



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Robustness in Dynamic Factor Analysis Identification and Stochastic Optimal Control

Ph.D. candidate

Lucia Falconi

Advisor

Professor Augusto Ferrante

Director & Coordinator

Professor Fabio Vandin

Ph.D. School in
Information Engineering

Department of
Information Engineering
University of Padova

2023

Abstract

This dissertation is divided into two main parts, the common thread being the prominent role of entropy-based metrics in the robust identification and control of stochastic models. The first part is concerned with the *dynamic factor analysis* problem. Factor analysis models boast a long tradition and find natural application in many engineering and scientific disciplines, including, for example, psychology, econometrics, system engineering, machine learning and statistics. In general, the attention for this kind of models is motivated by their effectiveness in complex-data representation. In this part of the thesis, inspired by the previous contributions on robust estimation and robust static factor analysis, we propose a novel approach to deal with the dynamic factor analysis problem for the case of zero-mean Gaussian stochastic processes. To robustly estimate the number of factors, we construct a confidence region centered at a finite sample estimate of the underlying model which contains the true model with a prescribed probability. In this confidence region, we seek for the most parsimonious factor model by solving a convex optimization problem. This paradigm is applied to the identification of moving-average factor models. The obtained result is then generalized to the larger class of autoregressive moving-average factor models by resorting to an iterative technique in which the autoregressive and the moving-average part of the model are identified separately.

The second part of the thesis deals with a finite-horizon *distributionally robust optimal control problem* for linear stochastic uncertain systems. The linear quadratic Gaussian optimal control, which is one of the most fundamental ideas in control theory, suffers from a major disadvantage in that it does not provide systematic means for addressing the issue of robustness. On the other hand, in designing any feedback control law, a fundamental requirement is that of robustness, that is the ability to maintain satisfactory performances even in presence of misspecifications and perturbances in the plant model. In this thesis, a new paradigm is proposed for the robustification of the linear quadratic Gaussian controller against distributional uncertainties on the noise process. Our controller optimizes the closed-loop performances in the worst possible scenario under the constraint that the noise distributional deviation does not exceed a certain threshold limiting the relative entropy pseudo-distance between the actual noise distribution and the nominal one. The main novelty is that the bounds on the distributional deviation can be arbitrarily distributed along the whole disturbance trajectory.

Contents

Abstract	iii
List of symbols and conventions	2
List of abbreviations	3
1 Introduction	5
1.1 Relative entropy and relative entropy rate	7
I Robust identification of dynamic factor analysis models	11
2 Introduction to factor analysis	13
2.1 Static factor analysis problem	14
2.2 Dynamic factor analysis problem	15
2.3 Contribution of the thesis	17
2.4 Motivating examples	19
3 A robust approach to MA factor modeling	21
3.1 Problem formulation	22
3.1.1 The choice of δ	24
3.2 Problem solution	26
3.2.1 Matricial reparameterization of the Problem	26
3.2.2 The dual problem	30
3.2.3 Existence of solutions	33
3.3 Solution to the primal problem	40
3.4 The proposed algorithm	43
3.5 Numerical simulations	47
4 A robust approach to ARMA factor modeling	51
4.1 Problem formulation	52

4.2	Problem solution	53
4.2.1	AR dynamics estimation	53
4.3	Numerical simulations	57
4.3.1	Synthetic example	58
4.3.2	Smart building dataset	59
A	Spectral density estimation	63
A.1	The f -truncated periodogram	64
A.1.1	Unidimensional case	64
A.1.2	Multidimensional case	66
B	Generalized dynamic factor analysis	73
B.1	State-space G DFA models	74
II	Distributionally robust control with distributed uncertainties	81
5	Introduction to optimal and robust control	83
5.1	Contribution of the thesis	85
6	D^2-LQG: worst performance analysis	89
6.1	Preliminary results	90
6.1.1	Gaussian integrals	90
6.1.2	Optimization and convexity	91
6.2	Stochastic uncertain System	92
6.3	Worst performance analysis	94
6.3.1	Least favorable model	98
6.4	Simulations	98
7	D^2-LQG: state feedback control	103
7.1	Stochastic uncertain Systems	104
7.1.1	Relation with the linear fractional transformation model . . .	106
7.2	Minimax optimal control	107
7.3	Simulations	113
7.3.1	Example 1	113
7.3.2	Example 2	115
C	LQG and risk-sensitive LQG control with state-feedback	121
C.1	LQG control	121

C.2	Risk-sensitive LQG control	122
	Summary and outlook	126
	References	137

List of symbols and conventions

Symbol	Description
\mathbb{N}	Set of natural numbers
\mathbb{Z}	Ring of integer numbers
\mathbb{R}	Field of real numbers
\mathbb{C}	Field of complex numbers
$\mathbb{R}^{n \times m}$	Set of $n \times m$ matrices with real entries.
$\mathbb{C}^{n \times m}$	Set of $n \times m$ matrices with complex entries.
\mathbb{Q}_n	Set of $n \times n$ real symmetric matrices
\mathbb{Q}_n^+	Cone of $n \times n$ real symmetric positive semi-definite matrices
$\mathbb{Q}_{m(n)}$	Set of real symmetric block-matrices with $n \times n$ square blocks of dimension $m \times m$
$\mathbb{E}[\cdot], \mathbb{E}$	Expected value of a random variable or vector
\mathcal{S}_m^+	Space of $m \times m$ matrix-valued coercive and bounded spectral densities defined on the unit circle $\{e^{i\theta}; \theta \in [-\pi, +\pi]\}$.
$\mathcal{Q}_{m,n}$	Set of $m \times m$ Hermitian pseudo-polynomial matrices of order n
A_{ij}	Element of matrix A in the i -th row and j -th column
A^\top	Transpose of matrix A
A^*	Hermitian transpose of matrix A
A^{-1}	Inverse of the square matrix A
A^{-T}	Transpose on the inverse of the square matrix A
$\det(A), A $	Determinant of the square matrix A
ker	Kernel of a matrix
rank	Rank of a matrix
tr	Trace of a square matrix
I_n	Identity matrix of dimension $n \times n$ (the subscript is omitted if clear by the context)
$\langle \cdot, \cdot \rangle$	Frobenius inner product, $\langle A, B \rangle := \text{tr}(A^\top B)$ for $A, B \in \mathbb{R}^{n \times n}$
$\ \cdot \ $	Frobenius norm induced by the Frobenius inner product

$\ \cdot\ _2$	Matrix spectral norm
\succeq, \succ	Partial ordering induced by the cone of symmetric positive semi-definite matrices
$\sigma(A)$	Spectrum of the matrix A
$\mathcal{N}(\mu, \Sigma)$	(Multivariate) normal distribution with mean μ and covariance Σ

Throughout the dissertation the abbreviation *a.e.* will stand for almost everywhere. For instance, a property that holds “*a.e.* in \mathbb{C} ” means that the set of point in \mathbb{C} where this property does not hold has zero-measure. Given a function $\Phi(e^{j\vartheta})$ defined over the unit circle $\{e^{j\vartheta} : \vartheta \in [-\pi, \pi]\}$, we write $\Phi(e^{j\vartheta}) \succ 0$ ($\succeq 0$) if $\Phi(e^{j\vartheta})$ is positive (semi)definite $\forall \vartheta \in [-\pi, \pi]$. In Chapter 3 of Part I we use the shorthand notation $\int \Phi$ to denote the integral of the function $\Phi(e^{j\vartheta})$ over the interval $[-\pi, \pi]$ with respect to the normalized Lebesgue measure $d\vartheta/2\pi$. We use both the symbols $x(t)$ and x_t , with $t \in \mathbb{Z}$, to denote the value at time t of the discrete-time stochastic process $x = \{x(t), t \in \mathbb{Z}\}$

List of abbreviations

Abbreviation	Meaning
ADMM	Alternating direction method of multipliers
AR	Autoregressive
ARMA	Autoregressive moving-average
AS	Averaging sequence
BIBO	Bounded-input bounded-output
DFA	Dynamic factor analysis
DGFA	Dynamic generalized factor analysis
DRC	Distributionally robust control
D^2 -LQG	Distributed uncertainty distributionally robust linear quadratic Gaussian
ETFE	Empirical transfer function estimate
FA	Factor analysis
GFA	Generalized factor analysis
LFT	Linear fractional transformation
LQG	Linear quadratic Gaussian
LTI	Linear time-invariant
MA	Moving-average
MIMO	Multiple-input multiple-output
ML	Maximum likelihood
PEM	Prediction error method
SISO	Single-input single-output

1

Introduction

The present dissertation focuses on two fundamental problems in Systems and Control theory: the *identification of dynamic factor analysis models* and the *optimal control of stochastic uncertain systems*. The *leitmotiv* connecting them is the prominent role of entropic measures in guaranteeing robustness of the performance.

In modern society, the formidable development of technology has led to a frenetic increase of the quantity of available data; the interpretation and understanding of the information contained in these data is, on the one hand, an extremely challenging task and, on the other one, a crucial step for any engineering application. For this reason, the research literature has recently witnessed an increasing interest in developing efficient methods to organize the available data in suitable structured models and to provide a concise and parsimonious representation of them. One of the classical methods for this purpose is based on *factor analysis*. This is a statistical tool which aims to describe the variability among a large number of observed correlated variables in terms of a potentially lower number of unobserved variables, called *latent factors*; the observed variables are modelled as linear combinations of the latent factors plus a noisy term.

The origin of factor models goes back to the beginning of the last century in psychology. Since then, they have been an active area of research with countless contributions in many disciplines, from econometrics to statistics to system engineering, to name just a few. Whereas the factor analysis problem was originally formulated in the static case, the idea has been later generalized to the dynamic case, where the observed data originate from stochastic processes. The solution to the dynamic factor analysis problem may be obtained by decomposing the spectral density matrix of the observed data as the sum of a diagonal spectral density matrix (accounting for the noisy component) and a positive semidefinite spectral density matrix, whose rank must be as small as possible since it equals the number of latent factors in the model.

Despite its long tradition, the problem of robustly identifying a dynamic factor analysis model in the realistic situation in which only a finite length realization of the process is available is still an open problem. This has a strong motivation on the fact that even if the underlying data generating process is genuinely low-rank, the minimum rank solution of the classical factor analysis problem rapidly degrades when a certain degree of uncertainty affects our knowledge of the spectral density. Part I of the present dissertation deals with this problem. This part starts off with a general overview on the topic in Chapter 2. Then, it proposes a strategy to robustly estimate dynamic factor models by introducing an adequate confidence region, defined in terms of an entropic measure, centered in a finite sample estimate of the underlying model and containing the true model with a prescribed probability. This paradigm is applied to the identification of MA factor models in Chapter 3 and then extended to the ARMA case in Chapter 4. Part I concludes with two appendices. Appendix A deals with the spectral estimation problem, which represents a preliminary step of our robust dynamic factor analysis procedure. It proposes an estimator of the spectral density which is proved to be mean-square consistent under extremely weak assumptions. Finally in Appendix B we discuss an interesting generalization of the *exact* factor models.

Part I is mostly based on the following works:

- F. Crescente, L. Falconi, F. Rozzi, *et al.*, “Learning AR factor models,” in *2020 59th IEEE Conference on Decision and Control (CDC)*, IEEE, 2020, pp. 274–279
- L. Falconi, A. Ferrante, and M. Zorzi, “A Robust Approach to ARMA Factor Modeling,” *IEEE Transactions on Automatic Control*, 2023
- L. Falconi, A. Ferrante, and M. Zorzi, “Mean-square consistency of the f -truncated M^2 -periodogram,” *Automatica*, vol. 147, p. 110 672, 2023
- G. Picci, L. Falconi, A. Ferrante, *et al.*, “Hidden factor estimation in dynamic generalized factor analysis models,” *Automatica*, vol. 149, p. 110 834, 2023

Stochastic optimal control involves the design of a control policy that minimizes an expected cost of interest under the assumption that the disturbance distribution is known a priori. The most well-known method of this type is the Linear Quadratic Gaussian (LQG) control, which assumes that the disturbance follows a Gaussian distribution with known mean and covariance. Thus, it ignores possible inaccuracies in the distribution information. However, estimating an accurate noise distribution is unrealistic, hence the assumption of a perfect knowledge of the noise description turns out to be unreasonable

in many practical situations. The use of these inaccurate models in the construction of an optimal controller may significantly decrease the control performance and cause unwanted behaviors. To overcome the issue of limited distribution information in stochastic control and make the LQG technique robust against variations and errors in the noise distribution, Part II of this dissertation investigates a *distributionally robust control* (DRC) approach. Instead of assuming a given distribution, the DRC method seeks a control policy that minimizes the expected cost under the worst-case distribution in a so-called ambiguity set. The philosophy behind this approach is to treat the worst-case scenario: if you do not know what you are up against, plan for the worst and optimize!

After a brief introduction in Chapter 5 on the robust and optimal control theory and the DRC approach, we formulate a novel DRC problem where the ambiguity set is specified, for each time instant, as a ball centered at the nominal Gaussian distribution and defined in the relative-entropy topology. In Chapter 6, we solve the worst performance analysis problem in order to characterize the worst-case distribution in the ambiguity set. Exploiting these results, we derive in Chapter 7 the optimal state-feedback control policy.

Part II is mostly based on the following works:

- L. Falconi, A. Ferrante, and M. Zorzi, “A new perspective on robust performance for LQG control problems,” in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 3003–3008
- L. Falconi, A. Ferrante, and M. Zorzi, “Distributionally robust LQG control under distributed uncertainty,” (preliminary version available at *arXiv preprint arXiv:2306.05227 (2023)*)

1.1 Relative entropy and relative entropy rate

The *fil rouge* between the DFA problem introduced in Part I and the DRC problem proposed in Part II is the use of entropic measures to define suitable confidence regions, with the aim of guaranteeing robustness of the performance when a certain degree of uncertainty affects the nominal model of the underlying process. Entropic functionals and divergence criteria have a long-standing tradition in the identification and control of stochastic models and systems, a very popular choice being the relative entropy or Kullback-Leibler pseudo-distance between probability density functions. Here, we briefly recall the definition and some well-known properties of the relative entropy; in doing so we refer mainly to [42]. We further review the concept of relative entropy rate, which may be seen as the generalization of the relative entropy to stochastic systems and may be understood as a distance between spectral density functions.

Definition 1.1.1. Given two probability measures \tilde{f} and f on \mathbb{R}^n , the *relative entropy* (or *Kullback-Leibler divergence*) between \tilde{f} and f is defined as

$$R(\tilde{f}||f) := \begin{cases} \int_{\mathbb{R}^n} \ln \frac{\tilde{f}(x)}{f(x)} \tilde{f}(x) dx & \text{if } \tilde{f} \ll f \\ +\infty & \text{otherwise} \end{cases}$$

where $\tilde{f} \ll f$ denotes the fact that the probability density \tilde{f} is absolutely continuous with respect to the probability density f .

In particular, for the case in which \tilde{f} and f are multivariate normal distributions the integral can be explicitly computed.

Lemma 1.1.2. *Suppose that $\tilde{f} = \mathcal{N}(\tilde{v}, \tilde{V})$ and $f = \mathcal{N}(v, V)$ are Gaussian density functions on \mathbb{R}^n with $V, \tilde{V} \succ 0$, then*

$$R(\tilde{f}||f) = \frac{1}{2} \left(\text{tr}(V^{-1}\tilde{V}) - n + (\tilde{v} - v)^\top V^{-1}(\tilde{v} - v) + \ln \left(\frac{\det V}{\det \tilde{V}} \right) \right).$$

$R(\cdot||\cdot)$ is not a distance since it is not symmetric, and, more importantly, it does not satisfy the triangle inequality. However it is a pseudo-distance as it satisfies the following property: $R(\tilde{f}||f) \geq 0$ and $R(\tilde{f}||f) = 0$ if and only if $\tilde{f}(x) = f(x)$ a.e. It is well known that, for a given probability density function f , $R(\tilde{f}||f)$ is a strictly convex function of \tilde{f} on the set of probability density functions such that $R(\tilde{f}||f) < \infty$. The next variational formula is closely related to some analysis in this dissertation and will be applied on numerous occasions throughout the book.

Lemma 1.1.3 (Relative Entropy and Free Energy duality). *For a given probability density function $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$ and a bounded measurable function $J : \mathbb{R}^n \rightarrow \mathbb{R}$, then*

$$\sup_{\tilde{f}(\cdot) \in \mathcal{P}} \int_{\mathbb{R}^n} J(x) \tilde{f}(x) dx - R(\tilde{f}||f) = \log \int_{\mathbb{R}^n} e^{J(x)} f(x) dx \quad (1.1)$$

where the supremum is taken over the set \mathcal{P} of all the probability density functions $\tilde{f}(x)$ on \mathbb{R}^n . Let $\tilde{f}^o(x)$ denote the probability measure on \mathbb{R}^n which is absolutely continuous with respect to f and satisfies

$$\tilde{f}^o(x) = f(x) \frac{e^{J(x)}}{\int_{\mathbb{R}^n} e^{J(\bar{x})} f(\bar{x}) d\bar{x}}. \quad (1.2)$$

Then, the supremum in the variational formula (1.1) is uniquely attained at $\tilde{f}^o(x)$. The quantity $\log \int_{\mathbb{R}^n} e^{J(x)} f(x) dx$ is called the free-energy of J with respect to f .

When dealing with discrete-time stochastic processes a natural generalization of the notion of relative entropy is the so-called relative entropy rate, interpretable, naively, as the rate of growth of the relative entropy. Let $y = \{y(t); t \in \mathbb{Z}\}$ and $z = \{z(t); t \in \mathbb{Z}\}$ be two zero-mean, jointly Gaussian, stationary processes taking values in \mathbb{R}^n . Let $y_{[-N,N]}$ and $z_{[-N,N]}$ be the random vectors obtained by considering the restriction of y and z to the time interval $\{-N, -N+1, \dots, 0, \dots, N-1, N\}$. Denote by $p_{y_{[-N,N]}}$ and $p_{z_{[-N,N]}}$ the corresponding joint density functions.

Definition 1.1.4. The *relative entropy rate* between y and z is defined as

$$R_r(y||z) := \lim_{N \rightarrow \infty} \frac{1}{2N+1} R(p_{y_{[-N,N]}} || p_{z_{[-N,N]}}).$$

provided that the limit exists.

The following profound, information-theoretic result provides an explicit formula expressing the relative entropy rate between y and z in terms of their spectral densities [70], [117].

Theorem 1.1.5. Let $y = \{y(t), t \in \mathbb{Z}\}$ and $z = \{z(t), t \in \mathbb{Z}\}$ be two zero-mean, jointly Gaussian, stationary processes taking values in \mathbb{R}^n with spectral density functions Φ_y and Φ_z , respectively. Assume that at least one of the following conditions is satisfied:

- (a) $\Phi_y \Phi_z^{-1}$ is bounded;
- (b) $\Phi_y \in L_2(\pi, \pi)$ and Φ_z is coercive, i.e. $\exists \alpha > 0$ such that $\Phi(e^{i\vartheta} - \alpha I_n) > 0$ a.e. on $\{e^{i\vartheta}, \vartheta \in [\pi, \pi]\}$.

Then,

$$R_r(y||z) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \log |\Phi_z(e^{i\vartheta}) \Phi_y^{-1}(e^{i\vartheta})| + \text{tr} \left[\Phi_z^{-1}(e^{i\vartheta}) (\Phi_y(e^{i\vartheta}) - \Phi_z(e^{i\vartheta})) \right] \right\} d\vartheta.$$

Observe that, in the case of scalar spectra, $R_r(y||z) = \frac{1}{2} \mathcal{S}_{IS}(\Phi_y || \Phi_z)$ where

$$\mathcal{S}_{IS}(\Phi_y || \Phi_z) := \int_{-\pi}^{\pi} \left\{ \frac{\Phi_y(e^{i\vartheta})}{\Phi_z(e^{i\vartheta})} - \log \frac{\Phi_y(e^{i\vartheta})}{\Phi_z(e^{i\vartheta})} - 1 \right\} d\vartheta$$

is the celebrated *Itakura-Saito distance* widely used in signal processing [71]. Therefore, the relative entropy rate may be viewed as a multivariate generalization of the classical Itakura-Saito distance.

Part I

**Robust identification of dynamic
factor analysis models**

2

Introduction to factor analysis

A main objective when modelling big data is to organize complex, high dimensional datasets in suitable structured models, providing a simple and parsimonious representation of them. *Factor Analysis* (FA) is a well-established and widely used technique to achieve this goal. It is a statistical tool that explains correlations among a set of observed variables in terms of a much smaller number of non-observed variables, called *latent* or *common factors*, that influence all the observations.

FA boasts a long history. Its origin can be traced back to the seminal work [113] on general intelligence at the beginning of the last century. At the time, the English psychologist Spearman developed a single-common-factor-model to intercorrelate the mental test scores of 36 boys on different topics such as French, English, mathematics and musical talent. Subsequently, in the early 1930s, Thurstone proposed a more general model allowing for more than one common factor, representing different mental abilities, [119]. From these first seeds a rich stream of literature was developed at the interface between psychology and mathematics: [15], [73], [79]–[81], [107], [120]. The interest for this kind of models has rapidly grown also outside the psychology community, and analysis of factor models has become an important tool in several branches of Statistics and Econometrics, as well as in many Engineering sciences: [9], [35], [36], [46], [51], [65], [66], [74], [78], [93], [94], [99] just to name a few.

In all these applications, the key point is that the observable variables are the result of a common, simple behavior plus local interactions. The typical visual representation of this situation is that of a flock of birds where the trajectory of each single bird is determined by the “average” trajectory of the flock and by a variation proper to the individual bird. With this premise, FA describes each of the observed variables as the sum of a component depending on few latent factors, common to all the observations, plus *idiosyncratic noise*, affecting each of the variables independently of the others. The latent factors capture the co-movements among the observations, whereas the idiosyncratic

disturbances account for measurement errors and special features that are specific to each variable. The purpose of FA consists in characterizing the common factors, representing the compressed information, and the idiosyncratic noise. Once the latent factors have been extracted, the statistical description of the complex system could be reasonably restricted to the common component and be based on the simple model thereof. This is of paramount importance in order to reduce data dimensionality, as well as to gain a better understanding and interpretation of the complex interrelationships among the measured variables.

The idea of FA is particularly important if the cross-sectional dimension is large with respect to sample size. In such situations, conventional identification methods may lead to problems in which the number of parameters can be of the same order of magnitude or even larger than the sample size. FA is believed to provide an answer to this issue as it concentrates the explication of the high-dimensional data in a (hopefully) small-dimensional vector of latent factors.

2.1 Static factor analysis problem

In the original formulation, a factor model is linear static model given by

$$y = W_L u + W_D w \quad (2.1)$$

where $W_L \in \mathbb{R}^{m \times r}$, with $r \ll m$, and $W_D \in \mathbb{R}^{m \times m}$ diagonal; $u := [u_1 \dots u_r]^\top$ and $w := [w_1 \dots w_m]^\top$ are Gaussian random vectors of dimension r and m , respectively, with zero mean and covariance matrix equal to the identity. The vectors u and w are assumed to be independent. Model (2.1) has the following interpretation: the m -dimensional random vector y is the *observed vector*; W_L is the *factor loading matrix*, u represents the (independent) *latent factors* and $\hat{y} := W_L u$ is the *latent variable*; finally $\tilde{y} := W_D w$ corresponds to the *idiosyncratic component*. Note that an essential part of the model specification is that the m components of the idiosyncratic error \tilde{y} are mutually uncorrelated. The aim of model (2.1) is to provide an explanation of the mutual correlations of the observed variables y_i in terms of a smaller number of common factors, in the sense that $\mathbb{E}[y_i y_j] = \mathbb{E}[\hat{y}_i \hat{y}_j]$ for all $i \neq j$.

From the FA model (2.1), it follows that y is itself a zero mean Gaussian random vector with covariance matrix Σ given by

$$\Sigma = L + D \quad (2.2)$$

where $L := W_L W_L^\top$ is a low-rank matrix - with rank equal to r - and $D := W_D W_D^\top$ is diagonal. Hence, in its original conception the construction of a factor model is mathematically equivalent to the matrix additive decomposition (2.2) where $\Sigma \succeq 0$ is given and the rank r of L must be as small as possible. This can be seen as a special kind of *low rank plus sparse* decomposition problem of the covariance matrix [29], [31], a diagonal matrix being, in intuitive terms, as sparse as one could possibly ask for.

Clearly model (2.2) is maximally parsimonious if the rank of L is minimum. The problem of minimizing the rank of L in decomposition (2.2) is known as *Frisch's problem*. It turns out that, in general, this is a formidable problem and, to date, no exact solution is actually available, with the only exception of the special case when this minimum rank is $r = m - 1$ [103]. This lack of explicit formulas has motivated a rich stream of literature, where countless variations have been considered and studied; we refer the reader to the recent papers [10], [16], [27], [92] where different approaches have been proposed. A first difficulty related to the Frisch's problem is the non-convexity of the rank function. A typical strategy to practically compute a decomposition of type (2.2) where the rank of L is small is to consider a convex relaxation of the problem, where, in place of the rank, the trace norm of L is minimized, see [37], [52], [53]. In addition, it is a well-known fact that estimation of FA models is an ill-posed problem. In fact, even a minuscule variation in the covariance matrix Σ of the observed data usually leads to a substantial variation of the number of hidden factors, which is the key feature of the modeling procedure. On the other hand, from an applications standpoint such a matrix must be estimated from finite data records, hence it is inevitably subject to errors. This sample covariance matrix may not have a low-rank decomposition even if the underlying true covariance matrix does. The problem of estimating the number of hidden factors from the sample covariance matrix is therefore of crucial importance. It has been addressed in [9] and [77] by means of statistical methods, and in [92] by making use of the 2-Wasserstein distance. An alternative optimization-based approach to robustly estimate the number of factors is proposed in [27] leveraging on a scale invariance property of the relative-entropy pseudo-distance.

2.2 Dynamic factor analysis problem

Whereas the initial approach to FA was oriented to the static case, the idea has been further generalized to data originating from stochastic processes. *Dynamic Factor Analysis* (DFA) has been addressed much more recently than its static counterpart, the first contribution to this field being apparently [66]. Subsequently, dynamic factor models

have been deeply investigated, in particular in the fields of Econometrics and Statistics. We refer to the surveys [10], [36], [84], [114] and to the recent paper [59] for an overview of the contributions provided by such a community over the recent years and a rich list of references. More recently, a growing interest on factor models has been witnessed also in the Systems and Control engineering community due to their ability of modeling time series data of large cross-sectional dimension, see e.g. [20], [27], [28], [135]. Indeed, several modern engineering applications are characterized by interconnected systems with hundreds or thousands of variables mainly “driven” by a small number of hidden factors, for which the DFA decomposition proves useful; we will provide some examples at the end of the Chapter.

The natural extension of model (2.1) to the dynamic case is

$$y(t) = W_L(e^{i\vartheta})u(t) + W_D(e^{i\vartheta})w(t), \quad t \in \mathbb{Z}, \quad (2.3)$$

where $W_L(e^{i\vartheta})$ and $W_D(e^{i\vartheta})$ are transfer functions in the form

$$W_L(e^{i\vartheta}) = \sum_{k=0}^{+\infty} W_{L,k} e^{-i\vartheta k}, \quad W_{L,k} \in \mathbb{R}^{m \times r} \quad (2.4)$$

$$W_D(e^{i\vartheta}) = \sum_{k=0}^{+\infty} W_{D,k} e^{-i\vartheta k}, \quad W_{D,k} \in \mathbb{R}^{m \times m} \text{ diagonal} \quad (2.5)$$

and $u = \{u(t), t \in \mathbb{Z}\}$ and $w = \{w(t), t \in \mathbb{Z}\}$ are i.i.d. Gaussian processes of dimension r and m , respectively, where $r \ll m$, with zero mean and covariance equal to the identity. It is assumed that $\mathbb{E}[u(t)w(s)^\top] = 0$ for all $t, s \in \mathbb{Z}$. Here, u is the process which describes the r latent factors not accessible to observations, W_L is the factor loading transfer matrix, $\hat{y} := W_L u$ is the latent variable and finally $\tilde{y} := W_D w$ is the idiosyncratic noise. Accordingly, the process $y = \{y(t), t \in \mathbb{Z}\}$ is a zero mean Gaussian process with power spectral density Φ given by

$$\Phi = \Phi_L + \Phi_D \quad (2.6)$$

with $\Phi_L := W_L W_L^*$ low-rank - where by rank we mean the normal rank, *i.e.* the rank almost everywhere - and $\Phi_D := W_D W_D^*$ diagonal. Therefore, y represents a DFA model if its spectral density can be decomposed as “low-rank plus diagonal” as in (2.6).

The basic formulation of DFA problem consists in determining the minimum r for which a DFA decomposition (2.6) holds. This problem is extremely challenging and it is still open. In the realistic situation in which only a finite sample estimate $\hat{\Phi}$ of the spectral density Φ of the observed data is available, solving this problem is even more difficult. Indeed, as already observed for the static case, the solution to the DFA

problem is extremely fragile so that the accuracy in the estimation of Φ may severely affect the decomposition (2.6). Note that this problem may be cast in the rich stream of literature devoted to the identification of dynamic graphical models with latent factors where a decomposition of the type “low-rank plus sparse” of a certain spectral density is considered, see for example [28], [87], [135]. Differently from the previous papers, however, in DFA we do not take into account the presence of an underlying graphical model and we try to find a “low-rank plus diagonal” decomposition.

A wide literature is available on dynamic factor models, with different strands of research evolving around the topic. A first fundamental issue is that of estimating the number r of latent factors. Up to now, the main contribution to this problem is given by [68], where the authors proposed an information-based criterion based on some asymptotic properties of the eigenvalues of the sample spectral density; they established, under appropriate assumptions, the consistency of their method as the cross-sectional dimension and the length of the observed series both tend to infinity.

Once r is known, the main research interests concern the identification of the factor model parameters and the estimation the common factors from the observed data. Different approaches are available: [45], [125] and [115] focused on time-domain techniques. They consider a special dynamic factor model wherein the common (dynamic) factors are only combined in a static way is considered. The parameters of the model are estimated by maximum likelihood, then a Kalman filter is used to obtain efficient estimates of the factors. More recently, principal component techniques have been introduced to consistently estimate the latent factors from the observable variables as the cross sectional dimension m and the sample size N both tend to infinity [61]–[64]. An alternative likelihood-based method has been shown in [41], where the factor model is cast in state-space form and the likelihood is maximized using the Expectation Maximization (EM) algorithm.

2.3 Contribution of the thesis

In this thesis, we propose a novel optimization-based paradigm for the estimation of dynamic factor models in the realistic situation in which only a finite sample estimate $\hat{\Phi}$ of the underlying model is available. We remark that the proposed method simultaneously solves the problem of estimating the number r of factors and identifying the DFA model parameters. This differentiates it from the existing literature where the solution to the DFA problem is typically performed in two steps, first the estimation of the number of factors and then the identification of the model. To cope with the fragility issue

characterizing the DFA problem and robustly estimate the number of factors, we define a confidence region in the *Itakura-Saito topology* centered at the sampled spectrum $\hat{\Phi}$ and containing the true model with a user-chosen probability. In this confidence region we seek for the most parsimonious factor model admitting a “low-rank plus diagonal” decomposition.

This paradigm is applied to the identification of MA factor models in Chapter 3. In more detail, given a finite sample estimate of the underlying data generating process, we propose a resampling-based method to construct a confidence region containing the true spectrum with prescribed probability. In this confidence region, the problem, formulated as a rank minimization of a suitable spectral density, is efficiently approximated via a trace norm convex relaxation. The latter is addressed by resorting to the Lagrange duality theory, which allows to prove the existence of solutions. Finally, a numerical algorithm to solve the dual problem is presented. The effectiveness of the proposed estimator is assessed through simulation studies with synthetic data. The results of Chapter 3 are published in

- L. Falconi, A. Ferrante, and M. Zorzi, “A Robust Approach to ARMA Factor Modeling,” *IEEE Transactions on Automatic Control*, 2023.

In Chapter 4 we extend the previous approach to ARMA processes. The proposed solution consists of two steps: an AR dynamic estimation step, for which we introduce a novel moment matching method, followed by a DFA identification step for MA models. Although the above procedure is suboptimal, numerical simulations both with synthetic and real data show that the resulting estimator performs well. Chapter 4 is based on

- F. Crescente, L. Falconi, F. Rozzi, *et al.*, “Learning AR factor models,” in *2020 59th IEEE Conference on Decision and Control (CDC)*, IEEE, 2020, pp. 274–279
- L. Falconi, A. Ferrante, and M. Zorzi, “A Robust Approach to ARMA Factor Modeling,” *IEEE Transactions on Automatic Control*, 2023.

The DFA techniques introduced in Chapter 3 and Chapter 4 build upon an estimate of the spectral density function of the underlying system from a finite length record of observed data. We address the spectral estimation problem in Appendix A, where we propose the *f-truncated periodogram*, that is a truncated periodogram where the truncation point is a suitable function f of the sample size. We provide a unified, conceptually simple and self-contained proof of the asymptotic consistency of the proposed estimator that holds for possibly multidimensional and multivariate random processes. The results of this Appendix are published in

- L. Falconi, A. Ferrante, and M. Zorzi, “Mean-square consistency of the f -truncated M^2 -periodogram,” *Automatica*, vol. 147, p. 110 672, 2023.

Finally, Appendix B discusses an interesting generalization of the factor model (2.3) which allows cross-correlation among the idiosyncratic components. For this kind of model, we address the problem of estimating the hidden factors from the observed data, given a state-space representation of the underlying system. In particular, we analyze the capability of Kalman estimators to tackle this problem. The results of this chapter are published in

- G. Picci, L. Falconi, A. Ferrante, *et al.*, “Hidden factor estimation in dynamic generalized factor analysis models,” *Automatica*, vol. 149, p. 110 834, 2023.

2.4 Motivating examples

We conclude this introductory chapter by presenting a selection of practical problems relevant to the Systems and Control community where the application of DFA models shows promise.

1) *Air Pollution Monitoring*: Assume that we want to monitor the concentration over time of two pollutants, namely benzene (C₆H₆) and carbon monoxide (CO), in a certain city by means of a large number of sensors spread all over the area. Assume that each sensor measures either the concentration of C₆H₆ or CO. This situation can be efficiently described by a DFA model where the latent variables correspond the average concentration of the two pollutants at time t , and the observed signals correspond to the outputs of the sensors. Since the concentration of the pollutants varies considerably within the city, the sensor output at time t is a measure of the average concentration corrupted by a random fluctuation related to its specific position and to accidental measurement error.

2) *Smart Building Modelling*: Consider the problem of reducing the energy consumption of a smart building which integrates a whole range of different technologies and a complex monitoring system. For example, suppose that the house has indoor sensors for temperature, humidity and carbon dioxide; outdoor sensors for lighting measurements, wind speed, rain, sun irradiance and temperature. In order to improve the energy efficiency, one may want to forecast the evolution of some signals of interest. To this end, it is necessary to derive a mathematical model of the system. In this scenario, we have a large number of signals collected by the sensors which are strongly correlated and

show a common behavior. Hence, it is reasonable to expect that most of the variability can be explained in terms of a smaller number of factors, say heating or air conditioning consumption, average temperature, humidity etc. The application of DFA techniques may help in providing a parsimonious model of the smart building system, which is easy to understand and interpret.

3) *Dynamic texture modelling*: Large-dimensional time series often occur in computer vision and dynamic image processing, where we observe a signal $y(t)$ with tens of thousands of components, obtained by stacking the intensities at time t of each pixel of an image. For instance, we may be interested in modelling and analyzing dynamic textures, which are sequences of images showing temporal regularity, such as smoke, flames, flowing water, or moving grass. Dynamic textures images may be described in terms of hidden variables, say the state of the texture, corrupted by additive noise, see [39]. Hence, we can use DFA techniques to identify these models and compress the high dimensional vector $y(t)$ into simple mathematical structures. Note that, in dynamic textures modeling, typically the number of samples that can be used for identification is in the same order or smaller than the data dimensionality. Indeed the number of images in the sequences is in the order of a few hundreds while the dimension of $y(t)$ (which is equal to the number of pixels of the image) is in the order of a few hundreds or thousands [39].

Other significant engineering applications for which factor models seem to have great potential are presented in [20].

3

A robust approach to MA factor modeling

This chapter deals with the DFA problem for MA processes. The aim is to extract, from a high dimensional stream of observed data, a linear MA model featuring a small number of hidden variables. This is important both from the point of view of the model simplicity and to uncover the structure of the mechanism generating the data.

Mathematically, the DFA problem may be formulated as that of decomposing the spectral density of the process generating the data as the sum of a diagonal spectral density and a low-rank one, with the minimum possible rank. One of the main difficulties is that the solution is inherently fragile; in fact, even if the underlying data generating process is genuinely low rank, the minimum rank solution to this problem rapidly degrades when a certain degree of uncertainty affects the spectral density function. In this chapter we propose a strategy to cope with this fragility issue by taking a point of view whose nature is similar to the approach used by [27] in the *static* FA setting. More precisely, given an estimate of the underlying model based on a finite data sample, we propose a method which accounts for the uncertainty in the estimation by computing a “confidence neighbourhood” containing the true model with a prescribed probability. We define this confidence neighbourhood as a ball centred at the estimated model and defined in the *Itakura-Saito topology*. In this region, we search for a structured model which admits the smallest number of latent factors.

The overall procedure may be summarized as follows:

- Given the observed data, we compute a raw estimate $\hat{\Phi}$ of the spectral density Φ generating the data (see Appendix A for a discussion on spectral density estimation).
- We compute a neighbourhood \mathcal{N} of $\hat{\Phi}$ that contains Φ with prescribed probability; clearly the size of \mathcal{N} depends on the sample size.
- We compute a refined estimate $\Phi^\circ \in \mathcal{N}$ by imposing that it admits an additive decomposition as a diagonal spectral density and a spectral density with the lowest

possible rank. To this end we set up an optimization problem that we address by resorting to duality. In particular, we prove existence of solutions and provide a numerical algorithm to compute a solution.

Notice that, whereas in the existing literature the solution to the DFA problem consists of two steps, first the estimation of the number of factors and then the identification of the model parameters, the proposed method simultaneously solves these two problems.

The present chapter is published in

- L. Falconi, A. Ferrante, and M. Zorzi, “A Robust Approach to ARMA Factor Modeling,” *IEEE Transactions on Automatic Control*, 2023

and it is outlined as follows. In Section 3.1 we introduce the DFA problem for moving-average MA models; in Section 3.2 we prove that such a problem admits a solution by means of duality theory; Section 3.3 shows how to reconstruct the solution to the primal problem from the dual one; in Section 3.4 we propose an algorithm to compute the solution to the dual problem; finally Section 3.5 presents some numerical results.

Throughout the chapter integrals are always defined from $-\pi$ to π with respect to the normalized Lebesgue measure $d\vartheta/2\pi$; to simplify the notation, the interval of integration and the normalized Lebesgue measure are omitted.

3.1 Problem formulation

Consider the MA factor model whose order is n :

$$y(t) = W_L u(t) + W_D w(t) \quad (3.1)$$

where

$$W_L(e^{i\vartheta}) = \sum_{k=0}^n W_{L,k} e^{-i\vartheta k}, \quad W_D(e^{i\vartheta}) = \sum_{k=0}^n W_{D,k} e^{-i\vartheta k},$$

$W_{L,k} \in \mathbb{R}^{m \times r}$, $W_{D,k} \in \mathbb{R}^{m \times m}$ diagonal; $u = \{u(t), t \in \mathbb{Z}\}$ and $w = \{w(t), t \in \mathbb{Z}\}$ are normalized white Gaussian noises of dimension r and m , respectively, such that $\mathbb{E}[u(t)w(s)^\top] = 0 \forall t, s$. The process u describes the r factors, with $r \ll m$, not accessible to observation; W_L is the factor loading transfer matrix; $W_L u(t)$ is the latent variable; $W_D w(t)$ is idiosyncratic noise. Accordingly, $y = \{y(t), t \in \mathbb{Z}\}$ is a m -dimensional Gaussian stationary stochastic process with power spectral density

$$\Phi = \Phi_L + \Phi_D \quad (3.2)$$

where $\Phi_L = W_L W_L^* \succeq 0$ and $\Phi_D = W_D W_D^* \succeq 0$ belong to the finite dimensional space:

$$\mathcal{Q}_{m,n} := \left\{ \sum_{k=-n}^n R_k e^{-i\vartheta k}, \quad R_k = R_{-k}^T \in \mathbb{R}^{m \times m} \right\}.$$

By construction, $\text{rank}(\Phi_L) = r$, where rank denotes the normal rank, and Φ_D is diagonal. Hence, y represents a factor model if its spectral density can be decomposed as “low rank + diagonal” as in (3.2).

