Dissertation

Decentralized Localization Based on Wave Fields

Particle Filters and Weiss-Weinstein Error Bounds

By
DI Florian Xaver
Institute of Telecommunications
Fakultät für Elektrotechnik und Informationstechnik, Technische Universität Wien

Submitted in partial fulfillment of the requirements for the degree of
Doktor der technischen Wissenschaften (Dr. techn.)

Advised by
Univ.Prof. Ing. Dipl.-Ing. Dr.-Ing. Christoph F. Mecklenbräuker

2nd examiner:
Peter Gerstoft, Ph.D. M.Sc. M.Sc.
Scripps Institution of Oceanography, University of California, San Diego

Rigorosum: Vienna, March 2013 — Update: Darmstadt, October 24, 2013

Abstract

A key-challenge in wireless sensor networks is the development of decentralized signal processing and algorithms, i.e. without the central fusion center. More specific, in my dissertation I have contributed to the localization of acoustic sources in acoustic wave fields. It contains three elements:

The physical model in terms of the acoustic wave equation is continuous and has to be discretized and decentralized. I utilize a stochastic model to incorporate noise and the lack of knowledge.

On top of this model, I use a decentralized maximum a-posteriori particle filter as an estimator. It supports the non-Gaussianity and non-linearity of my model. For the final global consensus of the source location, I additionally present a consensus algorithm.

Non-Gaussian and discrete distributions with finite support demand for general analytic Bayesian performance bounds to benchmark estimators. Thus, I derive the analytic sequential Weiss-Weinstein lower bound on the error variance of any estimator for a linear model and probability distributions: Gaussian distributions, discrete / continuous uniform distributions, exponential distributions, Laplace distributions, and discrete distributions with finite alphabet.

Eventually, I join these elements and, moreover, consider the perturbed communication between sensors. On that account, I generalize the sequential Weiss-Weinstein bound for my non-linear model.

Zusammenfassung

Zukünftige Funksensornetze lassen darauf hoffen, durch Denzentralisierung der Signalverarbeitung die Verwaltung und damit die Rechenleistung auf die Sensoren verteilen zu können. Dies hat zur Folge, dass Algorithmen und Verfahren auf ihre Dezentralisierbarkeit untersucht werden müssen. In meiner Dissertation erforsche ich die dezentrale Lokalisierung akustischer Quellen unter Zuhilfenahme der zugrunde liegenden Physik. Dies lässt sich in drei Bereiche gliedern:

Das physikalische Modell in Form der akustischen Wellengleichung muss zunächst diskretisiert und zerlegt werden. Rauschen und mangelnde Kenntnis des Raumes führen zu einem dezentralen probabilistischen Modell.

Für die Lokalisierung verwendet ein Schätzer nun dieses Modell. Da dem Modell eine Nichtlinearität innewohnt und die Zustandsvektoren kontinuierliche und diskrete Zufallsvariablen beinhalten, entwickle ich einen dezentralen Particle-Filter als lokalen Maximum A-posteriori Schätzer. Darauf aufbauend, sichert ein Konsensus-Algorithmus den globalen Konsensus zwischen den einzelnen Sensoren.

Zur Untersuchung der Performance eines Schätzers verwende ich Bayessche untere Schranken für die Fehlervarianz. Die Bayessche Beschreibung und diskrete/kontinuierliche Wahrscheinlichkeitsverteilungen verlangen nach Sequentiellen Weiss-Weinstein-Schranken. Unter Vorraussetzung einer linearen Zustandsraumdarstellung gebe ich analytische Lösungen für Gauß-Verteilungen, Gleichverteilungen, Exponentialverteilungen, Laplace-Verteilungen und diskrete Verteilungen mit endlichem Alphabet an.

Schlussendlich führe ich alle drei Bereiche zusammen und gehe auch auf die gestörte Kommunikation zwischen den Sensoren ein. Für mein nichtlineares Modell leite ich die Sequentielle Weiss-Weinstein-Schranke her.

Acknowledgments

In the first place, I give thanks to my parents and my partner for mental support during studies, research, and publishing toward my Ph.D.

I am indebted to my advisors Christoph F. Mecklenbräuker and Gerald Matz for placing their trust in me. I have appreciated the academic freedom they gave me.

Research and writing a dissertation is a lonely work. Discussions with colleagues and Peter Gerstoft have been very productive, gave new ideas and motivation. Some colleagues became close friends and that gave raise to the amicable environment of our department.

Not to mention the fact that my position have been external funded by Vienna's government (WWTF) whereas my university place and the University itself is funded by Austria's government. Software like LaTeX, Octave, and Linux are open source with a world-wide, keen community that is still growing.

Last but not least, dear Reader, I am glad that my thesis raised your interest.

"All human knowledge begins with intuitions, thence passes to concepts and ends with ideas."
— Emanuel Kant, Kritik der reinen Vernunft

"There is no branch of math, however abstract, which may not some day be applied to phenomena in the real world."
— Nikolái Lobachevsky

Contents

List of Figures

List of Tables

Acronyms and Notation

Acronyms

ACF	autocorrelation function
AMC	argumentum-maximi consensus
BD	Bhattacharyya distance
BZ	Bobrovsky-Zakai
CR	Cramér-Rao
DDPF	decentralized distributed particle filter
FDM	finite-difference method
FEM	finite-element method
KF	Kalman filter
MAP	maximum a-posteriori
MC	maximum consensus
MMSE	minimum mean-squared error
MPF	marginalized particle filter
MSE	mean-squared error
PD	probability density
PDE	partial differential equation
PDF	probability density function
PF	particle filter
PMF	probability mass function
PSD	power spectral density
SCR	sequential Cramér-Rao
SEM	spectral-element method
SIR	sampling-importance-resampling
SN	sensor network
SPDE	stochastic partial differential equation
SWW	sequential Weiss-Weinstein
WW	Weiss-Weinstein

General Notation

$\mathbf{0}$	Zero vector
$\mathbf{1}$	One vector
$\dot{\cdot}$	Elements of ... that correspond to nodes along the boundary.
$[\cdot]_{a,b}$	Element of row a and column b of
$[S]$	S is the representative of the equivalence class $[S]$
$\cdot^{\mathrm{c}}, \cdot^{\mathrm{d}}$	Continuous, discrete
$\boldsymbol{C}_{\boldsymbol{x}}$	Covariance matrix of $\boldsymbol{x}$
Covar $\{\cdot, \cdot\}$	Covariance of ...
$\Delta\cdot$	Difference operator
$\Delta_{\cdot}$	Difference constant
$\overset{1}{\underset{h}{\Delta}}_x$	Difference quotient of first order
$d(\cdot, \cdot)$	Geodesic distance
Exp $\{\boldsymbol{\alpha}_k\}$	Exponential distribution
N $\{\boldsymbol{m}_{\boldsymbol{x}_k}, \boldsymbol{C}_{\boldsymbol{x}_k}\}$	Gaussian distribution
La $\{\boldsymbol{m}_{\boldsymbol{x}_k}, \boldsymbol{b}_k\}$	Laplace distribution
Unif $\{\boldsymbol{r}_k, \boldsymbol{s}_k\}$	Uniform distribution
∂_t	Partial derivative with respect to time
$\partial\mathcal{X}$	Boundary of discrete/continuous set $\mathcal{X}$
$\|\cdot\|_2$	Euclidean norm of ...
$E\{\cdot\}$	Expectation of ...
$\|$	given
$\boldsymbol{\nabla}$	Gradient
$\ddot{\cdot}$	Elements of ... that correspond to 2nd order boundary nodes.
$\odot$	Hadamard (element-wise) product
$\hat{\cdot}$	Estimate of ...
:	history (timespan)
$\mathbb{1}_{\cdot}$	Indicator function
$\cdot$	Inner-product
λ	Measures assuming existence of PDs (Counting or Lebesgue)
∇^2	Laplace operator
$L(\cdot)$	Likelihood ratio (density)
$\cdot^{[l,m]}$	Cluster m and particle l
$\cdot^{[l]}$	Particle l
$\cdot^{(m)}$	Cluster m
$\cdot^{(m,m')}$	Transmission from Cluster m to m'
μ	Negative Bhattacharyya distance

$\succcurlyeq$	The difference between left and right hand side is a positive semi-definite matrix.
$\geqslant$	The elementwise "greater equal"
$>$	The elementwise "greater than"
$\preccurlyeq$	The difference between left and right hand side is a negative semi-definite matrix.
$<$	The elementwise "smaller than"
P	Probability measure
$P\{\cdot\}$	Probability of ...
v_x	Probability density
v	Probability density $v(x) \triangleq v_x(x)$
f_x	Probability density function
f	Probability density function $f(x) \triangleq f_x(x)$
p_x	Probability mass function
p	Probability mass function $p(x) \triangleq p_x(x)$
ρ	Bhattacharyya coefficient
Var $\{\cdot\}$	Variance of ...
$\langle \cdot, \cdot \rangle_{C_x}$	Weighted inner-product by matrix $\boldsymbol{C_x}$
$\|\cdot\|_{C_x}$	Weighted norm by matrix $\boldsymbol{C_x}$
$x(\cdot)$	Function
$\boldsymbol{X}$	Matrix
$\boldsymbol{\mathfrak{x}}$	Matrix composed by other matrices
x	Scalar
$x[\cdot]$	Sequence (sampled function)
$\mathcal{X}$	Set, manifold
$\boldsymbol{x}$	Vector

$\boldsymbol{\alpha}_k$	parameter of exponential distribution xvi, 60, 81
$\boldsymbol{A}_k$	Auxilary matrix for SWW 48
$\boldsymbol{b}_k$	parameter of Laplace distribution xvi, 61, 81
c	Speed of acoustic waves 13, 21, 37, 38
C	Observation/measurement function 9
$\boldsymbol{C}$	Observation/measurement matrix 20, 21, 51, 71
D	Distortion 73
δ	Delta distribution 18
$\boldsymbol{D}_{k+1}^{ij}$	Auxilary matrix for SWW 48, 73
$\boldsymbol{D}_k$	Noise matrix 21
$\boldsymbol{\varepsilon}$	Estimation error 9, 38, 45
$\tilde{\varepsilon}$	generalized scalar estimation error 9
$\boldsymbol{\epsilon}$	Heaviside step distribution 14
$\boldsymbol{\varepsilon}_k$	Estimation error at time k 73
$\boldsymbol{e}_l$	Vector $\boldsymbol{e}_l$ is the lth unit vector 20
fl	Flatness 41
$\boldsymbol{g}$	Score (sensitivity function) 10, 11, 21, 46
$\boldsymbol{\Gamma}$	Input matrix 21
$\boldsymbol{\gamma}_k^{(m)}$	Coupling input vector 25
G	Green function 14, 15
$\boldsymbol{H}$	Parameter matrix of BZ and WW bound 46, 47
$\boldsymbol{H}_k$	Parameter matrix of SWW bound at time k 47, 73
$\mathfrak{H}_k$	Block diagonal matrix of matrices $\boldsymbol{H}_0$ to $\boldsymbol{H}_k$ 47
$\mathfrak{h}_k$	Column vector of $\mathfrak{H}_k$ 47
$\boldsymbol{h}$	Parameter vector (BZ or WW bound) 46, 81
I	Integral 7
I	Number of sample points in the 1st space dimension 15, 23, 28, 38, 42
i	First space dimension (discrete) 15, 17, 23, 38
$\boldsymbol{\imath}$	Innovations vector 38, 72, 95
$\varrho_{\mathrm{a/g}}$	Itakura distance 96
$\boldsymbol{I}$	Unit matrix 93
j	2nd space dimension (discrete) 15, 17, 23, 38
J	Number of sample points in the 2nd space dimension 15, 23, 28, 38, 42
$\boldsymbol{J}$	Auxilary matrix for CR (Fischer), BZ, and WW 46
$\boldsymbol{J}_k$	Auxilary matrix for SWW 48

$\boldsymbol{K}_k$ Kalman gain at time k 40

$\mathcal{L}$ Lattice 15, 38
ℓ Geodesic length 96
L Number of particles 7, 22

$\boldsymbol{m}_{\boldsymbol{x}_k}$ Mean vector xvi, 54, 61, 81
M The number of clusters 24, 28
$\mu_k^{(i,m)}$ Messages due to migrating sources 28

$\mathcal{N}^{(m)}$ Neighbor set of cluster m 25, 27
$\boldsymbol{n}_k$ Sources' life-span vector at time k 19, 21
$\boldsymbol{n}$ Normal vector 17, 42
ν Frequency 31

Ω Region 17
θ Angular frequency 40, 95

P_s Probability mass of a source (depends on location and time) sort 23, 26
$\boldsymbol{\Phi}_k$ Transition/system matrix 21, 38, 50, 71
Φ Transition/system function 47
$\boldsymbol{p}_k$ State pressure vector at time k 17, 21, 38, 71, 72
$\mathcal{P}$ Manifold of PSDs 41, 96
P_channel Input power of the channel 73

q Importance function (proposal distribution) 16, 22
q $q(\boldsymbol{r},t) = \partial_t p(\boldsymbol{r},t)$ 38
Q Spatial covariance function 14
$\boldsymbol{q}_k$ State vector of time derivative of pressure 17, 21, 38, 72
Q Measure for importance sampling 7
Q Quantizer function 23

$\boldsymbol{R}$ Noise coupling matrix 25
$\boldsymbol{r}$ Location 13, 14, 37
$\boldsymbol{r}_k$ Minimum vector (uniform distribution) xvi, 57, 81
ρ_x^B Bhattacharyya coefficient of the Bernoulli distribution 62
ρ_x^C Bhattacharyya coefficient of the categorical distribution 62, 81
$\varrho_{\mathrm{a/g}}$ Degradation of the prediction-error variance 95
ρ_x^E Bhattacharyya coefficient of the exponential distribution 60, 81
ρ_x^G Bhattacharyya coefficient of the Gaussian distribution 54, 73, 81
ϱ Georgiou distance 41, 96
ρ_x^L Bhattacharyya coefficient of the Laplace distribution 61, 81
ρ_x^U Bhattacharyya coefficient of the uniform distribution 58, 81
$\mathcal{R}$ Set of Sensors $\mathcal{R} \subset \mathcal{L}$ 16

S The number of active sources, i.e. $S = |\mathcal{T}_k|$ sort 18, 28
s Source function 13, 14, 37

$\sharp$	Taylor approximation 66
$\varsigma_{\boldsymbol{x}_k}$	Width of the support 58
$\boldsymbol{s}_k$	Source vector at time k 17
$\boldsymbol{s}_k$	Maximum vector (uniform distribution) xvi, 57, 81
S	Power spectral density 40, 95
$\boldsymbol{T}$	State-coupling matrix 25
t	Continuous time 13, 14, 37
k	Discrete time 5, 16, 17, 23, 38
$\boldsymbol{w}_k$	Transition/driving noise 17, 39, 47, 67
$\mathcal{T}_k$	Set $\mathcal{T}_k \subset \mathcal{L}$ of active sources at time k 18, 19, 23
$\boldsymbol{u}_k$	Input vector at time k 21
$\boldsymbol{v}_k$	Observation/measurement noise 20, 21, 47, 67
$\omega_k^{[l]}$	Weight at time k of particle l 7, 22, 26
$\mathcal{W}_{k+1}$	Set of weights at time $k+1$ 9, 72
$\boldsymbol{W}_k$	SWW bound at time k 48, 73
W	Space-time Wiener process 14
$\boldsymbol{x}$	Parameter vector 5, 9
$\boldsymbol{\xi}_k^{(m)}$	Coupling state vector 24
$\boldsymbol{x}_k$	State vector at time k 6, 21, 47, 50, 66, 67
x	First space dimension (continuous) 15
$\mathcal{X}_k$	Set of particles at time k 9
$\boldsymbol{y}$	Observations/measurements 9, 45, 46
$\boldsymbol{y}_k$	Observations/measurements at time k 6, 20, 21, 47, 51, 67, 71
y	2nd space dimension (continuous) 15

1
Preface

In the recent past, decentralization of algorithms has become popular[1] due to the increased availability of (wireless) sensor networks (SNs)[2]. The estimation[3] of parameters and states in SNs is the key challenge that I address in my dissertation.

Before I go into details, I have to concrete three frequently mentioned terms which are often sloppily used. My nomenclature classifies three non-central types of systems (e.g. estimators):

- In control theory, models stemming from the numerical approximation of partial differential equations (PDEs) are called *distributed* if quantities of the PDEs depend on their location. This characterization passes on to systems (estimators) that are based on such models.
- In contrast, a system that has several dependent computational units without fusion center, is *decentralized*.
- If a system (or a part of it) consists of several independent computational units, then they work *in parallel*.

The localization scenario I am presenting in the following is distributed due to the model, decentralized due to the absence of a fusion center, and some parts even work in parallel. More specifically, I focus on decentralized localization based on the underlying physical field. Knowledge about the physics has mainly been exploit in ocean acoustics and geophysics[4].

Throughout my dissertation I consider the in-door localization scenario in Figure 1.1. I consider a hallway with microphones aligned along the walls. An acoustic source occurs somewhere in this hallway and emits propagating acoustic waves. Some acoustic rays are received directly by the microphones. Other rays are reflected at the walls and cause interference. The microphones observe the instantaneous acoustic pressure. The evolving field, i.e. the pressure, is a state. Since source position and starting time typically vary over time, they are also states. In contrast, parameters are quantities which are constant over time, e.g. the speed of sound in the hallway. My aim is designing a decentralized estimator inferring the source positions and occurrence time utilizing the observed signals.

[1] Coates 2004; Dokmanic et al. 2012; Farahmand et al. 2010; Hlinka et al. 2013; Liu et al. 2009; Lu et al. 2009; Oreshkin et al. 2010; Waterschoot et al. 2012.

[2] Akyildiz et al. 2002; Patwari et al. 2005; Raghavendra et al. 2006; Zhao et al. 2004.

[3] Boukerche et al. 2007; Sheng et al. 2005; Veeravalli et al. 2012.

[4] Candy et al. 1992, 1996, 2002; Jovanovic et al. 2009; Yardim et al. 2009.

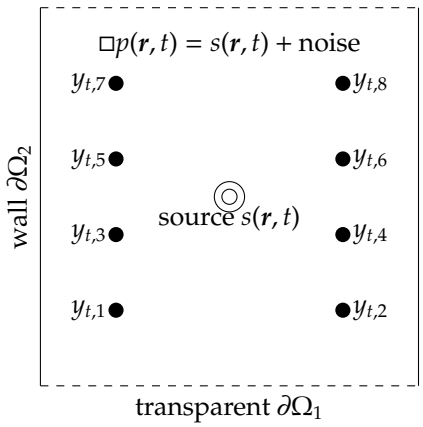


Figure 1.1: Acoustic source localization in a hallway. Microphones are aligned along the walls.

Generally, there are two classes of inference approaches:

1. Deterministic methods[5,6,7],
2. Probabilistic methods[8,9]

In the former case, a deterministic model description maps current states of sources to field observations of the microphones. This mapping is called the *forward model*. In an ideal case, a one-to-one mapping exists and the states can be computed. In the latter case, there are two probabilistic inference methods:

- The *Frequentist approach* posits that the prior states are deterministic. In other words, it ignores any available additional knowledge about source occurrence. Only the observation is modeled as random.
- The *Bayesian approach* also incorporates the statistics/knowledge of the source occurrence: the knowledge about the source is modeled by a probability distribution called prior. The inference utilizes the Bayesian rule.

In the remainder, I follow the Bayesian approach and seek for an underlying stochastic discrete-time model. I show how it derives from the acoustic-wave model described by a partial differential equation[10] (cf. a static formulation[11]). The final discrete-time model features loosely coupled state variables. Due to the loose coupling, these models are interpreted as reduced-order models[12].

Furthermore, I am seeking for a Bayesian lower bound for above's estimators. Sequential Cramér-Rao (SCR) bounds were developed[13] for continuous random states. However, the Bayesian bound shall be applicable jointly to discrete and continuous random state variables. Additionally, the bound shall support the corresponding probability densities with finite alphabet. It turns out that the regularity conditions for the applicability of the Bayesian SCR bound are too restrictive for discrete (quantized) states [cf. Duan et al. 2008b for the non-sequential Cramér-Rao (CR) bound].

[5] Jensen et al. 2011.
[6] Mattheij et al. 2005.
[7] Zhdanov 2002.
[8] Chow 2007.
[9] Dalang et al. 1998.

[10] Sawo 2009; Xaver et al. 2011.
[11] Reise et al. 2012.
[12] Mohammadi et al. 2012.
[13] Khan et al. 2008; Tichavsky et al. 1998.

This requirement guides us first to the Weiss-Weinstein (WW) bound[14]. The temporal evolution of states is described by a state-space model and motivates the extension of the WW bound to a sequential[15] formulation [sequential Weiss-Weinstein (SWW)]. My approach differs from earlier hybrid state-estimation formulations[16], which evaluate the SCR bound for continuous states depending on the discrete states. Apart from the underlying theory of SWW bounds[17] and the application to fault-prone systems[18], I am not aware of any explicit analytic results for specific probability densities nor their rigorous derivations.

[14] Renaux 2007; Renaux et al. 2008; Tran et al. 2012; Weiss et al. 1988.
[15] Rapoport et al. 2004b; Reece et al. 2005.
[16] Ristic et al. 2004; Washburn et al. 1985.
[17] Rapoport et al. 2007a; Reece et al. 2005.
[18] Rapoport et al. 2004a, 2007b.

1.1 Contribution

I demonstrate a decentralized localization approach of an acoustic source in an acoustic field. I address particle filtering which leads to a decentralized algorithm utilizing a consensus approach. Since my model contains discrete and non-Gaussian distributions, I use a generalization of the Bayesian SCR bound as performance bound for my estimator. The communication load demands for source coding of the exchanged signals inducing quantization noise. Finally, I join these issues to analyze my decentralized estimation approach.

Previous work on decentralized distributed estimation is summarized by Sawo 2009,

Starting point

Sawo, F. (2009). "Nonlinear state and parameter estimation of spatially distributed systems". PhD thesis. Universität Karlsruhe (see pp. 2, 3).

and on SWW bounds by Rapoport et al. 2004b,

Rapoport, I. and Y. Oshman (Dec. 2004b). "Recursive Weiss-Weinstein lower bounds for discrete-time nonlinear filtering". In: *43rd IEEE Conf. on Decision and Control*. Vol. 3, pp. 2662–2667 (see pp. 3, 49, 53, 92).

The latter paper provides a general integral-formulation of SWW bounds. My thesis is based on the following contributions I published (submitted):

Papers that directly contribute to my dissertation

Xaver, F., C. F. Mecklenbräuker, P. Gerstoft, and G. Matz (Nov. 2010). "Distributed state and field estimation using a particle filter". In: *Proc. 44th Asilomar Conf. Signals, Syst., Comput.* Pacific Grove, CA, pp. 1447–1451.

Xaver, F., G. Matz, P. Gerstoft, and C. F. Mecklenbräuker (2011). "Localization of acoustic sources using a decentralized particle filter". In: *EURASIP JWCN* 2011.1, 94ff (see pp. 2, 13, 21, 64).

Xaver, F., G. Matz, P. Gerstoft, and N. Görtz (Nov. 2012a). "Localization of acoustic sources utilizing a decentralized particle filter". In: *Proc. 46th Asilomar Conf. Signals, Syst., Comput.* Pacific Grove, CA (see p. 71).

Xaver, F., G. Matz, P. Gerstoft, and C. F. Mecklenbräuker (Mar. 2012b). "Predictive state vector encoding for decentralized field estimation in sensor networks". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. Kyoto, JP, pp. 2661–2664 (see pp. 37, 72).

1.2 *Organization of My Thesis*

Preliminaries & models

In Chapter 2, I summarize results of the measure theory in the probabilistic context, the Bayesian estimation, and particle filtering. In Chapter 3, I describe the stochastic wave equation and its approximation by finite differences. The subsequent chapters contribute to following three research areas (cf. Fig. 1.2):

Sequential Bayesian estimators

I present a particle filter (PF) on top of decentralized model, termed the decentralized distributed particle filter (DDPF), in Chapter 4 (2nd branch of Fig. 1.2).

Bayesian performance bounds

I derive analytic solutions of the linear SWW bound on the error covariance of any Bayesian filter in Chapter 6. The partly non-linear decentralized model and estimator demands for a generalization in Chapter 7 (2nd branch of Fig. 1.2). Some proofs depend on Lemmas from Appendix A. In Appendix B I prove that the CR bound is a special case of the WW and Bobrovsky-Zakai (BZ) bound.

Communication

Chapter 5 presents the whitening of the transmitted signals whereas Chapter 7 analyzes the influence of the whitening error on the global SWW bound. The definition of the error is derived in Appendix C.

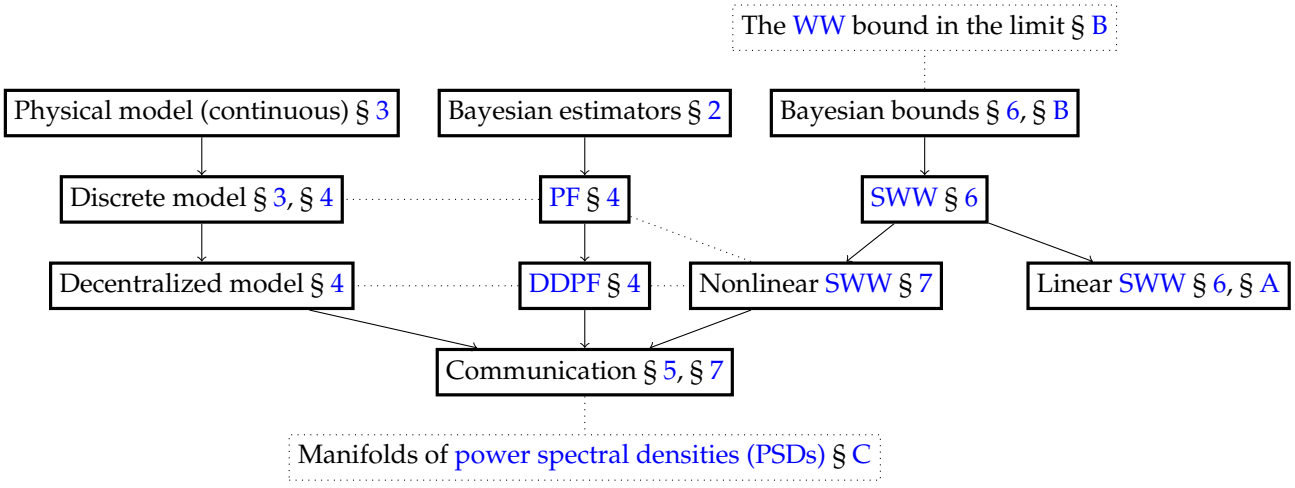


Figure 1.2: Topics and organization (three main branches).

2
Preliminaries

The purpose of this chapter is introducing notations, definitions, and theory.

2.1 Probability Theory

Let us assume a probability space $(\mathbb{R}^N, \mathcal{B}, P_x)$ with the sample space $\mathbb{R}^N$, the Borel algebra $\mathcal{B}$ and the measure $P_x : \mathcal{B} \to [0,1]$.

The expectation of a function $g(\boldsymbol{x})$ is defined using the probability measure $P_x(\mathcal{B}) = P\{\boldsymbol{x} \in \mathcal{B}\}$ by

$$E_x g(\boldsymbol{x}) \triangleq \int_{\mathbb{R}^N} g(\boldsymbol{x}) \mathrm{d}P_x(\boldsymbol{x}) \,. \tag{2.1}$$

Let us assume a probability measure consisting[1] of a continuous P_x^{c} and a discrete P_x^{d} part[2], i.e.,

$$P_x = c_1 P_x^{\mathrm{c}} + c_2 P_x^{\mathrm{d}} \,, \tag{2.2}$$

with $c_1 + c_2 = 1, c_1 \in [0,1]$. Inserting (2.2) into (2.1), the latter one splits into one integral with Lebesgue measure $\lambda^{\mathrm{c}}([a_1, b_1] \times \cdots \times [a_N, b_N]) = (b_1 - a_1) \cdots (b_N - a_N)$ and another one with counting measure $\lambda_C^{\mathrm{d}}(\mathcal{A}) = \sum_{\ell \in C} \mathbb{1}_\ell(\mathcal{A})$ where $C \in \mathcal{B}$ and $\mathbb{1}$ is the indicator function. We arrive at

$$\begin{aligned} &\int_{\mathbb{R}^N} c_1 g(\boldsymbol{x}) f_x(\boldsymbol{x}) \mathrm{d}P^{\mathrm{c}}(\boldsymbol{x}) + \int_{\mathbb{R}^N} c_2 g(\boldsymbol{x}) p_x(\boldsymbol{x}) \mathrm{d}\, \lambda_C^{\mathrm{d}}(\boldsymbol{x}) \\ &\qquad = \int_{\mathbb{R}^N} c_1 g(\boldsymbol{x}) f_x(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \sum_{\boldsymbol{x} \in C} c_2 g(\boldsymbol{x}) p_x(\boldsymbol{x}) \end{aligned} \tag{2.3}$$

with the probability density function (PDF) $f_x(\boldsymbol{x}) = \mathrm{d}P_x^{\mathrm{c}}(\boldsymbol{x}) / \mathrm{d}\, \lambda^{\mathrm{c}}(\boldsymbol{x})$ and the probability mass function (PMF) $p_x(\boldsymbol{x}) = \mathrm{d}P_x^{\mathrm{d}}(\boldsymbol{x}) / \mathrm{d}\, \lambda_C^{\mathrm{d}}(\boldsymbol{x})$. The derivative denotes the Radon-Nikodym derivative .

Furthermore, an adapted random process $\{\boldsymbol{x}(t) : t \in \mathbb{R}_+\}$ or $\{\boldsymbol{x}_k = x[k] : k \in \mathbb{N}_0\}$ is defined on an filtered probability space $(\mathbb{R}^N, \mathcal{B}, P_x)$ with a filtration $\mathcal{F}_t$ (or $\mathcal{F}_k$) contained in $\mathcal{B}$. If $t_1, t_2 \in \mathbb{R}_+$ with $t_1 < t_2$ then $\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}$. Similar for discrete-time processes.

In the following, we denote a hybrid continuous/discrete probability density by

$$v_x = c_1 f_x(\boldsymbol{x}) + c_2 p_x(\boldsymbol{x}), c_1 \in [0,1], c_1 + c_2 = 1 \tag{2.4}$$

[1] I neglect the singular continuous measure.

[2] Billingsley 2012; Burk 2007; Meintrup et al. 2004.

and call it simply probability density (PD). Especially when no measure is specified this notation allows the consideration of continuous and discrete random variables. We use the notation $\mathrm{d}\lambda_x$ whenever we assume the existence of a density for the random variable $\boldsymbol{x}$. To simplify notation, we use $\mathrm{E}\{\cdot\} \triangleq \mathrm{E}_{x,y}\{\cdot\}$, $f(\boldsymbol{x}) \triangleq f_x(\boldsymbol{x})$, $p(\boldsymbol{x}) \triangleq p_x(\boldsymbol{x})$ and $v(\boldsymbol{x}) \triangleq v_x(\boldsymbol{x})$. The expectation operator defines an inner product $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \mathrm{E}\left\{\boldsymbol{x}_1 \boldsymbol{x}_2^{\mathrm{T}}\right\}$ which in turn induces a norm $\|\boldsymbol{x}\| = \sqrt{\mathrm{E}\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}}$. Both satisfy the Cauchy-Schwarz inequality.

2.2 *Bayesian Estimation*

To perform Bayesian estimation (e.g. maximum a-posteriori (MAP) or minimum mean-squared error (MMSE)) of the state vector $\boldsymbol{x}_k$ given the past observations $\boldsymbol{y}_{1:k} = \begin{bmatrix} \boldsymbol{y}_1^T & \dots & \boldsymbol{y}_k{}^T \end{bmatrix}^T$, the posterior distribution $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ is computed sequentially.

Using the Bayesian theorem and the fact that $\boldsymbol{y}_{k+1}$ and $\boldsymbol{y}_{1:k}$ are conditionally independent (due to the Markov chain assumption) given $\boldsymbol{x}_{k+1}$, we have

$$\begin{aligned} f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1}) &= f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{k+1}, \boldsymbol{y}_{1:k}) \\ &= \frac{f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}, \boldsymbol{y}_{1:k}) f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k})}{f(\boldsymbol{y}_{k+1}|\boldsymbol{y}_{1:k})} \\ &= \frac{f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}) f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k})}{\int f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}) f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) \mathrm{d}\boldsymbol{x}_{k+1}} , \end{aligned} \tag{2.5}$$

which is known as the update step (cf. Fig. 2.1). While the measurement PDF $f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1})$ in (2.5) is known, $f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k})$ needs to be computed via the so-called prediction step,

$$f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) = \int f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \mathrm{d}\boldsymbol{x}_k . \tag{2.6}$$

Here, the transition PDF $f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$ is known and $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ has been computed in the previous time step $k-1$.

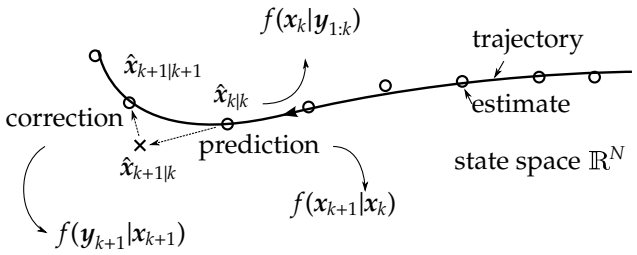


Figure 2.1: Bayesian estimation.

The MMSE estimate is defined by

$$\hat{\boldsymbol{x}}_{\mathrm{MMSE},k}(\boldsymbol{y}_{1:k}) \triangleq \mathrm{E}\left\{\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\right\} \tag{2.7}$$

whereas the MAP estimate is

$$\hat{\boldsymbol{x}}_{\mathrm{MAP},k}(\boldsymbol{y}_{1:k}) \triangleq \arg\max_{\boldsymbol{x}_k} f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) . \tag{2.8}$$

Since the integral in (2.6) is intractable , it is approximated using a Monte-Carlo technique known as importance sampling.

2.3 Importance Sampling and Monte-Carlo integration

Consider the expectation of a function $g(\boldsymbol{x}_k)$ of a random vector $\boldsymbol{x}_k$ under the measure $P(\boldsymbol{x}_k)$. Then the expectation is defined by

$$I \triangleq \mathrm{E}_{\boldsymbol{x}_k}\left\{g(\boldsymbol{x}_k)\right\} \triangleq \int g(\boldsymbol{x}_k)\mathrm{d}P(\boldsymbol{x}_k)\,. \tag{2.1}$$

It is convenient to integrate over another measure $Q(\boldsymbol{x}_k)$. Rewriting (2.1) gives

$$\int g(\boldsymbol{x}_k)\frac{\mathrm{d}P(\boldsymbol{x}_k)}{\mathrm{d}Q(\boldsymbol{x}_k)}\mathrm{d}Q(\boldsymbol{x}_k) = \int g(\boldsymbol{x}_k)\tilde{w}(\boldsymbol{x}_k)\mathrm{d}Q(\boldsymbol{x}_k) \tag{2.9}$$

where the density

$$\tilde{\omega}(\boldsymbol{x}_k) \triangleq \frac{\mathrm{d}P(\boldsymbol{x}_k)}{\mathrm{d}Q(\boldsymbol{x}_k)} = \frac{f(\boldsymbol{x}_k)}{q(\boldsymbol{x}_k)} \tag{2.10}$$

is the Radon-Nikodym derivative and hence a PD. density $q(\boldsymbol{x}_k)$ is termed the importance function. Observe that the support of $f(\boldsymbol{x}_k)$ must contain the support of $q(\boldsymbol{x}_k)$. In other words, the Radon-Nikodym derivative has to exist. The Integral (2.9) can be integrated by a Monte-Carlo approach, i.e.

$$I = \int g(\boldsymbol{x}_k)\tilde{\omega}(\boldsymbol{x}_k)\mathrm{d}Q(\boldsymbol{x}_k) \approx \frac{1}{L}\sum_{l=1}^{L} g(\boldsymbol{x}_k^{[l]})\tilde{\omega}(\boldsymbol{x}_k^{[l]}) = \hat{I} \tag{2.11}$$

where $\hat{I}$ is the Monte-Carlo estimate. Vectors $\boldsymbol{x}_k^{[l]}$ are drawn from $q(\boldsymbol{x}_k)$. Estimate $\hat{I}$ is unbiased, i.e.

$$\begin{aligned}\mathrm{E}\left\{\hat{I}\right\} &= \mathrm{E}\left\{\frac{1}{L}\sum_{l=1}^{L} g(\boldsymbol{x}_k^{[l]})\tilde{\omega}(\boldsymbol{x}_k^{[l]})\right\} \\ &= \frac{1}{L}\sum_{l=1}^{L}\mathrm{E}\left\{g(\boldsymbol{x}_k^{[l]})\tilde{\omega}(\boldsymbol{x}_k^{[l]})\right\} = \mathrm{E}\left\{g(\boldsymbol{x}_k)\tilde{\omega}(\boldsymbol{x}_k)\right\} = I\,.\end{aligned} \tag{2.12}$$

The variance is

$$\mathrm{Var}\left\{\hat{I}\right\} = \frac{1}{L^2}\sum_{l=1}^{L}\mathrm{Var}\left\{g(\boldsymbol{x}_k^{[l]})\tilde{\omega}(\boldsymbol{x}_k^{[l]})\right\} = \frac{1}{L}\mathrm{Var}\left\{g(\boldsymbol{x}_k)\tilde{\omega}(\boldsymbol{x}_k)\right\} \tag{2.13}$$

and goes to zero if L tends to infinite.

2.4 Sample-Importance-Resample Particle Filtering

A continuous posterior density $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ is approximated by a PMF with L weights $\omega_k^{[l]}$ and L particles $\boldsymbol{x}_k^{[l]}$, i.e.

$$f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \sum_{l=1}^{L}\omega_k^{[l]}\mathbb{1}_{\boldsymbol{x}_k-\boldsymbol{x}_k^{[l]}} \tag{2.14}$$

for all time k. Function $\mathbb{1}_{\cdot}$ indicates the indicator function. The normalized weights

$$\omega_k^{[l]} = \frac{\tilde{\omega}_k^{[l]}}{L} = \frac{1}{L}\frac{f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}{q(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})} \tag{2.15}$$

whereas the vectors $\boldsymbol{x}_k^{[l]}$ drawn from an importance function that is addressed later [cf. (2.11)]. Next, let us substitute (2.14) into (2.6),

$$\begin{aligned} f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) &= \int f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \mathrm{d}\boldsymbol{x}_k \\ &\approx \int f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) \left[\sum_{l=1}^{L} \omega_k^{[l]} \mathbb{1}_{\boldsymbol{x}_k - \boldsymbol{x}_k^{[l]}} \right] \mathrm{d}\boldsymbol{x}_k \\ &= \sum_{l=1}^{L} f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^{[l]}) \omega_k^{[l]} \end{aligned} \tag{2.16}$$

I now apply the importance sampling (2.9) and Monte-Carlo integration to the prediction step (2.6),

$$\begin{aligned} f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) &= \int f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) \frac{f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})}{q(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})} q(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \mathrm{d}\boldsymbol{x}_k \\ &= \int f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) \tilde{\omega}_k(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) q(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \mathrm{d}\boldsymbol{x}_k \\ &\approx \sum_{l=1}^{L} f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^{[l]}) \omega_k^{[l]} , \end{aligned} \tag{2.17}$$

where $\boldsymbol{x}_k^{[l]}$ are drawn from $q(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$. Observe the similarity of (2.16) and (2.17). Hence, the particle filter (PF) is an importance sampling filter. The update step (2.5) becomes

$$\begin{aligned} f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1}) &\propto f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}) f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) \\ &= f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}) \sum_{l=1}^{L} f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^{[l]}) \omega_k^{[l]} . \end{aligned} \tag{2.18}$$

The PDF $f(\boldsymbol{y}_{k+1}|\boldsymbol{y}_{1:k})$ does not depend on the state vector and thus is a constant. A popular importance function (2.16)

$$q(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1}) := \sum_{l=1}^{L} f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^{[l]}) \omega_k^{[l]} \tag{2.19}$$

This leads to

$$\omega_{k+1}^{[l]} \propto \frac{f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}^{[l]}) \frac{1}{L} \sum_{l'=1}^{L} f(\boldsymbol{x}_{k+1}^{[l]}|\boldsymbol{x}_k^{[l']}) \omega_k^{[l']}}{\frac{1}{L} \sum_{l'=1}^{L} f(\boldsymbol{x}_{k+1}^{[l]}|\boldsymbol{x}_k^{[l']}) w_k^{[l']}} = f(\boldsymbol{y}_{k+1}^{[l]}|\boldsymbol{x}_{k+1}^{[l]}) \tag{2.20}$$

with the approximated posterior density

$$f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1}) \approx p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1}) = \sum_{l=1}^{L} \omega_{k+1}^{[l]} \mathbb{1}_{\boldsymbol{x}_{k+1} - \boldsymbol{x}_{k+1}^{[l]}} \tag{2.21}$$

where $w_{k+1}^{[l]}$ is the normalized right side of (2.20). The particles $\boldsymbol{x}_{k+1}^{[l]}$ are drawn from the importance function $q(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1})$.

With (2.21), the result for the MMSE PF is

$$\begin{aligned} \hat{\boldsymbol{x}}_{\mathrm{MMSE},k+1}(\boldsymbol{y}_{1:k+1}) &= \int \boldsymbol{x}_{k+1} f(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1}) \mathrm{d}\boldsymbol{x}_{k+1} \\ &\approx \sum_{l=1}^{L} \boldsymbol{x}_{k+1}^{[l]} \omega_{k+1}^{[l]} \end{aligned} \tag{2.22}$$

and for MAP PF

$$\hat{\boldsymbol{x}}_{\mathrm{MAP},k+1}(\boldsymbol{y}_{1:k+1}) = \arg\max_{\boldsymbol{x}_{k+1}} w_{k+1}^{[l]} \mathbb{1}_{\boldsymbol{x}_{k+1}-\boldsymbol{x}_{k+1}^{[l]}} \,. \tag{2.23}$$

In theory, particles $\boldsymbol{x}_{k+1}^{[l]}$ and weights could be used for the subsequent prediction step. In practice, degeneration of particles arises: After some time most particles have weights close to zero. To forestall degeneration, the predict and update step is followed by a re-sample step. This step ensures that particles $\boldsymbol{x}_{k+1}^{[l]}$ with low weight $\omega_{k+1}^{[l]}$ are dropped and that the probability that a particle survives is equal to its weight.

Fig. 2.2 plots the algorithm of one iteration of the sampling-importance-resampling (SIR) PF. Sets $\mathcal{W}_{k+1}$ and $\mathcal{X}_k$ are the set of weights and particles.

input : $\mathcal{X}_k, \boldsymbol{y}_{k+1}$
output: $\mathcal{X}_{k+1}, \mathcal{W}_{k+1}, \hat{\boldsymbol{x}}_{k+1}$
for $i = 1$ **to** L **do**
  Draw $\boldsymbol{x}_{k+1}^{[l]} \sim f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$;
  $\bar{\omega}_{k+1}^{[l]} \leftarrow f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}^{[l]})$;
  normalization of $\bar{\omega}_{k+1}^{[l]}$;
end
$\hat{\boldsymbol{x}}_{k+1}$ computed by Equation (2.22) or (2.23);
$\mathcal{X}_{k+1}, \mathcal{W}_{k+1} \leftarrow \mathrm{resample}(\mathcal{X}_{k+1}, \mathcal{W}_{k+1})$;

Figure 2.2: One iteration of the sample-importance-resample particle filter.

2.5 *Bayesian Bounds*

In the sequel, I show the derivation of the Bayesian lower bound for the mean-squared error (MSE) of any[3] Bayesian estimator[4]. Vector $\boldsymbol{x}$ is the N-dimensional parameter vector to be inferred from the perturbed measurements

$$\boldsymbol{y} = C(\boldsymbol{x}) + \boldsymbol{v} \,, \qquad \boldsymbol{x} \sim v(\boldsymbol{x}) \,, \qquad \boldsymbol{v} \sim v(\boldsymbol{v}) \,, \tag{2.24}$$

with a mapping C and measurement noise $\boldsymbol{v}$. The extension to stochastic processes $\boldsymbol{x}_k$ is the subject of Chapter 6. With the estimation $\hat{\boldsymbol{x}}(\boldsymbol{y})$, the *estimation error* is defined by

$$\boldsymbol{\varepsilon} \triangleq \hat{\boldsymbol{x}}(\boldsymbol{y}) - \boldsymbol{x} \,. \tag{2.25}$$

Let $\alpha(\boldsymbol{y})$ and $\beta(\boldsymbol{x})$ be real-valued functions, and $\tilde{\varepsilon} = \alpha(\boldsymbol{y}) - \beta(\boldsymbol{x})$ a generalization of (2.25). Furthermore, $g(\boldsymbol{x}, \boldsymbol{y})$ being a measurable function satisfying $\mathrm{E}_{\boldsymbol{x}}\{g(\boldsymbol{x}, \boldsymbol{y})\} = 0$, the inner-product

$$\mathrm{E}_{\boldsymbol{x},\boldsymbol{y}}\{\alpha(\boldsymbol{y})g(\boldsymbol{x}, \boldsymbol{y})\} = \int \alpha(\boldsymbol{y}) \int g(\boldsymbol{x}, \boldsymbol{y}) v(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\,\lambda_{\boldsymbol{x}}\, \mathrm{d}\,\lambda_{\boldsymbol{y}} = 0 \,. \tag{2.26}$$

Subtracting $\mathrm{E}_{\boldsymbol{x},\boldsymbol{y}}\{\beta(\boldsymbol{x})g(\boldsymbol{x}, \boldsymbol{y})\}$ we obtain

$$\mathrm{E}_{\boldsymbol{x},\boldsymbol{y}}\{\tilde{\varepsilon} g(\boldsymbol{x}, \boldsymbol{y})\} = -\mathrm{E}_{\boldsymbol{x},\boldsymbol{y}}\{\beta \boldsymbol{x} g(\boldsymbol{x}, \boldsymbol{y})\} \,, \tag{2.27}$$

[3] biased and unbiased compared to classical (Frequentist) estimation theory

[4] Weiss et al. 1988.

and it follows

$$\mathrm{E}_{x,y}\left\{|\tilde{\varepsilon}||g(\boldsymbol{x},\boldsymbol{y})|\right\} \geq \left|\mathrm{E}_{x,y}\left\{\beta(\boldsymbol{x})g(\boldsymbol{x},\boldsymbol{y})\right\}\right| . \tag{2.28}$$

Next, let us utilize the Hölder's inequality,

$$\mathrm{E}\left\{|xy|\right\} \leq \left(\mathrm{E}\left\{|x|^p\right\}\right)^{1/p}\left(\mathrm{E}\left\{|y|^q\right\}\right)^{1/q} \tag{2.29}$$

with $p, q \in (1,\infty)$, $1/p + 1/q = 1$, on the left side of (2.28):

$$\left(\mathrm{E}\left\{|\tilde{\varepsilon}|^p\right\}\right)^{1/p}\left(\mathrm{E}\left\{|g(\boldsymbol{x},\boldsymbol{y})|^{1-1/p}\right\}\right)^{\frac{1}{1-1/p}} \geq \left|\mathrm{E}_{x,y}\left\{\beta(\boldsymbol{x})g(\boldsymbol{x},\boldsymbol{y})\right\}\right| . \tag{2.30}$$

Taking the inequality to the power of p gives

$$\mathrm{E}\left\{|\tilde{\varepsilon}|^p\right\} \geq \frac{\left|\mathrm{E}_{x,y}\left\{\beta(x)g(\boldsymbol{x},\boldsymbol{y})\right\}\right|^p}{\left(\mathrm{E}\left\{|g(\boldsymbol{x},\boldsymbol{y})|^{1-1/p}\right\}\right)^{p-1}} . \tag{2.31}$$

For $p = 2$, a scalar $\boldsymbol{x} = x$, $\alpha(y) = \hat{x}(y)$, and $\beta(x) = x$, (2.31) is a lower bound on the error covariance matrix $\mathrm{E}\left\{|\varepsilon|^2\right\}$.

