Robust Model Fitting via Convex Optimization Techniques

A Dissertation Presented

by

Yongfang Cheng

to

The Department of Electrical and Computer Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Northeastern University
Boston, Massachusetts

August 2016
Dissertation Title: Robust Model Fitting via Convex Optimization Techniques
Author: Yongfang Cheng
NUID: 000533263
Department: Electrical and Computer Engineering

Approved for Dissertation Requirements of the Doctor of Philosophy Degree

Dissertation Advisor
Dr. Mario Sznaier
Signature Date

Dissertation Committee Member
Dr. Octavia Camps
Signature Date

Dissertation Committee Member
Dr. Jennifer Dy
Signature Date

Dissertation Committee Member
Dr. Rifat Sipahi
Signature Date

Department Chair
Dr. Miriam Leeser
Signature Date

Associate Dean of Graduate School:
Dr. Sara Wadia-Fascetti
Signature Date
To my parents and my brother.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Acronyms</td>
<td>x</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xi</td>
</tr>
<tr>
<td>Abstract of the Dissertation</td>
<td>xii</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Challenges</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Previous work</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Contributions</td>
<td>6</td>
</tr>
<tr>
<td>1.3.1 Static Data</td>
<td>6</td>
</tr>
<tr>
<td>1.3.2 Dynamical Data</td>
<td>7</td>
</tr>
<tr>
<td><strong>2 Preliminaries</strong></td>
<td>9</td>
</tr>
<tr>
<td>2.1 Notations</td>
<td>9</td>
</tr>
<tr>
<td>2.2 Moments-Based Polynomial Optimization</td>
<td>10</td>
</tr>
<tr>
<td>2.2.1 Moment Extension Problem</td>
<td>11</td>
</tr>
<tr>
<td>2.2.2 K-Moment Problem</td>
<td>12</td>
</tr>
<tr>
<td>2.2.3 Sparse Polynomial Optimization</td>
<td>13</td>
</tr>
<tr>
<td>2.3 Support Vector Machine</td>
<td>14</td>
</tr>
<tr>
<td>2.3.1 Primal SVM</td>
<td>15</td>
</tr>
<tr>
<td>2.3.2 Dual SVM</td>
<td>16</td>
</tr>
<tr>
<td>2.4 Atomic Norms in LTI identification</td>
<td>16</td>
</tr>
<tr>
<td>2.5 Superstability and Equalized Performance</td>
<td>18</td>
</tr>
<tr>
<td>2.6 Robust Optimization</td>
<td>20</td>
</tr>
<tr>
<td><strong>3 Robust Linear Model Fitting</strong></td>
<td>23</td>
</tr>
<tr>
<td>3.1 Motivation</td>
<td>23</td>
</tr>
<tr>
<td>3.2 Problem Statement</td>
<td>25</td>
</tr>
<tr>
<td>3.3 A Convex Approach to Robust Linear Model Fitting</td>
<td>27</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>3.3.1</td>
<td>A Constrained POP Reformulation</td>
</tr>
<tr>
<td>3.3.2</td>
<td>A Sequence of Convex Relaxations</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Two Iterative Algorithms for Rank-1 Solution</td>
</tr>
<tr>
<td>3.4</td>
<td>Extensions: Handling Additional A-Priori Information</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Rank Deficient Constraint</td>
</tr>
<tr>
<td>3.5</td>
<td>Experiments</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Experimental Protocols</td>
</tr>
<tr>
<td></td>
<td>Results</td>
</tr>
<tr>
<td>3.6</td>
<td>Conclusions</td>
</tr>
<tr>
<td>4</td>
<td>Subspace Clustering with Priors</td>
</tr>
<tr>
<td>4.1</td>
<td>Motivation</td>
</tr>
<tr>
<td>4.2</td>
<td>Problem Statement</td>
</tr>
<tr>
<td>4.3</td>
<td>A Greedy Approach to Clustering</td>
</tr>
<tr>
<td>4.4</td>
<td>A Non-Greedy Approach to Clustering</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Clustering as a Nonlinear Feasibility Problem</td>
</tr>
<tr>
<td>4.4.2</td>
<td>A Sequence of Convex Relaxations</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Computational Complexity Analysis</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Handling Outliers</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Handling Additional A-Priori Information</td>
</tr>
<tr>
<td>4.4.6</td>
<td>Subspace Recovery Conditions</td>
</tr>
<tr>
<td>4.5</td>
<td>Experiments</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Synthetic Static Data</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Motion Segmentation</td>
</tr>
<tr>
<td>4.5.3</td>
<td>Planar Segmentation</td>
</tr>
<tr>
<td>4.6</td>
<td>Conclusions</td>
</tr>
<tr>
<td>5</td>
<td>Moments-Based Semi-Supervised Support Vector Machine</td>
</tr>
<tr>
<td>5.1</td>
<td>Motivation</td>
</tr>
<tr>
<td>5.2</td>
<td>Problem Statement</td>
</tr>
<tr>
<td>5.3</td>
<td>Moments-Based $S^3$ VM</td>
</tr>
<tr>
<td>5.3.1</td>
<td>A Sequence of Convex Relaxations</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Reweighted $\ell_1$ Relaxation</td>
</tr>
<tr>
<td>5.4</td>
<td>ALM Algorithm for $S^3$ VM</td>
</tr>
<tr>
<td>5.5</td>
<td>Extension: Kernel SVM</td>
</tr>
<tr>
<td>5.6</td>
<td>Experiments</td>
</tr>
<tr>
<td>5.7</td>
<td>Conclusions</td>
</tr>
<tr>
<td>6</td>
<td>EIV Identification of LTI Systems</td>
</tr>
<tr>
<td>6.1</td>
<td>Motivation</td>
</tr>
<tr>
<td>6.2</td>
<td>Problem Statement</td>
</tr>
<tr>
<td>6.3</td>
<td>A Convex Approach to EIV Identification of LTI Systems</td>
</tr>
<tr>
<td>6.3.1</td>
<td>A Constrained POP Formulation</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Gramian Based Rank Minimization</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Moments Based Convex Relaxation</td>
</tr>
</tbody>
</table>
## 6.4 Extensions

- **6.4.1 Handling Missing Data**
- **6.4.2 Handling Outliers**

## 6.5 Experiments

## 6.6 Conclusions

## 7 EIV Identification of SARX Systems

- **7.1 Motivation**
- **7.2 Problem Statement**
- **7.3 A Convex Approach to Identification of SARX Models**
  - **7.3.1 A Semi-Algebraic Optimization Reformulation**
  - **7.3.2 A Sequence of Convex Relaxations**
  - **7.3.3 Computational Complexity Analysis**
  - **7.3.4 Handling Additional A-Priori Information**
- **7.4 Extension: Identification with Minimum Number of Switches**
- **7.5 Experiments**
  - **7.5.1 Identification of W ARX Models of Equation Error Structure**
  - **7.5.2 General Identification of W ARX Models of EIV Structure**
  - **7.5.3 Additional A-Priori Information**
  - **7.5.4 Segmentation of ARX Models**
- **7.6 Conclusions**

## 8 Identification of LPV Systems with LFT Parametric Dependence

- **8.1 Motivation**
- **8.2 Problem Statement**
- **8.3 LPV Identification via Convex Optimization**
  - **8.3.1 Identification with noiseless measurements**
  - **8.3.2 The noisy measurements case**
- **8.4 Proposed Algorithms**
  - **8.4.1 A Single-Step Algorithm**
  - **8.4.2 An Efficient Multi-Step Relaxation**
- **8.5 Experiments**
- **8.6 Conclusions**

## 9 Model (In)validation of SARX Systems with Unknown Switches

- **9.1 Motivation**
- **9.2 Problem Statement**
- **9.3 Convex (In)Validation Certificates**
  - **9.3.1 A MILP reformulation**
  - **9.3.2 Sparsification Based Certificates**
  - **9.3.3 Moments Based Certificates**
- **9.4 Incorporating Structural Constraints**
- **9.5 Experiments**
  - **9.5.1 Academic Examples**
  - **9.5.2 Contextually Abnormal Activity Recognition**
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Data for the Toy Example (Left: Clean Data; Middle: Data Contaminated Noise vs. Right: Expected Clustering)</td>
</tr>
<tr>
<td>1.2</td>
<td>Subspaces without Outliers vs. Subspaces Skewed by Outliers</td>
</tr>
<tr>
<td>2.1</td>
<td>Hard-Margin SVM Vs. Soft-Margin SVM</td>
</tr>
<tr>
<td>3.1</td>
<td>2D Linear Regression Problem. (Left: clean data; Center: noisy data without outliers; Right: noisy data with outliers)</td>
</tr>
<tr>
<td>3.2</td>
<td>Sample Image Pairs from the VGG Dataset</td>
</tr>
<tr>
<td>3.3</td>
<td>Comparison for Image Pairs (House, Library, Merton I, Merton II, Merton III and Wadham from top to bottom, in each row, the figures are Precision, Recall, Hmean and Similarity, respectively.)</td>
</tr>
<tr>
<td>4.1</td>
<td>Greey Subspace Clustering</td>
</tr>
<tr>
<td>4.2</td>
<td>Sample Frames from the Hopkins 155 'cars9' Dataset (Frame 1, Frame 41, Frame 61)</td>
</tr>
<tr>
<td>4.3</td>
<td>Performance Comparison for the Motion Segmentation Example for Different Sets of Priors</td>
</tr>
<tr>
<td>4.4</td>
<td>Images for Planar Segmentation: Merton I</td>
</tr>
<tr>
<td>4.5</td>
<td>Average Performance of LRR and SSC over 20 Instances</td>
</tr>
<tr>
<td>4.6</td>
<td>Denoised Data vs. Noise (Left: LRR with $\lambda = 0.01$; Right: SSC with $\tau = 0.55$)</td>
</tr>
<tr>
<td>6.1</td>
<td>Basic Setups for LTI Systems (Left: AR Model; Central: ARX Model; Right: EIV Model)</td>
</tr>
<tr>
<td>7.1</td>
<td>Textured Image, Ground Truth of Labels, and Labels Obtained by Algorithm 9</td>
</tr>
<tr>
<td>7.2</td>
<td>Identified Active Mode $\sigma_t$ vs. Ground Truth for Example 7.5.2</td>
</tr>
<tr>
<td>7.3</td>
<td>Identified Active Mode $\sigma_5$ vs. Ground Truth for Example 7.5.3</td>
</tr>
<tr>
<td>7.4</td>
<td>Detection of Switches $s_t$ and Parameters $r_t$ for Example 7.5.4</td>
</tr>
<tr>
<td>8.1</td>
<td>Configuration of the LPV System</td>
</tr>
<tr>
<td>8.2</td>
<td>Equivalent LPV System with $0 \in \mathcal{P}$</td>
</tr>
<tr>
<td>8.3</td>
<td>Data for Identification</td>
</tr>
<tr>
<td>8.4</td>
<td>Data for Validation</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>9.1</td>
<td>Problem Setup</td>
</tr>
<tr>
<td>9.2</td>
<td>Size of LMIs in [127] - Size of LMIs in (9.17)</td>
</tr>
<tr>
<td>9.3</td>
<td>A Structurally Constrained Transition Graph (One-step transitions from $G_1$ to $G_2$ are not allowed)</td>
</tr>
<tr>
<td>9.4</td>
<td>Detecting Contextually Abnormal Activities in Video Sequences (For the purpose of this example we assumed that transitions from wait to running are not allowed. Hence the top sequence ($walk \to stop \to run$) should be labeled abnormal, while the bottom one ($stop \to walk \to run$) should not be flagged.)</td>
</tr>
<tr>
<td>10.1</td>
<td>Closed-loop System</td>
</tr>
<tr>
<td>10.2</td>
<td>Disturbance and Output Sequence of Different Controllers (from left to right, the first row: Distance $w_t$, $e_t$ from (10.16), $e_t$ from (10.22); the second row: $e_t$ from (10.32), $e_t$ from (10.33) and $e_t$ from (10.37))</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Mean and Standard Deviation of Precision (%)</td>
<td>38</td>
</tr>
<tr>
<td>3.2</td>
<td>Mean and Standard Deviation of Recall (%)</td>
<td>38</td>
</tr>
<tr>
<td>3.3</td>
<td>Mean and Standard Deviation of Hmean (%)</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>Mean and Standard Deviation of Similarity (×100)</td>
<td>38</td>
</tr>
<tr>
<td>4.1</td>
<td>Recovery Conditions (&quot;x&quot; denotes no certificates)</td>
<td>49</td>
</tr>
<tr>
<td>4.2</td>
<td>Performance Comparison for Different Synthetic Data Scenarios (n: dimension of data, d_i: dimension of each subspace, N_i: number of samples on each subspace, ε: error bound, µ and σ: mean and standard deviation of err_i, (*): experiments with outliers)</td>
<td>51</td>
</tr>
<tr>
<td>4.3</td>
<td>Average Performance for Planar Segmentation (%)</td>
<td>56</td>
</tr>
<tr>
<td>4.4</td>
<td>Running Time (sec)</td>
<td>56</td>
</tr>
<tr>
<td>5.1</td>
<td>Datasets used in Section 5.6, c: the number of classes, d: the dimensionality of the data, l: the number of labeled examples, and n: the number of all the data in the dataset</td>
<td>71</td>
</tr>
<tr>
<td>5.2</td>
<td>Experimental Results in Section 5.6 in ACCURACY (%): mean (std)</td>
<td>71</td>
</tr>
<tr>
<td>6.1</td>
<td>Estimates of Model Parameters under Different Noise Bounds</td>
<td>87</td>
</tr>
<tr>
<td>6.2</td>
<td>Estimates of Model Parameters with 10% Missing Data</td>
<td>87</td>
</tr>
<tr>
<td>6.3</td>
<td>Estimates of Model Parameters with 20% Missing Data</td>
<td>87</td>
</tr>
<tr>
<td>7.1</td>
<td>Results for Example 7.5.2</td>
<td>103</td>
</tr>
<tr>
<td>7.2</td>
<td>Results for Example 7.5.3</td>
<td>104</td>
</tr>
<tr>
<td>9.1</td>
<td>Invalidation Results for Example A.1 (Part 1)</td>
<td>135</td>
</tr>
<tr>
<td>9.2</td>
<td>Invalidation Results for Example A.1 (Part 2)</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>(both the actual noise bound and the prior knowledge on it are ε = 0.2)</td>
<td></td>
</tr>
<tr>
<td>9.3</td>
<td>Invalidation Results for Example A.1 (Part 3)</td>
<td>136</td>
</tr>
<tr>
<td>9.4</td>
<td>Complexity Analysis for Moments-Based Formulations (n_y = 1)</td>
<td>136</td>
</tr>
<tr>
<td>9.5</td>
<td>Invalidation under Structural Constraints (the same data as in Table 9.1)</td>
<td>138</td>
</tr>
<tr>
<td>9.6</td>
<td>Invalidation Results for Example B</td>
<td>140</td>
</tr>
<tr>
<td>10.1</td>
<td>Comparison between Different LP Formulations for Example 1</td>
<td>154</td>
</tr>
</tbody>
</table>
List of Acronyms

ALM  Augmented Lagrangian Method
AR  Auto-Regressive
ARX  Auto-Regressive with eXogenous input
EIV  Error-In-Variables
GPCA  Generalized Principal Component Analysis
LFT  Linear Fractional Transformation
LP  Linear Programming
LPV  Linear Parameter Varying
LTI  Linear Time Invariant
PCA  Principal Component Analysis
POP  Polynomial Optimization Problem
PSD  Positive Semi-Definite
SDP  Semi-Definite Programming
SVD  Singular Value Decomposition
SVM  Support Vector Machines
$S^3$VM  Semi-Supervised Support Vector Machines
SWARX  Switched Auto-Regressive with eXogenous input
Acknowledgments

A lot of people have contributed directly or indirectly to this dissertation and the scientific research associated with it.