Assume to collect a finite length realization of y defined in (3.1), say $y^N = \{y(1) \dots y(N)\}$, where the order n is known. We want to estimate the corresponding factor model, that is the decomposition in (3.2), as well as the number of factors r . To address this problem, given the data y^N , we first compute the sample covariance lags \hat{R}_j as

$$\hat{R}_j = \frac{1}{N} \sum_{t=0}^{N-j} y(t+j)y(t)^\top, \quad j = 0 \dots n.$$

Then, an estimate $\hat{\Phi}$ of Φ is obtained by the truncated periodogram:

$$\hat{\Phi} = \sum_{k=-n}^n \hat{R}_k e^{i\vartheta k}. \quad (3.3)$$

Notice that $\hat{\Phi}$ could be not positive definite for all ϑ ; in that case, we can add εI_m to the right side of Equation (3.3), with the constant $\varepsilon > 0$ chosen in such a way as to ensure the positivity of $\hat{\Phi}$. On the other hand, $\hat{\Phi}$ may not admit a low rank plus diagonal decomposition. Thus, we estimate directly the two terms Φ_L and Φ_D of the decomposition (3.2) by solving the following optimization problem:

$$\begin{aligned} & \min_{\Phi, \Phi_L, \Phi_D \in \mathcal{Q}_{m,n}} \text{rank}(\Phi_L) \\ & \text{subject to } \Phi_L + \Phi_D = \Phi, \\ & \Phi \succ 0 \text{ a.e.}, \quad \Phi_L, \Phi_D \succeq 0, \\ & \Phi_D \text{ diagonal}, \\ & \mathcal{S}_{IS}(\Phi || \hat{\Phi}) \leq \delta. \end{aligned} \quad (3.4)$$

The first three constraints impose that Φ_L and Φ_D provide a genuine spectral density decomposition of type (3.2). The last constraint, in which $\mathcal{S}_{IS}(\Phi || \hat{\Phi})$ is the Itakura-Saito

divergence defined by ([56], [71])

$$\mathcal{S}_{IS}(\Phi||\hat{\Phi}) := \int \log |\hat{\Phi}\Phi^{-1}| + \text{tr}(\hat{\Phi}^{-1}\Phi - I_m),$$

imposes that Φ belongs to a set “centered” at the nominal spectral density $\hat{\Phi}$ and with prescribed tolerance δ . Problem (3.4) is in general an extremely hard problem because of the non-convexity of the rank function. A tractable convex-relaxation of (3.4) is given by

$$\begin{aligned} \min_{\Phi, \Phi_L, \Phi_D \in \mathcal{Q}_{m,n}} \quad & \text{tr} \int \Phi_L \\ \text{subject to} \quad & \Phi_L + \Phi_D = \Phi, \\ & \Phi \succ 0 \text{ a.e.}, \quad \Phi_L, \Phi_D \succeq 0, \\ & \Phi_D \text{ diagonal}, \\ & \mathcal{S}_{IS}(\Phi||\hat{\Phi}) \leq \delta. \end{aligned} \tag{3.5}$$

The substitution of the rank with the trace norm from Problem (3.4) to Problem (3.5) is justified by the fact that $\text{tr} \int \Phi_L$ represents the convex hull of $\text{rank}(\Phi_L)$ (see [135]).

Notice that Φ_D is uniquely determined by Φ and Φ_L . Thus, Problem (3.5) can be rewritten by removing Φ_D :

$$\begin{aligned} (\Phi^\circ, \Phi_L^\circ) = \quad & \arg \min_{\Phi, \Phi_L \in \mathcal{Q}_{m,n}} \text{tr} \int \Phi_L \\ \text{subject to} \quad & \Phi \succ 0 \text{ a.e.}, \quad \Phi_L, \Phi - \Phi_L \succeq 0, \\ & \Phi - \Phi_L \text{ diagonal}, \\ & \mathcal{S}_{IS}(\Phi||\hat{\Phi}) \leq \delta. \end{aligned} \tag{3.6}$$

3.1.1 The choice of δ

Before solving our problem, we deal with the choice of the tolerance parameter δ appearing in the constraint of (3.6). This choice should reflect the accuracy of the estimate $\hat{\Phi}$ of Φ . This can be accomplished by choosing a desired probability $\alpha \in (0, 1)$ and considering a ball of radius δ_α (in the Itakura-Saito topology) centered in $\hat{\Phi}$ and containing the true spectrum Φ with probability α .

The estimation of δ_α is not an easy task because we do not know the true power spectral density Φ . Next, we propose a resampling-based method to estimate it. The idea

is to approximate Φ with $\hat{\Phi}$, and use this model to perform a resampling operation. Let

$$W(e^{i\vartheta}) = \sum_{k=0}^n W_k e^{-i\vartheta k}, \quad W_k \in \mathbb{R}^{m \times m}$$

be the minimum phase spectral factor of $\hat{\Phi}$ and define the process $\hat{y} = \{\hat{y}(t), t \in \mathbb{Z}\}$ as $\hat{y}(t) := W(e^{i\vartheta})e(t)$, where $e(t)$ is an m -dimensional normalized white noise. The truncated periodogram (understood as *estimator*) based on a sample of the process \hat{y} of length N is

$$\hat{\Phi}_r(e^{i\vartheta}) = \sum_{k=-n}^n e^{-i\vartheta k} \frac{1}{N} \sum_{t=0}^{N-k} \hat{y}(t+k)\hat{y}(t)^T,$$

where the subscript “ r ” stands for resampling, as it is the means by which we perform the resampling operation, and the boldface notation $\hat{\Phi}_r$ is used to highlight that this is an *estimator*, namely a random matrix and must not be confused with the corresponding *estimate* which is denoted by $\hat{\Phi}_r$. The latter is a deterministic matrix obtained by replacing the random process $\hat{y}(t)$ with the corresponding realization $\hat{y}(t)$. By generating a realization $\hat{y}^N = \{\hat{y}(1) \dots \hat{y}(N)\}$ from $\hat{\Phi}$ (i.e. by resampling the data), we can easily obtain a realization of the random variable $\mathcal{S}_{IS}(\hat{\Phi} || \hat{\Phi}_r)$. Accordingly, it is possible to compute numerically δ_α such that $\Pr(\mathcal{S}_{IS}(\hat{\Phi} || \hat{\Phi}_r) \leq \delta_\alpha) = \alpha$ by a standard Monte Carlo procedure. Numerical simulations show that this technique indeed provides a good estimate of δ .

It is worth noting that if the chosen α is too large with respect to the data length N , the resulting δ_α may be too generous yielding a diagonal Φ obeying $\mathcal{S}_{IS}(\Phi || \hat{\Phi}) \leq \delta_\alpha$. In this case Problem (3.6) admits the trivial solution $\Phi_L = 0$ and $\Phi_D = \Phi$. To rule out this trivial case, δ in (3.6) must be strictly smaller than the upper bound

$$\delta_{\max} := \min_{\substack{\Phi \in \mathcal{S}_m^+ \\ \Phi \text{ diagonal}}} \mathcal{S}_{IS}(\Phi || \hat{\Phi})$$

where \mathcal{S}_m^+ denotes the family of bounded and coercive functions defined on the unit circle and taking values in the cone of positive definite $m \times m$ Hermitian matrices. Since Φ must be diagonal, by denoting with ϕ_i and by $\hat{\gamma}_i$ the i -th element in the diagonal of Φ and of $\hat{\Phi}^{-1}$, respectively, we have

$$\delta_{\max} = \min_{\substack{\Phi \in \mathcal{S}_m^+ \\ \Phi \text{ diagonal}}} \int \log |\hat{\Phi}\Phi^{-1}| + \text{tr}(\hat{\Phi}^{-1}\Phi - I_m)$$

$$\begin{aligned}
&= \min_{\substack{\Phi \in \mathcal{S}_m^+ \\ \Phi \text{ diagonal}}} \int \log \left| \hat{\Phi} \left(\text{diag}^2(\hat{\Phi}^{-1}) \right) \left(\text{diag}^2(\hat{\Phi}^{-1}) \right)^{-1} \Phi^{-1} \right| + \text{tr} \left(\hat{\Phi}^{-1} \Phi - I_m \right) \\
&= \min_{\substack{\Phi \in \mathcal{S}_m^+ \\ \Phi \text{ diagonal}}} \int \log \left| \left(\text{diag}^2(\hat{\Phi}^{-1}) \right)^{-1} \Phi^{-1} \right| + \text{tr}[\hat{\Phi}^{-1} \Phi - I_m] + \int \log |\hat{\Phi} \text{diag}^2(\hat{\Phi}^{-1})| \\
&= \min_{\substack{\Phi \in \mathcal{S}_m^+ \\ \Phi \text{ diagonal}}} \int \log \left(\prod_{i=1}^m \hat{\gamma}_i^{-1} \phi_i^{-1} \right) + \sum_{i=1}^m \hat{\gamma}_i \phi_i - m + \int \log |\hat{\Phi} \text{diag}^2(\hat{\Phi}^{-1})| \\
&= \left[\sum_{i=1}^m \min_{\phi_i \in \mathcal{S}_1^+} \int \log(\hat{\gamma}_i^{-1} \phi_i^{-1}) + \hat{\gamma}_i \phi_i - 1 \right] + \int \log |\hat{\Phi} \text{diag}^2(\hat{\Phi}^{-1})| \\
&= \left[\sum_{i=1}^m \min_{\phi_i \in \mathcal{S}_1^+} \mathcal{S}_{IS}(\phi_i || \hat{\gamma}_i^{-1}) \right] + \int \log |\hat{\Phi} \text{diag}^2(\hat{\Phi}^{-1})|
\end{aligned}$$

where $\text{diag}^2(\cdot)$ is the (orthogonal projection) operator mapping a square matrix M into a diagonal matrix of the same size having the same main diagonal of M . Therefore, since the Itakura-Saito divergence is nonnegative, the solution corresponds to $\phi_i^{\text{opt}}(e^{i\vartheta}) = (\hat{\gamma}_i(e^{i\vartheta}))^{-1}$, $i = 1, \dots, m$ for which $\mathcal{S}_{IS}(\phi_i^{\text{opt}} || \hat{\gamma}_i^{-1}) = 0$. Accordingly,

$$\delta_{max} = \int \log |\hat{\Phi} \text{diag}^2(\hat{\Phi}^{-1})|. \quad (3.7)$$

The derivation of the aforementioned result is based on reasonings similar to [28, Section IV].

A more generous upper bound can be derived by assuming that Φ is the spectrum of an MA process of order n . However, numerical experiments showed that $\delta_{max} \gg \delta_\alpha$ even in the case that N is relatively small.

3.2 Problem solution

In this section we first provide a finite dimensional matrix parameterization of Problem (3.6). The latter is then analyzed by resorting to the Lagrange duality theory, which allows us to prove the existence of a solution.

3.2.1 Matricial reparameterization of the Problem

To study Problem (3.6) it is convenient to introduce a matricial parameterization for Φ , Φ_L and $\Phi - \Phi_L$. To this end, we first introduce the so-called shift operator:

$$\Delta(e^{i\vartheta}) := [I_m \quad e^{i\vartheta} I_m \quad \dots \quad e^{in\vartheta} I_m]; \quad (3.8)$$

and the finite dimensional space $\mathbb{Q}_{m(n+1)}$ of symmetric block-matrices with $(n+1) \times (n+1)$ square blocks of dimension $m \times m$. If $X \in \mathbb{Q}_{m(n+1)}$, X_{ij} denotes the block of X in position i, j with $i, j = 0, \dots, n$, so that

$$X = \begin{bmatrix} X_{00} & X_{01} & \dots & X_{0n} \\ X_{01}^\top & X_{11} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ X_{0n}^\top & X_{1n}^\top & \dots & X_{nn} \end{bmatrix}.$$

Moreover, $\mathbf{M}_{m,n}$ denotes the vector space of matrices of the form

$$Y := [Y_0 \ Y_1 \ \dots \ Y_n], \quad Y_0 \in \mathbb{Q}_m, \ Y_1, \dots, Y_n \in \mathbb{R}^{m \times m}. \quad (3.9)$$

The linear mapping $T : \mathbf{M}_{m,n} \rightarrow \mathbb{Q}_{m(n+1)}$ constructs a symmetric block-Toeplitz matrix from its first block row so that if Y is given by (3.9),

$$T(Y) = \begin{bmatrix} Y_0 & Y_1 & \dots & Y_n \\ Y_1^\top & Y_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & Y_1 \\ Y_n^\top & \dots & Y_1^\top & Y_0 \end{bmatrix}.$$

If $[D(X)]_0 = \sum_{h=0}^n X_{hh}$ and $[D(X)]_j = 2 \sum_{h=0}^{n-j} X_{hh+j}$, for $j = 1, \dots, n$, the adjoint of T is the mapping $D : \mathbb{Q}_{m(n+1)} \rightarrow \mathbf{M}_{m,n}$ defined by $D(X) = \left[[D(X)]_0 \ \dots \ [D(X)]_n \right]$. Given $X \in \mathbb{Q}_{m(n+1)}$, by direct computation we obtain

$$\Delta X \Delta^* = [D(X)]_0 + \frac{1}{2} \sum_{j=1}^n e^{-ij\vartheta} [D(X)]_j + e^{ij\vartheta} [D(X)]_j^\top, \quad (3.10)$$

thus $\Delta X \Delta^* \in \mathbb{Q}_{m,n}$. Conversely, since D is a surjective map, any element in $\mathbb{Q}_{m,n}$ may be parameterized as (3.10). We conclude that

$$\mathbb{Q}_{m,n} = \left\{ \Delta X \Delta^* \text{ s.t. } X \in \mathbb{Q}_{m(n+1)} \right\},$$

and we introduce the following matrix parameterization for Φ , Φ_L and $\Phi - \Phi_L$:

$$\begin{aligned}\Phi &= \Delta X \Delta^* \in \mathcal{Q}_{m,n} \\ \Phi_L &= \Delta L \Delta^* \in \mathcal{Q}_{m,n} \\ \Phi - \Phi_L &= \Delta(X - L) \Delta^* \in \mathcal{Q}_{m,n}\end{aligned}\tag{3.11}$$

with X and L matrices in $\mathbb{Q}_{m(n+1)}$.

Next, the objective is to provide a more convenient formulation of Problem (3.6) in terms of X and L . To this end, we have to take into account the following points.

1) Positivity Constraints $\Phi \succ 0$ a.e. and $\Phi_L, \Phi - \Phi_L \succeq 0$:

It can be shown (see for example [135, Appendix A]) that, for any $\Psi \in \mathcal{Q}_{m,n}$, $\Psi \succeq 0$ if and only if there exists a matrix $P \in \mathbb{Q}_{m(n+1)}$ such that $\Delta P \Delta^*$ and $P \succeq 0$. Therefore, we replace the conditions $\Phi_L \succeq 0$ with $L \succeq 0$, the condition $\Phi - \Phi_L \succeq 0$ with $X - L \succeq 0$. Note that these conditions only guarantee $X \succeq 0$ and thus Φ to be positive semidefinite, however we will show that this is sufficient to guarantee that $\Phi \succ 0$ a.e. at the optimum.

2) Constraint $\Phi - \Phi_L$ diagonal:

Let $\text{ofd} : \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$ denote the linear operator defined as follows: given $A \in \mathbb{R}^{m \times m}$, $\text{ofd}(A)$ is the matrix in which each off-diagonal element is equal to the corresponding element of A and each diagonal element is zero. We define the ‘‘block ofd’’ linear operator $\text{ofd}_B : \mathbf{M}_{m,n} \rightarrow \mathbf{M}_{m,n}$ as follows. Given $Z = [Z_0 \ Z_1 \ \dots \ Z_n] \in \mathbf{M}_{m,n}$, then $\text{ofd}_B(Z) = [\text{ofd}(Z_0) \ \text{ofd}(Z_1) \ \dots \ \text{ofd}(Z_n)]$. It is not difficult to show that ofd_B is a self-adjoint operator, since ofd is self-adjoint as well. Then, it is easy to see that the condition $\Phi - \Phi_L$ diagonal is equivalent to the condition $[D(X - L)]_j$ diagonal for $j = 0, \dots, n$, that is $\text{ofd}_B(D(X - L)) = 0$.

3) The Low Rank Regularizer:

We have

$$\text{tr} \int \Phi_L = \text{tr} \int \Delta L \Delta^* = \text{tr} \left(L \int \Delta^* \Delta \right) = \text{tr}(L)$$

where we exploited the fact that $\int e^{ij\vartheta} = 1$ if $j = 0$, and $\int e^{ij\vartheta} = 0$ otherwise.

4) The Divergence Constraint:

A convenient matrix parameterization of the Itakura-Saito divergence $\mathcal{S}_{IS}(\Phi || \hat{\Phi})$ can be obtained by making use of the following facts.

First, since $\Phi = \Delta X \Delta^*$ with $X \succeq 0$, there exists $A \in \mathbb{R}^{m \times m(n+1)}$ such that $X = A^\top A$.

Then, by using the Jensen-Kolmogorov formula we obtain

$$\int \log |\Phi| = \int \log |\Delta A^\top A \Delta^*| = \log |A_0^\top A_0| = \log |X_{00}| \quad (3.12)$$

which holds provided that $X_{00} \succ 0$ and Φ is coercive.¹ We need to generalize this result to spectral densities that may be singular on the unit circle. This is possible because the zeros of a rational spectral density, if any, have finite multiplicity so that the logarithm of the determinant of a rational spectral Φ is integrable as long as the normal rank of Φ is full.

Lemma 3.2.1. *Consider a power spectral density $\Phi \in \mathcal{Q}_{m,n}$ having full normal rank. Let $X \in \mathcal{Q}_{m(n+1)}$ be such that $X \succeq 0$, $X_{00} \succ 0$, and $\Phi = \Delta X \Delta^*$. Then $\int \log |\Phi| = \log |X_{00}|$.*

Proof. Since $\Phi = \Delta X \Delta^*$ with $X \succeq 0$, there exists $A \in \mathbb{R}^{m \times m(n+1)}$ such that $X = A^\top A$. The matrix A is such that $\Phi \succeq 0$ admits the spectral factorization $\Phi = W W^*$ where $W := \Delta A^\top$. Now, define $\Phi_n := \Phi + \frac{1}{n} I$ with $n \in \mathbb{N}$ and let $W_n := \Delta A_n$ be a spectral factor of Φ_n with $A_n \in \mathbb{R}^{m \times m(n+1)}$. Clearly, $\lim_{n \rightarrow +\infty} \Phi_n = \Phi$; accordingly, $\lim_{n \rightarrow +\infty} W_n = W$ and $\lim_{n \rightarrow +\infty} A_n = A$. Since $\Phi_n \succ 0 \forall \vartheta$ we can exploit (3.12) to obtain $\int \log |\Phi_n| = \log |A_{n0}^\top A_{n0}|$. Then, applying the limit operator to both sides, we have

$$\lim_{n \rightarrow +\infty} \int \log |\Phi_n| = \log |A_0^\top A_0| = \log |X_{00}|.$$

To conclude the proof, it remains to show that in the left side of the previous equation it is possible to interchange the limit and the integral operators. To this aim, we introduce the sequence $\{f_n\}_{n=1}^{+\infty}$ where $f_n(t) := \log |\Phi_n(\vartheta)|$ and the function $f(\vartheta) := \lim_{n \rightarrow +\infty} f_n(t) = \log |\Phi(\vartheta)|$. Observe that, since the interval of integration $[-\pi, \pi]$ is bounded and $f_1(\vartheta) < +\infty$ for any $\vartheta \in [-\pi, \pi]$, then $\int f_1(\vartheta) d\vartheta < +\infty$. We also define the sequence $\{g_n\}_{n=1}^{+\infty}$ as $g_n(\vartheta) := f_n(\vartheta) - f_1(\vartheta)$ and $g(\vartheta) := \lim_{n \rightarrow +\infty} g_n(\vartheta)$. $\{g_n\}$ is a pointwise non-increasing sequence of measurable non-positive functions,

$$\dots \leq g_2(\vartheta) \leq g_1(\vartheta) \leq 0, \quad \forall \vartheta \in [-\pi, +\pi]$$

converging to $g(\vartheta)$ from above. Hence, it satisfies all the hypotheses of Beppo-Levi's monotone convergence theorem (applied with opposite signs), from which it immediately follows that $\lim_{n \rightarrow +\infty} \int g_n(\vartheta) = \int g(\vartheta)$, and consequently

$$\lim_{n \rightarrow +\infty} \int f_n(\vartheta) = \int g(\vartheta) + \int f_1(\vartheta). \quad (3.13)$$

¹A spectral density Φ is said to be *coercive* if it is positive definite on the unit circle.

Now, since $f_1(\vartheta) < +\infty$ for all ϑ ,

$$g(\vartheta) = f(\vartheta) - f_1(\vartheta), \quad (3.14)$$

and, by plugging (3.14) into (3.13), we finally obtain $\lim_{n \rightarrow +\infty} \int f_n(\vartheta) = \int f(\vartheta)$. ■

A second observation in order to conveniently parameterize the Itakura-Saito divergence constraint is that, by exploiting the cyclic property of the trace,

$$\int \text{tr}(\hat{\Phi}^{-1}\Phi) = \int \text{tr}(\hat{\Phi}^{-1}\Delta X \Delta^*) = \text{tr}\left(X \int \Delta^* \hat{\Phi}^{-1} \Delta\right) = \langle X, T(\hat{P}) \rangle,$$

where \hat{P} is defined from the expansion $\hat{\Phi}^{-1} = \sum_{k=-\infty}^{\infty} \hat{P}_k e^{-i\vartheta k}$ as $\hat{P} := [\hat{P}_0 \dots \hat{P}_n]$.

Summing up, we get the following matrix re-parameterization of Problem (3.6):

$$\begin{aligned} (X^\circ, L^\circ) = \arg \min_{X, L \in \mathbb{Q}_{m(n+1)}} \quad & \text{tr}(L) \\ \text{subject to} \quad & X_{00} \succ 0, L \succeq 0, X - L \succeq 0, \\ & \text{ofd}_B(D[X - L]) = 0, \\ & -\log |X_{00}| + \int \log |\hat{\Phi}| + \langle X, T(\hat{P}) \rangle - m \leq \delta. \end{aligned} \quad (3.15)$$

We remark once again that to prove the equivalence between (3.6) and (3.15) we still need to show that $\Phi \succ 0$ a.e. at the optimum: this fact will be established after the variational analysis.

3.2.2 The dual problem

We reformulate the constrained minimization problem in (3.15) as an unconstrained problem by means of Duality Theory, [21].

If we use $V, U \in \mathbb{Q}_{m(n+1)}$, $V, U \succeq 0$ as the multipliers associated with the constraints on the positive semi-definiteness of $X - L$ and L , respectively, $Z \in \mathbf{M}_{m,n}$ as the multiplier associated with the constraint $\text{ofd}_B(D(X - L)) = 0$ and $\lambda \in \mathbb{R}, \lambda \geq 0$, as the multiplier

associated with the Itakura-Saito divergence, then the Lagrangian is

$$\begin{aligned}
\mathcal{L}(X, L, \lambda, U, V, Z) &= \text{tr}(L) - \langle V, X - L \rangle - \langle U, L \rangle + \langle Z, \text{ofd}_B(D(X - L)) \rangle \\
&\quad + \lambda \left(-\log |X_{00}| + \int \log |\hat{\Phi}| + \langle X, T(\hat{P}) \rangle - m - \delta \right) \\
&= \langle L, I \rangle - \langle V, X - L \rangle - \langle U, L \rangle + \langle T(\text{ofd}_B(Z)), X - L \rangle + \langle X, \lambda T(\hat{P}) \rangle \\
&\quad - \lambda \left(\log |X_{00}| - \int \log |\hat{\Phi}| + m + \delta \right) \\
&= \langle L, I - U + V - T(\text{ofd}_B(Z)) \rangle + \langle X, T(\text{ofd}_B(Z)) - V + \lambda T(\hat{P}) \rangle \\
&\quad - \lambda \left(\log |X_{00}| - \int \log |\hat{\Phi}| + m + \delta \right).
\end{aligned} \tag{3.16}$$

where in the first equality of the previous equation we have first exploited the fact that the operator ofd_B is self-adjoint and then the fact that the mappings T and D are adjoints. Note that we have not included the constraint $X_{00} \succ 0$ because, as we will show later on, this condition is automatically met by the solution to the dual problem.

The dual function is defined as the infimum of \mathcal{L} over X and L . Thanks to the convexity of the Lagrangian, we rely on standard variational methods to characterize the minimum.

- **Partial minimization with respect to L :**

The Lagrangian \mathcal{L} depends on L only through $\langle L, I - U + V - T(\text{ofd}_B(Z)) \rangle$ which is bounded from below only if

$$I - U + V - T(\text{ofd}_B(Z)) = 0. \tag{3.17}$$

Thus, we get that

$$\inf_L \mathcal{L} \begin{cases} = \langle X, T(\text{ofd}_B(Z)) - V + \lambda T(\hat{P}) \rangle - \\ \lambda \left(\log |X_{00}| - \int \log |\hat{\Phi}| + m + \delta \right) & \text{if (3.17)} \\ = -\infty & \text{otherwise.} \end{cases}$$

- **Partial minimization with respect to X :**

The terms in X_{00} are bounded from below only if

$$\left[T(\text{ofd}_B(Z)) - V + \lambda T(\hat{P}) \right]_{00} \succ 0 \tag{3.18}$$

If (3.18) holds, by taking convexity into account, the matrix X_{00} achieving the minimum is easily obtained by imposing the optimality condition $\delta \mathcal{L}(X_{00}; \delta X_{00}) = 0$

for any $\delta X_{00} \in \mathbb{Q}_m$. This is equivalent to

$$\begin{aligned} \delta \mathcal{L}(X_{00}; \delta X_{00}) &= \text{tr} \left(\left[T(\text{ofd}_B(Z)) - V + \lambda T(\hat{P}) \right]_{00} \delta X_{00} \right) - \lambda \text{tr} \left(X_{00}^{-1} \delta X_{00} \right) \\ &= \left\langle \left[T(\text{ofd}_B(Z)) - V + \lambda T(\hat{P}) \right]_{00} - \lambda X_{00}^{-1}, \delta X_{00} \right\rangle \\ &= 0, \quad \forall \delta X_{00}. \end{aligned}$$

It follows that the terms in X_{00} are minimized if $\lambda > 0$ and

$$X_{00} = \left(\left[T(\hat{P}) + \lambda^{-1} (T(\text{ofd}_B(Z)) - V) \right]_{00} \right)^{-1}. \quad (3.19)$$

The Lagrangian is linear in the remaining variables X_{lh} , for $(l, h) \neq (0, 0)$, and therefore bounded from below only if

$$\left[T(\text{ofd}_B(Z)) - V + \lambda T(\hat{P}) \right]_{lh} = 0 \quad \forall (l, h) \neq (0, 0). \quad (3.20)$$

Summing up, the minimization of the Lagrangian with respect to X and L is finite if and only if (3.17), (3.18), and (3.20) hold true in which case

$$\min_{X, L} \mathcal{L} = -\lambda \left(-\log \left| \left[T(\hat{P}) + \lambda^{-1} (T(\text{ofd}_B(Z)) - V) \right]_{00} \right| - \int \log |\hat{\Phi}| + \delta \right).$$

Otherwise the Lagrangian has no minimum and its infimum is $-\infty$.

To simplify the notation, let us define the vector space

$$\mathcal{O} := \{Z \in \mathbf{M}_{m,n} : \text{ofd}_B(Z) = Z\};$$

since Z always appears in the form $\text{ofd}_B(Z)$, we can replace it with $Z \in \mathcal{O}$. Then, we can formulate the dual problem for the Lagrangian (3.16) as

$$\max_{(\lambda, U, V, Z) \in \tilde{\mathcal{C}}} \tilde{J} := \lambda \left(\log \left| \left[T(\hat{P}) + \lambda^{-1} (T(Z) - V) \right]_{00} \right| + \int \log |\hat{\Phi}| - \delta \right) \quad (3.21)$$

where the feasible set $\tilde{\mathcal{C}}$ is given by:

$$\begin{aligned} \tilde{\mathcal{C}} := \{ & (\lambda, U, V, Z) : U, V \in \mathbb{Q}_{m(n+1)}, U, V \succeq 0, Z \in \mathcal{O}, I - U + V - T(Z) = 0, \lambda \in \mathbb{R}, \\ & \lambda > 0, [\lambda T(\hat{P}) + T(Z) - V]_{00} \succ 0, [\lambda T(\hat{P}) + T(Z) - V]_{lh} = 0 \quad \forall (l, h) \neq (0, 0) \}. \end{aligned}$$

Note that the constraints $I - U + V - T(Z) = 0$ and $U \succeq 0$ are equivalent to the constraint $I + V - T(Z) \succeq 0$. Thus, we can eliminate the redundant variable U ; moreover, by

changing the sign to the objective function \tilde{J} and observing that

$$\left[T(\hat{P}) + \lambda^{-1}(T(Z) - V) \right]_{00} = \hat{P}_0 + \lambda^{-1}(Z_0 - V_{00}),$$

we can rewrite (3.21) as a minimization problem

$$\min_{(\lambda, V, Z) \in \mathcal{C}} J := \lambda \left(-\log \left| \hat{P}_0 + \lambda^{-1}(Z_0 - V_{00}) \right| - \int \log |\hat{\Phi}| + \delta \right) \quad (3.22)$$

where the corresponding feasible set \mathcal{C} is:

$$\mathcal{C} := \left\{ (\lambda, V, Z) : V \in \mathbb{Q}_{m(n+1)}, V \succeq 0, Z \in \mathcal{O}, \lambda \in \mathbb{R}, \lambda > 0, I + V - T(Z) \succeq 0, \right. \\ \left. [\lambda \hat{P}_0 + Z_0 - V_{00}] \succ 0, [\lambda(T(\hat{P})) + T(Z) - V]_{lh} = 0 \quad \forall (l, h) \neq (0, 0) \right\}.$$

3.2.3 Existence of solutions

The aim of this section is to show that (3.22) admits a solution. The set \mathcal{C} is not compact, as it is neither closed nor bounded. We show that we can restrict the search of the minimum of J over a compact set. Then, since the objective function is continuous over \mathcal{C} (and hence over the restricted compact set), we can use Weierstrass Theorem to conclude that the problem does admit a minimum.

The first step consists in showing that we can restrict \mathcal{C} to a subset where $\lambda \geq \varepsilon$ with ε is a positive constant.

Proposition 3.2.2. *Let $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{N}}$ be a sequence of elements in \mathcal{C} such that*

$$\lim_{k \rightarrow \infty} \lambda^{(k)} = 0. \quad (3.23)$$

Then, such a sequence cannot be an infimizing sequence.

Proof. The proof exploits arguments similar to [28, Proposition 6.1]. We consider two possible scenarios separately.

Scenario 1: Let $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{N}}$ be such that, besides (3.23), we have $\| (\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) \| \rightarrow +\infty$. Since we are dealing with symmetric matrices, this is equivalent to

$$\lim_{k \rightarrow \infty} \max_{\alpha^{(k)} \in \sigma \left((\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) \right)} |\alpha^{(k)}| = +\infty. \quad (3.24)$$

Next, we show that (3.24) leads to

$$\lim_{k \rightarrow \infty} \min_{\alpha^{(k)} \in \sigma \left((\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) \right)} \alpha^{(k)} = -\infty. \quad (3.25)$$

Indeed, (3.24) implies that at least one of the two conditions (3.25) and

$$\lim_{k \rightarrow \infty} \max_{\alpha^{(k)} \in \sigma \left((\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) \right)} \alpha^{(k)} = +\infty \quad (3.26)$$

holds. To show that (3.26) implies (3.25), notice that, since $V^{(k)} \succeq 0$ and $\lambda^{(k)} > 0 \forall k$, then $\max \left\{ \alpha^{(k)} : \alpha^{(k)} \in \sigma \left((\lambda^{(k)})^{-1} [Z^{(k)}]_0 \right) \right\}$ is greater than or equal to the argument of the limit in the left-side of (3.26). Thus (3.26) implies

$$\lim_{k \rightarrow \infty} \max_{\alpha^{(k)} \in \sigma \left((\lambda^{(k)})^{-1} [Z^{(k)}]_0 \right)} \alpha^{(k)} = +\infty. \quad (3.27)$$

But $(\lambda^{(k)})^{-1} [Z^{(k)}]_0$ is traceless $\forall k$, hence the sum of its eigenvalues is zero and

$$\lim_{k \rightarrow \infty} \min_{\alpha^{(k)} \in \sigma \left((\lambda^{(k)})^{-1} [Z^{(k)}]_0 \right)} \alpha^{(k)} = -\infty. \quad (3.28)$$

By the same argument as before, from $V^{(k)} \succeq 0$ and $\lambda^{(k)} > 0$ we obtain that $\forall k$ $\min \left\{ \alpha^{(k)} : \alpha^{(k)} \in \sigma \left((\lambda^{(k)})^{-1} [Z^{(k)}]_0 \right) \right\}$ is greater than or equal to the argument in the left side of (3.25). This fact, together with (3.28), leads to (3.25). By (3.25), we obtain that for k sufficiently large $[Z^{(k)}]_0 - [V^{(k)}]_{00} + \lambda^{(k)} \hat{P}_0 = (\lambda^{(k)}) \left((\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{P}_0 \right)$ has at least a negative eigenvalue, so the sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})$ is not in \mathcal{C} .

Scenario 2: Consider now a sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{N}}$ for which, besides (3.23), we have $\| (\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) \| \rightarrow c$ with $0 \leq c < \infty$. Since $\| (\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) \|$ is bounded, there exists $l \in \mathbb{R}$ such that for all k

$$(\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{P}_0 \preceq lI_m.$$

Then, since for any two matrices $A, B \in \mathbf{Q}_m$, if $0 \preceq A \preceq B$ then $\det(A) \leq \det(B)$, we have

$$-\log |(\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{P}_0| - \int \log |\hat{\Phi}| + \delta \geq -m \log l - \int \log |\hat{\Phi}| + \delta.$$

If $l_1 := -m \log l - \int \log |\hat{\Phi}| + \delta$, it holds that $J(\lambda^{(k)}, V^{(k)}, Z^{(k)}) \geq \lambda^{(k)} l_1$ for any k . Note

that $\lambda^{(k)}l_1 \rightarrow 0$ thus $\forall \varepsilon > 0, \exists \bar{k}$ such that

$$J(\lambda^{(k)}, V^{(k)}, Z^{(k)}) \geq \lambda^{(k)}l_1 > -\varepsilon, \quad \forall k > \bar{k}$$

Now, it is sufficient to find a triple $(\bar{\lambda}, \bar{V}, \bar{Z}) \in \mathcal{C}$ with $J(\bar{\lambda}, \bar{V}, \bar{Z})$ strictly negative to conclude that such a sequence is not an infimizing sequence. Consider $\bar{\lambda}$ sufficiently small, but strictly greater than zero and $\bar{Z}_j = -\bar{\lambda} \text{ofd}(\hat{P}_j)$ for all $j = 0, \dots, n$. Moreover, let $\hat{P}_{j,d} = \text{diag}^2(\hat{P}_j)$, $j = 0, \dots, n$, $\hat{P}_d = [\hat{P}_{0,d} | \dots | \hat{P}_{n,d}]$ and $\mathcal{T}_{n+1,d} = T(\hat{P}_d)$. Observe that $\mathcal{T}_{n+1,d}$ is defined from $T(\hat{P})$ by the same ‘‘block by block diagonalization’’ procedure defined in [28, Lemma A.2], so it is positive definite. Partition $\mathcal{T}_{n+1,d}$ as

$$\mathcal{T}_{n+1,d} = \begin{bmatrix} \hat{P}_{0,d} & \mathcal{K}_d \\ \mathcal{K}_d^\top & \mathcal{T}_{n,d} \end{bmatrix}$$

which defines the matrices \mathcal{K}_d and $\mathcal{T}_{n,d}$ and let

$$\bar{V} = \bar{\lambda} \begin{bmatrix} \mathcal{K}_d \mathcal{T}_{n,d}^{-1} \mathcal{K}_d^\top & \mathcal{K}_d \\ \mathcal{K}_d^\top & \mathcal{T}_{n,d} \end{bmatrix}.$$

As already noticed, in view of [28, Lemma A.2], $\mathcal{T}_{n,d}$ is positive definite, and from the Schur Complement \bar{V} is positive semi-definite. It is possible to see that the triple $(\bar{\lambda}, \bar{V}, \bar{Z})$ belongs to \mathcal{C} for $\bar{\lambda}$ sufficiently small. Moreover, $\text{diag}^2(\hat{\Phi}^{-1})$ is the power spectral density of the process whose covariance lags are $\hat{P}_{d,j}$. Hence, in view of [28, Lemma A.1] we have that

$$\begin{aligned} J(\bar{\lambda}, \bar{V}, \bar{Z}) &= -\bar{\lambda} \log |\hat{P}_{0,d} - \mathcal{K}_d \mathcal{T}_{n,d}^{-1} \mathcal{K}_d^\top| - \bar{\lambda} \int \log |\hat{\Phi}| + \bar{\lambda} \delta \\ &\leq -\bar{\lambda} \int \log |\text{diag}^2(\hat{\Phi}^{-1})| - \bar{\lambda} \int \log |\hat{\Phi}| + \bar{\lambda} \delta \\ &= \bar{\lambda} \left(\delta - \int \log |\hat{\Phi} \text{diag}^2(\hat{\Phi}^{-1})| \right) = \bar{\lambda}(\delta - \delta_{max}) \\ &< 0, \end{aligned}$$

where in the last equality we have taken into account the expression (3.7) while the last inequality follows from the assumption $\delta_{max} > \delta$.

The only remaining possibility is the case in which $\lim_{k \rightarrow \infty} \|(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00})\|$ does not exist. However, in this case it is always possible to consider a subsequence for which the limit exists (finite or infinite) and we can therefore reduce to one of the

previous cases. ■

As a consequence of Proposition 3.2.2, minimizing the dual functional over the set \mathcal{C} is equivalent to minimize it over the set

$$\mathcal{C}_1 := \left\{ (\lambda, V, Z) : V \in \mathbb{Q}_{m(n+1)}, V \succeq 0, Z \in \mathcal{O}, I + V - T(Z) \succeq 0, \lambda \in \mathbb{R}, \right. \\ \left. \lambda \geq \varepsilon, [\lambda \hat{P}_0 + Z_0 - V_{00}] \succ 0, [\lambda(T(\hat{P})) + T(Z) - V]_{lh} = 0 \forall (l, h) \neq (0, 0) \right\}.$$

Next, we show that we can restrict \mathcal{C}_1 to a subset in which both $(T(Z) - V)$ and λ cannot diverge.

Proposition 3.2.3. *Let $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{N}}$ be a sequence of elements in \mathcal{C}_1 such that either*

$$\lim_{k \rightarrow \infty} \|T(Z^{(k)}) - V^{(k)}\| = +\infty \quad (3.29)$$

or

$$\lim_{k \rightarrow \infty} \lambda^{(k)} = +\infty \quad (3.30)$$

or both. Then, such a sequence cannot be an infimizing sequence.

Proof. This proof follows arguments similar to the proof of Proposition 6.2 in [28], with a few small differences. First, we show that (3.29) holds if and only if (3.30) holds as well. We are assuming that the estimated model has a non-trivial dynamics, *i.e.* there exists $i \neq 0$ such that $\hat{P}_i \neq 0$. Thus, $[T(\hat{P})]_{\bar{l}\bar{h}} \neq 0$ for some $(\bar{l}, \bar{h}) \neq (0, 0)$. From $[\lambda^{(k)}T(\hat{P}) + T(Z^{(k)}) - V^{(k)}]_{\bar{l}\bar{h}} = 0$, which is one of the conditions for the sequence to be in \mathcal{C}_1 , (3.30) holds if and only if $[T(Z^{(k)}) - V^{(k)}]_{\bar{l}\bar{h}}$ diverges. It remains to show that if

$$\lim_{k \rightarrow \infty} \|[Z^{(k)}]_0 - [V^{(k)}]_{00}\| = +\infty, \quad (3.31)$$

then (3.30) holds. Since we are dealing with symmetric matrices, (3.31) is equivalent to

$$\lim_{k \rightarrow \infty} \max_{\alpha^{(k)} \in \sigma\left(\left([Z^{(k)}]_0 - [V^{(k)}]_{00}\right)\right)} |\alpha^{(k)}| = +\infty, \quad (3.32)$$

which in turn implies at least one of the two conditions

$$\lim_{k \rightarrow \infty} \min_{\alpha^{(k)} \in \sigma\left(\left([Z^{(k)}]_0 - [V^{(k)}]_{00}\right)\right)} \alpha^{(k)} = -\infty \quad (3.33)$$

$$\lim_{k \rightarrow \infty} \max_{\alpha^{(k)} \in \sigma\left(\left([Z^{(k)}]_0 - [V^{(k)}]_{00}\right)\right)} \alpha^{(k)} = +\infty. \quad (3.34)$$

With similar reasonings to the proof of Proposition 3.2.2, it is possible to show that (3.34) implies (3.33). Consequently, since $(\lambda^{(k)}\hat{P}_0 + [Z^{(k)}]_0 - [V^{(k)}]_{00}) \succ 0$, we conclude that (3.30) holds. So far we have seen the equivalence between (3.29) and (3.30) and the fact that (3.31) implies (3.30). Next, we show that (3.30) implies not only (3.31), but the stronger condition

$$\lim_{k \rightarrow \infty} \frac{\| [Z^{(k)}]_0 - [V^{(k)}]_{00} \|}{\lambda^{(k)}} \neq 0. \quad (3.35)$$

Assume by contradiction that (3.35) does not hold. Then the corresponding sequence does not belong to \mathcal{C}_1 as the constraint on the positive semi-definiteness of $I + V^{(k)} - T(Z^{(k)})$ fails for k sufficiently large. Indeed, consider a principal minor of order 2 of $I + V^{(k)} - T(Z^{(k)})$ as follows. Select a block \hat{P}_h with $h \neq 0$ and an element in position (p, q) such that $[\hat{P}_h]_{(p,q)} \neq 0$. Then, consider the 2×2 sub-matrix of $I + V^{(k)} - T(Z^{(k)})$

$$\begin{bmatrix} 1 + [V^{(k)}]_{00} - [Z^{(k)}]_0 & \lambda^{(k)}[\hat{P}_h]_{(p,q)} \\ \lambda^{(k)}[\hat{P}_h]_{(p,q)} & 1 + \lambda^{(k)}[\hat{P}_0]_{(q,q)} \end{bmatrix}, \quad (3.36)$$

where $[\hat{P}_0]_{(q,q)} > 0$ and the off-diagonal terms are obtained by employing the constraint $[\lambda^{(k)}T(\hat{P}) + T(Z^{(k)}) - V^{(k)}]_{lh} = 0$ for $(l, h) \neq (0, 0)$. It is possible to see that the determinant of (3.36) diverges to $-\infty$ as $k \rightarrow +\infty$. Recalling that a symmetric matrix is positive semidefinite if and only if every principal minor is non-negative, we conclude that the constraint on positive semidefiniteness of $I + V^{(k)} - T(Z^{(k)})$ fails for k sufficiently large, leading to a contradiction. Therefore the proof reduces to ruling out the following two possible cases.