Suppose that $p > 2$,

$$g(\boldsymbol{x},\boldsymbol{y}) := \sum_{n=1}^{N} a_n g_n(\boldsymbol{x},\boldsymbol{y}) , \tag{2.32a}$$

$$\beta(\boldsymbol{x}) := \sum_{n=1}^{N} b_n \beta_n(\boldsymbol{x}) = \sum_{n=1}^{N} b_n [\boldsymbol{x}]_n , \tag{2.32b}$$

with scalars a_n and b_n. Again $\mathrm{E}_x\left\{g_n(\boldsymbol{x},\boldsymbol{y})\right\} = 0$. Substituting (2.32) into (2.31) gives

$$\boldsymbol{b}^{\mathrm{T}}\,\mathrm{E}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right\}\boldsymbol{b} \geq \frac{\left(\boldsymbol{b}^{\mathrm{T}}\,\mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}\boldsymbol{a}\right)^2}{\boldsymbol{a}^{\mathrm{T}}\,\mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}\boldsymbol{a}} \tag{2.33}$$

where $\mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}$ is a non-singular matrix, $\boldsymbol{a} \triangleq [a_1,\cdots,a_N]^{\mathrm{T}}$, $\boldsymbol{g} \triangleq [g_1,\cdots,g_N]^{\mathrm{T}}$, and $\boldsymbol{b} \triangleq [b_1,\cdots,b_N]^{\mathrm{T}}$. Since $\boldsymbol{a}$ is arbitrary, we seek for the optimal vector that maximizes the right sight of (2.33). Let us define $\boldsymbol{a}' = \mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}^{1/2}$ so that

$$\begin{aligned}\frac{\left(\boldsymbol{b}^{\mathrm{T}}\,\mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}\boldsymbol{a}\right)^2}{\boldsymbol{a}^{\mathrm{T}}\,\mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}\boldsymbol{a}} &= \frac{\left(\boldsymbol{b}^{\mathrm{T}}\,\mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}\mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}^{-1/2}\boldsymbol{a}'\right)^2}{(\boldsymbol{a}')^{\mathrm{T}}\boldsymbol{a}'} \\ &= \frac{\left(\boldsymbol{b}^{\mathrm{T}}\,\mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}\mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}^{-1/2}\mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}^{-1/2}\mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}^{\mathrm{T}}\boldsymbol{b}\right)\left((\boldsymbol{a}')^{\mathrm{T}}\boldsymbol{a}'\right)}{(\boldsymbol{a}')^{\mathrm{T}}\boldsymbol{a}'} .\end{aligned} \tag{2.34}$$

For the last equality, I have used the Cauchy - Schwarz inequality with equality if

$$\boldsymbol{a}' = \mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}^{-1/2}\mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}\boldsymbol{b} \tag{2.35}$$

or, respectively,

$$\boldsymbol{a} = \mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}^{-1}\mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}\boldsymbol{b} . \tag{2.36}$$

Combining (2.33) and (2.34) gives the lower bound of the error covariance, i.e.

$$\mathrm{E}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right\} \geqslant \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\} \mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}^{-1} \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}^{\mathrm{T}} \tag{2.37}$$

The elements of all matrices must be finite. The relational operator $\geqslant$ indicates that the difference between left and right hand sides is a positive semi-definite matrix. The function $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})$ is a *sensitivity function* termed *score* which defines specific Bayesian bounds.

3
Models

THE AIM OF THIS CHAPTER is to define a source localization problem for which, subsequently, I develop a model. Most parts were published in Xaver et al. 2011.

"The best model of a cat is another cat, or better yet, the cat itself."
— Norbert Wiener

There are two approaches. In the first, I describe the wave field deterministically by a partial differential equation (PDE) and turn the resulting discrete-time model into a stochastic one. This has the advantage that we are able to specify any type of noise and prior. In the second, I describe the wave field probabilistically by a stochastic partial differential equation (SPDE) and assume a particular Wiener noise process. Then I approximate it by finite differences to get a discrete-time model where the continuous noise induces the discrete-time noise.

In the following, I address both approaches. They lead to the same discrete-time model which is used in the subsequent chapters. I first present the acoustic PDE closely followed by the acoustic SPDE and their numerical approximations. Note that every (stochastic) PDE system can be approximated by a discrete system, but that a discrete system is not necessarily a (stochastic) PDE system in the limit. The remainder is devoted to source and measurement models.

3.1 Forward Model of the Spatio-Temporal Field

A Deterministic Forward Model

In the following, let us consider an acoustic problem characterized by the hyperbolic PDE[1,2,3,4] (scalar wave equation):

$$\frac{1}{c^2}\partial_t^2 p(\boldsymbol{r},t) - \nabla^2 p(\boldsymbol{r},t) = s(\boldsymbol{r},t), \quad \boldsymbol{r} = [x,y]^\mathrm{T} \in \Omega, \tag{3.1}$$

Here, $p(\boldsymbol{r},t)$ denotes pressure, ∂_t is the partial derivative with respect to time, ∇^2 the Laplace operator, c the speed of sound, $s(\boldsymbol{r},t)$ is the source, and $\Omega \subset \mathbb{R}^2$ is the 2-D region of interest.

[1] Jensen et al. 2011.
[2] Mattheij et al. 2005.
[3] Zhdanov 2002.
[4] Tarantola 2005.

Given a point in time and space, the light cone[5] demonstrates the temporal and spatial dependency (see Fig. 3.1). A point at the present is influenced by the region indicated by the cone below the hypersurface of the present. Moreover, this point influences the region indicated by the cone at the top. When we discretize the wave equation

[5] Minkowski 1909.

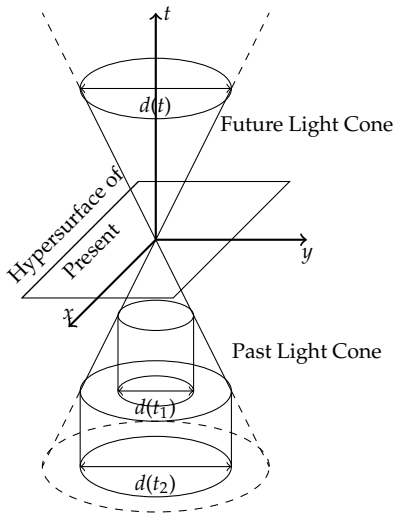


Figure 3.1: Light cone in two spatial dimensions. The cylinders shows a discrete approximation of the light cone at discrete time t_k.

[6] Morse et al. 1953, p. 842ff.

[7] Gardner 1990 provides a general introduction on processes. The spatio-temporal extension for diffusion and wave equations has been published in Chow 2007; Dalang et al. 1998; Jentzen et al. 2009, and Hausenblas 2010.

by a numerical method with sample period Δ_t, we approximate the cones by cylinders. Here, variable d denotes the diameter of influence and the time $t_2 = t_1 + \Delta_t$.

The solution of (3.1) at location $\boldsymbol{r}_\ell$ and time t is given by

$$\begin{aligned} y_{t,\ell} &\triangleq p(\boldsymbol{r}_\ell, t) \\ &= \int_\Omega \int_0^t G(\boldsymbol{r}_\ell, \boldsymbol{r}, t - t') s(\boldsymbol{r}, t') \mathrm{d}t' \mathrm{d}\boldsymbol{r} \,, \end{aligned} \tag{3.2}$$

where $G(\boldsymbol{r}_\ell, \boldsymbol{r}, t)$ is the Green function to the problem (3.1). If $\Omega = \mathbb{R}^2$ then

$$G(\boldsymbol{r}_\ell, \boldsymbol{r}, t) = \begin{cases} \frac{2c}{\sqrt{c^2(t-t_0)^2 - |\boldsymbol{r}_\ell - \boldsymbol{r}|^2}} \,, & |\boldsymbol{r}_\ell - \boldsymbol{r}| < c(t - t_0) \,, \\ 0 \,, & \text{else} \,, \end{cases} \tag{3.3}$$

and for $\Omega = \mathbb{R}$, i.e. $\boldsymbol{r} = x$,

$$G(\boldsymbol{r}_\ell, \boldsymbol{r}, t) = \begin{cases} 2\pi c \left[1 - \epsilon\left(\frac{|x - x_\ell|}{c} - (t - t_0)\right)\right] \,, & |x - x_0| < c(t - t_0) \,, \\ 0 \,, & \text{else} \,. \end{cases} \tag{3.4}$$

Distribution ϵ denotes the Heaviside step.[6]

B *Stochastic Forward Model*

This part is devoted to the stochastic counterpart of (3.1) for the Wiener noise process with covariance function Q. Let $\mathcal{H} \triangleq L^2(\Omega)$ and $(\mathbb{R}^N, \mathcal{B}, \mathcal{F}, P_x)$ be the probability space with a filtration $\mathcal{F}_t$ defined in Section 2.1. The stochastic wave equation is given by

$$\frac{1}{c^2}\partial_t^2 p(\boldsymbol{r}, t) - \nabla^2 p(\boldsymbol{r}, t) = s(\boldsymbol{r}, t) + \tilde{s}(\boldsymbol{r}, t)\partial_t W(\boldsymbol{r}, t) \,, \quad \boldsymbol{r} \in \Omega, \tag{3.5}$$

with a deterministic source $s(\boldsymbol{r}, t)$ and noise $\tilde{s}(\boldsymbol{r}, t)\partial_t W(\boldsymbol{r}, t)$ (cf. Fig. 1.1). Process $W = \{W(\boldsymbol{r}, t) : \boldsymbol{r} \in \mathbb{R}^2, t \in \mathbb{R}_+\}$ is a $\mathcal{H}$-valued Wiener process[7] with covariance function $Q(\boldsymbol{r}_1, \boldsymbol{r}_2)$ so that

$$\mathrm{E}\{W(\boldsymbol{r}_1, t_1)\} = 0 \,, \tag{3.6}$$

$$\mathrm{Covar}\{W(\boldsymbol{r}_1, t_1), W(\boldsymbol{r}_2, t_2)\} = \min(t_1, t_2) Q(\boldsymbol{r}_1, \boldsymbol{r}_2) \,. \tag{3.7}$$

with finite trace

$$\int_\Omega Q(\boldsymbol{r}, \boldsymbol{r}) \mathrm{d}\boldsymbol{r} < \infty \,. \tag{3.8}$$

In the sequel, let $0 \le t_0 < t_1 < t_2 < t_3 < t_4$. Then the variance

$$\mathrm{Var}\{W(\boldsymbol{r}, t)\} = tQ(\boldsymbol{r}, \boldsymbol{r}) \,. \tag{3.9a}$$

The mean of the spatio-temporal increments is

$$\mathrm{E}\{W(\boldsymbol{r}_2, t_2) - W(\boldsymbol{r}_1, t_1)\} = 0 \tag{3.9b}$$

and its variance

$$\begin{aligned} &\mathrm{E}\left\{(W(\boldsymbol{r}_2, t_2) - W(\boldsymbol{r}_1, t_1))^2\right\} \\ &\qquad = t_2 Q(\boldsymbol{r}_2, \boldsymbol{r}_2) + t_2 Q(\boldsymbol{r}_1, \boldsymbol{r}_1) - 2t_1 Q(\boldsymbol{r}_1, \boldsymbol{r}_2) \,. \end{aligned} \tag{3.9c}$$

The covariance of the increments is

$$\begin{aligned}&\mathrm{E}\left\{(W(\boldsymbol{r}_4,t_4)-W(\boldsymbol{r}_3,t_3))\,(W(\boldsymbol{r}_2,t_2)-W(\boldsymbol{r}_1,t_1))\right\}\\&\quad=t_2\,(Q(\boldsymbol{r}_2,\boldsymbol{r}_4)-Q(\boldsymbol{r}_2,\boldsymbol{r}_3))+t_1\,(Q(\boldsymbol{r}_1,\boldsymbol{r}_3)-Q(\boldsymbol{r}_1,\boldsymbol{r}_4))\end{aligned}\tag{3.9d}$$

which equals to zero if $\boldsymbol{r}_1=\boldsymbol{r}_2=\boldsymbol{r}_3=\boldsymbol{r}_4$.

The weak solution at location $\boldsymbol{r}_\ell$ and time t is given by the Itō process

$$\begin{aligned}y_{t,\ell}&\triangleq p(\boldsymbol{r}_\ell,t)\\&=\int_\Omega\int_0^t G(\boldsymbol{r}_\ell,\boldsymbol{r},t-t')s(\boldsymbol{r},t')\mathrm{d}t'\mathrm{d}\boldsymbol{r}\\&+\int_\Omega\int_0^t G(\boldsymbol{r}_\ell,\boldsymbol{r},t-t')\tilde{s}(\boldsymbol{r},t')W(\mathrm{d}t'\mathrm{d}\boldsymbol{r})\,.\end{aligned}\tag{3.10}$$

The first two lines are similar to the solution of the deterministic wave equation. The integral in the fourth line is a stochastic Itō[8] integral[9]. Current research avoids the Itō formula and goes beyond this standard case as summarized in Jentzen et al. 2009.

[8] The use of the Stratonovich integral would give an additional term in (3.10).

[9] Burk 2007; Kloeden et al. 2011; Øksendal 2010.

3.2 *Numerical Solutions*

Equations (3.2) and (3.10) represent the so called measurement model. It has one important disadvantage: The integral $\int_0^t \mathrm{d}t$ has to be computed with an increasing computational effort over time.

Some authors[10] circumvent this problem and approximate (3.10) for one source by

[10] Hlinka et al. 2012.

$$y_{t,\ell}\approx\frac{s'(\boldsymbol{r}_\mathrm{s},t-(1/c)|\boldsymbol{r}_\ell-\boldsymbol{r}_\mathrm{s}(t)|)}{|\boldsymbol{r}_\ell-\boldsymbol{r}_\mathrm{s}(t)|}+\text{noise}\,,\tag{3.11}$$

with source function s' and its location $\boldsymbol{r}_\mathrm{s}$. They use a discretized version as measurement equation and a time-evolution model of the source location $\boldsymbol{r}_\mathrm{s}$ as transition equation. This approach ignores wave phenomenons.

Instead, I use the forward model describing the evolution of the wave field in the sequel. Each time step a sensor measures the field, only a snapshot at the sensor's position is taken. Both, the forward model and the measurement model form a state-space model. Until now, I have only considered continuous-time problems. In the remainder of this subsection, I summarize the finite-difference method (FDM) to get a discrete representation.

To obtain a space-time discrete model, the differential operators are approximated by finite differences (FDM), see Figure 3.2. We assume a rectangular region in two dimensions (i.e., $\boldsymbol{r}=(x,y)$) and use a spatial sampling set given by the finite square lattice $\mathcal{L}=\{(i\Delta_r,j\Delta_r):\ i=1,\ldots,I,\ j=1,\ldots,J\}$, where Δ_r is the spatial sampling interval. For simplicity, we assume identical sampling intervals in both coordinates, but using different sampling intervals for each coordinate is straightforward (Different sampling intervals influence the accuracy of the field approximation only but not the principal features of the

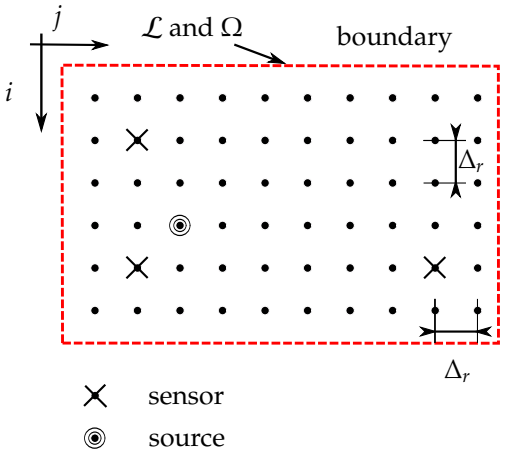


Figure 3.2: Lattice due to the FDM with boundaries, sources, and sensors. Set $\mathcal{L}$ is the sampling lattice while Ω denotes the area.

decentralized estimator). For simplicity, we assume that there are R sensors whose locations form a subset $\mathcal{R}$ of the lattice $\mathcal{L}$.

First, we introduce the auxiliary function $q(\boldsymbol{r}, t)$. Then we recast (3.5) to

$$\frac{1}{c^2}\partial_t q(\boldsymbol{r}, t) - \nabla^2 p(\boldsymbol{r}, t) = s(\boldsymbol{r}, t) + \tilde{s}(\boldsymbol{r}, t)\partial_t W(\boldsymbol{r}, t)\,, \quad \boldsymbol{r} \in \Omega, \tag{3.12a}$$

$$\partial_t p(\boldsymbol{r}, t) = q(\boldsymbol{r}, t)\ . \tag{3.12b}$$

For the Laplace operator, we then obtain the discrete approximation

$$\begin{aligned}\nabla^2 p(i\Delta_r, j\Delta_r, t) &\approx \underset{\Delta_r,\Delta_r}{\overset{2}{\Delta}}{}_{x,y} p(i\Delta_r, j\Delta_r, t) \\ &\triangleq \frac{1}{\Delta_r^2}\Big(p\big((i-1)\Delta_r, j\Delta_r, t\big) \\ &\quad + p\big((i+1)\Delta_r, j\Delta_r, t\big) + p\big(i\Delta_r, (j-1)\Delta_r, t\big) \\ &\quad + p\big(i\Delta_r, (j+1)\Delta_r, t\big) - 4p(i\Delta_r, j\Delta_r, t)\Big).\end{aligned} \tag{3.13}$$

For the first-order temporal derivative we have

$$\begin{aligned}\partial_t p(i\Delta_r, j\Delta_r, t) &\approx \underset{\Delta_t}{\overset{1}{\Delta}}{}_t p(i\Delta_r, j\Delta_r, t) \\ &\triangleq \frac{1}{\Delta_t}\left(p(i\Delta_r, j\Delta_r, (k+1)\Delta_t) - p(i\Delta_r, j\Delta_r, k\Delta_t)\right)\,,\end{aligned} \tag{3.14}$$

similar for q. Here, k is the discrete time index, and Δ_t is the temporal sampling period. It is upper bounded by Δ_r/c to ensure numerical stability. The right choice of Δ_t is beyond the scope of our paper, so that we refer our reader to Jensen et al. 2011.

The finite difference approach is also applicable to SPDEs[11]. Equation (3.12) with (3.13) and (3.14) gives the discrete model

$$\begin{aligned}q[i, j, k+1] &= q[i, j, k] \\ &\quad + \frac{\Delta_t c^2}{\Delta_r}\left(p[i+1, j, k] + p[i-1, j, k]\right. \\ &\quad \left. + p[i, j+1, k] + p[i, j-1, k] - 4p[i, j, k]\right) \\ &\quad + \Delta_t c^2 s[i, j, k] + c^2\tilde{s}[i, j, k](W[i, j, k+1] - W[i, j, k])\,,\end{aligned} \tag{3.15a}$$

$$p[i, j, k+1] = p[i, j, k] + \Delta_t q[i, j, k]\,. \tag{3.15b}$$

[11] Jentzen et al. 2009; Kloeden et al. 2011; McDonald 2007; Walsh 2006.

The main difference to the ordinary FDM is that we only need to analyze the random temporal increments

$$\Delta_t W(\boldsymbol{r}, t) = W(\boldsymbol{r}, \Delta_t k + \Delta_t) - W(\boldsymbol{r}, \Delta_t k) \tag{3.16}$$

in the last term of (3.15a). Due to (3.9) the mean of the temporal increments is

$$\mathrm{E}\left\{\Delta_t W(\boldsymbol{r}, t)\right\} = \mathrm{E}\left\{\Delta_x W(x, y, t)\right\} = \mathrm{E}\left\{\Delta_y W(x, y, t)\right\} = 0 \tag{3.17}$$

and the variance is

$$\mathrm{E}\left\{(\Delta_t W(\boldsymbol{r}, t))^2\right\} = \Delta_t Q(\boldsymbol{r}, \boldsymbol{r})\,, \tag{3.18a}$$

$$\begin{aligned}\mathrm{E}\left\{(\Delta_x W(x, y, t))^2\right\} &= tQ((x + \Delta_x, y), (x + \Delta_x, y)) + tQ((x, y), (x, y)) \\ &\quad - 2tQ((x + \Delta_x, y), (x, y))\end{aligned} \tag{3.18b}$$

$$\begin{aligned}\mathrm{E}\left\{(\Delta_y W(x, y, t))^2\right\} &= tQ((x, y + \Delta_y), (x, y + \Delta_y)) + tQ((x, y), (x, y)) \\ &\quad - 2tQ((x, y + \Delta_y), (x, y))\end{aligned} \tag{3.18c}$$

In my thesis, I only use the FDM, which is a fast numerical approximation. Another more advanced method is the finite-element method (FEM)[12] or the promising spectral-element method (SEM)[13].

[12] Hausenblas 2010; Kovacs et al. 2010.

[13] Komatitsch et al. 2005.

3.3 Initial and Boundary Conditions

Hereafter, let the initial conditions be

$$p(\boldsymbol{r}, t) = 0, \quad \boldsymbol{r} \in \Omega,\ t = 0, \tag{3.19}$$

$$\partial_t p(\boldsymbol{r}, t) = 0, \quad \boldsymbol{r} \in \Omega,\ t = 0. \tag{3.20}$$

I address two boundary conditions (cf. Fig. 1.1)

$$\frac{1}{c}\partial_t p(\boldsymbol{r}, t) - \nabla p(\boldsymbol{r}, t) \cdot \boldsymbol{n} = 0, \quad \boldsymbol{r} \in \partial\Omega_1, \tag{3.21}$$

$$\partial_t p(\boldsymbol{r}, t) = 0, \quad \boldsymbol{r} \in \partial\Omega_2, \tag{3.22}$$

to model a hallway. I denote the inner-product by $\cdot$ and the gradient by ∇. Notation $\partial\Omega_1$ is the transparent part of the boundary of Ω (with normal vector $\boldsymbol{n}$) modeling an infinite domain for the behind uncovered area. Boundary $\partial\Omega_2$ (disjoint from $\partial\Omega_1$) models walls. The choice[14] of these boundary conditions indeed affects the resulting state-space model but does not change the general formulation of the decentralized approach.

[14] Another boundary condition

$$p(\boldsymbol{r}, t) = 0, \quad \boldsymbol{r} \in \partial\Omega_3\,,$$

models windows.

3.4 Discrete Transition Equation

We introduce the pressure vector $\boldsymbol{p}_k = \mathrm{vec}\{\boldsymbol{P}_k\}$ with $[\boldsymbol{P}_k]_{i,j} = p(i\Delta_r, j\Delta_r, k\Delta_t)$. The source vector $\boldsymbol{s}_k$ and the pressure derivative vector $\boldsymbol{q}_k$ are defined similarly. Then the vector formulation of (3.15) is given by

$$\begin{bmatrix}\boldsymbol{q}_{k+1} \\ \boldsymbol{p}_{k+1}\end{bmatrix} = \underbrace{\begin{bmatrix}\boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{21} & \boldsymbol{I}\end{bmatrix}}_{\boldsymbol{\Phi}_{\mathrm{FDM}}}\begin{bmatrix}\boldsymbol{q}_k \\ \boldsymbol{p}_k\end{bmatrix} + \Delta_t c^2 \begin{bmatrix}\boldsymbol{s}_k \\ \boldsymbol{0}\end{bmatrix} + \boldsymbol{w}_k\,. \tag{3.23}$$

The diagonal matrix $\mathbf{\Phi}_{11}$ results from the boundary condition (3.21). Its diagonal elements are

$$[\mathbf{\Phi}_{11}]_{ii} = \begin{cases} 1-2\kappa & \text{for nodes on the boundary } \partial\Omega_1, \\ 1, & \text{else,} \end{cases}$$

where $\kappa = c/\Delta_r$. Also the diagonal matrix

$$[\mathbf{\Phi}_{21}]_{ii} = \begin{cases} 1 & \text{for inner nodes and nodes on the boundary } \partial\Omega_1, \\ 0 & \text{nodes on the boundary } \partial\Omega_3 \end{cases}$$

depends on the boundary condition (3.22). Similarly, the sparse matrix $\mathbf{\Phi}_{12}$ stems from (3.1) and is given by

$$[\mathbf{\Phi}_{12}]_{i,j} = \begin{cases} -4\kappa^2, & i = j, \\ 2\kappa^2, & |i-j| = 1 \text{ for nodes on } \partial\Omega_1, \\ \kappa^2, & |i-j| = 1 \vee |i-j| = I \text{ for inner nodes,} \\ 0, & \text{else.} \end{cases}$$

The vector $\boldsymbol{s}_k$ is the sampled source modeled in the following section. In case of an approximated SPDE,

$$\boldsymbol{w}_k = \begin{bmatrix} \boldsymbol{w}_k^{\mathrm{q}} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} W[1,1,k+1] - W[1,1,k] \\ \cdots \\ W[I,J,k+1] - W[I,J,k] \\ \mathbf{0} \end{bmatrix}, \tag{3.24}$$

where $\boldsymbol{w}_k^{\mathrm{q}}$ models noise on top of the source and the lack of knowledge of the field. In case of an approximated PDE, we are free to model the noise. I generalize (3.24) to

$$\boldsymbol{w}_k \triangleq \begin{bmatrix} \boldsymbol{w}_k^{\mathrm{q}} \\ \boldsymbol{w}_k^{\mathrm{p}} \end{bmatrix}, \tag{3.25}$$

with $\boldsymbol{w}_k^{\mathrm{p}}$ only modeling the lack of knowledge and use this in the sequel. The lack of knowledge of the field may have different causes:

- Inhomogeneous material (e.g. air),
- Unknown boundary conditions,
- Approximation errors,
- Neglect sources

Furthermore, an additional noise term enables the use of Bayesian performance bounds (see Sections 6.9 and 7.4).

3.5 *Source and Location Models*

We assume that there are S sources whose positions form a subset $\mathcal{T}_k$ of the discretization lattice $\mathcal{L}$, i.e.,

$$s[i,j,k] = \sum_{l=1}^{S} s_0[k-k_\ell]\,\delta(i-i_\ell, j-j_\ell)\,, \tag{3.26}$$

where δ is the Delta distribution. Function $s_0[k]$ is known, but the positions (i_ℓ, j_ℓ) and the activation/occurrence times k_ℓ are unknown. In the remainder of my thesis I use two different models: Source Model I gives rise to a decentralized estimator in Chapter 4 whereas Source Model II is more convenient for the performance bound of estimators in Chapter 7. If the number of sources $S = 1$, both models are similar. Both models implicitly assume that S does not change with time.

A Source Model I

I use Model I in Chapter 4 for the algorithm of a decentralized estimator. It features a distributed state $n[i, j, k]$ and is suitable for small S.

Here, the unknowns i_ℓ, j_ℓ and k_ℓ are captured via the integer variables $n[i, j, k]$ that describe, for a lattice point (i, j), the time between the source occurrence and the current time instant k, i.e., for the lth source there is $n[i_\ell, j_\ell, k] = \max\{k - k_\ell, 0\}$. If there is no source at position (i, j), then $n[i, j, k] = 0$. Clearly, the source life span satisfies the state transition equation

$$n[i, j, k+1] = \begin{cases} n[i, j, k] + 1, & (i, j) \in \mathcal{T}_k, \\ 0, & \text{else}, \end{cases}$$

where $\mathcal{T}_k = \{(i_\ell, j_\ell) | k \geq k_l\}$ is the set of sources active at time k. Arranging the variables $n[i, j, k]$ into a vector $\boldsymbol{n}_k$ similarly to $\boldsymbol{p}_k$, $\boldsymbol{q}_k$, and $\boldsymbol{s}_k$, we obtain

$$\boldsymbol{n}_{k+1} = \boldsymbol{n}_k + \mathbb{1}_{\mathcal{T}_k}, \tag{3.27}$$

where the elements of $\mathbb{1}_{\mathcal{T}_k}$ are zero or one depending on whether a source is active at the corresponding position and at time instant k, i.e.,

$$[\mathbb{1}_{\mathcal{T}_k}]_{i+(j-1)I} \triangleq \begin{cases} 1, & (i, j) \in \mathcal{T}_k, \\ 0, & \text{else}. \end{cases} \tag{3.28}$$

Function $\mathbb{1}_\cdot$ is the vector valued indicator function. Note that the state vector $\boldsymbol{n}_k$ has at most S non-zero elements. Using the convention $s_0[0] = 0$, the source vector $\boldsymbol{s}_k$ in (3.23) is rewritten as

$$\boldsymbol{s}_k \triangleq \boldsymbol{s}_0[\boldsymbol{n}_k], \tag{3.29}$$

thereby linking the state-equation (3.27) and the forward model (3.23).

Since the source function depends on time and space, these quantities suffer from noise and are modeled in the following: The temporal noise models the perturbation of a source's lifespan by an additional term in (3.27) while this is not possible for the spatial perturbation. This is due to the fact that the position of sources are coded into the sub-vector $\boldsymbol{n}_k$ by placing its elements. From a practical perspective this is done by a time dependent matrix $\boldsymbol{D}_k$ which displaces the elements of a vector to other positions (jitter) according to the mapping between grid and sub-vector $\boldsymbol{n}_k$.

Equation (3.27) becomes

$$\boldsymbol{n}_{k+1} = \boldsymbol{D}_k\Big(\boldsymbol{n}_k + \mathbb{1}_{\mathcal{T}_k} + \mathbb{1}_{\mathcal{T}_k} \odot \boldsymbol{n}'\Big). \tag{3.30}$$

Here, $\boldsymbol{n}'$ is a random integer perturbation, $\odot$ is the Hadamard (element-wise) product, and the lth column of the displacement matrix $\boldsymbol{D}_k$ is given by $\boldsymbol{e}_{l+d(l)}$, with the canonical column unit vector

$$[\boldsymbol{e}_l]_n \triangleq \begin{cases} 1, & l = n, \\ 0, & \text{else}, \end{cases}$$

and a random integer jitter $d(l)$ whose probability mass is concentrated about zero.

Because of linearity, (3.30) is rewritten as

$$\boldsymbol{n}_{k+1} = \boldsymbol{D}_k\boldsymbol{n}_k + \boldsymbol{D}_k\mathbb{1}_{\mathcal{T}_k} + \boldsymbol{D}_k\mathrm{diag}\,(\mathbb{1}_{\mathcal{T}_k})\,\boldsymbol{n}'. \tag{3.31}$$

B Source Model II

I use Model II in Chapter 7 to analyze the algorithm of the estimator introduced in Chapter 4.

Here, the positions (i_ℓ, j_ℓ) and the lifespan n_ℓ of S sources are stacked into a discrete vector $\boldsymbol{x}_k^{\mathrm{d}}$. Then

$$\boldsymbol{s}_k \triangleq \boldsymbol{s}_0(\boldsymbol{x}_k^{\mathrm{d}}) \triangleq [s_0[i_1, j_1, n_1], \cdots, s_0[i_S, j_S, n_S]]^{\mathrm{T}}\ . \tag{3.32}$$

Random vector $\boldsymbol{x}_k^{\mathrm{d}}$ evolves over time modeled by following transition equation

$$\boldsymbol{x}_{k+1}^{\mathrm{d}} = \boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{w}_k^{\mathrm{d}}\ , \tag{3.33}$$

similar to (3.31). The distributions of the prior $\boldsymbol{x}_0^{\mathrm{d}}$ and the noise $\boldsymbol{w}_k^{\mathrm{d}}$ are discrete. The noise describes the spatial and temporal jitter of the sources.

3.6 Measurement Model

Since the evolution of the field is governed by the forward model, the measurement model picks the pressure at the location of the sensors. The actual measurements $\boldsymbol{y}_k$ are given by noisy samples of the pressure field at the sensor positions $(i_l', j_l') \in \mathcal{R}$. Thus the measurement equation is

$$\boldsymbol{y}_k = \tilde{\boldsymbol{C}}\boldsymbol{x}_k + \boldsymbol{v}_k = \boldsymbol{C}\boldsymbol{p}_k + \boldsymbol{v}_k\ , \tag{3.34}$$

where $\boldsymbol{v}_k$ denotes measurement noise and

$$\tilde{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{C} & \boldsymbol{0} \end{bmatrix}\ , \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{e}_{i_1'+(j_1'-1)I}^{\mathrm{T}} \\ \vdots \\ \boldsymbol{e}_{i_R'+(j_R'-1)I}^{\mathrm{T}} \end{bmatrix}, \tag{3.35}$$

with $\boldsymbol{e}_l$ denoting the lth unit vector.

4
Decentralized Particle Filtering

This chapter is devoted to the decentralized and distributed particle filter (PF) published in Xaver et al. 2011. My approach features the use of the discretized wave equation. I open with a central augmented state-space model followed by a decentralization step. This provides a basis for a decentralized maximum a-posteriori (MAP) PF that uses a consensus algorithm. Finally, I present the overall algorithm and some simulation results.

4.1 Augmented State-Space Model

Model I from Section 3.5 offers a spatially distributed location vector that is convenient for the following decentralized PF. For that reason, I first combine the state-space model (3.23) with (3.29) and (3.31) to obtain an augmented state-space model for the extended state vector

$$\boldsymbol{x}_k = \begin{bmatrix} \boldsymbol{q}_k \\ \boldsymbol{p}_k \\ \boldsymbol{n}_k \end{bmatrix},$$

i.e.,

$$\boxed{\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{\Phi}_k \boldsymbol{x}_k + \boldsymbol{\Gamma}_k \boldsymbol{u}_k + \boldsymbol{G}_k \boldsymbol{n}'_k\,, \\ \boldsymbol{y}_k &= \boldsymbol{C}\boldsymbol{p}_k + \boldsymbol{v}_k \end{aligned}} \qquad \begin{aligned} &(4.1) \\ &(3.34) \end{aligned}$$

with

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} & \boldsymbol{0} \\ \Delta_t \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D}_k \end{bmatrix}, \quad \boldsymbol{\Gamma}_k = \begin{bmatrix} \Delta_t c^2 \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D}_k \end{bmatrix}, \tag{4.2}$$

and

$$\boldsymbol{G}_k = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{D}_k \operatorname{diag}(\mathbb{1}_{\mathcal{T}_k}) \end{bmatrix}, \quad \boldsymbol{u}_k = \begin{bmatrix} s_0[\boldsymbol{n}_k] \\ \boldsymbol{0} \\ \mathbb{1}_{\mathcal{T}_k} \end{bmatrix}. \tag{4.3}$$

Note that non-linearity is inherent in (4.1).

4.2 Bayesian Estimation

In this section, I recall some important facts of Bayesian estimation and particle filtering primarily introduced in Sections 2.2 and 2.4. Furthermore, this section addresses the implementation of a PF.

A Particle Filter

The approximate sequential computation of the posterior distribution $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ based on importance sampling using the transition probability density function (PDF) $f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ as importance (or, proposal) distribution $q(\boldsymbol{x}_k)$ leads to the particle filter. Here, the desired probability mass functions (PMFs) are approximated in terms of particles, i.e., samples $\boldsymbol{x}_k^{[l]}$ and associated weights $\omega_k^{[l]}$, hence

$$f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \sum_{l=1}^{L} \omega_k^{[l]} \mathbb{1}_{\boldsymbol{x}_k - \boldsymbol{x}_k^{[l]}}, \tag{4.4}$$

where L is the number of particles. The new samples for the subsequent time instant are generated using the *proposal distribution*[1]

[1] Also known as importance function

$$q(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k = \boldsymbol{x}_k^{[l]}),$$

where for the generation of each new particle $\boldsymbol{x}_{k+1}^{[l]}$ the previous particle $\boldsymbol{x}_k^{[l]}$ is chosen randomly with probability $\omega_k^{[l]}$. Sampling from $q(\boldsymbol{x}_{k+1})$ can be achieved by generating a noise realization $\boldsymbol{w}_k^{[l]}$ and invoking the state transition equation (4.1), i.e.,

$$\boldsymbol{x}_{k+1}^{[l]} = \boldsymbol{\Phi}_k^{[l]} \boldsymbol{x}_k^{[l]} + \boldsymbol{\Gamma}_k^{[l]} \boldsymbol{u}_k^{[l]} + \boldsymbol{G}_k^{[l]} \boldsymbol{n}'^{[l]}_k \,. \tag{4.5}$$

Vector $\boldsymbol{u}_k^{[l]}$ can be computed from the particle $\boldsymbol{x}_k^{[l]}$ according to (4.3). The dependency of the matrices on k issues from spatial noise.

The unnormalized weight for each new particle is

$$\tilde{\omega}_{k+1}^{[l]} = \omega_k^{[l]} f(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}^{[l]}) = \omega_k^{[l]} f_v(\boldsymbol{y}_{k+1} - \tilde{\boldsymbol{C}} \boldsymbol{x}_{k+1}^{[l]}), \tag{4.6}$$

where $f_v(\boldsymbol{v}_k)$ is the distribution of the measurement noise and we used the measurement equation (3.34). For i.i.d. Gaussian measurement noise with variance σ_v^2

$$\tilde{\omega}_{k+1}^{[l]} = \omega_k^{[l]} \exp\Big(- \frac{1}{2\sigma_v^2} \big\| \boldsymbol{y}_{k+1} - \tilde{\boldsymbol{C}} \boldsymbol{x}_{k+1}^{[l]} \big\|_2^2 \Big). \tag{4.7}$$

Once all unnormalized weights have been obtained, the actual weights are computed via the normalization $\omega_{k+1}^{[l]} = \tilde{\omega}_{k+1}^{[l]} / \sum_{l'=1}^{M} \tilde{\omega}_{k+1}^{[l']}$. Particle filters suffer from a general problem termed sample degeneracy, i.e., after some time only few particles have non-negligible weights. This problem is circumvented using resampling[2]. With sampling-importance-resampling (SIR), new samples are drawn from the distribution $\sum_{l=1}^{L} \omega_k^{[l]} \mathbb{1}_{\boldsymbol{x}_k - \boldsymbol{x}_k^{[l]}}$ and all weights are identical, i.e., $\omega_k^{[l]} = 1/L$ (cf. Fig. 2.2.).

[2] Hol et al. 2006.

To obtain initial particles $\boldsymbol{x}_0^{[l]}$ samples of the state vector is needed. S random realizations of source positions and activation times are generated according to the prior distributions. Then we apply the noise-free version of the state-space model (4.1) k_{start} times, i.e.,

$$\boldsymbol{x}_0^{[l]} = \boldsymbol{\Phi}^{k_{\text{start}}} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{n}_0^{[l]} \end{bmatrix} + \sum_{\ell=0}^{k_{\text{start}}-1} \boldsymbol{\Phi}^{k_{\text{start}}-1-\ell} \boldsymbol{\Gamma} \boldsymbol{u}_\ell^{[l]}, \tag{4.8}$$

where $\boldsymbol{n}_0^{[l]}$ and $\boldsymbol{u}_\ell^{[l]}$ are determined by the realizations of the source parameters [cf. (4.3) and Subsection 3.5.A]. The random variable k_{start} denotes the time duration between source occurrence and activation of the estimator.

B Source Localization

Using (4.4), the posterior PDF of $\boldsymbol{n}_k$ (i.e., the last IJ elements of $\boldsymbol{x}_k$) is approximated as

$$f(\boldsymbol{n}_k|\boldsymbol{y}_{1:k}) \approx \sum_{l=1}^{L} \omega_k^{[l]} \mathbb{1}_{\boldsymbol{n}_k - \boldsymbol{n}_k^{[l]}}. \tag{4.9}$$

(Note that $\boldsymbol{n}_k$ contains all information about position and activation time of the sources.) The probability $\mathrm{P}\{\mathcal{T}_k|\boldsymbol{y}_{1:k}\}$ for sources to be active at the coordinate set $\mathcal{T}_k$ at time k is obtained via marginalization:

$$\mathrm{P}\{\mathcal{T}_k|\boldsymbol{y}_{1:k}\} = \sum_{l\in\Lambda_k} \omega_k^{[l]}, \quad \Lambda_k = \left\{l : Q\left(\boldsymbol{n}_k^{[l]}\right) = \mathbb{1}_{\mathcal{T}_k}\right\}. \tag{4.10}$$

Here, the function $Q : \mathbb{R}^{IJ} \to \{0, 1\}$ sets all entries of $\boldsymbol{n}_k^{[l]}$ to 1 which are unequal to 0. In the case of one source and a SIR PF with $w_k^{[l]} = 1/L$, the probability for a source at position (i, j) at time k is approximately obtained as

$$P_{\mathrm{s}}(i, j, k) = \mathrm{P}\{\text{source at } (i, j, k)|\boldsymbol{y}_{1:k}\} = \frac{L_{i,j,k}}{L}, \tag{4.11}$$

where $L_{i,j,k}$ is the number of particles for which $[\boldsymbol{n}_k^{[l]}]_{i+(j-1)I} > 0$.

4.3 Decentralized Scheme

The particle filter developed in the previous section is centralized in nature since it requires all pressure measurements, the observation modalities described by the globally assembled likelihood function and operates on the full state vector $\boldsymbol{x}_k$ in a fusion center. Additionally, the computed estimates are inherently unknown on the individual sensor nodes. In a sensor network (SN) context, such constraints are undesirable since they imply a large communication overhead to collect the measured data, a high computational effort due to the high-dimensional state vector, a feedback to the sensors nodes to spread the estimates and a central knowledge of measurement noise. Therefore, a decentralized scheme that distributes the data collection and computational costs among several clusters of sensor nodes is developed. This is achieved by splitting the state-space model (4.1), (3.34) into lower-dimensional sub-models (each corresponding to a cluster), cf. Sawo et al. 2006 and Sawo et al. 2008. Due to the sparsity of the state-space matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma}$, these sub-models are only loosely coupled, thus a decentralized PF that requires little communication between the clusters can be developed.

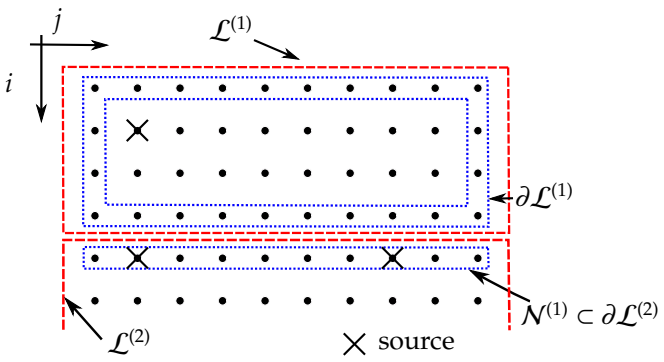


Figure 4.1: Vertices collected in 2 clusters $\mathcal{L}^{(\cdot)}$, their boundary sets $\partial\mathcal{L}^{(\cdot)}$ and neighbor sets $\mathcal{N}^{(\cdot)}$.

A *Clusters and Partitioned State-Space Model*

I start with partitioning the region of interest Ω into M disjoint subregions $\Omega^{(m)}$. The sampling lattice corresponding to each subregion is given by $\mathcal{L}^{(m)} = \mathcal{L} \cap \Omega^{(m)}$ with its boundary nodes $\partial\mathcal{L}^{(m)}$, see Fig. 4.1. The sensors within each subregion form clusters, denoted by $\mathcal{R}^{(m)} = \mathcal{R} \cap \Omega^{(m)} \subset \mathcal{L}^{(m)}$. To each subregion let us associate a subset of elements of the state vector $\boldsymbol{x}_k$ given by

$$\boldsymbol{x}_k^{(m)} = \begin{bmatrix} \boldsymbol{q}_k^{(m)} \\ \boldsymbol{p}_k^{(m)} \\ \boldsymbol{n}_k^{(m)} \end{bmatrix} \tag{4.12}$$

where

$$\boldsymbol{p}_k^{(m)} = \left[p(i\Delta_r, j\Delta_r, k\Delta_t) \right]_{(i,j)\in\mathcal{L}^{(m)}}$$

and the superscript $^{(m)}$ refers to region m.

