "Robust control of mobility and communications in autonomous robot teams," *IEEE Access*, vol. 1, pp. 290–309, 2013.

[76] J. Fink, A. Ribeiro, and V. Kumar, "Robust control for mobility and wireless communication in cyber-physical systems with application to robot teams," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 164–178, Jan 2012.

[77] Y. Yan and Y. Mostofi, "To go or not to go: On energy-aware and communication-aware robotic operation," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 218–231, Sept 2014.

[78] A. Ghaffarkhah and Y. Mostofi, "Path planning for networked robotic surveillance," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3560–3575, July 2012.

[79] A. Ghaffarkhah and Y. Mostofi, "Communication-aware motion planning in mobile networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2478–2485, Oct 2011.

[80] Regina C. Elandt-Johnson and Norman Lloyd Johnson, *Survival Models and Data Analysis*, Wiley, 1999.

[81] Rick Durrett, *Probability: theory and examples*, Cambridge university press, 2010.

[82] J. L. Speyer and W. H. Chung, *Stochastic Processes, Estimation, and Control*, vol. 17, Siam, 2008.

[83] K. J. Astrom, *Introduction to Stochastic Control Theory*, vol. 70, New York: Academic Press, 1970.

[84] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, vol. 317, Springer Science & Business Media, 2009.

[85] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory (MPS-SIAM Series on Optimization)*, SIAM-Society for Industrial and Applied Mathematics, 1st edition, 2009.

[86] D. P. Bertsekas, *Dynamic Programming & Optimal Control*, vol. II: Approximate Dynamic Programming, Athena Scientific, Belmont, Massachusetts, 4th edition, 2012.

[87] D. P. Bertsekas and S. E. Shreve, *Stochastic optimal control: The discrete time case*, vol. 23, Academic Press New York, 1978.

[88] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, vol. 2, Cambridge university press Cambridge, 1996.

[89] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-Time Nonlinear Filtering Algorithms Using Gauss-Hermite Quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, May 2007.

[90] G. H. Golub and J. H. Welsch, "Calculation of Gauss Quadrature Rules," *Mathematics of computation*, vol. 23, no. 106, pp. 221–230, 1969.

[91] B. W. Levinger, "The Square Root of a 2×2 Matrix," *Mathematics Magazine*, vol. 53, no. 4, pp. 222–224, 1980.

[92] Charalambos D. Aliprantis and Kim Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer Science & Business Media, 2006.

[93] George W Mackey, "Borel structure in groups and their duals," *Transactions of the American Mathematical Society*, vol. 85, no. 1, pp. 134–165, 1957.

[94] L. Meier, R. Larson, and A. Tether, "Dynamic programming for stochastic control of discrete systems," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 767–775, Dec 1971.

[95] G. B. Folland, *Real Analysis: Modern Techniques and their Applications*, John Wiley & Sons, 2nd edition, 1999.

[96] Ralph E Strauch, "Negative dynamic programming," *The Annals of Mathematical Statistics*, vol. 37, no. 4, pp. 871–890, 1966.

[97] D. S. Kalogerias and A. P. Petropulu, "Mobile Beamforming & Spatially Controlled Relay Communications," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 6405–6409.