Scenario 1: Consider the case of a sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{N}}$ such that, besides (3.30), we also have

$$\lim_{k \rightarrow \infty} \frac{\| [Z^{(k)}]_0 - [V^{(k)}]_{00} \|}{\lambda^{(k)}} = +\infty.$$

We can repeat the same reasoning of the Proof of Proposition 3.2.2 to conclude that at least one eigenvalue of $(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00})$ tends to $-\infty$ as $k \rightarrow \infty$. This implies that for k sufficiently large the positivity of $[Z^{(k)}]_0 - [V^{(k)}]_{00} + \lambda^{(k)}\hat{P}_0 = \lambda^{(k)}\left((\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{P}_0\right)$ fails, which rules out this case.

Scenario 2: Finally, consider a sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{N}}$ in \mathcal{C}_1 for which $\| [Z^{(k)}]_0 - [V^{(k)}]_{00} \| \rightarrow \infty$ at the same speed of $\lambda^{(k)}$ and $\| [T(Z^{(k)}) - V^{(k)}]_{lh} \|$. Since $(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{P}_0 \succeq 0$, it holds that $[V^{(k)}]_{00} - [Z^{(k)}]_0 = \lambda^{(k)}(\hat{P}_0 - C^{(k)})$ for a certain $C^{(k)} \succeq 0$.

The latter together with $[\lambda^{(k)}T(\hat{P}) + T(Z^{(k)}) - V^{(k)}]_{lh} = 0$ for $(l, h) \neq (0, 0)$ leads to

$$(\lambda^{(k)})^{-1}(V^{(k)} - T(Z^{(k)})) = \begin{bmatrix} \hat{P}_0 - C^{(k)} & \mathcal{K} \\ \mathcal{K}^\top & \mathcal{T} \end{bmatrix}$$

with $\mathcal{K} := [\hat{P}_1 \dots \hat{P}_n]$ and $\mathcal{T} := T([\hat{P}_0 \dots \hat{P}_{n-1}])$. Therefore, we get that

$$(\lambda^{(k)})^{-1}(I + V^{(k)} - T(Z^{(k)})) = \begin{bmatrix} \hat{P}_0 - C^{(k)} & \mathcal{K} \\ \mathcal{K}^\top & \mathcal{T} \end{bmatrix} + \mathcal{O}\left(\frac{1}{\lambda^{(k)}}\right).$$

By using the Schur complement, we get $\hat{P}_0 - C^{(k)} - \mathcal{K}\mathcal{T}^{-1}\mathcal{K}^\top + \mathcal{O}\left(\frac{1}{\lambda^{(k)}}\right) \succeq 0$, thus $C^{(k)} \preceq C_{max}^{(k)}$ with $C_{max}^{(k)} := \hat{P}_0 - \mathcal{K}\mathcal{T}^{-1}\mathcal{K}^\top + \mathcal{O}\left(\frac{1}{\lambda^{(k)}}\right)$. Consequently, $\hat{P}_0 + (\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) \preceq C_{max}^{(k)}$ and ,

$$\begin{aligned} J^{(k)} &:= J(\lambda^{(k)}, V^{(k)}, Z^{(k)}) = \lambda^{(k)} \left(-\log \left| \hat{P}_0 + (\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) \right| - \int \log |\hat{\Phi}| + \delta \right) \\ &\geq \lambda^{(k)} \left(-\log |C_{max}^{(k)}| - \int \log |\hat{\Phi}| + \delta \right) = \lambda^{(k)} \left(-\log |C_{max}^{(k)}| + \int \log |\hat{\Phi}^{-1}| + \delta \right). \end{aligned}$$

Notice that $\hat{\Phi}^{-1} = \sum_{k=-\infty}^{\infty} \hat{P}_k e^{-i\theta k}$ is the power spectral density of an AR process of order n , then by using [28, Lemma A.1], we get

$$\begin{aligned} J^{(k)} &\geq \lambda^{(k)} \left(-\log |C_{max}^{(k)}| + \log |\hat{P}_0 - \mathcal{K}\mathcal{T}^{-1}\mathcal{K}^\top| + \delta \right) \\ &= \lambda^{(k)} \left(\delta + \mathcal{O}\left(\frac{1}{\lambda^{(k)}}\right) \right) \rightarrow +\infty \text{ as } k \rightarrow \infty. \end{aligned}$$

Thus $(\lambda^{(k)}, V^{(k)}, Z^{(k)})$ cannot be an infimizing sequence. ■

It follows from Proposition 3.2.3 that there exists $\beta \in \mathbb{R}$ with $|\beta| < \infty$ such that $T(Z) - V \succeq \beta I$, and $0 < \gamma < \infty$ such that $\lambda \leq \gamma$. Therefore, the set \mathcal{C}_1 can be further restricted to the set

$$\begin{aligned} \mathcal{C}_2 &:= \left\{ (\lambda, V, Z) : V \in \mathbb{Q}_{m(n+1)}, V \succeq 0, Z \in \mathcal{O}, \lambda \in \mathbb{R}, \gamma \geq \lambda \geq \varepsilon, \right. \\ &\quad \beta I \preceq T(Z) - V \preceq I, [\lambda \hat{P}_0 + Z_0 - V_{00}] \succ 0, \\ &\quad \left. [\lambda(T(\hat{P})) + T(Z) - V]_{lh} = 0 \forall (l, h) \neq (0, 0) \right\}. \end{aligned}$$

In addition, we prove that it is not possible for V and Z to diverge while keeping the difference $T(Z) - V$ finite. Accordingly, we can further restrict the search for the optimal solution to a subset \mathcal{C}_3 in which neither V nor Z can diverge:

Proposition 3.2.4. *Let $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{N}}$ be a sequence of elements in \mathcal{C}_2 such that*

$$\lim_{k \rightarrow \infty} \|V^{(k)}\| = +\infty \quad (3.37)$$

or

$$\lim_{k \rightarrow \infty} \|Z^{(k)}\| = +\infty \quad (3.38)$$

or both. Then, such a sequence cannot be an infimizing sequence.

Proof. Consider a sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{N}}$ in \mathcal{C}_2 . We first show that $[Z^{(k)}]_0$ cannot diverge. Indeed, assume by contradiction that $\lim_{k \rightarrow \infty} \| [Z^{(k)}]_0 \| = +\infty$. Since it is a symmetric and traceless matrix, this implies

$$\lim_{k \rightarrow \infty} \min_{\alpha^{(k)} \in \sigma([Z^{(k)}]_0)} \alpha^{(k)} = -\infty.$$

Therefore, since $\lambda^{(k)} \hat{P}_0$ is bounded and $V^{(k)}$ positive semidefinite $\forall k$, then $(\lambda^{(k)} \hat{P}_0 + [Z^{(k)}]_0 - [V^{(k)}]_{00})$ has at least a negative eigenvalue for k sufficiently large, so that the sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})$ is not in \mathcal{C}_2 . We conclude that $\lim_{k \rightarrow \infty} \| [Z^{(k)}]_0 \| < \infty$. As a consequence, since $\beta I \preceq T(Z^{(k)}) - V^{(k)} \preceq I$ (which is one of the condition for the sequence to be in \mathcal{C}_2), and $[T(Z^{(k)})]_{hh} = [Z^{(k)}]_0$ by construction, it holds that $\forall k \| [V^{(k)}]_{hh} \| < \infty$, for $h = 0, \dots, n$. Then, from $V^{(k)} \succeq 0$ it follows that also the off-diagonal blocks of $V^{(k)}$ must be bounded $\forall k$, i.e.

$$\| [V^{(k)}]_{hl} \| < \infty, \quad l \neq h, \quad l, h = 0, \dots, n. \quad (3.39)$$

Finally, by boundedness of $(T(Z^{(k)}) - V^{(k)})$ and by (3.39) we obtain that $\| [Z^{(k)}]_h \| < \infty$, for $h = 1, \dots, n$, concluding the proof. \blacksquare

Thus, the minimization over \mathcal{C}_2 is equivalent to the minimization over the subset

$$\mathcal{C}_3 := \left\{ (\lambda, V, Z) : V \in \mathbb{Q}_{m(n+1)}, \alpha I \succeq V \succeq 0, Z \in \mathcal{O}, \beta I \preceq T(Z) - V \preceq I, \lambda \in \mathbb{R}, \right. \\ \left. \gamma \geq \lambda \geq \varepsilon, [\lambda \hat{P}_0 + Z_0 - V_{00}] \succ 0, [\lambda(T(\hat{P})) + T(Z) - V]_{lh} = 0 \forall (l, h) \neq (0, 0) \right\}$$

for a certain $\alpha > 0$.

Finally, we consider a sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k \in \mathbb{Z}} \in \mathcal{C}_3$ such that $[(\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{P}_0]$ tends to be singular as $k \rightarrow \infty$. This implies that $|(\lambda^{(k)})^{-1} ([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{P}_0|$ tends to zero and hence $J \rightarrow +\infty$. Thus, such a sequence cannot be an

infimizing sequence. Therefore, the final set \mathcal{C}_C is

$$\mathcal{C}_C := \{(\lambda, V, Z) : V \in \mathbb{Q}_{m(n+1)}, \alpha I \succeq V \succeq 0, Z \in \mathcal{O}, \beta I \preceq T(Z) - V \preceq I, \lambda \in \mathbb{R}, \\ \gamma \geq \lambda \geq \varepsilon, [\lambda \hat{P}_0 + Z_0 - V_{00}] \succeq \mu I, [\lambda(T(\hat{P})) + T(Z) - V]_{lh} = 0 \forall (l, h) \neq (0, 0)\}$$

where $\alpha, \beta, \gamma, \varepsilon$ and μ such that $|\alpha|, |\beta|, |\gamma|, |\varepsilon|$ and $|\mu| < +\infty$.

Theorem 3.2.5. *Problem (3.22) is equivalent to*

$$\min_{(\lambda, V, Z) \in \mathcal{C}_C} J(\lambda, V, Z)$$

and it admits solution.

Proof. Equivalence of the two problems has already been proven by the previous arguments. Since \mathcal{C}_C is closed and bounded, hence compact, and J is continuous over \mathcal{C}_C , by Weierstrass' Theorem the minimum exists. \blacksquare

3.3 Solution to the primal problem

In this section, after proving that the primal problem (3.6) and its matrix reformulation (3.15) are equivalent, we show how to recover the solution to the primal problem.

Let $(\lambda^\circ, V^\circ, Z^\circ)$ be a solution to (3.22) and (X°, L°) be the corresponding solution to (3.15). Since X_{00}° is positive definite, $\log |X_{00}^\circ|$ is finite. By Lemma 3.2.1, at the optimum $\int \log |\Phi|$ must be finite as well; this implies that $\Phi(e^{i\vartheta})$, with $\vartheta \in [-\pi, \pi]$, may be singular at most on a set of zero measure, or, in other terms, $\Delta X^\circ \Delta^* \succ 0$ a.e.. This observation leads to the following proposition:

Proposition 3.3.1. *Let (X°, L°) be a solution to (3.15). Then $\Delta X^\circ \Delta^* \succ 0$ a.e.. Accordingly, (3.6) and (3.15) are equivalent.*

Now we are ready to show how to recover the solution to the primal problem; to this aim we need the following result, see [111].

Lemma 3.3.2. *Let $Z \in \mathbf{M}_{m,n}$ and $W \in \mathbb{Q}_m$. If $W \succ 0$ is such that*

$$T(Z) \succeq \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \tag{3.40}$$

then $T(Z) \succ 0$.

Exploiting the constraints $[\lambda(T(\hat{P})) + T(Z) - V]_{lh} = 0, \forall (l, h) \neq (0, 0)$ and $[\lambda\hat{P}_0 + Z_0 - V_{00}] \succ 0$, it is not difficult to see that

$$V^\circ = \lambda^\circ T(\hat{P}) + T(Z^\circ) - \begin{bmatrix} W^\circ & 0 \\ 0 & 0 \end{bmatrix} \quad (3.41)$$

where

$$W^\circ := Z_0^\circ - V_{00}^\circ + \lambda^\circ \hat{P}_0 \succ 0. \quad (3.42)$$

Since $V^\circ \succeq 0$ and in view of Lemma 3.3.2, $\lambda^\circ T(\hat{P}) + T(Z^\circ) \succ 0$. Hence, V° has rank at least equal to mn .

Since the duality gap between (3.15) and (3.22) is equal to zero, we have that $\langle V^\circ, X^\circ - L^\circ \rangle = 0$, which in turn implies

$$V^\circ(X^\circ - L^\circ) = 0 \quad (3.43)$$

because $V^\circ, X^\circ - L^\circ \succeq 0$. Recalling that $\text{rank}(V^\circ) \geq mn$, in view of (3.43) the matrix $X^\circ - L^\circ$ has rank at most equal to m . Let $\text{rank}(X^\circ - L^\circ) = \tilde{m} \leq m$. Then, there exists a full-row rank matrix $A \in \mathbb{R}^{\tilde{m} \times m(n+1)}$ such that

$$X^\circ - L^\circ = A^\top A. \quad (3.44)$$

By (3.43), it follows that $V^\circ A^\top = 0$. Let $Y_D := [v_o \ v_1 \ \dots \ v_l] \in \mathbb{R}^{m(n+1) \times l}$ denote the matrix whose columns form a basis of $\ker(V^\circ)$. Note that the dimension l of the null space of V° is at least \tilde{m} because $\text{Im}(A^\top) \subseteq \ker(V^\circ)$ and $\text{rank}(A^\top) = \tilde{m}$; also $l \leq m$ because $\text{rank}(V^\circ) \geq mn$. Rewriting the matrix A^\top as $A^\top = Y_D S$ with $S \in \mathbb{R}^{l \times \tilde{m}}$, from (3.44) we obtain

$$X^\circ - L^\circ = Y_D Q_D Y_D^\top, \quad (3.45)$$

with $Q_D := S S^\top \in \mathbb{Q}_l$ unknown.

In a similar fashion, by the zero duality gap between (3.15) and (3.22), the *complementary slackness condition* for the multiplier associated to the positive semi-definiteness of L reads as $\langle U^\circ, L^\circ \rangle = 0$, which in turn implies $U^\circ L^\circ = 0$. Repeating the same reasoning as before, it can be seen that, if the dimension of the null space of U° is \tilde{r} with $\tilde{r} \geq r$ and $Y_L := [u_o \ u_1 \ \dots \ u_{\tilde{r}}] \in \mathbb{R}^{m(n+1) \times \tilde{r}}$ is a matrix whose columns form a basis of $\ker(U^\circ)$,

then L° can be written as

$$L^\circ = Y_L Q_L Y_L^\top \quad (3.46)$$

with $Q_L \in \mathbb{Q}_{\tilde{r}}$ unknown. Plugging (3.46) into (3.45), we then obtain

$$X^\circ - Y_L Q_L Y_L^\top = Y_D Q_D Y_D^\top. \quad (3.47)$$

Assume now that each block of $X^\circ - L^\circ$ is diagonal, namely

$$\text{ofd}\left(\left[Y_D Q_D Y_D^\top\right]_{hk}\right) = 0 \quad h, k = 0, \dots, n. \quad (3.48)$$

Remark 3.3.3. We can make the previous assumption without loss of generality. Indeed, let $(\Phi^\circ, \Phi_L^\circ)$ be the solution to Problem (3.6) and $\Phi_D^\circ = \Phi^\circ - \Phi_L^\circ$; X , L and $D = X - L$ are any matrices in $\mathbb{Q}_{m(n+1)}$ such that $\Phi^\circ = \Delta X \Delta^*$, $\Phi_L^\circ = \Delta L \Delta^*$ and $\Phi_D^\circ = \Delta D \Delta^*$. We can always consider a different matrix parameterization $(\tilde{X}, \tilde{L}, \tilde{D})$ for Φ° , Φ_L° and Φ_D° as follows. First notice that there always exists a matrix \tilde{D} with all diagonal blocks such that $\Phi_D^\circ = \Delta \tilde{D} \Delta^*$; in other words, we can always find $\delta D \in \mathbb{Q}_{m(n+1)}$ such that $\Delta \delta D \Delta^* = 0$ and $\tilde{D} := D + \delta D$ satisfies $\text{ofd}\left(\left[\tilde{D}\right]_{hk}\right) = 0$ for $h, k = 0, \dots, n$. Now, let $\delta X \in \mathbb{Q}_{m(n+1)}$ such that $\Delta \delta X \Delta^* = 0$ and $\tilde{X} := X + \delta X$ satisfies (3.19). Define $\tilde{L} = \tilde{X} - \tilde{D} = X - D + \delta L$ with $\delta L := \delta X - \delta D$. It is easy to see that $\Phi^\circ = \Delta \tilde{X} \Delta^*$ and $\hat{\Phi}_L = \Delta \tilde{L} \Delta^*$. This means that (\tilde{X}, \tilde{L}) is still a solution to Problem (3.15) and it allows us to restrict to solutions of (3.15) for which (3.48) holds.

By applying the ofd operator to both sides of (3.47) and exploiting the assumption (3.48), it is not difficult to obtain

$$\text{ofd}\left(\left[Y_L Q_L Y_L^\top\right]_{00}\right) = \text{ofd}(X_{00}^\circ) \quad (3.49)$$

which is a system of $m(m-1)/2$ linear equations in the $\tilde{r}(\tilde{r}+1)/2$ unknowns Q_L . Notice that X_{00} is given by (3.19). Finally, once L° is computed, in order to retrieve Q_D we exploit (3.48) and the following system of $m(m+1)/2$ linear equations:

$$\left[Y_D Q_D Y_D^\top\right]_{00} = X_{00}^\circ - L_{00}^\circ. \quad (3.50)$$

Since both the dual and the primal problem admit solution, the resulting systems of equations (3.48), (3.49) and (3.50) do admit solutions.

3.4 The proposed algorithm

In this section we propose an algorithm to solve numerically the dual problem.

To start with, as observed in Section 3.3, we rewrite (3.22) in a different fashion by getting rid of the slack variable $V \in \mathbb{Q}_{m(n+1)}$. This is done by introducing a new variable $W \in \mathbb{Q}_m$ defined, similarly to (3.42), as $W := Z_0 - V_{00} + \lambda \hat{P}_0 \succ 0$ such that, as in (3.41), the variable V can be expressed as

$$V = \lambda T(\hat{P}) + T(Z) - \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix}. \quad (3.51)$$

Accordingly, the dual problem (3.22) can be expressed in terms of the variables λ , W and Z as follows:

$$\min_{(\lambda, W, Z) \in \mathcal{C}} J \quad (3.52)$$

where $J := \lambda \left(-\log |\lambda^{-1} W| - f \log |\hat{\Phi}| + \delta \right)$ and the corresponding feasible set \mathcal{C} is:

$$\mathcal{C} := \left\{ (\lambda, W, Z) : W \in \mathbb{Q}_m, W \succ 0, Z \in \mathcal{O}, \lambda \in \mathbb{R}, \lambda > 0, \right. \\ \left. \lambda T(\hat{P}) + T(Z) - \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \succeq 0, I + \lambda T(\hat{P}) - \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \succeq 0 \right\}.$$

We can further simplify our problem as follows. First, we observe that the constraint

$$V = \lambda T(\hat{P}) + T(Z) - \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \succeq 0 \quad (3.53)$$

can be rewritten as

$$\lambda T(\hat{P}) + T(Z) \succeq \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix}$$

and then, by Lemma 3.3.2, $\lambda T(\hat{P}) + T(Z) \succ 0$. Now, we can easily rewrite (3.53) recalling the characterization of a symmetric positive semidefinite matrix using the Schur complement, see [132]. To this aim, it is convenient to introduce the linear operators $T_{0,0} : \mathbf{M}_{m,n} \rightarrow \mathbb{Q}_m$, $T_{0,1:n} : \mathbf{M}_{m,n} \rightarrow \mathbf{M}_{m,n-1}$ and $T_{1:n,1:n} : \mathbf{M}_{m,n} \rightarrow \mathbb{Q}_{mn}$ that, for a given matrix $H \in \mathbf{M}_{m,n}$ construct a symmetric block-Toeplitz matrix and extract the blocks in position $(0,0)$, $(0,1:n)$ and $(1:n,1:n)$, respectively. With this notation, we

have

$$T(Z + \lambda\hat{P}) = \begin{bmatrix} T_{0,0}(Z + \lambda\hat{P}) & T_{0,1:n}(Z + \lambda\hat{P}) \\ T_{0,1:n}^\top(Z + \lambda\hat{P}) & T_{1:n,1:n}(Z + \lambda\hat{P}) \end{bmatrix}$$

and the constraint (3.53) is equivalent to $T_{1:n,1:n}(Z + \lambda\hat{P}) \succ 0$ and $W \preceq Q(\lambda, Z)$ with

$$Q(\lambda, Z) := T_{0,0}(Z + \lambda\hat{P}) - T_{0,1:n}(Z + \lambda\hat{P})T_{1:n,1:n}^{-1}(Z + \lambda\hat{P})T_{0,1:n}^\top(Z + \lambda\hat{P}).$$

In a similar fashion, the last matricial inequality constraint in \mathcal{C} can be equivalently expressed as $W \preceq R(\lambda)$ where

$$R(\lambda) := I + T_{0,0}(\lambda\hat{P}) - T_{0,1:n}(\lambda\hat{P})\left(I + T_{1:n,1:n}(\lambda\hat{P})\right)^{-1}T_{0,1:n}^\top(\lambda\hat{P}).$$

Therefore, Problem (3.22) can be formulated as

$$\min_{(\lambda, W, Z) \in \mathcal{C}} J = \lambda \left(-\log |\lambda^{-1}W| - \int \log |\hat{\Phi}| + \delta \right) \quad (3.54)$$

where

$$\mathcal{C} := \left\{ (\lambda, W, Z) : Z \in \mathcal{O}, \lambda \in \mathbb{R}, \lambda > 0, T_{1:n,1:n}(Z + \lambda\hat{P}) \succ 0, \right. \\ \left. W \in \mathbb{Q}_m, W \succ 0, W \preceq Q(\lambda, Z), W \preceq R(\lambda) \right\}.$$

Remark 3.4.1. Before solving this problem, notice that J in (3.54) is jointly convex in (λ, W) , and at each feasible point (λ_0, W_0) (where $\lambda_0 \neq 0$) it is strictly convex in all directions except for the direction equal to the point itself, i.e. (λ_0, W_0) . Thus, if $(\lambda_{\text{opt}}, W_{\text{opt}})$ is an optimal value for λ and W , all the other optimal values must lie on the non-strictly convex direction. As a consequence there exist $\alpha_{\min} \leq 0$ and $\alpha_{\max} \geq 0$ such that $((1 + \alpha)\lambda_{\text{opt}}, (1 + \alpha)W_{\text{opt}})$ are optimal values for (λ, W) for any $\alpha \in [\alpha_{\min}, \alpha_{\max}]$. Moreover, if $\alpha_{\min} = \alpha_{\max} = 0$ then the pair $(\lambda_{\text{opt}}, W_{\text{opt}})$ that, together with a certain Z , solves (3.54) is unique. This is indeed the case. In fact, at the optimum J is strictly negative so that at the optimum the derivative of J along the only non-strictly convex direction $(\lambda_{\text{opt}}, W_{\text{opt}})$ is not equal to zero. In other words, at the optimum and along the only non-strictly convex direction J is not constant which implies $\alpha_{\min} = \alpha_{\max} = 0$.

The uniqueness of Z is a much more problematic issue. Indeed, we have observed in simulations that in some cases it may happen that there are different optimal values of Z . The corresponding number of identified latent factors, however, is not affected and the predictive powers of the identified models are essentially the same.

Solving Problem (3.54) simultaneously for λ , W , and Z is not trivial because the inequality constraints $W \preceq Q(\lambda, Z)$ and $W \preceq R(\lambda)$ both depend on λ . On the other hand, once we fix the dual variable λ to a positive constant $\bar{\lambda} > 0$, the problem:

$$\min_{(W,Z) \in \mathcal{C}_{\bar{\lambda}}} J(\bar{\lambda}, W, Z) \quad (3.55)$$

with

$$\mathcal{C}_{\bar{\lambda}} = \{(W, Z) : Z \in \mathcal{O}, W \in \mathbb{Q}_m, T_{1:n,1:n}(Z + \bar{\lambda}\hat{P}) \succ 0, W \succ 0, W \preceq Q(\bar{\lambda}, Z), W \preceq R(\bar{\lambda})\}$$

can be efficiently solved by resorting to the ADMM algorithm, [22]. To this aim, we rewrite Problem (3.55) by introducing a new variable $Y \in \mathbb{Q}_m$ defined as $Y = Q(\bar{\lambda}, Z) - W$:

$$\begin{aligned} \min_{\substack{(W,Z) \in \mathcal{C}_{W,Z} \\ Y \in \mathbb{Q}_m^+}} J &= \bar{\lambda} \left(-\log |\bar{\lambda}^{-1}W| - \int \log |\hat{\Phi}| + \delta \right) \\ \text{subject to } Y &= Q(\bar{\lambda}, Z) - W \end{aligned} \quad (3.56)$$

where $\mathcal{C}_{W,Z} := \{(W, Z) : Z \in \mathcal{O}, W \in \mathbb{Q}_m, W \succ 0, W \preceq R(\bar{\lambda}), T_{1:n,1:n}(Z + \bar{\lambda}\hat{P}) \succ 0\}$ and \mathbb{Q}_m^+ denotes the cone of symmetric positive semidefinite matrices of size $m \times m$. If $M \in \mathbb{Q}_m$ is the Lagrange multiplier, and $\rho > 0$ is the *penalty parameter*, the *augmented Lagrangian* for (3.56) is

$$\begin{aligned} \mathcal{L}_\rho(W, Z, Y, M) &:= \bar{\lambda} \left(-\log |\bar{\lambda}^{-1}W| - \int \log |\hat{\Phi}| + \delta \right) + \langle M, Y - Q(\bar{\lambda}, Z) + W \rangle \\ &\quad + \frac{\rho}{2} \|Y - Q(\bar{\lambda}, Z) + W\|^2 \end{aligned}$$

Accordingly, given the initial guesses $W^{(0)}$, $Z^{(0)}$, $Y^{(0)}$ and $M^{(0)}$, the ADMM updates are:

$$(W^{(k+1)}, Z^{(k+1)}) = \arg \min_{(W,Z) \in \mathcal{C}_{W,Z}} \mathcal{L}_\rho(W, Z, Y^{(k)}, M^{(k)}) \quad (3.57)$$

$$Y^{(k+1)} = \arg \min_{Y \in \mathbb{Q}_m^+} \mathcal{L}_\rho(W^{(k+1)}, Z^{(k+1)}, Y, M^{(k)}) \quad (3.58)$$

$$M^{(k+1)} = M^{(k)} + \rho \left(Y^{(k+1)} - Q(\bar{\lambda}, Z^{(k+1)}) + W^{(k+1)} \right).$$

Problem (3.57) does not admit a closed form solution, therefore we approximate the optimal solution by a gradient projection step

$$W^{(k+1)} = \Pi \left(W^{(k)} - t_k \nabla_W \mathcal{L}_\rho(W^{(k)}, Z^{(k)}, Y^{(k)}, M^{(k)}) \right)$$

$$Z^{(k+1)} = \Pi_{\mathcal{O}} \left(Z^{(k)} - t_k \nabla_Z \mathcal{L}_\rho(W^{(k)}, Z^{(k)}, Y^{(k)}, M^{(k)}) \right)$$

where:

- $\nabla_W \mathcal{L}_\rho(W, Z, Y, M)$ denotes the gradient of the augmented Lagrangian with respect to W , which is $\nabla_W \mathcal{L}_\rho = -\bar{\lambda}W^{-1} + M + \rho(Y - Q + W)$.
- $\nabla_Z \mathcal{L}_\rho(W, Z, Y, M)$ denotes the gradient of the augmented Lagrangian with respect to Z , $\nabla_Z \mathcal{L}_\rho = [I_m \quad -T_{0,1:n}T_{1:n,1:n}^{-T}]^\top$, where the omitted argument of the operators $T_{0,1:n}$ and $T_{1:n,1:n}$ is intended to be equal to $(Z + \bar{\lambda}\hat{P})$.
- $\Pi_{\mathcal{O}}$ denotes the projection operator onto \mathcal{O} , that is $\Pi_{\mathcal{O}}(A) = \text{ofd}_B(A)$.
- Π denotes the projection operator onto the convex cone $\{S \in \mathbb{Q}_m : S \preceq R(\bar{\lambda})\}$. It is not difficult to see that $\Pi(A) = R(\bar{\lambda}) - \Pi_+(R(\bar{\lambda}) - A)$, where Π_+ is the projection operator onto the cone \mathbb{Q}_m^+ .
- the step-size t_k is determined at each step k in an iterative fashion: we start by setting $t_k = 1$ and we decrease it progressively of a factor β , with $0 < \beta < 1$, until the conditions $W^{(k+1)} \succ 0$ and $T_{1:n,1:n}(Z^{(k+1)} + \bar{\lambda}\hat{P}) \succ 0$ are met and the Armijo's condition [21] is satisfied.

Problem (3.58) admits a closed form solution, which can be easily computed as:

$$Y^{(k+1)} = \Pi_+ \left(Q(\bar{\lambda}, Z^{(k+1)}) - W^{(k+1)} - \frac{1}{\rho} M^{(k)} \right).$$

To define the stopping criterion, we need to introduce the following quantities

$$R^P = Y - Q(\bar{\lambda}, Z^{(k+1)}) + W^{(k+1)}$$

$$R^D = D \left(\begin{bmatrix} I_m \\ -T_{1:n,1:n}^{-1} T_{0,1:n}^\top \end{bmatrix} \rho(Y^{(k+1)} - Y^{(k)}) \begin{bmatrix} I_m & -T_{0,1:n} T_{1:n,1:n}^{-1} \end{bmatrix} \right)$$

which are referred to as the primal and dual residual, respectively. Notice that the omitted argument of the operators $T_{0,1:n}$ and $T_{1:n,1:n}$ is intended to be equal to $(Z^{(k+1)} + \bar{\lambda}\hat{P})$. Then, the algorithm stops when the following conditions are met:

$$\|R^P\| \leq m\varepsilon^{\text{ABS}} + \varepsilon^{\text{REL}} \max \left\{ \|W^{(k)}\|, \|Q(\bar{\lambda}, Z^{(k)})\|, \|Y^{(k)}\| \right\}$$

$$\|R^D\| \leq m\sqrt{(n+1)}\varepsilon^{\text{ABS}} + \varepsilon^{\text{REL}} \left\| D \left(\begin{bmatrix} I_m \\ -T_{1:n,1:n}^{-1} T_{0,1:n}^\top \end{bmatrix} M^{(k)} \begin{bmatrix} I_m & -T_{0,1:n} T_{1:n,1:n}^{-1} \end{bmatrix} \right) \right\|$$

where ε^{ABS} and ε^{REL} are the desired absolute and relative tolerances.

It remains to determine the optimal value λ° for λ which solves Problem (3.54). To this aim, we exploit the following result, see [21, pp.87-88]:

Proposition 3.4.2. *If f is convex in (x, y) and \mathcal{C} is a convex non-empty set, then the function $g(x) = \inf_{y \in \mathcal{C}} f(x, y)$ is convex in x , provided that $g(x) > -\infty$ for some x . The domain of g is the projection of $\text{dom}(f)$ on its x -coordinates.*

This result guarantees that the function $g(\lambda) = \min_{(W, Z) \in \mathcal{C}_\lambda} J(\lambda, W, Z)$ is convex in λ . Hence, in order to determine $\lambda^\circ = \arg \min_{\lambda > 0} g(\lambda)$ we can choose an initial interval of uncertainty $[a, b]$ containing λ° , then we progressively reduce it by evaluating $g(\lambda)$ at two points within the interval placed symmetrically, each at distance $h > 0$ from the midpoint. This is repeated until the width of the uncertainty interval is smaller than a certain tolerance $l > 0$.

The overall procedure to solve the dual problem (3.54) is summarized in Algorithm 1.

Algorithm 1 Dual problem solution

Input: $b > a > 0, l > 0, h > 0$

Output: $(\lambda^\circ, W^\circ, Z^\circ)$

```

1: repeat
2:    $\tilde{a} = (a + b)/2 - h;$ 
3:    $\tilde{b} = (a + b)/2 + h.$ 
4:   Compute  $g(\tilde{a})$  by applying the ADMM with  $\lambda = \tilde{a}$ .
5:   Compute  $g(\tilde{b})$  by applying the ADMM with  $\lambda = \tilde{b}$ .
6:   if  $g(\tilde{a}) < g(\tilde{b})$  then
7:      $b = \tilde{b}$ 
8:   else
9:      $a = \tilde{a}$ 
10:  end if
11: until  $b - a < l$ 
12:  $\lambda^\circ = (a + b)/2.$ 
13: Compute  $(W^\circ, Z^\circ)$  by applying the ADMM with  $\lambda = \lambda^\circ.$ 

```

3.5 Numerical simulations

In this section, we test the efficiency of the proposed approach to estimate the number of factors through numerical simulations. We compare our method to the algorithm proposed by [68].

We consider a Monte Carlo study composed by 50 experiments, where for each experiment:

1. We build an MA factor model (3.1) of order $n = 2$, with m manifest variables and r latent factors, computed by randomly generating the zeros of the transfer functions $[W_L]_{(i,j)}$'s and $[W_D]_{(i,i)}$'s for $i = 1, \dots, m, j = 1, \dots, r$ within the circle with center at the origin and radius 0.95 on the complex plane. The model is generated in such a way that $\int \|\Phi_L(e^{i\theta})\| / \int \|\Phi_D(e^{i\theta})\| = 2$, that is the idiosyncratic component is not negligible with respect to the latent variable.
2. We generate from the model a sample y^N of length $N = 5000$.
3. We apply the proposed identification procedure to estimate the number of common factors. More precisely, we define

$$s_j := \int \frac{\sigma_j(\Phi_L^\circ(e^{i\theta}))}{\sigma_1(\Phi_L^\circ(e^{i\theta}))} \quad (3.59)$$

where $\sigma_j(\Phi_L^\circ(e^{i\theta}))$ denotes the j -th largest singular value of Φ_L° at frequency θ . It is clear that s_j represents the integral of the j -th largest normalized singular value of Φ_L° over the unit circle. In order to determine the rank of Φ_L° , we search for a “knee” in the bar plot of s_j . In mathematical terms, this is equivalent to determining the index i maximizing the ratio s_i/s_{i+1} . More in details, let i_{max} be the first i such that $s_{i+1} < 0.05$ and i_{min} the last i such that $s_i > 0.3$. Then, we define the “numerical rank” of Φ_L° as

$$r^\circ := \max_{i_{min} \leq i \leq i_{max}} s_i/s_{i+1}. \quad (3.60)$$

Here, the role of i_{max} and i_{min} is to exclude the normalized singular values which are extremely small and the ones which are sufficiently large.

4. We compute the number of factors from the data sequence y^N by applying the method proposed by Hallin and Liška, [68].
5. We assess the performance of the two estimators in terms of the mean absolute error

$$\bar{e} = \frac{1}{50} \sum_{i=1}^{50} |r - \hat{r}|$$

where \hat{r} is one of the two previous estimates and r is the true rank of the data generating process.

In all the simulations, the parameter δ is computed according to the empirical procedure of Section 3.1.1 with $\alpha = 0.5$. Then, Problem (3.54) is solved by applying

	Proposed method	Hallin and Liška method
m = 20, r=2	0.32	0.72
m = 30, r=2	0.12	0.22
m = 40, r=2	0.02	0.02
m = 20, r=4	1.1	2.98
m = 30, r=4	0.62	2.04
m = 40, r=4	0.32	1.64

Table 3.1: Mean absolute error between the estimated rank and the true rank r .

Algorithm 1 with $l = 7$ and $h = 3$. In regard to the ADMM algorithm, we set $\varepsilon^{\text{ABS}} = 10^{-4}$, $\varepsilon^{\text{REL}} = 10^{-4}$ and the penalty term $\rho = 0.05$.

Table I shows the mean absolute error \bar{e} when $r = 2, 4$ and $m = 20, 30, 40$. We see that in the six Monte Carlo studies the proposed method outperforms the Hallin and Liška's algorithm. In particular, the performances of the two estimators are comparable when the number m of observed variables is much higher than the number r of latent factors. The problem becomes more challenging when the ratio m/r decreases; improvements by our method are more significant in these situations.

Figures 3.1 and 3.2 plot the quantities s_j obtained by applying our estimation method in one of the previous Monte Carlo experiment with $m = 40$ manifest variables and $r = 2$ and $r = 4$ latent variables, respectively. We can notice that there is a knee point at $j = 2$ and at $j = 4$ in Figure 3.1 and Figure 3.2, respectively, so that we can recover the exact number of common factors in both cases.

The simulations results provide evidence of a good performance.

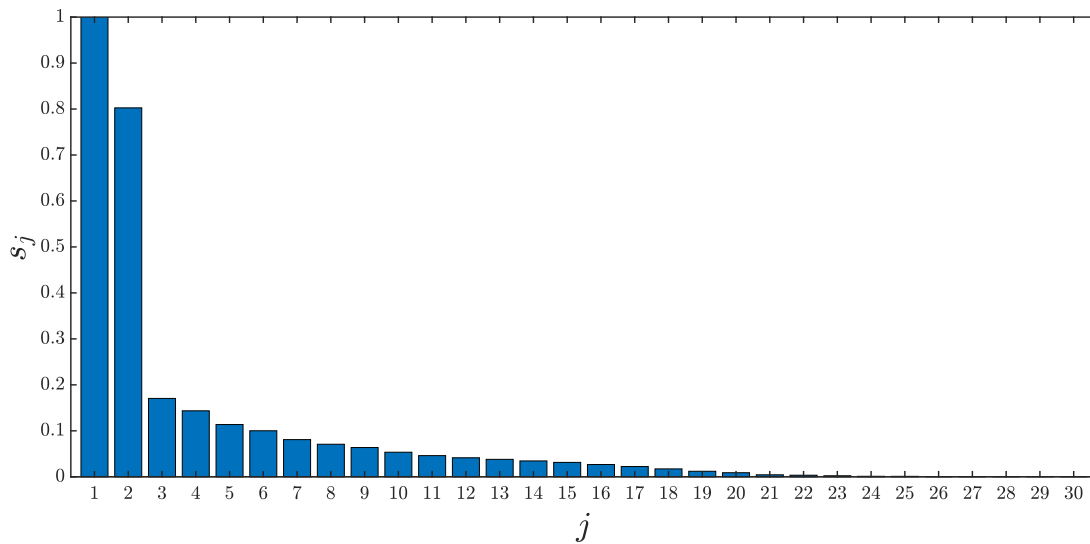


Figure 3.1: Estimated MA factor model with $n = 2$, $m = 40$, and $r = 2$. Integral over the unit circle of the first 30 normalized singular values of Φ_L° with $N = 5000$.

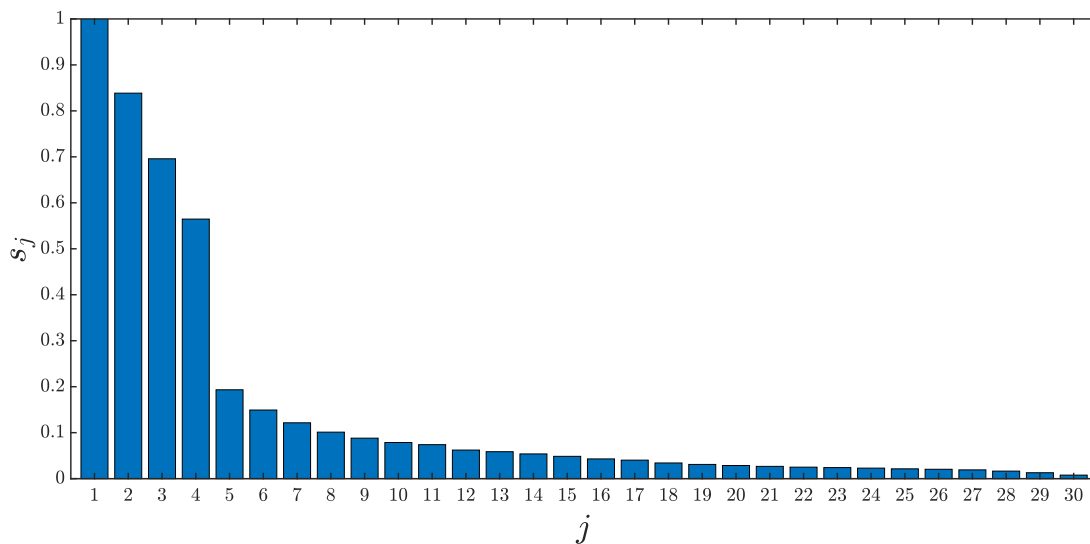


Figure 3.2: Estimated MA factor model with $n = 2$, $m = 40$, and $r = 4$. Integral over the unit circle of the first 30 normalized singular values of Φ_L° with $N = 5000$.