First of all, I would like to extend my sincere thanks to my adviser Prof. Mario Sznaier, for his endless passion and rigorous attitude for academic research setting me a good example of a qualified researcher, for his smartness and wide knowledge which makes me feel “jealous” and always encourages me to work hard, and for his patience with encouraging me to be optimistic and happy whenever I am frustrated by the tough nuts in my life.

I would also like to thank Prof. Octavia Camps, for her patience with explaining concepts in computer vision to a layman like me over and over again. Her expertise and suggestions play a key role in finishing submissions to conferences on computer vision and pattern recognition. I would also like to thank Prof. Constantino Lagoa at Penn State University, who gave me a lot of suggestions during our discussion on polynomial optimization and superstability.

I would like to thank Prof. Jennifer Dy and Prof. Rifat Sipahi who serve as the committee members of my PhD program, for their precious advise and challenging problems on my dissertation, without which, it would not be as good as it is.

I owe a special debt to Prof. Gilead Tadmor, for encouraging me to follow my heart when I came to Northeastern University six years ago. Without his help, I would have no chance working with my current adviser and I would not be as happy as I am now.

I would like to thank all my labmates in Robust Systems Lab at Northeastern University. I feel so fortunate to work with these excellent guys for six years and build up life long friendship with them. I have learnt so much from them during the ordinary discussion on various research topics. Thanks to them for their openness and kindness injecting some funny element to the long journey of academic research.

Last but not least, many thanks from the bottom of my heart go to my family, my parents and my brother. Thank my parents for their endless love and selfless support to me whenever I am successful or defeated, and for their on-time daily calls for six years in spite of physical distance and jet lag. Thank my brother for letting me know he is always by my side and ready for solving the troubles for me whenever I need help. The unconditional love from my family supports me to keep on fathoming the unknown and making progress.
Abstract of the Dissertation

Robust Model Fitting via Convex Optimization Techniques

by

Yongfang Cheng

Doctor of Philosophy in Electrical and Computer Engineering
Northeastern University, August 2016
Dr. Mario Sznaier, Adviser

The past few years have witnessed an unprecedented growth in data acquisition capabilities due to the tremendous advances in information sensing. Although these developments have numerous potential applications, ranging from smart grids to self-aware environments and smart cities, they present new challenges to traditional data analysis techniques including search, storage, transfer and visualization, to name a few. These challenges originate from two aspects: one is the curse of dimensionality caused by the huge amount of the data; the other is the quality degradation of the data caused by various noise.

As a simple and efficient way of compressing data, regression model fitting, has obtained increasing interests in pattern analysis and signal processing due to the fact that the data of huge amount often consists of several types of structure that are easy for processing. Most of the existing methods inherit two weaknesses. One is that the noise is assumed to be stochastic, therefore, the accuracy of the result highly depends on whether the data is sufficient and consistent with the assumption of its statistical distribution; the other is that for switched regression models, the clustering and identification are achieved alternatively instead of simultaneously, ending up with a local optimal solution.

To circumvent these challenges and difficulties, in this dissertation, we aim to develop efficient algorithms for model fitting from noisy data appealing to convex optimization techniques, such as polynomial optimization, atomic norm minimization and robust optimization. Firstly, we start with the simplest case, robust single regression model identification, followed by semi-supervised support vector machines and switched regression models, which fit the data to a finite number of regression models. Secondly, we extend the results on static data to scenarios with dynamical data which are more difficult due to the interdependence of samples because of the temporal correlation. For dynamical data, we also consider the linear parameter varying model governed by a continuous
mode variable in the form of linear fractional transformation rather than a discrete mode variable in switched regression model. Thirdly, we consider the model invalidation problem which is a key step before further usage of the model. Finally, contrary to build a model from the data, we directly establish a robust superstable controller for all possible models consistent with the experimental data skipping the model identification step.
Chapter 1

Introduction

Many problems of practical interest involve fitting models to a given set of sample points, which is known as subspace clustering. Examples of applications include, among others, image compression [1], face recognition under varying illumination [2], motion segmentation [3], segmentation of video sequences [4] in computer vision and machine learning societies, and system identification [5] in control community.

As we know, if the data are free from noise and the assumptions on the models for instance the number of submodels and the order of submodels are correct, then the problem of model fitting is equivalent to solving a set of linear regression equations. However, in many scenarios of practical interest, the data are corrupted by intensive noise and even outliers which may be caused by faults in sensors, making the model fitting problem nontrivial. Next, by a toy example, we summarize the challenges which make this problem difficult.

1.1 Challenges

Consider the problem of modeling the data points \( \{ x_j \in \mathbb{R}^2 \}_{j=1}^{N_p} \), illustrated in Figure 1.1 with a union of subspaces.

- *The curse of dimensionality*. Recent advances in sensing have made feasible a huge amount of data, driving us into the era of “Big Data”, challenging the computational tractability of the existing approaches to model identification which scale at least polynomially in the number of data points.
CHAPTER 1. INTRODUCTION

Figure 1.1: Data for the Toy Example (Left: Clean Data; Middle: Data Contaminated Noise vs. Right: Expected Clustering)

- The number of subspaces and their dimensions are unknown. If there is no constraint on these two parameters, then we can always fit a single subspace to all the data points by standard Principal Component Analysis (PCA); or we can fit the data with \( N_p \) different affine subspaces, that is, one subspace per data point. Neither of the two trivial solutions are satisfying, while the ideal clustering is to fit three one-dimensional subspaces to all the data as shown in Figure 1.1.

- The requirement of clustering data and identifying subspaces simultaneously. Under the circumstance that the clustering is known, then the subspaces can be got by fitting one subspace to each cluster of data. On the other hand, under the circumstance that subspaces are known, it is much easier to determine to which subspace a data point belongs. However, in most practical cases, data clustering and subspaces identification are required simultaneously.

- Data are usually contaminated by noise. Noise on data is not uncommon due to the limited resolution of measuring instruments or calibration, as shown in Figure 1.1. More seriously, false manipulation by humans or fault of sensors may lead to outliers (as shown by larger dots in Figure 1.2), denoting those data which live too far away from the subspaces the other data belongs to and which can distort the subspaces severely.

- Structure of data. The data considered can be classified into static data and dynamical data. Static data means the data points are independent from each other, while dynamical data means the data correlate with each other, which are ubiquitous in control community, for instance signals generated by physical systems. For dynamical data, noise corrupts the data in two ways:
equation error structure and Error-In-Variables (EIV) structure. For the first case, a single error term is added to the model equation taking into account all possible sources of uncertainty affecting the output. In contrast, for the second case, the data are corrupted by noise entry-wise, and consequentially the correlation between data points changes the sparsity pattern exhibited by the problem, further affecting the computational complexity of corresponding algorithms.

![Figure 1.2: Subspaces without Outliers vs. Subspaces Skewed by Outliers](image)

**1.2 Previous work**

Due to its importance and the challenges enumerated above, during the past few decades, substantial research efforts have been devoted to subspace clustering, leading to many algorithms, which can be roughly classified into statistical [6,7], algebraic [5,8–10], spectral clustering based methods [4,11–13] and support vector regression methods [14–19]. Next, we will analyze the advantages and disadvantages of several representative methods.

- **Statistical methods.** RANdom SAmple Consensus (RANSAC) [6] is an iterative approach that at each iteration one subspace is fitted to randomly sampled data points, and then among the other points as many as possible points are selected as the inliers of the subspace. The process is repeated until the percentage of inliers has exceeded a given threshold. While in principle this method is robust against outliers, the number of iterations it needs may increase exponentially with the increasing of the number of subspaces and the percentage taken up by the outliers, due to the fact the probability of choosing several data points defining one original subspace decreases. Conversely, because of its random nature, limiting the number of
iterations may lead to arbitrarily bad results. Mixtures of Probabilistic Principal Component Analyzer (MPPCA) is another statistical method which assumes the distribution of the data inside each subspace and the noise are both Gaussian, and parameters are estimated using an Expectation-Maximization (EM) algorithm. Despite its simplicity, its main drawback is once the assumption on the distribution is inconsistent with the real case, it may lead to unsatisfactory results.

- **Algebraic methods.** Generalized Principle Component Analysis (GPCA) \[5,8,9\], making use of the hybrid decoupling constraint, exploits the properties of subspace clustering by reducing the problem to estimating the coefficients of a multivariate polynomial from the noisy measurements of its zeros. Once the polynomial is determined, the parameters of subspaces can be recovered via polynomial differentiation at each data point. While GPCA can perfectly recover the subspaces in case of clean data, its performance deteriorates as the noise level increases. This drawback motivated the robust version of GPCA, the moments based GPCA in \[10,20,21\], where the original noisy data is denoised via rank minimization, prior to implement GPCA. However, in the presence of noise, all these methods entail a common risk that there are no guarantees that the resulting polynomial can be factorized as a product of several linear polynomials.

- **Self-representation based methods.** To be robust against noise, several methods proposed recently, including Sparse Subspace Clustering (SSC) \[12\], Low Rank Representation (LRR) \[11\], Fixed Rank Representation (FRR) \[13\], and Robust Subspace Clustering (RSC) \[4\], exploits the geometric properties of subspaces by reducing the problem to looking for sparse or low rank solutions to a set of linear equations that encode the fact that subspaces are self-expressive (i.e., each point on a subspace can be expressed as a linear combination of other points on it). Despite the advantage that they can handle data of much higher dimensions than GPCA, these methods share some common weaknesses. Firstly, all these methods typically involve relaxations, such as nuclear norm for rank and $\ell_1$ norm for sparsity, to obtain tractable convex problems. However, in the presence of noise there is no guarantee that these relaxations are exact, especially in cases arising in systems identification, where the data points are correlated. Secondly, recovering parameters of subspaces requires performing spectral clustering first to classify the data. Thus, there is no direct control on the fitting error.
CHAPTER 1. INTRODUCTION

- **Support vector regression methods.** A common feature of the methods mentioned above is that the clustering and the identification of subspaces are performed in sequence instead of simultaneously. Thus, the parameters of subspaces are not optimized directly. In control community, a lot of existing algorithms for hybrid system identification are based on the idea of support vector regression. That is, points on subspaces are equivalently described as linear regressions, and the problem is recast in a framework of minimization of either the minimum or the product of the loss functions [15][16][18][19]. The advantage of these methods is the optimization is directly performed with respect to the parameters of subspaces. On the other hand, the nonconvex nature of this framework needs various kinds of relaxation techniques to find a satisfactory solution. For instance, the Multilevel Coordinate Search (MCS) algorithm [22] is used in [16], Difference of Convex functions programming (DC) is used in [19], and heuristic of reweighted ℓ₁ norm is used in [17], [15] and [18]. Usually the equivalence between the relaxation and the original nonconvex problem depends on some specific conditions on the data points, which is too difficult to verify in practice.

The existing methods are able to circumvent some of the difficulties mentioned in Section 1.1 and produce satisfactory results in some cases. However, the general subspace clustering solutions can further be improved in many ways, such as

- Most of the existing methods of subspace clustering, except [10][20][21], assume the noise is stochastic, and do not enforce any hard constraints on the noise level. Thus, the resulted subspace may be severely skewed once an outlier is mistakenly clustered as an inlier. Although this drawback is made up for in [10][20][21] by enforcing constraints on noise, as extensions of GPCA, they also inherit some of the drawbacks of GPCA such as the sensitiveness to outliers and the factorability of the resulting high-order polynomial to product of several linear polynomials. Therefore, to exploit an algorithm which can recover subspaces and restrict the fitting error within a certain level is meaningful.

- Practically, the source of data can give us some prior knowledge on the distribution of the data, for instance, in surveillance videos, we can estimate the percentage of data fitting to different subspaces according to the areas of background, pedestrians, and moving vehicles. Also in biological systems, due to the fact that many mechanisms are not invertible, the submodels used to describe different mechanisms are prohibited to switch arbitrarily. How to make use of these types of prior information to increase the accuracy of modeling is promising.
CHAPTER 1. INTRODUCTION

- For dynamical data, while most of literatures on hybrid system identification propose methods robust against error in equations, the identification of hybrid systems from data corrupted by noise in EIV structure is still an open issue.