Except for $\boldsymbol{\Phi}_{12}$, all of the blocks in the state-space matrices $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Gamma}_k$ are diagonal or zero [cf. (4.2)]. Thus there is no coupling between the sub-vectors $\boldsymbol{p}_k^{(m)}$ from different subregions and similarly for the sub-vector $\boldsymbol{q}_k^{(m)}$. Coupling between state vectors from different regions, induced by the non-diagonal structure of $\boldsymbol{\Phi}_{12}$, is between the sub-vectors $\boldsymbol{q}_k^{(m)}$ in one subregion and the sub-vectors $\boldsymbol{p}_k^{(m)}$ in the adjacent subregions (in fact, this coupling is limited to samples at the boundaries of the subregions). The same applies for the sub-vectors $\boldsymbol{n}_k^{(m)}$ due to the spatial noise. This gives

$$\boxed{\begin{aligned} \boldsymbol{x}_{k+1}^{(m)} &= \boldsymbol{\Phi}_k^{(m)} \boldsymbol{x}_k^{(m)} + \boldsymbol{\xi}_k^{(m)} \\ &\quad + \boldsymbol{\Gamma}_k^{(m)} \boldsymbol{u}_k^{(m)} + \boldsymbol{\gamma}_k^{(m)} \\ &\quad + \boldsymbol{G}_k^{(m)} \boldsymbol{n'}_k^{(m)}, \\ \boldsymbol{y}_k^{(m)} &= \boldsymbol{C}^{(m)} \boldsymbol{p}_k^{(m)} + \boldsymbol{v}_k^{(m)}. \end{aligned}} \tag{4.13}$$

This coupling Equation (4.13) is only possible for the time-independent part of these matrices. However for uncorrelated noise between clusters, the time-dependent part, i.e. $\boldsymbol{D}_k$, is calculated separately according to Subsection 3.5.A on every cluster at each time step, see below.

The coupling terms between neighboring subregions are given by

$$\boldsymbol{\xi}_k^{(m)} = \sum_{m' \in \mathcal{N}^{(m)}} \boldsymbol{T}_k^{(m,m')} \boldsymbol{x}_k^{(m')}, \tag{4.14}$$

with

$$\boldsymbol{T}_k^{(m,m')} = \begin{bmatrix} \mathbf{0} & \boldsymbol{\Phi}_{12}^{(m,m')} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{D}_k^{(m,m')} \end{bmatrix}, \tag{4.15}$$

and, analogously,

$$\boldsymbol{\gamma}_k^{(m)} = \sum_{m' \in \mathcal{N}^{(m)}} \boldsymbol{R}_k^{(m,m')} \boldsymbol{u}_k^{(m')}, \tag{4.16}$$

with

$$\boldsymbol{R}_k^{(m,m')} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{D}_k^{(m,m')} \end{bmatrix}. \tag{4.17}$$

Here, $\mathcal{N}^{(m)}$ is the set of subregions adjacent to $\Omega^{(m)}$ and $\boldsymbol{\Phi}_{12}^{(m,m')}$ is obtained from $\boldsymbol{\Phi}_{12}$ by extracting the rows and columns corresponding to $\mathcal{L}^{(m)}$ and $\mathcal{L}^{(m')}$. The off-diagonals of $\boldsymbol{\Phi}_{12}$ are extremely sparsely populated; in fact, (4.14) contains only few non-zero terms corresponding to adjacent pressure samples and the change of sources from one to another cluster. $\boldsymbol{D}_k^{(m,m')}$ is generated from every cluster m' such that the composition of all sub-matrices $\boldsymbol{D}_k^{(m)}$ and $\boldsymbol{D}_k^{(m,m')}$ equals $\boldsymbol{D}_k$. From a practical perspective, elements of $\boldsymbol{D}_k^{(m)}$ are calculated separately on every cluster by means of spatial noise with additional triggering a message to neighbor clusters whenever a source hop (migration) from one cluster to another is detected [this takes over the purpose of $\boldsymbol{D}_k^{(m,m')}$ and supersedes (4.16)]. Furthermore, the coupling term $\boldsymbol{\xi}_k^{(m)}$ means that pressure samples at subregion boundaries are exchanged between neighboring clusters in order to compute the finite differences.

Boundary conditions do not play a role in the decomposition step as long as (i) they do not depend on adjacent neighbors and (ii) their numerical solution fits into (3.23). In the first situation, an additional term $\boldsymbol{\Phi}_{11}^{(m,m')}$ or $\boldsymbol{\Phi}_{21}^{(m,m')}$ arises in matrix $\boldsymbol{T}_k^{(m,m')}$.

B Decentralized Particle Filter

For the decentralized PF, we need to distribute the sampling (particle generation) step and the weight computation step. Based on the local particles and weights, each cluster can then compute posterior source probabilities in a similar manner as in Subsection 4.3.B.

Sub-particles $\boldsymbol{x}_k^{[l,m]}$ within cluster $\mathcal{R}^{(m)}$ are generated according to (4.13), cf. also (4.5),

Particle Generation

$$\boxed{\begin{aligned} \boldsymbol{x}_{k+1}^{[l,m]} = {} & \boldsymbol{\Phi}_k^{(m)} \boldsymbol{x}_k^{[l,m]} + \boldsymbol{\xi}_k^{[l,m]} \\ & + \boldsymbol{\Gamma}_k^{(m)} \boldsymbol{u}_k^{[l,m]} + \boldsymbol{\gamma}_k^{[l,m]} \\ & + \boldsymbol{G}_k^{[l,m]} \boldsymbol{n'}_k^{[l,m]} . \end{aligned}} \tag{4.18}$$

Here, $\boldsymbol{x}_k^{[l,m]}$ is a randomly chosen previous particle and $\boldsymbol{n'}_k^{[l,m]}$ is a (local) noise vector realization. Furthermore, $\boldsymbol{\xi}_k^{[l,m]} = \sum_{m' \in \mathcal{N}^{(m)}} \boldsymbol{T}_k^{(m,m')} \boldsymbol{x}_k^{[l,m']}$ and

$\xi_k^{[l,m]} = \sum_{m' \in \mathcal{N}^{(m)}} \boldsymbol{R}_k^{(m,m')} \boldsymbol{u}_k^{[l,m']}$, respectively. In order to compute the latter, only elements of $\boldsymbol{x}_k^{[l,m']}$ that correspond to pressure samples from the boundaries of adjacent subregions are exchanged and in the event of source hopping from one to another cluster a message is sent.

Assuming independent measurement noise in the individual subregions, i.e., $f_v(\boldsymbol{v}_k) = \prod_{m=1}^{M} f_{v^{(m)}}\left(\boldsymbol{v}_k^{(m)}\right)$, the weight update (4.6) is computed in each cluster as

Weights

$$\tilde{\omega}_{k+1}^{[l]} = \omega_k^{[l]} \prod_{m=1}^{M} \bar{\omega}_k^{[l,m]}, \tag{4.19}$$

where the partial weight

$$\bar{\omega}_k^{[l,m]} = f_{v^{(m)}}(\boldsymbol{y}_{k+1}^{(m)} - \tilde{\boldsymbol{C}}^{(m)} \boldsymbol{x}_{k+1}^{[l,m]})$$

is computed within each cluster and then are shared among all clusters to obtain the final unnormalized weight. Farahmand et al. 2010 and Oreshkin et al. 2010 are treating the issue of computation of the global factorizable likelihood by means of distributed protocols. If these take longer than the time-span between two estimator iterations, the particle filter converts to a particle predictor.

A remaining problem with the decentralized PF is that the sampling (particle generation) step (4.18) requires that the clusters pick local particles $\boldsymbol{x}_k^{[l,m]}$, $m = 1, \ldots, M$, that correspond to the same global particle $\boldsymbol{x}_k^{[l]}$. This choice is made at random according to the weights $\omega_k^{[l]}$. The same problem occurs for the resampling procedure. Since a central random number generator whose output is distributed to each cluster incurs a large communication overhead, we propose to use identical pseudo-number generators in all clusters and initialize those with the same seed, thereby ensuring that all clusters perform the same (re)sampling.[3,4]

(Re)sampling

[3] Farahmand et al. 2010.
[4] Coates 2004.

4.4 *Decentralized Source Localization*

The PF yields the posterior PDF of the sources' position and lifespan. To obtain the current MAP position estimates

$$\boxed{(\hat{i}_k, \hat{j}_k) = \arg \max_{(i,j) \in \mathcal{L}} P_s(i, j, k),} \tag{4.20}$$

the maximum and the maximizing state of the posterior PDF $P_s(i, j, k)$ in (4.11) must be found. In the decentralized scheme, each cluster disposes only of the local posterior PDF for the state sub-vector $\boldsymbol{x}_k^{(m)}$. To find the global maximizing state, each cluster determines the local maximizing state and afterwards the clusters use a distributed consensus protocol to determine the global maximum. For simplicity, this procedure is here developed for one source.

For the centralized PF, the posterior probability for a source to be active at time k at position (i, j) is given by (4.11). In the decentralized case, each cluster determines a similar probability according to

$$P_s^{(m)}(i, j, k) = \begin{cases} \frac{L_{i,j,k}^{(m)}}{L}, & (i, j) \in \mathcal{L}^{(m)}, \\ 0, & \text{else}, \end{cases}$$

where $L_{i,j,k}^{(m)}$ denotes the number of particles $\boldsymbol{x}_k^{[l,m]}$ for which $[\boldsymbol{n}_k^{[l,m]}]_{i+(j-1)I} > 0$. Since the probabilities $P_{\mathrm{s}}^{(m)}(i,j,k)$ have disjoint support, the maximization underlying the MAP estimates (4.20) is

$$P_{k,\max} = \max_{(i,j)\in\mathcal{L}} P_{\mathrm{s}}(i,j,k) = \max_m P_{k,\max}^{(m)}$$

with

$$P_{k,\max}^{(m)} = \max_{(i,j)\in\mathcal{L}^{(m)}} P_{\mathrm{s}}^{(m)}(i,j,k). \tag{4.21}$$

While the local maxima with regard to $\mathcal{L}^{(m)}$ can be determined within each cluster, the global maximization with regard to m requires communication between the clusters. Since sharing the local maxima among all clusters via broadcast transmissions requires a large coordinated transmission, we compute the global maximum via the maximum consensus (MC) algorithm[5]. For the MC algorithm we assume that only neighboring clusters communicate with each other. Thus each cluster sends to the adjacent clusters a message which contains the local maximum and the position for which the local maximum is achieved. In the subsequent steps, each cluster compare the incoming "maximum" messages with their current estimate of the global position and retain the most likely and its associated position. In the next iteration this message will be sent to the neighboring clusters.

[5] Bauso et al. 2006.

Denote the current estimate of the maximum $P_{k,\max}$ for cluster m by $\hat{P}_{k,\max}^{(m)}$ and let $(\hat{i}_k^{(m)}, \hat{j}_k^{(m)})$ be the associated position estimate (initially, $\hat{P}_{k,\max}^{(m)} = P_{k,\max}^{(m)}$). In our MC algorithm, termed argumentum-maximi consensus (AMC), at time instant k each cluster performs the following steps:

1. Send a message containing the estimates $\hat{P}_{k,\max}^{(m)}$ and $(\hat{i}_k^{(m)}, \hat{j}_k^{(m)})$ to the neighbor clusters $\mathcal{N}^{(m)}$.

2. Receive corresponding messages from the neighbor cluster; if a neighbor $m' \in \mathcal{N}^{(m)}$ remains silent, then $\hat{P}_{k,\max}^{(m')} = \hat{P}_{k-1,\max}^{(m')}$.

3. Update the maximum probability and position as

$$\hat{P}_{k+1,\max}^{(m)} = \hat{P}_{k,\max}^{(m_0)}, \quad (\hat{i}_{k+1}^{(m)}, \hat{j}_{k+1}^{(m)}) = (\hat{i}_k^{(m_0)}, \hat{j}_k^{(m_0)}),$$

   with $m_0 = \arg\max_{m'\in\{m\}\cap\mathcal{N}^{(m)}} P_{k,\max}^{(m')}$.

4. If $\hat{P}_{k+1,\max}^{(m)} \neq \hat{P}_{k,\max}^{(m)}$ goto 1), otherwise goto 2).

When the maximum is fixed, all clusters converges to the true maximum after some iterations (depending on the diameter of the cluster communication graph). Here, the position of the maximum moves as the distributed PF evolves and the AMC will then allow the clusters to jointly track the maximum.

4.5 Algorithm Summary

A Dimensions and Trade-offs

Since we are estimating the two-dimensional position and activation time for each of the S sources, the number of unknowns equals $3S$. This is relevant for the choice of the number of particles, cf. Leeuwen 2009. For the calculation of the forward model (state transition), however, the dimension of the state vector $\boldsymbol{x}_k$ is relevant which equals $3IJ$. In the decentralized case, the computational complexity of the forward model is distributed across all clusters.

We now face the behavior of a high number of clusters. Generally, the volume of a polytope (cluster) $\mathcal{L}^{(m)}$ with edge lengths $e_i(m)$ in a d-dimensional lattice $\mathcal{L} \subset \mathbb{Z}^d$ is given by $|\mathcal{L}^{(m)}| = \prod_{i=1}^{d} e_i^{(m)}$ while its $(d-1)$-dimensional surface equals $|\partial\mathcal{L}^{(m)}| = 2\sum_{j=1}^{d} \partial_j \prod_{i=1}^{d} e_i^{(m)}$.

Generally, the dimension per cluster of the equation system to be calculated is $3|\mathcal{L}^{(m)}|$ which, in comparison, equals in the centralized case $3|\mathcal{L}|$.

In our 2-D problem, let the lattice $\mathcal{L}$ be partitioned into $M = M_\mathrm{i} M_\mathrm{j}$ clusters of same size, M_i clusters in i-direction and M_j clusters in j-direction. Then $e_1 = I/M_\mathrm{i}$ and $e_2 = J/M_\mathrm{j}$. Furthermore, the volume $|\mathcal{L}^{(m)}| = IJ/M_\mathrm{i}M_\mathrm{j}$. When $M \to \infty$ then the dimension of the equation system, which specifies the amount of computation, becomes[6] in $\mathcal{O}(1/M)$. Thus the computational effort per cluster decreases when the number of clusters increases. On the other hand, an increasing number of clusters leads to a larger number of boundaries and hence to a larger communication overhead (i.e., message exchange between adjacent clusters).

[6] Knuth 1976.

Figure 4.2: Global initialization

```
generate priors X_0 ;                                  // Equation (4.8)
decompose X_0 to {X_0^(m)} ;                           // Equation (4.12)
choose seed (Section B);
for m = 1 to M parallel do
    DD-SIR-PF(X_0^(m), seed) of cluster m;
end
```

B Communication Between Clusters

The variables which are broadcast by cluster m are summarized by the set

$$\left\{\bar{\mathcal{W}}_k^{(m)}, \mathcal{P}_k^{(m)}, \mu_k^{(i,m)}, \hat{P}_{k,\max}^{(m)}, \hat{\mathcal{T}}_k^{(m)}\right\} . \tag{4.22}$$

The first subset $\bar{\mathcal{W}}_k^{(m)} = \left\{\bar{\omega}_k^{[1,m]}, \cdots, \bar{\omega}_k^{[L,m]}\right\}$ collects the local PF weights while $\mathcal{P}_k^{(m)} = \left\{[\boldsymbol{p}_k^{[1,m]}]_{i+(j-1)I} \,|\, (i,j) \in \partial\mathcal{L}^{(m)}\right\}_{l=1}^{L}$ collects all pressure substate particles on the boundary. The third, $\mu_k^{(i,m)}$, signifies a message about sources which migrate across boundaries from one cluster to another. Every message includes the new location and the current

input : $\mathcal{X}_0^{(m)}$,seed
$k \leftarrow 1$;
wait while no signal sensed and no wake-up call;
send wake-up call to other clusters;
while *estimating* **do**
  observe: $\boldsymbol{y}_k^{(m)}$
  $\left\{\bar{\mathcal{W}}_k^{(m)}, \mathcal{X}_k^{(m)}\right\} \leftarrow$ `SI(`$\mathcal{X}_{k-1}^{(m)}, \boldsymbol{y}_k^{(m)}$`)`;
  transmit $\left\{\bar{\mathcal{W}}_k^{(m)}, \mathcal{P}_k^{(m)}, \hat{P}_{k-1,\max}^{(m)}, \hat{\mathcal{T}}_{k-1}^{(m)}\right\}$;
  wait until reception from other clusters;
  $\left\{\mathcal{W}_k, \mathcal{X}_k^{(m)}\right\} \leftarrow$ `modify(`$\bar{\mathcal{W}}_k^1, \cdots, \bar{\mathcal{W}}_k^M, \mathcal{X}_k^{(m)}, \mathcal{P}_k^{(\mathcal{N}^{(m)})}$`)`
  calculate $\left\{\hat{P}_{k,\max}^{(m)}, \hat{\mathcal{T}}_k^{(m)}\right\}$; `// Equation (4.21)`
  $\mathcal{X}_k^{(m)} \leftarrow$ `resampling(`$\mathcal{W}_k, \mathcal{X}_k^{(m)}$, *seed*`)` ;
  $\mathcal{W}_k^{(m)} \leftarrow \{1/L\}_{\ell=1}^{L}$;
  $k \leftarrow k+1$;
end

Figure 4.3: `DD-SIR-PF()`: Decentralized distributed SIR particle filter of cluster m.

time duration since the occurrence of the sources. The last two terms stem from the AMC algorithm where $\hat{\mathcal{T}}_k^{(m)} = (\hat{i}_k^{(m)}, \hat{j}_k^{(m)})$.

Note that the cardinality of (4.22) which is a measure of the amount of transmission per cluster is given by the sum

$$\begin{aligned} L & \quad (\bar{\mathcal{W}}_k^{(m)} \text{ to all clusters}) \\ +|\partial\mathcal{L}^{(m)}|L & \quad (\mathcal{P}_k^{(m)} \text{ to adjacent clusters}) \\ +2M & \quad (\hat{P}_{k,\max}^{(m)} \text{ and } \hat{\mathcal{T}}_k^{(m)} \text{ to adjacent clusters}) \end{aligned}$$

Here, the $\mu_k^{(i,m)}$ messages are disregarded. The amount of transmission in the decentralized case to adjacent neighbors for $M_\mathrm{i} \to \infty$ and $M_\mathrm{j} \to \infty$ is in $O(1/M_\mathrm{i})$ and $O(1/M_\mathrm{j})$, respectively. The transmission of weights is in $O(M)$ for $M \to \infty$ while the overall communication load is in $O(M^2)$.

Note, that there is no approximation compared to the centralized method and thus, neither source coding nor approximations reducing the weight communication have been considered. For the communication of the weights the graph needs to be either fully connected or the clusters needs to act as relay. A summary is drawn in Table 4.1.

	neighbor	not neighbor
$\boldsymbol{p}_k$	boundary elements	
$\boldsymbol{n}_k$	source migration	
$\omega_k^{[l,m]}$	all	(all if not relaying/forwarding)
$\hat{\mathcal{T}}_k^{(m)}$	all	
$P_{k,\max}^{(m)}$	all	

Table 4.1: Necessary message exchange. Source migration denotes the information that a source changes from one cluster to another.

C Algorithm

The algorithm of the decentralized and distributed SIR PF together with the AMC is drawn in Figs. 4.2 to 4.5. Compare it with that one in Arulampalam et al. 2002 and note that the for-loop can be parallelized.

The joint setup of the computational nodes is shown in Fig. 4.2 which consists of the calculation of the priors and the synchronization of the pseudo-random generator. Subsequently, each individual PF is launched (see Fig. 4.3). Two important sub-routines are plotted in their own tableaus:

- Fig. 4.4 calculates particles and sends messages when a source jumps over to another cluster.
- Fig. 4.5 adds states from the neighbor clusters according to (4.13) and calculates the overall weight (4.19).

Figure 4.4: `SI()`: sample importance part

input : $\mathcal{X}_{k-1}^{(m)}, \boldsymbol{y}_k^{(m)}$
output: $\left\{\bar{\mathcal{W}}_k^{(m)}, \mathcal{X}_k^{(m)}\right\}$
for $i = 1$ **to** L **do**
  Draw $\boldsymbol{x}_k^{[l,m]} \sim f(\boldsymbol{x}_k^{(m)} | \boldsymbol{x}_{k-1}^{(m)})$;
  if *source(s) cross(es) boundary* **then**
    send message to adjacent cluster
  end
  $\bar{\omega}_k^{[l,m]} \leftarrow f(\boldsymbol{y}_k^{(m)} | \boldsymbol{x}_k^{[l,m]})$;
end

Figure 4.5: `modify()`: contribution of the neighbors. $\boldsymbol{T}^{(m)}$ is a mapping from neighbors' pressure sub-states to the own sub-states with $\boldsymbol{T}^{(m)}\mathcal{P}^{(\mathcal{N}^{(m)})}{}_k$ assembles to $\{\xi_k^{[l,m]}\}_{l=1}^{L}$.

input : $\left\{\bar{\mathcal{W}}_k^1, \cdots, \bar{\mathcal{W}}_k^M, \mathcal{X}_k^{(m)}, \mathcal{P}_k^{(\mathcal{N}^{(m)})}\right\}$
output: $\left\{\mathcal{W}_k, \mathcal{X}_k^{(m)}\right\}$
$\mathcal{X}_k^{(m)} \leftarrow \mathcal{X}_k^{(m)} + \boldsymbol{T}^{(m)}\mathcal{P}_k^{(\mathcal{N}^{(m)})}$; `// Equation (4.15)`
$\hat{\mathcal{W}}_k \leftarrow \bar{\mathcal{W}}_k^1 \cdots \bar{\mathcal{W}}_k^M$; `// Equation (4.19)`
normalize $\hat{\mathcal{W}}_k$;

4.6 Simulations

In this section we present simulations illustrating the performance of the proposed algorithms 4.2 to 4.5. The configuration used in the simulations is shown in Fig. 4.6 with parameters in Table 4.2 ($\mathrm{N}\left\{\mu, \sigma^2\right\}$ denotes the Gaussian distribution with mean μ and variance σ^2). In particular, we used $M = 5$ subregions $\Omega^{(m)}$ corresponding to 5 clusters each with 2 sensors. We considered a single source located in $\Omega^{(3)}$ at the lattice point $(i_0, j_0) = (25, 25)$; it is modeled by choosing the source function as $s_0[n] = s_0(n\Delta_t)$ where $s_0(t)$ is a time-shifted Ricker wavelet. A Ricker wavelet[7] is defined by the negative second derivative of a Gaussian function such that

$$\mathrm{ricker}(t) = \left(1 - 2\pi^2\nu^2t^2\right)\exp\left(-\pi^2\nu^2t^2\right). \tag{4.23}$$

[7] Ryan 1994.

Here, ν is approximately the peak frequency. A Ricker wavelet shifted by 16.7 ms with $\nu = 60$ Hz is used, i.e. $s_0(t) = \text{ricker}(t - 16.7\,\text{ms})$, see Fig. 4.7. The acoustic pressure field is simulated using the finite-difference method (FDM) introduced in Section 3.2. A snapshot of the field at time $k = 160$ is shown in Fig. 4.8.

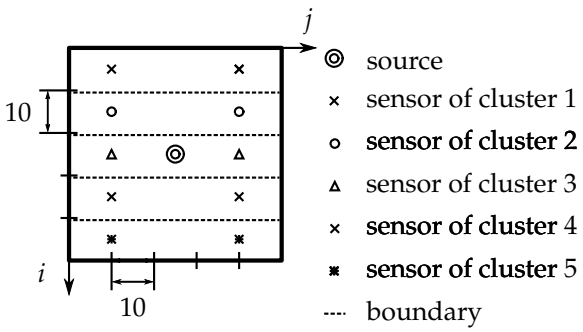


Figure 4.6: Simulation setup comprising sensors, a single source, and SN cluster structure.

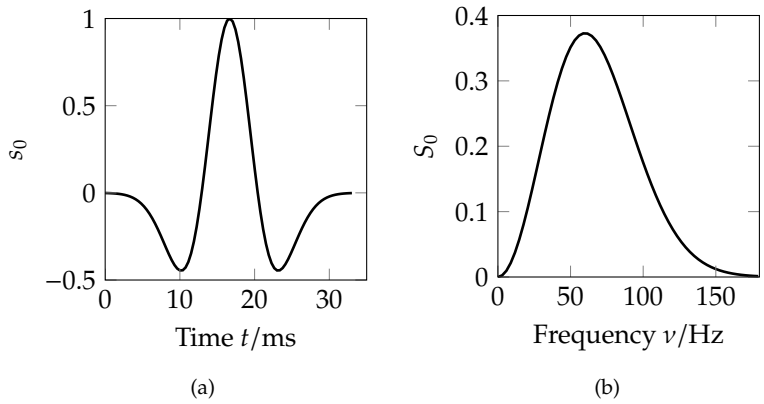


Figure 4.7: Ricker wavelet shifted by 16.7 ms with $\nu = 60$ Hz (a) in the time-domain and (b) its Fourier transform.

The parameters used in the decentralized PF are summarized in Table 4.3 (Unif $\{a, b\}$ represents a discrete uniform PDF with support $[a, b]$). For the fixed source position, we used a discrete uniform distribution on the 50×50 lattice. The spatio-temporal noise and the observation noise are drawn from a Gaussian distribution. The PF is initialized at time $k = 0$ and the source is assumed to become active at time instant $k < 0$. The maximum value of the random variable k_{start} is a prior and is proportional to the maximal possible time duration between source arise and first detection [cf. (4.8)]. Larger values of k_{start} necessitate a larger number of particles to cover the time interval $[-k_{\text{start}}, 0]$ and thus to achieve the same approximation accuracy.

A *Estimation of Posterior PDF*

For the centralized PF, Fig. 4.9a shows an example of the posterior PDF $P_{\text{s}}(i, j, k)$ for the source position obtained with the centralized particle filter at time instant $k = 160$ [cf. (4.11)]. For comparison, Fig. 4.9b shows the result obtained with the decentralized PF, i.e., the composition $\sum_{m=1}^{5} P_{\text{s}}^{(m)}(i, j, k)$ of the local posterior PDF obtained by

Table 4.2: Hallway simulation: settings of the model.

FDM	Δ_t	371 ns
	Δ_r	12.24 cm
	$I \times J$	50×50
speed	c	340 m/s
noise	w	i.i.d. $\mathrm{N}\left\{0, 100\,\mathrm{pPa/s^2}\right\}$
	v	i.i.d. $\mathrm{N}\{0, 100\,\mathrm{pPa}\}$
source	$s_0(t)$	ricker$(t - 16.7\,\mathrm{ms})$
	(i_0, j_0)	(25, 25)
sensors	setup	Fig. 4.6

Table 4.3: Hallway simulation: settings of the PF.

particles	L	20000
space/time jitter	x, y	$\mathrm{N}\left\{0, \Delta_r^2/8^2\right\}$
	t	$\mathrm{N}\left\{0, \Delta_t^2/8^2\right\}$
	v	i.i.d. $\mathrm{N}\{0, 5\,\mathrm{mPa}\}$
priors	k_{start}	Unif $\{0, 41345\}$
	i, j	Unif $\{0, 50\}$

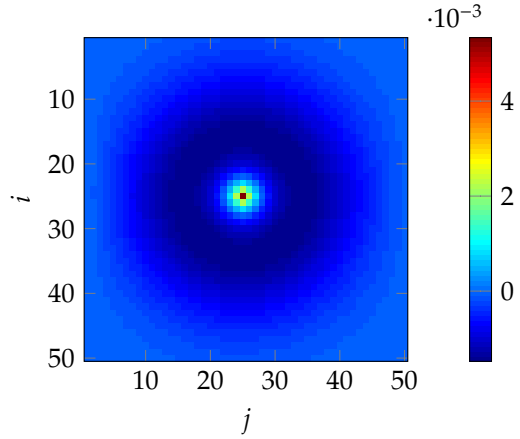
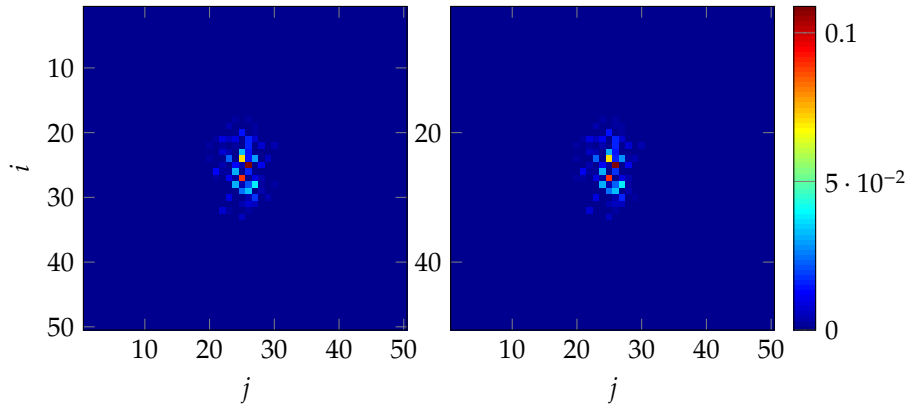


Figure 4.8: Pressure field from finite difference modeling after 41750 time steps (corresponding to estimation time k=160).



(a) (b)

Figure 4.9: Posterior source position PDF $P_{\mathrm{s}}(i, j, k)$ at time $k = 160$ obtained with (a) centralized and (b) decentralized PF.

each cluster. It is seen that the centralized and the decentralized PF obtain similar results and both yield a posterior PDF which is well-concentrated about the true position $(i_0, j_0) = (25, 25)$ of the source.

Figs. 4.13a and 4.13b show the MAP and MMSE of the source's i coordinate and j coordinate, respectively. The MAP estimates $(\hat{i}_k, \hat{j}_k)$ are given by (4.20); the minimum mean-squared error (MMSE) estimates $(\hat{i}_k^{\mathrm{MMSE}}, \hat{js}_k^{\mathrm{MMSE}})$ are obtained as conditional means of the source coordinates obtained with the conditional posterior PMF $P_{\mathrm{s}}(i, j, k)$ for given k. Since the prior of the source location is a discrete uniform distribution, MMSE estimates at $k = 0$ equal $(\hat{i}_k^{\mathrm{MMSE}}, \hat{js}_k^{\mathrm{MMSE}}) = (I/2, J/2) = (25, 25)$. Hence, in this specific case the MMSE estimates outperform the MAP estimates for small k. After a certain number of PF iterations (around $k > 6$), however, the MAP estimates match the true source position better than the MMSE estimates. The variance of $P_{\mathrm{s}}(i, j, k)$ for any given k (which can be interpreted as MMSE) is shown in Fig. 4.10 and corroborates that for small-to-medium k the i coordinate estimate is more reliable; this can be attributed to the specific sensor arrangements which favors better i-resolution (cf. Fig. 4.6).

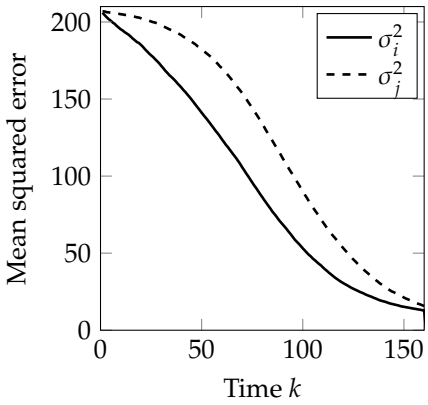


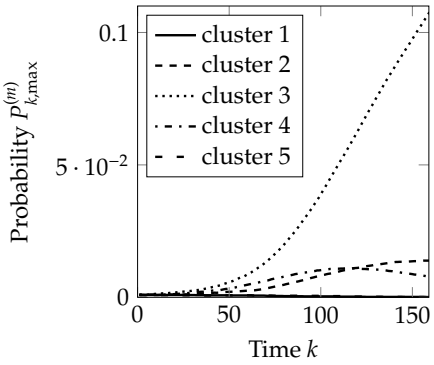
Figure 4.10: Variance of the posterior distribution $P_{\mathrm{s}}(i, j, k)$ with respect to i and j coordinates.

B *Decentralized MAP Source Localization*

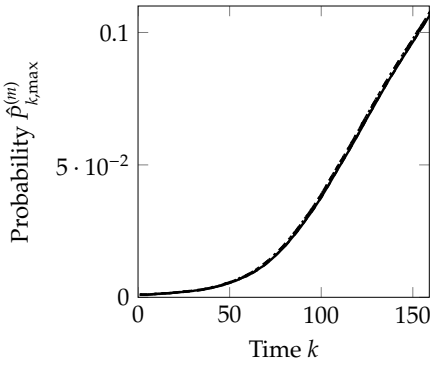
This subsection illustrates the decentralized source localization using the AMC algorithm proposed in Section 4.4 (simulation setup unchanged). Recall that with AMC, each cluster has estimates $\hat{P}_{k,\max}^{(m)}$ of the MAP probability and $(\hat{i}_k^{(m)}, \hat{j}_k^{(m)})$ of the associated position. Fig. 4.11 shows the local MAP probabilities $P_{k,\max}^{(m)}$ [cf. (4.21)] for all five clusters; clearly, only the third cluster builds up a distinguished maximum over time, which indicates that the source is located within $\Omega^{(3)}$.

All clusters track the global MAP probability, Figs. 4.11 and 4.12, and eventually agree on the source position provided by cluster 3 whose behavior over time resembles the global estimates using the centralized PF (cf. Figs. 4.13).

After about 6 iterations, the PF achieves a localization accuracy on the order of the lattice spacing Δ_r. These estimates could be further improved (with higher computational complexity) by refining the discretization lattice and increasing the number of particles.

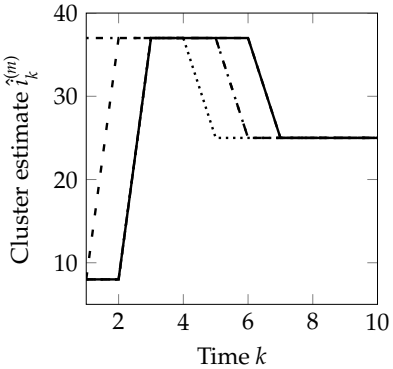


(a)

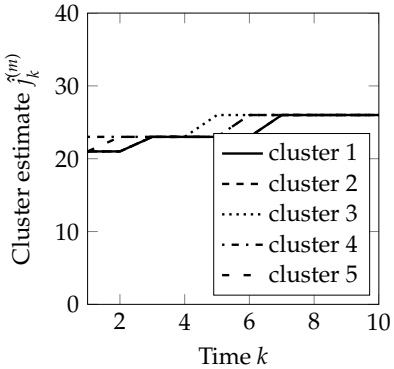


(b)

Figure 4.11: Local (a) MAP probabilities $P^{(m)}_{k,\max}$ in contrast to (b) MAP probability estimates $\hat{P}^{(m)}_{k,\max}$ obtained by the individual clusters using the AMC algorithm.



(a)



(b)

Figure 4.12: Source coordinate estimates (a) $\hat{i}^{(m)}_k$ and (b) $\hat{j}^{(m)}_k$ of the individual clusters.

4.7 *Conclusions of this Chapter*

I proposed a scheme for the localization of multiple acoustic sources in a SN. The method uses an augmented non-linear non-Gaussian state-space model for the acoustic field and on a PF for sequential Bayesian estimation of source positions. This state-space representation for the wave equation gives additional prior physical knowledge and incorporates perturbations and distortion like echoes, thereby resulting in improved estimation accuracy. In addition to the source positions, my PF implicitly provides an estimate of the acoustic field itself. I further developed a decentralized PF in which the computational complexity is distributed over several clusters of the SN. The decentralized PF exploits the sparsity of the matrices involved in the state-space model. In fact, the loose coupling between the components of the state vector allows separate and parallel computation of equation sub-systems of much smaller dimension in each cluster heads. To determine the global MAP estimate of the position of a source, I proposed an argumentum-maximi-consensus algorithm in which the clusters exchange their best MAP probability and source position.

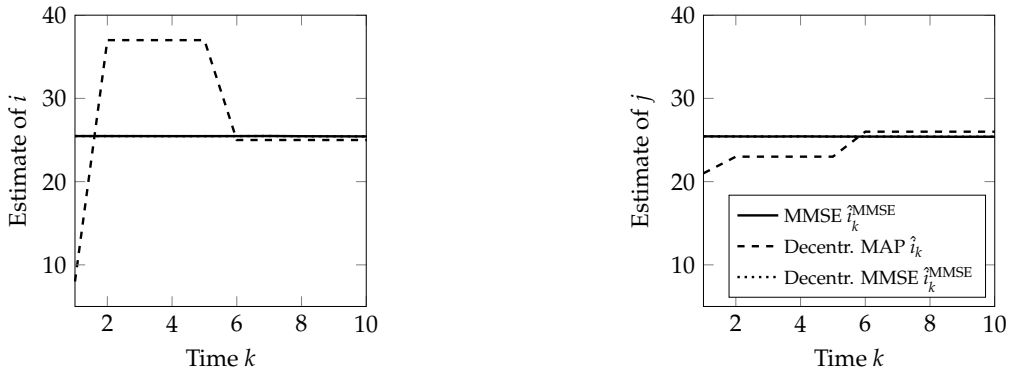


Figure 4.13: MAP and MMSE estimate of the i and j coordinate of the source (note that the lines of the centralized and decentralized MMSE estimations are close together).

5
Communication

This chapter was previously published in Xaver et al. 2012b. I address the communication overhead of particles of the decentralized distributed particle filter (DDPF) introduced in Chapter 4. Again, I use the illustrative example of a 2D acoustic field in a hallway along with an estimator for the position of an acoustic source (see Fig. 1.1 on p. 2) but use Source Model II from Section 3.5.

In the sequel, I propose to use differential encoding based on a Kalman predictor that exploits the spatio-temporal field correlation via the underlying state space model. The correlation stems from the wave equation and is illustrated by the light cone in Fig. 3.1 of Section 3.1. With differential encoding[1], only the difference between the predicted signal and the measurement is transmitted and the receiver reconstructs the original signal using the prediction error. Ideally, the prediction error is a white innovation signal and hence has a flat power spectrum.

[1] For details on linear prediction of discrete-time vector processes see Vaidyanathan 2008, and Anderson et al. 1979.

Due to the discretized partial differential equation (PDE), the global field is modeled by an autoregressive process of order one and thus is predictable. However, using the state space model to this end is not trivial in the context of decentralized estimation. In this case, the hyperbolic structure of the PDE becomes relevant and the order increases with the iteration of the sequential estimator. The white innovations/prediction error signal has a smaller dynamic range and hence can be sent using less transmit power. This is particularly desirable in battery-operated wireless sensor networks (SNs). However, the actual source encoding (i.e., quantization and bit allocation) of the innovations signal is beyond the scope of my thesis.

5.1 Problem Definition

In the remainder of this chapter, I use the deterministic forward model of Chapter 3. The two-dimensional acoustic field in a hallway is described by following scalar wave equation[2], i.e.

[2] Jensen et al. 2011.

$$\frac{1}{c^2}\partial_t^2 p(\boldsymbol{r},t) - \nabla^2 p(\boldsymbol{r},t) = s(\boldsymbol{r},t), \quad \boldsymbol{r} \in \Omega. \tag{3.1}$$

This is a linear hyperbolic second-order PDE, where $p(\boldsymbol{r},t)$ denotes pressure dependent on location $\boldsymbol{r}$ and time t, ∂_t is the partial derivative

with respect to time, ∇^2 is the Laplace operator, c is the sound speed, $s(\boldsymbol{r}, t)$ is a (random) source, and $\Omega \subset \mathbb{R}^2$ is the 2-dimensional region of interest. For the boundary and initial conditions I refer to Section 5.3.

I define $q(\boldsymbol{r}, t) = \partial_t p(\boldsymbol{r}, t)$ and approximate the wave equation via a finite-difference method (FDM)[3]. This results in the state transition model [cf. (3.23)]

$$\begin{bmatrix} \boldsymbol{q}_{k+1} \\ \boldsymbol{p}_{k+1} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \Delta_t \boldsymbol{I} & \boldsymbol{I} \end{bmatrix}}_{\boldsymbol{\Phi}_k} \begin{bmatrix} \boldsymbol{q}_k \\ \boldsymbol{p}_k \end{bmatrix} + \Delta_t c^2 \begin{bmatrix} \boldsymbol{s}_k \\ \boldsymbol{0} \end{bmatrix}, \tag{5.1}$$

[3] Jensen et al. 2011.

where the pressure vector is defined as $\boldsymbol{p}_k = \mathrm{vec}\{\boldsymbol{P}_k\}$ with $[\boldsymbol{P}_k]_{ij} = p(i\Delta_r, j\Delta_r, k\Delta_t); i = 0, \cdots, I, j = 0, \cdots, J; k \in \mathbb{N}_0$ and similar for $\boldsymbol{q}_k$ and source $\boldsymbol{s}_k$. The set of tuples (i, j), termed nodes, is denoted by $\mathcal{L} = \{(i, j) \in \mathbb{N}^2 : (i\Delta_r, j\Delta_r) \in \Omega\}$ (cf. Figure 5.1a).

The model (5.1) consists of two parts:

- The matrix $\boldsymbol{\Phi}_k$ maps the state $[\boldsymbol{q}_k^{\mathrm{T}}\ \boldsymbol{p}_k^{\mathrm{T}}]^{\mathrm{T}}$ deterministically to its *prediction* $[\boldsymbol{q}_{k+1|k}^{\mathrm{T}}\ \boldsymbol{p}_{k+1|k}^{\mathrm{T}}]^{\mathrm{T}}$ in the following time step. (Note that an optimal predictor additionally considers the redundancy of the source.)
- The second term on the right-hand side involving $\boldsymbol{s}_k$ perturbs the state vectors, is assumed to be random, and thus represents the *innovations* process.

[4] Our discussion extends straightforwardly to the case of more than two clusters.

We now partition the nodes $\mathcal{L}$ into two[4] disjoint clusters $\mathcal{L}^{(1)}, \mathcal{L}^{(2)} \subset \mathcal{L}$. The symbol $\bar{\cdot}$ will be used to denote the elements of a vector that correspond to nodes along the boundary between both the clusters and $\check{\cdot}$ signifies that second-order boundary nodes, i.e., neighbors of boundary nodes, are also included in the respective sub-vector. Figure 5.1b specifies the various neighborhood sets used in what follows.

Due to the sparse structure of the global matrix $\boldsymbol{\Phi}_{12}$ (nodes depend only on their neighbors), only the elements of $\bar{\boldsymbol{p}}_k$ need to be exchanged between the clusters for the decentralized estimation of the field (cf. Chapter 4). For this state sub-vector, (5.1) implies

$$\bar{\boldsymbol{p}}_{k+1} = \bar{\boldsymbol{p}}_k + \Delta_t \bar{\boldsymbol{\Phi}}_{11} \bar{\boldsymbol{q}}_{k-1} + \Delta_t \check{\boldsymbol{\Phi}}_{12} \check{\boldsymbol{p}}_{k-1} \tag{5.2}$$

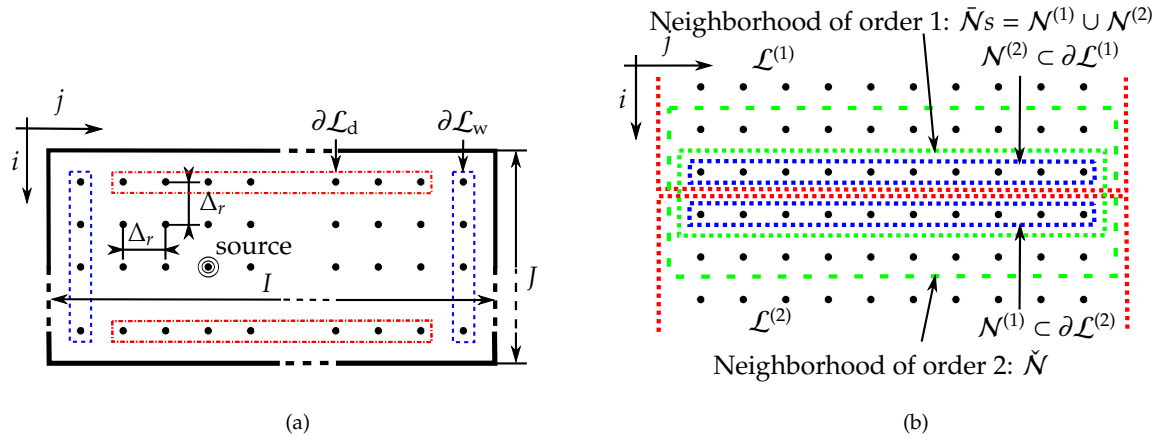
where the source term cancels due to the assumption of sources far away from the boundary. Here $\boldsymbol{s}_k$ does not denote the source of innovation as in the central case, but rather the unknown pressure in the neighborhood $\check{\mathcal{N}} \setminus \bar{\mathcal{N}}$. Note that no approximation is performed.

The main idea of this paper is to signal only the innovations vector

$$\boldsymbol{\iota}_{k+1} = \bar{\boldsymbol{p}}_{k+1} - \bar{\boldsymbol{p}}_{k+1|k} \tag{5.3}$$

rather than the actual state vector $\bar{\boldsymbol{p}}_{k+1}$. This is advantageous since $\boldsymbol{\iota}_{k+1}$ can be better compressed because it is white and has smaller power. The receiving cluster reverts the differential encoding by adding the received innovation vector to the local prediction, i.e.

$$\bar{\boldsymbol{p}}_{k+1} = \boldsymbol{\iota}_{k+1} + \bar{\boldsymbol{p}}_{k+1|k}\,. \tag{5.4}$$



(a) (b)

Figure 5.1: The discretized area $\mathcal{L}$ with disjoint boundary $\partial\mathcal{L} = \bigcup_{\ell=1}^{4} \partial\mathcal{L}_\ell$ is shown in (a). The nodes correspond to sample points of the field. In (b) the area is decomposed into two clusters corresponding to $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$. Various neighbor sets are defined in this sketch.

5.2 *Prediction by the Kalman Filter*

Linear prediction theory exploits the statistical structure of the source which in our context corresponds to the stochastic transition model (5.1). But due to (5.2) it does not seem feasible without high prediction order and with the unknown second order neighbor states[5]. Instead we use the computationally efficient Kalman filter (KF).

The KF[6] falls within the scope of *Bayesian estimators* and is a combination of a sequential *linear minimum mean square error* estimator combined with a state space model, i.e. a state transition model of random states as in (5.1) with additive noise and a observation model, respectively. It estimates the states through observations $\boldsymbol{y}_k$ and is optimum if and only if the priors and noise are Gaussian.