4

A robust approach to ARMA factor modeling

This chapter extends the robust approach to MA factor modeling presented in Chapter 3 to ARMA factor models. A 2-step procedure is proposed to robustly estimate the number of factors as well as to identify the parameters of the model: a step of *AR dynamics estimation* is followed by a *robust MA DFA* step. While this method hinges on a suboptimal procedure, it provides accurate estimation of the number of factors and of the dynamics of ARMA models: some simulation studies, both with synthetic and real data, indeed show good performances of the method, even with a limited number of data points.

The contribution of this chapter is twofold. Firstly, we propose a procedure to estimate the coefficients of an AR model by means of moment matching. We show that the estimated model has stability guarantees and coincides with the (unique) solution of a certain maximum entropy problem. Secondly, we derive an identification method to estimate the number of factors and the parameters of an ARMA factor model describing the observed data.

The results of the present chapter are published in

- F. Crescente, L. Falconi, F. Rozzi, *et al.*, “Learning AR factor models,” in *2020 59th IEEE Conference on Decision and Control (CDC)*, IEEE, 2020, pp. 274–279;
- L. Falconi, A. Ferrante, and M. Zorzi, “A Robust Approach to ARMA Factor Modeling,” *IEEE Transactions on Automatic Control*, 2023.

The rest of the chapter is outlined as follows. In Section 4.1 we introduce the DFA problem for ARMA models. Section 4.2 presents a 2-step algorithm to solve the problem; in particular, Subsection 4.2.1 proposes a moment matching method for the identification of the parameters of AR models. Finally Section 4.3 shows some numerical results with synthetic data and then with a real dataset extracted from a smart building.

4.1 Problem formulation

Consider the ARMA factor model:

$$y(t) = a^{-1}(W_L u(t) + W_D w(t)) \quad (4.1)$$

where

$$a(e^{i\vartheta}) = \sum_{k=0}^p a_k e^{-i\vartheta k}, \quad a_k \in \mathbb{R} \quad (4.2)$$

and

$$\begin{aligned} W_L(e^{i\vartheta}) &= \sum_{k=0}^n W_{L,k} e^{-i\vartheta k}, & W_{L,k} &\in \mathbb{R}^{m \times r} \\ W_D(e^{i\vartheta}) &= \sum_{k=0}^n W_{D,k} e^{-i\vartheta k}, & W_{D,k} &\in \mathbb{R}^{m \times m} \text{ diagonal.} \end{aligned}$$

The processes $u = \{u(t), t \in \mathbb{Z}\}$ and $w = \{w(t), t \in \mathbb{Z}\}$ are normalized white Gaussian noises of dimension r and m , respectively, such that $\mathbb{E}[u(t)w(s)^\top] = 0 \forall t, s$. It is assumed that the number r of latent factor is much smaller than the cross-sectional dimension m . Notice that

$$y_{MA}(t) := ay(t) = W_L u(t) + W_D w(t)$$

is a MA process of order n whose spectral density $\Phi = W_L W_L^* + W_D W_D^* \in \mathcal{Q}_{m,n}$ admits a low rank plus diagonal decomposition.

Assume to collect a realization $y^N = \{y(1) \dots y(N)\}$ of length N of the process y defined by (4.1) and suppose that the order p of the AR part and the order n of the MA part are known. We face the problem of estimating the parameters of the factor model (4.1) and the number of factors r .

Remark 4.1.1. It is not restrictive to assume that the autoregressive part in (4.1) is characterized by a scalar filter a ; indeed, any ARMA factor model can be written in the form of (4.1).

Remark 4.1.2. There is an identifiability issue in the problem. Indeed, if we multiply a , W_L and W_D by an arbitrary non-zero real number, the model remains the same. We can easily eliminate this uninteresting degree of freedom by normalizing the polynomial $a(z)$, so that from now on we assume $a_0 = 1$. In addition, for the identifiability of model (4.1), we assume that there are not zero/pole cancellations between the roots of the denominator $a(z)$ and the numerator matrices $W_L(z)$ and $W_D(z)$. In other words, we assume that (4.1) is a minimal representation of the ARMA model.

4.2 Problem solution

In this Section, we propose a 2-step procedure to solve the proposed ARMA DFA problem. The idea is to estimate first the parameters of the filter a , and then the spectral densities Φ_L and Φ_D by preprocessing y^N through a . In more details, the proposed solution consists of the following two steps:

1. The AR dynamic estimation:

Given the realization y^N , we estimate the p parameters of the filter a . To accomplish this task, we next propose a maximum likelihood estimator for AR models. In doing so, we are estimating an AR process whose spectral density is $a^{-1}(a^{-1})^* I_m$.

2. The MA dynamic factor analysis:

Let y_{MA}^N be the finite length trajectory obtained by passing through the estimated filter $a^\circ(e^{i\theta})$ the trajectory y^N with zero initial conditions. We apply the MA DFA procedure presented in Chapter 3 with the dataset y_{MA}^N . More in details, after computing the truncated periodogram $\hat{\Phi} \in \mathcal{Q}_{m,n}$ from y_{MA}^N , we solve the optimization problem (3.6) with $\hat{\Phi}$ in order to recover the number of latent factors and identify the MA parameters Φ° , Φ_L° and $\Phi - \Phi_L^\circ$.

Although the above procedure is suboptimal, the numerical simulations show that the resulting estimator of the number of factors performs well, see Section 4.3.1.

4.2.1 AR dynamics estimation

The first step is the AR dynamics estimation, for which we propose a moment matching method. Consider the AR process of order p

$$y(t) = a(z)^{-1}e(t) \quad (4.3)$$

where $e = \{e(t), t \in \mathbb{Z}\}$ is a m -dimensional normalized white Gaussian noise. The process $y(t)$ is obtained by stacking together the output of m identical (scalar) filters driven by independent (scalar) white noises. Therefore $y(t)$ may be viewed as a multivariate process with m independent channels $y_i(t)$ all of which feature the same probability description. This will be a key feature in what follows.

Given a finite-length realization $y^N = \{y(1) \dots y(N)\}$ of the AR process (4.3), the aim is to estimate the coefficients of the filter $a(z)$, namely $a = [a_1 \dots a_p]^\top \in \mathbb{R}^p$ (as we have fixed $a_0 = 1$). To address this problem we resort to the maximum-likelihood (ML)

principle and compute the estimate as

$$\mathbf{a}_{ML} := \arg \min_{\mathbf{a}} \ell(\mathbf{y}^N; \mathbf{a}) \quad (4.4)$$

where the negative log-likelihood $\ell(\mathbf{y}, \mathbf{a})$ is defined as

$$\ell(\mathbf{y}^N; \mathbf{a}) := -\log f\left(y(N), \dots, y(p+1) \mid y(p), \dots, y(1)\right).$$

In other words, we estimate the parameter vector \mathbf{a} in such a way that the model (4.3) maximizes the likelihood of producing the finite trajectory \mathbf{y}^N .

Scalar case: Firstly, we consider the scalar case $m = 1$. Since we are dealing with an AR model the solution can be obtained by standard arguments in closed form. In fact, by taking (4.2) into account, we can rewrite (4.3) as

$$\sum_{k=0}^p a_k y(t-k) = e(t),$$

so that

$$y(t) = -\sum_{k=1}^p a_k y(t-k) + y(t).$$

Therefore,

$$f(y(t) \mid y(t-1) \dots y(t-p)) \sim \mathcal{N}\left(-\begin{bmatrix} 0 & \mathbf{a}^\top \end{bmatrix} Y(t), 1\right) \quad (4.5)$$

and

$$f(y(N) \dots y(p+1) \mid y(p) \dots y(1)) = \prod_{t=p+1}^N f(y(t) \mid y(t-1) \dots y(t-p)) \quad (4.6)$$

where $Y(t) := \begin{bmatrix} y(t) & y(t-1) & \dots & y(t-p) \end{bmatrix}^\top$. Then, the negative log-likelihood (up to constant terms) results

$$\begin{aligned} \ell(\mathbf{y}^N; \mathbf{a}) &= \frac{1}{2} \sum_{t=p+1}^N \left(\begin{bmatrix} 1 & \mathbf{a}^\top \end{bmatrix} Y(t) \right)^2 = \frac{1}{2} \sum_{t=p+1}^N \text{tr} \left(\begin{bmatrix} 1 & \mathbf{a}^\top \end{bmatrix} Y(t) Y(t)^\top \begin{bmatrix} 1 \\ \mathbf{a} \end{bmatrix} \right) \\ &= \frac{1}{2} \begin{bmatrix} 1 & \mathbf{a}^\top \end{bmatrix} \left[\sum_{t=p+1}^N Y(t) Y(t)^\top \right] \begin{bmatrix} 1 \\ \mathbf{a} \end{bmatrix}. \end{aligned}$$

We now define the matrix

$$\hat{T} := \frac{1}{N-p} \sum_{t=p+1}^N Y(t)Y(t)^\top = \begin{bmatrix} \tilde{\tau}_0 & z^\top \\ z & \hat{T}_{22} \end{bmatrix}$$

where the partition is such that $\tilde{\tau}_0$ is a scalar and z is a column vector. In this way, we have

$$\ell(y^N; \mathbf{a}) = \frac{N-p}{2} \begin{bmatrix} 1 & \mathbf{a}^\top \end{bmatrix} \hat{T} \begin{bmatrix} 1 \\ \mathbf{a} \end{bmatrix} = \frac{1}{2} \left[\tilde{\tau}_0 + 2z^\top \mathbf{a} + \mathbf{a}^\top \hat{T}_{22} \mathbf{a} \right].$$

Since $\ell(y^N; \mathbf{a})$ is clearly convex in \mathbf{a} , Problem (4.4) is solved by annihilating the gradient of $\ell(y^N; \mathbf{a})$ with respect to \mathbf{a} , i.e. by imposing that

$$\frac{\partial \ell(y^N; \mathbf{a})}{\partial \mathbf{a}} = 2z^\top + 2\hat{T}_{22}\mathbf{a} = 0. \quad (4.7)$$

Since $\ell(y^N; \mathbf{a})$ is a quadratic form in \mathbf{a} , (4.7) provides a closed form formula which finally yields

$$\mathbf{a}_{ML} = -\hat{T}_{22}^{-1}z. \quad (4.8)$$

Multivariate case: Of course, the interesting case is the multivariate one, *i.e.* $m > 1$. To address this case, we recall that the m components $y_i(t)$ of the vector process $y(t)$ are independent scalar processes, *i.e.* $y_k(t) \perp y_l(s)$, $\forall k \neq l$, $\forall s, t$, and they all have the same probabilistic description, *i.e.* all the $y_k(t)$'s have the same spectral density

$$\Phi_{y_k}(z) = \frac{1}{a(z)a(z^{-1})}, \quad \forall i = 1, \dots, m. \quad (4.9)$$

Hence, the multivariate case can be addressed as that of a scalar process with m -times as many data. In fact, in view of the independence of the components of $y(t)$, the likelihood is

$$f(y(N) \dots y(p+1) | y(p) \dots y(1)) = \prod_{k=1}^m f(y_k(N) \dots y_k(p+1) | y_k(p) \dots y_k(1)).$$

Moreover, in view of (4.9), we can repeat for each k the argument that led to (4.6) to obtain an expression for $f(y_k(N) \dots y_k(p+1) | y_k(p) \dots y_k(1))$. This yields

$$f(y(N) \dots y(p+1) | y(p) \dots y(1)) = \prod_{k=1}^m \prod_{t=p+1}^N f(y_k(t) | y_k(t-1) \dots y_k(t-p)).$$

We can now repeat the previous computation and obtain

$$\ell(\mathbf{y}^N; \mathbf{a}) = \frac{1}{2} \sum_{k=1}^m \begin{bmatrix} 1 & \mathbf{a}^\top \end{bmatrix} \begin{bmatrix} \sum_{t=p+1}^N Y_k(t) Y_k(t)^\top \\ \mathbf{1} \\ \mathbf{a} \end{bmatrix}$$

where $Y_k(t) := \begin{bmatrix} y_k(t) & y_k(t-1) & \dots & y_k(t-p) \end{bmatrix}^\top$. We now define the matrix

$$\hat{T} := \frac{1}{m(N-p)} \sum_{k=1}^m \sum_{t=p+1}^N Y_k(t) Y_k(t)^\top = \begin{bmatrix} \tilde{\tau}_0 & z^\top \\ z & \hat{T}_{22} \end{bmatrix} \quad (4.10)$$

where, as for the scalar case, the partition is such that $\tilde{\tau}_0$ is a scalar and z is a column vector. In this way, we are exactly in the situation discussed for the scalar case and the solution is thus given again by (4.8) with T and z now provided by (4.10). Such a solution, however, is not guaranteed to correspond to a stable model (*i.e.* a model such that all the zeros of $a(z)$ are inside the unit circle). Notice that \hat{T} is an estimate of the Toeplitz matrix $T = \mathbb{E}[Y_k(t) Y_k(t)^\top] \succ 0$. Although $\hat{T} \rightarrow T$ almost surely as $N \rightarrow \infty$, $\hat{T} \succ 0$ is *not* Toeplitz for finite values of N . To address such an issue, we consider the biased estimate

$$\hat{T}_b = \begin{bmatrix} \tau_0 & \tau_1 & \dots & \tau_p \\ \tau_1 & \tau_0 & \ddots & \\ & \ddots & \ddots & \tau_1 \\ \tau_p & & \tau_1 & \tau_0 \end{bmatrix} = \begin{bmatrix} \tau_0 & z_b^\top \\ z_b & \hat{T}_{b,22} \end{bmatrix} \quad (4.11)$$

where

$$\tau_l = \frac{1}{mN} \sum_{k=1}^m \sum_{t=k+1}^N y_k(t) y_k(t-l), \quad l = 0 \dots p.$$

It is not difficult to see that $\hat{T}_b \succ 0$ generically. Accordingly, we can choose as estimate of \mathbf{a} :

$$\mathbf{a}_{ME} = -\hat{T}_{b,22}^{-1} z_b. \quad (4.12)$$

It is worth noting that (4.12) is the solution to a Yule-Walker equation, see [116]. Accordingly, $a_{ME}(z) = 1 + [z^{-1} \dots z^{-p}] \mathbf{a}_{ME}$ is a stable polynomial. Hence, the estimated spectral density of each $\bar{y}_k(t)$ is $\Phi_{ME}(z) = (a_{ME}(z) a_{ME}(z^{-1}))^{-1}$.

Proposition 4.2.1. Let $\check{\tau}_l = \tau_l v^\top \hat{T}_b^{-1} v$ with $v = [1 \ 0 \ \dots \ 0]^\top$. Then, Φ_{ME} is the unique solution to the following maximum entropy problem:

$$\begin{aligned} \Phi_{ME} = \operatorname{argmax}_{\Phi} \int_{-\pi}^{\pi} \log \det \Phi(e^{j\vartheta}) d\vartheta \\ \text{s.t. } \int_{-\pi}^{\pi} e^{j\vartheta l} \Phi(e^{j\vartheta}) d\vartheta = \check{\tau}_l, \quad l = 0 \dots p. \end{aligned} \quad (4.13)$$

Proof. Let $\check{T}_b := v^\top \hat{T}_b^{-1} v \hat{T}_b$. It is well known (see for example [116]) that the solution to (4.13) is $\Phi(z) = \sigma^2 (a(z)a(z^{-1}))^{-1}$ with $a(z) = 1 + [z^{-1} \dots z^{-p}]a$ such that

$$\check{T}_b \begin{bmatrix} 1 \\ a \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \end{bmatrix}. \quad (4.14)$$

Notice that in (4.14) is a system of $p+1$ equations. Consider the subsystem composed by the second equation up to the last equation: since \check{T}_b is invertible, its solution is (4.12). It remains to show that $\sigma^2 = 1$. Substituting (4.12) in the first equation, we have

$$\sigma^2 = v^\top \check{T}_b \begin{bmatrix} 1 & a_{ME}^\top \end{bmatrix}^\top = (\tau_0 + z_b^\top a_{ME}) (v^\top \hat{T}_b^{-1} v) = (\tau_0 - z_b^\top \hat{T}_{b,22}^{-1} z_b) (v^\top \hat{T}_b^{-1} v) = 1$$

where the last equality is due by the fact that $\tau_0 - z_b^\top \hat{T}_{b,22}^{-1} z_b$ is the Schur complement of the block $\hat{T}_{b,22}$ of \hat{T}_b . \blacksquare

Algorithm 2 summarizes the AR estimation procedure. It is clear from Proposition 4.2.1 that the resulting AR estimator matches the rescaled moments of y^N .

Algorithm 2 AR dynamics Estimation

Input: p, y^N

Output: a_{ME}

- 1: Let $Y_k(t) := [y_k(t) \ \dots \ y_k(t-p)]^\top$, $k = 1 \dots m$.
 - 2: Compute \hat{T}_b and thus $\hat{T}_{b,22}$, z_b from $Y_k(t)$
 - 3: Compute $a_{ME} = -\hat{T}_{b,22}^{-1} z_b$
-

4.3 Numerical simulations

In this section, we test the performance of the proposed 2-step procedure for ARMA DFA both to synthetic and real data. As regards the second step of the procedure, *i.e.* the MA DFA step, in all the simulations we apply the robust identification procedure presented in Chapter 3 with the following parameters: the parameter δ is computed according to

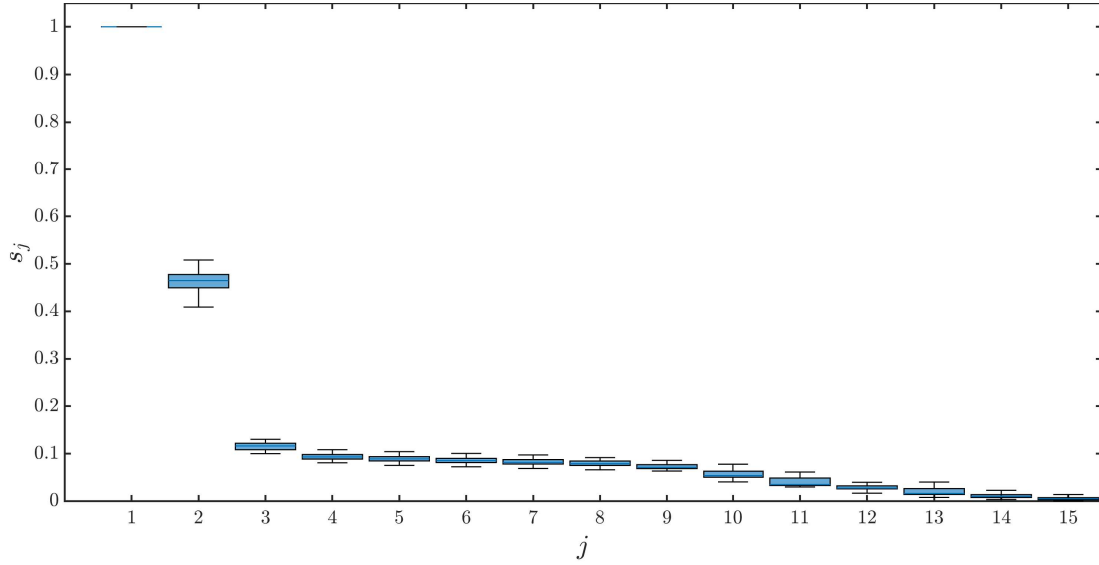


Figure 4.1: Estimated ARMA factor model with $m = 40$, $r = 2$, $n = 2$ and $p = 2$. Box-plot of the integral over the unit circle of the first 15 normalized singular values of Φ_L° with $N = 5000$.

the empirical procedure of Section 3.1.1 with $\alpha = 0.5$. Then, Problem (3.54) is solved by applying Algorithm 1 with $l = 7$ and $h = 3$. In regard to the ADMM algorithm, we set $\varepsilon^{\text{ABS}} = 10^{-4}$, $\varepsilon^{\text{REL}} = 10^{-4}$ and the penalty term $\rho = 0.05$.

4.3.1 Synthetic example

We consider a Monte Carlo simulation study composed of 50 experiments as follows. We randomly build an ARMA factor model (4.1) with $m = 40$, $r = 2$, $n = 2$ and $p = 2$; without loss of generality we fix $a_0 = 1$. Then, for each Monte Carlo experiment:

1. We randomly generate a data sequence of length $N = 5000$ from the ARMA model;
2. We perform the 2-step ARMA factor model identification procedure proposed in Section 4.2.
3. Given the estimated Φ_L° , we compute the integral of the normalized singular values s_j defined in (3.59) for $j = 1, \dots, m$.

The boxplot of the quantities s_j for the estimated Φ_L° 's are shown in Figure 4.1 and it reveals that the proposed identification procedure is able to successfully recover the number of latent factors.

4.3.2 Smart building dataset

The SMLsystem is a house built in Valencia at the Universidad CEU Cardenal Herrera (CEU-UCH). It is a modular house that integrates a whole range of different technologies to improve energy efficiency, with the objective to construct a near zero-energy house. A complex monitoring system has been used in the SMLsystem: it has indoor sensors for temperature, humidity and carbon dioxide; outdoor sensors are also available for lighting measurements, wind speed, rain, sun irradiance and temperature. We refer the reader to [130] for a detailed description of the building and its monitoring system. Two datasets from the SMLsystem are available for download at the UCI Machine Learning repository <http://archive.ics.uci.edu/ml>. We take into account $m = 17$ sensor signals extracted from these datasets: the indoor temperature (in °C) of the dining room and of the bedroom, the weather forecast temperature (in °C), the carbon dioxide (in ppm) in the dining room and in the bedroom, the relative humidity (in %) in the dining room and the bedroom, the lighting in the dining room and the bedroom (in lx), the sun dusk, the wind (in cm/sec), the sun light (in klx) in the west, east and south facade, the sun irradiance (in dW), the outdoor temperature (in °C) and finally the outdoor relative humidity (in %). The data are sampled with a period of $T = 15\text{min}$ and each sample is the mean of the last 15 minutes, reducing in this way the signal noise. The first dataset $y^{N_1} = \{y(1), \dots, y(N_1)\}$ was captured during March 2011 and has $N_1 = 2764$ points (≈ 28 days), while the second dataset $y^{N_2} = \{y(N_1 + 1), \dots, y(N_1 + N_2)\}$ has $N_2 = 1373$ points (≈ 14 days) collected in June 2011.

It is reasonable to expect that the variability of the considered signals may be successfully explained by a smaller number of factors. Motivated by this reason, we apply the ARMA factor model identification procedure with parameters $n = 2$ and $p = 2$ using the realization y^{N_1} . As shown in Figure 4.2, we obtain an estimate of 4 latent factors. For the sake of comparison, we also use the Matlab function `armax()` of the System Identification Toolbox to compute the prediction-error method (PEM) estimate for an ARMA model with polynomials $A(z)$ and $C(z)$, diagonal, of order 2 from the realization y^{N_1} . It is well known that the PEM estimate has guarantees of optimality, at least for large samples, for the identification of linear dynamic systems, so that it is interesting to use it as a benchmark to which we can compare the prediction capability of our model. We also compare our model with the factor model proposed in [41] where, however, the number r of latent factors is assumed to be a given input. We computed this input by preliminarily applying the Hallin and Liska's information criterion (IC). The paper [41] assumes an underlying AR factor model where the idiosyncratic components are temporarily and cross-sectionally uncorrelated and it employs the expectation maximization algorithm

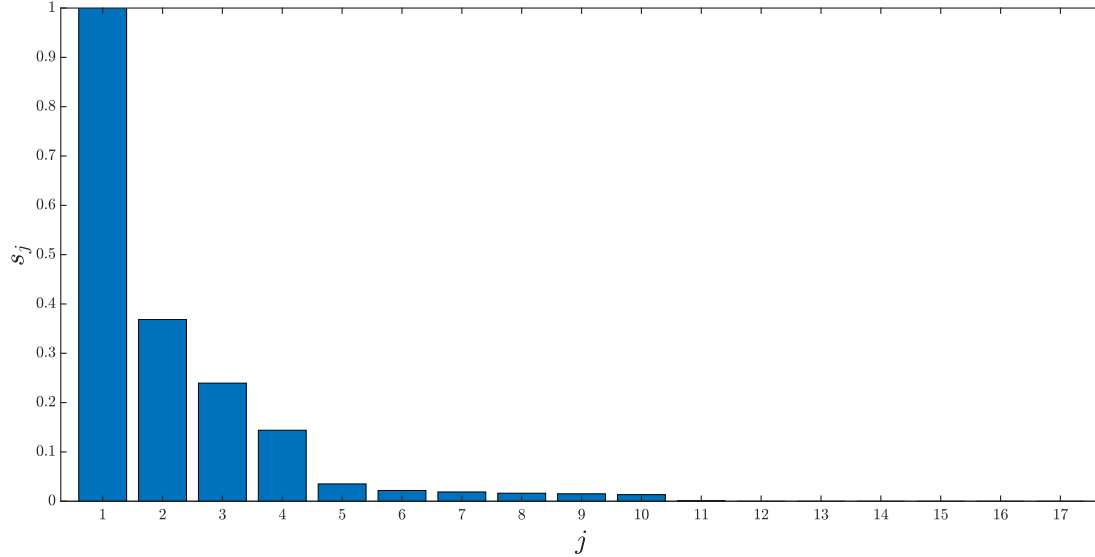


Figure 4.2: SML system dataset: Application of the ARMA factor models identification procedure by using the measurements y^{N_1} from the SMLsystem as training data. The figure shows the integral over the unit circle of the normalized singular values of Φ_L^c .

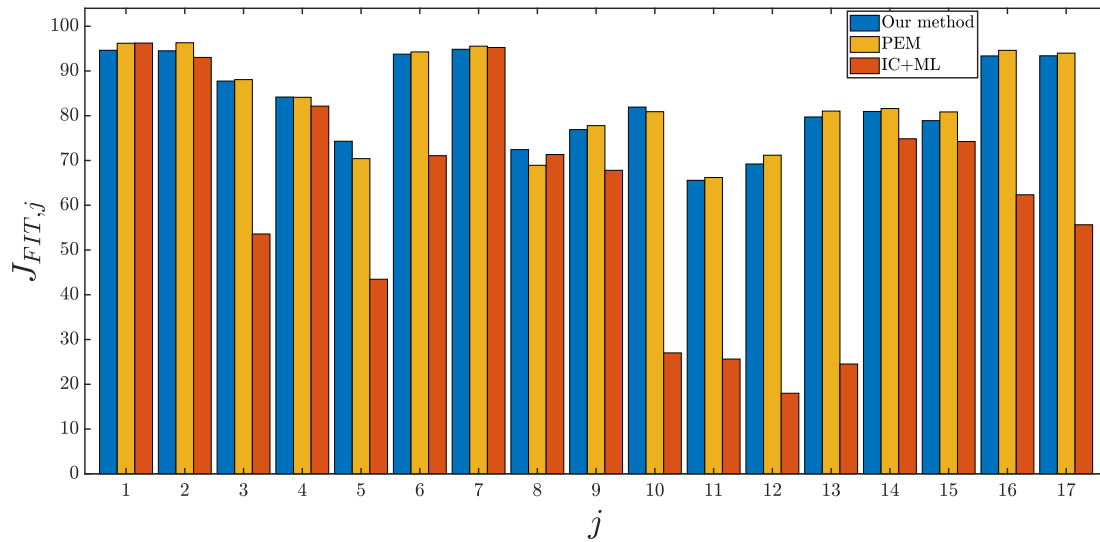


Figure 4.3: SML system dataset: Fit (in percentage) term $J_{FIT,j}$ for each output channel for the model estimated via our ARMA DFA method, via PEM and via the IC+ ML approach. The models are estimated using the dataset y^{N_1} , with $N_1 = 2764$, whereas the fit values are computed by using the measurements y^{N_2} from the SMLsystem as validation data.

to compute the ML estimator. The selection of the model order both for our method and for the method proposed in [41] is obtained by applying the Bayesian information criterion (BIC). Notice that since the IC has a random step, it gives different results when repeatedly applied to the same dataset y^{N_1} . In this particular case the IC method estimates either 1 or 4 latent factors: to be more than fair, we provided the method proposed in [41] with the input $r = 4$ since the corresponding model explains the training data much better than the one corresponding to $r = 1$.

The second dataset y^{N_2} is used in the validation step to test the prediction capability of the three estimated models. The results are summarized in Figure 4.3 which displays for each output channel $j = 1, \dots, m$ the fit (percentage) term:

$$J_{FIT,j} := 100 \left(1 - \frac{\sqrt{\sum_{t=N_1+1}^{N_1+N_2} (y_j(t) - \hat{y}_j(t|t-1))^2}}{\sqrt{\sum_{t=N_1+1}^{N_1+N_2} (y_j(t) - \bar{y}_j)^2}} \right)$$

where $\bar{y}_j := \frac{1}{N_2} \sum_{t=N_1+1}^{N_1+N_2} y_j(t)$ and $\hat{y}_j(t|t-1)$ is the one-step ahead prediction at time t computed with zero initial conditions for the three estimated models. The figure shows that the ARMA factor model estimated with the proposed method matches quite well the measurement data y^{N_2} , reaching fit values that are essentially equal to the benchmark PEM estimates. This allows us to conclude that the available smart building dataset can be successfully modeled with the proposed method. It is a remarkable result since the factor model is parameterized by 257 coefficients, much less than the 612 coefficients of the PEM estimate. Not only the factor model is more parsimonious, but it is also able to organize the complex, high dimensional dataset in a suitable structured model, which is easier to understand and interpret. On the contrary, the PEM model is much more complex and does not give us any intuitive explanation of the underlying dynamics. Finally we notice the AR factor model obtained by applying the IC and the ML principle fails to provide good performances for several output channels.

Another important consequence of dealing with simple models, such as those considered in our setting, is the possibility of identifying systems from a limited dataset. Indeed simple models have few parameters and hence highly reduce the risk of overfitting. To concretely show this advantage in our setting, we repeat the previous simulation by assuming that we have only access to the first 800 measurements (≈ 8 days) of the Smart Building dataset y^{N_1} for the estimation step. We then compute the one-step ahead prediction capabilities of our ARMA factor model, the PEM model and the IC+ML model on the second dataset y^{N_2} . Figure 4.4 shows the results. We find out that the proposed ARMA factor model provides the best performances, reaching an average fit

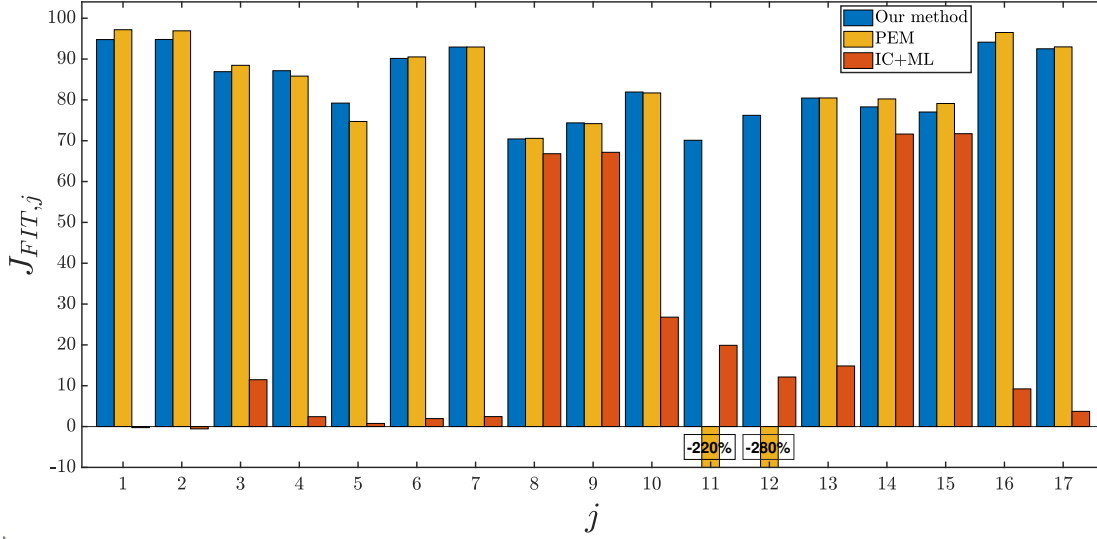


Figure 4.4: SML system dataset: Fit (in percentage) term $J_{FIT,j}$ for each output channel for the model estimated via our ARMA DFA method, via PEM and via the IC+ ML approach. The models are estimated using the first 800 samples of the dataset y^{N_1} , whereas the fit values are computed by using the measurements y^{N_2} from the SMLsystem as validation data. Fit terms lower than -10% are not shown in the figure and they are specified by a label rounded by a rectangular box on the corresponding bar.

term equal to 83.6% against the 45.9% and the 22.5% of the PEM and the IC+ML methods, respectively. In particular the PEM method completely fails to predict the channels number 11 and 12. We conclude that, differently from the PEM model, our ARMA factor model does not suffer from overfitting.

We have repeated the numerical simulations with the SMLsystem dataset for different values of the probability α , specifically for $\alpha = 0.2$ and $\alpha = 0.8$. We obtain that the estimated number of factors is still equal to 4 and the prediction capabilities of the model remain essentially the same: the algorithm appears to be robust with respect to the choice of α . We have also tested the proposed DFA method on the Smart Building dataset by changing the values of the tolerance parameters ε^{ABS} and ε^{REL} . By decreasing the value of the tolerances the computation time grows, whereas the resulting model remains essentially the same. On the other hand, for larger values of the tolerance parameters the ADMM algorithm may stop before reaching an accurate solution. As a matter of fact, by setting $\varepsilon^{\text{REL}} = \varepsilon^{\text{ABS}} = 10^{-3}$, the DFA procedure still recovers the exact number of latent factors, but the resulting model shows poorer performances.



Spectral density estimation

This appendix deals with the problem of estimating the spectral density function of a stationary random process (or random field) from a finite length record of observed data. This problem is of paramount importance in control, signal processing and time-series analysis. Just to highlight a few of the countless important applications where spectral estimation plays a key role, we mention the problem of estimating the impulse response of a single-input single-output (SISO) linear time-invariant (LTI) system through empirical transfer function estimate (ETF), [85], [110], the reconstruction of the topology of dynamic network, [2], [8], [32], [136] and targets detection using radar signals, [43], [44]. Even more complex and refined procedures, such as the THREE and THREE-like methods [23], [56], [57] or the robust DFA paradigm introduced in Chapters 3 and 4, start from a coarse estimate of the spectral density.

Spectral estimation has been extensively studied and many different approaches, both parametric and non parametric, have been proposed; we refer the reader to [116] for an overview of the literature and a rich list of references. The most basic and widely used estimation method is the periodogram. Its statistical properties have been extensively studied over the years. It is well known that the periodogram provides an asymptotically unbiased estimate of the underlying spectrum, but it has some problems: indeed, it is a mean-square inconsistent spectral estimator¹ so that, contrarily to what one would expect, the estimate does not converge to the “true” spectral density as the number of samples increases. Furthermore, the periodogram values at adjoining frequencies are asymptotically uncorrelated, so that the estimate may exhibit an erratic behavior. Finally, a crucial issue is enforcing positivity of the estimated spectral density. This can be achieved by the so-called biased version of the periodogram. The latter, however, forces the rank of the estimated spectral density to be equal to one which is a serious

¹Recall that an estimator is said to be mean-square consistent if the mean square error of the estimator tends to zero as the sample size tends to infinity.

problem in many situations where multivariate processes are considered both in the unidimensional and in the multidimensional case.

A traditional way of dealing with these problems is to smooth the periodogram across frequencies; several methods to smooth the periodogram have been considered [11], [12], [18], [33]. A viable and simple strategy to smooth the periodogram consists in truncating the sum in its definition formula, *i.e.* in windowing the periodogram with a rectangular weight function. The effect of the truncation is to reduce the variance of the estimator at the price of a higher bias. Hence, a proper choice of the truncation point is crucial to correctly balance the tradeoff between bias and variance. This choice, however, is far from being obvious.

In this appendix, we propose the f -truncated periodogram, *i.e.* a truncated periodogram where the truncation point is a suitable function f of the sample size. By focusing on discrete-time, second-order, stationary, Gaussian, possibly multivariate and multidimensional random processes, we show that, under the only assumption that the spectral density of the underlying process has absolutely summable Fourier coefficients, our truncation method provides an estimator which is mean-square consistent. We provide a unified, direct, conceptually simple and self-contained proof of the consistency of the f -truncated periodogram that holds for multidimensional and multivariate random processes. The results of this appendix are published in

- L. Falconi, A. Ferrante, and M. Zorzi, “Mean-square consistency of the f -truncated M^2 -periodogram,” *Automatica*, vol. 147, p. 110 672, 2023.

A.1 The f -truncated periodogram

We give separate discussion for unidimensional random processes (Subsection A.1.1) and multidimensional random fields (Subsection A.1.2). Although the methodology is the same, the unidimensional case is analyzed to better understand the paradigm.

A.1.1 Unidimensional case

Suppose that we have the m -valued, zero-mean, stationary random process

$$y(t) = \sum_{\sigma=0}^{\infty} M(\sigma)e(t - \sigma), \quad t \in \mathbb{Z}, \quad (\text{A.1})$$

where $e = \{e(t), t \in \mathbb{Z}\}$ is normalized white Gaussian noise of dimension p . The impulse response $M(\cdot) : \mathbb{Z} \rightarrow \mathbb{R}^{m \times p}$ is such that $\sum_{\sigma=-\infty}^{\infty} \|M(\sigma)\| = \sum_{\sigma=0}^{\infty} \|M(\sigma)\| < \infty$, meaning

that the system is causal and bounded-input bounded-output (BIBO) stable, which is the case in most practical applications. Note that in (A.1) we have assumed that the system generating the process $y(t)$ is causal, i.e. $M(\sigma) = 0$ for all $\sigma < 0$, because this is the most common situation in the unidimensional case, where t typically represents the time variable. However, this restriction is not necessary and the results can be easily extended to non-causal systems.

It is well-known that the autocovariance sequence $R_k := \mathbb{E}[y(t+k)y(t)^\top]$ depends only on the lag k and it enjoys the property $R_k = R_{-k}^\top$. Define the spectrum of the process as the Fourier transform of the covariance function

$$\Phi(e^{i\theta}) := \sum_{k=-\infty}^{k=\infty} R_k e^{-i\theta k}, \quad \theta \in [0, 2\pi). \quad (\text{A.2})$$

Then, a simple estimator of the spectrum is

$$\hat{\Phi}(e^{i\theta}) := \sum_{k=-n}^{k=n} \hat{R}_k e^{-i\theta k} \quad (\text{A.3})$$

where \hat{R}_k is an estimate of the covariance lag R_k obtained from the available sample $y^N = \{y(t), t = 1, \dots, N\}$. There are two standard ways to obtain the sample covariances required in (A.3)

$$\hat{R}_k := \frac{1}{N-k} \sum_{t=1}^{N-k} y(t+k)y(t)^\top \quad k = 0, \dots, n, \quad (\text{A.4})$$

$$\hat{R}_k := \frac{1}{N} \sum_{t=1}^{N-k} y(t+k)y(t)^\top \quad k = 0, \dots, n, \quad (\text{A.5})$$

which are the unbiased and the biased covariance estimates, respectively.

Next, we address the fundamental problem of selecting the parameter n in (A.3). A typical choice is $n = N - 1$, in which case the estimator (A.3) is called the periodogram. The main problem with the periodogram lies in its large variations about the true spectrum, even for very large data samples. This effect can be reduced by truncating the periodogram, *i.e.* by choosing $n < N - 1$. We may expect that the smaller the n , the larger the reduction in variance and the lower the resolution. Hence, the choice of n should be based on a trade-off between spectral resolution and statistical variance. Here, we propose to select the length n of the truncated periodogram (A.3) as a function $f(N)$ of the numerosity N of the data sample. Clearly, the performance of the obtained estimator is determined by the chosen function f ; to stress this dependence the estimator

will hereafter be referred to as f -truncated periodogram. If

$$n := f(N), \quad (\text{A.6})$$

where f is a function taking values in \mathbb{N} such that

$$\lim_{N \rightarrow \infty} f(N) = \infty, \quad (\text{A.7})$$

$$\lim_{N \rightarrow \infty} \frac{f(N)^2}{N} = 0, \quad (\text{A.8})$$

then the following theorem can be stated.

Theorem A.1.1. *Given any stochastic process y of the form (A.1), the f -truncated periodogram (A.3) with n defined by (A.6)-(A.7)-(A.8) and \hat{R}_k estimated through (A.4) or (A.5) is a uniformly mean-square consistent estimator of the spectral density, that is*

$$\lim_{N \rightarrow \infty} \mathbb{E} \|\Delta(e^{i\theta})\|^2 = 0 \quad \text{uniformly over } [0, 2\pi),$$

where $\Delta(e^{i\theta}) := \Phi(e^{i\theta}) - \hat{\Phi}(e^{i\theta})$.