1.3 Contributions

Motivated by the weakness of the existing methods, in this dissertation, under the assumption that the noise of the data is deterministic and $\ell_\infty$-norm bounded, we address several nonconvex problems arising in model fitting by resorting to recent developments in convex optimization. Firstly we consider modeling the given data into three types of models: single linear regression model, switched linear regression model, and Linear Parameter Varying (LPV) systems with Linear Fractional Transformation (LFT) parametric dependence. For the dynamical data, we consider noise corrupting the data in two ways: error in equation and error in variables. Besides, we also consider the model (in)validation problem which is of great importance before further usage of the identified model. Finally we give some inspiration on establishing controllers robust to the entire consistent set of models directly without model fitting. Specifically, we aim to solve the following problems.

1.3.1 Static Data

- Robust Linear Regression Model Fitting. In spite of a lot of existing methods on identifying a linear regression model, their performance deteriorates dramatically when the data is mixed with a large percentage of outliers. By introducing a binary variable for each sample to indicate whether it is an outlier or not, we reformulate the model fitting problem as a constrained Polynomial Optimization Problem (POP) to seek for the model minimizing the number of outliers. Theoretically we show that by this formulation the global optimality is guaranteed, and practically we propose efficient algorithms to solve the nonconvex problem by its moments-based convex relaxation.

- Semi-supervised support vector machine ($S^3$VM). Two-class Support Vector Machine (SVM) aims to find two parallel hyperplanes which classifies a cloud of data into two clusters in the sense of maximizing margin. The standard SVM, where labels of all the training data is known, is a convex quadratic programming. On the other hand, in cases where a large portion of training data is unlabeled, the general framework of SVM used to find a classifier and
CHAPTER 1. INTRODUCTION

determine the unknown labels simultaneously, the problem named $S^3$VM becomes nonconvex, but also a constrained POP, which can be solved by moments-based convex relaxation.

- **Subspace clustering with priors.** Under the assumption that the number of subspaces and their dimensions are known, given a bound on the $\ell_\infty$ of the noise, we aim to develop a general framework for subspace clustering which can cluster the data and recover the subspaces simultaneously and a set of noise consistent with the given bound is guaranteed to exist. The proposed framework is also aimed to flexibly incorporate existing prior information.

1.3.2 Dynamical Data

- **EIV Linear Time Invariant (LTI) system identification.** In the case of EIV noise, the noise terms entering the regression equation are interdependent from each other, which makes the problem of model fitting harder. Comparing to the existing work which looks for a low rank approximation to a matrix of the size equal to the length of the data, we attempt to find a low rank approximation to the Gramian matrix whose size is independent from the size of the data. As a alternative, we also formulate the problem as a constrained POP problem, which is versatile of incorporating additional constraints like co-occurrence and rank deficient constraints on the model parameters.

- **EIV Switched Auto-Regressive eXogenous (SARX) system identification.** We aim to extend the general framework of subspace clustering for static data to SARX modeling for dynamical data corrupted by measurement noise. Compared to the error in equation structure, the resulted POP is of higher order due to the product term of the model parameter and the noise, increasing the minimum order of relaxation based on moments theory. By enforcing low rank and sparse solution, we develop an algorithm to obtain a model consistent with the data based on the lowest order convex relaxation.

- **Identification of LPV systems with LFT parametric dependence.** In contrast to SARX model where the data is fitted to a finite number of linear submodels and the model variable is a finite integer, in many scenarios of practical interest an alternative to modeling dynamical data is LPV systems with LFT parametric dependence where the mode variable, named scheduling parameter, is a continuous variable instead. We develop a simple approach to identifying LPV systems with LFT dependence on a measurable and manipulatable scheduling parameter from the input/output data via atomic norm minimization.
CHAPTER 1. INTRODUCTION

• **EIV SARX model (in)validation.** A key step before using the identified models is to (in)validate them by extra experimental data. For EIV SARX models, in the case where the discrete mode variable is unavailable directly, the (in)validation problem is a special case of EIV SARX model identification with known model parameters and also nonconvex. We aim to formulate this problem into the general framework of constrained POP and provide necessary and sufficient convex (in)validation certificates.

• **Robust superstable controller design for LTI systems.** The framework of model identification in this dissertation aims to obtain a single model in the consistent set defined by the given data. In control theory, before the identified model is used for controller synthesis, generally an uncertainty bound will be estimated, which is usually approximated from outside by a simple set for instance, by a polyhedron or an ellipsoid, introducing the conservativeness. In the last part of this dissertation we propose a framework to build a controller stabilizing all the plants in the original consistent set directly without identifying a model for the plant.

The dissertation is organized as follows: In Chapter 2 we summarize the notations and some knowledge on convex optimization which plays a key role in deriving the main results in this dissertation; In Chapters 3-5 the model fitting problem for static data is considered; In Chapters 6-9 we consider the model fitting problem for dynamical data; In Chapter 10 we consider designing a robust controller stabilizing all the plants in the original consistent set defined by the experimental data without model identification step.
Chapter 2

Preliminaries

For ease of reference, in this chapter we summarize the notations used throughout this dissertation and recall some basic concepts in classification and clustering and some results in convex optimization techniques required to establish the results in the following chapters.

2.1 Notations

⊗ Kronecker product
C set of complex number
R set of real number
N positive integers
Z non-negative integers
N\{\text{i : }\in \mathbb{N}, 1 \leq i \leq k\} \{i :\in \mathbb{N}, 1 \leq i \leq k\}
x a vector in \mathbb{R}^n
M a matrix in \mathbb{R}^{n \times m}
I Identity matrix
<M, N> trace(M^T N)
\sigma_i(M) the \text{i}-th largest singular value of M
\|x\|_\infty \ell_\infty\text{-norm of the vector } x \in \mathbb{R}^n, \|x\|_\infty = \sup_i |x_i|
\|x\|_p (0 < p < \infty) \ell_p\text{-norm of the vector } x \in \mathbb{R}^n, \|x\|_p = \left( \sum_i |x_i|^p \right)^{1/p}
CHAPTER 2. PRELIMINARIES

\[ \|x\|_0 = \ell_0 \text{ quasinorm of the vector } x \in \mathbb{R}^n \]
\[ \|x\|_0 = \text{number of non-zero elements in } x \]
\[ M \succeq N \]
\[ \text{the matrix } M - N \text{ is positive semidefinite} \]
\[ E_\mu[f(x)] = \text{expectation of } f(x), \mu(x) \text{ is a Borel measure of } x, \ E_\mu[f(x)] = \int f(x)d\mu(x) \]
\[ m(x^\alpha) = \text{moment of order } \alpha \text{ of } \mu(x), \ m(x^\alpha) = \int x^\alpha d\mu(x) \]
\[ \mathbb{D}_\rho = \text{a closed disc in } \mathbb{C} \text{ centered at the origin with radius } \rho \]
\[ \text{conv}(A) = \text{convex hull of the set } A \]
\[ \Re(x), \Im(x) = \text{real part of } x \in \mathbb{C}, \text{imaginary part of } x \in \mathbb{C} \]
\[ h \ast u = \text{convolution between } h \text{ and } u \]
\[ T_x = \text{lower triangular block toeplitz matrix associated with any column vector} \]
\[ x = [x_0 \ x_1 \ \cdots \ x_{n-1}]^T \]
\[ T_x = \begin{bmatrix} x_0 & 0 & \cdots & 0 \\ x_1 & x_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ x_{n-1} & x_{n-2} & \cdots & x_0 \end{bmatrix} \]
\[ \text{Proj}_x \mathcal{P}(x, y) = \{ x : \forall x \in \text{Proj}_x \mathcal{P}(x, y), \exists y, \text{ such that } (x, y) \in \mathcal{P}(x, y) \} \]

2.2 Moments-Based Polynomial Optimization

**Definition 1.** A polynomial \( p(x) \) is said to be a sum-of-squares polynomial (SoS), if it can be written as \( p(x) = \sum_{j=1}^{m} u_j(x)^2 \) for some polynomials \( u_1(x), \ldots, u_m(x) \).

**Definition 2.** A \( s \)-atomic measure is a measure with \( s \) atoms, that is, a linear positive combination of \( s \) Dirac measures. Specifically, a measure with a single atom is a Dirac measure.

**Definition 3.** Let

\[ \left\{ x^\alpha = \prod_{i=1}^{n} x_i^{\alpha_i} \right\}, \forall \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{Z}^n \text{ and } |\alpha| = \sum_{i=1}^{n} \alpha_i \leq N \quad (2.1) \]

be a basis for \( N \)-degree real-valued polynomial and its dimension is \( s_N = \binom{N+n}{N} \). The multiindices \( \alpha \) are arranged according to grevlex order on monomials so that \( 0 = \alpha^{(1)} < \alpha^{(2)} < \cdots < \alpha^{(s_N)} \). Then for a given Borel measure \( \mu(x) \), its moments sequence of order up to \( N \) is defined as
CHAPTER 2. PRELIMINARIES

\( m_N \in \mathbb{R}^{sN} \)

\[ \forall_{i=1}^{sN} : m_N(i) = \int x^{\alpha(i)} \, d\mu(x) = m_{\alpha(i)} \]  

(2.2)

and its \( N \)-th order moment matrix \( M_N(m_{2N}) \in \mathbb{R}^{sN \times sN} \) is defined as

\[ \forall_{i=1}^{sN} \forall_{j=1}^{sN} : M_N(i, j) = \int x^{\alpha(i) + \alpha(j)} \, d\mu(x) = m_{\alpha(i) + \alpha(j)} \]  

(2.3)

**Definition 4.** Given a polynomial of the form \( g(x) = \sum_{k=1}^{l} g_k x^{\beta(k)} \), let \( \delta \) denote the degree of \( g(x) \), then its \((N - \lceil \frac{\delta}{2} \rceil)\)-th localizing matrix \( L_{N - \lceil \frac{\delta}{2} \rceil}(g m_{2N}) \in \mathbb{R}^{sN \times sN} \) is defined as

\[ \forall_{i=1}^{sN} \forall_{j=1}^{sN} : L_{N - \lceil \frac{\delta}{2} \rceil}(i, j) = \int \sum_{k=1}^{l} g_k x^{\beta(k) + \alpha(i) + \alpha(j)} \, d\mu(x) = \sum_{k=1}^{l} g_k m_{\beta(k) + \alpha(i) + \alpha(j)} \]  

(2.4)

For instance, when \( n = 2 \) and \( N = 2 \), one obtains that

\[ m_2 = [1, m_{0,1}, m_{1,0}, m_{0,2}, m_{1,1}, m_{2,0}] \]
\[ m_4 = [1, m_{0,1}, m_{1,0}, m_{0,2}, m_{1,1}, m_{2,0}, m_{0,3}, m_{1,2}, m_{2,1}, m_{3,0}] \]  

(2.5)

and

\[ M_2(m_4) = \begin{bmatrix} 1 & m_{0,1} & m_{1,0} & m_{0,2} & m_{1,1} & m_{2,0} \\ m_{0,1} & m_{0,2} & m_{1,1} & m_{0,3} & m_{1,2} & m_{2,1} \\ m_{1,0} & m_{1,1} & m_{2,0} & m_{1,2} & m_{2,1} & m_{3,0} \\ m_{0,2} & m_{0,3} & m_{1,2} & m_{0,4} & m_{1,3} & m_{2,2} \\ m_{1,1} & m_{1,2} & m_{2,1} & m_{1,3} & m_{2,2} & m_{3,1} \\ m_{2,0} & m_{2,1} & m_{3,0} & m_{2,2} & m_{3,1} & m_{4,0} \end{bmatrix} \]  

(2.6)

The 1st order localizing matrix associated with the polynomial \( g(x) = a - x_1^2 - x_2^2 \) is

\[ L_1(g m_4) = \begin{bmatrix} a - m_{2,0} - m_{0,2} & am_{0,1} - m_{2,1} - m_{0,3} & am_{1,0} - m_{3,0} - m_{1,2} \\ am_{0,1} - m_{2,1} - m_{0,3} & am_{0,2} - m_{2,2} - m_{0,4} & am_{1,1} - m_{3,1} - m_{1,3} \\ am_{1,0} - m_{3,0} - m_{1,2} & am_{1,1} - m_{3,1} - m_{1,3} & am_{2,0} - m_{4,0} - m_{2,2} \end{bmatrix} \]  

(2.7)

**2.2.1 Moment Extension Problem**

Given a finite multi-sequence of scalars \( m_{2N} = \{ m_\alpha \} \in \mathbb{R}^{s2N} \), with \( \alpha \in \mathbb{Z}^n \) and \( |\alpha| \leq 2N \), build the matrix \( M_N(m_{2N}) \) as in (3), and \( M_N(m_{2N}) \succeq 0 \) holds. The moment extension problem is defined as follows: is it possible to extend the sequence \( m_{2N} \) with new scalars...
CHAPTER 2. PRELIMINARIES

\{m_\beta, 2N < |\beta| \leq 2(N + 1)\} to \mathbf{m}_{2(N+1)}, such that \mathbf{M}_{N+1}(\mathbf{m}_{2(N+1)}) \succeq 0?

If such an extension exists, then \mathbf{M}_{N+1} is called a positive extension of \mathbf{M}_N. If in addition, \text{rank}\{\mathbf{M}_{N+1}\} = \text{rank}\{\mathbf{M}_N\} holds, then \mathbf{M}_{N+1} is called a flat extension of \mathbf{M}_N.

**Theorem 1.** (Theorem 3.7 in [24]) Let \(m_{2N} \in \mathbb{R}^{s_{2N}}, \) with \(\alpha \in \mathbb{Z}^n\) and \(|\alpha| \leq 2N\).