A *Model of the Decentralized System*

We now propose the use of the KF to predict the states $\bar{\boldsymbol{p}}_{k+1}$. For this, let the states of the KF be $[\bar{\boldsymbol{q}}_k^{\mathrm{T}}\, \check{\boldsymbol{p}}_k^{\mathrm{T}}]^{\mathrm{T}}$ with the associated transition model

$$\begin{bmatrix} \bar{\boldsymbol{q}}_{k+1} \\ \check{\boldsymbol{p}}_{k+1} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\Phi}}_{11} & \check{\boldsymbol{\Phi}}_{12} \\ \Delta_t \bar{\boldsymbol{I}} & \check{\boldsymbol{I}} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{q}}_k \\ \check{\boldsymbol{p}}_k \end{bmatrix} + \boldsymbol{G}\boldsymbol{w}_k\,, \tag{5.5a}$$

where $\boldsymbol{w}_k$ is the driving noise. The matrix $\boldsymbol{G} : \mathbb{R}^{|\check{\mathcal{N}}\setminus\bar{\mathcal{N}}|} \to \mathbb{R}^{|\check{\mathcal{N}}|}$ ensures that the driving noise is only added to pressure states in the second-order cluster boundary $\check{\mathcal{N}} \setminus \bar{\mathcal{N}}$. These states are modeled as unknowns and are estimated by the KF in both clusters in the same way. We note that the linear approach underlying the KF is optimal only if $\boldsymbol{w}_k$ is Gaussian.

The aim of matrix $\boldsymbol{G}$ combined with an observation model is to maintain correct state sub-vectors $\bar{\boldsymbol{p}}_k$. This in turn improves the estimates of their neighbors $\check{\mathcal{N}} \setminus \bar{\mathcal{N}}$. With this in mind, let the observation model of cluster m be

$$\boldsymbol{y}_{k+1}^{(m)} = \bar{\boldsymbol{p}}_{k+1}\big(\bar{\boldsymbol{p}}_{k+1}^{(m)}, \boldsymbol{\varepsilon}_{k+1}^{(\overline{m})}, \bar{\boldsymbol{p}}_{k+1|k}^{(\overline{m})}\big). \tag{5.5b}$$

where cluster $\overline{m}$ is the neighbor of cluster m. Here $\bar{\boldsymbol{p}}_{k+1}(\cdot)$ is viewed as a vector-valued linear function depending on the own pressure states

[5] Anderson et al. 1979.

[6] Anderson et al. 1979; Kalman 1960; Kay 1993; Simon 2006; Yardim et al. 2011.

$\bar{\boldsymbol{p}}_{k+1}^{(m)}$ at the boundary and the sum $\boldsymbol{\varepsilon}_{k+1}^{(\overline{m})} + \bar{\boldsymbol{p}}_{k+1|k}^{(\overline{m})}$ (received signal plus prediction). Its sole purpose is to shift the elements of the vectors appropriately. As a direct consequence of the measurement model, the correction step of the KF adjusts $\bar{\boldsymbol{p}}_{k+1|k}$ to the exact values.

B *Decentralized Predictive Encoding Algorithm*

In the following, the subscript $k|k$ emphasizes that estimates at time k are based on observations up to time k. The complete state vector $[\boldsymbol{q}_1^{\mathrm{T}}\ \boldsymbol{p}_1^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{p}_0$ with $\boldsymbol{q}_1 = \boldsymbol{q}_1(\boldsymbol{p}_0)$ are assumed to be known as prior for the KF and the covariance matrices are estimated via the *empirical covariance function*. The latter allows the computation of the Kalman gain matrix $\boldsymbol{K}_k$ for every iteration $k > 1$.

With our method, cluster m performs the following steps starting at time $k = 1$ (decompositions and compositions of vectors are not stated explicitly):

1. Compute the prediction

$$\begin{bmatrix} \bar{\boldsymbol{q}}_{k+1|k} \\ \check{\boldsymbol{p}}_{k+1|k} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\Phi}}_{11} & \check{\boldsymbol{\Phi}}_{12} \\ \Delta_t \bar{\boldsymbol{I}} & \check{\boldsymbol{I}} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{q}}_{k|k} \\ \check{\boldsymbol{p}}_{k|k} \end{bmatrix}. \tag{5.6a}$$

2. Determine the innovation vector

$$\boldsymbol{\varepsilon}_{k+1}^{(m)} = \bar{\boldsymbol{p}}_{k+1}^{(m)} - \bar{\boldsymbol{p}}_{k+1|k}^{(m)}. \tag{5.6b}$$

3. Send $\boldsymbol{\varepsilon}_{k+1}^{(m)}$ to cluster $\overline{m}$ and receive $\boldsymbol{\varepsilon}_{k+1}^{(\overline{m})}$ from cluster $\overline{m}$.

4. Correct the prediction via

$$\begin{bmatrix} \bar{\boldsymbol{q}}_{k+1|k+1} \\ \check{\boldsymbol{p}}_{k+1|k+1} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{q}}_{k+1|k} \\ \check{\boldsymbol{p}}_{k+1|k} \end{bmatrix} + \boldsymbol{K}_k \boldsymbol{\varepsilon}_{k+1}^{(\overline{m})}. \tag{5.6c}$$

5. Use $\bar{\boldsymbol{p}}_{k+1} = \bar{\boldsymbol{p}}_{k+1|k+1}$ for the decentralized estimation of the field $[\boldsymbol{p}_{k+2}^{(m)\mathrm{T}}\ \boldsymbol{p}_{k+2}^{(m)\mathrm{T}}]^{\mathrm{T}}$.

6. Increase k by one and go to step 1.

The predictor (5.6) scales straightforwardly with the number $\check{N}$ of the neighboring clusters. A larger number of neighborhoods entails a larger number of unknown pressure states which have to be estimated in each time step. This increases the dimension of the state vector $[\bar{\boldsymbol{q}}_k^{\mathrm{T}}\ \check{\boldsymbol{p}}_k^{\mathrm{T}}]^{\mathrm{T}}$ in (5.5a) but still exploits the structure of the global $\boldsymbol{\Phi}_{12}$.

5.3 *Numerical Results*

The level of redundancy in a random signal is specified by the auto-correlation and, equivalently, the power spectral density (PSD). For a white signal, the former equals the delta function while the latter is a constant. Beyond that, several methods are used to show the flatness of a PSD and to define a distance between two of them. We recall briefly those definitions.

Let $S(\theta)$ denote the PSD of a discrete time process defined on $[-\pi, \pi)$. In the sequel we use the notion of the distance between

two PSDs S_1 and S_2 from Georgiou 2007 (see Appendix C). There, Georgiou's distance $\varrho(S_1, S_2)$ is defined by

$$\ln\left(\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{S_1(\theta)}{S_2(\theta)}\mathrm{d}\theta\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{S_2(\theta)}{S_1(\theta)}\mathrm{d}\theta\right)\right) . \tag{5.7a}$$

It induces a metric tensor in a manifold $\mathcal{P}$ of PSDs (up to scaling factors). Integration of the metric tensor over a geodesic line between two PSDs gives the path length between them in the manifold,

$$d(S_1, S_2) = \mathrm{Var}\left\{\ln S_1(\theta) - \ln S_2(\theta)\right\} . \tag{5.7b}$$

To measure the whiteness of a signal we choose one of the PSDs to be constant. In addition, we consider a traditional metric used to measure the flatness of a PSD[7],

[7] Vaidyanathan 2008.

$$\mathrm{fl}(S_1) = \frac{e^{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln S_1(\theta)\mathrm{d}\theta}}{\frac{1}{2\pi}\int_{-\pi}^{\pi}S_1(\theta)\mathrm{d}\theta} . \tag{5.8}$$

As illustrative example, a 2-D rectangular hallway from Chapter 4 is simulated using the FDM from above. This hallway and thus the nodes are portioned into two clusters in common with Fig. 5.1. The source s_k is modeled by a Ricker wavelet with additive white Gaussian noise (see Section 4.6). All parameter values are summarized in Tab.5.1.

In this example, 10^4 time steps were simulated with the goal of comparing the statistics of the original signal $\bar{\boldsymbol{p}}_k^{(1)}$ and the whitened signal $\boldsymbol{\varepsilon}_k^{(1)}$, $k = 1, \ldots, 10^4$. Figure 5.2 presents the temporal and spatial empirical autocorrelation function (ACF) averaged over space and time, respectively for the actual state vector and the innovations vector (in the temporal case, the source ACF is also shown). The (small) residual noise in Figure 5.2a corresponds to estimation errors of the KF and is neglected in the following. Clearly, our KF-based predictor succeeds in decorrelating the state vector both temporally and spatially. At time lag $k = \pm 1$ there are two minima that stem from the structure of the state transition model (5.1).

The distance and flatness metrics introduced above are applied to the original state vector and to the innovations vector, averaged over all boundary nodes in cluster 1. The results are summarized in Figure 5.3. In particular, S_i is the discrete-time Fourier transform of a modified innovation ACF where the amplitudes between ± 0.1 are truncated. Hence the high-frequency component in Fig. 5.2a are removed. S_0 = constant denotes the reference PSD of a perfect white signal while S_p is the PSD of the actual state vector.

Furthermore, the improvement achieved by our approach strongly impacts the error resulting from the subsequent quantization. Consider an 8-bit quantizer whose dynamic range is matched to minimum and maximum of the corresponding signals. In the example considered, the quantization mean-squared error (MSE) incurred with the innovations approach is $-100\,\mathrm{dB}$. In comparison, the MSE resulting from quantization of the original state equals $-13.5\,\mathrm{dB}$.

Table 5.1: Settings of the simulated hallway

quantity	*notation*	*value*
rectangular area	$I \times J$	50×50
	Δ_t	371 ns
	Δ_r	12.24 cm
rectangular area	$I \times J$	50×50
acoustic speed	c	340 m/s
source shape	$s_0(t)$	ricker($t - 16.7$ ms)
source location	(i_0, j_0)	$(25, 25)$
Gaussian noise	σ_s	$0.001/(\Delta_t c^2)$
8 sensors		$\{(i, j) : i = 1, \cdots, 4; j = 1, 4\}$
cluster boundary		betw. $i = 10$ and 11
rigid walls $\boldsymbol{r} \in \partial\mathcal{L}_\mathrm{w}$		$\partial_t p(\boldsymbol{r}, t) = 0$
open doors $\boldsymbol{r} \in \partial\mathcal{L}_\mathrm{d}$		$\frac{1}{c}\partial_t p(\boldsymbol{r}, t) - \nabla p(\boldsymbol{r}, t) \cdot \boldsymbol{n} = 0$
initial conditions		$\partial_t p(\boldsymbol{r}, 0) = 0, p(\boldsymbol{r}, 0) = 0$

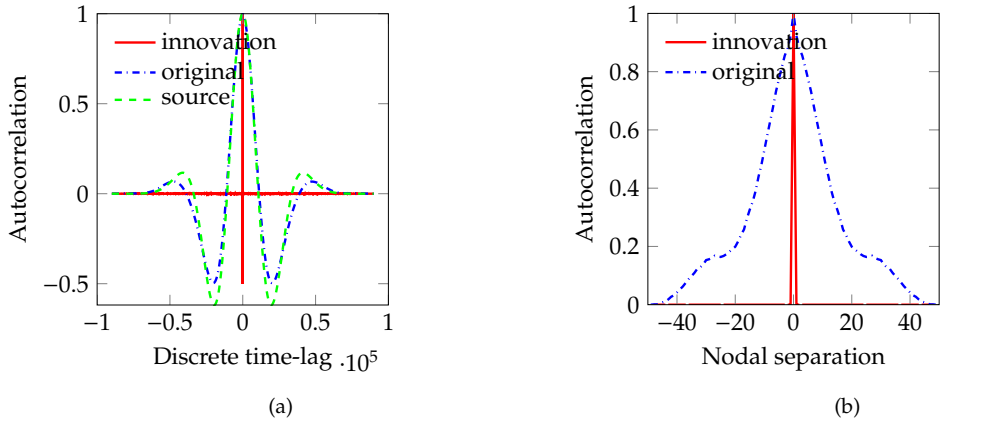
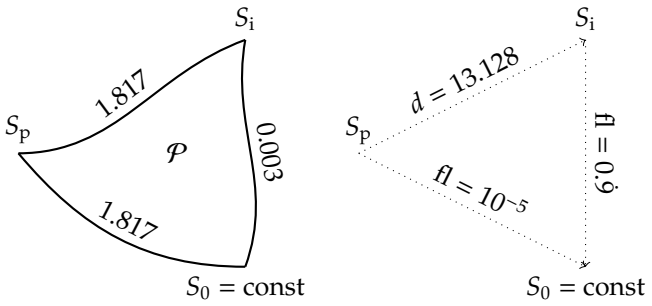


Figure 5.2: Empirical ACFs normalized to unit power. The temporal ACF (a) is averaged over nodes $\partial\mathcal{L}^{(1)}$. The spatial ACF (b) is a function of nodes $\partial\mathcal{L}^{(1)}$ and is averaged over time.

Figure 5.3: Geodesic distance $d(\cdot, \cdot)$ in a manifold $\mathcal{P}$ of PSDs (left) vs. flatness (right) related to the empirical ACF over time (cf. with Fig. 5.2a). S_0 = constant is the PSD of a perfect white signal, S_p is the PSD of the pressure states and S_i is the modified PSD of the innovation.



5.4 Conclusions of this Chapter

Exploiting the spatio-temporal field dependencies of the field reduces the communication overhead in a clustered sensor network. The differential field-state-vector encoding builds on a Kalman prediction step governing the field's spatio-temporal evolution. Only the innovations/prediction error vector needs to be exchanged among clusters. The resulting decentralized sequential algorithm uses the empirical covariances and Kalman gain matrix $\boldsymbol{K}_k$ which are pre-computed and stored in a look-up table. As verified for the example of an acoustic field, our method succeeds in decorrelating the relevant state subvector. The approach reduces transmit power and quantization errors in clustered sensor networks for physics-based field estimation.

6
The Weiss-Weinstein Lower Bound

Although my proposed estimator of Chapter 4 utilizes a particle filter, this Monte-Carlo approach does not obtain any analytic result of the Bayesian estimation problem. Furthermore, the state-space model of Section 4.1 induces continuous and discrete probability densities (PDs) so that the famous sequential Cramér-Rao (SCR) bound is not applicable to get a performance bound on the error variance. This leads to the sequential Weiss-Weinstein (SWW) bound. I have developed the SWW bound in my recently submitted paper[1] which is the basis of this chapter. The results are applicable to a general estimation scenario and are used in Chapter 7 for the localization scenario presented in Chapter 3 and 4.

[1] Xaver et al. 2013.

In Section 6.1, the use of the Weiss-Weinstein (WW) bound is introduced leading to the general formulation of the SWW bound. In addition to the referred literature, I motivate the use of the SWW bound for hybrid continuous/discrete distributions and densities of finite alphabet. I provide a general description of the bound utilizing the expectation operator. Furthermore, I emphasize foundations, which I need for the proofs in the subsequent sections. Section 6.2 introduces the definitions of quantized and hybrid models. Sections 6.3 – 6.8 give analytic solutions for the continuous, the discrete, and the hybrid models. I study the models for Gaussian, uniform, exponential, Laplace, and categorical distributions and priors as well as a mixed scenario. Section 6.9 develops and discusses several practical issues. Supporting simulations complete the analysis in Section 6.10. Several lemmas are summarized and proved in Appendix A.

6.1 *Bayesian Lower Bounds*

This section uses the notation of the Bayesian lower bound introduced in Section 2.2.5, i.e. the inequality

$$\mathrm{E}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right\} \geqslant \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\} \mathrm{E}\left\{\boldsymbol{g}\boldsymbol{g}^{\mathrm{T}}\right\}^{-1} \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{g}^{\mathrm{T}}\right\}^{\mathrm{T}}, \tag{6.1}$$

with the estimation error $\boldsymbol{\varepsilon}$ defined in (2.25), measurements $\boldsymbol{y}$, and the score $\boldsymbol{g}$. Furthermore, I utilize the hybrid PD v_x, which is defined in (2.4).

For the Cramér-Rao (CR) the score of a continuous random parameter $\boldsymbol{y}$ is defined by

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = \partial_x \ln v(\boldsymbol{x},\boldsymbol{y}) = \frac{\partial_x v(\boldsymbol{x},\boldsymbol{y})}{v(\boldsymbol{x},\boldsymbol{y})} \,. \tag{6.2}$$

with the assumption that $\lim_{[x]_\ell\to\pm\infty}[\boldsymbol{x}]_\ell v(\boldsymbol{x}|\boldsymbol{y}) = 0$ for all $\ell = 1,\cdots,N$ and $\boldsymbol{y}$. The ℓth element of $\boldsymbol{x}$ is denoted by $[\boldsymbol{x}]_\ell$. Furthermore, the first and second derivatives of $v(\boldsymbol{x},\boldsymbol{y})$ with respect to $\boldsymbol{x}$ must exist and be absolutely integrable[2]. Inserting (6.2) into (2.37) gives[3]

$$\mathrm{E}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right\} \geqslant \mathrm{E}\left\{\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})^{\mathrm{T}}\right\}^{-1} \triangleq (\boldsymbol{J})^{-1} \tag{6.3}$$

with $\boldsymbol{J}$ being the Fisher information matrix.

For discrete $\boldsymbol{x}$, the ∂_x in (6.2) is approximated by the difference quotient $\overset{1}{\underset{h}{\Delta}}_x$, i.e.

$$\overset{1}{\underset{\boldsymbol{h}}{\Delta}}_{\boldsymbol{x}} v(\boldsymbol{x}) \triangleq \left(\overset{1}{\underset{h_1}{\Delta}}_{x_1} v(\boldsymbol{x}),\cdots,\overset{1}{\underset{h_N}{\Delta}}_{x_N} v(\boldsymbol{x})\right)^{\mathrm{T}} \tag{6.4}$$

where $\overset{1}{\underset{h_\ell}{\Delta}}_{x_\ell} v(\boldsymbol{x}) \triangleq (v(\boldsymbol{x}+h_\ell\boldsymbol{e}_\ell) - v(\boldsymbol{x}))/h_\ell$, and only the ℓ^{th} elements of the unit vector $\boldsymbol{e}_\ell$ is unity. Variables h_ℓ are called test points and specify the sample period if the densities are discrete approximations of continuous ones. This allows the use of hybrid continuous/discrete densities $v(\boldsymbol{x},\boldsymbol{y})$. One alternative[4] to the score (6.2) is

$$\partial_x \ln v(\boldsymbol{x},\boldsymbol{y}) \approx \frac{1}{v(\boldsymbol{x},\boldsymbol{y})}\overset{1}{\underset{\boldsymbol{h}}{\Delta}}_{\boldsymbol{x}} v(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) \tag{6.6}$$

with $\boldsymbol{h} = [h_1,\cdots,h_N]^{\mathrm{T}} \in \mathbb{R}^N$. This score is a special case of Bobrovsky and Zakai's[5] choice of score,

$$g_u = L(\boldsymbol{x}+\boldsymbol{h}_u,\boldsymbol{x},\boldsymbol{y}) - 1 \qquad\qquad u = 1,\cdots,N\,, \tag{6.7}$$

Here, L is the *likelihood ratio*[6]

$$L(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}) \triangleq \frac{v(\boldsymbol{x}_1,\boldsymbol{y})}{v(\boldsymbol{x}_2,\boldsymbol{y})} = \frac{v(\boldsymbol{x}_1,\boldsymbol{y})}{\tilde{v}(\boldsymbol{x}_1,\boldsymbol{y})} = \frac{\mathrm{d}P^{(1)}_{\boldsymbol{x},\boldsymbol{y}}}{\mathrm{d}P^{(2)}_{\boldsymbol{x},\boldsymbol{y}}} \tag{6.8}$$

which is equivalent to the Radon-Nikodym derivative of probability measure $P^{(1)}$ with respect to $P^{(2)}$. The Bobrovsky-Zakai (BZ) lower bound is[7]

$$\mathrm{E}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right\} \geqslant \boldsymbol{H}\boldsymbol{J}^{-1}\boldsymbol{H}^{\mathrm{T}} \tag{6.9}$$

where

$$[\boldsymbol{J}]_{ab} := \mathrm{E}\left\{L(\boldsymbol{x}+\boldsymbol{h}_a,\boldsymbol{x},\boldsymbol{y})L(\boldsymbol{x}+\boldsymbol{h}_b,\boldsymbol{x},\boldsymbol{y})\right\} - 1\,, \tag{6.10}$$

$$\boldsymbol{H} \triangleq [\boldsymbol{h}_1,\cdots,\boldsymbol{h}_N]\,, \qquad a,b = 1,\cdots N\,. \tag{6.11}$$

The specific choices of $\boldsymbol{h}_a$ and $\boldsymbol{h}_b$ influence the lower bound on the error co-variance of elements a and b.

The Radon-Nikodym derivative (6.8) exists if and only if $P^{(1)}$ is absolutely continuous with respect to $P^{(2)}$. This means that the support

[2] Thus, $v(\boldsymbol{x},\boldsymbol{y}) = f(\boldsymbol{x},\boldsymbol{y})$.

[3] Weiss et al. 1988.

[4] A more natural approximation would use the chain rule for the difference quotient (Milne-Thomson 1933), i.e.

$$\left[\overset{1}{\underset{\boldsymbol{h}}{\Delta}}_{\boldsymbol{x}} \ln v(\boldsymbol{x},\boldsymbol{y})\right]_\ell = \ln\left(1+\frac{v(\boldsymbol{x},\boldsymbol{y})}{h_\ell}\right)\overset{1}{\underset{h_\ell}{\Delta}}_{\boldsymbol{x}} v(\boldsymbol{x},\boldsymbol{y}) \tag{6.5}$$

but is not valuable for our problem.

[5] Bobrovsky et al. 1975.

[6] We note that the common notation $f(\boldsymbol{x}) = \sum_{\boldsymbol{x}'} p(\boldsymbol{x}')\delta(\boldsymbol{x}-\boldsymbol{x}')$ for discrete distributions is wrong.

[7] Bobrovsky et al. 1975.

of $\tilde{v}$ is part of the support of v. This is not the case for truncated densities such as the uniform density. Thus a more general bound is necessary.

The WW lower bound is a generalization of the BZ bound. In the sequel, we use the score

$$g_u(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{L(\boldsymbol{x} + \boldsymbol{h}_u, \boldsymbol{x}, \boldsymbol{y})} - \sqrt{L(\boldsymbol{x} - \boldsymbol{h}_u, \boldsymbol{x}, \boldsymbol{y})} \tag{6.12}$$

where $u = 1, \cdots, N$ (cf. **vu2011weiss**; Bell et al. 2006; Weiss et al. 1988 with $s_1 = s_2 = 1/2$). Inserting (6.12) into (2.37), the WW bound is given by

$$\mathrm{E}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right\} \geqslant \boldsymbol{H}\boldsymbol{J}^{-1}\boldsymbol{H}^{\mathrm{T}} \tag{6.13}$$

where

$$[\boldsymbol{J}]_{a,b} := 2\frac{e^{\mu(\boldsymbol{h}_a,-\boldsymbol{h}_b)} - e^{\mu(\boldsymbol{h}_a,\boldsymbol{h}_b)}}{e^{\mu(\boldsymbol{h}_a,\boldsymbol{0})}e^{\mu(\boldsymbol{0},\boldsymbol{h}_b)}}\,, \tag{6.14}$$

with the negative non-metric Bhattacharyya distance (BD) between $v(\boldsymbol{x} + \boldsymbol{h}_a, \boldsymbol{y})$ and $v(\boldsymbol{x} - \boldsymbol{h}_b, \boldsymbol{y})$,[8],

$$\mu(\boldsymbol{h}_a, \boldsymbol{h}_b) = \ln \mathrm{E}\left\{\frac{\sqrt{v(\boldsymbol{x} + \boldsymbol{h}_a, \boldsymbol{y})v(\boldsymbol{x} - \boldsymbol{h}_b, \boldsymbol{y})}}{v(\boldsymbol{x}, \boldsymbol{y})}\right\}\,. \tag{6.15}$$

[8] Basseville 1989; Bhattacharyya 1943; Kailath 1967.

The corresponding Bhattacharyya coefficient $\rho = \exp(\mu(\boldsymbol{h}_a, \boldsymbol{h}_b))$ lies between zero and unity. The more uniform the density $v(\boldsymbol{x}, \boldsymbol{y})$ is, the closer is ρ to unity. The more general WW score in[9] is linked to the more general α-Chernoff divergence and its coefficient[10].

[9] Weiss et al. 1988.

[10] Basseville 1989; Chernoff 1952.

In Appendix B I prove that the BZ bound (Theorem 31) and the CR (Theorems 32 and 33) are limits of the WW bound.

A *Sequential Weiss-Weinstein Bound*

The sequential Weiss-Weinstein bound is the extension of the WW bound to a process $\boldsymbol{x} = \{\boldsymbol{x}_k\}$ with discrete time $k \in \mathbb{N}$. The evolution over time is described by a state-space model

$$\boldsymbol{x}_{k+1} = \Phi(\boldsymbol{x}_k) + \boldsymbol{w}_k\,, \qquad \boldsymbol{w}_k \sim v(\boldsymbol{w}_k)\,, \tag{6.16a}$$

$$\boldsymbol{y}_k = C(\boldsymbol{x}_k) + \boldsymbol{v}_k\,, \qquad \boldsymbol{v}_k \sim v(\boldsymbol{v}_k)\,, \tag{6.16b}$$

with a mapping Φ and state noise $\boldsymbol{w}_k$. We first consider the joint WW bound for the prior and history of states $\boldsymbol{x}_{0:k} = [\boldsymbol{x}_0, \cdots, \boldsymbol{x}_k]^{\mathrm{T}}$ for deriving a recursive algorithm to iteratively compute the WW bound of every time step k. A block-diagonal matrix defines the $kN \times kN$ test-point matrix

$$\boldsymbol{\mathfrak{H}}_k \triangleq \begin{bmatrix} \boldsymbol{H}_0 & & \\ & \ddots & \\ & & \boldsymbol{H}_k \end{bmatrix} = [\boldsymbol{\mathfrak{h}}_0, \cdots, \boldsymbol{\mathfrak{h}}_{kN}]\,. \tag{6.17}$$

The matrix $\boldsymbol{H}_\ell = [\boldsymbol{h}_{\ell,1}, \cdots, \boldsymbol{h}_{\ell,N}]$ corresponds to $\boldsymbol{H}$ in (6.13) at time ℓ. Using the error vector $\boldsymbol{\varepsilon}_{0:k} = \hat{\boldsymbol{x}}_{0:k}(\boldsymbol{y}_{1:k}) - \boldsymbol{x}_{0:k}$, the error covariance matrix

$$\mathrm{E}\left\{\boldsymbol{\varepsilon}_{0:k}\boldsymbol{\varepsilon}_{0:k}^{\mathrm{T}}\right\} \geqslant \boldsymbol{\mathfrak{H}}_k\boldsymbol{\mathfrak{J}}_k^{-1}\boldsymbol{\mathfrak{H}}_k^{\mathrm{T}}\,. \tag{6.18}$$

The overall matrix $\mathfrak{J}_k$ can be partitioned into

$$\mathfrak{J}_k = \left[\begin{array}{cc|c} & & \mathfrak{o} \\ & \mathfrak{A}_{k-1} & \boldsymbol{B}_k^{01} \\ \hline \mathfrak{o} & \boldsymbol{B}_k^{10} & \boldsymbol{B}_k^{11} \end{array}\right] . \tag{6.19}$$

with $\mathfrak{A}_{k-1} = \text{blockdiag}\,(\boldsymbol{A}_0, \cdots, \boldsymbol{A}_{k-1})$. Matrix $\mathfrak{A}_{k-1}$ captures information from the times $[0, k-1]$, $\boldsymbol{B}_k^{11}$ the time k and $\boldsymbol{B}_k^{01} = (\boldsymbol{B}_k^{10})^{\mathrm{T}}$ the transition between them. The $\mathfrak{o}$ matrices in (6.19) are due to the Markovian property, i.e.

$$v(\boldsymbol{x}_{0:k}, \boldsymbol{y}_{1:k}) = v(\boldsymbol{y}_k|\boldsymbol{x}_k)v(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})v(\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_{1:k-1}) .$$

For the time $k = 0$, we have $\boldsymbol{B}_0^{11} = \mathfrak{J}_0 = \boldsymbol{J}_0$, i.e. the bound of the prior.

In the remainder of this section, we derive a recursive update for the WW bound at time k, i.e.

$$\boxed{\mathrm{E}\left\{\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^{\mathrm{T}}\right\} \succcurlyeq \boldsymbol{W}_k \triangleq \boldsymbol{H}_k(\boldsymbol{J}_k)^{-1}\boldsymbol{H}_k^{\mathrm{T}} .} \tag{6.20}$$

In addition to (6.19), we consider the time interval $[0, k+1]$ and partition the overall matrix

$$\mathfrak{J}_{k+1} = \left[\begin{array}{cc|cc|c} & & & \mathfrak{o} & \mathfrak{o} \\ & \mathfrak{A}_{k-1} & & \boldsymbol{D}_{k+1}^{01} & \boldsymbol{0} \\ \hline \mathfrak{o} & \boldsymbol{D}_{k+1}^{10} & & \boldsymbol{D}_{k+1}^{11} & \boldsymbol{D}_{k+1}^{12} \\ \hline \mathfrak{o} & \boldsymbol{0} & & \boldsymbol{D}_{k+1}^{21} & \boldsymbol{D}_{k+1}^{22} \end{array}\right] . \tag{6.21}$$

Matrix $\boldsymbol{D}_{k+1}^{11}$ captures the time k, $\boldsymbol{D}_{k+1}^{22}$ the time $k+1$ and the others the transition between the time instances. Using the Schur complement, the right lowest part of $\mathfrak{J}_{k+1}^{-1}$ is given by the inverse of

$$\begin{aligned} \boldsymbol{J}_{k+1} &= \boldsymbol{D}_{k+1}^{22} - \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{D}_{k+1}^{21} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} & \mathfrak{A}_{k-1} & \mathfrak{o} \\ & & \boldsymbol{D}_k^{01} \\ \mathfrak{o} & \boldsymbol{D}_k^{10} & \boldsymbol{D}_k^{11} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{D}_{k+1}^{12} \end{bmatrix} \\ &= \boldsymbol{D}_{k+1}^{22} - \boldsymbol{D}_{k+1}^{21} \Big(\underbrace{\boldsymbol{D}_{k+1}^{11} - \boldsymbol{D}_{k+1}^{10}\boldsymbol{A}_{k-1}^{-1}\boldsymbol{D}_{k+1}^{01}}_{\equiv \boldsymbol{A}_k} \Big)^{-1} \boldsymbol{D}_{k+1}^{12} . \end{aligned} \tag{6.22}$$

We compare it with

$$\boldsymbol{J}_k = \boldsymbol{D}_k^{22} - \boldsymbol{D}_k^{21}\boldsymbol{A}_{k-1}^{-1}\boldsymbol{D}_k^{12} , \tag{6.23}$$

The sequential update becomes

$$\boldsymbol{A}_k = \boldsymbol{D}_{k+1}^{11} - \boldsymbol{D}_{k+1}^{10}\boldsymbol{A}_{k-1}^{-1}\boldsymbol{D}_{k+1}^{01} , \tag{6.24a}$$

$$\boldsymbol{J}_{k+1} = \boldsymbol{D}_{k+1}^{22} - \boldsymbol{D}_{k+1}^{21}\boldsymbol{A}_k^{-1}\boldsymbol{D}_{k+1}^{12} , \tag{6.24b}$$

for all $k = 0, 1, \cdots$. Matrix $\boldsymbol{A}_{-1}^{-1} := \boldsymbol{0}$ whereas $\boldsymbol{J}_0^{-1}$ is set to the co-variance of the prior. According to (6.15) and (6.21),

$$[\boldsymbol{D}_{k+1}^{ij}]_{m,n} = 2\frac{e^{\mu_1} - e^{\mu_2}}{e^{\mu_3}e^{\mu_4}} , \qquad i, j \in \{0, 1, 2\} , \tag{6.24c}$$

with

$$\mu_1 = \mu(\mathfrak{h}_{(k-2+i)N+m}, -\mathfrak{h}_{(k-2+j)N+n}) , \tag{6.24d}$$
$$\mu_2 = \mu(\mathfrak{h}_{(k-2+i)N+m}, \mathfrak{h}_{(k-2+j)N+n}) , \tag{6.24e}$$
$$\mu_3 = \mu(\mathfrak{h}_{(k-2+i)N+m},) , \tag{6.24f}$$
$$\mu_4 = \mu(, \mathfrak{h}_{(k-2+j)N+n}) , \tag{6.24g}$$

the initial conditions

$$v(\boldsymbol{y}_0|\boldsymbol{x}_0)v(\boldsymbol{x}_0|\boldsymbol{x}_{-1}) := v(\boldsymbol{y}_0)v(\boldsymbol{x}_0) , \tag{6.24h}$$

and (6.24i) in Fig. 6.1.

$$\mu(\mathfrak{h}_a, \mathfrak{h}_b) = \ln \mathrm{E}\left\{ \frac{\prod_{\ell=0}^{k+1} v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a})^{1/2} v(\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a}|\boldsymbol{x}_{\ell-1} + \boldsymbol{h}_{\ell-1,a})^{1/2} v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b})^{1/2} v(\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b}|\boldsymbol{x}_{\ell-1} - \boldsymbol{h}_{\ell-1,b})^{1/2}}{v(\boldsymbol{x}_{0:k+1}, \boldsymbol{y}_{1:k+1})} \right\} \tag{6.24i}$$

Figure 6.1: Negative BD of the SWW bound.

Inspecting (6.24i) and (6.21) leads to

Proposition 1

Given a time-invariant state space model with time-invariant noise distributions and sub-matrices $\boldsymbol{H}_k := \boldsymbol{H}_0$. Then $\boldsymbol{D}_k^{10} = \boldsymbol{D}_k^{21}$ and $\boldsymbol{D}_k^{01} = \boldsymbol{D}_k^{12} = (\boldsymbol{D}_k^{21})^{\mathrm{T}}$ for $k > 2$.

B *Linear Models*

For a linear transition and/or measurement equation, (6.24) can be further simplified. In addition to[11], we will provide and stress important facts that will be used in the subsequent sections for the proofs.

[11] Rapoport et al. 2004b.

Lemma 2

If the expectation in (6.24i) can be factored into independent expectations, i.e.

$$\mu(\mathfrak{h}_a, \mathfrak{h}_b) = \ln(E_0 \cdots E_{k+1}) \tag{6.25}$$

where

$$\begin{aligned} E_\ell \triangleq \mathrm{E}\Big\{ & \frac{v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a})^{1/2} v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b})^{1/2}}{v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell)} \\ & \times \frac{v(\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a}|\boldsymbol{x}_{\ell-1} + \boldsymbol{h}_{\ell-1,a})^{1/2}}{v(\boldsymbol{x}_\ell|\boldsymbol{x}_{\ell-1})} \\ & \times v(\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b}|\boldsymbol{x}_{\ell-1} - \boldsymbol{h}_{\ell-1,b})^{1/2} \Big\} \end{aligned} \tag{6.26}$$

then

$$\boldsymbol{D}_{k+1}^{01} = (\boldsymbol{D}_{k+1}^{10})^{\mathrm{T}} = \boldsymbol{B}_k^{01} = (\boldsymbol{B}_k^{10})^{\mathrm{T}} . \tag{6.27}$$

Proof. Let us focus on (6.24i). We first recast (6.24i) as in Fig. 6.2 and omit all zero vectors $\boldsymbol{h}_{\ell,a}$ and $\boldsymbol{h}_{\ell,b}$. To compute $\boldsymbol{B}_k$, Part (6.28a) and (6.28b) are separable. Part (6.28a) is an expectation $E_{k+1}(\boldsymbol{h}_{k,a}) = E_{k+1}(\boldsymbol{0}) = 1$. To compute $\boldsymbol{D}^{10}_{k+1}$ and $\boldsymbol{D}^{01}_{k+1}$ we assume independent expectations (6.25). Thus, Part (6.28a) and Part (6.28b) are also separable. Part (6.28a) is an expectation $E_{k+1}(\boldsymbol{h}_{k,a})$. For $\mu(,\mathfrak{h}_b)$ in (6.24c), the expectation $E_{k+1}(\boldsymbol{h}_{k,a}) = E_{k+1}(\boldsymbol{0}) = 1$. For $\mu(\mathfrak{h}_a,-\mathfrak{h}_b)$, $\mu(\mathfrak{h}_a,\mathfrak{h}_b)$, and $\mu(\mathfrak{h}_a,)$, the expectations $E_{k+1}(\boldsymbol{h}_{k,a})$ are equal. Thus the $E_{k+1}(\boldsymbol{h}_{k,a})$ cancels in (6.24c). What raises is identical to $\boldsymbol{B}^{01}_k$. ■

$$\mu(\mathfrak{h}_a,\mathfrak{h}_b) = \ln \mathrm{E}\Big\{ \frac{v(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k+\boldsymbol{h}_{k,a})^{1/2} v(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)^{1/2}}{v(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)} \tag{6.28a}$$

$$\times \frac{v(\boldsymbol{y}_k|\boldsymbol{x}_k+\boldsymbol{h}_{k,a})^{1/2} v(\boldsymbol{y}_k|\boldsymbol{x}_k)^{1/2} v(\boldsymbol{x}_k+\boldsymbol{h}_{k,a}|\boldsymbol{x}_{k-1})^{1/2} v(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}-\boldsymbol{h}_{k-1,b})^{1/2}}{v(\boldsymbol{y}_k|\boldsymbol{x}_k) v(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})}$$

$$\times \frac{v(\boldsymbol{y}_{k-1}|\boldsymbol{x}_{k-1})^{1/2} v(\boldsymbol{y}_{k-1}|\boldsymbol{x}_{k-1}-\boldsymbol{h}_{k-1,b})^{1/2} v(\boldsymbol{x}_{k-1}|\boldsymbol{x}_{k-2})^{1/2} v(\boldsymbol{x}_{k-1}-\boldsymbol{h}_{k-1,b}|\boldsymbol{x}_{k-2})^{1/2}}{v(\boldsymbol{y}_{k-1}|\boldsymbol{x}_{k-1}) v(\boldsymbol{x}_{k-1}|\boldsymbol{x}_{k-2})} \frac{v(\boldsymbol{x}_{0:k-2},\boldsymbol{y}_{1:k-2})}{v(\boldsymbol{x}_{0:k-2},\boldsymbol{y}_{1:k-2})} \Big\} \tag{6.28b}$$

Figure 6.2: Negative BD with separated densities.

Lemma 3: Linear transition equation

Given a linear state-transition equation

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Phi}_k \boldsymbol{x}_k + \boldsymbol{w}_k \,, \qquad \boldsymbol{w}_k \sim v(\boldsymbol{w}_k) \,. \tag{6.29}$$

Then the conditions for (6.25) are fulfilled.

Proof. Integrating over the transition densities (as in (2.1))

$$\int \frac{v(\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a}|\boldsymbol{x}_{\ell-1} + \boldsymbol{h}_{\ell-1,a})^{1/2}}{v(\boldsymbol{x}_\ell|\boldsymbol{x}_{\ell-1})} \times v(\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b}|\boldsymbol{x}_{\ell-1} - \boldsymbol{h}_{\ell-1,b})^{1/2} \mathrm{d}P_{\boldsymbol{x}_\ell|\boldsymbol{x}_{\ell-1}}$$

$$= \int \frac{v_{\boldsymbol{w}_\ell}(\boldsymbol{w}_\ell + \boldsymbol{h}_{\ell,a} - \boldsymbol{\Phi}_\ell \boldsymbol{h}_{\ell-1,a})^{1/2}}{v_{\boldsymbol{w}_\ell}(\boldsymbol{w}_\ell)} \times v_{\boldsymbol{w}_\ell}(\boldsymbol{w}_\ell - \boldsymbol{h}_{\ell,b} + \boldsymbol{\Phi}_\ell \boldsymbol{h}_{\ell-1,b})^{1/2} \mathrm{d}P_{\boldsymbol{w}_\ell} \tag{6.30}$$

with $\boldsymbol{w}_\ell = \boldsymbol{x}_\ell - \boldsymbol{\Phi}_k \boldsymbol{x}_{\ell-1}$ and the conditional probability measure $P_{\boldsymbol{x}_\ell|\boldsymbol{x}_{\ell-1}}$. Observe that (6.30) is independent of time $\ell - 1$. ■

Additionally to the transition equation in Lemma 3, we address the measurement equation.

Corollary 4: Linear measurement equation

Given the linear transition equation (6.29) and the measurement equation

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{v}_k\,, \qquad \boldsymbol{v}_k \sim v(\boldsymbol{v}_k)\,. \tag{6.31}$$

Let the state and measurement noise be independent.

Then

$$\mu(\mathfrak{h}_a, \mathfrak{h}_b) = \ln(E_0 \cdots E_{k+1} E_0' \cdots E_{k+1}')\,, \tag{6.32}$$

with

$$E_\ell \triangleq \mathrm{E}\Big\{\frac{v(\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a}|\boldsymbol{x}_{\ell-1} + \boldsymbol{h}_{\ell-1,a})^{1/2}}{v(\boldsymbol{x}_\ell|\boldsymbol{x}_{\ell-1})} \times v(\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b}|\boldsymbol{x}_{\ell-1} - \boldsymbol{h}_{\ell-1,b})^{1/2}\Big\}\,, \tag{6.33a}$$

$$E_\ell' \triangleq \mathrm{E}\Big\{\frac{v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a})^{1/2} v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b})^{1/2}}{v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell)}\Big\}\,, \tag{6.33b}$$

i.e. the expectation over $\boldsymbol{x}_{0:k+1}, \boldsymbol{y}_{1:k+1}$ splits into expectations w.r.t. $\boldsymbol{x}_\ell, \boldsymbol{y}_\ell$.

Proof. The factorization corresponding to the transition densities have been proved with Lemma 2. Dually, the factorization of the integrals concerning the measurement noise are proved in the following.

Due to the additivity of the measurement function,

$$\int \frac{v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a})^{1/2} v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b})^{1/2}}{v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell)} \mathrm{d}P_{\boldsymbol{y}_\ell|\boldsymbol{x}_\ell} = \int \frac{v_{\boldsymbol{v}_\ell}(\boldsymbol{v}_\ell - \boldsymbol{C}_\ell \boldsymbol{h}_{\ell,a})^{1/2} v_{\boldsymbol{v}_\ell}(\boldsymbol{v}_\ell + \boldsymbol{C}_\ell \boldsymbol{h}_{\ell,b})^{1/2}}{v_{\boldsymbol{v}_\ell}(\boldsymbol{v}_\ell)} \mathrm{d}P_{\boldsymbol{v}_\ell} \tag{6.34}$$

with $\boldsymbol{v}_\ell = \boldsymbol{y}_\ell - \boldsymbol{C}_k \boldsymbol{x}_\ell$.

Due to the independence of $\boldsymbol{v}_\ell$ and $\boldsymbol{w}_\ell$ and their independence from time $\ell - 1$, the equation in Fig. 6.2 is separable into factors due to innovation and measurement noise. ∎

Next we assume independence between continuous and discrete random sub-vectors of the innovation noise vector, say

$$v(\boldsymbol{w}_k^{\mathrm{c}}, \boldsymbol{w}_k^{\mathrm{d}}) = f(\boldsymbol{w}_k^{\mathrm{c}}) p(\boldsymbol{w}_k^{\mathrm{d}})\,. \tag{6.35}$$

Corollary 5

With (6.29), (6.31), and (6.35), Equation (6.32) factorizes further into

$$\mu(\mathfrak{h}_a, \mathfrak{h}_b) = \ln(E_0^{\mathrm{c}} \cdots E_{k+1}^{\mathrm{c}} E_0^{\mathrm{d}} \cdots E_{k+1}^{\mathrm{d}} E_0' \cdots E_{k+1}') \,, \tag{6.36}$$

where $E_\cdot^{\mathrm{c}}$ denotes the expectation over continuous probability distributions whereas $E_\cdot^{\mathrm{d}}$ denotes the expectation over discrete

ones.

In the remainder of our paper, we compute the expectations in (6.36) for different noise and priors.

C *Sequential WW Bound for the Linear Transition Model*

Recursion (6.24) simplifies if the system function $\Phi_k = \boldsymbol{\Phi}_k$ is linear. Applying the matrix inversion lemma to (6.19) gives

$$\boldsymbol{J}_k = \boldsymbol{B}_k^{11} - \boldsymbol{B}_k^{10}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k^{01} \,. \tag{6.37}$$

Substitution of (6.37) and (6.27) into (6.22) leads[12] to

[12] Rapoport et al. 2004b.

$$\boxed{\boldsymbol{J}_{k+1} = \boldsymbol{D}_{k+1}^{22} - \boldsymbol{D}_{k+1}^{21}(\boldsymbol{D}_{k+1}^{11} + \boldsymbol{J}_k - \boldsymbol{B}_k^{11})^{-1}\boldsymbol{D}_{k+1}^{12}} \tag{6.38}$$

with $\boldsymbol{B}_0^{11} = \boldsymbol{J}_0$ and $\boldsymbol{B}_k^{11} = \boldsymbol{D}_k^{22}, k = 1, 2, \cdots$.

6.2 *Models*

In the remainder, we use Corollary 4 and 5 to derive analytic SWW bounds for different noise and prior. The solutions are general in the sense that the structure are the same for different distributions. Furthermore, we investigate the SWW bound for the case of states and noise quantized uniformly from continuous distributions. We prove that SWW bounds of continuous and uniformly quantized states are equal for suitable choices of $\mathfrak{H}_k$. We assume uniform quantization with step size Δ_x, i.e. the probability densities are sampled and normalized.