The proof is given in the next subsection, where the result is stated in the more general setting of a multidimensional and multivariate random field (see Theorem A.1.2 for the case in which unbiased estimates of the covariance sequence are considered and Theorem A.1.3 for the biased estimates case).

A.1.2 Multidimensional case

We extend and prove the consistency of the f -truncated periodogram to the class of multidimensional random fields. For the sake of clarity, here we first discuss the case in which the f -truncated periodogram is built with the unbiased sample covariances and then with the biased covariance estimates.

Consider the second-order stationary random field

$$y(t) = \sum_{\sigma \in \mathbb{Z}^d} M(\sigma) e(t - \sigma), \quad t = (t_1, \dots, t_d) \in \mathbb{Z}^d \quad (\text{A.9})$$

where the positive integer d is the dimension of the index set, $M(\cdot) : \mathbb{Z}^d \rightarrow \mathbb{R}^{m \times p}$ is the impulse response of a BIBO stable filter, *i.e.*

$$\sum_{\sigma \in \mathbb{Z}^d} \|M(\sigma)\| < \infty, \quad (\text{A.10})$$

and $e = \{e(t), t \in \mathbb{Z}^d\}$ is a p -dimensional normalized white Gaussian noise. Accordingly, $y = \{y(t), t \in \mathbb{Z}^d\}$ is a real-valued zero mean random field of dimension m . The covariance matrix, defined as

$$R_k := \mathbb{E}[y(t+k)y(t)^\top] \quad (\text{A.11})$$

does not depend on t by stationary and satisfies the symmetric property $R_k = R_{-k}^\top$. The spectral density of the random field is the Fourier transform of the matrix field $\{R_k, k \in \mathbb{Z}^d\}$, *i.e.*

$$\Phi(e^{i\theta}) := \sum_{k \in \mathbb{Z}^d} R_k e^{-i\langle k, \theta \rangle}, \quad (\text{A.12})$$

where $\theta = (\theta_1, \dots, \theta_d)$ takes values in $\mathbb{T}^d := [0, 2\pi)^d$, $e^{i\theta}$ is a shorthand notation for $(e^{i\theta_1}, \dots, e^{i\theta_d})$ and $\langle k, \theta \rangle := k_1\theta_1 + \dots + k_d\theta_d$ is the usual inner product in \mathbb{R}^d .

Suppose that we observe the finite-length realization of the field $\{y(t), 1 \leq t_j \leq N \text{ for } j = 1, \dots, d\}$ and we want to estimate the spectrum from these observations. The standard unbiased estimate of the covariance sequence is

$$\hat{R}_k := \frac{1}{N_k} \sum_{t \in \Xi_{N,k}} y(t+k)y(t)^\top, \quad (\text{A.13})$$

where each component of the index t satisfies

$$\begin{cases} 1 \leq t_j \leq N - k_j & \text{if } k_j \geq 0 \\ -k_j + 1 \leq t_j \leq N & \text{if } k_j < 0 \end{cases} \quad (\text{A.14})$$

and hence the set $\Xi_{N,k} := \{t \in \mathbb{Z}^d : t_j \text{ satisfies (A.14) for } j = 1, \dots, d\}$ with cardinality $N_k := |\Xi_{N,k}| = (N - |k_1|) \dots (N - |k_d|)$.

Consider a real function f satisfying the assumptions (A.7) and (A.8), and let $n := f(N)$. Then, the f -truncated periodogram is defined as

$$\hat{\Phi}(e^{i\theta}) := \sum_{k \in \Lambda_n} \hat{R}_k e^{-i\langle k, \theta \rangle} \quad (\text{A.15})$$

with $\Lambda_n := \{k \in \mathbb{Z}^d : |k_j| \leq n \text{ for } j = 1, \dots, d\}$. The following theorem guarantees consistency of $\hat{\Phi}$.

Theorem A.1.2. *Given any random field y of the form (A.9), the f -truncated periodogram (A.15) with n defined by (A.6)-(A.7)-(A.8) and \hat{R}_k estimated through (A.13) is a uniformly mean-square consistent estimator of the spectral density, that is, given*

$$\Delta(e^{i\theta}) := \Phi(e^{i\theta}) - \hat{\Phi}(e^{i\theta}),$$

$$\lim_{N \rightarrow \infty} \mathbb{E} \|\Delta(e^{i\theta})\|^2 = 0 \quad \text{uniformly over } \mathbb{T}^d.$$

Proof. By plugging (A.9) into (A.11), it is easy to get $R_k = \sum_{\sigma \in \mathbb{Z}^d} M(\sigma + k)M(\sigma)^\top$ and therefore,

$$\|R_k\|^2 = \sum_{\sigma_a, \sigma_b} \text{tr} \left(M(\sigma_a)M(\sigma_a + k)^\top M(\sigma_b + k)M(\sigma_b)^\top \right).$$

Notice that the BIBO stability assumption (A.10) implies

$$\sum_{k \in \mathbb{Z}^d} \|R_k\| \leq \sum_{k, \sigma \in \mathbb{Z}^d} \|M(\sigma + k)\| \|M(\sigma)\| = \sum_{\sigma \in \mathbb{Z}^d} \|M(\sigma)\| \sum_{k \in \mathbb{Z}^d} \|M(\sigma + k)\| < \infty. \quad (\text{A.16})$$

From (A.9) and (A.13), we immediately obtain

$$\hat{R}_k = \frac{1}{N_k} \sum_{t \in \Xi_{N,k}} \sum_{\sigma_a, \sigma_b \in \mathbb{Z}^d} M(\sigma_a)e(t + k - \sigma_a)e(t - \sigma_b)^\top M(\sigma_b)^\top,$$

and clearly $\mathbb{E}[\hat{R}_k] = R_k$ for any k and N . Moreover,

$$\begin{aligned} \mathbb{E} \|\hat{R}_k\|^2 &= \mathbb{E} \left[\text{tr} \left(\frac{1}{N_k^2} \sum_{t_a, t_b} \sum_{\substack{\sigma_a, \sigma_b, \\ \sigma_c, \sigma_d}} M(\sigma_b)e(t_a - \sigma_b)e(t_a + k - \sigma_a)^\top M(\sigma_a)^\top \right. \right. \\ &\quad \left. \left. \times M(\sigma_c)e(t_b + k - \sigma_c)e(t_b - \sigma_d)^\top M(\sigma_d)^\top \right) \right] \\ &= \frac{1}{N_k^2} \sum_{t_a, t_b} \sum_{\substack{\sigma_a, \sigma_b, \\ \sigma_c, \sigma_d}} \sum_{q, z=1}^m \sum_{i, j, h, l=1}^p M_{qi}(\sigma_b)M_{zj}(\sigma_a)M_{zh}(\sigma_c)M_{ql}(\sigma_d) \\ &\quad \times \mathbb{E} \left[e_i(t_a - \sigma_b)e_j(t_a + k - \sigma_a)e_h(t_b + k - \sigma_c)e_l(t_b - \sigma_d) \right]. \end{aligned}$$

Now, in order to evaluate the expectation in the above formula, by recalling that e is a white Gaussian noise, we distinguish the following cases: if

$$\begin{cases} i = j, & h = l \\ t_a - \sigma_b = t_a + k - \sigma_a, \\ t_b + k - \sigma_c = t_b - \sigma_d \\ t_a - \sigma_b \neq t_b - \sigma_d \end{cases} \quad \text{or}$$

$$\left\{ \begin{array}{l} i = h, \quad j = l \\ t_a - \sigma_b = t_b + k - \sigma_c, \\ t_a + k - \sigma_a = t_b - \sigma_d \\ t_a - \sigma_b \neq t_b - \sigma_d \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{l} i = l, \quad j = h \\ t_a - \sigma_b = t_b - \sigma_d, \\ t_a + k - \sigma_a = t_b + k - \sigma_c \\ t_a - \sigma_b \neq t_a + k - \sigma_a \end{array} \right. ,$$

then the expected value is 1. Moreover, if $t_a - \sigma_b = t_a + k - \sigma_a = t_b + k - \sigma_c = t_b - \sigma_d$, then the expectation is 1 if $i = j$, $h = l$, $i \neq h$ or if $i = h$, $j = l$, $i \neq j$ or if $i = l$, $j = h$, $i \neq j$, whereas the expectation is equal to 3 if $i = j = h = l$. In all the other cases the expected value is zero. In view of these observations, we have:

$$\begin{aligned} \mathbb{E}\|\hat{R}_k\|^2 &= \frac{1}{N_k^2} \left(\sum_{t_a, t_b} \sum_{\sigma_b, \sigma_d} \sum_{q, z=1}^m \sum_{i, h=1}^p M_{qi}(\sigma_b) M_{zi}(\sigma_b + k) M_{zh}(\sigma_d + k) M_q(\sigma_d) \right. \\ &\quad + \sum_{t_a} \sum_{\sigma_a, \sigma_b, \sigma_d} \sum_{q, z=1}^m \sum_{i, j=1}^p M_{qi}(\sigma_b) M_{zj}(\sigma_a) M_{zi}(\sigma_b + \sigma_d - \sigma_a + 2k) M_{qj}(\sigma_d) \\ &\quad \left. + \sum_{t_a} \sum_{\sigma_a, \sigma_b, \sigma_c} \sum_{q, z=1}^m \sum_{i, j=1}^p M_{qi}(\sigma_b) M_{zj}(\sigma_a) M_{zj}(\sigma_c) M_{qi}(\sigma_b + \sigma_c - \sigma_a) \right) \\ &\leq \|R_k\|^2 + \frac{C_1}{N_k} + \frac{C_2}{N_k} = \|R_k\|^2 + \frac{C}{N_k} \end{aligned} \quad (\text{A.17})$$

where the inequality follows from the BIBO stability assumption (A.10) and $C := C_1 + C_2$ is a constant independent of k and N . Consequently,

$$\mathbb{E}\|R_k - \hat{R}_k\|^2 = \mathbb{E}\|\hat{R}_k\|^2 - \|R_k\|^2 \leq \frac{C}{N_k}. \quad (\text{A.18})$$

Now, define

$$S_1(e^{i\theta}) := \sum_{k \in \Lambda_n} (R_k - \hat{R}_k) e^{-i\langle \theta, k \rangle} \quad (\text{A.19})$$

$$S_2(e^{i\theta}) := \sum_{k \in \mathbb{Z}^d \setminus \Lambda_n} R_k e^{-i\langle \theta, k \rangle}, \quad (\text{A.20})$$

so that $\Delta = S_1 + S_2$. From (A.7) and (A.16) it follows that $\lim_{N \rightarrow \infty} \|S_2(e^{i\theta})\| = 0$.

Moreover, $\mathbb{E}\|S_1(e^{i\theta})\| = 0$ and

$$\begin{aligned} \mathbb{E}\|S_1(e^{i\theta})\|^2 &\leq \sum_{k_a, k_b} \sqrt{\mathbb{E}\|R_{k_a} - \hat{R}_{k_a}\|^2 \mathbb{E}\|R_{k_b} - \hat{R}_{k_b}\|^2} \leq \sum_{k_a, k_b \in \Lambda_n} \sqrt{\frac{C}{N_{k_a}} \frac{C}{N_{k_b}}} \\ &\leq \sum_{k_a, k_b \in \Lambda_n} \frac{C}{(N - \max\{k_{a_1}, \dots, k_{a_d}, k_{b_1}, \dots, k_{b_d}\})^d} \leq (2n+1)^{2d} \frac{C}{(N-n)^d} \end{aligned}$$

where the first inequality is a consequence of the Cauchy–Schwarz inequality, and the second inequality comes from (A.18). Notice that under the assumption (A.8), the right hand side of the final inequality tends to zero as N grows to infinity. This suffices to conclude the proof. Indeed, for each θ , $\lim_{N \rightarrow \infty} \mathbb{E}\|\Delta(e^{i\theta})\|^2 = \lim_{N \rightarrow \infty} \{\mathbb{E}\|S_1(e^{i\theta})\|^2 + \|S_2(e^{i\theta})\|^2\} = 0$. ■

Next, we prove that the same consistency result of Theorem A.1.2 holds when, in place of (A.13), we consider the biased covariance estimates

$$\hat{R}_k := \frac{1}{N_0} \sum_{t \in \Xi_{N,k}} y(t+k)y(t)^\top, \quad (\text{A.21})$$

with $N_0 := N^d$, in the truncated periodogram (A.15).

Theorem A.1.3. *Given any random field y of the form (A.9), the f -truncated periodogram (A.15) with n defined by (A.6)-(A.7)-(A.8) and \hat{R}_k estimated through (A.21) is a uniformly mean-square consistent estimator of the spectral density.*

Proof. The proof is similar to the proof of Theorem A.1.2. The main difference is that

$$\hat{R}_k := \frac{1}{N_0} \sum_{t \in \Xi_{N,k}} \sum_{\sigma_a, \sigma_b \in \mathbb{Z}^d} M(\sigma_a) e(t+k-\sigma_a) e(t-\sigma_b)^\top M(\sigma_b)^\top$$

and then $\mathbb{E}[\hat{R}_k] = \frac{N_k}{N_0} R_k$. Repeating the same reasoning as in the proof of Theorem A.1.2, it is not difficult to see that inequality (A.17) becomes

$$\mathbb{E}\|\hat{R}_k\|^2 \leq \left(\frac{N_k}{N_0}\right)^2 \|R_k\|^2 + \frac{N_k C_1}{N_0^2} + \frac{N_k C_2}{N_0^2} \leq \left(\frac{N_k}{N_0}\right)^2 \|R_k\|^2 + \frac{C}{N_0}.$$

Consequently,

$$\mathbb{E}\|R_k - \hat{R}_k\|^2 = \left(1 - 2\frac{N_k}{N_0}\right) \|R_k\|^2 + \mathbb{E}\|\hat{R}_k\|^2 \leq \left(1 - \frac{N_k}{N_0}\right)^2 \|R_k\|^2 + \frac{C}{N_0}. \quad (\text{A.22})$$

Let S_1 and S_2 be defined by (A.19) and (A.20) respectively, and $\Delta = S_1 + S_2$. We have already observed in Theorem A.1.2 that $\lim_{N \rightarrow \infty} \|S_2(e^{i\theta})\| = 0$. Next, we prove that also $\|S_1(e^{i\theta})\|$ converges to zero in the mean-square sense. To this end, we notice that

$$\begin{aligned} \mathbb{E}\|S_1(e^{i\theta})\|^2 &\leq \sum_{k_a, k_b} \sqrt{\mathbb{E}\|R_{k_a} - \hat{R}_{k_a}\|^2 \mathbb{E}\|R_{k_b} - \hat{R}_{k_b}\|^2} \\ &\leq \sum_{k_a \in \Lambda_n} \sqrt{\left[\left(1 - \frac{N_{k_a}}{N_0}\right)^2 \|R_{k_a}\|^2 + \frac{C}{N_0} \right]} \sum_{k_b \in \Lambda_n} \sqrt{\left[\left(1 - \frac{N_{k_b}}{N_0}\right)^2 \|R_{k_b}\|^2 + \frac{C}{N_0} \right]} \\ &\leq \left(\sum_{k \in \Lambda_n} \left(1 - \frac{N_k}{N_0}\right) \|R_k\| \right)^2 + 2\sqrt{\frac{C}{N_0}} \sum_{k \in \Lambda_n} \left(1 - \frac{N_k}{N_0}\right) \|R_k\| + (2n+1)^{2d} \frac{C}{N_0}, \end{aligned}$$

where the first inequality is a consequence of the Cauchy–Schwarz inequality, and second inequality comes from (A.22). The right side of the last inequality is given by the sum of three terms, of which the last one obviously tends to zero as N grows to infinity because of (A.8). As regards the other two terms, notice that

$$\sum_{k \in \Lambda_n} \left(1 - \frac{N_k}{N_0}\right) \|R_k\| \leq \sum_{j=0}^d \binom{d}{j} \left(\frac{n}{N}\right)^j \sum_{k \in \Lambda_n} \|R_k\|,$$

where $\binom{d}{j}$ is the binomial coefficient. The right-hand side goes to zero if condition (A.8) is satisfied since the infinite sum $\sum_{k \in \mathbb{Z}^d} \|R_k\|$ is assumed to be convergent. Hence $\lim_{N \rightarrow \infty} \mathbb{E}\|S_1(e^{i\theta})\|^2 = 0$ and, as a consequence, also $\lim_{N \rightarrow \infty} \mathbb{E}\|S_1(e^{i\theta})\| = 0$. This concludes the proof. Indeed, for each θ

$$\lim_{N \rightarrow \infty} \mathbb{E}\|\Delta(e^{i\theta})\|^2 = \lim_{N \rightarrow \infty} \left\{ \mathbb{E}\|S_1(e^{i\theta})\|^2 + \|S_2(e^{i\theta})\|^2 + 2 \operatorname{tr} \mathbb{E}[S_1(e^{i\theta})] S_2(e^{i\theta})^\top \right\} = 0. \quad \blacksquare$$

So far we have assumed that y is a real-valued signal. However, the result can be easily generalized to the complex-valued case by replacing the transpose operator $(\cdot)^\top$ with the complex-conjugate operator $(\cdot)^*$ in the previous definitions and computations.

In [50], in order to illustrate the theory and to highlight the importance in practice of the proposed estimator, we have performed numerical simulations concerning three concrete identification problems: the impulse response estimation of a SISO system, the problem of learning undirected graphical models and the target parameter estimation in automotive radars. The results confirm that the proposed spectral estimator is indeed effective.

B

Generalized dynamic factor analysis

One essential feature of the *exact* factor models (2.1) and (2.3) is that the the idiosyncratic disturbances are mutually uncorrelated at all leads and lags. However, in a number of practical cases, assuming orthogonality of the noise components is too restrictive. *generalized factor analysis* (GFA) and *generalized dynamic factor analysis* (GDFA) models mitigate this requirement, allowing for a “weak” correlation between the noise components. Obviously, this tends to make the FA problem ill-defined as the basic goal of uniquely splitting the observed signal into a noiseless component plus “additive noise” is made vacuous, unless some extra assumptions are made on the model and on the very notion of “noise”. Quite surprisingly, for models describing an *infinite* number of observables a meaningful weakening of the uncorrelation property of the noise components can be introduced, so as to guarantee uniqueness of the decomposition.

The static generalization of factor models was first rigorously discussed by Chamberlain and Rothschild [24], [25]. Later, Forni, Lippi and collaborators introduced GDFA models and their estimation in a series of widely quoted papers: [60]–[62], [65]. Although GDFA models have been primarily of interest to econometricians, this modeling paradigm has stimulated interest in the System Identification community, see [3], [4], [20], [34], [36].

In this Appendix, we shortly review the structure of GDFA models and we investigate the application of state-space estimation techniques to tackle the hidden factors estimation problem in GDFA models. We show that the one-step ahead Kalman predictor does not provide a perfect asymptotic estimate of the hidden factors. As a consequence, a GDFA model cannot be a predictor-based innovation model. We prove that the pure filter, on the contrary, provides a perfect asymptotic estimate of the latent state variable. Moreover, under reasonable assumptions, the filter estimation error converges weakly to the idiosyncratic noise generating the data so that GDFA models are weakly equivalent to a pure-filter type innovation models. The content of this appendix is based on

- G. Picci, L. Falconi, A. Ferrante, *et al.*, “Hidden factor estimation in dynamic generalized factor analysis models,” *Automatica*, vol. 149, p. 110 834, 2023,

to which we refer the reader for a thorough discussion.

B.1 State-space GDFA models

Before introducing the GDFA model, we need to introduce the concept of idiosyncratic sequence. We formally write the covariance matrix of an infinite-dimensional zero mean vector $y = [y_1 \ y_2 \ \dots]^\top$ as $\Sigma := \mathbb{E}[yy^\top]$. We let Σ_N indicate the top-left $N \times N$ block of Σ , equal to the covariance matrix of the subvector made of the first N components of y denoted y_N . The inequality $\Sigma > 0$ means that all submatrices Σ_N of Σ are positive definite, which we shall always assume in the following. Let $\ell^2(\Sigma)$ denote the Hilbert space of infinite sequences $a := \{a_k, k \in \mathbb{N}\}$ such that $\|a\|_\Sigma^2 := a^\top \Sigma a < \infty$. When $\Sigma = I$, we simply use the symbol ℓ^2 and denote the corresponding norm with the symbol $\|\cdot\|$.

Definition B.1.1 ([65]). Let $\{a_n, n \in \mathbb{N}\}$ be a sequence of elements of the space $\ell^2 \cap \ell^2(\Sigma)$. We say that $\{a_n, n \in \mathbb{N}\}$ is an *averaging sequence* (AS) if $\lim_{n \rightarrow \infty} \|a_n\| = 0$.

For instance, the sequence of elements in ℓ^2 defined as

$$a_n = \frac{1}{n} [\underbrace{1 \ \dots \ 1}_n \ 0 \ \dots]^\top$$

is an averaging sequence. An AS can be seen just as a sequence of linear functionals in $\ell^2 \cap \ell^2(\Sigma)$ converging strongly to zero. The definition is instrumental to the concept of *idiosyncratic sequence* of random variables which will be introduced next.

Definition B.1.2 ([65]). We say that the random sequence y is *idiosyncratic* if for any averaging sequence $\{a_n \in \ell^2 \cap \ell^2(\Sigma)\}$ it holds that $\lim_{n \rightarrow \infty} a_n^\top y = 0$. The limit is understood in mean square.

For example, a zero-mean sequence whose variance is a bounded operator in ℓ^2 is idiosyncratic. In fact let the operator norm $\|\Sigma\|$ be bounded by $\alpha > 0$. Then $\Sigma \leq \alpha I$ where I is the identity operator, so that,

$$\mathbb{E}[(a_n^\top y)^2] = a_n^\top \Sigma a_n \leq \alpha \|a_n\|^2 \rightarrow 0$$

for any sequence $\{a_n\}$ tending to zero in norm. In particular, uncorrelated (white) sequences of zero-mean random variables having uniformly bounded variance are idiosyncratic. The characterization actually goes both ways:

Proposition B.1.3. *A zero-mean sequence of random variables is idiosyncratic if and only if its variance matrix is a bounded operator in ℓ^2 . It follows, as remarked in [65], that a sequence y is idiosyncratic if and only if there exists M such that $\forall N > 0$, $\|\Sigma_N\| \leq M$, where Σ_N is the covariance of the vector y_N obtained by selecting the first N components of y (see [20] for more details).*

We are now ready to introduce the GDFA model. Let $t \in \mathbb{Z}$ be the time variable and denote by $y := \{y(t), t \in \mathbb{Z}\}$ a zero-mean, stationary vector process of infinite cross-sectional dimension so that at each time t the random vector $y(t)$ has countably infinite random components. We consider the following finite-dimensional¹ dynamic model where $y(t)$ depends on a vector of n common factors, x , evolving according to a linear dynamics of the form²

$$\begin{cases} x(t+1) = Ax(t) + v(t) \\ y(t) = Cx(t) + w(t). \end{cases} \quad (\text{B.1})$$

The study of GDFA's (B.1) will be undertaken by considering sequences of truncated models of increasing cross-sectional dimension N , each describing the subvector y_N made of the first N components of the original output vector y . More precisely, we consider a class of truncated models of the form

$$\begin{cases} x(t+1) = Ax(t) + v(t) \\ y_N(t) = C_N x(t) + w_N(t), \end{cases} \quad (\text{B.2})$$

where the state dimension n of the model (B.1) is fixed (and therefore does not grow with the output dimension N). Each output matrix $C_N \in \mathbb{R}^{N \times n}$ is the top submatrix of C of dimension $N \times n$ so that C_N has the nested structure $C_{N+t} = \begin{bmatrix} C_N^\top & \tilde{C}_k^\top \end{bmatrix}^\top$, and the noise vectors $w_N(t)$ have a similar nested structure. Notice that, by these assumptions, each $y_N(t)$ is a N -vector stationary process. We assume that

1. The n -dimensional latent factor $x = \{x(t); t \in \mathbb{Z}\}$ follows a stationary Markov evolution described by the first equation in (B.1), where $A \in \mathbb{R}^{n \times n}$ is an asymptotically stable matrix (all its eigenvalues have modulus strictly less than one) and $v(t)$ a white noise process of dimension n whose covariance is denoted by Q . Hence the steady-state

¹Of course, more general infinite-dimensional GDFA models are possible that are not included in our systems' class.

²For consistency with standard system-engineering notations (see e.g. [83]), we have denoted the factor loading (also called output or observation) matrix by C instead of W_L as in the previous Factor Analysis models, and with $x \in \mathbb{R}^n$ instead of $u \in \mathbb{R}^r$ the common factor vectors.

variance of $x(t)$ is the unique solution of the Stein (discrete-time Lyapunov) equation $\Sigma = A\Sigma A^\top + Q$.

2. The infinite-dimensional white noise vector $\{w(t), t \in \mathbb{Z}\}$ is not assumed to have uncorrelated components as in standard Factor Analysis models and (without loss of generality) is assumed to be uncorrelated with $\{v(t), t \in \mathbb{Z}\}$ at all times.

The following assumptions regarding the asymptotic behavior for $N \rightarrow \infty$ of the sequence of models (B.2) will be made:

3. C is an $\infty \times n$ matrix with *strongly linearly independent columns*. This means that $\lim_{N \rightarrow \infty} \lambda_{\min}[C_N^\top C_N] = +\infty$, where $\lambda_{\min}[\cdot]$ denotes the smallest eigenvalue (see [100, Appendix B] for a discussion and for more details on this).
4. The noise $w(0)$ is an idiosyncratic sequence (with respect to the cross sectional dimension).

Since $w_N(t)$ is a stationary white noise process (in t) for all N , it is easy to check that Assumption 4 implies that, for any given t , $w(t)$ is also idiosyncratic. To avoid technicalities, we also assume that

5. The model is wide-sense stationary and minimal so that the pair (A, Q) , where Q is the covariance of the white noise v , is reachable.
6. The covariance R_N of the output noise $w_N(t)$ is positive definite, i.e. $R_N > 0, \forall N$.

The Kalman predictor

For each finite N one can estimate the latent variable $x(t)$ in the GDFA model (B.2) by Kalman filtering, see e.g. [75]. The usual understanding of Kalman filtering leads to compute the one-step ahead estimate $\hat{x}_N(t|t-1)$ of the hidden variable $x(t)$ based on previous outputs up to time $t-1$. The estimator can obviously be implemented for finite truncations of the model of increasing dimension N and, assuming steady state, leads to the following sequence of innovation models:

$$\begin{cases} \hat{x}_N(t+1|t) = A\hat{x}_N(t|t-1) + K_N e_N(t) \\ y_N(t) = C_N \hat{x}_N(t|t-1) + e_N(t), \end{cases} \quad (\text{B.3})$$

where the innovation $e_N(t) := y_N(t) - C_N \hat{x}_N(t|t-1)$ has covariance

$$\Lambda_N = C_N P_N C_N^\top + R_N \quad (\text{B.4})$$

with P_N being the stabilizing solution of the algebraic Riccati equation

$$P_N = A[P_N - P_N C_N^\top (C_N P_N C_N^\top + R_N)^{-1} C_N P_N] A^\top + Q \quad (\text{B.5})$$

from which one can compute the *Kalman gain* $K_N := A P_N C_N^\top \Lambda_N^{-1}$. The question is how the estimates behave for $N \rightarrow \infty$. In particular the question is whether the (stationary) innovation representation (B.3) of $y_N(t)$ is a legitimate G DFA representation for $N \rightarrow \infty$. Given the standing assumption on C , this will be true if and only if the limit innovation process $e_N(t)$ is idiosyncratic. The answer is given by the following theorem.

Theorem B.1.4. *Consider a class of truncated models of the form (B.2) with Assumptions (1) to (6). For $N \rightarrow \infty$ the innovation process e_N in the steady-state innovation representation (B.3) does not tend to an idiosyncratic process. Hence for $N \rightarrow \infty$ the innovation model (B.3) does not tend to a legitimate G DFA model.*

Proof. By minimality of the model (B.2), (A, Q) is reachable and hence the stabilizing solution P_N of the algebraic Riccati equation (B.5) is positive definite for each fixed N . Then we can rewrite (B.5) as

$$P_N = A \left(P_N^{-1} + C_N^\top R_N^{-1} C_N \right)^{-1} A^\top + Q \quad (\text{B.6})$$

and hence, for each fixed N , we have $P_N \geq Q$. Since in the original G DFA model w is idiosyncratic, the noise covariances R_N are uniformly bounded, i.e. there exists α (independent of N) such that $R_N \leq \alpha I$. Therefore,

$$P_N^{-1} + C_N^\top R_N^{-1} C_N \geq \frac{1}{\alpha} C_N^\top C_N \geq \frac{\lambda_{\min}[C_N^\top C_N]}{\alpha} I$$

and, as a consequence,

$$\left(P_N^{-1} + C_N^\top R_N^{-1} C_N \right)^{-1} \leq \frac{\alpha I}{\lambda_{\min}[C_N^\top C_N]} \xrightarrow{N \rightarrow \infty} 0. \quad (\text{B.7})$$

This inequality together with (B.6) implies that P_N converges monotonically to Q . Hence, the perturbation term $C_N P_N C_N^\top$ of R_N in (B.4) must have at least one eigenvalue tending to infinity (actually as many as the rank of Q) and must therefore tend to an unbounded operator so that $e_N(t)$ is *not* an idiosyncratic process. In conclusion, the innovation model (B.3) does *not* satisfy the conditions of a G DFA model. ■

Since, as shown in the previous proof, P_N converges to Q , which is not the zero matrix, we have the following corollary.

Corollary B.1.5. *Under the assumptions of Theorem B.1.4, the steady state prediction error of the state does not converge to zero (in mean square) as N diverges. In particular, the one step ahead predictor of the common component vector $\chi_N(t) := C_N x(t)$ does not converge either to $Cx(t)$ or to the measured signal $y(t)$, as $N \rightarrow \infty$.*

Proof. The one step ahead predictor of $\chi_N(t)$, $\hat{\chi}(t | t-1) := C_N \hat{x}_N(t | t-1)$ is just the one step ahead predictor of $y_N(t)$ and one can write $y_N(t) = \hat{\chi}_N(t | t-1) + e_N(t)$ where e_N is the output prediction error i.e. the innovation. Since, as we have shown, the covariance matrix of the prediction error e_N becomes *unbounded* as $N \rightarrow \infty$, the predictor $\hat{\chi}_N(t | t-1)$ cannot be consistent in mean square. In fact, as we have already seen, the term $C_N P_N C_N^\top$, namely the steady state covariance matrix of the prediction error $C_N x(t) - C_N \hat{x}_N(t | t-1)$ must have at least one eigenvalue tending to infinity and there must then be at least one direction along which the error covariance diverges in mean square. This is *a fortiori* true for the covariance matrix $C_N P_N C_N^\top + R_N$ of the difference $y_N(t) - \hat{\chi}_N(t | t-1)$. ■

The fact that the innovation model (B.3) cannot in the limit be interpreted as a valid G DFA model has important consequences. Among them, it suggests that the application of standard subspace identification methods to the identification of G DFA models (see e.g. [88]), while effective in deriving a good generative model for the observed data, may lack the capability of extracting the very G DFA feature of the model itself. In fact, these methods are based (typically, via standard canonical correlation analysis of the future onto the strict past of the process $y(t)$) on the construction of a particular basis on the predictor space, [83]. Indeed, the results of the following subsection seem to suggest that considering the pure filter estimator in place of the predictor may be advantageous.

We may attempt a frequency-domain analysis of our result. Obviously the matrix transfer function of each model of the type (B.3) is the unique³ outer (i.e. stable and minimum-phase) spectral factor of the spectral density of the truncated process $y_N(t)$. For $N \rightarrow \infty$ these spectral factors (which are $N \times N$ rational outer matrices of full rank N) have zeros inside the open unit circle. Hence, the limit spectral factor will have all of its zeros (at most) in the closed unit circle. Therefore the limit spectral factor could also be called outer, or (weakly) minimum phase. However, since the innovation process corresponding to this factor is not idiosyncratic, the limit model cannot be a G DFA model. This could be stated by saying that *there cannot exist prediction-error innovation models in the class of G DFA descriptions*.

³Up to uninteresting multiplication on the right side by an orthogonal matrix.

The pure filter estimator

Since the output noise covariance of the one step-ahead innovation model is in a sense “too big” as it has an unbounded component, one may wonder if to obtain a viable GDFAs model, one could choose the (pure) filter estimate $\mathbb{E}[x(t) | y^t]$, where y^t is the infinite past up to the time t , instead of the one-step ahead state predictor. It is in fact well-known that this estimate leads to a model with smaller state error variance. To verify this conjecture, consider the steady state estimate $\hat{x}_N(t) := \mathbb{E}[x(t) | y_N^t]$, given the infinite past up to the time t , which satisfies the recursion

$$\hat{x}_N(t+1) = \mathbb{E}[x(t+1) | y_N^t] + \mathbb{E}[x(t+1) | e_N(t+1)]$$

which for a model with uncorrelated state and output noises, yields

$$\hat{x}_N(t+1) = A\hat{x}_N(t) + L_N e_N(t+1)$$

where $L_N := P_N C_N^\top \Lambda_N^{-1}$. This yields the *filtered innovation model*

$$\begin{cases} \hat{x}_N(t+1) &= A\hat{x}_N(t) + L_N e_N(t+1) \\ y_N(t) &= C_N \hat{x}_N(t) + \hat{e}_N(t) \end{cases} \quad (\text{B.8})$$

where (note the hatted symbol) $\hat{e}_N := y_N(t) - C_N \hat{x}_N(t)$ is the *filter innovation* which is a white noise process. In fact e_N and \hat{e}_N are related by the formula

$$e_N(t) = [I - C_N L_N]^{-1} \hat{e}_N(t) \quad (\text{B.9})$$

which follows from (B.8) as

$$\hat{e}_N(t) := y_N(t) - C_N(\hat{x}_N(t | t-1) + L_N e_N(t)) = e_N(t) - C_N L_N e_N(t).$$

This agrees with the fact that the noise term in the second equation of (B.8) is uncorrelated with y_N^t and hence with $\hat{x}_N(t)$. The variance of $\hat{e}_N(t)$ has the representation

$$\hat{\Lambda}_N := \mathbb{E}[\hat{e}_N(t)\hat{e}_N(t)^\top] = [I - C_N P_N C_N^\top \Lambda_N^{-1}] \Lambda_N [I - C_N P_N C_N^\top \Lambda_N^{-1}]^\top = R_N \Lambda_N^{-1} R_N.$$

Theorem B.1.6. *Consider a class of truncated models of the form (B.2) with Assumptions (1) to (6). Then, the state and output noises in the associated model (B.8) are uncorrelated and the output noise variance $\hat{\Lambda}_N$ tends to a bounded operator as $N \rightarrow \infty$. Therefore (B.8) converges to a legitimate GDFAs representation of the process y .*

Proof. That $v_N(t) := e_N(t+1)$ and $w_N(t) := \hat{e}_N(t)$ are uncorrelated follows readily from the equation (B.9) because $\hat{e}_N(t)$ is white. Moreover, all the eigenvalues of $\Lambda_N = C_N P_N C_N^\top + R_N$ are positive and bounded below. Hence, $\hat{\Lambda}_N$ remains bounded for $N \rightarrow \infty$. ■

A remarkable property of the pure filter realization which follows already from the calculation in (B.7) of the previous paragraph, is recast in the following statement.

Corollary B.1.7. *Under the assumptions of Theorem B.1.6, consider the (steady-state) covariance matrix of the state filtering error*

$$\Pi_N := \mathbb{E} [x(t) - \hat{x}_N(t)] [x(t) - \hat{x}_N(t)]^\top = \left[P_N^{-1} + C_N^\top R_N^{-1} C_N \right]^{-1}.$$

Then, as $N \rightarrow \infty$, the filter error covariance Π_N converges to the zero matrix.

These results imply that the limit for $N \rightarrow \infty$ of the (steady-state) filtered state estimate must converge to the true state $x(t)$, in other words we may say that the filtered estimate is a *consistent estimator*. This should not be surprising since it agrees with the previous general observation that in any *bona fide* G DFA model (B.2) the hidden variable $x(t)$ can asymptotically be recovered exactly as a linear functional of the infinite cross sectional history of the process.

Observe that even if the filtered state estimation error converges to zero (in mean square), in general we cannot recover the original idiosyncratic noise. In fact, if the output noise covariance is uniformly coercive (that is there is a $c > 0$ independent of N such that $R_N \geq cI_N$) then the steady-state covariance of $\hat{\delta}_N(t) := \hat{y}_N(t) - w_N(t)$ does not converge to zero as it is a rank n matrix whose n non-zero eigenvalues are bounded from below by a positive constant. However, we can show that if C_N is uniformly bounded, then for each fixed i the i -th component of $\hat{\delta}_N(t)$ converges to zero in mean square.

Part II

**Distributionally robust control
with distributed uncertainties**

5

Introduction to optimal and robust control

Optimal control is arguably one of the most fundamental ideas in control theory, with applications in virtually all the fields of the discipline, from robotics to bioengineering to finance [17], [19], [76], to name but a few. Traditional optimal control techniques are concerned with designing controllers which are optimal, with respect to some performance measure, for a system evolving over time, on the basis of a mathematical model of it. However, plant variability and uncertainty are formidable adversaries and, soon after their introduction, it was recognized that, as for a kind of waterbed effect, optimality makes these controllers vulnerable to model uncertainties [40].

The potential lack of robustness of traditional optimal controllers has motivated a formidable stream of research. Today, the examination of uncertainty in the mathematical model of a system is a central part of feedback control design and controllers which guarantee an adequate level of performance even in the presence of model misspecifications and inaccuracies are called *robust controllers*. A major stepping stone, which is often considered as the beginning of modern robust control, was the formulation of the H_∞ control theory. Initiated in the early 1980s by George Zames with the pioneering work [129], the objective of H_∞ optimization is to obtain a controller that minimizes the H_∞ norm of a certain transfer function, which is related to the unknown disturbance. This paradigm includes problems of disturbance attenuation, model matching, and tracking. From this first seed, a multitude of alternative approaches to address the robust optimal control problem has been proposed. We mention the mixed H_2/H_∞ control methods, which minimize an H_2 performance criterion subject to a prespecified H_∞ constraint on the closed-loop transfer function [67], [90], [133]; the guaranteed cost control approach, which involves the minimization of a performance measure upper bound to obtain a control law that guarantees a given level of performance for all admissible parameter variations [26], [124]; the optimal control problem for uncertain systems subject to integral quadratic constraints [89], [104], to name just a few. We refer the reader to [13], [38],

[97] and the references therein for a rather comprehensive review on the topic.

Whereas robust control has been traditionally dealt with deterministic approaches, in the last two decades stochastic optimization under uncertainty has received considerable attention. In particular, one emerging paradigm in this area is represented by the *Distributionally Robust Control* (DRC), [96], [122], [127]. Traditional stochastic optimal control techniques are based on the assumption that the probability distribution of the noise variables is fully known. However, in practice, such a probability distribution is usually only vaguely (or even very vaguely) known, and the use of inaccurate distributions may drastically decrease the control performance and cause unwanted system behaviors including instability, see for example [91]. DRC has emerged as a promising method to hedge against this effect. Instead of assuming a given distribution function, DRC considers a specified *ambiguity set* containing the possible distributions. Then, it designs a control policy which minimizes a given cost under the worst-case distribution in the ambiguity set.

The ambiguity set is a key ingredient of any DRC problem. A good ambiguity set should be rich enough to contain the true data-generating distribution with high confidence, and at the same time small enough to exclude pathological distributions, which would incentives overly conservative decisions. Moreover, the ambiguity set should facilitate a tractable reformulation of the DRC problem. A wide range of ambiguity sets was suggested and analyzed in the literature of DRC and distributionally robust optimization [14], [101], [106]. They can be broadly classified into moment-based ambiguity sets and discrepancy-based ambiguity sets. The use of moment ambiguity sets goes back to a pioneering paper by Scarf [105] where it was applied to inventory modeling. They contain all distributions that satisfy certain moment constraints, see for example [123] and the references therein. An attractive alternative is to define the ambiguity set as a ball in the space of probability distributions by using a probability distance function. It is assumed that there is a reference probability distribution and the ambiguity set consists of all probability measures which are “close” to it. Several possible choices for quantifying the concept of closeness between probability densities have been proposed, such as the total variational distance [121], the Wasserstein metric [106], [127], and ϕ -divergences [14] such as the Kullback-Leibler divergence [96]. By adjusting the radius of the ambiguity set, the modeler can control the degree of conservatism of the controller and trade off optimality and robustness. If the radius drops to zero, then the ambiguity set shrinks to a singleton that contains only the nominal distribution, and the DRC problem reduces to a traditional optimal stochastic control problem without uncertainty. On the other, by increasing the radius of the ambiguity set, we improve robustness of the controller.

5.1 Contribution of the thesis

This thesis proposes a new paradigm for the robustification of the linear quadratic Gaussian (LQG) control problem [5], [6], which we call the *Distributed uncertainty Distributionally robust Linear Quadratic Gaussian (D^2 -LQG)* controller. *Distributionally robust* because, by adopting the DRC approach, the proposed control scheme proves robust against variations in the probability density function underlying the system. *Distributed uncertainty* because the modeler is allowed to arbitrarily distribute the uncertainty along the time interval. In detail, in our problem we consider discrete-time, linear, stochastic, uncertain systems defined over a finite time interval. These systems are described in terms of a nominal system, driven by white Gaussian noise, and a class of perturbations that may affect the noise distribution. At each time instant, a relative entropy tolerance is used to model a limit on the admissible noise distributions. Our controller optimizes the closed-loop performances with respect to a quadratic cost function in the worst possible scenario.