If such an extension exists, then \(\mathbf{M}_{N+1}\) is called a positive extension of \(\mathbf{M}_N\). If in addition,

\[
\text{rank}\{\mathbf{M}_{N+1}\} = \text{rank}\{\mathbf{M}_N\}
\]

holds, then \(\mathbf{M}_{N+1}\) is called a flat extension of \(\mathbf{M}_N\).

### 2.2.2 \(K\)-Moment Problem

Consider the general constrained polynomial optimization problem in the form of

\[
\mathbf{\tilde{p}}^* := \min_{\mu \in \mathcal{P}(K)} \int p(x) d\mu(x) := \min_{\mu \in \mathcal{P}(K)} \mathbb{E}_\mu[p(x)] = \min_{\mu \in \mathcal{P}(K)} \mathbb{E}_\mu[p(x)]
\]

where \(p(x)\) is an \(m\)-degree real valued polynomial, and \(K\) is a compact semi-algebraic set defined by polynomial inequalities \(g_k(x) = \sum_{\beta} g_{k,\beta} x^\beta \geq 0\) of degree \(\delta_k, k = 1, \ldots, d,\) which can describe a lot of constraints of interest, for instance, \(x_i\) should be binary is equivalent to \(x_i - x_i^2 \geq 0\) and \(x_i^2 - x_i \geq 0.\) Generally, Problem (P1) is nonconvex, and difficult to solve. Instead, we consider a related problem

\[
\tilde{p}^* := \min_{\mu \in \mathcal{P}(K)} \mathbb{E}_\mu[p(x)]
\]

where \(\mathcal{P}(K)\) is the space of Borel measures with support constrained on \(K.\) Although (P2) is of infinite dimension, it is, in contrast to (P1), convex. As shown in [25] (Proposition 2.1), (P2) is equivalent to (P1) in the sense that

- \(p^* = \tilde{p}^*;\)
- for every optimal solution \(\mu^*\) of (P2), \(p(x) = p^*, \mu^*-\text{almost everywhere.}\)

Note that \(\mathbb{E}_\mu[p(x)] = \sum_\alpha p_\alpha \int x^\alpha d\mu(x) = \sum_\alpha p_\alpha m_\alpha,\) which is the linear functional of the moment sequence \(\{m_\alpha\}.\) Inspired by the results in Section 2.2.1, it follows that (P1) can be reduced to a sequence of Linear Matrix Inequalities (LMI) optimization problems in terms of the moment
sequence of some unknown Borel measure of the form

$$p_N^* = \min_{m_{2N}} \sum_{\alpha} p_\alpha m_\alpha$$

subject to

$$M_N \preceq 0$$
$$L_N - \left\lceil \delta_k \right\rceil (g_k m_{2N}) \succeq 0, \quad \forall k = 1, \ldots, d$$

(2.8)

Further it can be shown (see [25, 26]) that as $N$ goes to infinity, $p_N^*$ approaches $p^*$ from below. The next result gives a necessary and sufficient condition for the relaxation corresponding to a given finite $N$ to be exact.

**Theorem 2.** (Theorem 4.2 in [25]) The relaxation of order $N$ is exact if and only if there exist sum of square polynomials $u_0$ of degree at most $2N$ and $u_k$ of degree at most $2N - \delta_k$ such that

$$p(x) - p^* = u_0(x) + \sum_{k=1}^d u_k(x) g_k(x)$$

**Lemma 1.** Given $a$ and $b \in \mathbb{R}^n$, let $p^* = \min_{\|x\|_\infty \leq \epsilon} \|a + b^T x\|_2^2$. Then, for any $q \leq p^*$, the polynomial

$$p(x) = \|a + b^T x\|_2^2 - q$$

has a representation of the form $p(x) = p_0(x) + \sum_\iota \lambda_\iota (\epsilon^2 - x_\iota^2)$ where $p_0$ is a SoS polynomial of at most degree 2 and $\lambda_\iota \geq 0$.

**Proof.** From Lemma 3.10 in [27] we have that $\|a + b^T x\|_2^2 - p^* = \tilde{p}_0(x) + \sum_\iota \lambda_\iota (\epsilon^2 - x_\iota^2)$, for some $\lambda_\iota \geq 0$ and some second order SoS polynomial $\tilde{p}_0$. Hence

$$p = (a + b^T x)^2 - q = \tilde{p}_0 + p^* - q + \sum_\iota \lambda_\iota (\epsilon^2 - x_\iota^2)$$

Since $\tilde{p}_0$ is SoS and $q \leq p^*$, then $p_0$ is also SoS.

2.2.3 Sparse Polynomial Optimization

**Definition 5.** Considering a polynomial optimization problem of the form (P1), assume that the polynomial $p$ can be partitioned into $p = p_1 + \cdots + p_l$ such that each $p_j$ and each $g_k$ contain only variables in a subset $v_j \subseteq x$, and $\bigcup_{j=1}^l v_j = x$. If there exists a reordering $v'$ of $v_j$ such that for
CHAPTER 2. PRELIMINARIES

every \( i' = 1, \ldots, l - 1 \):

\[
\mathbf{v}_{i' + 1} \cap \bigcup_{j=1}^{i'} \mathbf{v}_j \subseteq \mathbf{v}_s \quad \text{for some } s \leq i'
\]  

(2.9)

then the running intersection property is satisfied.

The significance of this property is that it allows for substantial computational complexity reduction. Specifically, in the case of generic polynomials, a moments relaxation of order \( N \) entails considering moment and localizing matrices containing \( O(n^{2N}) \) variables denoted by \( \mathbf{m}_{2N} \) in (2.8).

On the other hand, if the running intersection property holds, it can be shown (in [26, 28]) that it is possible to define \( l \) sets of smaller sized matrices each containing only variables in \( \mathbf{v}_j \) (i.e. number of variables is \( O(\kappa^{2N}) \), denoted by \( \mathbf{m}_{j,2N} \), \( j = 1, \ldots, l \) in (2.10), where \( \kappa \) is the maximum cardinality of \( I_i \)). Since typically \( \kappa \ll n \), this leads to substantial reduction in the number of variables in the optimization problem (and hence its computational complexity) of the form

\[
p^*_N = \min_{\mathbf{m}_{j,2N}} \sum_{j=1}^{l} c_j^T \mathbf{m}_{j,2N}
\]

subject to

\[
\forall_{j=1}^l : \quad \mathbf{M}_{j,N} \left( \mathbf{m}_{j,2N} \right) \succeq 0
\]

\[
\forall_{k=1}^l : \quad \mathbf{L}_{j,N - \lceil \frac{\delta}{2} \rceil} \left( g_k \mathbf{m}_{j,2N} \right) \succeq 0, \quad \{ \text{variables in } g_k \} = \mathbf{v}_j.
\]  

(2.10)

2.3 Support Vector Machine

Support Vector Machine (SVM) [29] is a supervised learning model for classification and regression analysis of data or patterns based on finding maximum margin hyperplanes. A two-class classification problem by SVM is described as: given the training data set comprising \( N_p \) pairs of data, \( \{ \mathbf{x}_i, t_i \}_{i=1}^{N_p} \), with \( \mathbf{x}_i \in \mathbb{R}^n \) and \( t_i \in \{-1, +1\} \), find a linear model of the form

\[
y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b
\]  

(2.11)

where \( \phi(\mathbf{x}) \) denotes a fixed feature-space transformation, such that new data points \( \mathbf{x} \) are classified according to the sign of \( y(\mathbf{x}) \).
CHAPTER 2. PRELIMINARIES

2.3.1 Primal SVM

Under the assumption that the given data are linearly separable in the feature space \( \phi(x) \), the hard-margin SVM is to solve the optimization problem

\[
\{w^*, b^*\} = \arg\min_{w, b} \frac{1}{2} ||w||^2 \quad \text{subject to} \quad t_i(w^T \phi(x_i) + b^*) - 1 \geq 0 \quad \text{for} \quad i = 1, \ldots, N_p.
\]  

(2.12)

Consider the toy example as shown in the left figure in Figure 2.1, \( \phi(x) = x \), the hyperplane \( w^T x + b^* = 0 \) is the decision boundary. The margin, which is defined as the perpendicular distance between the decision boundary and the data point closest to it, \( 1/||w||_2 \) is maximized by solving (2.12). Those data points locating on the hyperplanes (margin boundaries) \( t(x)(w^T x + b^*) = 1 \) are known as support vectors, which are indicated by green circles. No data points are locating within the gap between the two hyperplanes.

The resulted SVM from (2.12) gives exact separation of the training data in the original input space \( x \). However, in practice, the class-conditional distribution may overlap, in which case exact separation of the training data can lead to poor generalization. Therefore, soft-margin SVM is developed, in which data points are allowed to be on the “wrong side” of the margin boundary, but with a penalty that increases with the distance from the boundary. Mathematically, it is formulated as the optimization problem

\[
\{w^*, b^*\} = \arg\min_{w, b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N_p} \xi_i \quad \text{subject to} \quad t_i(w^T \phi(x_i) + b) - 1 + \xi_i \geq 0, \quad \forall_{i=1}^{N_p} \xi_i \geq 0
\]  

(2.13a)

(2.13b)

(2.13c)

where the parameter \( C > 0 \) controls the trade-off between the penalty and the margin. As shown in the right figure in Figure 2.1, the soft-margin SVM increases the margin, the points locating on and within the gap of two hyperplanes \( t(x)(w^T \phi(x) + b^*) = 1 \) are support vectors (in green circles), and those points with \( \xi_i > 1 \) are misclassified.
CHAPTER 2. PRELIMINARIES

2.3.2 Dual SVM

Let \( K \in \mathbb{R}^{N_p \times N_p} \), with \( K(i, j) = \langle \phi(x_i), \phi(x_j) \rangle \), denote the kernel matrix for the training data. The Lagrangian dual problem associated with (2.13) is

\[
\{\alpha^*_i\} = \arg \min_{\alpha_i} \frac{1}{2} \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} t_i t_j \alpha_i \alpha_j K(i, j) - \sum_{i=1}^{N_p} \alpha_i \quad \text{subject to } 0 \leq \alpha_i \leq C, \forall i \in \{1, 2, \ldots, N_p\}
\]  

(2.14)

where \( \{\alpha_i\}^{N_p}_{i=1} \) are the dual variables associated with the constraints (2.13b). Those \( \phi(x_i) \) corresponding to \( \alpha^*_i > 0 \) are support vectors, and the primal variables can be calculated by

\[
w^* = \sum_{i=1}^{N_p} \alpha^*_i t_i \phi(x_i)
\]

\[
b^* = \frac{\sum_{j \in I} (1-t_j \phi(x_j)-\xi_j)}{\sum_{j \in I} t_j}, I = \{i : \alpha^*_i > 0\}
\]

(2.15)

One of the major advantages of the dual formulation lies in its capability of handling the patterns \( \phi(x) \) of infinite dimensions.

2.4 Atomic Norms in LTI identification

Definition 6. (30) Let \( A \) be a collection of atoms that is a compact subset of \( \mathbb{R}^p \). Assume that \( A \) is centrally symmetric about the origin and that no element \( a \in A \) lies in the convex hull of the other

\[16\]
elements \( \text{conv}(A \setminus a) \). Then the gauge function of \( A \), \( \| \cdot \|_A \),

\[
\| x \|_A = \inf \left\{ t > 0 : x \in t \text{conv}(A) \right\} = \inf \left\{ \sum_{a \in A} c_a : x = \sum_{a \in A} c_a a, c_a \geq 0 \forall a \in A \right\},
\]

is called the atomic norm induced by \( A \).

Atomic norm regularization was first introduced for identification of low order LTI models in [31], where \( A \) consists of all the single poles resulted from \( \epsilon \)-net discretization of the disk \( D_\rho \). In this case a low order model that interpolates the experimental data is obtained by \( \ell_1 \) regularization which enforces the sparsity of the resulting representation. The idea was further extended in [32] by taking into account the fact that the poles of real rational transfer functions are complex conjugate, and thus the set \( A \) is composed by

\[
\begin{align*}
\mathcal{A} &= \mathcal{A}_+ \cup (-\mathcal{A}_+), \quad \mathcal{A}_+ &= \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4 \\
\mathcal{A}_1 &= \left\{ \Psi_p(z) = \frac{1 - |p|^2}{2} \left( \frac{1}{z - p} + \frac{1}{z - p^*} \right) : p \in D_\rho, p \notin \mathbb{R} \right\} \\
\mathcal{A}_2 &= \left\{ \Psi_p(z) = \frac{1 - |p|^2}{2} \left( \frac{-j}{z - p} + \frac{j}{z - p^*} \right) : p \in D_\rho, p \notin \mathbb{R} \right\} \\
\mathcal{A}_3 &= \{ \Psi_p(z) = 1 \} \\
\mathcal{A}_4 &= \left\{ \Psi_p(z) = \frac{1 - |p|^2}{z - p} : z \in [-\rho, \rho] \right\}
\end{align*}
\]

where \( p^* \) denotes the complex conjugate of \( p \). With the choice of \( A \) as in (2.17), each proper rational transfer function \( G(z) \) can be written as a linear combination of atoms with real coefficients, i.e.,

\[
G(z) = \sum_{a \in A} c_a \Psi_p(z), \quad \text{where } c_a \in \mathbb{R}.
\]

Correspondingly, the impulse response of \( G(z) \), \( h \), can be represented as the linear combination of the impulse response of the atoms, \( H[\Psi_p(z)] \), i.e.,

\[
h = H[G(z)] = \sum_{a \in A} c_a H[\Psi_p(z)],
\]
CHAPTER 2. PRELIMINARIES

where for all \( i = 0, 1, 2, \ldots \),

\[
H[\Psi_p(z)]_i = \begin{cases} 
\pm (1 - |p|^2)\Re(p^i) & \text{for } a \in A_1 \\
\pm (1 - |p|^2)\Im(p^i) & \text{for } a \in A_2 \\
\pm \delta(i) & \text{for } a \in A_3 \\
\pm (1 - |p|^2)p^i & \text{for } a \in A_4.
\end{cases}
\]

(2.20)

Clearly, the smaller the cardinality of \( c_a \) is, the lower order \( G(z) \) has. Due to the fact that the forced response to input \( \{u_i\}_{i=0}^{N-1} \) is

\[
y_i = (h * u)_i = \sum_{k=0}^{i} h_{i-k} u_k = \sum_{k=0}^{i} h_k u_{i-k},
\]

(2.21)

which can be written in a compact form

\[
y = T_h u = T_u h
\]

(2.22)

where \( y = [y_0, y_1, \cdots, y_{N-1}]^T \), \( u = [u_0, u_1, \cdots, u_{N-1}]^T \) and \( h = [h_0, h_1, \cdots, h_{N-1}]^T \), the identification problem of LTI systems from experimental data \( \{u_i, y_i\}_{i=1}^{N-1} \) can be formulated as the constrained optimization problem

\[
\begin{align*}
\text{minimize} & \quad f(h) = \frac{1}{2}||y - T_u h||^2_2 \\
\text{subject to} & \quad ||h||_A \leq \tau
\end{align*}
\]

(2.23)

and efficiently solved using a Franke-Wolfe type algorithm [32].