Derived from the linear state-space model (6.29) and (6.31), we define the *quantized model*

$$\begin{aligned}
\boldsymbol{x}_{k+1}^{\mathrm{d}} &= \boldsymbol{\Phi}_k \boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{w}_k^{\mathrm{d}} \,, & \boldsymbol{x}_k^{\mathrm{d}} &\in \mathbb{Z}^N \,, && \text{(6.39a)}\\
\boldsymbol{y}_k &= \Delta_x \boldsymbol{C} \boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{v}_k \,, && && \text{(6.39b)}\\
\boldsymbol{w}_k^{\mathrm{d}} &\sim \frac{1}{c''} f_{w_k}(\boldsymbol{w}_k^{\mathrm{d}} \Delta_x) \,, & \boldsymbol{x}_0^{\mathrm{d}} &\sim \frac{1}{c'''} f_{x_0}(\boldsymbol{x}_0^{\mathrm{d}} \Delta_x) \,, && \text{(6.39c)}
\end{aligned}$$

and the *hybrid model*

$$\begin{aligned}
\boldsymbol{x}_{k+1}^{\mathrm{c}} &= \boldsymbol{\Phi}^{\mathrm{c}} \boldsymbol{x}_k^{\mathrm{c}} + \boldsymbol{\Phi}^{\mathrm{cd}} \boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{w}_k^{\mathrm{c}} \,, & \boldsymbol{x}_k^{\mathrm{c}} &\in \mathbb{R}^{N^{\mathrm{d}}} \,,\\
\boldsymbol{x}_{k+1}^{\mathrm{d}} &= \boldsymbol{\Phi}^{\mathrm{d}} \boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{w}_k^{\mathrm{d}} \,, & \boldsymbol{x}_k^{\mathrm{d}} &\in \mathbb{Z}^{N^{\mathrm{c}}} \,,\\
\boldsymbol{y}_k &= \boldsymbol{C}^{(1)} \boldsymbol{x}_k^{\mathrm{c}} + \boldsymbol{C}^{\mathrm{d}} \boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{v}_k \,, && && \text{(6.40)}\\
\boldsymbol{w}_k^{\mathrm{c}} &\sim f(\boldsymbol{w}_k^{\mathrm{c}}) \,, \boldsymbol{x}_0^{\mathrm{c}} \sim f(\boldsymbol{x}_k^{\mathrm{c}}) \,, & \boldsymbol{v}_k &\sim f_{v_k}(\boldsymbol{v}_k) \,,\\
\boldsymbol{w}_k^{\mathrm{d}} &\sim \frac{1}{c''} f_{w_k^{\mathrm{c}}}(\boldsymbol{w}_k^{\mathrm{d}} \Delta_x) \,, & \boldsymbol{x}_0^{\mathrm{d}} &\sim \frac{1}{c'''} f_{x_0^{\mathrm{c}}}(\boldsymbol{x}_0^{\mathrm{d}} \Delta_x) \,,
\end{aligned}$$

where $f_{x_0^{\mathrm{c}}}$, $f_{w_k^{\mathrm{c}}}$ and f_{v_k} are probability density functions (PDFs) of interest. Factors c'' and c''' normalize the densities. Without loss of generality, we set $N = N^{\mathrm{c}} + N^{\mathrm{d}}$.

6.3 *Analytic Solution for Gaussian Noise / Prior*

In this section we derive lower bounds for Gaussian[13] noise and priors

[13] Leemis et al. 2008.

$\mathrm{N}\{\boldsymbol{m}_{\boldsymbol{x}_k}, \boldsymbol{C}_{\boldsymbol{x}_k}\}$, i.e.

$$f(\boldsymbol{x}_k) \triangleq \frac{1}{(2\pi)^{N/2}\,(\det \boldsymbol{C}_{\boldsymbol{x}_k})^{1/2}} e^{-\frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{m}_{\boldsymbol{x}_k}\|^2_{\boldsymbol{C}_{\boldsymbol{x}_k}^{-1}}} . \tag{6.41}$$

It is convenient to use the definition

$$\rho_{\boldsymbol{x}}^{\mathrm{G}}(\boldsymbol{h}) \triangleq e^{-\frac{1}{8}\|\boldsymbol{h}\|^2_{\boldsymbol{C}_{\boldsymbol{x}}^{-1}}} , \tag{6.42}$$

with the co-variance matrix $\boldsymbol{C}_{\boldsymbol{x}}$ and the weighted norm $\|\boldsymbol{h}\|_{\boldsymbol{C}_{\boldsymbol{x}}^{-1}} \triangleq (\boldsymbol{h}^{\mathrm{T}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{h})^{1/2}$. The weighted norm is induce by the weighted inner-product $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle_{\boldsymbol{C}_{\boldsymbol{x}}^{-1}} \triangleq \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{x}_2$. We make extensive use of Lemmas formulated in Appendix A.1.

Theorem 6: SWW bound / Gaussian distributions

Consider a linear continuous, quantized, or hybrid state-space model. Let the prior, the innovation noise, and the likelihood function be Gaussian (6.41) and statistically independent.

Then the SWW lower bound (6.20) for $k \in \mathbb{N}_0$ is computed by (6.38) and (6.43) in Fig. 6.3 with $\rho(\boldsymbol{h}) := \rho^{\mathrm{G}}(\boldsymbol{h})$.

For the initial $k = 0$, matrix $\boldsymbol{D}_1^{11}$ for (6.38) is given by (6.43d) with the Bhattacharyya coefficient $\rho_{\boldsymbol{x}_0}(\boldsymbol{h}) := \rho_{\boldsymbol{x}_0}^{\mathrm{G}}(\boldsymbol{h})$.

$$\begin{aligned}
\left[\boldsymbol{D}_{k+1}^{11}\right]_{a,b} &= 2\frac{\rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,a} - \boldsymbol{\Phi h}_{k,b})\rho_{\boldsymbol{v}_k}(\boldsymbol{C h}_{k,a} - \boldsymbol{C h}_{k,b})\rho_{\boldsymbol{w}_{k-1}}(\boldsymbol{h}_{k,a} - \boldsymbol{h}_{k,b})}{\rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,a})\rho_{\boldsymbol{v}_k}(\boldsymbol{C h}_{k,a})\rho_{\boldsymbol{w}_{k-1}}(\boldsymbol{h}_{k,a})\rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,b})\rho_{\boldsymbol{v}_k}(\boldsymbol{C h}_{k,b})\rho_{\boldsymbol{w}_{k-1}}(\boldsymbol{h}_{k,b})} \\
&\quad - 2\frac{\rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,a} + \boldsymbol{\Phi h}_{k,b})\rho_{\boldsymbol{v}_k}(\boldsymbol{C h}_{k,a} + \boldsymbol{C h}_{k,b})\rho_{\boldsymbol{w}_{k-1}}(\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k,b})}{\rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,a})\rho_{\boldsymbol{v}_k}(\boldsymbol{C h}_{k,a})\rho_{\boldsymbol{w}_{k-1}}(\boldsymbol{h}_{k,a})\rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,b})\rho_{\boldsymbol{v}_k}(\boldsymbol{C h}_{k,b})\rho_{\boldsymbol{w}_{k-1}}(\boldsymbol{h}_{k,b})} && \text{(6.43a)} \\
\left[\boldsymbol{D}_{k+1}^{12}\right]_{a,b} &= \left[\boldsymbol{D}_{k+1}^{21}\right]_{b,a} = 2\frac{\rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,a} + \boldsymbol{h}_{k+1,b}) - \rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,a} - \boldsymbol{h}_{k+1,b})}{\rho_{\boldsymbol{w}_k}(\boldsymbol{\Phi h}_{k,a})\rho_{\boldsymbol{w}_k}(\boldsymbol{h}_{k+1,b})} && \text{(6.43b)} \\
\left[\boldsymbol{D}_{k+1}^{22}\right]_{a,b} &= 2\frac{\rho_{\boldsymbol{v}_{k+1}}(\boldsymbol{C h}_{k+1,a} - \boldsymbol{C h}_{k+1,b})\rho_{\boldsymbol{w}_k}(\boldsymbol{h}_{k+1,a} - \boldsymbol{h}_{k+1,b}) - \rho_{\boldsymbol{v}_{k+1}}(\boldsymbol{C h}_{k+1,a} + \boldsymbol{C h}_{k+1,b})\rho_{\boldsymbol{w}_k}(\boldsymbol{h}_{k+1,a} + \boldsymbol{h}_{k+1,b})}{\rho_{\boldsymbol{v}_{k+1}}(\boldsymbol{C h}_{k+1,a})\rho_{\boldsymbol{w}_k}(\boldsymbol{h}_{k+1,a})\rho_{\boldsymbol{v}_{k+1}}(\boldsymbol{C h}_{k+1,b})\rho_{\boldsymbol{w}_k}(\boldsymbol{h}_{k+1,b})} && \text{(6.43c)} \\
\left[\boldsymbol{D}_{1}^{11}\right]_{a,b} &= 2\frac{\rho_{\boldsymbol{w}_0}(\boldsymbol{\Phi h}_{0,a} - \boldsymbol{\Phi h}_{0,b})\rho_{\boldsymbol{x}_0}(\boldsymbol{h}_{0,a} - \boldsymbol{h}_{0,b}) - \rho_{\boldsymbol{w}_0}(\boldsymbol{\Phi h}_{0,a} + \boldsymbol{\Phi h}_{0,b})\rho_{\boldsymbol{x}_0}(\boldsymbol{h}_{0,a} + \boldsymbol{h}_{0,b})}{\rho_{\boldsymbol{w}_0}(\boldsymbol{\Phi h}_{0,a})\rho_{\boldsymbol{x}_0}(\boldsymbol{h}_{0,a})\rho_{\boldsymbol{w}_0}(\boldsymbol{\Phi h}_{0,b})\rho_{\boldsymbol{x}_0}(\boldsymbol{h}_{0,b})} && \text{(6.43d)}
\end{aligned}$$

Figure 6.3: General structure of the SWW's solution.

Proof. First, the negative BD (6.24c) is re-cast into (6.24i). According to (6.17), $\boldsymbol{h}_{\ell,a} = 0$ if $a < \ell N$ and $a > (\ell + 1)N$. Thus we may remove them from (6.24i). Due to linearity we invoke Corollary 4. We apply successively Lemmas 19, 20, and 21. Together with Definition 6.42, we get the analytic solution of the negative BD. Inserting this four-times into (6.24c) gives us one element of $\boldsymbol{D}_{k+1}^{12}$.

Further details for $k > 0$ are given in (6.44) in Fig. 6.4. There the negative BD for the elements of $\boldsymbol{D}_{k+1}^{12}$ is derived. From the beginning we consider the existence of a PD, either PDF or probability mass function (PMF). It remains the enumerator integral over states and

$$
\begin{aligned}
\boldsymbol{D}_{k+1}^{12}:\ \mu(\mathfrak{h}_a,\mathfrak{h}_b) &= \ln \iint v(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1})^{1/2} v(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k+\boldsymbol{h}_{k,a})^{1/2} v(\boldsymbol{y}_k|\boldsymbol{x}_k+\boldsymbol{h}_{k,a})^{1/2} v(\boldsymbol{x}_k+\boldsymbol{h}_{k,a}|\boldsymbol{x}_{k-1})^{1/2} \\
&\quad \times v(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}-\boldsymbol{h}_{k+1,b})^{1/2} v(\boldsymbol{x}_{k+1}-\boldsymbol{h}_{k+1,b}|\boldsymbol{x}_x)^{1/2} v(\boldsymbol{y}_k|\boldsymbol{x}_k)^{1/2} v(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})^{1/2} \mathrm{d}P_{\boldsymbol{x}_{k:k+1}} \mathrm{d}P_{\boldsymbol{y}_{k:k+1}} \\
&\overset{\text{Corollary 4}}{=} \ln \iiint f_{v_{k+1}}(\boldsymbol{y}_{k+1}-\boldsymbol{C}\boldsymbol{x}_{k+1})^{1/2} f_{v_{k+1}}(\boldsymbol{y}_{k+1}-\boldsymbol{C}(\boldsymbol{x}_{k+1}-\boldsymbol{h}_{k+1,b}))^{1/2} \mathrm{d}P_{\boldsymbol{y}_{k+1}} \\
&\quad \times f_{w_k}(\boldsymbol{x}_{k+1}-\boldsymbol{\Phi}(\boldsymbol{x}_k+\boldsymbol{h}_{k,a}))^{1/2} f_{w_k}(\boldsymbol{x}_{k+1}-\boldsymbol{h}_{k+1,b}-\boldsymbol{\Phi}\boldsymbol{x}_k)^{1/2} \mathrm{d}P_{\boldsymbol{x}_{k+1}} \\
&\quad \times f_{v_k}(\boldsymbol{y}_k-\boldsymbol{C}(\boldsymbol{x}_k+\boldsymbol{h}_{k,a}))^{1/2} f_{v_k}(\boldsymbol{y}_k-\boldsymbol{C}\boldsymbol{x}_k)^{1/2} \mathrm{d}P_{\boldsymbol{y}_k} \\
&\quad \times f_{w_{k-1}}(\boldsymbol{x}_k+\boldsymbol{h}_{k,a}-\boldsymbol{\Phi}\boldsymbol{x}_{k-1})^{1/2} f_{w_{k-1}}(\boldsymbol{x}_k-\boldsymbol{\Phi}\boldsymbol{x}_{k-1})^{1/2} \mathrm{d}P_{\boldsymbol{x}_k} \\
&\overset{\text{Lemma 19 \& 21}}{=} \ln \rho_{v_{k+1}}(\boldsymbol{C}\boldsymbol{h}_{k+1,b}) \rho_{w_k}(\boldsymbol{\Phi}\boldsymbol{h}_{k,a}-\boldsymbol{h}_{k+1,b}) \rho_{v_k}(\boldsymbol{C}\boldsymbol{h}_{k,a}) \rho_{w_{k-1}}(\boldsymbol{h}_{k,a})
\end{aligned}
\tag{6.44}
$$

Figure 6.4: Derivation of the negative BD for the elements of $\boldsymbol{D}_{k+1}^{12}$.

measurements [first line in (6.44)]. Finally, we insert the last line four times into (6.24c).

The negative BDs for elements of the other matrices are attained similarly [see Figs. 6.5 and 6.6].

For $k = 0$ we use the fact that $v(\boldsymbol{y}_0|\boldsymbol{x}_0) := v_{\boldsymbol{y}_0}(\boldsymbol{y}_0)$, and $v(\boldsymbol{x}_0|\boldsymbol{x}_{-1}) := v(\boldsymbol{x}_0)$. The main difference to $k > 0$ is the equality

$$
\int v_{\boldsymbol{x}_0}(\boldsymbol{x}_0+\boldsymbol{h}_{0,a})^{1/2} v_{\boldsymbol{x}_0}(\boldsymbol{x}_0-\boldsymbol{h}_{0,b})^{1/2} \mathrm{d}P_{\boldsymbol{x}_0} = \rho_{\boldsymbol{x}_0}^{\mathrm{G}}(\boldsymbol{h}_{0,a}+\boldsymbol{h}_{0,b}) \ . \tag{6.47}
$$

The derivation is demonstrated in (6.46).

■

Observe that the hybrid and quantized models assume that $c_1, c_2 \in \{0,1\}$ in (2.2), i.e. the densities are either continuous or discrete. For hybrid densities with $c_1, c_2 \in (0,1)$ and due to (2.3), the integrals would split into discrete and continuous parts.

In the next sections, we observe that the structure of (6.38), (6.43) and (6.43d) is similar for other distributions. Hence, (6.43) is discussed in detail.

Let us compare ρ^{G} with (8) and (60) in Kailath 1967, where we set $p_1(x)$ to Gaussian $\mathrm{N}\{\boldsymbol{0}, \boldsymbol{C}_x\}$ and $p_2(x)$ to $\mathrm{N}\{\boldsymbol{h}, \boldsymbol{C}_x\}$. This shows that Function ρ^{G} is the Bhattacharyya coefficient $\rho \in [0,1]$. In (6.43), ρ quantifies the non-constancy of the densities. The sharper a density is, the lower ρ is.

We observe that the structure of (6.43) stems from (6.24c). Matrix $\boldsymbol{D}_{k+1}^{11}$ reflects the influence of innovation and measurement noise at time k on $k+1$. Therefore, system matrix $\boldsymbol{\Phi}$ and measurement matrix $\boldsymbol{C}$ arise. Matrix $\boldsymbol{D}_{k+1}^{12} = (\boldsymbol{D}_{k+1}^{21})^{\mathrm{T}}$ addresses the transition between k and $k+1$. Thus, it is independent of the measurements and there is no function ρ_{v_k}. Matrix $\boldsymbol{D}_{k+1}^{22}$ addresses only time $k+1$. The structure is the same as of $\boldsymbol{D}_{k+1}^{11}$ except that no $\rho_{w_{k+1}}$ occurs due to causality.

Under one condition, the SWW bound for the continuous, the quantized and the hybrid models are equal:

Proposition 7: Equality of bounds

$$
\begin{aligned}
D_{k+1}^{11}: \quad \mu(\mathfrak{h}_a, \mathfrak{h}_b) &= \ln \iint v_{x_{k+1}|x_k}(x_{k+1}|x_k + h_{k,a})^{1/2} v_{y_k|x_k}(y_k|x_k + h_{k,a})^{1/2} v_{x_k|x_{k-1}}(x_k + h_{k,a}|x_{k-1})^{1/2} \\
&\quad \times v_{x_{k+1}|x_k}(x_{k+1}|x_k - h_{k,b})^{1/2} v_{y_k|x_k}(y_k|x_k - h_{k,b})^{1/2} v_{x_k|x_{k-1}}(x_k - h_{k,b}|x_{k-1})^{1/2} \mathrm{d}\lambda_{x_{k:k+1}} \mathrm{d}\lambda_{y_k} \\
&\overset{\text{Corollary 4}}{=} \ln \int v_{w_k}(x_{k+1} - \Phi(x_k + h_{k,a}) - \Gamma u_k)^{1/2} v_{w_k}(x_{k+1} - \Phi(x_k - h_{k,b}) - \Gamma u_k)^{1/2} \mathrm{d}\lambda_{w_{k+1}} \\
&\quad \times \int v_{v_k}(y_k - C(x_k + h_{k,a}))^{1/2} v_{v_k}(y_k - C(x_k - h_{k,b}))^{1/2} \mathrm{d}\lambda_{v_k} \\
&\quad \times \int v_{w_{k-1}}(x_k + h_{k,a} - \Phi x_{k-1} - \Gamma u_{k-1})^{1/2} v_{w_{k-1}}(x_k - h_{k,b} - \Phi x_{k-1} - \Gamma u_{k-1})^{1/2} \mathrm{d}\lambda_{w_k} \\
&\overset{\text{Lemmas 19 \& 21}}{=} \ln \rho_{C_{w_k}}(\Phi h_{k,a} + \Phi h_{k,b}) \rho_{C_{v_{k+1}}}(C h_{k,a} + C h_{k,b}) \rho_{C_{w_{k-1}}}(h_{k,a} + h_{k,b})
\end{aligned} \tag{6.45a}
$$

$$
\begin{aligned}
D_{k+1}^{21}: \quad \mu(\mathfrak{h}_a, \mathfrak{h}_b) &= \ln \iint v_{y_{k+1}|x_{k+1}}(y_{k+1}|x_{k+1} + h_{k+1,a})^{1/2} v_{x_{k+1}|x_k}(x_{k+1} + h_{k+1,a}|x_k)^{1/2} \\
&\quad \times v_{y_k|x_k}(y_k|x_k)^{1/2} v_{x_k|x_{k-1}}(x_k|x_{k-1})^{1/2} \\
&\quad \times v_{y_{k+1}|x_{k+1}}(y_{k+1}|x_{k+1})^{1/2} v_{x_{k+1}|x_k}(x_{k+1}|x_x - h_{k,b})^{1/2} \\
&\quad \times v_{y_k|x_k}(y_k|x_k - h_{k,b})^{1/2} v_{x_k|x_{k-1}}(x_k - h_{k,b}|x_{k-1})^{1/2} \mathrm{d}\lambda_{x_{k:k+1}} \mathrm{d}\lambda_{y_{k:k+1}} \\
&\overset{\text{Corollary 4}}{=} \ln \int v_{v_{k+1}}(y_{k+1} - C(x_{k+1} + h_{k+1,a}))^{1/2} v_{v_{k+1}}(y_{k+1} - C x_{k+1})^{1/2} \mathrm{d}\lambda_{v_{k+1}} \\
&\quad \times \int v_{w_k}(x_{k+1} + h_{k+1,a} - \Phi x_k - \Gamma u_k)^{1/2} v_{w_k}(x_{k+1} - \Phi(x_k - h_{k,b}) - \Gamma u_k)^{1/2} \mathrm{d}\lambda_{w_{k+1}} \\
&\quad \times \int v_{v_k}(y_k - C x_k)^{1/2} v_{v_k}(y_k - C(x_k - h_{k,b}))^{1/2} \mathrm{d}\lambda_{v_k} \\
&\quad \times \int v_{w_{k-1}}(x_k - \Phi x_{k-1} - \Gamma u_k)^{1/2} v_{w_{k-1}}(x_k - h_{k,b} - \Phi x_{k-1})^{1/2} \mathrm{d}\lambda_{w_k} \\
&\overset{\text{Lemmas 19 \& 21}}{=} \ln \rho_{C_{v_{k+1}}}(C h_{k+1,a}) \rho_{C_{w_k}}(h_{k+1,a} - \Phi h_{k,b}) \rho_{C_{v_k}}(C h_{k,b}) \rho_{C_{w_{k-1}}}(h_{k,b})
\end{aligned} \tag{6.45b}
$$

$$
\begin{aligned}
D_{k+1}^{22}: \quad \mu(\mathfrak{h}_a, \mathfrak{h}_b) &= \ln \iint v_{y_{k+1}|x_{k+1}}(y_{k+1}|x_{k+1} + h_{k+1,a})^{1/2} v_{x_{k+1}|x_k}(x_{k+1} + h_{k+1,a}|x_k)^{1/2} \\
&\quad \times v_{y_{k+1}|x_{k+1}}(y_{k+1}|x_{k+1} - h_{k+1,b})^{1/2} v_{x_{k+1}|x_k}(x_{k+1} - h_{k+1,b}|x_k)^{1/2} \mathrm{d}\lambda_{x_{k+1}} \mathrm{d}\lambda_{y_{k+1}} \\
&\overset{\text{Corollary 4}}{=} \ln \int v_{v_{k+1}}(y_{k+1} - C(x_{k+1} + h_{k+1,a}))^{1/2} v_{v_{k+1}}(y_{k+1} - C(x_{k+1} - h_{k+1,b}))^{1/2} \mathrm{d}\lambda_{v_{k+1}} \\
&\quad \times \int v_{w_k}(x_{k+1} + h_{k+1,a} - \Phi x_k)^{1/2} v_{w_k}(x_{k+1} - h_{k+1,b} - \Phi x_k)^{1/2} \mathrm{d}\lambda_{w_{k+1}} \\
&\overset{\text{Lemmas 19 \& 21}}{=} \ln \rho_{C_{v_{k+1}}}(C h_{k+1,a} + C h_{k+1,b}) \rho_{C_{w_k}}(h_{k+1,a} + h_{k+1,b})
\end{aligned} \tag{6.45c}
$$

Figure 6.5: Negative Bhattacharyya distance for $k > 0$.

$$
\begin{aligned}
\boldsymbol{D}_1^{11}: \qquad \mu(\mathfrak{h}_a, \mathfrak{h}_b) &= \ln \iint v_{x_1|x_0}(x_1|x_0 + h_{0,a})^{1/2} v_{y_0|x_0}(y_0|x_0 + h_{0,a})^{1/2} v_{x_0}(x_0 + h_{0,a})^{1/2} \\
&\quad \times v_{x_1|x_0}(x_1|x_0 - h_{0,b})^{1/2} v_{y_0|x_0}(y_0|x_0 - h_{0,b})^{1/2} v_{x_0}(x_0 - h_{0,b})^{1/2} \mathrm{d}\lambda_{x_{0:1}} \mathrm{d}\lambda_{y_0} \overset{\text{Corollary 4}}{=} \\
&\quad \ln \int v_{w_1}(x_1 - \boldsymbol{\Phi}(x_0 + h_{0,a}) - \boldsymbol{\Gamma} u_0)^{1/2} v_{w_1}(x_1 - \boldsymbol{\Phi}(x_0 - h_{0,b}) - \boldsymbol{\Gamma} u_0)^{1/2} \mathrm{d}\lambda_{w_1} \\
&\quad \times \int v_{x_0}(x_0 + h_{0,a})^{1/2} v_{x_0}(x_0 - h_{0,b})^{1/2} \mathrm{d}\lambda_{x_0} \\
&\overset{\text{Lemma 19}}{=} \ln \rho_{C_{w_1}}(\boldsymbol{\Phi} h_{0,a} + \boldsymbol{\Phi} h_{0,b}) \rho_{C_{x_0}}(h_{0,a} + h_{0,b})
\end{aligned} \tag{6.46}
$$

Figure 6.6: Negative Bhattacharyya distance for $k = 0$.

Given the continuous model (6.29), the quantized model (6.39), and the hybrid model (6.40). Let all distributions be either continuous or quantized Gaussian. If

$$\mathfrak{H}_k^{\mathrm{c}} = \Delta_x \mathfrak{H}_k^{\mathrm{d}} \tag{6.48}$$

then the SWW bound of all three models are equal.

Proof. Consider the proof of Theorem 6. First, we address the prior. We compare the integral with respect to the Lebesque measure for the continuous model with the integral with respect to the counting measure for quantized and hybrid models. If

$$\mathfrak{H}_0^{\mathrm{c}} = \Delta_x \mathfrak{H}_0^{\mathrm{d}} \tag{6.49}$$

then

$$\rho_{x_0}^{\mathrm{G}}(h_{0,a}^{\mathrm{c}} + h_{0,b}^{\mathrm{c}}) = \rho_{x_0}^{\mathrm{G}}(\Delta_x h_{0,a}^{\mathrm{d}} + \Delta_x h_{0,b}^{\mathrm{d}})$$

Next we consider the innovation noise. Inspecting Lemma 19 gives

$$
\begin{aligned}
&\rho_{w_k^{\mathrm{c}}}^{\mathrm{G}}(h_{k+1,a}^{\mathrm{c}} - \boldsymbol{\Phi} h_{k,a}^{\mathrm{c}} + h_{k+1,b}^{\mathrm{c}} - \boldsymbol{\Phi} h_{k,b}^{\mathrm{c}}) \\
&\qquad = \rho_{w_k^{\mathrm{d}}}^{\mathrm{G}}\left(\Delta_x(h_{k+1,a}^{\mathrm{d}} - \boldsymbol{\Phi} h_{k,a}^{\mathrm{d}} + h_{k+1,b}^{\mathrm{d}} - \boldsymbol{\Phi} h_{k,b}^{\mathrm{d}})\right) .
\end{aligned} \tag{6.50}
$$

and this leads to $\boldsymbol{D}_1^{11,\mathrm{c}} = \boldsymbol{D}_1^{11,\mathrm{d}}$ [cf. (6.43d)]. Additionally, inspecting Lemma 21 gives

$$\rho_{v_k}^{\mathrm{G}}(-\boldsymbol{C} h_{k+1,a}^{\mathrm{c}} - \boldsymbol{C} h_{k+1,b}^{\mathrm{c}}) = \rho_{v_k}^{\mathrm{G}}(\Delta_x(-\boldsymbol{C} h_{k+1,a}^{\mathrm{d}} - \boldsymbol{C} h_{k+1,b}^{\mathrm{d}})) . \tag{6.51}$$

The $\boldsymbol{D}_k^{ij}$-matrices become equal for all three models. ∎

6.4 *Analytic Solution for Uniform Distributions*

Similar to previous section, we now provide the analytic SWW bound for multivariate, independent, uniform densities[14,15] Unif$\{r_k, s_k\}$, i.e.

$$v(x_k) \triangleq \prod_{\ell=1}^{N} \frac{1}{[s_k - r_k]_\ell} \mathbb{1}_{[r_k]_\ell \leq [x_k]_\ell, [x_k]_\ell \leq [s_k]_\ell} . \tag{6.52}$$

[14] Leemis et al. 2008.
[15] Either continuous or discrete.

We utilize

$$\rho_x^{\mathrm{U}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \begin{cases} \left[1 - \frac{|[\boldsymbol{h}]_\ell|}{[\varsigma_{x_k}]_\ell}\right], & |[\boldsymbol{h}]_\ell| \leq [\varsigma_{x_k}]_\ell \,, \\ 0 \,, & \text{else} \,. \end{cases} \tag{6.53}$$

The width of the support is

$$\varsigma_{x_k} \triangleq \begin{cases} \boldsymbol{s}_k - \boldsymbol{r}_k \,, & v(\boldsymbol{x}_k) \text{ cont.} \,, \\ \boldsymbol{s}_k - \boldsymbol{r}_k + \mathbf{1} \,, & v(\boldsymbol{x}_k) \text{ disc.} \end{cases} \tag{6.54}$$

For the i.i.d. continuous uniform distribution

$$\begin{aligned} \boldsymbol{r}_k &= \boldsymbol{m}_{x_k} - 1/2 \sqrt{12 \operatorname{diag}(\boldsymbol{C}_{x_k})} \,, \\ \boldsymbol{s}_k &= \boldsymbol{m}_{x_k} + 1/2 \sqrt{12 \operatorname{diag}(\boldsymbol{C}_{x_k})} \,. \end{aligned} \tag{6.55}$$

whereas for the i.i.d. discrete uniform distribution

$$\begin{aligned} \boldsymbol{r}_k &= \boldsymbol{m}_{x_k} - 1/2\mathbf{1} - 1/2 \sqrt{\mathbf{1} + 12 \operatorname{diag}(\boldsymbol{C}_{x_k})} \,, \\ \boldsymbol{s}_k &= \boldsymbol{m}_{x_k} - 1/2\mathbf{1} + 1/2 \sqrt{\mathbf{1} + 12 \operatorname{diag}(\boldsymbol{C}_{x_k})} \,. \end{aligned} \tag{6.56}$$

Vector $\boldsymbol{m}_{x_k}$ denotes the mean of $\boldsymbol{x}_k$ and $\mathbf{1}$ the one-vector. This leads to

$$\varsigma_{x_k} = \begin{cases} \sqrt{12 \operatorname{diag}(\boldsymbol{C}_{x_k})} \,, & v(\boldsymbol{x}_k) \text{ cont.} \,, \\ \sqrt{\mathbf{1} + 12 \operatorname{diag}(\boldsymbol{C}_{x_k})} \,, & v(\boldsymbol{x}_k) \text{ disc.} \,. \end{cases} \tag{6.57}$$

Theorem 8: SWW bound / uniform distributions

Consider a linear continuous, quantized, or hybrid state-space model. Let the innovation noise, the measurement noise and the prior be uniform (6.52) and independent. Furthermore, let the elements of the vectors be statistically independent. Then the SWW bound (6.20) is given by (6.38), (6.43), and (6.43d) on Page 54 where all $\rho := \rho^{\mathrm{U}}$.

Proof. The derivation proceeds as in the proof of Theorem 6 but uses Lemmas 22, 23, and 24 from the appendices. ∎

Corollary 9: Uniform prior, Gaussian noise

Consider a linear continuous, quantized, or hybrid state-transition equation. Let $v(\boldsymbol{x}_0|\boldsymbol{x}_{-1}) := v(\boldsymbol{x}_0)$ be uniform, and both the measurement and the innovation noise be Gaussian.

Then

$$\rho_{\boldsymbol{x}_0} := \rho^{\mathrm{U}}_{\boldsymbol{x}_0}\,, \qquad \rho_{\boldsymbol{w}_0} := \rho^{\mathrm{G}}_{\boldsymbol{w}_0} \tag{6.58}$$

in (6.43d).

Proof. The derivation proceeds as in the proof of Theorem 6 but uses Lemma 23. ∎

The *finite support* of the uniform distribution induces bounds on the test-point matrix $\boldsymbol{H}_k$:

Proposition 10: Box conditions

Given a linear state-space model with multivariate independent uniform noise and prior. Then for all $k \geq 0$

$$-\boldsymbol{\varsigma}_{\boldsymbol{w}_k} \preccurlyeq \boldsymbol{h}_{k,a} \pm \boldsymbol{h}_{k,b} \preccurlyeq \boldsymbol{\varsigma}_{\boldsymbol{w}_k}\,, \tag{6.59a}$$
$$-\boldsymbol{\varsigma}_{\boldsymbol{w}_k} \preccurlyeq \boldsymbol{\Phi}\boldsymbol{h}_{k,a} \pm \boldsymbol{\Phi}\boldsymbol{h}_{k,b} \preccurlyeq \boldsymbol{\varsigma}_{\boldsymbol{w}_k}\,, \tag{6.59b}$$
$$-\boldsymbol{\varsigma}_{\boldsymbol{w}_k} \preccurlyeq \boldsymbol{\Phi}\boldsymbol{h}_{k,a} \pm \boldsymbol{h}_{k,b} \preccurlyeq \boldsymbol{\varsigma}_{\boldsymbol{w}_k}\,, \tag{6.59c}$$
$$-\boldsymbol{\varsigma}_{\boldsymbol{v}_k} \preccurlyeq \boldsymbol{C}\boldsymbol{h}_{k,a} \pm \boldsymbol{C}\boldsymbol{h}_{k,b} \preccurlyeq \boldsymbol{\varsigma}_{\boldsymbol{v}_k}\,. \tag{6.59d}$$

Furthermore,

$$\boldsymbol{h}_{k,a} \neq \boldsymbol{0}, \qquad \boldsymbol{h}_{k,b} \neq \boldsymbol{0}\,. \tag{6.59e}$$

Proof. Bounds (6.59a) to (6.59c) stem from (A.15) in Lemma 22. Bound (6.59b) and (6.59c) stem in a similar way from (A.20) in Lemma 24.

If both $\boldsymbol{h}_{k,a} \to \boldsymbol{0}$ and $\boldsymbol{h}_{k,b} \to \boldsymbol{0}$, the SWW bound collapses to the SCR bound[16]. For uniform distributions, the SCR bound does not exist because of the finite support and this leads to (6.59e). ∎

[16] Rapoport et al. 2007a; Weiss et al. 1988.

The upper bounds are important constraints on $\mathfrak{H}_k$. Assume that $v(\boldsymbol{w}_k), k \in \mathbb{N}_0$, has a much larger support than the support of all other $v(\boldsymbol{w}_{k'}), k' \in \mathbb{N}_0 \setminus \{k\}$. Then the maximum possible $\boldsymbol{H}_{k'}$ is defined by the minimum $\boldsymbol{H}_k$ through (6.59).

Proposition 11: Equality of bounds

Given the continuous model (6.29), the quantized model (6.39), and the hybrid model (6.40). Let the PDs be independently uniformly distributed. If

$$\mathfrak{H}^{\mathrm{c}}_k = \Delta_{\boldsymbol{x}}\mathfrak{H}^{\mathrm{d}}_k\,, \qquad \boldsymbol{\varsigma}^{\mathrm{c}}_{\boldsymbol{x}_k} = \Delta_{\boldsymbol{x}}(\boldsymbol{\varsigma}^{\mathrm{d}}_{\boldsymbol{x}_k} - \boldsymbol{1})\,, \tag{6.60}$$

then the SWW bound of the discrete, the continuous, and the hybrid models are equal.

Proof. The proof proceeds as that of Proposition 7 but uses Lemmas 22 to 24 instead of Lemmas 19 to 21. ■

Further note that the area of the support of $v(\boldsymbol{x})$ increases with increasing time.

6.5 *Analytic Solution for Exponential Distributions*

This section is devoted to the analytic SWW bound for models with multivariate independent exponential[17] densities $\operatorname{Exp}\{\boldsymbol{\alpha}_k\}$, i.e.

[17] Leemis et al. 2008.

$$f(\boldsymbol{x}_k) \triangleq \prod_{\ell=1}^{N} \begin{cases} [\boldsymbol{\alpha}_k]_\ell e^{-[\boldsymbol{\alpha}_k]_\ell [\boldsymbol{x}_k]_\ell}, & [\boldsymbol{x}_k]_\ell \geq 0\,, \\ 0, & \text{else}\,. \end{cases} \tag{6.61}$$

It is convenient to define

$$\rho_{\boldsymbol{x}}^{\mathrm{E}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \begin{cases} e^{-\alpha_\ell/2[\boldsymbol{h}]_\ell}\,, & [\boldsymbol{h}]_\ell \geq 0\,, \\ e^{\alpha_\ell/2[\boldsymbol{h}]_\ell}\,, & [\boldsymbol{h}]_\ell < 0\,, \end{cases} \tag{6.62}$$

with parameter $\alpha_\ell \triangleq [\boldsymbol{\alpha}]_\ell \geq 0$. Note that α_ℓ is the inverse of the mean and standard deviation of $[\boldsymbol{x}]_\ell$.

Theorem 12: SWW bound / exponential distributions

Consider a linear continuous, quantized, or hybrid model. Let the noise and the prior be defined by a multivariate independent exponential distribution (6.61).

Then the SWW bound (6.20) and (6.38) for the state vector $\boldsymbol{x}_k$ is given by (6.43) and (6.43d) on Page 54 where

$$\rho := \rho^{\mathrm{E}}\,. \tag{6.63}$$

Proof. The derivation of the SWW lower bound for Gaussian noise and prior (Theorem 6) leads to the proof: Starting with (6.24c), the BD (6.24i) is computed for the noise under consideration. A re-cast of the latter one is derived in Fig. (6.2) on Page 50. Next we use Corollary 5 and get multiplications of expectations. They compute as in Lemmas 25 and 26. Finally, we get (6.43) on Page 54 whereby $\rho_{\boldsymbol{x}} = \rho_{\boldsymbol{x}}^{\mathrm{E}}$. ■

Corollary 13: Prior

Consider a linear continuous, quantized, or hybrid state-transition equation. Let $v(\boldsymbol{y}_0|\boldsymbol{x}_0) := v_{\boldsymbol{y}_0}(\boldsymbol{y}_0)$ and $v(\boldsymbol{x}_0|\boldsymbol{x}_{-1}) := v(\boldsymbol{x}_0)$ be independently exponentially distributed. Then the SWW bound is given by Theorem 12 except that we utilize $\boldsymbol{D}_1^{11}$ in (6.43d) with

$$\rho_{\boldsymbol{x}_0}(\boldsymbol{h}) := \rho_{\boldsymbol{x}_0}^{\mathrm{E}}(\boldsymbol{h})\,. \tag{6.64}$$

Additionally, Proposition 7 for Gaussian distributions is also applicable for exponential distributions.

6.6 Analytic Solution for Laplace Distributions

Here we present the analytic SWW bound for models with multivariate independent Laplace densities[18] La $\{\boldsymbol{m}_{\boldsymbol{x}_k}, \boldsymbol{b}_k\}$, i.e.

[18] Leemis et al. 2008.

$$f(\boldsymbol{x}_k) \triangleq \prod_{\ell=1}^{N} \frac{1}{2[\boldsymbol{b}_k]_\ell} e^{-\left|[\boldsymbol{x}_k - \boldsymbol{m}_{\boldsymbol{x}_k}]_\ell\right| / 2[\boldsymbol{b}_k]_\ell} \,. \tag{6.65}$$

In the sequel, we utilize

$$\rho_{\boldsymbol{x}}^{\mathrm{L}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \begin{cases} \left(1 + \frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}]_\ell}\right) e^{-\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}]_\ell}} \,, & [\boldsymbol{h}]_\ell > 0 \,, \\ \left(1 - \frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}]_\ell}\right) e^{\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}]_\ell}} \,, & [\boldsymbol{h}]_\ell \leq 0 \,. \end{cases} \tag{6.66}$$

with parameter $\boldsymbol{b} > \boldsymbol{0}$.

Theorem 14: SWW bound / Laplace distributions

Consider a linear model. Let the noise and the prior be defined by a multivariate independent Laplace distribution (6.65).

Then the SWW bound (6.20) and (6.38) for the state vector $\boldsymbol{x}_k$ is given by (6.43) and (6.43d) on Page 54 where

$$\rho := \rho^{\mathrm{L}} \,. \tag{6.67}$$

Proof. The derivation of the WW lower bound for Gaussian noise and prior (Theorem 6) leads to the proof: Starting with (6.24c), the BD (6.24i) is computed for the noise under consideration. A re-cast of the latter one is derived in Fig. 6.2 on Page 50. Next we use Corollary 5 and get multiplications of expectations. They compute as in Lemmas 27 and 28. Finally, we get (6.43) on Page 54 whereby $\rho_{\boldsymbol{x}} = \rho_{\boldsymbol{x}}^{\mathrm{L}}$. ∎

6.7 Categorical Distributions

Multivariate discrete distributions of finite support $[0, N_1] \times \cdots \times [0, N'_N]$ with corresponding probability masses $\{p_{\ell,0}, \cdots, p_{\ell,N'_\ell}\}_{\ell=1}^{N}$ are termed categorical distributions, i.e. discrete random vector $\boldsymbol{x}_k$ has the PMF

$$p(\boldsymbol{x}_k) \triangleq \prod_{\ell=1}^{N} \left[\sum_{\ell'=0}^{N'_\ell} p_{\ell,\ell'} \, \mathbb{1}_{[\boldsymbol{x}_k]_\ell = \ell'} \right] . \tag{6.68}$$

Here, I neglected the time index k. The sum of the masses

$$\sum_{\ell'=0}^{N'_\ell} p_{\ell,\ell'} = 1 \,. \tag{6.69}$$

The computation of (6.33a) and (6.33b) uses the counting measure which is computable due to the finite support of the density. With

definition

$$\rho_x^{\mathrm{C}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \left[\sum_{\ell'=0}^{N'_\ell} p_{\ell'}^{1/2} p_{\ell'+[\boldsymbol{h}]_\ell}^{1/2} \right] \tag{6.70}$$

we formulate

Theorem 15: SWW bound / categorical distributions

Consider a linear model. Let the noise and the prior be defined by a multivariate independent categorical distribution (6.68).

Then the SWW bound (6.20) and (6.38) for the state vector $\boldsymbol{x}_k$ is given by (6.43) on Page 54 where

$$\rho := \rho^{\mathrm{C}} \,. \tag{6.71}$$

Proof. The derivation of the WW lower bound for Gaussian noise and prior (Theorem 6) leads to the proof: Starting with (6.24c), the BD (6.24i) is computed for the noise under consideration. A re-cast of the latter one is derived in Fig. (6.2) on Page 50. Next we use Corollary 5 and get multiplications of expectations. They compute as in Lemmas 29 and 30. Finally, we get (6.43) on Page 54 whereby $\rho_{\boldsymbol{x}} = \rho_x^{\mathrm{C}}$. ∎

Discrete uniform distributions with support $[0, N_1] \times \cdots \times [0, N'_N]$ are categorical distributions with $p_{\ell,\ell'} = p_{\ell,\ell''} \forall \ell', \ell'' \in \{0, \cdots, N'_\ell\}$. Due to the same finite support, the box conditions in Proposition 10 apply for categorical distributions. Note that the support of the state $\boldsymbol{x}_k$ increases with increasing k. Due to the common definition (6.68), the PMF $p(\boldsymbol{x}_k)$ is greater than zero if and only if $\boldsymbol{x} \geqslant \boldsymbol{0}$. Thus, in practical applications, it may be convenient to shift the categorical distribution by its mean.

6.8 Bernoulli Distributions

The Bernoulli distribution $\{p_{\ell,0}, p_{\ell,1} = 1 - p_{\ell,0}\}_{\ell=1}^{N}$ is a special case of the categorical distribution and thus their features are equal. We get

$$\rho_x^{\mathrm{C}}(\boldsymbol{h}) = \rho_x^{\mathrm{B}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \begin{Bmatrix} 1 \,, & [\boldsymbol{h}]_\ell = 0 \\ p_{\ell,0}^{1/2} p_{\ell,1}^{1/2} \,, & [\boldsymbol{h}]_\ell = 1 \\ 0 \,, & \text{else} \end{Bmatrix} \in [0,1] \,. \tag{6.72}$$

This implies that the non-zero elements of test-point matrix $\mathfrak{H}_k$ may only be 1.

6.9 Practical Issues

In the sequel, I address practical problems arising. Note that the test-point matrix $\boldsymbol{H}_k$ defines a specific SWW bound of the SWW family.

A Computational Effort

The non-sequential WW bound (6.18) computes $KN \times KN$ elements of $\mathfrak{J}_k$, where $K \in \mathbb{N}_+$ is the discrete time duration. This bound is quadratic in time whereas the SWW is constant, linear, or quadratic:

Consider the sequential WW bound (6.20) for a linear state-space model with an analytic solution (6.43). With (6.38), it requires the computation of $N \times N$ elements in each of the $3K$ matrices $\boldsymbol{D}_{k+1}^{11}$, $\boldsymbol{D}_{k+1}^{12}$, and $\boldsymbol{D}_{k+1}^{22}$. The number of operations to compute each element is independent of K. Hence, $3KN^2$ elements are computed and the effort is linear in time.

Moreover, if $v(\boldsymbol{y}_k|\boldsymbol{x}_k)$, $v(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$, and the test-point matrix $\boldsymbol{H}_k$ is constant for $k > 1$, then $\boldsymbol{D}_{k+1}^{ij} = \boldsymbol{D}_k^{ij}$, $i, j \in \{1, 2\}$. The computational effort is constant over time.

Consider the general SWW bound (6.20) with (6.24), without closed-form solution ρ, and a state-space model with discrete multivariate distributions of finite support $[r\mathbf{1}, s\mathbf{1}]$. The expectation (6.24i) then simplifies to $K + 1$ sums each $N\varsigma = N(s - r + 1)$ summands. At each $k = 1, \cdots, K$, Equation (6.24i) is computed for 4 matrices $\boldsymbol{D}_{k+1}^{01}$, $\boldsymbol{D}_{k+1}^{11}$, $\boldsymbol{D}_{k+1}^{12}$, and $\boldsymbol{D}_{k+1}^{22}$ of size $N \times N$. Thus, we obtain $4KN^2 \times (K + 1)N\varsigma$, i.e. the effort is quadratic in time (cf. Section 6.7 and Chapter 7).