The ambiguity set based on relative entropy pseudo-distance has several attractive features. First, relative entropy is a prominent metric in information theory which satisfies important structural properties. In addition, it is the natural “metric” between systems when they are identified from data, [69], [137]. Finally, it turns out that minimax LQG control problem with a relative entropy constraint admits closed-form solutions in the form of a parameterized, risk-sensitive LQG optimal controller (see [72], [126] for a thorough description of the risk-sensitive LQG problem). Thus it provides a new interpretation for the risk-sensitive paradigm.

Our work is closely related to the seminal paper [96] by Petersen *et al.* that proposed an entropy-based DRC optimal control problem for stochastic uncertain systems. A major difference between our paper and [96] concerns the distribution of the uncertainty along the time interval. Indeed, [96] assumes a *single* constraint on the overall noise distribution on the entire time interval. Conversely, our approach allows to arbitrarily distribute the uncertainty along the time interval. Indeed, it imposes a constraint to each time step, taking a point of view whose nature is similar to the approach used by [82], [134], [1], [128] in the robust filtering setting. Petersen’s approach is the natural way to model the uncertainties when the discrepancy between the nominal and the actual system is due to the action of an adversary who can manage a limited mismatch budget to perturb the nominal system and, when convenient, is allowed to concentrate most (or all) such a budget in a few time points. However, in most practical situations the model mismatch is a consequence of modelling approximations and random fluctuations, and

there is no a real adversary. In these cases, since the same effort is usually made to model each time step, it is more realistic to define a specific ambiguity set at each step of the time interval. For this reason, we address the latter problem. The reason for our choice is that in Petersen's approach it may happen that most of the uncertainty is concentrated on few time steps, leading to unrealistic scenarios and to overly conservative conclusions. We believe that this point may better clarified by two examples. Assume that we want to design a robust LQG optimal controller for an airplane. The Gaussian noise models the effects of the wind which, in practice, is not Gaussian. Therefore, we need to guarantee robustness against different distributions of the wind disturbance. By adopting the approach in [96], it may happen that the worst-case distribution differs from the nominal one at just one time interval but in that time interval the difference is unreasonably large: the controller must be able to counteract a totally unrealistic wind concentrated at a single time point with potential degradation of the performances. The possibility of distributing the uncertainty along the whole trajectory and to select different uncertainties radii depending on the reliability of the wind forecasts at each point is clearly an interesting feature in this scenario. A second example is the case when the difference between the nominal Gaussian noise and the actual one is used to account for errors in the model parameters. Also in this case, the possibility of distributing the uncertainty along the whole trajectory is clearly advantageous with respect to a potential worst-case where the model mismatch may be unreasonably concentrated at a single time instant.

By adopting the D^2 -LQG paradigm, in Chapter 6 we solve the worst-case performance analysis problem for autonomous systems. The problem is formulated as a stochastic constrained maximization problem in which the objective function is maximized with respect to the uncertain distributions of the system. At each time instant, a relative entropy tolerance is used to model a limit on the admissible noise distributions. The main result of this chapter consists in showing that the solution to the robust performance analysis problem takes the form of a risk-sensitive cost with time-varying risk-sensitive parameter and that the least favorable probability distribution is Gaussian, with a perturbation of both the mean and the variance with respect to the nominal one. The results of this chapter are published in

- L. Falconi, A. Ferrante, and M. Zorzi, "A new perspective on robust performance for LQG control problems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 3003–3008.

The results of Chapter 6 put the basis for the state-feedback D^2 -LQG controller synthesis faced in Chapter 7. Our controller optimizes the closed-loop performances in the worst possible scenario under the constraint that the noise distributional aberrance

does not exceed a certain threshold limiting the relative entropy pseudo-distance between the actual noise distribution and the nominal one. The bounds on the distributional aberrance can be arbitrarily distributed along the whole disturbance trajectory. Using the results of Chapter 6, we were able to transform the D^2 -LQG problem into an equivalent risk-sensitive LQG optimal control problem with state-feedback depending on a time-varying parameter. Chapter 7 is based on

- L. Falconi, A. Ferrante, and M. Zorzi, “Distributionally robust LQG control under distributed uncertainty,” (preliminary version available at *arXiv preprint arXiv:2306.05227 (2023)*).

Finally, Appendix C summarizes the main ideas and the formulas arising in the LQG and risk-sensitive LQG problem.

Notice that the results presented in this thesis rely on the hypothesis of linearity of the underlying system. Nevertheless, this theory is extremely relevant even when dealing with non-linear systems. Indeed, by linearizing the actual dynamics of the system, we can apply the proposed D^2 -LQG procedure and, in doing so, we provide a local approximation of the optimal control strategy for the underlying non-linear system.

6

D^2 -LQG: worst performance analysis

LQG control represents a well-known and profound optimal control design method [5], [6], [118] (see also [54], [55], [58] for the deterministic counterpart). It is indeed the most widely used technique for control of multiple-input and multiple-output (MIMO) systems and the basis for many advanced control techniques, such as all the MPC and MPC-like methods [102]. The LQG method involves the design of a controller which optimizes the closed-loop performance, with respect to a quadratic cost function, assuming that the underlying process is affected by a Gaussian distributed disturbance with known mean and covariance. However, assuming Gaussianity and complete knowledge of the noise description is unrealistic in most practical situations. Therefore, the LQG control may lead to poor performances when applied to the real system, because it does not guarantee robustness with respect to real disturbances and/or modelling uncertainties [40]. This potential lack of robustness has motivated a formidable stream of research aiming to ensure that the controller provides satisfactory performances even in the presence of model misspecifications and inaccuracies.

The results of this chapter can be placed in the broad area of robustification of LQG control and, more specifically, in the framework of DRC, which takes into account uncertainty in the probability distribution underlying a stochastic system. Instead of assuming a given distribution, DRC methods design a control policy which minimizes the expected value of a given cost under the worst-case distribution in a certain *ambiguity set*. In our work, we express the ambiguity set as a ball defined in the relative-entropy topology and centered at the nominal Gaussian distribution. The main novelty with respect to the existing literature on DRC with relative entropy [95], [96] is that the model uncertainty can be arbitrarily distributed along the system trajectory. This is accomplished by imposing a relative entropy constraint for each time instant, instead of assuming a *single* constraint for the entire disturbance trajectory as in [96]. For these reasons, we call our approach the *distributed uncertainty distributionally robust* LQG

control problem.

Within the *D*²-LQG framework, in this chapter we consider the worst-case performance analysis problem, that is the problem of evaluating the worst performance with respect to the admissible noise probability distributions. We consider a class of discrete-time, linear, stochastic, uncertain systems defined over a finite time interval, which are described in terms of a nominal system, driven by white Gaussian noise with known statistics, and a perturbed system, in which a general class of stochastic noise processes are allowed. At each time instant, a relative entropy ambiguity set is used to model a limit on the admissible noise distributions. The robust performance analysis problem is formulated as a stochastic constrained maximization problem in which the objective function is maximized with respect to the uncertain distributions of the system. To address this problem, we adopt a dynamic programming technique to break down the overall problem into simpler subproblems, one for each time-instant, starting from the last timestep and going backward in time. Each subproblem consists in a constrained stochastic maximization problem; we address it by exploiting the Lagrange duality theory. Finally, by using the duality between relative entropy and free energy we are able to convert this problem into the problem of evaluating a LQG risk-sensitive cost for autonomous systems with a time-varying risk-sensitive parameter [72], [126].

The proposed worst-case performance analysis problem is a preliminary step towards the solution of the problem of synthesizing a *D*²-LQG controller that we will address in the following Chapter.

This chapter is outlined as follows. Section 6.1 collects some preliminary results; Section 6.2 defines the class of stochastic uncertain system considered in this work; the worst case performance analysis problem is treated in Section 6.3 and the least-favorable state-space model is derived in Subsection 6.3.1; an illustrative example is presented in Section 6.4 which shows the advantages of distributing the model uncertainty.

6.1 Preliminary results

We collect some relevant results underlying the solution to the *D*²-LQG problems proposed in this thesis.

6.1.1 Gaussian integrals

Lemma 6.1.1. [131] *If $A \in \mathbb{Q}_n$, $A \succ 0$, $x, b \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then*

$$\int_{\mathbb{R}^n} e^{-\frac{1}{2}x^\top Ax + b^\top x + c} dx = \sqrt{\frac{(2\pi)^n}{|A|}} e^{\frac{b^\top A^{-1}b}{2} + c}.$$

Lemma 6.1.2. *Let $f(x)$ be the probability density function of a n -dimensional normal random vector $x \sim \mathcal{N}(\mu, \Sigma)$ with $\Sigma \succ 0$, and let $J(x) = \frac{1}{2}x^\top Gx$ where $G \in \mathbb{Q}_n^+$. Assume $(\Sigma^{-1} - G) \succ 0$. Then*

$$\tilde{f}(x) = f(x) \frac{e^{J(x)}}{\int_{\mathbb{R}^n} e^{J(\bar{x})} f(\bar{x}) d\bar{x}}$$

is a Gaussian density function, $\tilde{f} = \mathcal{N}(\tilde{\mu}, \tilde{\Sigma})$, with $\tilde{\mu} = (I - \Sigma G)^{-1}\mu$ and $\tilde{\Sigma} = (I - \Sigma G)^{-1}\Sigma$.

Proof. By applying Lemma 6.1.1, it is easy to obtain that

$$\begin{aligned} \int_{\mathbb{R}^n} e^{J(\bar{x})} f(\bar{x}) d\bar{x} &= \int_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp \left\{ \frac{1}{2} \bar{x}^\top G \bar{x} - \frac{1}{2} (\bar{x} - \mu)^\top \Sigma^{-1} (\bar{x} - \mu) \right\} d\bar{x} \\ &= \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \int \exp \left\{ \frac{1}{2} \bar{x}^\top (G - \Sigma^{-1}) \bar{x} + \mu^\top \Sigma^{-1} \bar{x} - \frac{1}{2} \mu^\top \Sigma^{-1} \mu \right\} d\bar{x} \\ &= \frac{1}{\sqrt{|I - \Sigma G|}} \exp \left\{ \frac{1}{2} \mu^\top \Sigma^{-1} \left((I - \Sigma G)^{-1} - I \right) \mu \right\} \end{aligned}$$

where in the third equality we have used Lemma 6.1.1. Moreover, by completing the square, we can see that

$$f(x) e^{J(x)} = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp \left\{ -\frac{1}{2} (x - \tilde{\mu})^\top \tilde{\Sigma}^{-1} (x - \tilde{\mu}) + \frac{1}{2} \mu^\top \Sigma^{-1} \left((I - \Sigma G)^{-1} - I \right) \mu \right\}$$

where $\tilde{\mu} = (I - \Sigma G)^{-1}\mu$ and $\tilde{\Sigma} = (I - \Sigma G)^{-1}\Sigma$. The thesis follows immediately. \blacksquare

6.1.2 Optimization and convexity

Lemma 6.1.3. [21, p.81] *Let X be a normed space. Let $\{f_\alpha(x) | \alpha \in I\}$ be a collection of functions with the same domain K . If K is a convex subset of X and $f_\alpha(x)$ is a convex function for each α , then $g(x) := \sup_{\alpha \in I} f_\alpha(x)$ is also convex.*

Lemma 6.1.4. [21, p.87] *Let $f(x, y)$ be convex in (x, y) and K be a convex non-empty set. Then the function $g(x) = \inf_{y \in K} f(x, y)$ is convex in x provided that $g(x) > -\infty$ for some x .*

Theorem 6.1.5 (Von Neumann's Minmax Theorem, [7]). *Let X and Y be linear vector spaces, and Ω and Ψ be convex subsets of X and Y , respectively. If $f(x, y)$ is concave in x and convex in y , then $\max_{x \in \Omega} \min_{y \in \Psi} f(x, y) = \min_{y \in \Psi} \max_{x \in \Omega} f(x, y)$.*

Theorem 6.1.6 (Strong duality, [86]). *Let X be a linear vector space and Ω be a convex subset of X . Let f be a real-valued concave function on Ω and $g : X \rightarrow \mathbb{R}$ a convex map.*

Suppose there exists $x_1 \in \Omega$ such that $g(x_1) < 0$. Let

$$\mu_0 = \sup_{\substack{x \in \Omega \\ g(x) \leq 0}} f(x)$$

and assume μ_0 is finite. Then,

$$\sup_{\substack{x \in \Omega \\ g(x) \leq 0}} f(x) = \min_{\tau \geq 0} \sup_{x \in \Omega} [f(x) - \tau g(x)]$$

and the minimum in the right side is achieved for a $\tau^o \geq 0$.

Lemma 6.1.7. [6, Lemma 3.1 p. 260] Let X and U be linear vector spaces. Assume that the function $f(x, u)$ has a unique minimum with respect to $u \in U$ for all $x \in X$. Then

$$\min_{u(x)} \mathbb{E} [f(x, u)] = \mathbb{E} [\min_u f(x, u)].$$

6.2 Stochastic uncertain System

We consider a class of discrete-time stochastic uncertain system defined over a finite time interval, described in terms of a *reference* (or *nominal*) *system* and a *perturbed system*.

1) *Reference system*: The reference system is described by

$$x_{t+1} = Ax_t + v_t, \quad t = 0, \dots, N \quad (6.1)$$

where $x_t \in \mathbb{R}^n$ is the state vector and $\{v_t, t = 0, \dots, N\}$ is a zero-mean, white Gaussian noise sequence with covariance matrix V . We assume that the noise v_t affects all the components of the dynamics so that $V \succ 0$; this is a typical assumption in problems where the relative entropy is used to measure the proximity of statistical model (see for example [82]). The matrices A and V have appropriate dimensions and, for ease of notation, they are assumed time independent. Generalization to time-varying systems only entails heavier notation; the conclusions however can be reached by the same arguments. Model (6.1) specifies the nominal transition density function

$$f(x_t | x_{t-1}) \sim \mathcal{N}(Ax_{t-1}, V), \quad t = 1, \dots, N + 1. \quad (6.2)$$

Assume that the initial state $x_0 = \bar{x}_0$ is a deterministic quantity. Then, the joint nominal probability density of the state sequence $x_{0:N+1} = [x_0^\top \dots x_{N+1}^\top]^\top$ can be expressed as

$$f(x_{0:N+1}) = \delta(x_0 - \bar{x}_0) \prod_{t=1}^{N+1} f(x_t|x_{t-1}).$$

2) *Perturbed system:* The corresponding perturbed system is described by the state equation

$$x_{t+1} = Ax_t + \tilde{v}_t, \quad t = 0, \dots, N \quad (6.3)$$

where the noise input sequence $\{\tilde{v}_t, t = 0, \dots, N\}$ is defined by an unknown probability density function. The joint probability measure of the perturbed system is denoted by $\tilde{f}(x_{0:N+1})$ and it is assumed to admit a similar Markov structure of the form

$$\tilde{f}(x_{0:N+1}) = \delta(x_0 - \bar{x}_0) \prod_{t=1}^{N+1} \tilde{f}(x_t|x_{t-1}).$$

Hereafter, we use the symbol $\mathbb{E}[\cdot]$ to denote the expectation with respect to the nominal conditional density $f(\cdot)$, whereas $\tilde{\mathbb{E}}[\cdot]$ denotes the expectation with respect to the perturbed density $\tilde{f}(\cdot)$.

3) *Relative Entropy Constraints:* We use the relative entropy between the true and the nominal densities to measure modeling errors. In our setting, instead of applying a single constraint to the relative entropy of the true and nominal probability densities of the state sequence $x_{0:N+1}$ over the whole time interval as done in the seminal paper [96], we specify separate modeling tolerances for each time step of the transition density (6.2). In other words, we use the relative entropy between the true and nominal transition densities at time t , namely $\tilde{f}(x_t|x_{t-1})$ and $f(x_t|x_{t-1})$, to measure modeling errors. Specifically, after introducing the quantity

$$R_t := \int \ln \left(\frac{\tilde{f}(x_t|x_{t-1})}{f(x_t|x_{t-1})} \right) \tilde{f}(x_t|x_{t-1}) dx_t,$$

the model mismatch at time t , with $t = 1, \dots, N + 1$, is expressed by the constraint

$$\tilde{\mathbb{E}}[R_t] \leq d_t, \quad (6.4)$$

where we recall that the expectation is taken with respect to the true marginal

$$\tilde{f}(x_{t-1}) = \int \tilde{f}(x_{t-1}|x_{t-2}) \dots \tilde{f}(x_1|\bar{x}_0) dx_1 \dots dx_{t-2}.$$

The ambiguity set is defined by the convex ball of functions $\mathcal{B}_t := \{\tilde{f}(x_t|x_{t-1}) \in \mathcal{P} : \tilde{\mathbb{E}}[R_t] \leq d_t\}$ where \mathcal{P} is the set of all probability density function on \mathbb{R}^n . Note that \mathcal{B}_t is parameterized by the density $\tilde{f}(x_{0:t-1})$. The tolerance parameter $d_t > 0$ quantifies the mismatch budget at time t .

6.3 Worst performance analysis

We consider the problem of characterizing, for the above stochastic uncertain system, the worst case performance with respect to the cost functional

$$J(x_{0:N+1}) = \frac{1}{2} \sum_{t=0}^N x_t^\top Q x_t + \frac{1}{2} x_{N+1}^\top Q_{N+1} x_{N+1}, \quad (6.5)$$

where $Q \succeq 0$, $Q_{N+1} \succ 0$. This is the typical cost function in LQG problems for autonomous systems like (6.3). More precisely, the problem under consideration is to find

$$\max_{\tilde{f}(x_1|x_0) \in \mathcal{B}_1} \dots \max_{\tilde{f}(x_{N+1}|x_N) \in \mathcal{B}_{N+1}} \tilde{\mathbb{E}}[J]. \quad (6.6)$$

Assumption 6.3.1. The couple (A, Q) is observable.

Because of the constraints' and the objective function's special structure, we can obtain the overall solution by optimizing a sequence of $N + 1$ single-variable constrained optimization problems by adopting a *backward dynamic programming* technique [102, ch. I]. We start by optimizing the objective function with respect to $\tilde{f}(x_{N+1}|x_N)$. More precisely, the problem to be solved at this last stage is

$$\max_{\tilde{f}(x_{N+1}|x_N) \in \mathcal{B}_{N+1}} \tilde{\mathbb{E}}[J] \quad (6.7)$$

for a given $\tilde{f}(x_{0:N})$. Exploiting the duality theory, we show that Problem (6.7) can be replaced by an equivalent unconstrained optimization problem. Let $\tau_{N+1} \in \mathbb{R}$ be the Lagrange multiplier associated with the constraint (6.4) for $t = N + 1$ and define the Lagrangian $L_{N+1}(\tau_{N+1}) := J - \tau_{N+1}[R_{N+1} - d_{N+1}]$. We denote by $W_{N+1}(\tau_{N+1})$ the value of the following unconstrained optimization problem

$$W_{N+1}(\tau_{N+1}) = \max_{\tilde{f}(x_{N+1}|x_N) \in \mathcal{P}} \tilde{\mathbb{E}}[L(\tau_{N+1})]. \quad (6.8)$$

By applying Theorem 6.1.6 to the current optimization problem (6.7), we obtain the following theorem.

Theorem 6.3.2. *Consider the stochastic uncertain system (6.1), (6.3) and (6.4) with cost functional J given by (6.5). Then the value of the stochastic optimization problem (6.7) is finite if and only if $\Gamma_{N+1} := \{\tau_{N+1} > 0 \mid W_{N+1}(\tau_{N+1}) < \infty\}$ is non empty. In this case,*

$$\max_{\tilde{f}(x_{N+1}|x_N) \in \mathcal{B}_{N+1}} \tilde{\mathbb{E}}[J] = \min_{\tau_{N+1} > 0} W_{N+1}(\tau_{N+1}). \quad (6.9)$$

Proof. The theorem is obtained by sensibly applying Theorem 6.1.6 to the current optimization problem with X as the linear vector space of functions from \mathbb{R}^n to \mathbb{R} and Ω the set of probability density functions on \mathbb{R}^n . The objective function corresponds to $\tilde{\mathbb{E}}[J]$ which is a linear (thus concave) map of $\tilde{f}(x_{N+1}|x_N)$. The function $g(\cdot)$ corresponding to $\tilde{\mathbb{E}}[R_{N+1} - d_{N+1}]$ is a convex function of $\tilde{f}(x_{N+1}|x_N)$ thanks to the properties of the relative entropy. Finally, the transition probability density function $\tilde{f}(x_{N+1}|x_N) = f(x_{N+1}|x_N)$ satisfies $\tilde{\mathbb{E}}[R_{N+1} - d_{N+1}] = -d_{N+1} < 0$. Hence, the conditions of the lemma are satisfied. Now, the main clarification is that we can rule out the case in which the minimum in (6.9) is achieved at $\tau_{N+1} = 0$. Indeed, if $\tau_{N+1} = 0$ we have $L(\tau_{N+1}) = J$. Since Q_{N+1} is strictly positive, it is straightforward to verify that $W_{N+1} = \infty$ for any $\tilde{f}(x_{0:N})$. ■

Now, we evaluate the quantity $W_{N+1}(\tau_{N+1})$ for $\tau_{N+1} > 0$. By exploiting first Lemma 1.1.3 and then Lemma 6.1.1, it follows

$$\begin{aligned} W_{N+1} &= \tilde{\mathbb{E}} \left[\frac{1}{2} \sum_{t=0}^N x_t^\top Q x_t + \tau_{N+1} \left(\max_{\tilde{f}(x_{N+1}|x_N) \in \mathcal{P}} \tilde{\mathbb{E}} \left[\frac{x_{N+1}^\top Q_{N+1} x_{N+1}}{2\tau_{N+1}} - R_{N+1} \mid x_N \right] \right) \right] \\ &\quad + \tau_{N+1} d_{N+1} \\ &= \tilde{\mathbb{E}} \left[\frac{1}{2} \sum_{t=0}^N x_t^\top Q x_t + \tau_{N+1} \ln \left(\int \exp \left\{ \frac{x_{N+1}^\top Q_{N+1} x_{N+1}}{2\tau_{N+1}} \right\} f(x_{N+1}|x_N) dx_{N+1} \right) \right] \\ &\quad + \tau_{N+1} d_{N+1} \\ &= \tilde{\mathbb{E}} \left[\frac{1}{2} \sum_{t=0}^{N-1} x_t^\top Q x_t + x_N^\top \Pi_N x_N \right] - \frac{\tau_{N+1}}{2} \ln \left| I - \frac{Q_{N+1} V}{\tau_{N+1}} \right| + \tau_{N+1} d_{N+1} \quad (6.10) \end{aligned}$$

where

$$\Pi_N = Q + A^\top \left(Q_{N+1}^{-1} - \frac{V}{\tau_{N+1}} \right)^{-1} A \quad (6.11)$$

provided that

$$\left(Q_{N+1}^{-1} - \frac{V}{\tau_{N+1}} \right) \succ 0. \quad (6.12)$$

The latter condition is necessary for W_{N+1} to be finite. We know from Theorem 6.3.2 that Problem (6.7) is finite if and only if the set Γ_{N+1} is non-empty, hence if and only if there exists $\tau_{N+1} > 0$ such that (6.12) is satisfied.

Lemma 6.3.3. *If (6.12) is satisfied, then Π_N defined by (6.11) is positive definite.*

Proof. It is clear that $\Pi_N \succeq 0$ as it is the sum of three positive semidefinite matrices. We must show that it is invertible. Assume by contradiction that there exists $y \neq 0$ such that $y^\top \Pi_N y = 0$. It follows that $Qy = 0$ and $Ay = 0$. But then, $y \neq 0$ belongs to the non-observable subspace, contradicting Assumption 6.3.1. ■

In view of (6.9), if Γ_{N+1} is non-empty, Problem (6.6) can be rewritten as

$$\max_{\tilde{f}(x_1|x_0) \in \mathcal{B}_1} \dots \max_{\tilde{f}(x_N|x_{N-1}) \in \mathcal{B}_N} \min_{\tau_{N+1} > 0} W_{N+1}(\tau_{N+1}).$$

Notice that W_{N+1} is a dual function, thus it is convex with respect to the Lagrange multiplier τ_{N+1} [21]. Since W_{N+1} is convex in τ_{N+1} and linear (thus concave) in $\tilde{f}(x_N|x_{N-1})$, in view of Theorem 6.1.5 we can switch the minimization and maximization operation without affecting the result. Hence, Problem (6.6) is equivalent to

$$\max_{\tilde{f}(x_1|x_0) \in \mathcal{B}_1} \dots \min_{\tau_{N+1} > 0} \max_{\tilde{f}(x_N|x_{N-1}) \in \mathcal{B}_N} W_{N+1}(\tau_{N+1}).$$

The knowledge of the structure of W_{N+1} as derived in (6.10) allows us to move to the next stage of the dynamic programming recursion. In the next stage, if we neglect the constant terms, we are concerned with the solution to the problem

$$\max_{\tilde{f}(x_N|x_{N-1}) \in \mathcal{B}_N} \tilde{\mathbb{E}} \left[\frac{1}{2} \sum_{t=0}^{N-1} x_t^\top Q x_t + x_N^\top \Pi_N x_N \right]$$

with $\tilde{f}(x_{0:N-1})$ as a parameter. This problem is identical in structure to the problem (6.7) that we have just solved, thus we can rearrange Theorem 6.3.2 and Lemma 6.3.3 and write out the solution by simply renaming variables. By repeating the dynamic programming recursion $N + 1$ times we derive the following result.

Theorem 6.3.4. *Consider the stochastic uncertain system (6.1)-(6.3), the model mismatch constraints (6.4) and the cost functional (6.5). Introduce the risk-sensitive-like Riccati recursion*

$$\Pi_t = Q + A^\top \left(\Pi_{t+1}^{-1} - \frac{V}{\tau_{t+1}} \right)^{-1} A, \quad t = N, \dots, 0. \quad (6.13)$$

with $\Pi_{N+1} = Q_{N+1}$, which is required to satisfy the condition

$$\left(\Pi_{t+1}^{-1} - \frac{V}{\tau_{t+1}} \right)^{-1} \succ 0 \quad t = N, \dots, 0. \quad (6.14)$$

The minimax optimal control problem (6.6) is finite if and only if there exist τ_t 's strictly positive such that (6.14) is satisfied. In this a case, Problem (6.6) is equivalent to

$$\min_{\tau_1 > 0, \dots, \tau_{N+1} > 0} W_1(\tau_1, \dots, \tau_{N+1}) := \frac{1}{2} \bar{x}_0^\top \Pi_0 \bar{x}_0 - \sum_{t=1}^{N+1} \left(\frac{\tau_t}{2} \log |I - \Pi_t \frac{V}{\tau_t}| - \tau_t d_t \right). \quad (6.15)$$

To prove Theorem 6.3.4, we need the following preliminary lemma.

Lemma 6.3.5. *For any probability density $\tilde{f}(x_{0:t})$, the function $W_{t+1}(\tau_{t+1}, \dots, \tau_{N+1})$ defined as*

$$W_{t+1} := \tilde{\mathbb{E}} \left[\sum_{k=0}^{t-1} \frac{x_k^\top Q x_k}{2} + \frac{x_t^\top \Pi_t x_t}{2} \right] - \sum_{k=t+1}^{N+1} \left(\frac{\tau_k}{2} \log |I - \frac{\Pi_k V}{\tau_k}| - \tau_k d_k \right)$$

with Π_t 's given by (6.13) is jointly convex in $(\tau_{t+1}, \dots, \tau_{N+1})$ for any $t = 0, \dots, N$.

Proof. The function $W_{t+1}(\tau_{t+1}, \dots, \tau_{N+1})$ can be equivalently written as

$$W_{t+1} = \max_{\tilde{f}(x_{t+1}|x_t) \dots \tilde{f}(x_{N+1}|x_N) \in \mathcal{P}} \tilde{\mathbb{E}} \left[J - \sum_{k=t+1}^{N+1} \tau_k (R_k - d_k) \right].$$

Thus, since W_{t+1} is the pointwise maximum of a set of affine functions in the variables $\tau_{t+1}, \dots, \tau_{N+1}$, it is convex by Lemma 6.1.3. ■

Proof of Theorem (6.3.4). The proof follows from the line of reasoning leading up to the theorem statement. The only necessary clarification is that at each time step t , with $t = 0, \dots, N$,

$$\max_{\tilde{f}(x_t|x_{t-1}) \in \mathcal{B}_t} \min_{\tau_{t+1}, \dots, \tau_{N+1} > 0} W_{t+1} = \min_{\tau_{t+1}, \dots, \tau_{N+1} > 0} \max_{\tilde{f}(x_t|x_{t-1}) \in \mathcal{B}_t} W_{t+1}$$

for any given $\tilde{f}(x_{0:t-1})$. This follows from Theorem 6.1.5 since the function W_{t+1} is linear in $\tilde{f}(x_t|x_{t-1})$ and convex in $(\tau_{t+1}, \dots, \tau_{N+1})$ by Lemma 6.3.5. ■

We immediately recognize that our solution has the same structure of the solution to the robust performance analysis problem considered in [96, Section IV.B]. However, instead of using a constant Lagrange multiplier τ , here τ_t is time-dependent. Indeed, whereas in the previous work [96] a single relative entropy constrained was imposed, resulting in a unique Lagrange multiplier, here each $\tilde{f}(x_{t+1}|x_t)$ has an associated relative entropy constraint. Moreover, by writing $\theta_t = \tau_t^{-1}$, we easily see that the solution to our problem takes the form of a risk-sensitive cost (see the state feedback control results given in [72] in the special case where there is no control input). However, while in

standard risk-sensitive control problems the risk-sensitive parameter θ appearing in the exponential of the quadratic loss function is constant, here θ_t is time-dependent. As a matter of fact, in the present paper we tackle the robust analysis problem by evaluating $N + 1$ risk-sensitive costs with a time-varying risk-sensitive parameter, namely

$$\int \exp \left\{ \frac{1}{2\tau_{t+1}} x_t^\top \Pi_t x_t \right\} f(x_t | x_{t-1}) dx_t \quad t = N + 1, \dots, 1.$$

As a result we obtain a risk-sensitive Riccati recursion with a time-dependent sensitivity parameter $1/\tau_t$.

6.3.1 Least favorable model

The solution scheme presented in the previous section does not require an explicit computation of the maximizing players $f^\circ(x_{t+1}|x_t)$'s of Problem (6.6). However, for simulations and performance evaluation, it is instructive to construct the least favorable model. Let $\tau_1^\circ, \dots, \tau_{N+1}^\circ$ be the solutions to Problem (6.15) and $\Pi_0^\circ, \dots, \Pi_{N+1}^\circ$ be the matrices obtained from the Riccati recursion (6.13) with $\tau_t = \tau_t^\circ$. Note that in the above calculations we leveraged on Lemma 1.1.3 in order to evaluate the worst scenario for each time step t . As a consequence, the least favorable model is given by

$$\tilde{f}^\circ(x_t | x_{t-1}) = \frac{\exp \left\{ \frac{1}{2\tau_t^\circ} x_t^\top \Pi_t^\circ x_t \right\} f(x_t | x_{t-1})}{\int \exp \left\{ \frac{1}{2\tau_t^\circ} \bar{x}_t^\top \Pi_t^\circ \bar{x}_t \right\} f(\bar{x}_t | x_{t-1}) d\bar{x}_t}.$$

By Lemma 6.1.2, we conclude that $\tilde{f}^\circ(x_t | x_{t-1}) = \mathcal{N}(\tilde{\mu}_t, \tilde{\Sigma}_t)$ with

$$\tilde{\mu}_t = \left(I - \frac{\Pi_t^\circ V}{\tau_t^\circ} \right)^{-1} A x_{t-1}, \quad \tilde{\Sigma}_t = \left(I - \frac{\Pi_t^\circ V}{\tau_t^\circ} \right)^{-1} V. \quad (6.16)$$

Therefore, the least-favorable density of the noise \tilde{v}_t is itself again Gaussian, but it involves a perturbation of both the mean and the variance with respect to the nominal distribution.

6.4 Simulations

We present a numerical example in order to point out the differences between the uncertain system presented in this chapter and the one proposed by Petersen, James and Dupuis in [96]. Consider the system (6.1) with $A = 1.3$ and initial condition $\bar{x}_0 = 0$ over the finite time interval $[0, N + 1]$ with $N = 49$; the reference noise signal is assumed to be a white Gaussian noise with covariance $V = 10^{-4}$; the quadratic cost functional (6.5) is

defined by $Q = 2.5 \cdot 10^{-4}$ and $Q_{N+1} = 5 \cdot 10^{-5}$; finally we set the tolerance parameters $d_t = 10^{-2}$ for all $t = 1, \dots, N$. We adopt the coordinate descend algorithm [112] to solve Problem (6.15). At each iteration of the algorithm, we have to solve a single variable optimization problem in τ_t . In order to do this, we solve the Riccati equations (6.13) for different values of τ_t , we then evaluate the quantity $W(\tau_1, \dots, \tau_{N+1})$ as a function of τ_t and we choose the value of the variable for which the minimum is achieved. By applying the aforementioned algorithm we find that the solution to the worst performance analysis problem (6.6) is $1.28 \cdot 10^4$. It is interesting to compare this value with the result obtained by considering the stochastic uncertain system where a single relative-entropy constraint is applied, *i.e.* where the joint probability density function $\tilde{f}(x_{0:N+1})$ is assumed to satisfy

$$R(\tilde{f}(x_{0:N+1})||f(x_{0:N+1})) \leq d, \quad (6.17)$$

where

$$d = \sum_{t=1}^{N+1} d_t \quad (6.18)$$

is set to $d = 0.5$. The procedure to solve this problem is presented by Petersen, James and Dupuis in [96, Section IV.B]. We find that the result is equal to $2.33 \cdot 10^4$, which is considerably higher than ours. Figure 6.1 gives a deeper insight to the difference between the two models. It compares the relative entropy distance

$$\int_{\mathbb{R}^n} R(\tilde{f}^o(x_{t+1}|x_t)||f(x_{t+1}|x_t)) \tilde{f}^o(x_t) dx_t$$

at each time step for the worst model obtained by applying the single relative entropy constraint (6.17) and the worst model for which the uncertainty is expressed incrementally (6.4). It is evident that in the standard model the maximizing player allocates most of the mismatch modeling budget to the first time intervals. This leads to unrealistic and extremely pessimistic conclusions. On the other hand, imposing a relative entropy constraint to each time step promotes an equal distribution of the uncertainty during the time interval, which is more reasonable in most practical situations.

It is also significant to analyze the uncertainty distribution over the interval $[1, N + 1]$ of the two least-favourable models for progressively larger values of the tolerance parameter d_t (the constant d is updated according to the usual relation (6.18)). Figure 6.2 illustrates this comparison for $d_t = 0.002$, $d_t = 0.01$ and $d_t = 0.02$. The figures reveal that Petersen's, James' and Dupuis' least favourable model considerably changes as d_t increases, with a concentration of the model mismatch budget in few time steps that becomes more and more evident. On the other hand, our model is less sensitive to parameter variations.

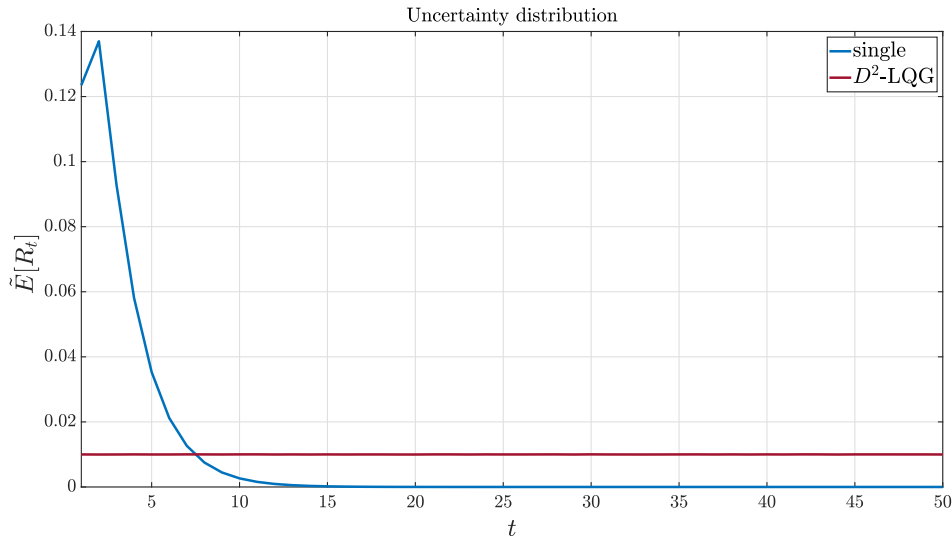


Figure 6.1: Relative entropy distance $\int_{\mathbb{R}^n} R(\tilde{f}^o(x_{t+1}|x_t)||f(x_{t+1}|x_t))\tilde{f}^o(x_t)dx_t$ between the worst case conditional distribution $\tilde{f}^o(x_{t+1}|x_t)$ and the nominal model $f(x_{t+1}|x_t)$ for $t = 1, \dots, 50$. We compare the case in which the worst model is obtained by considering a single relative-entropy constraint (6.17) for the whole time interval (in blue) and the case in which a distributed relative entropy constraint (6.4) is considered (in red).

Finally, we remark that the strategy proposed in this paper gives the opportunity to set a time-varying tolerance level d_t . This can be useful if we know that the modeler has not exercised the same level of effort to characterize each time component of the model.

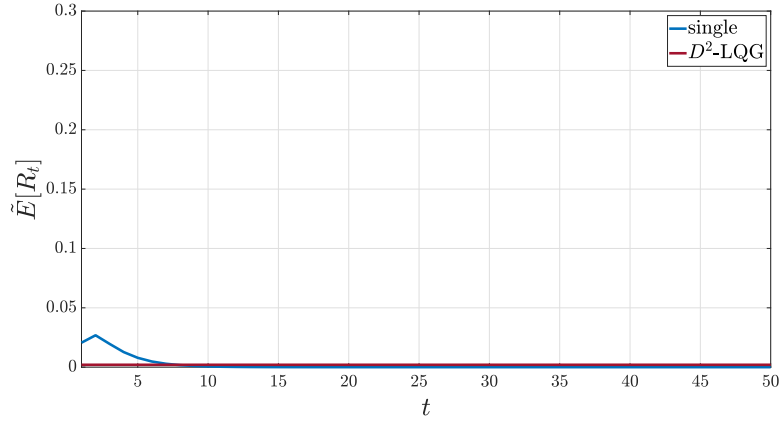
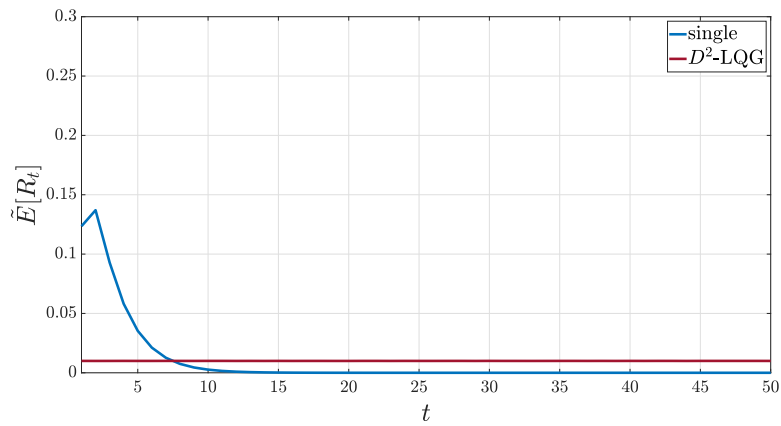
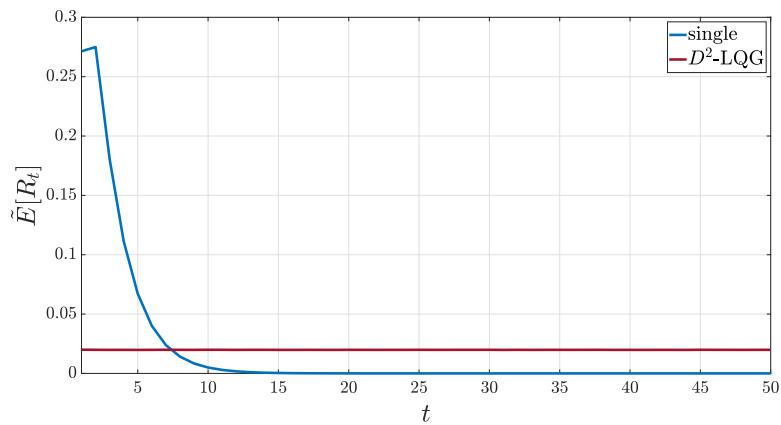
(a) $d_t = 0.002$ (b) $d_t = 0.01$ (c) $d_t = 0.02$

Figure 6.2: Relative entropy distance $\int_{\mathbb{R}^n} R(\tilde{f}^o(x_{t+1}|x_t)||f(x_{t+1}|x_t))\tilde{f}^o(x_t)dx_t$ between the worst case conditional distribution $\tilde{f}^o(x_{t+1}|x_t)$ and the nominal model $f(x_{t+1}|x_t)$ for $t = 1, \dots, 50$ by considering the single relative-entropy constraint (6.17) (in blue) and the distributed relative entropy constraints (6.4) (in red) for increasing values of the tolerance parameter d_t .