2.5 Superstability and Equalized Performance

Definition 7. ([33]) A Linear Time Invariant (LTI) discrete-time system represented by the following ARMA model:

\[
e_t = -\sum_{k=1}^{n} a_k e_{t-k} + \sum_{k=0}^{n} b_k w_{t-k}
\]

(2.24)

where \( e_t \) and \( w_t \) denote the output and exogenous disturbance, respectively is said to have (finite) equalized \( N \)-performance less than \( \mu \) iff for \(|e_t| \leq \mu, t = 0, 1, \cdots, N - 1, \) and all sequences \( w_t, \)

18
CHAPTER 2. PRELIMINARIES

\[ t = 0, 1, \cdots, |w_t| \leq 1 \text{ compatible with } e_t, t = 0, 1, \cdots, N - 1, \text{ we have that} \]

\[ |e_t| \leq \mu, \quad \forall t \geq N. \quad (2.25) \]

Roughly speaking the definition above states that a plant achieves an \( N \)-equalized performance level \( \mu \) if, whenever the magnitude of the output is below \( \mu \) for \( N \) consecutive time instants, then this situation continues into the future, provided that the \( \ell^\infty \) norm of the disturbance input remains bounded by 1.

**Definition 8.** (\cite{34}) System (2.24) is said to be superstable, if its characteristic polynomial \( D(\lambda) \) is super stable, i.e. \( \sum_1^n |a_k| < 1 \).

As shown in \cite{35}, every stable plant has finite equalized \( N \)-performance, for some sufficiently large \( N \). However, only superstable systems achieve finite \( n \)-equalized performance, where \( n \) is the McMillan degree of the plant. This fact follows as a corollary of the following result:

**Theorem 3.** (\cite{35}) Let \( \mu \geq 0 \). The system (2.24) with McMillan degree \( n \) has equalized \( n \)-performance less than \( \mu \) if and only if the following condition holds

\[ \mu ||a||_1 + ||b||_1 \leq \mu. \quad (2.26) \]

Therefore the equalized \( n \)-performance level \( \mu^n \) of the system is given by

\[ \mu^n = \frac{||b||_1}{1 - ||a||_1}, \quad (2.27) \]

where \( a = \begin{bmatrix} a_1 \ldots a_n \end{bmatrix}^T \in \mathbb{R}^n \) and \( b = \begin{bmatrix} b_0 \ldots b_n \end{bmatrix}^T \in \mathbb{R}^{n+1} \).

**Proof.** If \( ||b||_1 = 0 \), then (2.24) reduces to \( e_t = -\sum_{k=1}^n a_k e_{t-k} \), and (2.26) reduces to \( \mu ||a||_1 \leq \mu \).

Then for each \( t \), if \( |e_{t-k}| \leq \mu \) for \( \forall k = n, n - 1, \ldots, 1 \), we have

\[ |e_t| = \left| -\sum_{k=1}^n a_k e_{t-k} \right| \leq \sum_{k=1}^n |a_k| |e_{t-k}| \leq \mu \sum_{k=1}^n |a_k| = \mu ||a||_1 \leq \mu, \quad (2.28) \]

meaning the system (2.24) has the equalized \( n \)-performance \( \mu \) and \( ||a|| \leq 1 \). Further, when \( ||b||_1 = 0 \), if \( ||a||_1 < 1 \), \( \mu^n \) defined in (2.27) is also 0. Since 0 is a feasible solution to (2.26), the superstable system has equalized \( n \)-performance level 0.
CHAPTER 2. PRELIMINARIES

If $\|b\|_1 \neq 0$, then from (2.26), we have $0 < \|b\|_1 \leq \mu (1 - \|a\|_1)$, combining the fact that $\mu$ is positive, it can be obtained that $\|a\|_1 < 1$ and $\mu^n = \frac{\|b\|_1}{1 - \|a\|_1} > 0$. From (2.26), for each $t$, if $|e_{t-k}| \leq \mu$ for $\forall k = n, n-1, \ldots, 1$, we have that for any disturbance sequence $w_{t-k}$, if $\|w_{t-k}\| \leq 1$ for each $k = 0, 1, \ldots, n$, the following holds:

$$
|e_t| = |- \sum_{k=1}^{n} a_k e_{t-k} + \sum_{k=0}^{n} b_k w_{t-k}| \\
\leq \sum_{k=1}^{n} |a_k| |e_{t-k}| \\
\leq \mu \sum_{k=1}^{n} |a_k| + \sum_{k=0}^{n} |b_k| |w_{t-k}| = \mu \|a\|_1 + \|b\|_1 \leq \mu,
$$

meaning the system (2.24) is superstable and has equalized $n$-performance $\mu$. Since $\mu^n$ is a feasible solution to $\mu$ in (2.26), thus, the equalized $n$-performance level is $\mu^n$. 

2.6 Robust Optimization

The goal of robust optimization is to guarantee that, in problems subject to uncertainty, a set of “hard” constraints is satisfied for all possible values of the uncertain parameters. Notably, as shown for instance in [36], several optimization problems and uncertainty region characterizations allow for reformulating the Robust Counterpart (RC) of the original nonlinear problem as a tractable optimization problem. Consider a nonlinear constraint

$$
f(a, x) \leq 0 \quad (2.30)
$$

where $x \in \mathbb{R}^n$ is the optimization variable, $f(\cdot, x)$ is concave, in its first argument, for all $x \in \mathbb{R}^n$, and $a \in \mathbb{R}^m$ is an uncertain vector, which is only known to reside in a set $U$. The RC of (2.30) is

$$
f(a, x) \leq 0, \quad \forall a \in U. \quad (2.31)
$$

The following theorem, combining results in [36] and [37], serves as a theoretical foundation for the subsequent derivation in Chapter 10.

**Theorem 4.** Assume that

(i) the nonlinear function $f$ is of the form

$$
f(a, x) = (h_x + Qa)^T x + h_a^T a + r; \quad (2.32)
$$
(ii) the uncertain set $\mathcal{U}$ is a compact, strictly feasible polyhedron of the form

$$\mathcal{U} \doteq \{ a \in \mathbb{R}^m : Da + q \geq 0, D \in \mathbb{R}^{d \times m}, q \in \mathbb{R}^d \}. \quad (2.33)$$

where $h_x, Q, h_a, r, D,$ and $q$ are deterministic constants of appropriate dimensions.

Then the vector $x \in \mathbb{R}^n$ satisfies (2.31) if and only if there exist $z \in \mathbb{R}^d$ satisfying the set of (in)equalities:

$$\begin{cases} q^T z + h^T_x x + r \leq 0 \\ -D^T z = h_a + Q^T x \\ z \geq 0 \end{cases} \quad (2.34)$$

Proof. RC (2.31) is equivalent to the worst-case reformulation

$$F(x) \leq 0, \quad F(x) \doteq \max_{a \in \mathcal{U}} f(a, x). \quad (2.35)$$

Using the indicator function $\delta(a|\mathcal{U})$ defined as

$$\delta(a|\mathcal{U}) = \begin{cases} 0 & \text{if } a \in \mathcal{U} \\ \infty & \text{otherwise,} \end{cases} \quad (2.36)$$

and Fenchel duality (38), we have

$$F(x) = \max_{a \in \mathbb{R}^m} \{ f(a, x) - \delta(a|\mathcal{U}) \} \leq F_d(x) \quad (2.37)$$

where $F_d(x)$ is the optimal dual function

$$F_d(x) \doteq \min_{v \in \mathbb{R}^m} \{ \delta^*(v|\mathcal{U}) - f_*(v, x) \}, \quad (2.38)$$

$\delta^*$ is convex conjugate function of $\delta$ in (2.36), and $f_*$ is the partial concave conjugate function of $f(a, x)$ with respect to $a$. Specifically,

$$f_*(v, x) = \min_{a \in \mathbb{R}^m} v^T a - f(a, x) = -h_x^T x - r, \text{ subject to } v = h_a + Q^T x \quad (2.39)$$
and due to assumption (ii), strong duality holds for

\[
\delta^*(v|U) = \max_{a \in U} v^T a \\
= \min_{z \geq 0} z^T q, \text{ subject to } v = -D^T z
\] (2.40)

therefore, combining (2.38)-(2.40), we have

\[
F_d(x) = \min_z z^T q + h^T x + r \\
\text{subject to } -D^T z = h_n + Q^T x, \ z \geq 0
\] (2.41)

Obviously if there exist a feasible solution \(z_0 \in \mathbb{R}^d\) to (2.34), then we have

\[
F(x) \leq F_d(x) \leq z_0^T q + h^T x + r \leq 0.
\]

\[\square\]
Chapter 3

Robust Linear Model Fitting

3.1 Motivation

Robust linear model fitting has gained ever-increasing research interest due to its extensive applications in the communities of signal processing, image processing, and computer vision, for instance, compressive sensing [39], planar surface fitting and fundamental matrix learning for stereo camera calibration and 3D reconstruction [40].

The difficulties in linear model fitting lie in two aspects. The first one comes from the noise contaminated data. In most of the existing results are established upon the assumptions that the noise is stochastic, and that the data is sufficient so that the estimated model is unbiased. As we know if the noise is independently normally distributed, then Least Squares (LS) can provide good model estimation [41]. However, the least squares estimator is quite sensitive to outliers. To resist the effects of outliers, the sum of squares is replaced by a function of the fitting error which performs locally like LS for small errors, and tapering off to a constant for large errors, resulting M-estimator. Although M-estimator performs well for cases where outliers amount to a small portion of the data, its performance substantially degrades as the outliers account for larger percentages. Thus, a large number of robust methods have been developed to explicitly take into account the presence of outliers while estimating the model parameters. Among them include randomized methods such as RANdom Sampling Consensus estimator (RANSAC) [6] and its variants, which attempt to find outlier-free data by repeatedly randomly selecting the minimal number of samples needed to generate a model, and selecting the best one, according to some optimality criteria. For instance, RANSAC selects the solution with the largest percentage of inliers in the complete dataset. MSAC [42], a randomized M-estimator [43], penalizes both the squared fitting error of inliers and number of outliers. Least
Median of Squares (LMS) selects the estimate which gives the least median fitting error. Finally, Maximum Likelihood Estimation SAmple Consensus (MLESAC) attempts to find the maximum likelihood estimate of the noise-free data. Random sampling based methods are attractive due to their simplicity and the existence of theoretical bounds on the number of samples required to guarantee a given probability of success. However, they suffer from several weaknesses: Firstly, for a given probability of success, the number of iterations needed grows very fast with the number of outliers. Secondly, since the bounds explicitly depend on the number of outliers, this quantity must be known or estimated accurately, since stopping the algorithm prematurely can lead to arbitrarily bad solutions.

To avoid the weaknesses induced by the combinatorial nature of randomized methods, motivated by compressive sensing theory, a number of sparse regularization based approaches have been developed (see for instance and references therein), where the noise variables is expressed as sum of two independent terms, with one representing the sparse outlier noise and the other representing the dense inlier noise, and the \( \ell_0 \) norm (cardinality) of outlier noise is minimized to find the model fitting as many inliers as possible. \( \ell_1 \) norm is minimized instead due to the nonconvexity of \( \ell_0 \) norm, leading to computationally efficient algorithms. It is pointed out in that the performance of the approach depends on the smallest principal angle between the regressor matrix and the outlier subspaces. However, all these theoretical results are based on the assumption that the outliers are sparse compared to inliers, thus their performance deteriorates with the percentage of outliers increases.

The second difficulty of linear model fitting comes from its wide application which may require the flexibility of the formulation to incorporate some prior information on the specific model, for instance, in estimation of fundamental matrix in computer vision, as we know the fundamental matrix is required to be rank deficient. However, the approaches mentioned above cannot handle this information by simply adding or removing constraints, instead they just approximate a rank deficient solution posteriorly. However, in it has been shown that ignoring the rank constraint degrades the accuracy in terms of the covariance matrix of the first-order variation of the solution. Thus, it motivates researchers to exploit formulations considering these extra constraints directly in order to improve accuracy.

Motivated by these challenges mentioned above, in this dissertation we propose a novel single-step framework for robust linear model fitting problem in the form of a constrained polynomial optimization problem, which can be solved by appealing to recent results on sparse polynomial optimization. The advantages of the proposed approach vis-à-vis existing techniques includes the abilities to: (1) explicitly impose extra constraints for properties of specific models, for example,
the rank deficient constraint; (2) handle noise and a very large percentage of outliers, without the need for additional assumptions such as bounds on the number of outliers; (3) certificate certify that a given convex relaxation has indeed found an optimal estimate of the model (the true model for noiseless data and the one that maximizes the number of inliers in the case of noisy measurements).