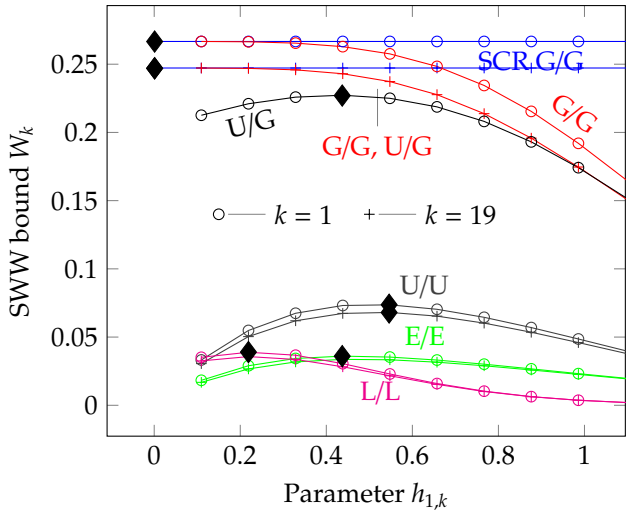


Figure 6.7: Impact of $h_{1,k}$ on the SWW bound for the state x_k (6.73a) for Gaussian prior/noise (G/G), uniform prior (U/G) / Gaussian noise, uniform prior/noise (U/U), exponential prior/noise (E/E), and Laplace prior/noise (L/L). For small $h_{1,k}$, the G/G SWW bounds approach the SCR bounds. Markers indicate the optimal $h_{1,k}$ for maximum SWW.

B Impact of the Test-Point Matrix

The optimal choice of the test-point matrix $\mathfrak{H}_k$ maximizes the WW bound. Even without a general optimal solution to this maximization, we provide some useful guidelines. To keep the discussion simple, an one-dimensional linear transition model is considered with Gaussian, uniform, exponential, and Laplace distributions, i.e.

$$x_{k+1} = x_k + w_k \,, \tag{6.73a}$$

$$y_k = x_k + v_k \,, \tag{6.73b}$$

with $\sigma_{x_0}^2 = 0.4$, $\sigma_{w_k}^2 = 0.4$, $\sigma_{v_k}^2 = 0.4$. For Gaussian and uniform distributions $\mu_{x_0} = \mu_{w_k} = 0$ whereas for the exponential distributions

$\mu_{x_0} = 1/\sigma_{x_0}$, $\mu_{w_k} = 1/\sigma_{w_k}$, and $\mu_{v_k} = 1/\sigma_{v_k}$ and for the Laplace distributions $b_{x_0}^2 = \sigma_{x_0}^2/2$, $b_{w_k}^2 = \sigma_{w_k}^2/2$, and $b_{v_k}^2 = \sigma_{v_k}^2/2$. Fig. 6.7 plots the SWW bounds and SCR[19] bounds vs. $h_{1,k}$ at two time steps $k = 1$ and $k = 19$.

[19] Ristic et al. 2004.

The SCR bounds only exists for the twice differentiable Gaussian density. In that case, when $h_{1,k} \to 0$, the SWW bound approaches the SCR bound which is the optimum. For uniform distributions, the test point $h_{1,k}$ is box constrained by (6.59). Fig. 6.7 shows only the positive part of this allowed interval $(0, s]$ and the point of maximum SWW bound is close to $(s - r)/4$. Notice that at $k = 19$, where the influence of the prior is small, that the uniform prior / Gaussian noise case approaches the all-Gaussian case, i.e. the influence of the prior fades with time. The markers in Fig. 6.7 show the optimal test points $h_{1,k}$ obtained numerically. Observe that the high mode of the exponential density at $w_k = v_k = x_0 = 0_+$ lowers its bound.

For dimensions greater than one, it is more difficult to obtain optimal matrices $\boldsymbol{H}_k, k \in \mathbb{N}_+$. The ℓth row of the system matrix $\boldsymbol{\Phi}$ specifies its dependency on all states. Similarly, the ℓth column of the $\boldsymbol{H}_k$ specifies, which states are considered for the computation of the ℓth-state's SWW bound. This suggests that the positions of non-zero elements in $\boldsymbol{H}_k$ should agree with $\boldsymbol{\Phi}^{\mathrm{T}}$.

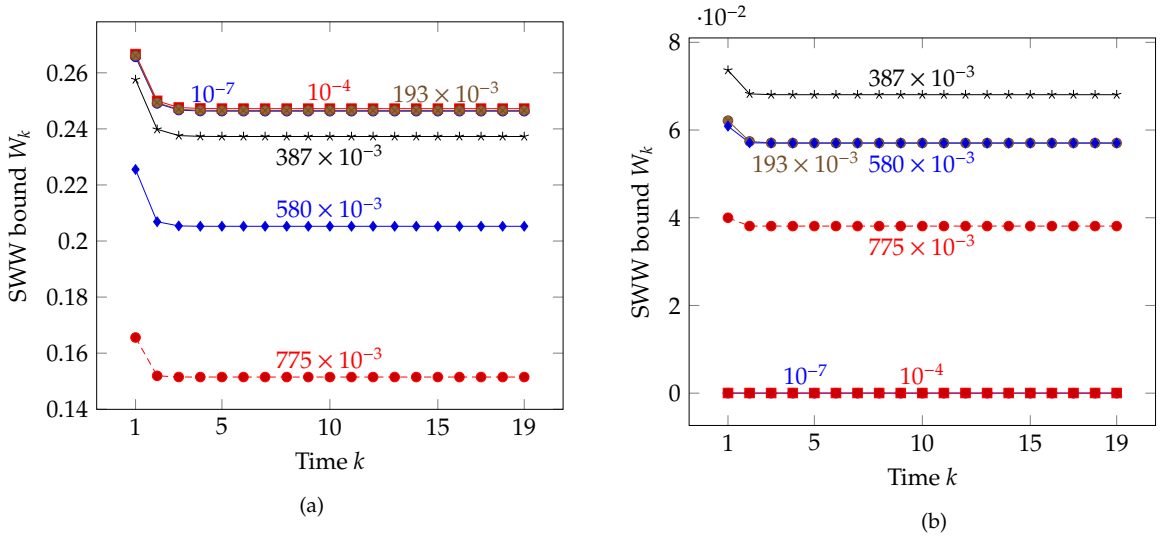


Figure 6.8: SWW lower bounds for different $h_{1,k}$ in (a) for the Gaussian prior/noise case and in (b) for the Uniform prior/noise case (cf. Fig. 6.7).

Another simulation shows the corresponding recursive WW lower bound in the G/G, U/U, E/E, and L/L case for different $h_{1,k}$ in Figs. 6.8 and 6.9.

C *Computation of the Prior*

Consider a hybrid model (6.40) where the state $\boldsymbol{x}_0^{\mathrm{c}}$ is modeled by

$$\boldsymbol{x}_0^{\mathrm{c}} = \sum_{\ell=-K}^{-1} (\boldsymbol{\Phi}^{\mathrm{c}})^{1-\ell} \Phi^{\mathrm{cd}}(\boldsymbol{x}_\ell^{\mathrm{d}}) + \boldsymbol{w}_\ell \,. \tag{6.74}$$

[20] Xaver et al. 2011.

with time horizon K. Function Φ^{cd} might be a source[20] in an acoustic

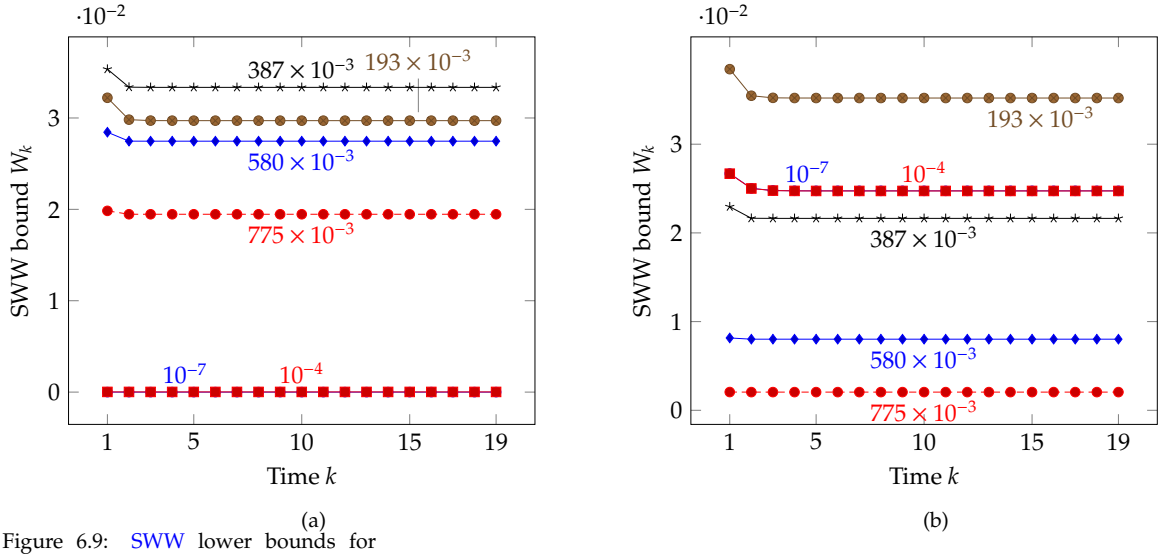


Figure 6.9: SWW lower bounds for different $h_{1,k}$ for (a) the exponential prior/noise and (b) the Laplace prior/noise case.

field with the sum representing the evolution of the corresponding acoustic field during K time steps. The prior $v(\boldsymbol{x}_0^{\mathrm{c}}, \boldsymbol{x}_0^{\mathrm{d}})$ is computed by marginalizing the joint probability density

$$v(\boldsymbol{x}_{-K:0}) = v(\boldsymbol{x}_{-K}) \prod_{\ell=-K}^{-1} v(\boldsymbol{x}_{\ell+1}|\boldsymbol{x}_\ell) \,. \tag{6.75}$$

Fortunately, the explicit computation of the marginal $v(\boldsymbol{x}_0)$ is not necessary in our context since we are only interested in the lower bound of the error variance and not in the PD itself. Therefore we assume a known PD at time $-K$, i.e. it carries over the role of the prior. The SWW bound (6.20) recursively computes the WW bound until time 0. Clearly, in this time interval no measurements influence the bound, i.e.

$$v(\boldsymbol{y}_\ell|\boldsymbol{x}_\ell) = v(\boldsymbol{y}_\ell) \,, \qquad \forall \ell \leq 0 \,. \tag{6.76}$$

Due to the existence of a density and the independence of the states, the expectations (6.33b) reduces to $\int v(\boldsymbol{y}_\ell) \mathrm{d}P_{\boldsymbol{y}_\ell} = 1$ for $\ell = -K, \cdots, 0$. This causes $\rho_{v_\ell} = 1$ in (6.43). Briefly speaking, our approach uses a simplified version of the SWW recursion instead of the explicit computation of the prior at time zero.

D *Partly-deterministic Transition Equations*

An interesting problem occurs when some parts of the transition equation [e.g. (6.16a)] are deterministic, i.e. no noise is added. This results in a singular matrix $\mathrm{E}\left\{\boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}}\right\}$. This causes the Bayesian bounds to become singular (cf. Section 2.5). For SCR bounds, Tichavsky et al. 1998 perform regularization by assuming additive noise with small variance. This may meet most physical problems, so does a discretized physical field.

E Linear Approximation of a Non-linearity

If the state-transition model is non-linear, Lemma 3 does not hold in general. If no closed-form solution is possible, a linear approximation of the system matrix gains importance. I emphasize that the approximation of the system matrix gives a *non-approximated* SWW.

Consider following state-space model,

$$\begin{bmatrix} \boldsymbol{x}^{\mathrm{c}}_{k+1} \\ \boldsymbol{x}^{\mathrm{d}}_{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}^{\mathrm{c}} \boldsymbol{x}^{\mathrm{c}}_{k} + \boldsymbol{\phi}^{\mathrm{cd}}(\boldsymbol{x}_k{}^{\mathrm{d}}) \\ \boldsymbol{\Phi}^{\mathrm{d}} \boldsymbol{x}^{\mathrm{d}}_{k} \end{bmatrix} + \boldsymbol{w}_k \,, \tag{6.77}$$

$$\boldsymbol{y}_k = \boldsymbol{C} \boldsymbol{x}^{\mathrm{c}}_{k} + \boldsymbol{v}_k \,, \tag{6.78}$$

where $\boldsymbol{\phi}^{\mathrm{cd}}$ is a vector-valued non-linear continuous mapping which will be approximated by means of Taylor, i.e.

$$[\boldsymbol{\phi}^{\mathrm{cd}}(\boldsymbol{x}_k)]_i \approx [\boldsymbol{\phi}^{\mathrm{cd}}]_i(\boldsymbol{a}) + (\boldsymbol{x}_k - \boldsymbol{a})^{\mathrm{T}} \nabla [\boldsymbol{\phi}^{\mathrm{cd}}(\boldsymbol{a})]_i \,. \tag{6.79}$$

We set $\boldsymbol{a} = \boldsymbol{x}_k$, such that

$$\begin{aligned} &v_{\boldsymbol{x}_\ell | \boldsymbol{x}_{\ell-1}}(\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a} | \boldsymbol{x}_{\ell-1} + \boldsymbol{h}_{\ell-1,a}) \approx \\ &\quad v_{\boldsymbol{w}^{\mathrm{c}}_\ell}(\boldsymbol{x}^{\mathrm{c}}_\ell + \boldsymbol{h}_{\ell,a} - \boldsymbol{\Phi}^{c} \boldsymbol{x}^{\mathrm{c}}_{\ell-1} - \boldsymbol{\Phi}^{c} \boldsymbol{h}^{\mathrm{c}}_{\ell-1} - \boldsymbol{\phi}^{\mathrm{cd}}(\boldsymbol{x}^{\mathrm{d}}_{\ell-1}) - \sharp) \times \\ &\quad\quad v_{\boldsymbol{w}^{\mathrm{d}}_\ell}(\boldsymbol{x}^{\mathrm{d}}_\ell + \boldsymbol{h}^{\mathrm{d}}_{\ell,a} - \boldsymbol{\Phi}^{d} \boldsymbol{x}^{\mathrm{d}}_{k} - \boldsymbol{\Phi}^{d} \boldsymbol{h}^{\mathrm{d}}_{\ell-1,a}) = \\ &\quad\quad v_{\boldsymbol{w}^{\mathrm{c}}_\ell}(\boldsymbol{w}^{\mathrm{c}}_\ell + \boldsymbol{h}^{\mathrm{c}}_{\ell,a} - \boldsymbol{\Phi}^{c} \boldsymbol{h}^{\mathrm{c}}_{\ell-1} - \sharp) \times \\ &\quad\quad\quad v_{\boldsymbol{w}^{\mathrm{d}}_\ell}(\boldsymbol{x}^{\mathrm{d}}_\ell + \boldsymbol{h}^{\mathrm{d}}_{\ell,a} - \boldsymbol{\Phi}^{d} \boldsymbol{h}^{\mathrm{d}}_{\ell-1,a}) \end{aligned} \tag{6.80}$$

where for $\sharp$ we use either

$$[\sharp_{\max}]_i = \max_{\boldsymbol{x}_k} [\boldsymbol{h}^{\mathrm{d,T}}_{\ell-1,a} \boldsymbol{\phi}^{\mathrm{cd}}(\boldsymbol{x}^{\mathrm{d}}_k)]_i \,, \tag{6.81}$$

$$[\sharp_{\min}]_i = \min_{\boldsymbol{x}_k} [\boldsymbol{h}^{\mathrm{d,T}}_{\ell-1,a} \boldsymbol{\phi}^{\mathrm{cd}}(\boldsymbol{x}^{\mathrm{d}}_k)]_i \,, \tag{6.82}$$

and $\boldsymbol{h}_{\ell-1,a} = [\boldsymbol{h}^{\mathrm{c,T}}_{\ell-1,a}, \boldsymbol{h}^{\mathrm{d,T}}_{\ell-1,a}]^{\mathrm{T}}$. What we need are strict bounds for the discrete and continuous state vectors using the approximation above.

Theorem 16

Consider Model (6.78) with the Taylor approximation (6.79). Then $\sharp := \sharp_{\max}$ gives a SWW lower bound for the discrete state vector $\boldsymbol{x}^{\mathrm{d}}_k$ and $\sharp := \sharp_{\min}$ that for the continuous state vector $\boldsymbol{x}^{\mathrm{c}}_k$.

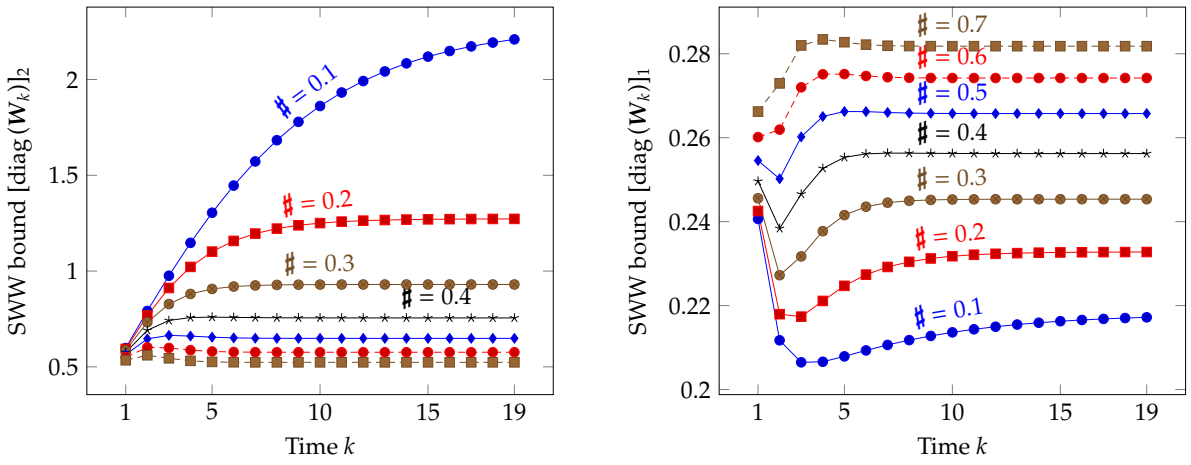
Proof. Let us focus on the second part of the state-space model, i.e. the part corresponding to $\boldsymbol{x}^{\mathrm{d}}_k$. Consider the worst case, $\sharp := 0$. Then there is no coupling, and the measurements do not influence the inference. Therefore the lower bound for $\boldsymbol{x}^{\mathrm{d}}_k$ is not corrected. On the other hand $\boldsymbol{x}^{\mathrm{c}}_k$ is independent of $\boldsymbol{x}^{\mathrm{d}}_k$ and therefore its ambiguity is not increased.

Let $\sharp := \sharp_{\max}$. Here the measurements correct the bound of $\boldsymbol{x}^{\mathrm{d}}_k$ because of the strong coupling to $\boldsymbol{x}^{\mathrm{c}}_k$: the lower bound of $\boldsymbol{x}^{\mathrm{d}}_k$ reduces. Otherwise the ambiguity of $\boldsymbol{x}^{\mathrm{d}}_k$ causes the lower bound of $\boldsymbol{x}^{\mathrm{c}}_k$ to be higher. ∎

Table 6.1: Settings.

$\boldsymbol{\Phi}^{\mathrm{c}}, \boldsymbol{\Phi}^{\mathrm{d}}, \boldsymbol{C}$	$= 1$	$\sharp$	$= 0.1, \cdots, 0.7$
$\boldsymbol{C}_{x_k}$	$= \mathrm{diag}\,(0.4, 0.4)$	$\boldsymbol{C}_{w_k}$	$= \mathrm{diag}\,(0.2, 0.2)$
$\boldsymbol{C}_{v_k}$	$= \mathrm{diag}\,(0.4, 0.4)$	$\boldsymbol{h}_k$	$= 0.01\boldsymbol{I}$

Fig. 6.10 illustrates the impact of different vectors $\sharp$ using settings in Table 6.1 where the noise and prior are zero-mean Gaussian. With increasing $\sharp$, Fig. 6.10a illustrates the decreasing bound for $\boldsymbol{x}_k^{\mathrm{d}}$ whereas the bound for $\boldsymbol{x}_k^{\mathrm{c}}$ increases conversely in Fig. 6.10b. Notice that al-



(a) (b)

Figure 6.10: SWW bound for corresponding (a) discrete and (b) continuous uncoupled States and Gaussian distributions.

though Theorem 16 gives lower bounds there is no statement on the quality of the bounds. The better the approximation the tighter the bound. Beyond the analytic solution, the approximation avoids the computation of $\boldsymbol{D}_{k+1}^{10}, \boldsymbol{D}_{k+1}^{01}$.

6.10 Simulation

In this section, the following linear state-space model demonstrates the bounds derived for different distributions:

$$\boldsymbol{x}_{k+1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{x}_k + \boldsymbol{w}_k \,, \tag{6.83a}$$

$$\boldsymbol{y}_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{x}_k + \boldsymbol{v}_k \,. \tag{6.83b}$$

The first state $[\boldsymbol{x}_k^{\mathrm{c}}]_1$ depends on itself and the second $[\boldsymbol{x}_k^{\mathrm{c}}]_2$ whereas the others depend only on themselves. Equation (6.83b) measures $[\boldsymbol{x}_k^{\mathrm{c}}]_1$ and $[\boldsymbol{x}_k^{\mathrm{c}}]_3$.

We plot the diagonals of the SWW bound $\boldsymbol{W}_k$ with the test-point matrix

$$\boldsymbol{H}_k = \begin{bmatrix} \boldsymbol{h}_{k,1} & \boldsymbol{h}_{k,2} & \boldsymbol{h}_{k,3} \end{bmatrix} := \begin{bmatrix} h_{\mathrm{opt}} & 0 & 0 \\ h_{\mathrm{opt}}/10 & h_{\mathrm{opt}} & 0 \\ 0 & 0 & h_{\mathrm{opt}} \end{bmatrix} . \tag{6.84}$$

The computation of Element $[\boldsymbol{W}_k]_{a,b}$ utilizes vectors $\boldsymbol{h}_{k,a}$ and $\boldsymbol{h}_{k,b}$ for $a, b \in \{1, 2, 3\}$. Although Fig. 6.11 show only the diagonals of $\boldsymbol{W}_k$, i.e. the bound on the error variances, Update (6.38) demands for the non-diagonal elements of $\boldsymbol{W}_k$.

We discuss four settings for continuous distributions in Fig. 6.11: the all-Gaussian, the uniform prior / Gaussian noise, the all-uniform, and the all-exponential case. Their covariance matrices are $\boldsymbol{C}_{x_0} = \boldsymbol{C}_{w_k} = 0.4\boldsymbol{I}$ and $\boldsymbol{C}_{v_k} = 0.4\boldsymbol{I}$. The Gaussian and uniform distributions have zero-mean. The means of the exponential distributions equal their standard deviations. According to Fig. 6.7, the optimal $h_{\text{opt}} = h^{\text{c}}_{\text{opt}} = 0.55$ for the all-uniform case.

The all-Gaussian case is plotted in Fig. 6.11a. The SCR bound exists and is shown as reference[21]. If $h_{\text{opt}} \to 0$, the SWW bound approaches the SCR bound. The SCR bound is achievable using a Kalman filter. State $\ell = 3$ is observed and has the lowest bound. State $\ell = 1$ depends additionally on state $\ell = 2$ and hence has a higher bound. State $\ell = 2$ is not directly observed and thus has the highest bound.

[21] Ristic et al. 2004.

The all-uniform case, Fig. 6.11b, is similar to the all-Gaussian case except that the observed states are close together. The test-point h_{opt} is the value of the highest SWW bound. The all-exponential case is demonstrated in Fig. 6.11c shows the same behavior. Fig. 6.11d shows the SWW bound for uniform prior and Gaussian noise. Compared with the all-Gaussian case we only see a difference at time $k = 1$ (initial phase).

We use model (6.83) and test-point matrix (6.84) again for quantized Gaussian, quantized (discrete) uniform, and quantized exponential densities. We seek for settings leading to the same SWW bounds for quantized and continuous distributions.

The width (6.57) of the continuous uniform density computes to $\varsigma^{\text{c}}_{x_k} = 2.19 \times \mathbf{1}$. Let the discrete uniform density have a width of the support $\varsigma^{\text{d}}_{x_k} = 20 \times \mathbf{1}$. Then with (6.60), the quantization step size is $\Delta_x = 0.9$ and the covariance matrix of the discrete uniform distributions are $\boldsymbol{C}_{x_0} = \boldsymbol{C}_{w_k} = 1/12(\varsigma^{\text{d},2}_{x_k} - 1)\boldsymbol{I}$ and $\boldsymbol{C}_{v_k} = 1/12(\varsigma^{\text{d},2}_{x_k} - 1)\boldsymbol{I}$. With (6.60), the discrete $h_{\text{opt}} := h^{\text{d}}_{\text{opt}} = h^{\text{c}}_{\text{opt}}/\Delta_x = 5$. Now the SWW bound for the discrete (quantized) uniform distribution Unif$\{-\varsigma^{\text{d}}_{x_k}/2, \varsigma^{\text{d}}_{x_k}/2\}$ equals that of the continuous case.

Since Gaussian and exponential densities have infinite support, using their quantized versions with quantization step size $\Delta_x = 0.9$ and $h^{\text{d}}_{\text{opt}} = 5$ give the same SWWs as the continuous cases (see Fig. 6.11).

6.11 *Conclusions of this Chapter*

The family of Weiss-Weinstein bounds enables the use of hybrid discrete and continuous state-vectors. The use of the Chernoff or Bhattacharyya distance allows to give a general recursion for the sequential bound. If there is a closed-form solution of the distance for particular probability densities, analytic solutions of the SWW bound exists. Solutions of the continuous Gaussian, uniform, exponential, and Laplace distributions and of the discrete uniform, categorical

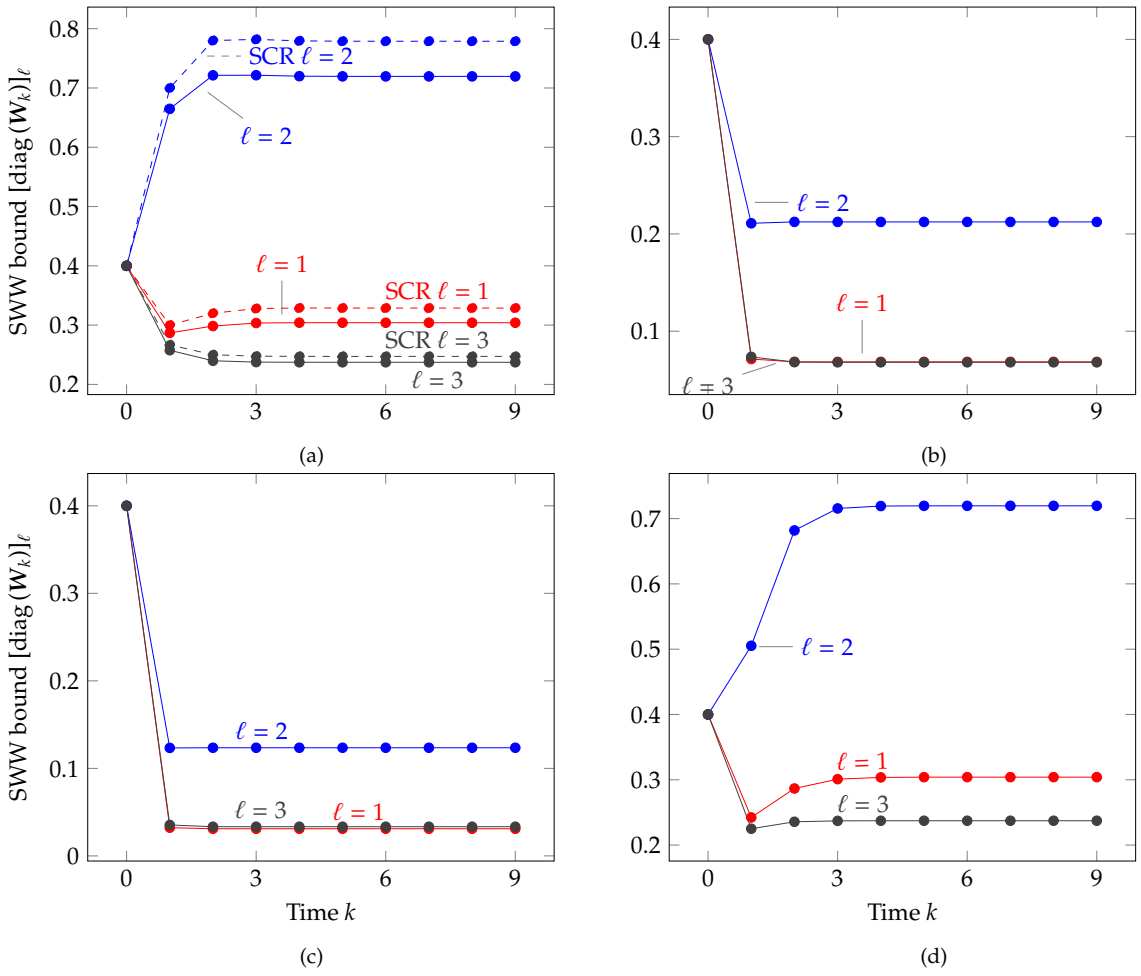


Figure 6.11: The SWW lower bounds for corresponding $[x_k]_\ell$ [cf. (6.83)] for (a) Gaussian prior/noise, (b) uniform prior/noise, (c) exponential prior/noise, and (d) uniform prior / Gaussian noise. Time $k = 0$ shows the prior error variance.

(Bernoulli), quantized Gaussian, and quantized exponential distributions are summarized in Tab. 8.1.

The optimal elements of $\mathfrak{H}_k$ for uniform distributions are in contrast to small values for Gaussian distributions. The finite support of uniform densities causes box constrains on $\mathfrak{H}_k$. The shape of the system matrix describes the dependency between states. Thus, it influences the choice of $\mathfrak{H}_k$ that describes the influence of noise on these states. The SWW bounds for continuous and quantized states are equal for specific choices of the bound's test points $\mathfrak{H}_k$. I highlight that the derivations concerning quantized states are applicable for quantized measurements as well.

Further results are related to practical issues. For linear state-space models with analytic solutions the computational effort is linear in time. Additionally, if the statistics are constant, then the effort is constant. If the prior density stems from a recursion, it is possible to compute the SWW bound without explicit prior. Regularization of deterministic sub-state equations is applicable in this framework.

7
Bounds on States of the Localization Model

This chapter is devoted to the sequential Weiss-Weinstein (SWW) bound for my state-space model including the non-linear source function (Chapter 3). Despite the possibility of the approximation of non-linearities (see Section 6.9.E), I am interested in a far tighter bound.

I address following issues formerly published in Xaver et al. 2012a:

1. Conditions for the existence of analytic SWW solutions.
2. Behavior of the SWW bound on pressure states and sensor location (cf. Section 3).
3. Sensitivity of the SWW bound to the distortion of the channel (cf. Section 5).

In Section 7.1, I recap and concretize the state-space model. The imperfect channel is specified in Section 7.2. Next, I give an analytic SWW bound (Section 7.3) and numerical results (Section 7.4). This chapter makes heavily use of the SWW's integral formulation and the Bhattacharyya coefficients computed in Chapter 6.

7.1 Model

In this section, I reformulate the state-space model of an acoustic source in a hallway (3.23), (3.33), and (3.34) with source Model II (3.32). Here, I denote the model of the acoustic wave field by $\boldsymbol{\Phi}^{\mathrm{c}}$, the non-linear vector-valued source by $\boldsymbol{\phi}^{\mathrm{cd}}$, the path of one source by $\boldsymbol{\Phi}^{\mathrm{d}}$, and the measurements by $\boldsymbol{C}$. This emphasizes the general solution of the SWW bounds beyond my illustrative application (see Fig. 1.1) and the way of looking at the problem. We get

$$\begin{bmatrix} \boldsymbol{x}_{k+1}^{\mathrm{c}} \\ \boldsymbol{x}_{k+1}^{\mathrm{d}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_k{}^{\mathrm{c}} \boldsymbol{x}_k^{\mathrm{c}} + \boldsymbol{\phi}^{\mathrm{cd}}(\boldsymbol{x}_k^{\mathrm{d}}) \\ \boldsymbol{\Phi}^{\mathrm{d}} \boldsymbol{x}_k^{\mathrm{d}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{w}_k^{\mathrm{c}} \\ \boldsymbol{w}_k^{\mathrm{d}} \end{bmatrix}, \tag{7.1a}$$

$$\boldsymbol{y}_k = \boldsymbol{C} \boldsymbol{x}_k^{\mathrm{c}} + \boldsymbol{v}_k, \tag{7.1b}$$

with the continuous N-dimensional state vector $\boldsymbol{x}_k^{\mathrm{c}} = [\boldsymbol{q}_k^{\mathrm{T}}, \boldsymbol{p}_k{}^{\mathrm{T}}]^{\mathrm{T}}$ and the discrete source position vector $\boldsymbol{x}_k^{\mathrm{d}}$ (cf. Model II in Section 3.4).

The vector x_k^{c} captures the pressure on a grid by p_k whereas its time derivative is denoted by q_k. Both are distributed over space and p_k represents the sampled wave field. The measurements are denoted by y_k. The continuous random vector w_k^{c} models the process noise of the acoustic wave, vector w_k^{d} the discrete position jitter, and v_k the continuous measurement noise.

In the decentralized case, $\mathbf{\Phi}^{\mathrm{c}}$ is decomposed into sub-matrices. Every sensor m has an estimator using following sub-state space model (cf. Section 4.B):

$$\begin{bmatrix} x_{k+1}^{\mathrm{c},(m)} \\ x_{k+1}^{\mathrm{d},(m)} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}^{\mathrm{c},(m)} x_k^{\mathrm{c},(m)} + \bar{x}_k^{\mathrm{c},(m)} + \phi^{\mathrm{cd}}(x_k^{\mathrm{d},(m)}) \\ \mathbf{\Phi}^{\mathrm{d}} x_k^{\mathrm{d},(m)} \end{bmatrix} + \begin{bmatrix} w_k^{\mathrm{c},(m)} \\ w_k^{\mathrm{d},(m)} \end{bmatrix}, \tag{7.2a}$$

$$y_k^{(m)} = C^{(m)} x_k^{\mathrm{c},(\mathrm{m})} + v_k^{(m)}. \tag{7.2b}$$

Vectors $x_k^{\mathrm{c},(m)}, m = 1, \cdots, M$ are disjoint sub-vectors of x_k^{c} whereas $\bar{x}_k^{\mathrm{c},(m)}$ is the coupling term between the different sub-state space models (see Fig. 7.1 for the one-dimensional case). Similar for vector $y_k^{(m)}$.

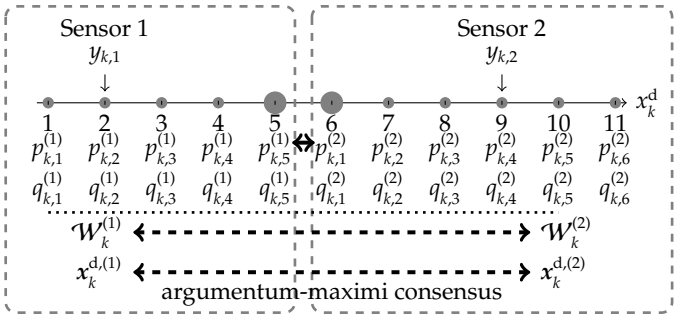


Figure 7.1: Decentralized model capturing an 1-D acoustic wave equation. The sample points of the wave model are assigned to two sensors. Arrows denote message exchange. The weights $\mathcal{W}_k^{(m)}$ of the PF and the local belief $x_k^{\mathrm{d},(m)}$ are not exchanged at every time k. The radii of the filled circles denote different noise variance.

Every sensor m has its own copy $x_k^{\mathrm{c},(m)}$ of the global source position vector x_k^{d}. Thus the local maximum a-posteriori (MAP) estimator of sensor m has a local estimate of the position. A final argumentum-maximi consensus algorithm ensures a joint belief of the position (cf. Chapter 4). Observe that the linear sub-structures of (7.1) and (7.2) allow the use of the very efficient marginalized particle filter[1].

[1] Schön et al. 2005.

Chapters 4 and 5 assume perfect channels between the sensors, i.e. the coupling term $\bar{x}_k^{\mathrm{c},(m)}$ in (7.2) is not perturbed. In the sequel we relax this assumption and analyze the impact of imperfect channels.

7.2 *Imperfect Channels Between Sensors*

Per se, there are no approximation errors inherent in the decentralized model (7.2) compared to (7.1). Thus, I use the global model (7.1) for the following analysis of the estimation error. Additionally, I assume an imperfect channel in the decentralized model which adds distortion to $\bar{x}_k^{\mathrm{c},(m)}$ in (7.2). This corresponds to additional noise in the centralized model (7.1), i.e. the imperfect channel perturbs some elements of x_k^{c}.

[2] Xaver et al. 2012b.

On that account, let us assume[2] that decorrelated signals ι_k =

decorr $\bar{x}_k^{c,(m)}$ are exchanged instead of $\bar{x}_k^{c,(m)}$. At the sensor's boundary, noise $\sim \mathrm{N}\{0, D\}$ is added to w_k^c.

Ie use following worst-case assumptions:

- Elements of vector ι_k are Gaussian distributed and are transmitted over an additive Gaussian channel with noise variance $\sigma^2_{\text{channel}}$.
- The channel can be used once per ι_k.
- The average channel input power P_{channel} is limited.

Then the distortion[3]

[3] Goertz 2007.

$$D = \frac{\sigma^2_{\iota_k}}{1 + P_{\text{channel}}/\sigma^2_{\text{channel}}} . \tag{7.3}$$

7.3 *Sequential Weiss-Weinstein Bound*

The hybrid continuous/discrete probability densities and the finite support of the source-position density demand for a more general Bayesian bound than the sequential Cramér-Rao bound: the sequential Weiss-Weinstein bound (Chapter 6). Additionally, I must use a general definition of the expectation operator (2.1) with some measurable function $g(\boldsymbol{x}, \boldsymbol{y})$ and the continuous/discrete probability measure $P_{\boldsymbol{x},\boldsymbol{y}}$. I define the N-dimensional vector $\boldsymbol{x}_k \triangleq [\boldsymbol{x}_k^{c,\mathrm{T}}, \boldsymbol{x}_k^{d,\mathrm{T}}]^{\mathrm{T}}$. The sequential WW bound

$$\boldsymbol{W}_k = \boldsymbol{H}_k \boldsymbol{J}_k^{-1} \boldsymbol{H}_k^{\mathrm{T}} \tag{7.4}$$

of the estimation error $\boldsymbol{\varepsilon}_k = \hat{\boldsymbol{x}}_k(\boldsymbol{y}_k) - \boldsymbol{x}_k$ lower bounds the error covariance[4]

[4] Rapoport et al. 2007a; Xaver et al. 2013.

$$\mathrm{E}\left\{\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^{\mathrm{T}}\right\} \succcurlyeq \boldsymbol{W}_k . \tag{7.5}$$

Symbol $\succcurlyeq$ denotes that the right side subtracted from the left gives a positive semi-definite matrix. Matrix $\boldsymbol{W}_k$ correspond to the inverse Fisher information matrices of the sequential Cramér-Rao bound. Test-point matrix $\boldsymbol{H}_k = [\boldsymbol{h}_{k,1}, \cdots, \boldsymbol{h}_{k,N}]$ is a parameters of the SWW bound at time k. Optimal matrices $\boldsymbol{H}_k$ give the maximum lower bound of all possible lower bounds of the SWW family.

Matrix $\boldsymbol{J}_k$ is sequentially computed by equations (6.24) on Page 48. My aim is to derive the analytic solution of the negative Bhattacharyya distance[5] μ for (7.1).

[5] Kailath 1967.

Theorem 17: SWW of non-linear transition model

Given the non-linear state-space model (7.1) where $\boldsymbol{x}_k^c \sim f(\boldsymbol{x}_0^c)$, $\boldsymbol{x}_0^d \sim p(\boldsymbol{x}_0^d)$, $\boldsymbol{w}_k^c \sim f(\boldsymbol{w}_k^c)$, $\boldsymbol{w}_k^d \sim p(\boldsymbol{w}_k^d)$ and $\boldsymbol{v}_k \sim f(\boldsymbol{v}_k)$.

Then the negative Bhattacharyya distances for the elements of all matrices $\boldsymbol{D}_{k+1}^{ij}$ are given in Figs. 7.2 and 7.3 using the Bhattacharyya coefficient

$$\rho_{\boldsymbol{x}}(\boldsymbol{h}) = \rho_{\boldsymbol{x}}^{\mathrm{G}}(\boldsymbol{h}) = e^{-\frac{1}{8}\|\boldsymbol{h}\|^2_{\mathrm{C}_{\boldsymbol{x}}}} \in [0, 1] \tag{7.6}$$

and the initial conditions

$$v(\boldsymbol{y}_0|\boldsymbol{x}_0)v(\boldsymbol{x}_0|\boldsymbol{x}_{-1}) := v(\boldsymbol{y}_0)v(\boldsymbol{x}_0)\,. \tag{7.7}$$

$$\begin{aligned}
D_1^{11} : \mu(\mathfrak{h}_a,\mathfrak{h}_b) &= \ln \sum_{\boldsymbol{x}_1^{\mathrm{d}}}\sum_{\boldsymbol{x}_0^{\mathrm{d}}} \rho_{w_0^{\mathrm{c}}}\left(-\boldsymbol{\Phi}^{\mathrm{c}}\boldsymbol{h}_{0,a}^{\mathrm{c}} - \boldsymbol{\Phi}^{\mathrm{c}}\boldsymbol{h}_{0,b}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(\boldsymbol{x}_0^{\mathrm{d}} + \boldsymbol{h}_{0,a}^{\mathrm{d}}) + \Phi^{\mathrm{cd}}(\boldsymbol{x}_0^{\mathrm{d}} - \boldsymbol{h}_{0,b}^{\mathrm{d}})\right)\rho_{x_0^{\mathrm{c}}}(\boldsymbol{h}_{0,a}^{\mathrm{c}} + \boldsymbol{h}_{0,b}^{\mathrm{c}}) \\
&\quad \times p(\boldsymbol{x}_1^{\mathrm{d}}|\boldsymbol{x}_0^{\mathrm{d}} + \boldsymbol{h}_{0,a}^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_1^{\mathrm{d}}|\boldsymbol{x}_0^{\mathrm{d}} - \boldsymbol{h}_{0,b}^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_0^{\mathrm{d}} + \boldsymbol{h}_{0,a}^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_0^{\mathrm{d}} - \boldsymbol{h}_{0,b}^{\mathrm{d}})^{1/2} \qquad (7.8a) \\
D_2^{11} : \mu(\mathfrak{h}_a,\mathfrak{h}_b) &= \ln \rho_{v_1^{\mathrm{c}}}(\boldsymbol{C}\boldsymbol{h}_{1,a} + \boldsymbol{C}\boldsymbol{h}_{1,b}) \\
&\quad \times \sum_{\boldsymbol{x}_2^{\mathrm{d}}} \cdots \sum_{\boldsymbol{x}_0^{\mathrm{d}}} \rho_{w_1^{\mathrm{c}}}\left(-\boldsymbol{\Phi}^{\mathrm{c}}\boldsymbol{h}_{1,a}^{\mathrm{c}} - \boldsymbol{\Phi}^{\mathrm{c}}\boldsymbol{h}_{1,b}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(\boldsymbol{x}_1^{\mathrm{d}} + \boldsymbol{h}_{1,a}^{\mathrm{d}}) + \Phi^{\mathrm{cd}}(\boldsymbol{x}_1^{\mathrm{d}} - \boldsymbol{h}_{1,b}^{\mathrm{d}})\right)\rho_{w_0^{\mathrm{c}}}(\boldsymbol{h}_{1,a}^{\mathrm{c}} + \boldsymbol{h}_{1,b}^{\mathrm{c}}) \\
&\quad \times p(\boldsymbol{x}_2^{\mathrm{d}}|\boldsymbol{x}_1^{\mathrm{d}} + \boldsymbol{h}_{1,a}^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_2^{\mathrm{d}}|\boldsymbol{x}_1^{\mathrm{d}} - \boldsymbol{h}_{1,b}^{\mathrm{d}})^{1/2} \\
&\quad \times p(\boldsymbol{x}_1^{\mathrm{d}} + \boldsymbol{h}_{1,a}^{\mathrm{d}}|\boldsymbol{x}_0^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_1^{\mathrm{d}} - \boldsymbol{h}_{1,b}^{\mathrm{d}}|\boldsymbol{x}_0^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_0^{\mathrm{d}}) \qquad (7.8b) \\
D_{k+1}^{11} : \mu(\mathfrak{h}_a,\mathfrak{h}_b) &= \ln \rho_{v_k^{\mathrm{c}}}(\boldsymbol{C}\boldsymbol{h}_{k,a} + \boldsymbol{C}\boldsymbol{h}_{k,b}) \\
&\quad \times \sum_{\boldsymbol{x}_{k+1}^{\mathrm{d}}} \cdots \sum_{\boldsymbol{x}_0^{\mathrm{d}}} \rho_{w_k^{\mathrm{c}}}\left(-\boldsymbol{\Phi}^{\mathrm{c}}\boldsymbol{h}_{k,a}^{\mathrm{c}} - \boldsymbol{\Phi}^{\mathrm{c}}\boldsymbol{h}_{k,b}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(\boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{h}_{k,a}^{\mathrm{d}}) + \Phi^{\mathrm{cd}}(\boldsymbol{x}_k^{\mathrm{d}} - \boldsymbol{h}_{k,b}^{\mathrm{d}})\right)\rho_{w_{k-1}^{\mathrm{c}}}(\boldsymbol{h}_{k,a}^{\mathrm{c}} + \boldsymbol{h}_{k,b}^{\mathrm{c}}) \\
&\quad \times p(\boldsymbol{x}_{k+1}^{\mathrm{d}}|\boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{h}_{k,a}^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_{k+1}^{\mathrm{d}}|\boldsymbol{x}_k^{\mathrm{d}} - \boldsymbol{h}_{k,b}^{\mathrm{d}})^{1/2} \\
&\quad \times p(\boldsymbol{x}_k^{\mathrm{d}} + \boldsymbol{h}_{k,a}^{\mathrm{d}}|\boldsymbol{x}_{k-1}^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_k^{\mathrm{d}} - \boldsymbol{h}_{k,a}^{\mathrm{d}}|\boldsymbol{x}_{k-1}^{\mathrm{d}})^{1/2} \\
&\quad \times \prod_{\ell=0}^{k-2} p(\boldsymbol{x}_{\ell+1}^{\mathrm{d}}|\boldsymbol{x}_\ell^{\mathrm{d}}) f_{x_0^{\mathrm{d}}}(\boldsymbol{x}_0^{\mathrm{d}}) \qquad (7.8c) \\
D_1^{22} : \mu(\mathfrak{h}_a,\mathfrak{h}_b) &= \ln \rho_{v_1^{\mathrm{c}}}(\boldsymbol{C}\boldsymbol{h}_{1,a} + \boldsymbol{C}\boldsymbol{h}_{1,b}) \sum_{\boldsymbol{x}_1^{\mathrm{d}}}\sum_{\boldsymbol{x}_0^{\mathrm{d}}} \rho_{w_0^{\mathrm{c}}}(\boldsymbol{h}_{1,a}^{\mathrm{c}} + \boldsymbol{h}_{1,b}^{\mathrm{c}}) \\
&\quad \times p(\boldsymbol{x}_1^{\mathrm{d}} + \boldsymbol{h}_{1,a}^{\mathrm{d}}|\boldsymbol{x}_0^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_1^{\mathrm{d}} - \boldsymbol{h}_{1,a}^{\mathrm{d}}|\boldsymbol{x}_0^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_0^{\mathrm{d}}) \qquad (7.8d) \\
D_{k+1}^{22} : \mu(\mathfrak{h}_a,\mathfrak{h}_b) &= \ln \rho_{v_{k+1}^{\mathrm{c}}}(\boldsymbol{C}\boldsymbol{h}_{k+1,a} + \boldsymbol{C}\boldsymbol{h}_{k+1,b}) \sum_{\boldsymbol{x}_{k+1}^{\mathrm{d}}} \cdots \sum_{\boldsymbol{x}_0^{\mathrm{d}}} \rho_{w_k^{\mathrm{c}}}(\boldsymbol{h}_{2,a}^{\mathrm{c}} + \boldsymbol{h}_{2,b}^{\mathrm{c}}) \\
&\quad \times p(\boldsymbol{x}_{k+1}^{\mathrm{d}} + \boldsymbol{h}_{k+1,a}^{\mathrm{d}}|\boldsymbol{x}_k^{\mathrm{d}})^{1/2} p(\boldsymbol{x}_{k+1}^{\mathrm{d}} - \boldsymbol{h}_{k+1,a}^{\mathrm{d}}|\boldsymbol{x}_k^{\mathrm{d}})^{1/2} \\
&\quad \times \prod_{\ell=0}^{k-1} p(\boldsymbol{x}_{\ell+1}^{\mathrm{d}}|\boldsymbol{x}_\ell^{\mathrm{d}}) p(\boldsymbol{x}_0^{\mathrm{d}}) \qquad (7.8e)
\end{aligned}$$

Figure 7.2: The negative Bhattacharyya distance for the elements of matrices D_{k+1}^{ij} (part 1).