7

D^2 -LQG: state feedback control

This chapter extends the worst-performance analysis problem presented in Chapter 6 to the problem of synthesizing a state-feedback D^2 -LQG control policy. The main contributions of this chapter can be summarized as follows. We propose a generalization of the stochastic uncertain systems considered in Chapter 6 which allows for more flexibility in modelling the system uncertainty. The uncertain systems are described in terms of a nominal system, driven by white Gaussian noise, and a perturbed system, in which a general class of stochastic noise processes are allowed. At each time instant, a relative entropy tolerance is used to model a limit on the admissible noise distributions. For this class of uncertain systems, we propose a finite-horizon state-feedback D^2 -LQG control problem. To address this problem, we adopt a dynamic programming technique, where each subproblem consists in an infinite-dimensional minimax optimization problem over a time interval of length one. By exploiting the Lagrange duality theory and the duality between free-energy and relative entropy, we are able to convert this problem into an equivalent risk-sensitive LQG optimal control problem with full-state information which can be solved by using existing results in [72]. The overall solution takes the form of a risk-sensitive LQG controller with a time-varying risk-sensitive parameter.

The outline of the chapter is as follows. Section 7.1 introduces the class of stochastic uncertain systems and the relative-entropy ambiguity set, providing a generalization of the uncertain systems considered in Chapter 6. The proposed uncertain system is interpreted in terms of the linear fractional transformation (LFT) model in Subsection 7.1.1. In Section 7.2 we address the problem of synthesizing the D^2 -LQG controller with state feedback. In Section 7.3 we present some simulations to show the effectiveness of the proposed control strategy in trading off optimality and robustness.

7.1 Stochastic uncertain Systems

We extend to actuated systems and we generalize the class of uncertain systems defined in Chapter 6.

1) *Reference system*: The reference system is described by

$$x_{t+1} = Ax_t + Bu_t + v_t \quad t = 0, \dots, N \quad (7.1)$$

where $x_t \in \mathbb{R}^n$ is the state vector, $u_t \in \mathbb{R}^m$ is the control input and $v = \{v_t \in \mathbb{R}^n, t = 0, \dots, N\}$ is a zero-mean, white Gaussian noise sequence with covariance matrix $V \succ 0$. The matrices A, B and V have appropriate dimensions, and, for the ease of notation, they are assumed time independent. The initial state $x_0 = \bar{x}_0$ is assumed to be a deterministic quantity.

2) *Perturbed system*: The corresponding perturbed system is modelled as

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + \tilde{v}_t, & t = 0, \dots, N \\ z_t &= E_1 x_t + E_2 u_t \end{aligned} \quad (7.2)$$

where the noise input sequence $\tilde{v} = \{\tilde{v}_t, t = 0, \dots, N\}$ is defined by an unknown probability density function. The signal $z_t \in \mathbb{R}^p$ is introduced to add more flexibility in modelling the system's uncertainty; it is used to define the set of allowable uncertain noise probability measures via the relative entropy constraints. One possible interpretation of this signal will be given at the end of the section. The matrices E_1 and E_2 are known matrices which, for ease of computation, are assumed to satisfy the relation $E_1^\top E_2 = 0$. Our results can be easily extended to the case in which we drop this assumption.

3) *Admissible Controller*: We suppose that the whole state is accessible with negligible measure noise and we consider state feedback control policies of the form

$$u_t = \pi_t(x_t), \quad t = 0, \dots, N \quad (7.3)$$

where π_t is a measurable map from \mathbb{R}^n to \mathbb{R}^m . As discussed later, if we enlarge the set of feasible control policies by allowing π_t to depend on the the whole past history of the state, the optimal solution remains the same. In other words, there is no limitation in considering only the control policies (7.3). This is indeed rather intuitive if we consider the Markovian property of the state.

For any admissible control law $\pi = \{\pi_t, t = 0, \dots, N\}$, the nominal joint probability

density function of the state sequence $x_{0:N+1} = [x_0^\top \dots x_{N+1}^\top]^\top$ can be expressed as

$$f_\pi(x_{0:N+1}) = \delta(x_0 - \bar{x}_0) \prod_{t=1}^{N+1} f_\pi(x_t|x_{t-1}) \quad (7.4)$$

where $f_\pi(x_t|x_{t-1}) = \mathcal{N}(Ax_{t-1} + Bu_{t-1}, V)$. We consider a similar decomposition for the joint probability measure $\tilde{f}_\pi(x_{0:N+1})$ of the perturbed system,

$$\tilde{f}_\pi(x_{0:N+1}) = \delta(x_0 - \bar{x}_0) \prod_{t=1}^{N+1} \tilde{f}_\pi(x_t|x_{t-1}).$$

Hereafter, we use the symbols $\mathbb{E}_\pi[\cdot]$ and $\tilde{\mathbb{E}}_\pi[\cdot]$ to denote the expectation with respect to the nominal conditional density $f_\pi(\cdot)$ and the perturbed density $\tilde{f}_\pi(\cdot)$, respectively, given the control law π .

3) *Relative Entropy Constraints:* Given an admissible control law π , we measure the model mismatch at time t between the perturbed and the nominal systems in terms of the relative entropy distance between $\tilde{f}_\pi(x_t|x_{t-1})$ and $f_\pi(x_t|x_{t-1})$,

$$R_t := R_t(\tilde{f}_\pi(x_t|x_{t-1})||f_\pi(x_t|x_{t-1})) = \int_{\mathbb{R}^n} \ln \left(\frac{\tilde{f}_\pi(x_t|x_{t-1})}{f_\pi(x_t|x_{t-1})} \right) \tilde{f}_\pi(x_t|x_{t-1}) dx_t.$$

Then, the admissible perturbed density functions $\tilde{f}_\pi(x_t|x_{t-1})$'s satisfy

$$\tilde{\mathbb{E}}_\pi[R_t] \leq \tilde{\mathbb{E}}_\pi \left[\frac{1}{2} \|z_{t-1}\|^2 \right] + d_t \quad (7.5)$$

where the expectation is taken with respect to the true marginal

$$\tilde{f}_\pi(x_{t-1}) = \int \tilde{f}_\pi(x_{t-1}|x_{t-2}) \dots \tilde{f}_\pi(x_1|\bar{x}_0) dx_1 \dots dx_{t-1}.$$

The ambiguity set is defined by the convex ball of functions

$$\mathcal{B}_t := \{ \tilde{f}_\pi(x_t|x_{t-1}) \in \mathcal{P} : \tilde{\mathbb{E}}_\pi[R_t] \leq \tilde{\mathbb{E}}_\pi \left[\frac{1}{2} \|z_{t-1}\|^2 \right] + d_t \}$$

where \mathcal{P} is the set of all probability density functions over \mathbb{R}^n . The tolerance parameter $d_t > 0$ quantifies the constant mismatch budget allowed at time t . Clearly, \mathcal{B}_t is parameterized by the density function $\tilde{f}_\pi(x_{t-1})$. It is worth noting that the model uncertainty is spread over the finite interval, unlike [96] where the model uncertainty is characterized using one relative entropy constraint regarding the entire interval. The latter characterization allows to concentrate the uncertainty in one step, which is not a

common situation in practice. In contrast, the former promotes an equal distribution of uncertainty along the time interval.

7.1.1 Relation with the linear fractional transformation model

We can motivate the above stochastic uncertainty description by interpreting it in terms of LFT models [98]. LFT is a very general and powerful tool to represent various sources of uncertainty in dynamical models, frequently used in robust control theory. It consists in separating the nominal model from the uncertainty in a feedback interconnection as shown in Figure 7.1. The uncertainty operator Δ is a quantity which is typically unknown but bounded in magnitude. This is a very general paradigm both conceptually and practically. In fact, in most cases the uncertainty on the system dynamics can be effectively modelled by suitably selecting the admissible structure of Δ and the way in which Δ is bounded in magnitude. We can interpret the signal z_t in (7.2) as the input of the uncertain block Δ in the LFT model. Then the uncertain system (7.1)-(7.2)-(7.5) allows for any uncertainty block Δ whose norm is bounded. Notice that Δ can be dynamic and time-varying: in this case the notation Δ_t is adopted.

Lemma 7.1.1. *Consider the stochastic uncertain system (7.1)-(7.2) with control input (7.3). Assume that the perturbed noise signal \tilde{v} is generated according to the LFT model of Figure 7.1 where the uncertainty block is a time-varying matrix $\Delta_t \in \mathbb{R}^{n \times p}$, i.e. $\tilde{v}_t = v_t + \bar{v}_t$ with $\bar{v}_t = v_t + \Delta_t z_t$ for $t = 0, \dots, N$. Then, the relative-entropy constraint (7.5) is satisfied at each time step t if*

$$\Delta_t \Delta_t^\top \preceq V, \quad t = 0, \dots, N. \quad (7.6)$$

Proof. From the definition of \tilde{v} , it follows that $\tilde{f}_\pi(x_{t+1}|x_t) = \mathcal{N}(Ax_t + Bu_t + \Delta_t z_t, V)$. Hence, given $x_t \in \mathbb{R}^n$,

$$R_{t+1} = \frac{1}{2} z_t^\top \Delta_t^\top V^{-1} \Delta_t z_t \leq z_t^\top z_t,$$

where the equality follows from Lemma 1.1.2 and the inequality from the norm bound (7.6). Now, the thesis is obtained by taking into account the monotonicity of the expected value operator and the positivity of d_{t+1} . ■

In conclusion, our paradigm appears to be the natural framework to address the robustification of LQG control solutions in all the situations in which the uncertainty may be represented as in Figure 7.1. The main advantage with respect to existing literature on DRC with relative entropy is that our formulation is ideally suited to capture the essence of the problem whenever the uncertainty is distributed along the whole time interval.

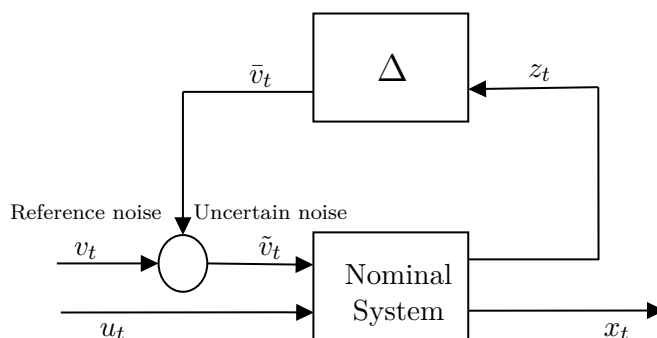


Figure 7.1: Linear fractional transformation (LTF) uncertain model.

7.2 Minimax optimal control

We consider the problem of constructing a state feedback controller which minimizes the worst case performance for the above stochastic uncertain system. The main result of this section is to show that this problem can be converted into an equivalent state-feedback risk-sensitive LQG optimal control problem with a time varying risk-sensitive parameter.

We consider the class of stochastic uncertain systems introduced in Section 7.1. The performance index is the standard symmetric quadratic function of state and control action

$$J = \frac{1}{2} \sum_{t=0}^N (x_t^\top Q x_t + u_t^\top R u_t) + \frac{1}{2} x_{N+1}^\top Q_{N+1} x_{N+1}, \quad (7.7)$$

where the matrices $Q \succeq 0$, $Q_{N+1} \succ 0$ and $R \succ 0$ are designed according to the control specifications. This index measures the state deviation from the origin at the final time step and during the transient phase, as well as the energy spent for the control. The lower the index takes value, the more the control strategy satisfies the specifications. We are concerned with the control problem

$$\min_{\{u_t = \pi_t(x_t), t=0, \dots, N\}} \max_{\{\tilde{f}(x_t | x_{t-1}) \in \mathcal{B}_t, t=1, \dots, N+1\}} \tilde{\mathbb{E}}_\pi[J]. \quad (7.8)$$

Assumption 7.2.1. The pair (A, Q) is observable.

Similarly to Chapter 6, we address this problem by adopting the dynamic programming technique. Consider the situation at time N : the state variable x_N has been observed and the control signal $u_N = \pi_N(x_N)$ should be determined by solving the minimax problem

$$\min_{u_N = \pi_N(x_N)} \max_{\tilde{f}(x_{N+1} | x_N) \in \mathcal{B}_{N+1}} \tilde{\mathbb{E}}_\pi[J] \quad (7.9)$$

where the density function $\tilde{f}(x_{0:N})$ and the input sequence $u_{0:N-1} := [u_0^\top; \dots; u_{N-1}^\top]^\top$ are considered as parameters. Problem (7.9) can be replaced by an equivalent unconstrained stochastic game by means of the duality theory.

Theorem 7.2.2. *Consider the uncertain system (7.1)-(7.5) with cost functional (7.7). Assume that $\tilde{f}(x_{0:N})$ and $u_{0:N-1}$ are fixed and let*

$$W_{N+1}(\tau_{N+1}) := \min_{u_N = \pi_N(x_N)} \max_{\tilde{f}(x_{N+1}|x_N) \in \mathcal{P}} \tilde{\mathbb{E}}_\pi[L_{N+1}(\tau_{N+1})], \quad (7.10)$$

where $L_{N+1}(\tau_{N+1}) := J - \tau_{N+1}[R_{N+1} - \frac{1}{2}\|z_N\|^2 - d_{N+1}]$ and $\tau_{N+1} \in \mathbb{R}$ is the Lagrange multiplier. Problem (7.9) is finite if and only if $\Gamma_{N+1} := \{\tau_{N+1} > 0 \mid W_{N+1}(\tau_{N+1}) < \infty\}$ is nonempty. In this case,

$$\min_{u_N = \pi_N(x_N)} \max_{\tilde{f}(x_{N+1}|x_N) \in \mathcal{B}_{N+1}} \tilde{\mathbb{E}}_\pi[J] = \min_{\tau_{N+1} > 0} W_{N+1}(\tau_{N+1}). \quad (7.11)$$

The proof follows similar steps to the proof of Theorem 6.3.2 of Section 6 and it is therefore omitted.

Now, we evaluate the quantity $W_{N+1}(\tau_{N+1})$ for $\tau_{N+1} > 0$. We have that

$$W_{N+1} = \min_{u_N} \tilde{\mathbb{E}}_\pi \left[\frac{1}{2} \sum_{t=0}^{N-1} (x_t^\top Q x_t + u_t^\top R u_t) + \tau_{N+1} \left(\max_{\tilde{f}(x_{N+1}|x_N)} \tilde{\mathbb{E}} \left[\frac{1}{2\tau_{N+1}} \left(x_N^\top \bar{Q}_{\tau_{N+1}} x_N + u_N^\top \bar{R}_{\tau_{N+1}} u_N + x_{N+1}^\top Q_{N+1} x_{N+1} \right) - R_{N+1} \mid x_N \right] \right) + \tau_{N+1} d_{N+1}, \right]$$

where $\bar{Q}_{\tau_{N+1}} := Q + \tau_{N+1} E_1^\top E_1$ and $\bar{R}_{\tau_{N+1}} := R + \tau_{N+1} E_2^\top E_2$. By applying Lemma 1.1.3 it follows that

$$W_{N+1} = \min_{u_N} \tilde{\mathbb{E}} \left[\frac{1}{2} \sum_{t=0}^{N-1} (x_t^\top Q x_t + u_t^\top R u_t) + \tau_{N+1} \log \mathbb{E} \left[J_{RS,N+1} \mid x_N \right] \right] + \tau_{N+1} d_{N+1},$$

where

$$J_{RS,N+1} := \exp \left\{ \frac{x_N^\top \bar{Q}_{\tau_{N+1}} x_N + u_N^\top \bar{R}_{\tau_{N+1}} u_N + x_{N+1}^\top Q_{N+1} x_{N+1}}{2\tau_{N+1}} \right\}.$$

Here the subscript RS stands for *risk-sensitive*. Indeed, this is a risk-sensitive criterion with risk-sensitivity parameter $1/\tau_{N+1}$. From Lemma 6.1.7, we can switch the minimization

with respect to $u_N = \pi_N(x_N)$ and the expectation with respect to x_N so that

$$W_{N+1} = \tilde{\mathbb{E}}_\pi \left[\frac{1}{2} \sum_{t=0}^{N-1} (x_t^\top Q x_t + u_t^\top R u_t) + \tau_{N+1} \min_{u_N} \log \mathbb{E}_\pi \left[J_{RS,N+1} | x_N \right] \right] + \tau_{N+1} d_{N+1}. \quad (7.12)$$

Notice that the problem $\min_{u_N} \mathbb{E}_\pi \log[J_{RS,N+1} | x_N]$ is a one-step risk-sensitive LQG control problem with state feedback and initial condition x_N . The solution to this problem is obtained by Jacobson [72] and involves the Riccati recursion

$$\Pi_N = \bar{Q}_{\tau_{N+1}} + A^\top \left(\Pi_{N+1}^{-1} + B \bar{R}_{\tau_{N+1}}^{-1} B^\top - \frac{V}{\tau_{N+1}} \right)^{-1} A \quad (7.13)$$

with $\Pi_{N+1} = Q_{N+1}$, which is required to satisfy the condition

$$(\Pi_{N+1}^{-1} - V/\tau_{N+1}) \succ 0. \quad (7.14)$$

Then, the optimal controller is

$$u_N = -\bar{R}_{\tau_{N+1}}^{-1} B^\top \left(\Pi_{N+1}^{-1} + B \bar{R}_{\tau_{N+1}}^{-1} B^\top - \frac{V}{\tau_{N+1}} \right)^{-1} A x_N$$

and, accordingly,

$$\min_{u_N} \log \mathbb{E}_\pi \left[J_{RS,N+1} | x_N \right] = \frac{x_N^\top \Pi_N x_N}{2\tau_{N+1}} - \frac{1}{2} \log \left| I - \frac{\Pi_{N+1} V}{\tau_{N+1}} \right|. \quad (7.15)$$

Lemma 7.2.3. *If (7.14) is satisfied, then Π_N defined by (7.13) is positive definite.*

The proof follows similar steps to Lemma 6.3.3 and it is therefore omitted.

By plugging (7.15) into (7.12), we obtain that for $\tau_{N+1} \in \Gamma_{N+1}$,

$$W_{N+1} = \tilde{\mathbb{E}}_\pi \left[\sum_{t=0}^{N-1} \frac{x_t^\top Q x_t + u_t^\top R u_t}{2} + \frac{x_N^\top \Pi_N x_N}{2} \right] - \frac{\tau_{N+1}}{2} \log \left| I - \frac{\Pi_{N+1} V}{\tau_{N+1}} \right| + \tau_{N+1} d_{N+1}. \quad (7.16)$$

In view of (7.11), if Γ_{N+1} is non-empty, Problem (7.8) can be rewritten as

$$\min_{\{u_t = \pi_t(x_t), t=0, \dots, N-1\}} \max_{\{\tilde{f}_\pi(x_t | x_{t-1}) \in \mathcal{B}_t, t=1, \dots, N\}} \min_{\tau_{N+1} > 0} W_{N+1}(\tau_{N+1}).$$

Now, in order to move to the next stage of the dynamical programming recursion, we switch the minimization over τ_{N+1} and the maximization over $\tilde{f}(x_N | x_{N-1})$, so that we

can consider the problem

$$\min_{u_{N-1}=\pi_{N-1}(x_{N-1})} \max_{\tilde{f}_\pi(x_N|x_{N-1}) \in \mathcal{B}_N} \tilde{\mathbb{E}}_\pi[W_{N+1}(\tau_{N+1})].$$

By looking at expression (7.16), we immediately recognize that this problem is identical in structure to Problem (7.9) (except for some terms which are constant with respect to $\tilde{f}(x_N|x_{N-1})$ and u_{N-1}). Therefore, we can rearrange Theorem 7.2.2 and repeat all the previous reasoning to tackle it. We hasten to remark that the switch between the minimization and the maximization operators may in general affect the result. Sufficient conditions to perform this switch without changing the result are provided in Theorem 6.1.5. In order to apply this theorem, we notice that linearity (thus concavity) of the function W_{N+1} with respect to $\tilde{f}_\pi(x_N|x_{N-1})$ for any fixed τ_{N+1} immediately follows from (7.16). As regards convexity with respect to τ_{N+1} given $\tilde{f}_\pi(x_N|x_{N-1})$, we have to distinguish two cases: when the matrix E_2 in (7.2) is equal to zero and when it is different from zero. We first analyze the case $E_2 = 0$.

Theorem 7.2.4. *Consider the stochastic uncertain system (7.1)-(7.5) with $E_2 = 0$ and the cost functional (7.7). Let $\Pi_{N+1} = Q_{N+1}$, and introduce the risk sensitive Riccati recursion*

$$\Pi_t = \bar{Q}_{\tau_{t+1}} + A^\top \left(\Pi_{t+1}^{-1} + B \bar{R}_{\tau_{t+1}}^{-1} B^\top - \frac{V}{\tau_{t+1}} \right)^{-1} A, \quad t = N, \dots, 0 \quad (7.17)$$

where

$$\bar{Q}_{\tau_{t+1}} := Q + \tau_{t+1} E_1^\top E_1, \quad (7.18)$$

$$\bar{R}_{\tau_{t+1}} := R, \quad (7.19)$$

provided that

$$\left(\Pi_{t+1}^{-1} - \frac{V}{\tau_{t+1}} \right) \succ 0 \quad t = N, \dots, 0. \quad (7.20)$$

Assume that there exist τ_t 's strictly positive such that (7.20) is satisfied. Then, the minimax optimal control problem (7.8) is finite and it is equivalent to

$$\min_{\tau_1 > 0, \dots, \tau_{N+1} > 0} W_1(\tau_1, \dots, \tau_{N+1}) := \frac{\bar{x}_0^\top \Pi_0 \bar{x}_0}{2} - \sum_{t=1}^{N+1} \left(\frac{\tau_t}{2} \log \left| I - \frac{\Pi_t V}{\tau_t} \right| - \tau_t d_t \right). \quad (7.21)$$

Moreover, if $\tau_1^\circ, \dots, \tau_{N+1}^\circ$ solve Problem (7.21) and $\Pi_0^\circ, \dots, \Pi_{N+1}^\circ$ are the matrices ob-

tained from (7.17) with $\tau_t = \tau_t^\circ$, then the optimal control input for problem (7.8) is

$$u_t^\circ = -\bar{R}_{\tau_{t+1}^\circ}^{-1} B^\top \left(\Pi_{t+1}^\circ{}^{-1} + BR_{\tau_{t+1}^\circ}^{-1} B^\top - \frac{V}{\tau_{t+1}^\circ} \right)^{-1} Ax_t. \quad (7.22)$$

To prove this theorem, we need the following preliminary Lemma.

Lemma 7.2.5. *Assume that $E_2 = 0$. Given the input sequence $u_{0:t}$ and the density function $\tilde{f}_\pi(x_{0:t-1})$, the function $W_{t+1}(\tau_{t+1}, \dots, \tau_{N+1})$ defined as*

$$W_{t+1} := \tilde{\mathbb{E}}_\pi \left[\sum_{k=0}^{t-1} \frac{x_k^\top Q x_k + u_k^\top R u_k}{2} + \frac{x_t^\top \Pi_t x_t}{2} \right] - \sum_{k=t+1}^{N+1} \left(\frac{\tau_k}{2} \log \left| I - \frac{\Pi_k V}{\tau_k} \right| - \tau_k d_k \right), \quad (7.23)$$

with Π_t 's given by (7.17), is jointly convex in $(\tau_{t+1}, \dots, \tau_{N+1})$ for any $t = 0, \dots, N$.

Proof. The function $W_{t+1}(\tau_{t+1}, \dots, \tau_{N+1})$ can be equivalently written as

$$W_{t+1} = \min_{u_t, \dots, u_N} \max_{\tilde{f}(x_{t+1}|x_t), \dots, \tilde{f}(x_{N+1}|x_N) \in \mathcal{P}} \tilde{\mathbb{E}}_\pi \left[J - \sum_{k=t+1}^{N+1} \tau_k (R_k - \frac{1}{2} \|z_k\|^2 - d_k) \right].$$

It is not difficult to see that the objective function in the previous problem is jointly convex in $(\tau_{t+1}, \dots, \tau_{N+1}, u_t, \dots, u_N)$ for any $(\tilde{f}(x_{t+1}|x_t), \dots, \tilde{f}(x_{N+1}|x_N))$ if $E_2 = 0$. Then, by Lemma 6.1.3, the maximization with respect to $(\tilde{f}(x_{t+1}|x_t), \dots, \tilde{f}(x_{N+1}|x_N))$ returns a convex function of $(\tau_{t+1}, \dots, \tau_{N+1}, u_t, \dots, u_N)$. Finally, in view of Lemma 6.1.4, the partial minimization over the variables (u_t, \dots, u_N) preserves convexity with respect to $(\tau_{t+1}, \dots, \tau_{N+1})$. \blacksquare

Proof of Theorem 7.2.4. The proof follows from the line of reasoning leading up to the theorem statement, repeating the dynamic programming recursion $N + 1$ times. The main clarification is that at each time step t , with $t = 0, \dots, N$, given $u_{0:t}$ and $\tilde{f}_\pi(x_{0:t-1})$,

$$\max_{\tilde{f}(x_t|x_{t-1}) \in \mathcal{B}_t} \min_{\tau_{t+1}, \dots, \tau_{N+1} > 0} W_{t+1} = \min_{\tau_{t+1}, \dots, \tau_{N+1} > 0} \max_{\tilde{f}(x_t|x_{t-1}) \in \mathcal{B}_t} W_{t+1}, \quad (7.24)$$

The saddle-point equality (7.24) derives from Theorem 6.1.5, since the function W_{t+1} is clearly linear in $\tilde{f}(x_t|x_{t-1})$ and jointly convex in $(\tau_{t+1}, \dots, \tau_{N+1})$ by Lemma 7.2.5. \blacksquare

If $E_2 \neq 0$, we can not use Lemma 7.2.5 to show convexity of W_{t+1} in $(\tau_{t+1}, \dots, \tau_{N+1})$. At each time step t , the switch between the max and the min operator may generate an inequality sign “ \leq ” (see [21, p.238]). Consequently, we have a weaker version of Theorem 7.2.4.

Theorem 7.2.6. Consider the stochastic uncertain system (7.1)-(7.5) with $E_2 \neq 0$ and the cost functional (7.7). Let the matrices $\Pi_t \in \mathbb{R}^{n \times n}$ be computed according to the Riccati difference equation (7.17) where formula (7.19) is substituted with

$$\bar{R}_{\tau_{t+1}} := R + \tau_{t+1} E_2^\top E_2.$$

Assume that there exist τ_t 's strictly positive such that (7.20) is satisfied. Then Problem (6.6) is finite and Problem (7.21) provides an upper bound to it.

Proof. The proof follows from the line of reasoning leading up to the theorem statement, repeating the dynamic programming recursion $N + 1$ times. Note that, differently from Theorem 7.2.4, at each time step t , given $\tilde{f}(x_{0:t-1})$ and $u_{0:t}$, we can only state that

$$\max_{\tilde{f}(x_t|x_{t-1}) \in \mathcal{B}_t} \min_{\tau_{t+1}, \dots, \tau_{N+1} > 0} W_{t+1} \leq \min_{\tau_{t+1}, \dots, \tau_{N+1} > 0} \max_{\tilde{f}(x_t|x_{t-1}) \in \mathcal{B}_t} W_{t+1}. \quad (7.25)$$

Therefore, when we switch the max and min operators to proceed in the dynamic programming recursion, we may in general generate an upper bound to Problem (6.6). ■

We propose a coordinate descend algorithm [112] to solve Problem (7.21). At each iteration of the algorithm, we have to solve a single variable optimization problem in τ_t . In order to do that, we solve the Riccati equations (6.13) for different values of τ_t , then we evaluate the quantity $W_1(\tau_1, \dots, \tau_{N+1})$ as a function of τ_t and we choose the value of the variable for which the minimum is achieved. If $E_2 = 0$, we have theoretical guarantees that Problem (7.21) is a convex optimization problem, thus the coordinate descent algorithm converges to the optimal point.

Remark 7.2.7. If the matrix $E_2 \neq 0$, we have already noticed that we can not extend the proof of Lemma 7.2.5 to show convexity of the function W_{t+1} in $(\tau_{t+1}, \dots, \tau_{N+1})$. Our conjecture, supported by numerical simulations, is that this function is quasi-convex [21] in $(\tau_{t+1}, \dots, \tau_{N+1})$ for each $t = 0, \dots, N$. If this conjecture holds true, then by *Sion's minimax theorem* [109] the inequality sign (7.25) can be substituted with an equality sign, and Problem (7.21) is actually equivalent to (6.6). Thus, if our conjecture holds true, the conclusions of Theorem 7.2.4 hold also for $E_2 \neq 0$. Numerical simulations in practical examples show that our paradigm provides good performances even when $E_2 \neq 0$. Example 1 of Section 7.3 is an example of such simulations.

Remark 7.2.8. We have assumed in Section 7.1 that the admissible control policies (7.22) are functions of the state vector only at the current time step t . We hasten to remark that **this hypothesis is not restrictive**. In fact, we can repeat the previous argument under the assumption that the controller has access to the whole past history $x_{0:t}$. The

computations are essentially the same and the optimal control policy does not change. In other words, no additional information about the future development of the system is obtained if past measurements are included.

7.3 Simulations

We present some numerical example in order to illustrate the theory developed above. We compare the solution obtained by our scheme both to the standard LQG solution and to the solution obtained by employing the state of the art method proposed by Petersen and James [96] to make the LQG solution resilient against model uncertainties.

7.3.1 Example 1

Consider the uncertain system (7.1)-(7.2) of dimension $n = 3$ over an interval of length $N = 100$ with state and input matrices

$$A = \begin{bmatrix} 0.5773 & -0.6335 & -0.0457 \\ 0.5477 & 1.7583 & 0.0524 \\ -0.4011 & -0.4754 & 1.0043 \end{bmatrix}, \quad B = \begin{bmatrix} 0.3212 \\ 0.3689 \\ -0.2741 \end{bmatrix},$$

and nominal noise variance $V = I$. The uncertainty input z_t is a scalar quantity defined by the matrices $E_1 = 0$ and $E_2 = 0.5$. The initial condition is $x_0 = [0.5 \quad 0.1 \quad -0.7]^\top$; the quadratic cost functional (7.7) is defined by the weight matrices

$$Q = 0.0025 \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad Q_{N+1} = 0.1 \cdot I, \quad R = 10^{-3}.$$

We set the tolerance parameters $d_t = 10^{-10}$ for all t in (7.5). We apply the D^2 -LQG procedure proposed in Section 7.2 to control the system. To test the performances, we compute the expected closed-loop cost when the real unknown noise process \tilde{v} is generated according to the LFT model of Figure 7.1. In particular, we assume that $\Delta = c [1 \quad 1 \quad 1]$ is a time-invariant matrix with $c \in \mathbb{R}$ ranging in the interval $[-1/\sqrt{3}, 1/\sqrt{3}]$. This ensures that $\|\Delta\|^2 \leq 1$, and consequently, in view of Lemma 7.1.1, that the relative entropy constraint (7.5) is satisfied at every time t . It is of interest to compare the results with the ones obtained by applying the standard LQG procedure, as well as the controller proposed by Petersen and James in [96, Section IV.B]. The authors of [96] propose a minimax

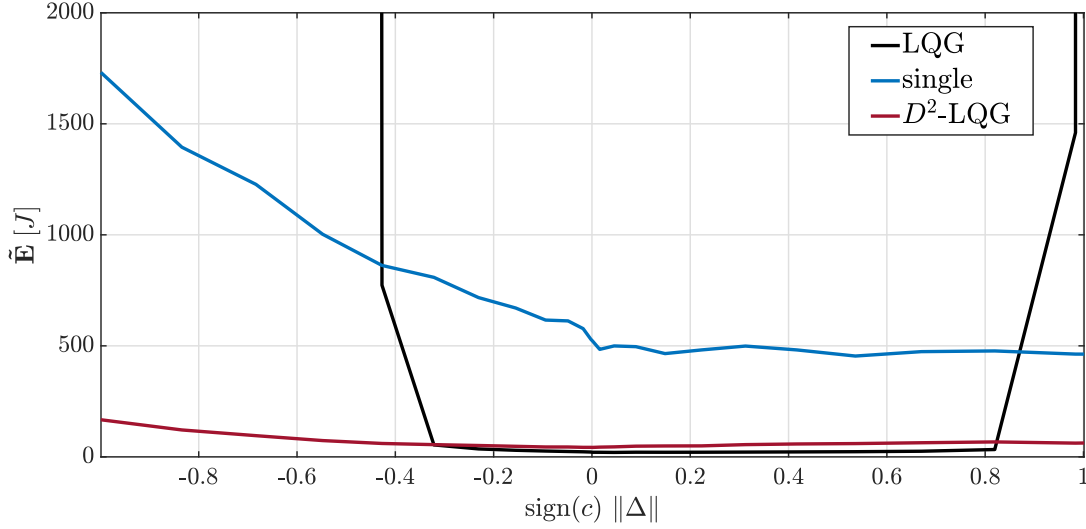


Figure 7.2: Expected closed loop cost versus the real uncertain parameter Δ . We compare the performances of the standard LQG controller (in black), the minimax LQG controller obtained by considering a single relative-entropy constraint (7.26) for the whole time interval (in blue) and the proposed D^2 -LQG controller in which an incremental relative entropy constraint (7.5) is considered (in red).

LQG controller for stochastic uncertain systems where the uncertainty is expressed in terms of a single relative-entropy constraint for the whole time interval. Namely, given the control law π and the nominal Gaussian density function $f_\pi(x_{0,N+1})$, the admissible perturbed density functions $\tilde{f}_\pi(x_{0,N+1})$ are assumed to satisfy the constraint

$$R(\tilde{f}_\pi(x_{0:N+1})||f(x_{0:N+1})) \leq d + \tilde{\mathbb{E}}_\pi \left[\frac{1}{2} \sum_{t=0}^{N+1} \|z_t\|^2 \right]. \quad (7.26)$$

The tolerance d is positive scalar, and, for a fair comparison, we set

$$d = \sum_{t=1}^{N+1} d_t. \quad (7.27)$$

A plot of the expected closed-loop cost (7.7) versus the uncertainty parameter Δ for the three controllers is given in Figure 7.2. When $\Delta = 0$ the perturbed system corresponds to the nominal one, for which the standard LQG controller is the optimal one. However, the performances of the standard LQG controller rapidly deteriorate when the magnitude of Δ increases, causing the cost of the system to increase dramatically. It is also evident that the minimax controller with a single relative entropy constraint [96] does not show a satisfactory behavior. The explanation lies in the fact that the maximizing player is

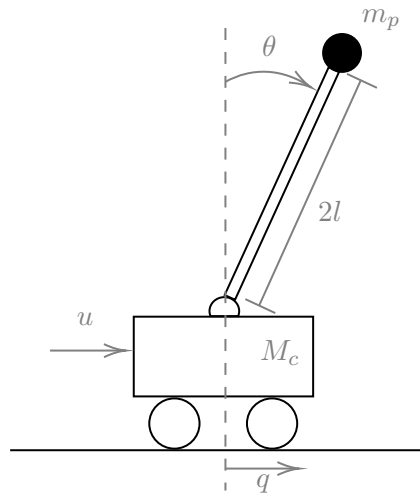


Figure 7.3: Inverted pendulum system.

Symbol	Parameters Name	Value
M_c	Mass of the Cart	3 kg
m_p	Mass of the Pendulum	1.5 kg
l	Half of Pendulum length	2 m
g	Gravity Acceleration	9.8 m/s

Table 7.1: Parameters of the inverted pendulum system.

allowed to allocate most of the mismatch modeling budget to the first time intervals; this leads to extremely pessimistic and conservative conclusions. Finally we notice that the proposed procedure is able to trade off optimality and resiliency. Indeed, when the real system coincides with the nominal one, the expected closed-loop cost is slightly larger than the optimal one. This cost remains almost constant when the perturbation matrix Δ is different from zero, giving evidence to the robustness properties of the control system.

7.3.2 Example 2

We consider the control of the inverted pendulum on a cart, which is a typical benchmark in control theory. The inverted pendulum system is depicted in Figure 7.3. It consists of a pendulum which is attached to a cart equipped with a motor driving it. The movement of the cart is constrained to the horizontal direction, whereas the pendulum can rotate in the vertical plane. The model parameters are listed in Table 7.1. The motion of the

system can be represented by [108]

$$\begin{aligned} m_p l \ddot{\theta} \cos \theta - m_p l \dot{\theta}^2 \sin \theta + (M_c + m_p) \ddot{q} &= u \\ J \ddot{\theta} - m_p l g \sin \theta + m_p l \cos \theta \ddot{q} &= 0, \end{aligned} \quad (7.28)$$

where θ (in rad) is the angle between the vertical axis and the rod of the pendulum, q (in m) is the horizontal displacement of the pendulum and u (in N) is the input acceleration. The quantity $J = (I + ml^2)$ is the pendulum moment of inertia around the pivot. When the rod of the pendulum is close to the unstable equilibrium point $\theta = 0$, any non-linearities in model (7.28) can be neglected. Thus, by defining the state variables $x = [\theta \quad \dot{\theta} \quad q \quad \dot{q}]^\top$, the system can be represented by the linear model [108]

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{3(M_c + m_p)g}{4M_c l + m_p l} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{3m_p g}{4M_c + m_p} & 0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ -\frac{3}{4M_c l + m_p l} \\ 0 \\ \frac{4}{4M_c + m_p} \end{bmatrix} u. \quad (7.29)$$

Assuming that the system lies in the linear region, the aim is to stabilize it at the equilibrium $x = [0 \ 0 \ 0 \ 0]^\top$. To solve this problem, we first discretize the continuous-time model (7.29) using zero-order hold on the input with sample time of $T_s = 0.2$ sec. The resulting state and input matrices are:

$$A = \begin{bmatrix} 1.0996 & 0.2066 & 0 & 0 \\ 1.0123 & 1.0996 & 0 & 0 \\ -0.0664 & -0.0044 & 1 & 0.2 \\ -0.6749 & -0.0664 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -0.0023 \\ -0.0230 \\ 0.0060 \\ 0.0597 \end{bmatrix}.$$

Then, we design a controller by applying the proposed D^2 -LQG control strategy. Specifically, we consider the initial state $x_0 = [0.1 \quad -0.1 \quad 0.05 \quad 0.02]^\top$ and the quadratic cost (7.7) for a time interval of length $N = 75$ with weights $Q = Q_{N+1} = \text{diag}(1, 1, 10, 1)$ and $R = 1$, where $\text{diag}(v)$ denotes the diagonal operator returning a square diagonal matrix

with the elements of vector v on the main diagonal. We set

$$E_1 = \begin{bmatrix} -0.0224 & 0.2295 & -0.1318 & -0.0035 \\ 0.1201 & 0.1169 & 0.0139 & -0.1209 \\ -0.1291 & -0.0602 & 0.2178 & -0.0656 \\ 0.1179 & -0.0579 & -0.09106 & 0.0530 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^\top,$$

and the tolerance $d_t = 0.0132$ for any t , such that $d = \sum d_t = 1$. Finally, v_t is assumed to be a white Gaussian noise with covariance $V = \text{diag}(0.1, 0.5, 0.1, 0.5)$. Figure 7.4 shows the closed-loop system response x and the control action u , together with the value of the cost J defined in (7.7), in the nominal (deterministic) case in which the underlying system is governed by the dynamics

$$x_{t+1} = Ax_t + Bu_t. \quad (7.30)$$

We compare the performances of the linear quadratic regulator, the minimax LQG controller proposed in [96] and the proposed D^2 -LQG strategy. From the figure, it is evident that the proposed procedure exhibits an intermediate behavior between the standard LQG control strategy and the single constraint minimax LQG one.

To evaluate the robustness of the three control schemes to model uncertainties and parameter perturbations, we consider the situation in which model (7.29) is an inaccurate representation of the underlying system. Specifically, we assume that the real underlying system is governed by the dynamics

$$x_{t+1} = (A + \Delta A)x_t + (B + \Delta B)u_t \quad (7.31)$$

with

$$\Delta A = \begin{bmatrix} 0.0269 & 0.0316 & -0.0243 & 0.0288 \\ -0.0296 & -0.0290 & -0.0163 & -0.0312 \\ 0.0332 & 0.0046 & 0.0348 & -0.0093 \\ -0.0246 & -0.0042 & -0.0284 & 0.0138 \end{bmatrix}, \quad \Delta B = \begin{bmatrix} -0.0004 \\ 0.0011 \\ 0.0012 \\ -0.0002 \end{bmatrix}.$$

The results are summarized in Figure 7.5. In this perturbed situation, the system controlled with the LQG technique lacks stability. This causes the state vector x and therefore the cost J to diverge. On the other hand, the robustness introduced by the minimax controllers prevents system instability. We can also observe that the proposed

technique exhibits better performances than the single constraint one, as it achieves a smaller value of the index J . Moreover, it requires the application of a smoother and lower-energy control signal, which is more suitable and realistic in practice.

Remark 7.3.1. The matrices ΔA and ΔB and E_1 defined above are randomly generated. By repeating the simulations for different E_1 , ΔA and ΔB (with approximately the same norms) we obtain similar results.