In addition, we provide theoretical results showing that, if in the optimization above a certain matrix containing only variables related to the fundamental matrix has rank 1, then the first order relaxation of the problem achieves global optimality. Combining these ideas with rank-minimization techniques leads to a computationally efficient algorithm with complexity comparable to robust regression based techniques, while still retaining the advantages noted above. These results are illustrated with several examples where the proposed algorithm is shown to consistently outperform existing approaches.

3.2 Problem Statement

Under the homogenous coordinates in \( \mathbb{R}^n \), consider the linear regression model of the form

\[
x^T r = 0, \quad r^T r = 1,
\]

(3.1)

where \( r \in \mathbb{R}^n \) is the model parameter. If a given sample \( x_j \in \mathbb{R}^n \) can be fitted by the model (3.1), then \( x_j^T r = 0 \) holds, otherwise, \( |x_j^T r| \) measures the distance from \( x_j \) to the hyperplane defined by the model (3.1).

As shown in Figure 3.1, given a set of samples \( \{x_j\}_{j=1}^{N_p} \), drawn from the model (3.1), then the difficulty in estimating the model parameters depends on the way the noise corrupts the data.

- In the case where the data is clean, from

\[
\begin{align*}
x_j^T r &= 0, \\
r^T r &= 1
\end{align*}
\]

(3.2)

the model parameter \( r \) can be determined by the null space of the regression matrix \( X = \begin{bmatrix} x_1 & x_2 & \ldots & x_N \end{bmatrix}^T \).

- In the case where the data is corrupted such that \( x_j \) deviates from the hyperplane no more than...
CHAPTER 3. ROBUST LINEAR MODEL FITTING

$\epsilon$, the linear model fitting can be formulated as finding a feasible solution to

$$\begin{cases} |x_j^T r| \leq \epsilon, \forall_{j=1}^{N_p} \\ r^T r = 1 \end{cases} \tag{3.3}$$

which is nonconvex due to the normalization condition for $r$. However, the nonconvexity can be circumvented by solving a set of convex problems. Let $\bar{r}_i$ denote a vector in $\mathbb{R}^n$ and its $i$-th entry is fixed as 1, then (3.3) is feasible is equivalent to

$$\forall_{i=1}^{n} : \{ \bar{r}_i^*, \epsilon_i^* \} = \arg \minimize_{\bar{r}_i, \epsilon_i} \epsilon_i \tag{3.4}$$

subject to

$$|x_j^T \bar{r}_i| \leq \epsilon_i, \forall_{j=1}^{N_p}$$

$$\bar{r}_i(i) = 1$$

and

$$\min \left\{ \frac{\epsilon_i^*}{||\bar{r}_i^*||_2}, \forall i = 1, \ldots, n \right\} \leq \epsilon \text{ is feasible.} \tag{3.5}$$

- In the case where the data corrupted such that some of them deviates from the hyperplane too far away, thus being considered as outliers, if the outliers are mistakenly considered as inliers, then the estimated model (shown as the dash line) would skew from the true model (shown as the dot line) dramatically. In this chapter, we will concentrate on this case.

![Figure 3.1: 2D Linear Regression Problem. (Left: clean data; Center: noisy data without outliers; Right: noisy data with outliers)\\x_j^T r = 0 \quad |x_j^T r| \leq \epsilon \quad \text{inlier:} |x_j^T r| \leq \epsilon \quad \text{outlier:} |x_j^T r| > \epsilon](image)

The problem considered in this chapter is formally stated as:

**Problem 1.** Given
CHAPTER 3. ROBUST LINEAR MODEL FITTING

- a set of noisy samples \( X = \{ x_j \in \mathbb{R}^n : x_j = \hat{x}_j + \eta_j, \forall j = 1, \ldots, N_p \} \), drawn from the hyperplane defined by the linear model \( S = \{ \hat{x} \in \mathbb{R}^n : \hat{x}^T r = 0 \} \), where \( r \in \mathbb{R}^n \), and \( r^T r = 1 \).

- prior information on the maximum distance from the inliers to the hyperplane, \( \epsilon \), that is, noisy samples with \( |x_j^T r| > \epsilon \) are considered to be outliers.

Find the parameter \( r \) such that

- \( r^T r = 1 \) holds;

- the number of inliers is maximized, that is, maximizing \( I(r) \doteq \) the number of samples satisfying \( |x_j^T r| \leq \epsilon \).

3.3 A Convex Approach to Robust Linear Model Fitting

In this section we present the main theoretical results of this chapter: (i) a reformulation of Problem 1 into a constrained polynomial optimization problem, (ii) a convergent sequence of convex relaxation, and (iii) a sufficient condition, given in terms of the rank of a small matrix, for the first element of this sequence to attain global optimality, leading to a computationally efficient algorithm. In this section, we will first consider the basic case described in Problem 1, and in Section 3.4 we will show the advantage our formulation in incorporating more specific constraint on the model parameter.

3.3.1 A Constrained POP Reformulation

By introducing indicator variables \( s_j \in \{0, 1\} \), Problem 1 can be reformulated as

\[
p^* = \min_{r, s_j} \sum_{i=1}^{N_p} (1 - s_j)
\]

subject to

\[
||r||_2^2 = 1
\]

\[
\forall j = 1, \ldots, N_p : \quad s_j = 1
\]

\[
\forall j = 1, \ldots, N_p : \quad s_j |x_j^T r| \leq s_j \epsilon
\]
In this formulation (3.6c) forces \( s_i \) to be binary. The last inequality, (3.6d), combined with (3.6c) enforces that \( s_i = 0 \) for outliers (that is points where \( |x_j^T r| > \epsilon \)). Thus, the cost function (3.6a) is meant to minimize the number of outliers.

3.3.2 A Sequence of Convex Relaxations

Clearly, (3.6) is nonconvex due to the bilinear terms and the binary variables in (3.6b)-(3.6d). However, it is a constrained POP problem, for which the moments based polynomial optimization techniques introduced in Chapter 2 can be used to obtain a sequence of convex relaxations of the form

\[
p^*_N = \min_{m_{2N}} N_p - \sum_{j=1}^{N_p} m(s_j)
\]

subject to

\[
M_N(m_{2N}) \succeq 0 \\
L_{N-1}(g_k m_{2N}) \succeq 0, \forall k = 1
\]

where \( m_{2N} \) denotes the moment sequence of the variables \( v = \{r, s_j, j = 1, \ldots, N_p\} \) of order up to \( 2N \), \( M_N \) and \( L_{N-1}(g_k m_{2N}) \) denote the \( N \)-th order moment matrix and \((N - 1)\)-th order localizing matrix associated with the constraint \( g_k \) that defines the feasibility set in (3.6b)-(3.6d), respectively. \( d \) represents the number of constraints in (3.6b)-(3.6d).

Considering the partitioning of the objective and the constraints in (3.6) of the form

\[
p = \sum_{j=0}^{N_p} p_j \quad \text{with} \quad p_0 = 0, \forall j = 1 : p_j = 1 - s_j
\]

\[K = \cap_{j=0}^{N_p} K_j \quad \text{with} \quad K_0 : \|r\|_2^2 = 1, \forall j = 1 : K_j : \begin{cases} s_j^2 = s_j \\ s_j |x_j^T r| \leq s_j \epsilon \end{cases}
\]

with each partition \( \{p_j, K_j\} \) associated with a sample, then it is easy to check that each partition contains variables in a subset \( v_j \), where \( v_0 = \{r\} \), and \( \forall j = 1 : v_j = \{r, s_j\} \). \( v_j \)'s satisfy the conditions \( \bigcup_{j=0}^{N_p} v_j = v \) and

\[
v_{j+1} \cap \bigcup_{l=0}^{j} v_l = \{r\} = v_0, \text{ for } \forall j = 0, \ldots, N_p - 1,
\]

meaning the problem exhibits the running intersection property and hence it can be simplified to
CHAPTER 3. ROBUST LINEAR MODEL FITTING

solving the sequence of reduced SDP problems of the form

\[ p_{m,N}^* = \min_{m_{j,2N}} \sum_{j=0}^{N_p} c_j^T m_{j,2N} \]

subject to

\[ \forall_{N_p}^{N_p} : M_{j,N}(m_{j,2N}) \succeq 0 \]

\[ \forall_{N_p}^{N_p} : L_{j,N-1}(g_{jk} m_{j,2N}) \succeq 0, \forall g_{jk} \in K_j \]

(3.9)

where \( m_{j,2N} \) and \( M_{j,2N} \) denote the moments sequence of variables in \( I_j \) of order up to \( 2N \) and the associated moment matrix of order \( N \), \( c_j^T m_{j,2N} \) is the expectation of \( p_j \), and \( L_{j,N-1} \) is the localizing matrix related to constraints in \( K_j \). Instead of considering a single moment matrix, \( M_N \), of size

\[
\begin{pmatrix}
 n + N_p + N \\
 N
\end{pmatrix}
\]

in (3.7), in (3.9) we consider \( N_p + 1 \) smaller moment matrices, \( M_{j,N} \), of size

\[
\begin{pmatrix}
 n + 1 + N \\
 N
\end{pmatrix}
\]

substantially reducing the dimension of moments sequence from \( \mathcal{O}((n + N_p)^2 N) \) to \( \mathcal{O}((n + 1)^2 N) \) in the case the number of samples \( N_p \) is much larger than the dimension of the data \( n \).

The results from Section 2.2.3 guarantee that this sequence of relaxation converges monotonically (from below) to the optimum. Nevertheless, from a practical standpoint its applicability is limited to relatively small problems, due to the fact that the number of variables increases combinatorially with \( N \) (even when exploiting the underlying sparse structure), and that, in principle the value of \( N \) required to achieve a flat extension, and hence certify optimality, can be large. In addition, once an optimal \( N \) has been found, extracting the solution \( r \) from the corresponding moment matrix \( M_{0,N} \) is far from trivial, unless \( \text{rank}(M_{0,N}) \) is low (see [49] for details). To circumvent these difficulties, in the sequel we will exploit the fact that moment matrices associated with atomic measures having a single atom have rank 1, since in this case the moments simply correspond to the powers of the variables, evaluated at the location of the atom, and in turn the variables are equal to their first order moments. In principle, one could try to exploit this observation by adding a low rank constraint on each of the moment matrices \( M_{i,N} \) to (3.7). However, this constraint would be computationally hard to enforce due to the large number of matrices involved. Surprisingly, as shown by the result below, enforcing a low rank constraint only on the moment matrix associated with \( r \) is enough to guarantee that the relaxation corresponding to \( N = 1 \) attains global optimality. Further, in this case the elements of \( r \) can be read directly from its associated moment matrix, without the need for additional computations.
CHAPTER 3. ROBUST LINEAR MODEL FITTING

Theorem 5. The first order reduced sized moment relaxation (3.9) of Problem (3.6) given by

\[
\tilde{p}_1^* = \min_{m_{j,2}} \quad N_p - \sum_{j=1}^{N_p} m(s_j) \\
\text{subject to} \\
\forall_{N_p}^{j=0} : \quad M_{j,1}(m_{j,2}) \succeq 0 \\
\forall_{N_p}^{j=0} : \quad L_{j,0}(g_{j,k}m_{j,2}) \succeq 0, \forall g_{j,k} \in K_j
\]

is exact if the moment matrix \( M_{0,1} \) associated with \( r \) has rank 1. Further, in this case the corresponding optimal first order moments sequence \( \{m(r)^*, m(s_j)^*\} \) is also an optimal solution to the original nonconvex problem (3.6).

Proof. Let

\[
M_{0,1} = m \begin{pmatrix} 1 \\ r \end{pmatrix} \begin{pmatrix} 1 & r^T \\ r & 1 \end{pmatrix} = \begin{pmatrix} 1 & m(r)^T \\ m(r) & m(rr^T) \end{pmatrix}
\]

and

\[
\forall_{N_p}^{j=0} : M_{j,1} = m \begin{pmatrix} 1 \\ r \end{pmatrix} \begin{pmatrix} 1 & r^T \\ r & 1 \end{pmatrix} = \begin{pmatrix} 1 & m(r)^T & m(s_j) \\ m(r) & m(rr^T) & m(r) \end{pmatrix}
\]

\[
= \begin{pmatrix} 1 & m(r)^T & m(s_j) \\ m(r) & m(rr^T) & m(r) \\ m(s_j) & m(r) & m(s_j) \end{pmatrix} = \begin{pmatrix} M_{0,1} & m(s_j) \\ m(s_j) & M_{0,1} \end{pmatrix}
\]

where the last entry of \( M_{j,1} \) results from the localizing matrix corresponding to \( s_j^2 = s_j \). From rank\(\{M_{0,1}\} = 1\), it follows that the first order moment variables \( m(r) \) satisfy the constraint (3.6b), i.e., \( ||m(r)||_2^2 = \text{trace}\{m(r)r^T\} = 1 \).