Proof. Due to (6.17), many $\boldsymbol{h}$ vectors in (6.24i) are zero and thus we omit them. For any time ℓ, the pairs of densities with $\boldsymbol{h}_{\ell,a} = \boldsymbol{h}_{\ell,b} = \boldsymbol{0}$ and $\boldsymbol{h}_{\ell-1,a} = \boldsymbol{h}_{\ell-1,b} = \boldsymbol{0}$ are separable from the rest. They cancel.

If either $\boldsymbol{h}_{\ell,a}$, $\boldsymbol{h}_{\ell,b}$, $\boldsymbol{h}_{\ell-1,a}$, or $\boldsymbol{h}_{\ell-1,b}$ has elements unequal zero, the product

$$\begin{aligned}
&v(\boldsymbol{x}_\ell + \boldsymbol{h}_{\ell,a}|\boldsymbol{x}_{\ell-1} + \boldsymbol{h}_{\ell-1,a}) v(\boldsymbol{x}_\ell - \boldsymbol{h}_{\ell,b}|\boldsymbol{x}_{\ell-1} - \boldsymbol{h}_{\ell-1,b}) \\
&\qquad = f(\boldsymbol{x}_\ell^{\mathrm{c}} + \boldsymbol{h}_{\ell,a}|\boldsymbol{x}_{\ell-1}^{\mathrm{c}} + \boldsymbol{h}_{\ell-1,a}^{\mathrm{c}}, \boldsymbol{x}_{\ell-1}^{\mathrm{d}} + \boldsymbol{h}_{\ell+1,a}^{\mathrm{d}}) \\
&\qquad \times f(\boldsymbol{x}_\ell^{\mathrm{c}} - \boldsymbol{h}_{\ell,b}|\boldsymbol{x}_{\ell-1}^{\mathrm{c}} - \boldsymbol{h}_{\ell-1,b}^{\mathrm{c}}, \boldsymbol{x}_{\ell-1}^{\mathrm{d}} - \boldsymbol{h}_{\ell-1,b}^{\mathrm{d}}) \\
&\qquad \times p(\boldsymbol{x}_\ell^{\mathrm{d}} + \boldsymbol{h}_{\ell,a}^{\mathrm{d}}|\boldsymbol{x}_{\ell-1}^{\mathrm{d}} + \boldsymbol{h}_{\ell-1,a}^{\mathrm{d}}) p(\boldsymbol{x}_\ell^{\mathrm{d}} - \boldsymbol{h}_{\ell,b}^{\mathrm{d}}|\boldsymbol{x}_{\ell-1}^{\mathrm{d}} - \boldsymbol{h}_{\ell-1,b}^{\mathrm{d}})\,. \qquad (7.10)
\end{aligned}$$

$$
\begin{aligned}
D_2^{10} : \mu(\mathfrak{h}_a, \mathfrak{h}_b) &= \ln \rho_{v_{k+1}^{\mathrm{c}}}(-\boldsymbol{C}\boldsymbol{h}_{1,a}) \sum_{x_2^{\mathrm{d}}} \cdots \sum_{x_0^{\mathrm{d}}} \rho_{w_1^{\mathrm{c}}}(-\boldsymbol{\Phi}^{\mathrm{c}} h_{2,a}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(x_1^{\mathrm{d}} + h_{1,a}^{\mathrm{d}}) + \Phi^{\mathrm{cd}}(x_1^{\mathrm{d}})) \\
&\quad \times \rho_{w_0^{\mathrm{c}}}(h_{1,a}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(x_0^{\mathrm{d}}) - \boldsymbol{\Phi}^{\mathrm{c}} h_{0,b}^{\mathrm{c}} + \Phi^{\mathrm{cd}}(x_0^{\mathrm{d}}) - h_{0,b}^{\mathrm{c}}) \rho_{x_0^{\mathrm{c}}}(h_{0,b}^{\mathrm{c}}) \\
&\quad \times p(x_2^{\mathrm{d}} | x_1^{\mathrm{d}} + h_{1,a}^{\mathrm{d}})^{1/2} p(x_2^{\mathrm{d}} | x_1^{\mathrm{d}})^{1/2} \\
&\quad \times p(x_1^{\mathrm{d}} + h_{1,a}^{\mathrm{d}} | \Phi^{\mathrm{d}}(x_0^{\mathrm{d}}))^{1/2} p(x_1^{\mathrm{d}} | x_0^{\mathrm{d}} - h_{0,b}^{\mathrm{d}})^{1/2} p(x_0^{\mathrm{d}})^{1/2} p(x_0^{\mathrm{d}} - h_{0,b}^{\mathrm{d}})^{1/2} \qquad (7.9a) \\
D_{k+1}^{10} : \mu(\mathfrak{h}_a, \mathfrak{h}_b) &= \ln \rho_{v_{k+1}^{\mathrm{c}}}(-\boldsymbol{C}\boldsymbol{h}_{k+1,a}) \rho_{v_k^{\mathrm{c}}}(\boldsymbol{C}\boldsymbol{h}_{k,b}) \\
&\quad \times \sum_{x_{k+1}^{\mathrm{d}}} \cdots \sum_{x_0^{\mathrm{d}}} \rho_{w_k^{\mathrm{c}}}(-\boldsymbol{\Phi}^{\mathrm{c}} h_{k,a}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(x_k^{\mathrm{d}} + h_{k,a}^{\mathrm{d}}) + \Phi^{\mathrm{cd}}(x_k^{\mathrm{d}})) \\
&\quad \times \rho_{w_{k-1}^{\mathrm{c}}}(h_{k,a}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(x_{k-1}^{\mathrm{d}}) - \boldsymbol{\Phi}^{\mathrm{c}} h_{k-1,b}^{\mathrm{c}} + \Phi^{\mathrm{cd}}(x_{k-1}^{\mathrm{d}}) - \boldsymbol{\Phi}^{\mathrm{c}} h_{k-1,b}^{\mathrm{c}}) \rho_{w_{k-2}^{\mathrm{c}}}(h_{k-1,b}^{\mathrm{c}}) \\
&\quad \times p(x_{k+1}^{\mathrm{d}} | x_k^{\mathrm{d}} + h_{k,a}^{\mathrm{d}})^{1/2} p(x_{k+1}^{\mathrm{d}} | x_k^{\mathrm{d}})^{1/2} \\
&\quad \times p(x_k^{\mathrm{d}} + h_{k,a}^{\mathrm{d}} | x_{k-1}^{\mathrm{d}})^{1/2} p(x_k^{\mathrm{d}} | x_{k-1}^{\mathrm{d}} - h_{k-1,b}^{\mathrm{d}})^{1/2} \\
&\quad \times \prod_{\ell=0}^{k-2} p(x_{\ell+1}^{\mathrm{d}} | x_\ell^{\mathrm{d}}) p(x_0^{\mathrm{d}}) \qquad (7.9b) \\
D_1^{12} : \mu(\mathfrak{h}_a, \mathfrak{h}_b) &= \ln \rho_{v_1^{\mathrm{c}}}(\boldsymbol{C}\boldsymbol{h}_{1,b}) \\
&\quad \times \sum_{x_1^{\mathrm{d}}} \sum_{x_0^{\mathrm{d}}} \rho_{w_1^{\mathrm{c}}}(-\boldsymbol{\Phi}^{\mathrm{c}} h_{0,a}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(x_0^{\mathrm{d}} + h_{0,a}^{\mathrm{d}}) + h_{1,b}^{\mathrm{c}} + \Phi^{\mathrm{cd}}(x_0^{\mathrm{d}})) \rho_{x_0^{\mathrm{c}}}(h_{0,a}^{\mathrm{c}}) \\
&\quad \times p(x_1^{\mathrm{d}} | x_0^{\mathrm{d}} + h_{0,a}^{\mathrm{d}})^{1/2} p(x_1^{\mathrm{d}} - h_{1,b}^{\mathrm{d}} | x_0^{\mathrm{d}})^{1/2} p(x_0^{\mathrm{d}} + h_{0,a}^{\mathrm{d}})^{1/2} p(x_0^{\mathrm{d}})^{1/2} \qquad (7.9c) \\
D_{k+1}^{12} : \mu(\mathfrak{h}_a, \mathfrak{h}_b) &= \ln \rho_{v_k^{\mathrm{c}}}(\boldsymbol{C}\boldsymbol{h}_{k,a}) \rho_{v_{k+1}^{\mathrm{c}}}(\boldsymbol{C}\boldsymbol{h}_{k+1,b}) \\
&\quad \times \sum_{x_{k+1}^{\mathrm{d}}} \cdots \sum_{x_0^{\mathrm{d}}} \rho_{w_k^{\mathrm{c}}}(-\boldsymbol{\Phi}^{\mathrm{c}} h_{k,a}^{\mathrm{c}} - \Phi^{\mathrm{cd}}(x_k^{\mathrm{d}} + h_{k,a}^{\mathrm{d}}) + h_{k+1,b}^{\mathrm{c}} + \Phi^{\mathrm{cd}}(x_k^{\mathrm{d}})) \rho_{x_{k-1}^{\mathrm{c}}}(h_{k,a}^{\mathrm{c}}) \\
&\quad \times p(x_{k+1}^{\mathrm{d}} | x_k^{\mathrm{d}} + h_{k,a}^{\mathrm{d}})^{1/2} p(x_{k+1}^{\mathrm{d}} - h_{k+1,b}^{\mathrm{d}} | x_k^{\mathrm{d}})^{1/2} \times p(x_k^{\mathrm{d}} + h_{k,a}^{\mathrm{d}} | x_{k-1}^{\mathrm{d}})^{1/2} p(x_k^{\mathrm{d}} | x_{k-1}^{\mathrm{d}})^{1/2} \\
&\quad \times \prod_{\ell=0}^{k-1} p(x_{\ell+1}^{\mathrm{d}} | x_\ell^{\mathrm{d}}) p(x_0^{\mathrm{d}}) \qquad (7.9d)
\end{aligned}
$$

Figure 7.3: The negative Bhattacharyya distance for the elements of matrices $\boldsymbol{D}_{k+1}^{ij}$ (part 2).

Inserting the definitions of the Gaussian distribution and the discrete uniform distribution gives the result. ■

We get an analytic result if the sums in Figs. 7.2 and 7.3 are finite. They are finite if and only if the transition probability mass function has finite support.

Corollary 18: Categorical distribution

Given a mixed continuous/discrete state space model with a non-linearity as in (7.1). If the probability density $p(\boldsymbol{x}_k^{\mathrm{d}})$ of the discrete random states is categorical, then an analytic sequential SWW bound exists.

7.4 *Numerical Results*

The SWW bound (7.4) of state-space model (7.1) with additional noise (7.3) from an imperfect channel is analyzed in the following.

Fig. 7.1 shows the setup for our analysis: Two sensors are in an one-dimensional grid of 11 nodes. The grid is partitioned into two similar sub-grids associated with the nearest sensors. On the left side, we assume a transparent boundary. On the right side, a wall is modeled. One source occurs at time $k = 1$. The settings are summarized in Tab. 7.1. Function $Q(\boldsymbol{A})$ replaces every non-zero element in $\boldsymbol{A}$ by 1 and $\boldsymbol{e}_\ell$ is a zero vector except the ℓ^{th} entry which is one.

Table 7.1: Simulation settings (cf. Fig. 7.1)

Densities

$f(\boldsymbol{q}_0) \sim \mathrm{N}\{0, 0.0001\boldsymbol{I}\}$ $\quad$ $p(x_0^{\mathrm{d}}) \sim \mathrm{Unif}\{4, 8\}$

$f(\boldsymbol{p}_0) \sim \mathrm{N}\{0, 0.01\boldsymbol{I}\}$ $\quad$ $f(\boldsymbol{w}_k^{\mathrm{c,p}}) \sim \mathrm{N}\{0, 0.01\boldsymbol{I}\}$

$f(\boldsymbol{w}_k^{\mathrm{c,q}}) \sim \mathrm{N}\{0, 0.25\boldsymbol{I}\}$ $\quad$ $f(\boldsymbol{v}_k) \sim \mathrm{N}\{0, 0.01\boldsymbol{I}\}$

Symmetric triangular density as transition density

$$p(x_k^{\mathrm{d}}|x_{k-1}^{\mathrm{d}}) = \begin{cases} 0.9\,, & x_k^{\mathrm{d}} = x_{k-1}^{\mathrm{d}} \\ 0.05\,, & x_k^{\mathrm{d}} = x_{k-1}^{\mathrm{d}} \pm 1\,, \\ 0.95\,, & x_k^{\mathrm{d}} = x_{k-1}^{\mathrm{d}} \in \{1, 11\}\,, \\ 0.1\,, & x_k^{\mathrm{d}} = 10,\ x_{k-1}^{\mathrm{d}} = 11\,, \\ 0.1\,, & x_k^{\mathrm{d}} = 2,\ x_{k-1}^{\mathrm{d}} = 1\,, \\ 0\,, & \text{else} \end{cases}$$

State-space model

ϕ^{cd}: source function, $\boldsymbol{\Phi}^{\mathrm{c}}$: discretized 1-D field, $\boldsymbol{\Phi}^{\mathrm{d}} = 1$

$\phi^{\mathrm{cd}} : x_k^{\mathrm{d}} \mapsto 0.68\boldsymbol{e}_{6+x_k^{\mathrm{d}}}$ $\quad$ $\boldsymbol{C}$ picks pressure state at 2 & 9

$[\boldsymbol{\Phi}^{\mathrm{c}}]_{1,1} = 0.98$ $\quad$ $[\boldsymbol{\Phi}^{\mathrm{c}}]_{i,i} = 1, i \in [2, 22]$

$[\boldsymbol{\Phi}^{\mathrm{c}}]_{i+11,i} = 5.8 \times 10^{-5}, i \in [1, 11]$

$[\boldsymbol{\Phi}^{\mathrm{c}}]_{i,i+11} = -3.4, i \in [1, 11]$

$[\boldsymbol{\Phi}^{\mathrm{c}}]_{i,i+12} = 1.7, i \in [2, 10]$ $\quad$ $[\boldsymbol{\Phi}^{\mathrm{c}}]_{1,13} = 3.4$

$[\boldsymbol{\Phi}^{\mathrm{c}}]_{i+1,i+11} = 1.7, i \in [1, 10]$

Parameters of the SWW for $k = 1, 2, \cdots$

$\boldsymbol{H}_0 = \boldsymbol{H}_k = \mathrm{blockdiag}\left(0.01Q\left(\boldsymbol{\Phi}^{\mathrm{c}\mathrm{T}}\right), 1\right)$

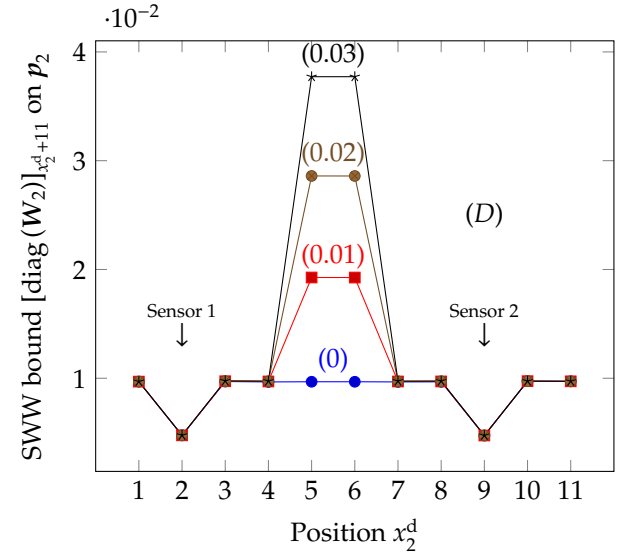
Fig. 7.4a shows the bound on the pressure states for one instant of time. It features low values at sensor positions and an error floor in-between. At the boundary between sensors, additional noise from the channel's distortion has a low influence on the neighborhood. Observe that the bound depends linearly on the additional distortion. The noise diffuses over several time steps.

The bound on the $\boldsymbol{q}$-states is plotted in Fig. 7.4b. Observe that the bound depends non-linearly on the additional distortion.

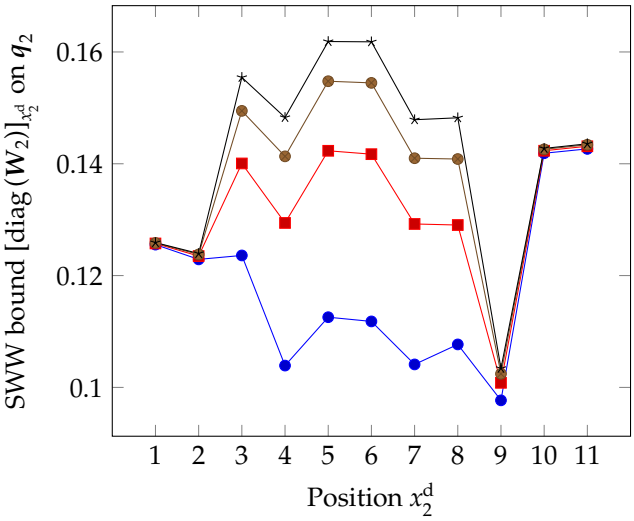
The bound on the localization-error variance is plotted in Fig. 7.4c. According to the equations in Figs. 7.2 and 7.3, the parameters $\boldsymbol{H}_k$ at time k also influence the computation at time $k \pm 1$. The minimum at time $k = 2$ is caused by different support of prior and transition density whereas $\boldsymbol{H}_0 = \boldsymbol{H}_1$.

7.5 *Conclusions of this Chapter*

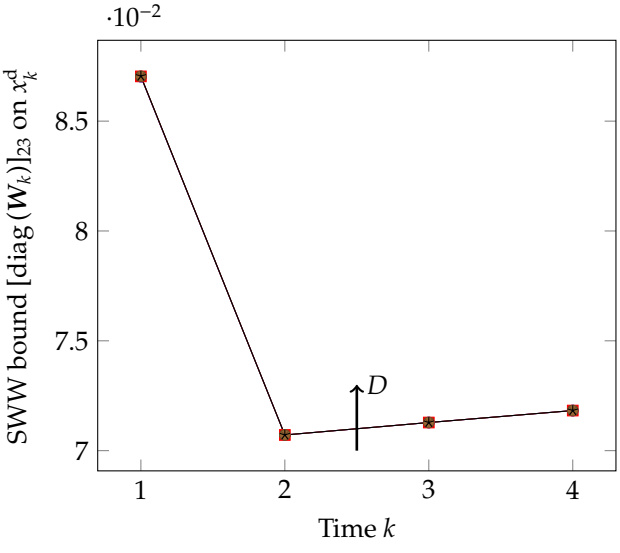
Hybrid continuous/categorical distributions of prior and noise demand for a more general sequential Bayesian bound than the sequential Cramér-Rao (SCR) bound: the SWW bound. Although a non-linearity is inherent in the state space model, there exist analytic solutions of the SWW bound. At a specific time step, the SWW bound of the pressure states stays approximately constant over location. The SWW bound of the source location is insensitive to noise introduced by the imperfect channel between sensors.



(a)



(b)



(c)

Figure 7.4: SWW bound on the error variance of (a) pressure states p_2, (b) q_2 states, and (c) the x_k^{d} state. Due to an imperfect channel, states $q_{2,5}$ and $q_{2,6}$ are added by Gaussian noise Unif $\{0, D\}$.

8
Conclusion

To close my thesis, I present the most important conclusions and give an outlook towards possible further research.

8.1 Conclusions

The discrete model stemming from the wave equation incorporates wave phenomenons (e.g. echoes) into the Bayesian estimation. The decentralized model exploits the sparsity of the system matrices. In fact, the loose coupling between the components of the state vector allows separate and parallel computation of equation sub-systems of much smaller dimension in each cluster heads. Due to the decomposition of the system matrix, decentralization distributes the computational effort over several clusters (sensors) which is very useful for large systems. If the finite difference method is replaced by another numerical method, the sparseness of the system matrix has to be ensured. The non-linear source couples the linear physical model with the linear model of the source's states. The spatio-temporal field dependencies are demonstrated by the light cone (Fig. 3.1). Exploiting these dependencies highly reduces the communication effort between clusters (sensors), transmit power, and quantization errors. My approach in Chapter 5 does a whitening of this spatio-temporal correlated signal up to some small noise.

The sequential Cramér-Rao (SCR) bound cannot be used as performance bound for the decentralized distributed estimation, due to hybrid discrete and continuous distributions of finite support. Sequential Weiss-Weinstein (SWW) bounds support those distributions. The recursion of the SWW bound uses Chernoff or Bhattacharyya coefficients. If there is a linear state-space model with closed-form solutions of the coefficients for particular probability densities, analytic solutions of the SWW bound exists. Table 8.1 summarizes the Bhattacharyya coefficients. For specific test-point matrices $\boldsymbol{H}_k$, the SWW bound of quantized[1] and original continuous distributions are equal. Probability densities (PDs) with finite support induce box constrains on the test-point matrix $\boldsymbol{H}_k$. An analytic solution of the SWW bound for non-linear systems only exists for specific non-linear state-space models [cf. (7.1)]. The computational effort depends on the state-space

[1] Quantized densities are discrete approximations of continuous densities.

model and the statistics:

- Consider the non-linear model (7.1). The computational effort is quadratic in time.
- For linear state-space models with analytic solutions, the computational effort is linear in time.
- Additionally, if the statistics are constant, then the effort is constant.

Consider the pressure-state vector. The SWW bound at one time step is approximately constant for every entry. This means that the bound is constant in space. The impact of noise due to imperfect communication mainly influences the bound of the exchanged pressure states. The SWW bound of the source location is insensitive to noise introduced by the imperfect channel between sensors.

8.2 *Outlook*

Based on my work, I present different interesting issues that are worth to follow.

Decentralized estimation

The finite-difference method (FDM) is a simple and fast numerical method obtaining a discrete model. There are quite a number of others, e.g. finite-element method (FEM). The decentralization approach, which I followed, demands for a very sparse system matrix of the discrete model. This is offered by the spectral-element method (SEM)[2] due to the clever use of basis functions associated with Gauss-Lobatto points and a matched Gaussian numerical integration.

[2] Komatitsch et al. 2005; Tromp et al. 2008.

Numerical approximation of stochastic partial differential equations (SPDEs) depends on the specific choice of the noise process. The theory of SPDEs need to be further developed together with numerical approximations[3]. More specifically, the stochastic wave equation beyond generalized Wiener processes needs attention.

[3] Dalang et al. 1998; Hausenblas 2010; Jentzen et al. 2009; Walsh 2006.

For decentralized distributed particle filter (DDPF), the gain of dimension reduction and the increase of communication are in opposition to each other. The use of marginalized particle filter (MPF)[4] would reduce both the computational effort and the communication load in each sensor (cluster):

[4] Schön et al. 2005.

- The MPF uses Kalman filters (KFs) for the linear and Gaussian part of state-transition model. This demands for fewer particles and, hence, less communication costs.
- Additionally, the accuracy of the estimate improves due to the analytic recursion of the KFs.
- The implementation of the KF is computationally more efficient than of the particle filter (PF).

The communication load could be further reduced if the weights of the DDPF are quantized. This leads to a different importance function used for the particle filtering.

Multivariate distribution	Definition	Bhattacharyya coeff. $\rho(\boldsymbol{h})$ for (6.43)		Test point $\boldsymbol{h} \neq \mathbf{0}$	Comments
Gaussian N $\{\boldsymbol{m}_x, \boldsymbol{C}_x\}$	$f(\boldsymbol{x}) \triangleq \frac{1}{(2\pi)^{N/2}(\det \boldsymbol{C}_x)^{1/2}} e^{-\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{m}_x\|^2_{\boldsymbol{C}_x^{-1}}}$	$\rho_x^{\mathrm{G}}(\boldsymbol{h}) \triangleq e^{-\frac{1}{8}\|\boldsymbol{h}\|^2_{\boldsymbol{C}_x^{-1}}}$	(6.42)	$\boldsymbol{h}$ small	$\boldsymbol{h} \to \mathbf{0}$: Cramér-Rao bound
Cont. uniform Unif $\{\boldsymbol{r}, \boldsymbol{s}\}$	$f(\boldsymbol{x}) \triangleq \prod_{\ell=1}^{N} \frac{1}{[\boldsymbol{s}-\boldsymbol{r}]_\ell} \mathbb{1}_{[\boldsymbol{x}]_\ell \geq [\boldsymbol{r}]_\ell, [\boldsymbol{x}]_\ell \leq [\boldsymbol{s}]_\ell}$	$\rho_x^{\mathrm{U}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \left[1 - \frac{\lvert[\boldsymbol{h}]_\ell\rvert}{[\boldsymbol{s}-\boldsymbol{r}]_\ell}\right]$	(6.53)	$\lvert[\boldsymbol{h}]_\ell\rvert \leq [\boldsymbol{s}-\boldsymbol{r}]_\ell$	$\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}^N, \boldsymbol{r} \leq \boldsymbol{s}$
Discrete uniform Unif $\{\boldsymbol{r}, \boldsymbol{s}\}$	$p(\boldsymbol{x}) \triangleq \prod_{\ell=1}^{N} \frac{1}{[\boldsymbol{s}-\boldsymbol{r}]_\ell} \mathbb{1}_{[\boldsymbol{x}]_\ell \geq [\boldsymbol{r}]_\ell, [\boldsymbol{x}]_\ell \leq [\boldsymbol{s}]_\ell}$	$\rho_x^{\mathrm{U}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \left[1 - \frac{\lvert[\boldsymbol{h}]_\ell\rvert}{[\boldsymbol{s}-\boldsymbol{r}]_\ell}\right]$	(6.53)	$\lvert[\boldsymbol{h}]_\ell\rvert \leq [\boldsymbol{s}-\boldsymbol{r}]_\ell$	$\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{r}, \boldsymbol{s} \in \mathbb{Z}^N, \boldsymbol{r} \leq \boldsymbol{s}$
Exponential Exp $\{\boldsymbol{\alpha}\}$	$f(\boldsymbol{x}) \triangleq \prod_{\ell=1}^{N} \begin{cases} [\boldsymbol{\alpha}]_\ell e^{-[\boldsymbol{\alpha}]_\ell [\boldsymbol{x}]_\ell}, & [\boldsymbol{x}]_\ell \geq 0 \\ 0, & \text{else} \end{cases}$	$\rho_x^{\mathrm{E}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \begin{cases} e^{-[\boldsymbol{\alpha}]_\ell/2[\boldsymbol{h}]_\ell}, & [\boldsymbol{h}]_\ell \geq 0 \\ e^{[\boldsymbol{\alpha}]_\ell/2[\boldsymbol{h}]_\ell}, & [\boldsymbol{h}]_\ell < 0 \end{cases}$	(6.62)		$\boldsymbol{\alpha} \geq \mathbf{0}$
Laplace La $\{\boldsymbol{m}_x, \boldsymbol{b}\}$	$f(\boldsymbol{x}) \triangleq \prod_{\ell=1}^{N} \frac{1}{2[\boldsymbol{b}]_\ell} e^{-\lvert[\boldsymbol{x}-\boldsymbol{m}_x]_\ell\rvert/2[\boldsymbol{b}]_\ell}$	$\rho_x^{\mathrm{L}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \left(1 + \frac{\lvert[\boldsymbol{h}]_\ell\rvert}{2[\boldsymbol{b}]_\ell}\right) e^{-\frac{\lvert[\boldsymbol{h}]_\ell\rvert}{2[\boldsymbol{b}]_\ell}}$,	(6.66)		$\boldsymbol{b} > \mathbf{0}$
Categorical $\{p_{\ell,0}, \cdots, p_{\ell,N'_\ell}\}_{\ell=1}^N$ Bernoulli $\{p_{\ell,0}, p_{\ell,1} = 1 - p_{\ell,0}\}_{\ell=1}^N$	$p(\boldsymbol{x}) \triangleq \prod_{\ell=1}^{N} \left[\sum_{\ell'=0}^{N'_\ell} p_{\ell,\ell'} \mathbb{1}_{[\boldsymbol{x}]_\ell = \ell'}\right]$	$\rho_x^{\mathrm{C}}(\boldsymbol{h}) \triangleq \prod_{\ell=1}^{N} \left[\sum_{\ell'=0}^{N'_\ell} p_{\ell,\ell'}^{1/2} p_{\ell,\ell'+[\boldsymbol{h}]_\ell}^{1/2}\right]$	(6.70) (6.72)	$\lvert[\boldsymbol{h}]_\ell\rvert \in \{0, \cdots, N'_\ell\}$ $\lvert[\boldsymbol{h}]_\ell\rvert \in \{0, 1\}$	masses $p_{\ell,\ell'} = P\{[\boldsymbol{x}]_\ell = \ell'\}$ support $[0, N'_1] \times \cdots \times [0, N'_N]$
Mixed distributions	$v = v([\boldsymbol{x}]_1) \cdots v([\boldsymbol{x}]_N)$	$\rho = \left(c_1 \rho_1^{\mathrm{c}} + c_2 \rho_1^{\mathrm{d}}\right) \cdots \left(c_1 \rho_N^{\mathrm{c}} + c_2 \rho_N^{\mathrm{d}}\right)$			$\begin{cases} v(x) = c_1 f(x) + c_2 p(x) \\ c_1 + c_2 = 1 \end{cases}$

Table 8.1: Summary of Bhattacharyya coefficients the SWW bound. The lower bound (6.20) is computed by (6.38) and (6.43) in Fig. 6.3 with the Bhattacharyya coefficients $\rho(\boldsymbol{h})$ presented in this table. [The underlying Bayesian score is (6.12).]

The boundary conditions of the forward model (wave equation) have been deterministic throughout my thesis. A stochastic approach allows a statement on the robustness of the localization regarding inaccurate boundary conditions.

SWW bounds

Scheduling of decentralized estimators in sensor networks (SNs) might depend on decentralized SWW bounds:

- Mohammadi et al. 2012 presented an approach for the decentralized Cramér-Rao (CR) bound.
- Conditional sequential Bayesian bounds for

$$\mathrm{E}\left\{\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^{\mathrm{T}} \mid \boldsymbol{y}_{1:k-1}\right\} \tag{8.1}$$

  are often more desired than unconditioned bounds due to their dependency on particular realizations.

[5] Ristic et al. 2004; Washburn et al. 1985.

Multiple-switching dynamic models (hybrid models) for linear models depend non-linearly on a state which evolves over time[5]. This is called the regime sequence. The conditional CR bound depends on a specific sequence. The unconditional CR bound is defined as the expectation of the conditional CR bound over the regime sequences. Since the SWW bound supports both discrete and continuous random states, it is a natural application for multiple-switching dynamic models.

The general structure for the linear SWW bound utilizes the Bhattacharyya coefficients of the corresponding PDs. Both, the non-sequential Weiss-Weinstein (WW) bound and the SWW bound would profit by closed-form solutions for different distributions.

[6] Duan et al. 2008a.

Furthermore, the SWW bound could be used to analyze models with quantized measurements[6].

A
Lemmas

In this chapter, I present Lemmas utilized in Chapter 6.

A.1 Gaussian Densities

The following Lemmas are independent of the discrete or continuous nature of the densities. The densities are either Gaussian densities or quantized Gaussian densities $p(\boldsymbol{w}_k^{\mathrm{d}}) = \frac{1}{c''} f_{w_k}(\boldsymbol{w}_k^{\mathrm{d}}\Delta_x)$ or $p(v_k^{\mathrm{d}}) = \frac{1}{c''} f_{v_k}(v_k^{\mathrm{d}}\Delta_x)$. The factor c'' normalizes the probability mass function (PMF).

Lemma 19: Gaussian innovation noise

For a Gaussian innovation noise, the solution of (6.33a) is

$$E_{k+1} = \rho_{\boldsymbol{w}_k}^{\mathrm{G}}(\boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,b}) \tag{A.1}$$

which is independent of $\boldsymbol{x}_k$ (cf. Lemma 3).

Proof. Let us insert the Gaussian density into (6.33a), i.e.

$$\mathrm{E}_{\boldsymbol{x}_{k+1}|\boldsymbol{x}_k}\left\{\frac{e^{-\frac{1}{4}\left\|\boldsymbol{x}_{k+1}+\boldsymbol{h}_{k+1,a}-\boldsymbol{\Phi}(\boldsymbol{x}_k+\boldsymbol{h}_{k,a})\right\|^2_{C_{w_k}^{-1}}}}{e^{-\frac{1}{2}\|\boldsymbol{x}_{k+1}-\boldsymbol{\Phi}\boldsymbol{x}_k\|^2_{C_{w_k}^{-1}}}} \times e^{-\frac{1}{4}\left\|\boldsymbol{x}_{k+1}-\boldsymbol{h}_{k+1,b}-\boldsymbol{\Phi}(\boldsymbol{x}_k-\boldsymbol{h}_{k,b})\right\|^2_{C_{w_k}^{-1}}}\right\}. \tag{A.2}$$

This simplifies to

$$\begin{aligned} c'_w \int_{-\infty}^{\infty} & e^{-1/2\|\boldsymbol{x}_{k+1}-\boldsymbol{\Phi}\boldsymbol{x}_k\|_{C_{w_k}^{-1}}} \\ & \times e^{-1/4\left\|\boldsymbol{h}_{k+1,a}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a}\right\|_{C_{w_k}} - 1/4\left\|-\boldsymbol{h}_{k+1,b}+\boldsymbol{\Phi}\boldsymbol{h}_{k,b}\right\|_{C_{w_k}^{-1}}} \\ & \times e^{-1/2\left\langle \boldsymbol{x}_{k+1}-\boldsymbol{\Phi}\boldsymbol{x}_k, \boldsymbol{h}_{k+1,a}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a}-\boldsymbol{h}_{k+1,b}+\boldsymbol{\Phi}\boldsymbol{h}_{k,b}\right\rangle_{C_{w_k}^{-1}}} \mathrm{d}P_{\boldsymbol{x}_{k+1}} \end{aligned} \tag{A.3}$$

where

$$c'_w \triangleq \begin{cases} (2\pi)^{-N/2} \det\left(\boldsymbol{C}_{w_k}\right)^{-1/2}, & \text{continuous}, \\ (2\pi)^{-N/2} \det\left(\boldsymbol{C}_{w_k}\right)^{-1/2} c'', & \text{quantized}. \end{cases} \tag{A.4}$$

We substitute

$$\begin{aligned}\|\boldsymbol{t}\|^2_{\boldsymbol{C}_{w_k}} &:= \|\boldsymbol{x}_{k+1} - \boldsymbol{\Phi}\boldsymbol{x}_k\|^2_{\boldsymbol{C}^{-1}_{w_k}} \\ &\quad + \left\langle \boldsymbol{x}_{k+1} - \boldsymbol{\Phi}\boldsymbol{x}_k, \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} - \boldsymbol{h}_{k+1,b} + \boldsymbol{\Phi}\boldsymbol{h}_{k,b} \right\rangle_{\boldsymbol{C}^{-1}_{w_k}} \\ &\quad + \frac{1}{4} \left\| \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} - \boldsymbol{h}_{k+1,b} + \boldsymbol{\Phi}\boldsymbol{h}_{k,b} \right\|^2_{\boldsymbol{C}^{-1}_{w_k}}\end{aligned}$$

and utilize

$$c'_w \int_{-\infty}^{\infty} e^{-1/2\|\boldsymbol{t}\|^2_{\boldsymbol{C}^{-1}_{w_k}}} \mathrm{d}P_{\boldsymbol{t}} = 1 \tag{A.5}$$

to obtain the final result. ■

Lemma 20: Gaussian prior

For a Gaussian prior, the solution of (6.33a) is

$$E_1 = \rho^{\mathrm{G}}_{\boldsymbol{x}_0}(\boldsymbol{h}_{0,a} + \boldsymbol{h}_{0,b}) \,. \tag{A.6}$$

Proof. The results follows from Lemma 19 where $v(\boldsymbol{x}_0|\boldsymbol{x}_{-1}) = v_{\boldsymbol{x}_0}(\boldsymbol{x}_0)$ and $\boldsymbol{h}_{-1,a} + \boldsymbol{h}_{-1,b} = 0$. ■

Lemma 21: Gaussian measurement noise

For a Gaussian measurement noise the solution of (6.33b) is

$$E'_{k+1} = \rho^{\mathrm{G}}_{v_{k+1}}(\boldsymbol{C}\boldsymbol{h}_{k+1,a} + \boldsymbol{C}\boldsymbol{h}_{k+1,b}) \tag{A.7}$$

which is independently of $\boldsymbol{x}_{k+1}$ (cf. Corollary 4).

Proof. Let us insert the Gaussian density into (6.33b), i.e.

$$\begin{aligned}\mathrm{E}_{\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}} \Bigg\{ & \frac{e^{-\frac{1}{4}\left\|\boldsymbol{y}_{k+1} - \boldsymbol{C}(\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a})\right\|^2_{\boldsymbol{C}^{-1}_{v_{k+1}}}}}{e^{-\frac{1}{2}\left\|\boldsymbol{y}_{k+1} - \boldsymbol{C}\boldsymbol{x}_{k+1}\right\|_{\boldsymbol{C}^{-1}_{v_{k+1}}}}} \\ & \times e^{-\frac{1}{4}\left\|\boldsymbol{y}_{k+1} - \boldsymbol{C}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b})\right\|^2_{\boldsymbol{C}^{-1}_{v_{k+1}}}} \Bigg\} .\end{aligned} \tag{A.8}$$

This simplifies to

$$\begin{aligned}c'_v \int_{-\infty}^{\infty} & e^{-1/2\left\|\boldsymbol{y}_{k+1} - \boldsymbol{C}\boldsymbol{x}_{k+1}\right\|_{\boldsymbol{C}^{-1}_{v_{k+1}}} - 1/4\left\|\boldsymbol{C}\boldsymbol{h}_{k,a}\right\|_{\boldsymbol{C}^{-1}_{v_{k+1}}} - 1/4\left\|\boldsymbol{C}\boldsymbol{h}_{k,b}\right\|_{\boldsymbol{C}^{-1}_{v_{k+1}}}} \\ & \times e^{-1/2\left\langle \boldsymbol{y}_{k+1} - \boldsymbol{C}\boldsymbol{x}_{k+1}, \boldsymbol{C}\boldsymbol{h}_{k,a} - \boldsymbol{C}\boldsymbol{h}_{k,b} \right\rangle_{\boldsymbol{C}^{-1}_{v_{k+1}}}} \mathrm{d}P_{\boldsymbol{y}_{k+1}}\end{aligned} \tag{A.9}$$

where

$$c'_v \triangleq \begin{cases} (2\pi)^{-N/2} \det\left(\boldsymbol{C}_{v_{k+1}}\right)^{-1/2} , & \text{continuous} \,, \\ (2\pi)^{-N/2} \det\left(\boldsymbol{C}_{v_{k+1}}\right)^{-1/2} c'' , & \text{quantized} \,. \end{cases} \tag{A.10}$$

We substitute

$$\begin{aligned}\|\boldsymbol{t}\|^2_{C_{v_{k+1}}} := &\left\|\boldsymbol{y}_{k+1} - \boldsymbol{C}\boldsymbol{x}_{k+1}\right\|^2_{C^{-1}_{v_{k+1}}} \\ &- \left\langle \boldsymbol{y}_{k+1} - \boldsymbol{C}\boldsymbol{x}_{k+1}, \boldsymbol{C}\boldsymbol{h}_{k+1,a} - \boldsymbol{C}\boldsymbol{h}_{k+1,b} \right\rangle_{C^{-1}_{v_{k+1}}} \\ &+ \frac{1}{4}\left\|\boldsymbol{C}\boldsymbol{h}_{k+1,a} - \boldsymbol{C}\boldsymbol{h}_{k+1,b}\right\|^2_{C^{-1}_{v_{k+1}}}\end{aligned}$$

and utilize

$$c'_v \int_{-\infty}^{\infty} e^{-1/2\|\boldsymbol{t}\|^2_{C^{-1}_{v_{k+1}}}} \mathrm{d}P_{\boldsymbol{t}} = 1 \tag{A.11}$$

to obtain the final result. ■

A.2 Uniform Densities

The following Lemmas are independent of the discrete or continuous nature of the densities. The densities are either continuous or discrete uniform densities.

Lemma 22: Uniform innovation noise

For an independent uniform density $v(\boldsymbol{w}_k)$, the solution of (6.33a) is

$$E_{k+1} = \rho^{\mathrm{U}}_{\boldsymbol{w}_k}(\boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,b}) \tag{A.12}$$

which is independent of $\boldsymbol{x}_k$.

Proof. Let us insert the uniform density into (6.33a), i.e.

$$\begin{aligned}\mathrm{E}_{\boldsymbol{x}_{k+1}|\boldsymbol{x}_k}\Big\{&\frac{v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}(\boldsymbol{x}_k + \boldsymbol{h}_{k,a}))^{1/2}}{v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} - \boldsymbol{\Phi}\boldsymbol{x}_k)} \\ &\times v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b}))^{1/2}\Big\}\end{aligned} \tag{A.13}$$

Due to the existence of a density, we have

$$\begin{aligned}&\int_{-\infty}^{\infty} v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}(\boldsymbol{x}_k + \boldsymbol{h}_{k,a}))^{1/2} \\ &\times v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b}))^{1/2} \mathrm{d}P_{\boldsymbol{x}_{k+1}} \\ &= \int_{\boldsymbol{r}_k}^{\boldsymbol{s}_k} \prod_{\ell=1}^{N} \frac{\mathbb{1}_{\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}(\boldsymbol{x}_k + \boldsymbol{h}_{k,a}) \in [\boldsymbol{r}_k, \boldsymbol{s}_k]}}{[\boldsymbol{\varsigma}_{\boldsymbol{w}_k}]_\ell} \\ &\times \mathbb{1}_{\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b}) \in [\boldsymbol{r}, \boldsymbol{s}]} \mathrm{d}P_{\boldsymbol{x}_k} \\ &= \prod_{\ell=1}^{N} \left[1 - \frac{|[\boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,b}]_\ell|}{[\boldsymbol{\varsigma}_{\boldsymbol{w}_k}]_\ell}\right]\end{aligned} \tag{A.14}$$

if

$$|[\boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,b}]_\ell| \leq [\boldsymbol{\varsigma}_{\boldsymbol{w}_k}]_\ell \,. \tag{A.15}$$

else zero. ■

Lemma 23: Uniform prior

For an independent uniform density $v(\boldsymbol{w}_0)$, the solution of (6.33a) is

$$E_1 = \rho^{\mathrm{U}}_{\boldsymbol{w}_k}(\boldsymbol{h}_{0,a} + \boldsymbol{h}_{0,b}) \,. \tag{A.16}$$

Proof. The results follows from Lemma 22 where $v(\boldsymbol{x}_0|\boldsymbol{x}_{-1}) = v_{\boldsymbol{x}_0}(\boldsymbol{x}_0)$ and $\boldsymbol{h}_{-1,a} + \boldsymbol{h}_{-1,b} = 0$. ∎

Lemma 24: Uniform measurement noise

For an independent uniform density $v(\boldsymbol{v}_k)$, the solution of (6.33b) is

$$E'_{k+1} = \rho^{\mathrm{U}}_{\boldsymbol{v}_k}(-\boldsymbol{C}\boldsymbol{h}_{k+1,a} - \boldsymbol{C}\boldsymbol{h}_{k+1,b}) \tag{A.17}$$

which is independent of $\boldsymbol{x}_{k+1}$.