Our simulation experiments suggest the following conclusions:

1. The standard LQG solution is quite fragile. In fact, the LQG optimal control policy applied to a system featuring modest perturbations with respect to the nominal model may yield an unstable closed-loop system. Both our method and the method proposed in [96] are, on the other hand, resilient to perturbations.
2. Our method to deal with uncertainties appears to be much less conservative than that in [96]. In fact, our performances are not significantly different from the optimal LQG ones for the nominal model while the performance index obtained by using the method in [96] is much larger. Moreover, even when the difference between the nominal and the actual model is significant, robustness against instability is guaranteed by both methods but in our method the value of the performance index is significantly smaller.

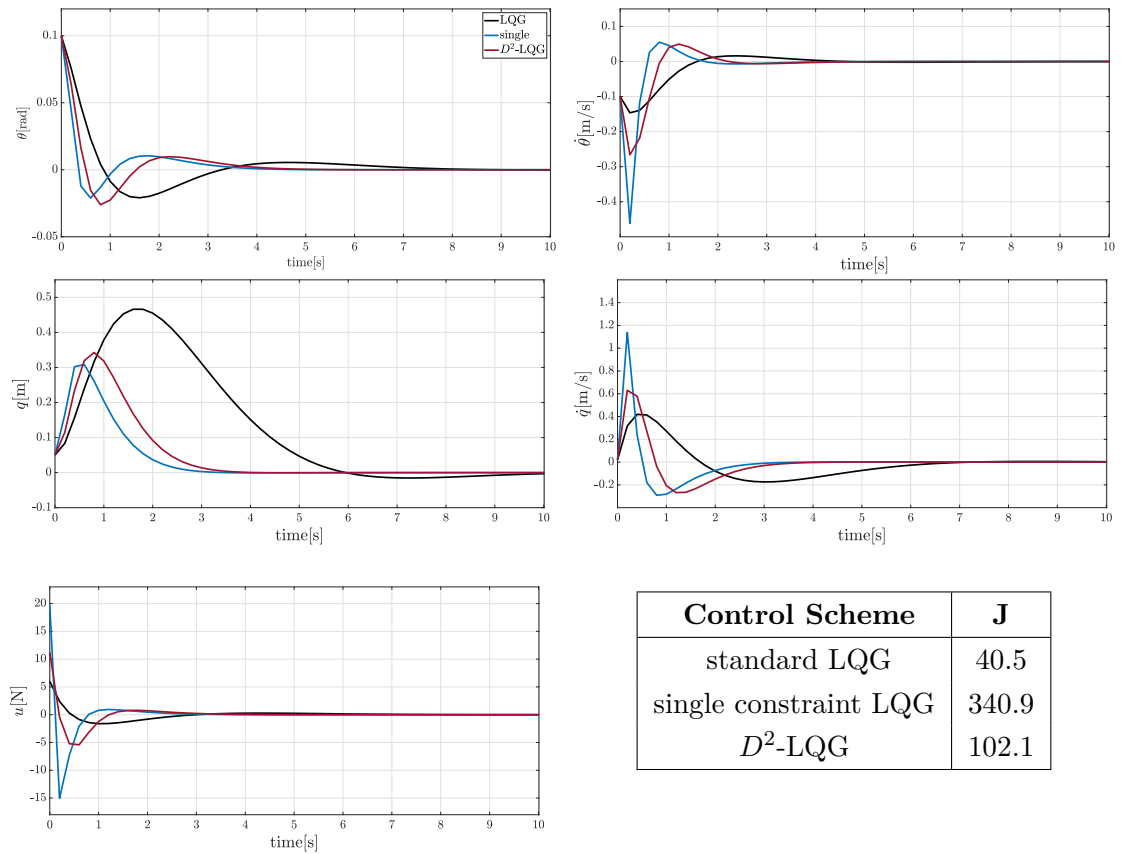


Figure 7.4: Inverted pendulum on a cart system: nominal case in which the underlying system is governed by the dynamics (7.30). Evolution of the closed-loop response x and the control input u over the time interval $[0, 10\text{s}]$, and value of the resulting cost J (7.7). We compare the performances of the standard LQG controller (in black), the minimax LQG controller obtained by considering a single relative-entropy constraint (7.26) for the whole time interval (in blue) and the proposed D^2 -LQG controller in which an incremental relative entropy constraint (7.5) is considered (in red).

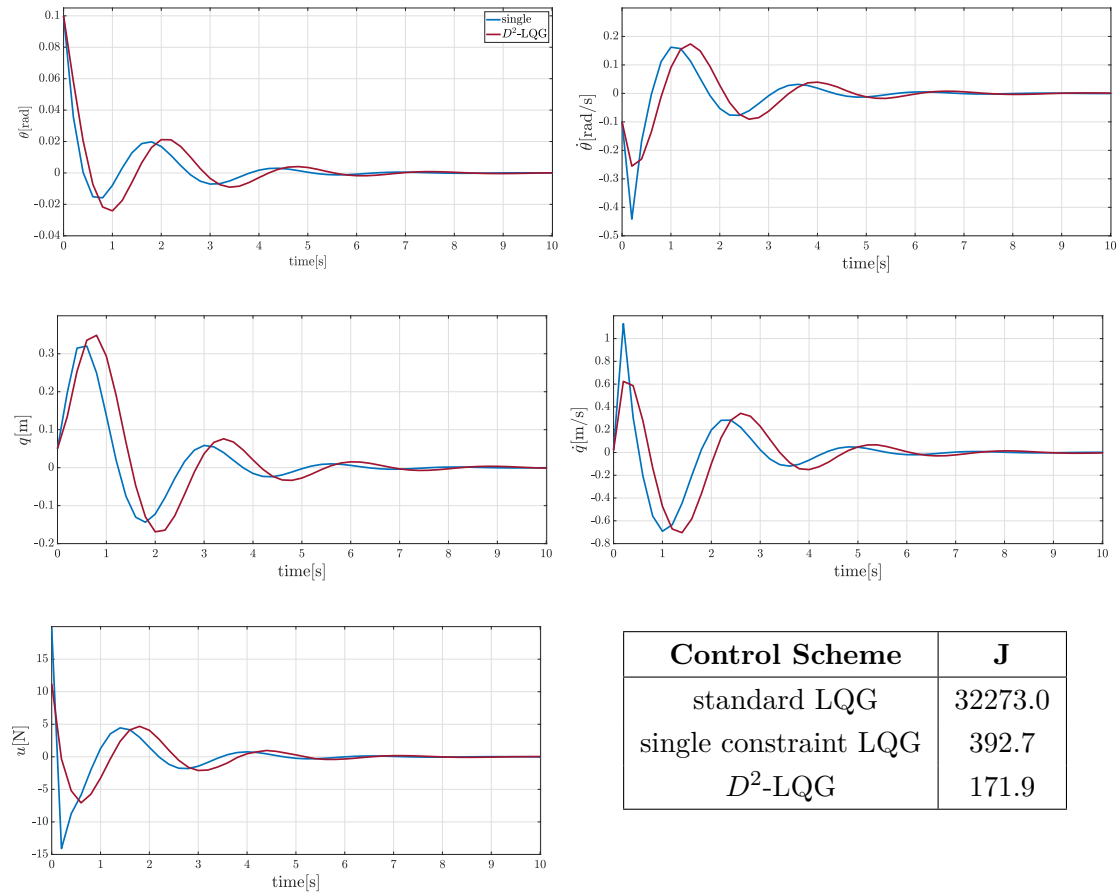


Figure 7.5: Inverted pendulum on a cart system: perturbed case in which the underlying system is governed by the dynamics (7.31). Evolution of the closed-loop response x and the control input u over the time interval $[0, 10\text{s}]$, and value of the resulting cost J (7.7). We compare the performances of the minimax LQG controller obtained by considering a single relative-entropy constraint (7.26) for the whole time interval (in blue) and the proposed D^2 -LQG controller in which an incremental relative entropy constraint (7.5) is considered (in red). The evolution of the system controlled with the standard LQG technique is not shown as the signals diverge.



LQG and risk-sensitive LQG control with state-feedback

In this appendix, we review two fundamental classes of discrete-time, finite horizon, linear optimal control problems: the LQG control and the risk-sensitive LQG (or linear exponential-of-quadratic Gaussian) control. After formulating the optimal control problems, we present formulas for optimal policies and optimal cost. In doing so, we mainly refer to [6], [126].

C.1 LQG control

Consider the discrete-time linear dynamic system

$$x_{t+1} = Ax_t + Bu_t + v_t \tag{C.1}$$

where $x_t \in \mathbb{R}^n$ represents the vector of state variables and $u_t \in \mathbb{R}^m$ is the control input. The system is affected by the white Gaussian noise $v = \{v_t, t \in \mathbb{Z}\}$ with $v_t \sim \mathcal{N}(0, V)$ and $V \succ 0$. The initial state $x_0 = \bar{x}_0$ is assumed to be a deterministic quantity. Suppose that the state is measurable and assume state-feedback control policies $u_t = \pi_t(x_t)$, where π_t is a measurable function mapping \mathbb{R}^n into \mathbb{R}^m . Introduce the quadratic cost function

$$J = \frac{1}{2} \sum_{t=0}^N \left(x_t^\top Q x_t + u_t^\top R u_t \right) + \frac{1}{2} x_{N+1}^\top Q_{N+1} x_{N+1}, \tag{C.2}$$

where $Q \succeq 0$ and $R, Q_{N+1} \succ 0$. Given the system (C.1) and the cost function (C.2), the objective of the LQG control is to find the control policy minimizing the expected cost

$$J_{LQG} = \min_{\{u_t, t=0, \dots, N\}} \mathbb{E}[J].$$

It turns out that the optimal LQG control policy is given by

$$u_t = -K_t x_t,$$

where the feedback gain matrix K_t is given by

$$K_t = R^{-1} B^\top (\Pi_{t+1}^{-1} + B R^{-1} B^\top)^{-1} A,$$

where the matrices Π_t can be determined by solving the backward Riccati difference equation

$$\begin{aligned} \Pi_t &= Q + A^\top \left(\Pi_{t+1}^{-1} + B R^{-1} B^\top \right)^{-1} A \\ \Pi_{N+1} &= Q_{N+1}. \end{aligned}$$

The corresponding minimum value of the cost function is

$$J_{LQG} = \frac{1}{2} \bar{x}_0^\top \Pi_0 \bar{x}_0 + \frac{1}{2} \sum_{t=1}^{N+1} \text{tr}(V \Pi_t).$$

Notice that, when $V = 0$ (deterministic case), the optimal policy remains the same, but the optimal cost decreases.

C.2 Risk-sensitive LQG control

In the previous LQG control problem, the optimal feedback is independent of the noise statistics and, indeed, it coincides with the control law of the deterministic linear quadratic regulator. An alternative control scheme, which shows enhanced robustness properties and explicitly takes into account the disturbance distribution, is the risk-sensitive LQG control introduced by Jacobson in 1973, [72].

Consider the linear time system (C.1) and the quadratic function (C.2). Suppose that the state is measurable and assume state-feedback control policies $u_t = \pi_t(x_t)$, where π_t is a measurable function mapping \mathbb{R}^n into \mathbb{R}^m . The risk-sensitive LQG problem seeks the control policy which minimizes the cost function

$$J_{RS} = \min_{\{u_t, t=0, \dots, N\}} \frac{1}{\theta} \log \mathbb{E}[e^{\theta J}], \quad (\text{C.3})$$

where $\theta > 0$ is the *risk-sensitive parameter* measuring the optimizer's aversion to risk.

It is obtained in [126] that the Riccati recursion for the risk-sensitive LQG becomes

$$\begin{aligned}\Pi_t &= Q + A^\top \left(\Pi_{t+1}^{-1} + BR^{-1}B^\top - \theta V \right)^{-1} A \\ \Pi_{N+1} &= Q_{N+1}\end{aligned}$$

as long as

$$\Pi_{t+1}^{-1} - \theta V \succ 0, \quad t = 0, \dots, N.$$

Then, the optimal control policy is $u_t = -K_t x_t$ where the feedback gain is given by

$$K_t = R^{-1}B^\top \left(\Pi_{t+1}^{-1} + BR^{-1}B^\top - \theta V \right)^{-1} A.$$

The corresponding optimal cost is

$$J_{RS} = \frac{1}{2} \bar{x}_0^\top \Pi_0 \bar{x}_0 - \frac{1}{2\theta} \sum_{t=1}^{N+1} \log |I - \theta \Pi_t V|.$$

Note that in case $\theta = 0$, the risk-sensitive LQG reduces to the standard LQG problem.

Summary and outlook

This thesis dealt with two different, yet intimately connected, problems. In the first part of the dissertation, we faced the problem of robustness in the identification of dynamic factor analysis models. We focused on discrete-time, linear, Gaussian, stochastic processes described in terms of few latent variables, and we considered the realistic situation in which only a finite sample estimate of the underlying model is available. Then, we proposed an entropic-based criterion to construct a confidence region around the given estimate and to take into account the uncertainty in the estimation. In this confidence region, we sought for the most parsimonious factor analysis model by solving a convex optimization problem. This paradigm was applied to the identification of moving-average factor models in Chapter 3 and it was extended to autoregressive moving-average factor models in Chapter 4. Simulations studies, both with synthetic and real data, confirmed the good performances of the procedure even when the cross-sectional dimension is large with respect to sample size. Some ancillary results that are both interesting per se and strictly related to the main topic were discussed in Appendix A and Appendix B. More in detail, in Appendix A a mean-square consistent spectral density estimator was proposed. In Appendix B we addressed the generalized dynamic factor analysis problem.

It is clear that the applications of the proposed robust approach are not limited to the dynamic factor analysis problem considered in this dissertation. In fact, it can be applied to a wider class of problems in which one may want to learn a structured second-order model having access to a finite-sample estimate of it.

One further application of this paradigm was presented in the second part of the dissertation, where we dealt with the problem of designing an optimal control policy for stochastic uncertain systems. We considered discrete-time, linear, stochastic systems for which only an inaccurate estimate of the underlying probability distribution function is available. We considered a confidence region in the relative entropy topology around the estimated nominal distribution, then we designed an optimal control policy for the “worst system” in the prescribed confidence region. The main novelty with respect to the existing literature on relative entropy distributionally robust control is that the bounds on the distributional uncertainty can be arbitrarily distributed along the whole system trajectory. In Chapter 6, we derived a closed-form expression of the worst admissible system, for which we designed a state-feedback controller in Chapter 7. Numerical simulations showed the advantages of the proposed procedure and its ability to trade off optimality and robustness.

Directions for future research regard, in particular, the generalization of the proposed control scheme to the case in which only some noisy linear combinations of the state are

available to the closed-loop controller. A second compelling direction is to integrate the proposed approach in more advanced control techniques, such as the model predictive control and the regret-optimal control paradigms.

References

- [1] S. Abadeh, V. Nguyen, D. Kuhn, and P. Esfahani, “Wasserstein distributionally robust Kalman filtering,” in *Advances in Neural Information Processing Systems*, 2018, pp. 8474–8483 (Cited in page 85).
- [2] D. Alpag0, M. Zorzi, and A. Ferrante, “A scalable strategy for the identification of latent-variable graphical models,” *IEEE Transactions on Automatic Control*, 2021 (Cited in page 63).
- [3] B. D. O. Anderson, A. Braumann, and M. Deistler, “Identification of generalized dynamic factor models from mixed-frequency data,” in *IFAC Papers On Line 51-15*, 2018, pp. 1008–1013 (Cited in page 73).
- [4] B. D. O. Anderson and M. Deistler, “Generalized linear dynamic factor models - a structure theory,” in *Proc. of the IEEE Decision and Control conference (CDC)*, 2008 (Cited in page 73).
- [5] B. Anderson and J. Moore, *Optimal Control: Linear Quadratic Methods* (Dover Books on Engineering). Dover Publications, 2007 (Cited in pages 85, 89).
- [6] K. Åström, *Introduction to Stochastic Control Theory* (Dover Books on Electrical Engineering). Dover Publications, 2006 (Cited in pages 85, 89, 92, 121).
- [7] J. Aubin and I. Ekeland, *Applied Nonlinear Analysis*. New York: John Wiley & Sons, 1984 (Cited in page 91).
- [8] E. Avventi, A. G. Lindquist, and B. Wahlberg, “Arma identification of graphical models,” *IEEE Transactions on Automatic Control*, vol. 58, no. 5, pp. 1167–1178, 2013 (Cited in page 63).
- [9] J. Bai and S. Ng, “Determining the number of factors in approximate factor models,” *Econometrica*, vol. 70, no. 1, pp. 191–221, 2002 (Cited in pages 13, 15).
- [10] J. Bai and S. Ng, “Large dimensional factor analysis,” *Foundations and Trends in Econometrics*, vol. 3, no. 2, pp. 89–163, 2008 (Cited in pages 15, 16).
- [11] M. S. Bartlett and J. Medhi, “On the efficiency of procedures for smoothing periodograms from time series with continuous spectra,” *Biometrika*, vol. 42, no. 1/2, pp. 143–150, 1955 (Cited in page 64).
- [12] M. S. Bartlett, “Smoothing periodograms from time-series with continuous spectra,” *Nature*, vol. 161, pp. 686–687, 1948 (Cited in page 64).
- [13] T. Başar and P. Bernhard, *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008 (Cited in page 83).

- [14] G. Bayraksan and D. K. Love, “Data-driven stochastic programming using phi-divergences,” in *The operations research revolution*, INFORMS, 2015, pp. 1–19 (Cited in page 84).
- [15] P. A. Bekker and J. de Leeuw, “The rank of reduced dispersion matrices,” *Psychometrika*, vol. 52, no. 1, pp. 125–135, 1987 (Cited in page 13).
- [16] D. Bertsimas, M. S. Copenhaver, and R. Mazumder, “Certifiably optimal low rank factor analysis,” *Journal of Machine Learning Research*, vol. 18, no. 29, pp. 1–53, 2017 (Cited in page 15).
- [17] D. Bertsimas and A. W. Lo, “Optimal control of execution costs,” *Journal of Financial Markets*, vol. 1, no. 1, pp. 1–50, 1998 (Cited in page 83).
- [18] R. B. Blackman and J. W. Tukey, “The measurement of power spectra from the point of view of communications engineering — part i,” *Bell System Technical Journal*, vol. 37, no. 1, pp. 185–282, 1958 (Cited in page 64).
- [19] J. Bobrow, S. Dubowsky, and J. Gibson, “Time-optimal control of robotic manipulators along specified paths,” *The International Journal of Robotics Research*, vol. 4, no. 3, pp. 3–17, 1985 (Cited in page 83).
- [20] G. Bottegal and G. Picci, “Modeling complex systems by generalized factor analysis,” *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 759–774, 2015 (Cited in pages 16, 20, 73, 75).
- [21] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, United Kingdom: Cambridge University Press, 2004 (Cited in pages 30, 46, 47, 91, 96, 111, 112).
- [22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011 (Cited in page 45).
- [23] C. Byrnes, T. T. Georgiou, and A. Lindquist, “A new approach to spectral estimation: A tunable high-resolution spectral estimator,” *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 3189–3205, 2000 (Cited in page 63).
- [24] G. Chamberlain, “Funds, factors and diversification in arbitrage pricing models,” *Econometrica*, vol. 51, no. 5, pp. 1305–1324, 1983 (Cited in page 73).
- [25] G. Chamberlain and M. Rothschild, “Arbitrage, factor structure and meanvariance analysis on large asset markets,” *Econometrica*, vol. 51, no. 5, pp. 1281–1304, 1983 (Cited in page 73).

- [26] S. Chang and T Peng, "Adaptive guaranteed cost control of systems with uncertain parameters," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 474–483, 1972 (Cited in page 83).
- [27] V. Ciccone, A. Ferrante, and M. Zorzi, "Factor models with real data: A robust estimation of the number of factors," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2412–2425, 2019 (Cited in pages 15, 16, 21).
- [28] V. Ciccone, A. Ferrante, and M. Zorzi, "Learning latent variable dynamic graphical models by confidence sets selection," *IEEE Transactions on Automatic Control*, vol. 65, no. 12, pp. 5130–5143, 2020 (Cited in pages 16, 17, 26, 33, 35, 36, 38).
- [29] V. Ciccone, A. Ferrante, and M. Zorzi, "Robust identification of "sparse plus low-rank" graphical models: An optimization approach," in *Proceedings of the 57th IEEE Conference on Decision and Control (CDC)*, Miami Beach, FL, USA, 2018 (Cited in page 15).
- [30] F. Crescente, L. Falconi, F. Rozzi, A. Ferrante, and M. Zorzi, "Learning AR factor models," in *2020 59th IEEE Conference on Decision and Control (CDC)*, IEEE, 2020, pp. 274–279 (Cited in pages 6, 18, 51).
- [31] F. Cucker and S. Smale, "Emergent behavior in flocks," *IEEE Transactions on Automatic Control*, vol. 52, no. 5, pp. 852–862, 2007 (Cited in page 15).
- [32] R. Dahlhaus, "Graphical interaction models for multivariate time series," *Metrika*, vol. 51, no. 2, pp. 157–172, 2000 (Cited in page 63).
- [33] P. J. Daniell, "Discussion on symposium on autocorrelation in time series," *Journal of the Royal Statistical Society*, vol. 8, no. 1, 1946 (Cited in page 64).
- [34] M. Deistler, B. D. O. Anderson, A. Filler, C. Zinner, and W. Chen, "Generalized linear dynamic factor models: An approach via singular autoregressions," *European Journal of Control*, vol. 3, pp. 211–224, 2010 (Cited in page 73).
- [35] M. Deistler, W. Scherrer, and B. D. Anderson, "The structure of generalized linear dynamic factor models," *Empirical economic and financial research: Theory, methods and practice*, pp. 379–400, 2015 (Cited in page 13).
- [36] M. Deistler and C. Zinner, "Modelling high-dimensional time series by generalized linear dynamic factor models: An introductory survey," 2007 (Cited in pages 13, 16, 73).
- [37] G. Della Riccia and A. Shapiro, "Minimum rank and minimum trace of covariance matrices," *Psychometrika*, vol. 47, pp. 443–448, 4 1982 (Cited in page 15).

- [38] P. Dorato, “A historical review of robust control,” *IEEE Control Systems Magazine*, vol. 7, no. 2, pp. 44–47, 1987 (Cited in page 83).
- [39] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, “Dynamic textures,” *International journal of computer vision*, vol. 51, pp. 91–109, 2003 (Cited in page 20).
- [40] J. Doyle, “Guaranteed margins for LQG regulators,” *IEEE Transactions on Automatic Control*, vol. 23, no. 4, pp. 756–757, 1978 (Cited in pages 83, 89).
- [41] C. Doz, D. Giannone, and L. Reichlin, “A quasi—maximum likelihood approach for large, approximate dynamic factor models,” *The Review of Economics and Statistics*, vol. 94, no. 4, pp. 1014–1024, 2012 (Cited in pages 17, 59, 61).
- [42] P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 2011 (Cited in page 7).
- [43] F. Engels, “Target shape estimation using an automotive radar,” in *Smart Mobile In-Vehicle Systems: Next Generation Advancements*, G. Schmidt, H. Abut, K. Takeda, and J. H. Hansen, Eds. New York, NY: Springer New York, 2014, pp. 271–290 (Cited in page 63).
- [44] F. Engels, P. Heidenreich, A. M. Zoubir, F. K. Jondral, and M. Wintermantel, “Advances in automotive radar: A framework on computationally efficient high-resolution frequency estimation,” *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 36–46, 2017 (Cited in page 63).
- [45] R. Engle and M. Watson, “A one-factor multivariate time series model of metropolitan wage rates,” *Journal of the American Statistical Association*, vol. 76, no. 376, pp. 774–781, 1981 (Cited in page 17).
- [46] B. D. Erson and M. Deistler, “Identification of dynamic systems from noisy data: Single factor case,” *Mathematics of Control, Signals and Systems*, vol. 6, no. 1, pp. 10–29, 1993 (Cited in page 13).
- [47] L. Falconi, A. Ferrante, and M. Zorzi, “A new perspective on robust performance for LQG control problems,” in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 3003–3008 (Cited in pages 7, 86).
- [48] L. Falconi, A. Ferrante, and M. Zorzi, “A Robust Approach to ARMA Factor Modeling,” *IEEE Transactions on Automatic Control*, 2023 (Cited in pages 6, 18, 22, 51).
- [49] L. Falconi, A. Ferrante, and M. Zorzi, “Distributionally robust LQG control under distributed uncertainty,” (preliminary version available at *arXiv preprint arXiv:2306.05227 (2023)*) (Cited in pages 7, 87).

- [50] L. Falconi, A. Ferrante, and M. Zorzi, “Mean-square consistency of the f -truncated M^2 -periodogram,” *Automatica*, vol. 147, p. 110 672, 2023 (Cited in pages 6, 19, 64, 71).
- [51] J. Fan, Y. Liao, and M. Mincheva, “Large covariance estimation by thresholding principal orthogonal complements,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 75, no. 4, pp. 603–680, 2013 (Cited in page 13).
- [52] M. Fazel, “Matrix rank minimization with applications,” *Elec. Eng. Dept. Stanford University*, vol. 54, pp. 1–130, 2002 (Cited in page 15).
- [53] M. Fazel, H. Hindi, and S. Boyd, “Rank minimization and applications in system theory,” in *Proceedings of the American Control Conference*, vol. 4, 2004, pp. 3273–3278 (Cited in page 15).
- [54] A. Ferrante and L. Ntogramatzidis, “A note on finite-horizon LQ problems with indefinite cost,” *Automatica*, vol. 52, pp. 290–293, 2015 (Cited in page 89).
- [55] A. Ferrante and L. Ntogramatzidis, “The extended symplectic pencil and the finite-horizon lq problem with two-sided boundary conditions,” *IEEE Transactions on Automatic Control*, vol. 58, no. 8, pp. 2102–2107, 2013 (Cited in page 89).
- [56] A. Ferrante, C. Masiero, and M. Pavon, “Time and spectral domain relative entropy: A new approach to multivariate spectral estimation,” *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2561–2575, 2012 (Cited in pages 24, 63).
- [57] A. Ferrante, M. Pavon, and F. Ramponi, “Hellinger versus Kullback–Leibler multivariable spectrum approximation,” *IEEE Transactions on Automatic Control*, vol. 53, no. 4, pp. 954–967, 2008 (Cited in page 63).
- [58] L. Ferrante A.and Ntogramatzidis, “Finite-horizon linear-quadratic optimal control with general boundary conditions,” in *Encyclopedia of Systems and Control*, J. Baillieul and T. Samad, Eds., London: Springer, 2019, pp. 1–7 (Cited in page 89).
- [59] G. Figá-Talamanca, S. Focardi, and M. Patacca, “Common dynamic factors for cryptocurrencies and multiple pair-trading statistical arbitrages,” *Decisions in Economics and Finance*, vol. 44, pp. 863–882, 2021 (Cited in page 16).
- [60] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, “The generalized dynamic factor model: Consistency and convergence rates,” *Journal of Econometrics*, vol. 119, pp. 231–235, 2004 (Cited in page 73).
- [61] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, “The generalized dynamic factor model: Identification and estimation,” *The review of Economic and Statistics*, vol. 65, pp. 453–473, 2000 (Cited in pages 17, 73).

- [62] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, “The generalized dynamic factor model: One-sided estimation and forecasting,” *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 830–840, 2005 (Cited in pages 17, 73).
- [63] M. Forni, M. Hallin, M. Lippi, and P. Zaffaroni, “Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis,” *Journal of Econometrics*, vol. 199, no. 1, pp. 74–92, 2017 (Cited in page 17).
- [64] M. Forni, M. Hallin, M. Lippi, and P. Zaffaroni, “Dynamic factor models with infinite-dimensional factor spaces: One-sided representations,” *Journal of Econometrics*, vol. 185, no. 2, pp. 359–371, 2015 (Cited in page 17).
- [65] M. Forni and M. Lippi, “The generalized dynamic factor model: Representation theory,” *Econometric theory*, vol. 17, no. 6, pp. 1113–1141, 2001 (Cited in pages 13, 73–75).
- [66] J. Geweke, “The dynamic factor analysis of economic time series,” in *Latent Variables in Socio-Economic Models*, D. A. A. Goldberger, Ed., North-Holland, 1977, pp. 365–383 (Cited in pages 13, 15).
- [67] W. M. Haddad, D. S. Bernstein, and D. Mustafa, “Mixed-norm H_2/H_∞ regulation and estimation: The discrete-time case,” *Systems & Control Letters*, vol. 16, no. 4, pp. 235–247, 1991 (Cited in page 83).
- [68] M. Hallin and R. Liška, “Determining the number of factors in the general dynamic factor model,” *Journal of the American Statistical Association*, vol. 102, no. 478, pp. 603–617, 2007 (Cited in pages 17, 47, 48).
- [69] L. Hansen and T. Sargent, *Robustness*. Princeton University Press, 2008 (Cited in page 85).
- [70] S. Ihara, *Information theory for continuous systems*. World Scientific, 1993, vol. 2 (Cited in page 9).
- [71] F. Itakura, “Analysis synthesis telephony based on the maximum likelihood method,” *Reports of the 6th Int. Cong. Acoust., 1968*, 1968 (Cited in pages 9, 24).
- [72] D. Jacobson, “Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games,” *IEEE Transactions on Automatic Control*, vol. 18, no. 2, pp. 124–131, 1973 (Cited in pages 85, 90, 97, 103, 109, 122).
- [73] H. F. Kaiser, “The varimax criterion for analytic rotation in factor analysis,” *Psychometrika*, vol. 23, no. 3, pp. 187–200, 1958 (Cited in page 13).

- [74] R. E. Kalman, *Identifiability and problems of model selection in econometrics*. US Government Printing Office, 1982 (Cited in page 13).
- [75] G. Kapetanios and M. Marcellino, “A parametric estimation method for dynamic factor models of large dimensions,” *Journal of Time Series Analysis*, vol. 30, pp. 208–238, 2009 (Cited in page 76).
- [76] A. D. Kuo, “An optimal control model for analyzing human postural balance,” *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 87–101, 1995 (Cited in page 83).
- [77] C. Lam and Q. Yao, “Factor modeling for high-dimensional time series: Inference for the number of factors,” *The Annals of Statistics*, pp. 694–726, 2012 (Cited in page 15).
- [78] D. N. Lawley and A. E. Maxwell, *Factor Analysis as a Statistical Method, Second ed.* London: Butterworths, 1971 (Cited in page 13).
- [79] D. N. Lawley, “The estimation of factor loadings by the method of maximum likelihood,” *Proceedings of the Royal Society of Edinburgh*, vol. 60, no. 1, pp. 64–82, 1940 (Cited in page 13).
- [80] W. Ledermann, “On a problem concerning matrices with variable diagonal elements,” *Proc. Royal Soc. Edinburgh*, vol. XL, pp. 1–17, 1939 (Cited in page 13).
- [81] W. Ledermann, “On the rank of the reduced correlation matrix in multiple factor analysis,” *Psychometrika*, vol. 2, pp. 85–93, 1937 (Cited in page 13).
- [82] B. C. Levy and R. Nikoukhah, “Robust state space filtering under incremental model perturbations subject to a relative entropy tolerance,” *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 682–695, 2013 (Cited in pages 85, 92).
- [83] A. Lindquist and G. Picci, *Linear stochastic systems: a geometric approach*. Springer Verlag, 2015 (Cited in pages 75, 78).
- [84] M. Lippi, M. Deistler, and B. Anderson, “High-dimensional dynamic factor models: A selective survey and lines of future research,” *Econometrics and Statistics*, 2022 (Cited in page 16).
- [85] L. Ljung, *System Identification: Theory for the User. Second Edition*. Prentice Hall PTR, 1999 (Cited in page 63).
- [86] D. G. Luenberger, *Optimization by Vector Space Methods*, 1st. USA: John Wiley & Sons, Inc., 1997 (Cited in page 91).

- [87] S. Maanan, B. Dumitrescu, and C. Giurcăneanu, “Maximum entropy expectation-maximization algorithm for fitting latent-variable graphical models to multivariate time series,” *Entropy*, vol. 20, p. 76, 2018 (Cited in page 17).
- [88] M. Marcellino, “An introduction to factor modelling,” Bocconi University, Lecture slides, 2017 (Cited in page 78).
- [89] A. Megretski and A. Rantzer, “System analysis via integral quadratic constraints,” *IEEE transactions on automatic control*, vol. 42, no. 6, pp. 819–830, 1997 (Cited in page 83).
- [90] D. Mustafa and D. S. Bernstein, “Lqg cost bounds in discrete-time h2/h-infinity control,” *Transactions of the Institute of Measurement and Control*, vol. 13, no. 5, pp. 269–275, 1991 (Cited in page 83).
- [91] A. Nilim and L. El Ghaoui, “Robust control of Markov decision processes with uncertain transition matrices,” *Operations Research*, vol. 53, no. 5, pp. 780–798, 2005 (Cited in page 84).
- [92] L. Ning, T. T Georgiou, A. Tannenbaum, and S. P. Boyd, “Linear models based on noisy data and the Frisch scheme,” *SIAM Review*, vol. 57, no. 2, pp. 167–197, 2015 (Cited in page 15).
- [93] D. Peña and G. Box, “Identifying a simplifying structure in time series,” *J. Amer. Stat. Ass.*, vol. 82, pp. 836–843, 1987 (Cited in page 13).
- [94] D. Peña and P. Poncela, “Nonstationary dynamic factor analysis,” *Journal of Statistical Planning and Inference*, vol. 136, pp. 1237–1257, 2006 (Cited in page 13).
- [95] I. Petersen, “Minimax LQG control,” *International Journal of Applied Mathematics and Computer Science*, vol. 16, no. 3, pp. 309–323, 2006 (Cited in page 89).
- [96] I Petersen, M. James, and P. Dupuis, “Minimax optimal control of stochastic uncertain systems with relative entropy constraints,” *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 398–412, 2000 (Cited in pages 84–86, 89, 93, 97–99, 105, 113, 114, 117, 118).
- [97] I. R. Petersen and R. Tempo, “Robust control of uncertain systems: Classical results and recent developments,” *Automatica*, vol. 50, no. 5, pp. 1315–1335, 2014 (Cited in page 83).
- [98] I. R. Petersen and R. Tempo, “Robust control of uncertain systems: Classical results and recent developments,” *Autom.*, vol. 50, pp. 1315–1335, 2014 (Cited in page 106).

- [99] G. Picci, “Parametrization of factor analysis models,” *Journal of Econometrics*, vol. 41, pp. 17–38, 1987 (Cited in page 13).
- [100] G. Picci, L. Falconi, A. Ferrante, and M. Zorzi, “Hidden factor estimation in dynamic generalized factor analysis models,” *Automatica*, vol. 149, p. 110 834, 2023 (Cited in pages 6, 19, 74, 76).
- [101] H. Rahimian and S. Mehrotra, “Distributionally robust optimization: A review,” *arXiv preprint arXiv:1908.05659*, 2019 (Cited in page 84).
- [102] J. Rawlings, D. Mayne, and M. Diehl, *Model Predictive Control: Theory, Computation, and Design*. Nob Hill Publishing, 2017 (Cited in pages 89, 94).
- [103] O. Reiersøl, “Identifiability of a linear relation between variables which are subject to error,” *Econometrica*, vol. 18, no. 4, pp. 375–389, 1950 (Cited in page 15).
- [104] A. V. Savkin and I. R. Petersen, “Minimax optimal control of uncertain systems with structured uncertainty,” *International Journal of Robust and Nonlinear Control*, vol. 5, no. 2, pp. 119–137, 1995 (Cited in page 83).
- [105] H. E. Scarf, K. Arrow, and S Karlin, *A min-max solution of an inventory problem*. Rand Corporation Santa Monica, 1957 (Cited in page 84).
- [106] A. Shapiro, “Distributionally robust stochastic programming,” *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2258–2275, 2017 (Cited in page 84).
- [107] A. Shapiro, “Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis,” *Psychometrika*, vol. 47, pp. 187–199, 1982 (Cited in page 13).
- [108] B. A. Sharif and A. Ucar, “State feedback and lqr controllers for an inverted pendulum system,” in *2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE)*, 2013, pp. 298–303 (Cited in page 116).
- [109] M. Sion, “On general minimax theorems.,” 1958 (Cited in page 112).
- [110] T. Söderström and P. Stoica, *System Identification*. Hemel Hempstead, UK: Prentice-Hall International, 1989 (Cited in page 63).
- [111] J. Songsiri, J. Dahl, and L. Vandenberghe, “Graphical models of autoregressive processes,” *Convex optimization in signal processing and communications*, pp. 89–116, 2010 (Cited in page 40).
- [112] J. C. Spall, “Cyclic seesaw process for optimization and identification,” *Journal of Optimization Theory and Applications*, vol. 154, pp. 187–208, 2012 (Cited in pages 99, 112).

- [113] C. Spearman, “The american journal of psychology,” *American Journal of Psychology*, vol. 15, p. 88, 1904 (Cited in page 13).
- [114] J. H. Stock and M. W. Watson, “35 Dynamic Factor Models,” in *The Oxford Handbook of Economic Forecasting*, Oxford University Press, 2011 (Cited in page 16).
- [115] J. H. Stock and M. W. Watson, “New indexes of coincident and leading economic indicators,” *NBER macroeconomics annual*, vol. 4, pp. 351–394, 1989 (Cited in page 17).
- [116] P. Stoica, R. L. Moses, *et al.*, *Spectral analysis of signals*. NJ: Prentice Hall, 2005 (Cited in pages 56, 57, 63).
- [117] A. Stoorvogel and J. Van Schuppen, “System identification with information theoretic criteria,” in *From Identification to Learning: the science of learning models from data*, Springer, 1996, pp. 289–338 (Cited in page 9).
- [118] T. Tanaka, P. M. Esfahani, and S. K. Mitter, “Lqg control with minimum directed information: Semidefinite programming approach,” *IEEE Transactions on Automatic Control*, vol. 63, no. 1, pp. 37–52, 2018 (Cited in page 89).
- [119] L. L. Thurstone, “Multiple factor analysis.,” *Psychological review*, vol. 38, no. 5, p. 406, 1931 (Cited in page 13).
- [120] L. R. Tucker and C. Lewis, “A reliability coefficient for maximum likelihood factor analysis,” *Psychometrika*, vol. 38, no. 1, pp. 1–10, 1973 (Cited in page 13).
- [121] I. Tzortzis, C. D. Charalambous, and T. Charalambous, “Dynamic programming subject to total variation distance ambiguity,” *SIAM Journal on Control and Optimization*, vol. 53, no. 4, pp. 2040–2075, 2015 (Cited in page 84).
- [122] B. P. G. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari, “Distributionally robust control of constrained stochastic systems,” *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 430–442, 2016 (Cited in page 84).
- [123] B. P. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari, “Distributionally robust control of constrained stochastic systems,” *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 430–442, 2015 (Cited in page 84).
- [124] D. Wang, J. Qiao, and L. Cheng, “An approximate neuro-optimal solution of discounted guaranteed cost control design,” *IEEE Transactions on Cybernetics*, vol. 52, no. 1, pp. 77–86, 2020 (Cited in page 83).
- [125] M. W. Watson and R. F. Engle, “Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models,” *Journal of Econometrics*, vol. 23, no. 3, pp. 385–400, 1983 (Cited in page 17).

-
- [126] P. Whittle, “Risk-sensitive linear/quadratic/gaussian control,” *Advances in Applied Probability*, vol. 13, no. 4, pp. 764–777, 1981 (Cited in pages 85, 90, 121, 123).
- [127] I. Yang, “Wasserstein distributionally robust stochastic control: A data-driven approach,” *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3863–3870, 2020 (Cited in page 84).
- [128] S. Yi and M. Zorzi, “Robust kalman filtering under model uncertainty: The case of degenerate densities,” *IEEE Transactions on Automatic Control*, vol. 67, no. 7, pp. 3458–3471, 2022 (Cited in page 85).
- [129] G. Zames, “Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses,” *IEEE Transactions on automatic control*, vol. 26, no. 2, pp. 301–320, 1981 (Cited in page 83).
- [130] F. Zamora-Martínez, P. Romeu, P. Botella-Rocamora, and J. Pardo, “On-line learning of indoor temperature forecasting models towards energy efficiency,” *Energy and Buildings*, vol. 83, pp. 162–172, 2014 (Cited in page 59).
- [131] A. Zee, *Quantum Field Theory in a Nutshell. Second Edition* (In a Nutshell). Princeton, NJ: Princeton Univ. Press, 2010 (Cited in page 90).
- [132] F. Zhang, *The Schur complement and its applications*. Springer Science & Business Media, 2006, vol. 4 (Cited in page 43).
- [133] K Zhou, K Glover, B Bodenheimer, and J Doyle, “Mixed H_2 and H_∞ performance objectives ii: Optimal control,” *IEEE Transactions on Automatic Control*, vol. 39, no. 8, pp. 1575–1587, 1994 (Cited in page 83).
- [134] M. Zorzi, “Robust kalman filtering under model perturbations,” *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2902–2907, 2017 (Cited in page 85).
- [135] M. Zorzi and R. Sepulchre, “AR identification of latent-variable graphical models,” *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2327–2340, 2016 (Cited in pages 16, 17, 24, 28).
- [136] M. Zorzi, “Autoregressive identification of Kronecker graphical models,” *Automatica*, vol. 119, p. 109 053, 2020 (Cited in page 63).
- [137] M. Zorzi, “On the robustness of the Bayes and Wiener estimators under model uncertainty,” *Automatica*, vol. 83, pp. 133–140, 2017 (Cited in page 85).