Let \( r(i) \) denote the \( i \)-th entry of the vector \( r \), for all \( i = 1, \ldots, n \). Since \( M_{j,1} \) is positive semi-definite, then all its principal minors should be positive semi-definite, thus, we have

\[
M_{ij,1} = m \begin{pmatrix} 1 \\ r(i) \\ s_j \end{pmatrix} \begin{pmatrix} 1 & r(i) & s_j \end{pmatrix} = \begin{pmatrix} 1 & m(r(i)) & m(s_j) \\ m(r(i)) & m(r(i)^2) & m(r(i)s_j) \\ m(s_j) & m(r(i)s_j) & m(s_j) \end{pmatrix} \succeq 0
\]

(3.11)
CHAPTER 3. ROBUST LINEAR MODEL FITTING

Following that \( M_{0,1} \) is rank 1, the equality \( m(r(i)^2) = m(r(i))^2 \) holds, and from

\[
\det(M_{ij,1}) = m(s_j)[m(r(i)m(r(i)s_j) - m(s_j)m(r(i)^2)] - m(r(i)s_j)[m(r(i)s_j) - m(s_j)m(r(i))] \\
= - [m(s_j)m(r(i)) - m(r(i)s_j)]^2 \\
\geq 0
\]

we have

\[ m(r(i)s_j) = m(s_j)m(r(i)). \] (3.12)

At the first order relaxation, the localizing matrix associated with (3.6d) reduces to linear inequality constraints

\[
x_j^T m(s_j r) \leq m(s_j) \epsilon \quad \text{and} \quad x_j^T m(s_j r)x_j \geq -m(s_j) \epsilon.
\] (3.13)

Substituting (3.12) into (3.13) leads to

\[
x_j^T m(s_j)m(r) \leq m(s_j) \epsilon \quad \text{and} \quad x_j^T m(s_j)m(r) \geq -m(s_j) \epsilon.
\] (3.14)

Thus, for any \( m(s_j) > 0 \), (3.13) reduces to

\[
x_j^T m(r) \leq \epsilon \quad \text{and} \quad x_j^T m(r) \geq -\epsilon,
\] (3.15)

implying that the sample \( x_j \) is an inlier to the hyperplane determined by the parameter \( m(r) \).

Next, note that, since (3.10) seeks to minimize \( -\sum_{j=1}^{N_p} m(s_j) \) term in the optimal solution will automatically increase to 1. Therefore, (3.10) is exactly equivalent to (3.6) (by replacing \( r, s_j \) with \( m(r), m(s_j) \)). Finally, note that the combination of the binary variables \( m(s_j)^* \) and \( \text{rank}\{M_{0,1}\} = 1 \) guarantees that \( M_{j,1} \) is also rank 1 for all \( j = 1, \ldots, N_p \), from where it follows that the first order moment sequence \( \{m(r)^*, m(s_j)^*\} \) is indeed an optimal solution of the original nonconvex problem (3.6).

3.3.3 Two Iterative Algorithms for Rank-1 Solution

Theorem \( \PageIndex{5} \) suggests that a computationally attractive algorithm can be obtained by simply adding the constraint \( \text{rank}\{M_{0,1}\} = 1 \) to (3.10). Unfortunately, the resulting problem is no longer convex. Nevertheless, a tractable convex relaxation can be obtained by using a (weighted) nuclear norm as a surrogate for rank\(^1\) and iteratively solving a sequence of regularized convex problems \( \cite{50} \).

\(^1\)This relaxation amounts to a local linearization of the matrix’s log-det function.
CHAPTER 3. ROBUST LINEAR MODEL FITTING

leading to Algorithm 1 outlined below. Note that each iteration has computational complexity roughly equivalent to that of regularized robust regression and robust low rank factorization based methods. Further, consistent numerical experiments show that typically only a few iterations are needed for convergence.

Algorithm 1 Robust Linear Model Fitting by Minimizing Reweighted Nuclear Norm

1: Initialize: \( k = 0 \), \( W^{(1)} = I \);
2: repeat
3: \( k = k + 1 \);
4: Solving
\[
M_{0,1}^{* (k)} = \arg \min_{m_{j,2}} \left\{ \lambda \text{trace}(W^{(k)}M_{0,1}) \right\}
\]
subject to
\[
\begin{align*}
\forall N_p^{j} = 0 & : M_{j,1}(m_{j,2}) \succeq 0 \\
\forall N_p^{j} = 0 & : L_{j,0}(g_{jk} m_{j,2}) \succeq 0, \forall g_{jk} \in K_j
\end{align*}
\]
5: Updating
\[
W^{(k+1)} = [M_{0,1}^{* (k)} + \sigma_2(M_{0,1}^{* (k)})]^{-1}
\]
6: until \( \sigma_2(M_{0,1}^{* (k)}) < 10^{-6} \sigma_1(M_{0,1}^{* (k)}) \).

An alternative approach to obtaining the rank-1 solution is based on the fact that the localizing matrix corresponding to \( \|r\|^2 = 1 \) is equivalent to

\[
\text{trace}\{M_{0,1}\} = \sum_{i=1}^{n+1} \sigma_i(M_{0,1}) = 2,
\]
combining \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{n+1} \geq 0 \), we have

\[
\|M_{0,1}\|_F = \sqrt{\sum_{i=1}^{n+1} \sigma_i^2} \leq \sqrt{\left(\sum_{i=1}^{n+1} \sigma_i \right)^2} = 2,
\]

where the equality holds if and only if \( \sigma_1 = 2 \) and \( \sigma_i = 0 \) for \( i = 2, \ldots, n + 1 \), meaning \( M_{0,1} \) is exactly rank 1. Thus, we could achieve rank-1 solution to (3.10) by adding an extra regularization...
term to the objective of (3.10), ending up with the problem

\[
\hat{p}_1^* = \min_{m_{j,2}} \left[ N_p - \sum_{j=1}^{N_p} m(s_j) \right] - \lambda \| M_{0,1} \|_F
\]

subject to

\[
\forall_{j=0}^{N_p} : M_{j,1}(m_{j,2}) \geq 0
\]
\[
\forall_{j=0}^{N_p} : L_{j,0}(g_{jk} m_{j,2}) \geq 0, \forall g_{jk} \in K_j
\]

which enforces the low rank constraint by maximizing the Frobenius norm of \( M_{0,1} \). However, to minimize (3.20) is nonconvex. As noted that the objective of (3.20) is a difference of two convex differentiable function, then we can resort to the Convex-Concave Procedure \[51\], we can obtain a local optimal solution to (3.20) by the iterative algorithm summarized in Algorithm 2, where \( W^{(k+1)} \) is the gradient of \( \| M_{0,1} \|_F \) at \( M_{0,1}^{(k)} \).

**Algorithm 2** Robust Linear Model Fitting by Maximizing Frobenius Norm

1: Initialize: \( k = 0, W^{(1)} = 0 \);
2: repeat
3: \( k = k + 1 \);
4: Solving

\[
M_{0,1}^{(k)} = \arg \min_{M_{0,1}} \left[ N_p - \sum_{j=1}^{N_p} m(s_j) \right] - \lambda \text{trace}\{ W^{(k)} M_{0,1} \}
\]

subject to Constraints in (3.20)

5: Updating

\[
W^{(k+1)} = \frac{M_{0,1}^{(k)}}{\| M_{0,1}^{(k)} \|_F}
\]

6: until \( \sigma_2(M_{0,1}^{(k)}) < 10^{-6} \sigma_1(M_{0,1}^{(k)}) \).

### 3.4 Extensions: Handling Additional A-Priori Information

A salient feature of the proposed framework is that it can handle many constraints that can be described as polynomial (in)equalities, which is important especially for cases where the data is insufficient and extra constraints are necessary to shrink the feasible set of the model. Next, we will take the fundamental matrix learning as an example to show the advantage of the proposed approach in handling additional a-priori information.
CHAPTER 3. ROBUST LINEAR MODEL FITTING

3.4.1 Rank Deficient Constraint

Given a pair of images of the same scene from two uncalibrated perspective views, the fundamental matrix \( F \in \mathbb{R}^{3 \times 3} \) is defined as the rank-2 matrix which satisfies the epipolar constraint

\[
x'^T F x = 0, \quad \forall x', x,
\]

(3.23)

where the homogenous coordinates \( x, x' \in \mathbb{R}^3 \) are the projections of the same 3D point in the two images. \( F \) has seven degrees of freedom due to the ambiguity caused by the scaling and singularity.

Among the existing references, only several of them considers the rank deficient constraint explicitly. For instance, In [52] an Levenberg-Marquard (LM) approach is proposed to optimize the singular value decomposition (SVD) of the fundamental matrix. In [53] and [54] the rank constraint is imposed by setting its determinant to 0, leading to a 3rd-order polynomial constraint. Alternatively, in [55] and [56] the estimation problem is reduced to one or several constrained polynomial optimization problems by imposing the constraint that the null space of the solution must contain a nonzero vector. The resulting optimization problems are solved by resorting to various optimization techniques, such as brand-and-bound ([53], [56]) approach or moments based convex relaxations ([55], [54]). Although these methods perform well under the assumption of small, suitably distributed (for instance Gaussian) measurement noise, their performance substantially degrade in the presence of even a few outliers (i.e. point mismatches).

Our goal here is to develop a computationally tractable algorithm that, starting from noisy point correspondences corrupted by outliers, simultaneously estimates a rank-2 fundamental matrix that maximizes the number of inliers, and, at the same time, explicitly identifies outliers, defined as those points whose distance from the surface defined by (3.23) is beyond a given bound. Specifically, we address the following problem:

**Problem 2.** Given a set of noisy point correspondences, \( \{x_i, x'_i\}, \ i = 1, \ldots, n \) drawn from two images of the same scenes, and a-priori bound on the fitting error \( |x_i'^T F x_i| \leq \epsilon \), determine a fundamental matrix \( F \) such that:

- \( \|F\|_F = 1 \);
- \( F \) is rank-2;
- The number of inliers is maximized, that is maximizing \( |\mathcal{I}(F)| \) where \( \mathcal{I}(F) \) denotes the number of samples satisfying \( |x_i'^T F x_i| \leq \epsilon \).
CHAPTER 3. ROBUST LINEAR MODEL FITTING

As we show in the sequel, the problem above can be recast as a linear model fitting problem and solved using the method presented in Section 3.3.

Let \( f \in \mathbb{R}^9 \) denote the vectorized \( F \), that is, \( f = \text{vec}(F) \) and \( F = \text{mat}(f) \). Then \( f^T f = \|F\|_F^2 \) and \( f^T (x \otimes x') = x'^TFx \). Therefore, comparing to Problem 1, the extra constraint in Problem 2 is the rank-2 constraint, which can be described as there exists a nonzero vector \( q \in \mathbb{R}^3 \) such that \( Fq = 0 \). Thus, Problem 2 can be formulated mathematically as

\[
p^* = \min_{q, F, s_j} \sum_{j=1}^{N_p} (1 - s_j)
\]

subject to

\[
\|F\|_F^2 = 1 \quad (3.24b)
\]

\[
s_j^2 = s_j, \forall j = 1 \quad (3.24c)
\]

\[
s_j |x'^T_i Fx_j| \leq \epsilon s_j, \forall j = 1 \quad (3.24d)
\]

\[
Fq = 0, q^T q = 1 \quad (3.24e)
\]

where (3.24e) is added to enforce the rank-2 constraint on \( F \). Compared to the formulation in (3.6), in (3.24) the extra constraint (3.24e) also consists of second order polynomials. Thus, all the analyses in Section 3.3.2 to Section 3.3.3 are applicable to solving (3.24) by just replacing \( r \) by \( [f^T q^T]^T \).

In Section 3.5, we will show the advantage of our approach by testing on real images.

3.5 Experiments

In this section, we describe a set of experiments used to verify our theoretical results. These experiments consist of estimating fundamental matrix from points correspondences from six pairs of real images: House, Merton I, Merton II, Merton III, Library, and Wadham, given by VGG, University of Oxford, as shown in Figure 3.2.

3.5.1 Experimental Protocols

In each experiment, before computing the fundamental matrix (no matter what method is used), the data are normalized in the way pointed out by [40] and [57].

Ground truth data. Given a pair of images, we first calculated a fundamental matrix from the correspondences \( (x_{i,t}, x'_{i,t}) \) provided with the dataset by minimizing the algebraic error. We considered the resulting \( F_t \) as the ground truth, with the threshold \( \epsilon \) given by \( \epsilon = \max_i |x'^T_i F_t x_{i,t}| \).
Then VLFeat toolbox \[58\] was used to obtain the SIFT features from the two images, and correspondences \((x_i, x'_i)\) were defined by those pairs of points whose \(\ell_2\) norm is not larger than 0.5. Finally, inliers and outliers were selected as those correspondences satisfying

\[ |x'_i^T F x_i| \leq \epsilon, \]

and those violating this beyond, respectively.

**Experimental set-up.** For each pair of images, we ran seven sets of experiments, with 100 correspondences and outliers ranging from 10% to 70%. For each set of experiment we ran 50 times by randomly picking \(N_{out}\) outliers and \(N_{in}\) inliers from \((x_i, x'_i)\) and compare with several state-of-the-art methods.

**Evaluation criteria.** We compared four performance indices as follows:

- **Precision** = \(\frac{\text{Ground truth inliers} \cap \text{identified inliers}}{\text{identified inliers}} \times 100(\%)\);

- **Recall** = \(\frac{\text{Ground truth inliers} \cap \text{identified inliers}}{\text{identified inliers}} \times 100(\%)\);

- **Hmeans** = \(\sqrt{\text{Precision} \times \text{Recall}} \times 100(\%)\);
CHAPTER 3. ROBUST LINEAR MODEL FITTING

- Similarity = |trace(F_t F^T)|. The closer Similarity to 1, the smaller the distance between the identified fundamental matrix F and the ground truth F_t.

Computational platform. All the competing algorithms are implemented in MATLAB and run on iMac with 3.4GHz CPU and 32G memory. The SDP solver used was SeduMi [59].

3.5.2 Results

In the experiments, we compared the performance of the proposed approach with that of RANSAC and its several extended versions, i.e., MSAC, MLESAC and LMEDS. For all these methods, the number of iterations was set to 500, and in each iteration the fundamental matrix was calculated using the eight-point algorithm. The results are summarized in Tables 3.1, 3.4 and illustrated in Figure 3.3. As shown there, the propose algorithm was consistently more robust than the SAC algorithms in the sense that it identified a larger number of inliers and yielded a fundamental matrix closer to the ground truth, in terms of a smaller subspace angle between vec(F) and vec(F_t).

Note that while in Tables 3.1, 3.3 the proposed method leads to larger variance, the corresponding mean value is higher, indicating that the other methods have consistently lower performance. Indeed, a quick computation assuming gaussian distributions shows that our method yields a higher objective value with probability larger than 0.7 even in the most unfavorable case.