Proof. We insert the uniform density into (6.33b) , i.e.

$$\begin{aligned} \mathrm{E}_{\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}}\Big\{ & \frac{v_{\boldsymbol{v}_k}(\boldsymbol{y}_{k+1} - \boldsymbol{C}(\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a}))^{1/2}}{v_{\boldsymbol{v}_k}(\boldsymbol{y}_{k+1} - \boldsymbol{C}\boldsymbol{x}_{k+1})} \\ & \times v_{\boldsymbol{v}_k}(\boldsymbol{y}_{k+1} - \boldsymbol{C}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b}))^{1/2}\Big\} \,, \end{aligned} \tag{A.18}$$

Due to the existence of a density, we have

$$\begin{aligned} & \int_{-\infty}^{\infty} v_{\boldsymbol{v}_{k+1}}(\boldsymbol{y}_{k+1} - \boldsymbol{C}(\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a}))^{1/2} \\ & \times v_{\boldsymbol{v}_{k+1}}(\boldsymbol{y}_{k+1} - \boldsymbol{C}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b}))^{1/2} \mathrm{d}P_{\boldsymbol{y}_{k+1}} \\ & = \int_{\boldsymbol{r}_{k+1}}^{\boldsymbol{s}_{k+1}} \prod_{\ell=1}^{N} \frac{\mathbb{1}_{\boldsymbol{y}_{k+1} - \boldsymbol{C}(\boldsymbol{x}_k + \boldsymbol{h}_{k+1,a}) \in [\boldsymbol{r}_{k+1}, \boldsymbol{s}_{k+1}]}}{[\varsigma_{\boldsymbol{v}_k}]_\ell} \\ & \times \mathbb{1}_{\boldsymbol{y}_{k+1} - \boldsymbol{C}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b}) \in [\boldsymbol{r}_{k+1}, \boldsymbol{s}_{k+1}]} \mathrm{d}P_{\boldsymbol{y}_{k+1}} \\ & = \prod_{\ell=1}^{N} \left[1 - \frac{|[\boldsymbol{C}\boldsymbol{h}_{k+1,a} + \boldsymbol{C}\boldsymbol{h}_{k+1,b}]_\ell|}{[\varsigma_{\boldsymbol{v}_k}]_\ell} \right] \end{aligned} \tag{A.19}$$

if

$$|[\boldsymbol{C}\boldsymbol{h}_{k+1,a} + \boldsymbol{C}\boldsymbol{h}_{k+1,b}]_\ell| \leq [\varsigma_{\boldsymbol{v}_k}]_\ell \tag{A.20}$$

else zero. ∎

A.3 Exponential Densities

We assume both exponential densities and quantized exponential densities $p(\boldsymbol{w}^{\mathrm{d}}_k) = \frac{1}{c''} f_{\boldsymbol{w}_k}(\boldsymbol{w}^{\mathrm{d}}_k \Delta_{\boldsymbol{x}})$ and $p(\boldsymbol{v}^{\mathrm{d}}_k) = \frac{1}{c''} f_{\boldsymbol{v}_k}(\boldsymbol{v}^{\mathrm{d}}_k \Delta_{\boldsymbol{x}})$. The factor $c'' = c''_1 \cdots c''_N$ normalizes the PMF.

Lemma 25: Innovation noise

Given a multivariate independent exponential density $v(\boldsymbol{w}_k)$, the solution of (6.33a) is

$$E_{k+1} = \rho^{\mathrm{E}}_{\boldsymbol{w}_k}(\boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,b}) \tag{A.21}$$

which is independent of $\boldsymbol{x}_k$.

Proof. Let us insert the density into (6.33a) and substitute $\boldsymbol{w} := \boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b})$, i.e.

$$\begin{aligned}
&\int_{-\infty}^{\infty} v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}(\boldsymbol{x}_k + \boldsymbol{h}_{k,a}))^{1/2} \\
&\times v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b}))^{1/2} \mathrm{d}P_{\boldsymbol{x}_{k+1}} \\
&= \prod_{\ell=1}^{N} \int_{-\infty}^{\infty} c'_\ell e^{-\alpha_\ell/2(\boldsymbol{w}+\boldsymbol{h}_{k+1,a}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a}+\boldsymbol{h}_{k+1,b}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a})-\alpha_\ell/2(\boldsymbol{w})} \\
&\times \mathbb{1}_{\boldsymbol{w}+\boldsymbol{h}_{k+1,a}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a}+\boldsymbol{h}_{k+1,b}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a}\in[0,\infty)} \mathbb{1}_{\boldsymbol{w}_k\in[0,\infty)} \mathrm{d}P_{\boldsymbol{w}}
\end{aligned} \tag{A.22}$$

where

$$c'_\ell = \begin{cases} \alpha_\ell\,, & v(\boldsymbol{w}_k) \text{ cont.}\,, \\ \alpha_\ell c''_\ell\,, & v(\boldsymbol{w}_k) \text{ discr.}\,, \end{cases} \tag{A.23}$$

normalizes the densities. With

$$c'_\ell \int_0^{\infty} e^{-\alpha_\ell [w]_l} \mathrm{d}P_{[w]_l} = 1 \tag{A.24}$$

we get

$$\prod_{\ell=1}^{N} e^{-\alpha_\ell/2(\boldsymbol{h}_{k+1,a}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a}+\boldsymbol{h}_{k+1,b}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a})} \tag{A.25}$$

if

$$[\boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,b}]_\ell \geq 0 \tag{A.26}$$

else

$$\prod_{\ell=1}^{N} e^{\alpha_\ell/2(\boldsymbol{h}_{k+1,a}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a}+\boldsymbol{h}_{k+1,b}-\boldsymbol{\Phi}\boldsymbol{h}_{k,a})}\,. \tag{A.27}$$

■

Lemma 26: Measurement noise

Given a multivariate independent exponential measurement noise $v(\boldsymbol{v}_k)$, the solution of (6.33b) is

$$E'_{k+1} = \rho^{\mathrm{E}}_{v_k}(-\boldsymbol{C}\boldsymbol{h}_{k+1,a} - \boldsymbol{C}\boldsymbol{h}_{k+1,b}) \tag{A.28}$$

which is independent of $\boldsymbol{x}_{k+1}$.

Proof. The proof is similar to that of Lemma 25 except $\boldsymbol{h}_{k+1,a} = \boldsymbol{h}_{k+1,b} := \boldsymbol{0}$ and the substitution of $\boldsymbol{\Phi}\boldsymbol{h}_{k,a}$ by $\boldsymbol{C}\boldsymbol{h}_{k+1,a}$, $\boldsymbol{\Phi}\boldsymbol{h}_{k,b}$ by $\boldsymbol{C}\boldsymbol{h}_{k+1,b}$ and $\boldsymbol{w}_k$ by $\boldsymbol{v}_{k+1}$. ■

A.4 Laplace Densities

Lemma 27: Innovation noise

Given multivariate independent exponential density $v(\boldsymbol{w}_k)$, the solution of (6.33a) is

$$E_{k+1} = \rho^{\mathrm{L}}_{w_k}(\boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,b}) \tag{A.29}$$

which is independent of $\boldsymbol{x}_k$.

Proof. Let us insert the density into (6.33a), substitute $\boldsymbol{w} := \boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b})$, and substitute $\boldsymbol{h} = \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a}$, i.e.

$$\begin{aligned}
&\int_{-\infty}^{\infty} v_{w_k}(\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}(\boldsymbol{x}_k + \boldsymbol{h}_{k,a}))^{1/2} \\
&\times v_{w_k}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b}))^{1/2} \mathrm{d}P_{\boldsymbol{x}_{k+1}} \\
&= \prod_{\ell=1}^{N} \int_{-\infty}^{\infty} \frac{1}{2[\boldsymbol{b}_k]_\ell} e^{-|[\boldsymbol{w}]_\ell|/2b - |[\boldsymbol{w}+\boldsymbol{h}]_\ell|/2b} \mathrm{d}\boldsymbol{w}
\end{aligned} \tag{A.30}$$

where normalizes the densities. If $\boldsymbol{h} \succcurlyeq \boldsymbol{0}$ we get

$$\prod_{\ell=1}^{N} \begin{Bmatrix} \frac{1}{2[\boldsymbol{b}_k]_\ell} e^{-\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}}, & [\boldsymbol{w}]_\ell \geq 0, [\boldsymbol{w}+\boldsymbol{h}]_\ell \geq 0 \\ \frac{1}{2[\boldsymbol{b}_k]_\ell} e^{-\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}}, & [\boldsymbol{w}]_\ell < 0, [\boldsymbol{w}+\boldsymbol{h}]_\ell < 0 \\ \frac{\boldsymbol{h}_\ell}{2[\boldsymbol{b}_k]_\ell} e^{-\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}}, & [\boldsymbol{w}]_\ell < 0, [\boldsymbol{w}+\boldsymbol{h}]_\ell \geq 0 \end{Bmatrix} = \prod_{\ell=1}^{N} \left(1 + \frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}\right) e^{-\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}} \tag{A.31}$$

If $\boldsymbol{h} \prec \boldsymbol{0}$ we get

$$\prod_{\ell=1}^{N} \begin{Bmatrix} \frac{1}{2[\boldsymbol{b}_k]_\ell} e^{\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}}, & [\boldsymbol{w}]_\ell \geq 0, [\boldsymbol{w}+\boldsymbol{h}]_\ell \geq 0 \\ \frac{1}{2[\boldsymbol{b}_k]_\ell} e^{\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}}, & [\boldsymbol{w}]_\ell < 0, [\boldsymbol{w}+\boldsymbol{h}]_\ell < 0 \\ -\frac{\boldsymbol{h}_\ell}{2[\boldsymbol{b}_k]_\ell} e^{\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}}, & [\boldsymbol{w}]_\ell \geq 0, [\boldsymbol{w}+\boldsymbol{h}]_\ell < 0 \end{Bmatrix} = \prod_{\ell=1}^{N} \left(1 - \frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}\right) e^{\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}} \tag{A.32}$$

For arbitrary $\boldsymbol{h} \neq \mathbf{0}$ we get

$$\rho_{\boldsymbol{w}}^{\mathrm{L}}(\boldsymbol{h}) = \prod_{\ell=1}^{N} \begin{cases} \left(1 + \frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}\right) e^{-\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}}, & [\boldsymbol{h}]_\ell > 0\,, \\ \left(1 - \frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}\right) e^{\frac{[\boldsymbol{h}]_\ell}{2[\boldsymbol{b}_k]_\ell}}, & [\boldsymbol{h}]_\ell \leq 0\,. \end{cases} \tag{A.33}$$

■

Lemma 28: Measurement noise

Given multivariate independent exponential measurement noise $v(\boldsymbol{v}_k)$, the solution of (6.33b) is

$$E'_{k+1} = \rho_{\boldsymbol{v}_k}^{\mathrm{L}}(-\boldsymbol{C}\boldsymbol{h}_{k+1,a} - \boldsymbol{C}\boldsymbol{h}_{k+1,b}) \tag{A.34}$$

which is independent of $\boldsymbol{x}_{k+1}$.

Proof. The proof is similar to that of Lemma 27 except $\boldsymbol{h}_{k+1,a} = \boldsymbol{h}_{k+1,b} := \mathbf{0}$ and the substitution of $\boldsymbol{\Phi}\boldsymbol{h}_{k,a}$ by $\boldsymbol{C}\boldsymbol{h}_{k+1,a}$, $\boldsymbol{\Phi}\boldsymbol{h}_{k,b}$ by $\boldsymbol{C}\boldsymbol{h}_{k+1,b}$, and $\boldsymbol{w}_k$ by $\boldsymbol{v}_{k+1}$.

■

A.5 *Categorical Densities*

Lemma 29: Innovation noise

Given a multivariate independent categorical density $v(\boldsymbol{w}_k)$, the solution of (6.33a) is

$$E_{k+1} = \rho_{\boldsymbol{w}_k}^{\mathrm{C}}(\boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,b}) \tag{A.35}$$

which is independent of $\boldsymbol{x}_k$.

Proof. Let us insert the categorical density into (6.33a), substitute $\boldsymbol{w} := \boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b})$, and substitute $\boldsymbol{h} = \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a} + \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}\boldsymbol{h}_{k,a}$, i.e.

$$\begin{aligned} &\int_{-\infty}^{\infty} v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} + \boldsymbol{h}_{k+1,a} - \boldsymbol{\Phi}(\boldsymbol{x}_k + \boldsymbol{h}_{k,a}))^{1/2} \\ &\times v_{\boldsymbol{w}_k}(\boldsymbol{x}_{k+1} - \boldsymbol{h}_{k+1,b} - \boldsymbol{\Phi}(\boldsymbol{x}_k - \boldsymbol{h}_{k,b}))^{1/2} \mathrm{d}P_{\boldsymbol{x}_{k+1}} \\ &= \prod_{\ell=1}^{N} \left[\sum_{\ell'=0}^{N'_\ell} p_{\ell,\ell'+[\boldsymbol{h}]_\ell}^{1/2} p_{\ell,\ell'}^{1/2} \mathbb{1}_{[\boldsymbol{x}]_\ell = \ell'} \right] \qquad \text{(A.36)} \\ &= \rho_{\boldsymbol{w}}^{\mathrm{C}}(\boldsymbol{h})\,. \qquad \text{(A.37)} \end{aligned}$$

Notice that $|[\boldsymbol{h}]_\ell| \in \{0, \cdots, N'_\ell\}$. ■

Lemma 30: Measurement noise

Given a multivariate independent exponential measurement den-

sity $v(\boldsymbol{v}_k)$, the solution of (6.33b) is

$$E'_{k+1} = \rho^{\mathrm{C}}_{\boldsymbol{v}_k}(-\boldsymbol{C}\boldsymbol{h}_{k+1,a} - \boldsymbol{C}\boldsymbol{h}_{k+1,b}) \tag{A.38}$$

which is independent of $\boldsymbol{x}_{k+1}$.

Proof. The proof is similar to that of Lemma 29 except $\boldsymbol{h}_{k+1,a} = \boldsymbol{h}_{k+1,b} := \boldsymbol{0}$ and the substitution of $\boldsymbol{\Phi}\boldsymbol{h}_{k,a}$ by $\boldsymbol{C}\boldsymbol{h}_{k+1,a}$, $\boldsymbol{\Phi}\boldsymbol{h}_{k,b}$ by $\boldsymbol{C}\boldsymbol{h}_{k+1,b}$, and $\boldsymbol{w}_k$ by $\boldsymbol{v}_{k+1}$. ■

B
The Weiss-Weinstein Bound in the Limit

Although the following facts are known I have not been able to find any publication of following proofs. Thus, I prove that the Weiss-Weinstein (WW) and Bobrovsky-Zakai (BZ) bounds approach the Cramér-Rao (CR) bound if $h \to 0$ and that the WW approaches the BZ bound if $s \to 0$, or $s \to 1$.

The error variance of the unbiased estimator $g(y) = \hat{x}$ is bounded according to Weiss et al. 1988 by

$$\mathrm{E}\left\{(\hat{x}-x)^2\right\} \geq \frac{h^2\,\mathrm{E}\left\{L(x+h,x,y)^s\right\}^2}{\mathrm{E}\left\{\left[L(x+h,x,y)^s - L(x-h,x,y)^{1-s}\right]^2\right\}} \tag{B.1}$$

with the likelihood ratio

$$L(x_1,x_2,y) \triangleq \frac{v(x_1,y)}{v(x_2,y)}\,. \tag{B.2}$$

Here the expectation is with respect to x and y.

Theorem 31: Limit case I of the scalar WW bound

For $s \to 0$ and $s \to 1$ the Weiss-Weinstein lower bound reduces to the BZ lower bound.

Proof. If $s \to 0$, the right side of the WW bound follows to

$$\frac{h^2}{\mathrm{E}\left\{\left[1 - L(x-h,x,y)\right]^2\right\}} \tag{B.3}$$

while for $s \to 1$

$$\frac{h^2}{\mathrm{E}\left\{\left[L(x+h,x,y) - 1\right]^2\right\}}\,. \tag{B.4}$$

In both cases, the argument of the expectation equals

$$\left[\frac{1}{v(x,y)}\overset{1}{\underset{h}{\Delta}}_x v(x,y)\right]^2 \tag{B.5}$$

which indeed is the BZ lower bound. ∎

The same arguments yield for the limits of the multivariate WW bound.

Rapoport et al. 2004b, (Theorem 3) proved the relation between sequential Weiss-Weinstein (SWW) and sequential Cramér-Rao (SCR). Although this includes the non-sequential version as a special case, the following proof is different and shorter.

Theorem 32: Limit case II of the scalar WW bound

Let exist a joint probability density function $f(x, y)$ of measurement y and parameter x with following properties:

1. $\lim_{x\to\pm\infty} xf(x|y) = 0$ for all x and y
2. $\partial_x \ln f(x, y)$ exists

Then for $h \to 0$ the Weiss-Weinstein lower bound reduces to the Cramér-Rao lower bound.

Proof. In the limit $h \to 0$, the expectation

$$\lim_{h\to 0} \mathrm{E}\left\{L(x+h, x, y)^s\right\} = 1\,. \tag{B.6}$$

Rewriting

$$\frac{1}{h}\left[L(x+h, x, y)^s - L(x-h, x, y)^{1-s}\right] \tag{B.7}$$

gives

$$\frac{1}{hv(x, y)}\left[v(x+h, y)^s v(x, y)^{1-s} - v(x-h, y)^{1-s} v(x, y)^s\right]\,. \tag{B.8}$$

Furthermore, this is equivalent to

$$v(x, y)^{1-s} \overset{1}{\underset{h}{\Delta}}_x v(x, y)^s + v(x, y)^s \overset{1}{\underset{h}{\Delta}}_x v(x, y)^{1-s}\,. \tag{B.9}$$

With the chain rule and $h \to 0$, above term becomes the score of the Cramér-Rao bound $\partial_x \ln f(x, y)$ which is independent of the parameter s. The result becomes

$$\mathrm{E}\left\{(\hat{x} - x)^2\right\} \geq \frac{1}{\mathrm{E}\left\{\partial_x \ln f(x, y) \partial_x \ln f(x, y)\right\}} \tag{B.10}$$

■

Theorem 33: Limit II of the multivariate WW bound

The multivariate WW bound is defined by

$$\mathrm{E}_{x,y}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right\} \succcurlyeq \boldsymbol{T}\boldsymbol{J}^{-1}\boldsymbol{T}^{\mathrm{T}}\,. \tag{B.11}$$

with

$$[\boldsymbol{J}]_{u,v} = \frac{\mathrm{E}\left\{\left[L(\boldsymbol{x}+\boldsymbol{h}_u,\boldsymbol{x},\boldsymbol{y})^{s_u} - L(\boldsymbol{x}-\boldsymbol{h}_u,\boldsymbol{x},\boldsymbol{y})^{1-s_u}\right]\right.}{\mathrm{E}\left\{L(\boldsymbol{x}+\boldsymbol{h}_u,\boldsymbol{x},\boldsymbol{y})^{s_u}\right\}} \frac{\left.\left[L(\boldsymbol{x}+\boldsymbol{h}_v,\boldsymbol{x},\boldsymbol{y})^{s_v} - L(\boldsymbol{x}-\boldsymbol{h}_v,\boldsymbol{x},\boldsymbol{y})^{1-s_v}\right]\right\}}{\mathrm{E}\left\{L(\boldsymbol{x}+\boldsymbol{h}_v,\boldsymbol{x},\boldsymbol{y})^{s_v}\right\}} . \quad \text{(B.12)}$$

and

$$\boldsymbol{T} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots] . \quad \text{(B.13)}$$

In the case of $\boldsymbol{T} = h\boldsymbol{I}$, letting $h \to 0$, the multivariate CR bound is obtained.

Proof. With the diagonal matrix $\boldsymbol{T}$, the vectors $\boldsymbol{h}_\ell = h\boldsymbol{e}_\ell$. This causes

$$\mathrm{E}_{\boldsymbol{x},\boldsymbol{y}}\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\right\} = h^2\boldsymbol{I}\boldsymbol{J}^{-1} . \quad \text{(B.14)}$$

The arguments are the same for every $[\boldsymbol{J}]_{u,v}$ as in Lemma 32. Furthermore, the partial difference quotients are with respect to $[\boldsymbol{x}]_u$ and $[\boldsymbol{x}]_v$, respectively. The result gives the information matrix of the CR bound, i.e.

$$[\boldsymbol{J}]_{u,v} = \mathrm{E}_{\boldsymbol{x},\boldsymbol{y}}\left\{\partial_{[\boldsymbol{x}]_u} \ln f(\boldsymbol{x},\boldsymbol{y}) \partial_{[\boldsymbol{x}]_v} \ln f(\boldsymbol{x},\boldsymbol{y})\right\} \quad \text{(B.15)}$$

The constrains of the CR bound have to be satisfied.

■

C
Manifolds of Power Spectral Densities

I START with a problem statement which leads to a manifold of power spectral densities (PSDs) and the geodesic distance between two PSDs.[1]

[1] Georgiou 2007.

Consider a time-discrete correlated second-order stationary random process (signal) u_k with assumed PSD $S(\theta)$ and a predictor

$$\hat{u}_k = \sum_{n=1}^{\infty} a_{n,f} u_{k-n} \ .$$

The aim is to generate a white signal (innovation) $u_k - \hat{u}_k$ with flat PSD, i.e. whitening of u_k.

Two questions arise:

- How shall the mismatch between assumed and true PSD be specified?
- How shall the whiteness of a signal u_k be defined?

In the literature there are three different possibilities:

- The *flatness* of a PSD does not allow to compare two different PSDs.
- The *Itakura-Saito*[2] (spectrum) *distance* between PSDs is only correct if both ly in a local neighborhood.

[2] Basseville 1989.

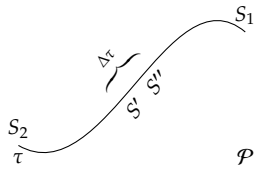
- *Georgiou's distance* is a geodesic distance of PSDs in a manifold. In the remainder, I summarize its derivation by Georgiou[3].

[3] Georgiou 2007.

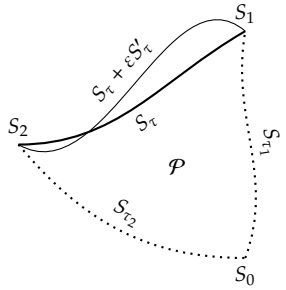
Let $u_k^{(1)}$ denote a random process with underlying PSD S_1. Consider a situation where a distinct PSD S_2 is used for filtering. Then the degradation of the prediction-error variance is defined by

$$\varrho_{\mathrm{a/g}}(S_1, S_2) \triangleq \frac{\mathrm{E}\left\{\left|u_0^{(1)} - \sum_{n=1}^{\infty} a_{n,S_2} u_{-n}^{(1)}\right|^2\right\}}{\mathrm{E}\left\{\left|u_0^{(1)} - \sum_{n=1}^{\infty} a_{n,S_1} u_{-n}^{(1)}\right|^2\right\}} \tag{C.1}$$

Function $\varrho_{\mathrm{a/g}}(S_1, S_2)$ may serve as *mismatch* between the two PSDs.



(a)



(b)

Figure C.1: Geodesic paths in the manifold $\mathcal{P}$. (a) arbitrary path and (b) geodesic path S_τ between S_1, S_2 and the white PSD S_0.

Furthermore

$$\varrho_{\mathrm{a/g}}(S_1, S_2) = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{S_1(\theta)}{S_2(\theta)} \mathrm{d}\theta}{e^{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left(\frac{S_1(\theta)}{S_2(\theta)}\right)\mathrm{d}\theta}} \geq 1 \,. \tag{C.2}$$

Its logarithm

$$\varrho_{\mathrm{a/g}}(S_1, S_2) \triangleq \ln \varrho_{\mathrm{a/g}}(S_1, S_2) \geq 0 \tag{C.3}$$

is termed the Itakura distance. Since $\varrho_{\mathrm{a/g}}(S_1, S_2) \neq \varrho_{\mathrm{a/g}}(S_2, S_1)$ it is symmetrized, i.e.

$$\varrho(S_1, S_2) \triangleq \varrho_{\mathrm{a/g}}(S_1, S_2) + \varrho_{\mathrm{a/g}}(S_2, S_1) \tag{C.4}$$

$$= \ln\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{S_1(\theta)}{S_2(\theta)}\mathrm{d}\theta \; \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{S_2(\theta)}{S_1(\theta)}\mathrm{d}\theta\right) . \tag{C.5}$$

It has two important properties:

1. $\varrho(S_1, S_2) \in \mathbb{R}_+ \cup \{\infty\}$.
2. $\varrho(S_1, S_2) = 0$ if and only if $S_1(\theta)/S_2(\theta)$ is constant.

Next a geometry and a natural metric is developed by collecting all PSDs into a manifold $\mathcal{P}$ where $\varrho(S_1, S_2)$ induces a pseudo-Riemannian metric in $\mathcal{P}$. In this manifold we are interested in the length of the shortest path between two PSDs called geodesic distance.

The differentiable manifold $(\mathcal{P}, 2^{\mathcal{P}})$ is defined by

$$\mathcal{P} = \{S : S \text{ differentiable on } (-\pi, \pi], S(\theta) > 0, \tag{C.6}$$

$$S, \frac{\mathrm{d}S}{\mathrm{d}\theta}, S^{-1} \text{ bounded}\} \tag{C.7}$$

Since $\varrho(S_1, S_2) = 0$ if and only if $S_1(\theta)/S_2(\theta)$ is constant those are in a equivalence class

$$[S] = \{S_1 \in \mathcal{P} : S_1 = cf, \, c \in \mathbb{R}_+\} \,. \tag{C.8}$$

The Taylor approximation of $\varrho(S_1, S_2)$ in (C.5) gives[4]

$$g_S : \mathcal{D} \to \mathbb{R}_+ : \Delta \mapsto g_S(\Delta) = \mathrm{Var}\left\{\frac{\Delta}{S}\right\}^{1/2} \tag{C.10}$$

which is a pseudo-Riemannian metric tensor with natural tangent space $\mathcal{D} = \{\Delta : \Delta, (\mathrm{d}\Delta/\mathrm{d}\theta) \text{ bounded}\}$ and $\Delta/S < 1$.

A small sub-interval of the path $S_\tau, \tau \in [0, 1]$, between PSDs S_1 and S_2 is

$$\Delta\ell = \sqrt{g_{S_\tau}(\dot{S}_\tau \Delta\tau)} \tag{C.11}$$

and hence the length

$$\ell = \int_0^1 \sqrt{g_{S_\tau}(\dot{S}_\tau \mathrm{d}\tau)} \,. \tag{C.12}$$

The geodesic distance demands for the extremum (minimum) of (C.12), see Fig. C.1. The calculus of variations gives the geodesic distance

$$d(S_1, S_2) = \mathrm{Var}\left\{\ln S_1 - \ln S_2\right\}^{1/2} \,. \tag{C.13}$$

Fig. C.1 illustrates the distances between both PSDs and a white signal.

[4]

$$\mathrm{Var}\{\cdot\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\cdot)^2 \mathrm{d}\theta - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \cdot \mathrm{d}\theta\right)^2 \tag{C.9}$$

Bibliography

Akyildiz, I. F., W Su, Y. Sankarasubramaniam, and E. Cayirci (Aug. 2002). "A survey on sensor networks". In: *IEEE Comm. Mag.* 40.8, pp. 102–114 (see p. 1).

Anderson, Brian D. O. and John B. Moore (1979). *Optimal filtering*. Inform. and System Sci. Series. Englewood Cliffs, New Jersey 07632: Rentice-hall, Inc. (see pp. 37, 39).

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (Feb. 2002). "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking". In: *IEEE Trans. Signal Process.* 50.2, pp. 174–188 (see p. 30).

Basseville, M. (1989). "Distance measures for signal processing and pattern recognition". In: *Signal Proc.* 18.4, pp. 349–369 (see pp. 47, 95).

Bauso, D., L. Giarré, and R. Pesenti (2006). "Non-linear protocols for optimal distributed consensus in networks of dynamic agents". In: *Systems & Control Letters* 55.11, pp. 918–928 (see p. 27).

Bell, K. L. and H. L. Van Trees (July 2006). "Combined Cramér-Rao/Weiss-Weinstein bound for tracking target bearing". In: *Fourth IEEE Workshop on Sensor Array and Multichannel Process.* Pp. 273–277 (see p. 47).

Bhattacharyya, A. (1943). "On a measure of divergence between two statistical populations defined by their probability distributions". In: *Bull. Calcutta Math. Soc* 35.99-109, p. 4 (see p. 47).

Billingsley, Patrick (2012). *Probability and measure*. 3rd. John Wiley & Sons, Inc., New York (see p. 5).

Bobrovsky, B. and M. Zakai (Dec. 1975). "A lower bound on the estimation error for Markov processes". In: *IEEE Trans. Autom. Control* 20.6, pp. 785–788 (see p. 46).

Boukerche, A., H. A. B. Oliveira, E. F. Nakamura, and A. A. F. Loureiro (2007). "Localization systems for wireless sensor networks". In: *IEEE Wireless Comm.* 14.6, pp. 6–12 (see p. 1).

Burk, Frank E. (2007). *A Garden of Integrals*. 1st ed. The Dolciani Math. Expositions 31. The Math. Assoc. of Am. (see pp. 5, 15).

Candy, J. V. and E. J. Sullivan (1992). "Ocean acoustic signal processing: a model-based approach". In: *Journal of the Acoust. Soc. Am.* 92.6, pp. 3185–3201 (see p. 1).

– (1996). "Model-based identification: an adaptive approach to ocean-acoustic processing". In: *IEEE Journal of Oceanic Eng.* 21.3, pp. 273–289 (see p. 1).

Candy, J. V. and E. J. Sullivan (2002). "Model-based identification: An adaptive approach to ocean-acoustic processing". In: *IEEE Journal of Oceanic Eng.*, 21.3, pp. 273–289 (see p. 1).

Chernoff, H. (1952). "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations". In: *The Ann. of Math. Stat.* 23.4, pp. 493–507 (see p. 47).

Chow, P. L. (2007). *Stochastic partial differential equations*. Vol. 11. Chapman & Hall/CRC (see pp. 2, 14).

Coates, M. (2004). "Distributed particle filters for sensor networks". In: *Proc. of the 3rd Int. Symp. on Inform. Process. in Sensor Networks*. ACM, pp. 99–107 (see pp. 1, 26).

Dalang, R. C. and N. E. Frangos (1998). "The stochastic wave equation in two spatial dimensions". In: *Ann. of Probability*, pp. 187–212 (see pp. 2, 14, 80).

Dokmanic, I. and M. Vetterli (Mar. 2012). "Room helps: Acoustic localization with finite elements". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. Kyoto, JP, pp. 2617–2620 (see p. 1).

Duan, Z., V. P. Jilkov, and X. R. Li (2008a). "Posterior Cramer-Rao bounds for state estimation with quantized measurement". In: *40th Southeastern Symposium on System Theory (SSST)*. IEEE, pp. 376–380 (see p. 82).

– (2008b). "State estimation with quantized measurements: Approximate MMSE approach". In: *11th Int. Conf. on Inform. Fusion*. IEEE, pp. 1–6 (see p. 2).

Farahmand, S., S. I. Roumeliotis, and G. B. Giannakis (2010). "Particle filter adaptation for distributed sensors via set membership". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. IEEE. Dallas,TX, pp. 3374–3377 (see pp. 1, 26).

Gardner, W. A. (1990). *Introduction to random processes with applications to signals and systems*. McGraw-Hill Publishing Company, New York (see p. 14).

Georgiou, T. T. (2007). "An intrinsic metric for power spectral density functions". In: *IEEE Signal Process. Letters* 14.8, pp. 561–563 (see pp. 41, 95).

Goertz, N. (2007). *Joint source-channel coding of discrete-time signals with continuous amplitudes*. Imperial College Press (see p. 73).

Hausenblas, Erika (2010). "Weak approximation of the stochastic wave equation". In: *Journal of Comp. and App. Math.* 235.1, pp. 33–58 (see pp. 14, 17, 80).

Hlinka, O., F. Hlawatsch, and P. M. Djurić (2013). "Distributed particle filtering in agent networks: A survey, classification, and comparison". In: *IEEE Signal Process. Mag.* 30.1, pp. 61–81 (see p. 1).

Hlinka, O., O. Slučiak, F. Hlawatsch, P. M. Djurić and, and M. Rupp (Aug. 2012). "Likelihood consensus and its application to distributed particle filtering". In: *IEEE Trans. Signal Process.* 60.8, pp. 4334–4349 (see p. 15).

Hol, Jeroen D., Thomas B. Schön, and Fredrik Gustafsson (2006). "On resampling algorithms for particle filters". In: *Nonlinear Statistical*

Signal Process. Workshop, 2006 IEEE. Ed. by sept. IEEE, pp. 79–82 (see p. 22).

Jensen, F., W. A. Kuperman, M. B. Porter, and H. Schmidt (June 2011). *Comp. Ocean Acoust*. 2nd. Springer (see pp. 2, 13, 16, 37, 38).

Jentzen, A. and P. E. Kloeden (2009). "The numerical approximation of stochastic partial differential equations". In: *Milan Journal of Math.* 77.1, pp. 205–244 (see pp. 14–16, 80).

Jovanovic, I., L. Sbaiz, and M. Vetterli (2009). "Acoustic tomography for scalar and vector fields: Theory and application to temperature and wind estimation". In: *Journal of Atmospheric and Oceanic Tech.* 26.8, pp. 1475–1492 (see p. 1).

Kailath, T. (Feb. 1967). "The divergence and Bhattacharyya distance measures in signal selection". In: *IEEE Trans. Commun. Technol.* 15.1, pp. 52–60 (see pp. 47, 55, 73).

Kalman, R. E. (1960). "A new approach to linear filtering and prediction problems". In: *Journal of Basic Eng.* 82.1, pp. 35–45 (see p. 39).

Kay, S. M. (1993). *Fundamentals of statistical signal process., estimation theory*. Vol. 1. NJ: Pearson Education (see p. 39).

Khan, U. A. and J. M. F. Moura (2008). "Distributing the Kalman filter for large-scale systems". In: *IEEE Trans. Signal Process.* 56.10, pp. 4919–4935 (see p. 2).

Kloeden, P. E. and E. Platen (2011). *Numerical solution of stochastic differential equations*. Vol. 23. Springer (see pp. 15, 16).

Knuth, D. E. (1976). "Big omicron and big omega and big theta". In: *ACM Sigact News* 8.2, pp. 18–24 (see p. 28).

Komatitsch, D., S. Tsuboi, and J. Tromp (2005). "The Spectral-Element Method in Seismology". In: *Seismic earth: Array analysis of broadband seismograms*. Ed. by Alan Levander and Guust Nolet. Vol. 157. Geophysical Monograph. Washington DC, USA: Amer. Geophysical Union, pp. 205–228 (see pp. 17, 80).

Kovacs, M., S. Larsson, and F. Saedpanah (2010). "Finite element approximation of the linear stochastic wave equation with additive noise". In: *SIAM Journal on Num. Analysis* 48.2, pp. 408–427 (see p. 17).

Leemis, L. M. and J. T. McQueston (2008). "Univariate distribution relationships". In: *The Amer. Statistician* 62.1, pp. 45–53 (see pp. 53, 57, 60, 61).

Leeuwen, P. J. van (2009). "Particle filtering in geophysical systems". In: *Monthly Weather Review* 137, p. 4089 (see p. 28).

Liu, H. Q., H. C. So, F. K. W. Chan, and K. W. K. Lui (2009). "Distributed particle filter for target tracking in sensor networks". In: *Progress in Electromagnetics Research* 11, pp. 171–182 (see p. 1).

Lu, Y. M. and M. Vetterli (2009). "Distributed spatio-temporal sampling of diffusion fields from sparse instantaneous sources". In: *3rd IEEE Int. Workshop on Comp. Adv. in Multi-Sensor Adaptive Process. (CAMSAP)*. IEEE, pp. 205–208 (see p. 1).

Mattheij, R. M. M., S. W. Rienstra, and J. H. M. ten Thije Boonkkamp (2005). *Partial Differential Equations: Modeling, Analysis, Computation*. Philadelphia: Soc. for Ind. and Applied Math. (see pp. 2, 13).

McDonald, S. (2007). *Finite difference approximation for linear stochastic partial differential equations with method of lines*. Tech. rep. University Library of Munich, Germany (see p. 16).

Meintrup, D. and S. Schäffler (Sept. 2004). *Stochastik: Theorie und Anwendungen*. Springer (see p. 5).

Milne-Thomson, L. M. (1933). *The calculus of finite differences*. Macmillian and Co., Limited (see p. 46).

Minkowski, Hermann (1909). "Raum und Zeit". In: *Jahresbericht der Deutschen Mathematiker-Vereinigung* 18, pp. 75–88 (see p. 13).

Mohammadi, A. and A. Asif (July 2012). "Decentralized sensor selection based on the distributed posterior Cramer-Rao lower bound". In: *15th Inter. Conf. on Inform. Fusion*, pp. 1668–1675 (see pp. 2, 82).

Morse, Philip M. and Herman Feshbach (1953). *Methods of theoretical physics*. McGraw-Hill Book Company, Inc. (see p. 14).

Øksendal, Bernt (2010). *Stochastic differential equations: An introduction with applications*. 6th. Springer (see p. 15).

Oreshkin, B. N. and M. J. Coates (July 2010). "Asynchronous distributed particle filter via decentralized evaluation of Gaussian products". In: *Proc. 6th Int. Conf. of Inform. Fusion*. IEEE. Edinburgh, Scotland (see pp. 1, 26).

Patwari, N., J. N. Ash, S. Kyperountas, III Hero A. O., R. L. Moses, and N. S. Correal (July 2005). "Locating the nodes: Cooperative localization in wireless sensor networks". In: *IEEE Signal Proc. Mag.* 22.4, pp. 54–69 (see p. 1).

Raghavendra, C. S., K. M. Sivalingam, and T. Znati (2006). *Wireless sensor networks*. Springer (see p. 1).

Rapoport, I. and Y. Oshman (Dec. 2004a). "A new estimation error lower bound for interruption indicators in systems with uncertain measurements". In: *IEEE Trans. Inf. Theory* 50.12, pp. 3375–3384 (see p. 3).

– (Dec. 2004b). "Recursive Weiss-Weinstein lower bounds for discrete-time nonlinear filtering". In: *43rd IEEE Conf. on Decision and Control*. Vol. 3, pp. 2662–2667 (see pp. 3, 49, 53, 92).

– (May 2007a). "Weiss-Weinstein lower bounds for Markovian systems. Part 1: Theory". In: *IEEE Trans. Signal Process.* 55.5, pp. 2016–2030 (see pp. 3, 59, 73).

– (May 2007b). "Weiss-Weinstein lower bounds for Markovian systems. Part 2: Applications to fault-tolerant filtering". In: *IEEE Trans. Signal Process.* 55.5, pp. 2031–2042 (see p. 3).

Reece, S. and D. Nicholson (July 2005). "Tighter alternatives to the Cramer-Rao lower bound for discrete-time filtering". In: *8th Int. Conf. on Inform. Fusion*. Vol. 1, 6 pp. (See p. 3).

Reise, G., G. Matz, and K. Gröchenig (Oct. 2012). "Distributed field reconstruction in wireless sensor networks based on hybrid shift-invariant spaces". In: *IEEE Trans. Signal Process.* 60.10, pp. 5426–5439 (see p. 2).

Renaux, A. (2007). "Weiss–Weinstein Bound for Data-Aided Carrier Estimation". In: *Signal Process. Letters, IEEE* 14.4, pp. 283–286 (see p. 3).

Renaux, A., P. Forster, P. Larzabal, C. D. Richmond, and A. Nehorai (Nov. 2008). "A fresh look at the Bayesian bounds of the Weiss-Weinstein family". In: *IEEE Trans. Signal Process.* 56.11, pp. 5334–5352 (see p. 3).

Ristic, B., S. Arulampalam, and N. Gordon (2004). *Beyond the Kalman filter: Particle filters for tracking applications*. Boston: Artech House (see pp. 3, 64, 68, 82).

Ryan, Harold (Sept. 1994). "Ricker, Ormsby, Klauder, Butterworth - A choice of wavelets". In: *CSEG Recorder* Sept. 09 (see p. 30).

Sawo, F. (2009). "Nonlinear state and parameter estimation of spatially distributed systems". PhD thesis. Universität Karlsruhe (see pp. 2, 3).

Sawo, F., V. Klumpp, and U. D. Hanebeck (2008). "Simultaneous state and parameter estimation of distributed-parameter physical systems based on sliced Gaussian mixture filter". In: *11th Int. Conf. on Inform. Fusion*, pp. 1–8 (see p. 23).

Sawo, F., K. Roberts, and U. D. Hanebeck (2006). "Bayesian estimation of distributed phenomena using discretized representations of partial differential equations". In: *3rd Int. Conf. on Informatics in Control, Automation and Robotics (ICINCO)*, pp. 16–23 (see p. 23).

Schön, T., F. Gustafsson, and P.-J. Nordlund (2005). "Marginalized particle filters for mixed linear/nonlinear state-space models". In: *IEEE Trans. Signal Process.* 53.7, pp. 2279–2289 (see pp. 72, 80).

Sheng, X. and Y. H. Hu (2005). "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks". In: *IEEE Trans. on Signal Process.* 53.1, pp. 44–53 (see p. 1).

Simon, D. (2006). *Optimal state estimation: Kalman, H [infinity] and nonlinear approaches*. Wiley-Interscience (see p. 39).

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Philadelphia: Soc. for Indust. and Applied Math. (see p. 13).

Tichavsky, P., C. H. Muravchik, and A. Nehorai (1998). "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering". In: *IEEE Trans. Signal Process.* 46.5, pp. 1386–1396 (see pp. 2, 65).

Tran, N. D., A. Renaux, R. Boyer, S. Marcos, and P. Larzabal (2012). "Weiss-Weinstein bound for MIMO radar with colocated linear arrays for SNR threshold prediction". In: *Signal Processing* 92.5, pp. 1353–1358 (see p. 3).

Tromp, Jeroen, Dimitri Komatitsch, and Qinya Liu (2008). "Spectral element and adjoint methods in seismology". In: *Comm. in Comp. Phys.* 3.1, pp. 1–32 (see p. 80).

Vaidyanathan, P. P. (2008). *The theory of linear prediction*. Synthesis Lect. on Signal Process. Morgan & Claypool Publishers (see pp. 37, 41).

Veeravalli, V.V. and P.K. Varshney (2012). "Distributed inference in wireless sensor networks". In: *Phil. Trans. of the Royal Soc. A: Math., Phys. and Eng. Sci.* 370.1958, pp. 100–117 (see p. 1).

Walsh, J. B. (2006). "On numerical solutions of the stochastic wave equation". In: *Illinois Journal of Mathematics* 50.1-4, pp. 991–1018 (see pp. 16, 80).

Washburn, R. B., T. G. Allen, and D. Teneketzis (June 1985). "Performance analysis for hybrid state estimation problems". In: *Am. Control Conf.* Pp. 1047–1053 (see pp. 3, 82).

Waterschoot, T. van and G. Leus (Mar. 2012). "Distributed estimation of static fields in wireless sensor networks using the finite element method". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. Kyoto, JP, pp. 2853–2856 (see p. 1).

Weiss, A. J. and E. Weinstein (1988). "A general class of lower bounds in parameter estimation". In: *IEEE Trans. Inf. Theory* 34.2, pp. 338–342 (see pp. 3, 9, 46, 47, 59, 91).

Xaver, F., P. Gerstoft, G. Matz, and C.F. Mecklenbräuker (2013). "Analytic Sequential Weiss-Weinstein Bounds". In: *IEEE Trans. Signal Process.* 61.20, pp. 5049–5062 (see pp. 45, 73).

Xaver, F., G. Matz, P. Gerstoft, and N. Görtz (Nov. 2012a). "Localization of acoustic sources utilizing a decentralized particle filter". In: *Proc. 46th Asilomar Conf. Signals, Syst., Comput.* Pacific Grove, CA (see p. 71).

Xaver, F., G. Matz, P. Gerstoft, and C. F. Mecklenbräuker (2011). "Localization of acoustic sources using a decentralized particle filter". In: *EURASIP JWCN* 2011.1, 94ff (see pp. 2, 13, 21, 64).

– (Mar. 2012b). "Predictive state vector encoding for decentralized field estimation in sensor networks". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. Kyoto, JP, pp. 2661–2664 (see pp. 37, 72).

Xaver, F., C. F. Mecklenbräuker, P. Gerstoft, and G. Matz (Nov. 2010). "Distributed state and field estimation using a particle filter". In: *Proc. 44th Asilomar Conf. Signals, Syst., Comput.* Pacific Grove, CA, pp. 1447–1451.

Yardim, C., P. Gerstoft, and W. S. Hodgkiss (2009). "Tracking of geoacoustic parameters using Kalman and particle filters". In: *The Journal of the Acoust. Soc. of Am.* 125, p. 746 (see p. 1).

Yardim, C., Z. H. Michalopoulou, and P. Gerstoft (2011). "An overview of sequential Bayesian filtering in ocean acoustics". In: *IEEE Oceanic Eng* 36(1), pp. 73–91 (see p. 39).

Zhao, F. and L. Guibas (2004). *Wireless sensor networks: An information processing approach*. Morgan Kaufmann (see p. 1).

Zhdanov, M. S. (2002). *Geophysical inverse theory and regularization problems*. Vol. 36. Amsterdam: Elsevier Science Ltd (see pp. 2, 13).

Index

Errata

March 18, 2013	Edition October 24, 2013
Figs. 6.5, 6.6	ξ replaced by ρ
Fig. 6.3	(typo)
Theorem 32	$\partial_x^2 \ln f(x, y)$ exists (removed)
Defs. (6.41), (6.42), Tab. 8.1, Sec. A.1	(typo $\cdot^{-1}$)
Def. (6.61) in Tab. 8.1	copy and past error
Sec. 2.5	removed *unbiased*, added footnote, (2.32) points to to equations now, fixed notation ($\boldsymbol{y} \to \boldsymbol{x}$)
Chapter 6	fixed general Bayesian bounds ($\boldsymbol{y} \to \boldsymbol{x}$), added term *test points*, substituted *parameter matrix/vector* by *test-point matrix / test points*
Chapter 7, 8	substituted *parameter matrix/vector* by *test-point matrix / test points*
Citations	corrected wrong citations