In terms of computational efficiency, since our algorithm requires solving semi-definite optimization problems, each iteration is more time consuming than those of SACs. On the other hand, consistent numerical experience shows that only a few iterations are needed for convergence (typically no more than 14). Since each iteration takes roughly 10 seconds when using 100 correspondences, the overall computational cost remains competitive vis-à-vis randomized methods, specially in cases involving large number of outliers.

3.6 Conclusions

In this chapter we consider the problem of robust linear model fitting from samples corrupted by noise and outliers. The main result is a novel framework in the form of constrained polynomial optimization problem, which is flexible to incorporate extra requirements on the model parameter such as rank deficient constraint, and also computationally tractable algorithms based on moments theory. These results are illustrated with an application in computer vision: fundamental matrix learning, showing that the proposed method performs well, even in scenarios characterized
CHAPTER 3. ROBUST LINEAR MODEL FITTING

by a very high percentage of outliers, consistently outperforming existing techniques. Research is
currently under way seeking implementations based on the first order methods (as opposed to the
interior point methods used by conventional SDP solvers), in order to further reduce the computational
burden.

Table 3.1: Mean and Standard Deviation of Precision (%)

<table>
<thead>
<tr>
<th>N_{out}</th>
<th>RANSAC</th>
<th>MSAC</th>
<th>LMEDS</th>
<th>MLESAC</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>97.7586 (1.7967)</td>
<td>97.6864 (1.7241)</td>
<td>97.8189 (1.9095)</td>
<td>97.9042 (1.7129)</td>
<td>98.2640 (1.9726)</td>
</tr>
<tr>
<td>20</td>
<td>94.7582 (3.0774)</td>
<td>94.6092 (2.3474)</td>
<td>94.5228 (3.0246)</td>
<td>94.6768 (3.9948)</td>
<td>96.6838 (2.8109)</td>
</tr>
<tr>
<td>30</td>
<td>90.3102 (5.1555)</td>
<td>90.6371 (3.8062)</td>
<td>90.1515 (4.8416)</td>
<td>90.6361 (5.0145)</td>
<td>93.5078 (4.6797)</td>
</tr>
<tr>
<td>40</td>
<td>83.8319 (6.8161)</td>
<td>83.6164 (5.6156)</td>
<td>83.9802 (7.3236)</td>
<td>84.0462 (6.6029)</td>
<td>90.1450 (6.5219)</td>
</tr>
<tr>
<td>50</td>
<td>74.9536 (10.7474)</td>
<td>74.7678 (6.6748)</td>
<td>74.7990 (10.0023)</td>
<td>75.8945 (9.5612)</td>
<td>83.8643 (8.7383)</td>
</tr>
<tr>
<td>60</td>
<td>58.3182 (12.7762)</td>
<td>59.4671 (9.1357)</td>
<td>60.3656 (12.4958)</td>
<td>60.9045 (10.9836)</td>
<td>75.8507 (14.0279)</td>
</tr>
<tr>
<td>70</td>
<td>43.5151 (13.0902)</td>
<td>42.9147 (12.2906)</td>
<td>44.8255 (12.9395)</td>
<td>44.0817 (13.3457)</td>
<td>62.6045 (16.8592)</td>
</tr>
</tbody>
</table>

Table 3.2: Mean and Standard Deviation of Recall (%)

<table>
<thead>
<tr>
<th>N_{out}</th>
<th>RANSAC</th>
<th>MSAC</th>
<th>LMEDS</th>
<th>MLESAC</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>75.6074 (4.8220)</td>
<td>71.9333 (4.5412)</td>
<td>69.1296 (5.1152)</td>
<td>72.9741 (6.2715)</td>
<td>89.2111 (7.8245)</td>
</tr>
<tr>
<td>20</td>
<td>74.1917 (5.5170)</td>
<td>70.8500 (4.7135)</td>
<td>70.6417 (5.3528)</td>
<td>71.6167 (6.5717)</td>
<td>85.5804 (8.9735)</td>
</tr>
<tr>
<td>30</td>
<td>72.1571 (6.3000)</td>
<td>70.0000 (5.1970)</td>
<td>70.6714 (6.1241)</td>
<td>69.9000 (7.3354)</td>
<td>80.1794 (10.3759)</td>
</tr>
<tr>
<td>40</td>
<td>69.4556 (7.7685)</td>
<td>67.1944 (7.7287)</td>
<td>68.7500 (8.4810)</td>
<td>67.6278 (7.8525)</td>
<td>77.3000 (10.9068)</td>
</tr>
<tr>
<td>50</td>
<td>64.9600 (10.5049)</td>
<td>62.4600 (9.8250)</td>
<td>62.3733 (10.8507)</td>
<td>63.4733 (9.6938)</td>
<td>71.9000 (11.1134)</td>
</tr>
<tr>
<td>60</td>
<td>57.1500 (13.4839)</td>
<td>56.9583 (10.6995)</td>
<td>55.7667 (13.8929)</td>
<td>57.3917 (10.8910)</td>
<td>66.7817 (15.0884)</td>
</tr>
<tr>
<td>70</td>
<td>50.2222 (14.3284)</td>
<td>47.5000 (16.3911)</td>
<td>45.7000 (14.0125)</td>
<td>48.8566 (14.7437)</td>
<td>62.0917 (18.6231)</td>
</tr>
</tbody>
</table>

Table 3.3: Mean and Standard Deviation of Hmean (%)

<table>
<thead>
<tr>
<th>N_{out}</th>
<th>RANSAC</th>
<th>MSAC</th>
<th>LMEDS</th>
<th>MLESAC</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>85.9271 (2.9264)</td>
<td>83.7512 (2.8941)</td>
<td>82.1771 (3.2714)</td>
<td>84.4428 (3.7745)</td>
<td>93.5321 (4.7464)</td>
</tr>
<tr>
<td>20</td>
<td>83.7798 (3.6833)</td>
<td>81.7846 (2.7879)</td>
<td>81.6613 (3.7210)</td>
<td>82.2578 (4.2533)</td>
<td>90.8371 (5.5753)</td>
</tr>
<tr>
<td>30</td>
<td>80.6653 (4.9393)</td>
<td>79.5672 (3.8809)</td>
<td>79.7593 (4.8613)</td>
<td>79.4978 (5.4053)</td>
<td>86.4264 (7.2137)</td>
</tr>
<tr>
<td>40</td>
<td>76.2084 (6.3890)</td>
<td>74.8478 (6.1565)</td>
<td>75.8708 (7.0401)</td>
<td>75.2759 (6.1798)</td>
<td>83.2783 (7.8551)</td>
</tr>
<tr>
<td>50</td>
<td>69.6849 (10.0422)</td>
<td>68.2157 (7.4362)</td>
<td>68.1352 (9.4998)</td>
<td>69.2968 (8.8621)</td>
<td>77.4227 (8.5625)</td>
</tr>
<tr>
<td>60</td>
<td>57.6164 (12.6155)</td>
<td>58.0626 (9.4320)</td>
<td>57.7907 (12.2474)</td>
<td>59.0019 (10.2399)</td>
<td>70.8664 (13.3851)</td>
</tr>
<tr>
<td>70</td>
<td>46.6358 (13.3631)</td>
<td>45.0231 (14.0348)</td>
<td>45.0193 (12.6929)</td>
<td>46.2706 (13.5550)</td>
<td>62.0409 (16.7669)</td>
</tr>
</tbody>
</table>

Table 3.4: Mean and Standard Deviation of Similarity (× 100)

<table>
<thead>
<tr>
<th>N_{out}</th>
<th>RANSAC</th>
<th>MSAC</th>
<th>LMEDS</th>
<th>MLESAC</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>99.7772 (0.7316)</td>
<td>99.6793 (0.0126)</td>
<td>99.4775 (2.2484)</td>
<td>99.9650 (0.0908)</td>
<td>99.9828 (0.0716)</td>
</tr>
<tr>
<td>20</td>
<td>99.6139 (1.7647)</td>
<td>99.1964 (0.0129)</td>
<td>99.2437 (3.3507)</td>
<td>99.9289 (0.3425)</td>
<td>99.9860 (0.0697)</td>
</tr>
<tr>
<td>30</td>
<td>99.4306 (2.6753)</td>
<td>99.2144 (0.0207)</td>
<td>99.6500 (1.4154)</td>
<td>99.8947 (0.3682)</td>
<td>99.9920 (0.0176)</td>
</tr>
<tr>
<td>40</td>
<td>99.6427 (1.0935)</td>
<td>99.0665 (0.0189)</td>
<td>99.5319 (2.5283)</td>
<td>99.8482 (0.3942)</td>
<td>99.9918 (0.0187)</td>
</tr>
<tr>
<td>50</td>
<td>98.5472 (5.1293)</td>
<td>96.9745 (0.3331)</td>
<td>99.4611 (3.4161)</td>
<td>99.6875 (1.3350)</td>
<td>99.9876 (0.0327)</td>
</tr>
<tr>
<td>60</td>
<td>97.6169 (8.7403)</td>
<td>94.9443 (0.0570)</td>
<td>99.0108 (5.0077)</td>
<td>98.8553 (5.3757)</td>
<td>99.9878 (0.0289)</td>
</tr>
<tr>
<td>70</td>
<td>96.6981 (9.0199)</td>
<td>94.5070 (0.2751)</td>
<td>98.7259 (5.6968)</td>
<td>98.5653 (5.0236)</td>
<td>99.9931 (0.0078)</td>
</tr>
</tbody>
</table>
Figure 3.3: Comparison for Image Pairs (House, Library, Merton I, Merton II, Merton III and Wadham from top to bottom, in each row, the figures are Precision, Recall, Hmean and Similarity, respectively.)
Chapter 4

Subspace Clustering with Priors

4.1 Motivation

In the previous chapter, we considered the problem of single linear model fitting from samples corrupted by noise and outliers. However, many complicated problems of practical interest involve fitting a given set of samples into several simple linear models. Take face recognition under varying illumination as an example, if the data associated with different persons were modeled by a single linear model, the resulting model would be very complicated, but on the other hand, the data associated with the same person can be modeled by a much simpler linear model, thus, face recognition prefers clustering the data according to person identity and establishing one model for each cluster.

In this chapter we consider the problem of subspace clustering from noisy data, potential corrupted with outliers. Most of the existing methods for solving this problem (for instance, GPCA, SSC, LRR, RSC, FRR) possess two drawbacks. One is that they do not have a direct control on the fitting error. The other is that they assume the underlying topology of the data is complete, i.e. any data point probably belongs to any subspace; conversely, these methods are unable to make use of any prior knowledge on the distribution of the data. Motivated by these drawbacks, we aim to propose a novel general approach which can identify a set of subspace parameters that restrain the fitting error below a presumed level and can also easily incorporate existing prior information, such as the relative frequency of each subspace and co-occurrence information (points known to be in the same of different subspaces).

In the first part of this chapter, we give a theoretical derivation of the main results. The main idea of the proposed approach is to recast the problem into a constrained polynomial optimization
problem, which in turn is relaxed to a sequence of convex optimization problems by appealing to moments based optimization techniques introduced in Chapter 2. As we show later this sequence of relaxation guarantees to converge to a subspace arrangement compatible with the \textit{a priori} information. However, with the order of relaxations increasing, the exponentially increased computation burden makes the problem intractable. To circumvent this difficulty, we provide convergence certificates for finite order relaxations, i.e. testable conditions guaranteeing that a finite order relaxation has indeed found a solution.

In the second part, the proposed approach is tested on several examples, including both synthetic and real data, where the ability to incorporate priors is key to obtaining the correct segmentation.

### 4.2 Problem Statement

The goal of this chapter is to recover a collection of subspaces from a cloud of noisy data such that certain priors are satisfied, or show that none exists. Formally, it can be stated as:

\begin{Problem}
Given:

- A set of noisy samples $X = \{x_j| x_j \in \mathbb{R}^n, x_j = \hat{x}_j + \eta_j, \forall j = 1, \ldots, N_p\}$ drawn from $N_s$ distinct linear subspaces $\{S_i \subset \mathbb{R}^n\}_{i=1}^{N_s}$ of dimension $n - 1$ of the form $S_i = \{\hat{x}|\hat{x} \in \mathbb{R}^n, r_i^T\hat{x} = 0\}$, where $r_i \in \mathbb{R}^n$, $\|r_i\|_2 = 1$ and $i = 1, \ldots, N_s$.

- A-priori information consisting of (i) a bound $\epsilon$ on the distance from the noisy sample to the corresponding subspace, (ii) a bound $N_o$ on the number of outliers, (iii) $N_{f_i}$, the relative frequency of each subspace, and (iv) point wise co-occurrence information.

Establish whether the observed data is consistent with the a-priori assumptions and, in that case, find a set of subspaces compatible with the a-priori information and assign the (inlier) samples to each subspace. That is, find a set of normals $\{r_i \in \mathbb{R}^n\}_{i=1}^{N_s}$ and a partition of the samples in $N_s + 1$ sets $\{X_i\}_{i=1}^{N_s}, X_o$ such that $\text{card}(X_o) \leq N_o$, $\text{card}(X_i) = N_{f_i} N_p$, and

$$|r_i^T x| \leq \epsilon \quad \text{for all} \quad x \in X_i.$$  
\end{Problem}
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

4.3 A Greedy Approach to Clustering

For Problem 3 if we consider samples drawn from one single subspace \( S_i \) as inliers to \( S_i \), and those drawn from other subspaces \( S_j \), with \( j \neq i \), as outliers to \( S_i \), then identifying the specific subspace \( S_i \) from \( X \) is reduced to the robust linear model fitting problem analyzed in Chapter 3. Motivated by this observation, we come up with a simple approach to subspace clustering in a greedy fashion as shown in Figure 4.1 where subspaces are determined step by step rather than simultaneously. As outlined in Algorithm 3, at each step a robust linear model fitting problem (3.6) is solved, and the identified inliers are removed from the samples used for next step.

![Figure 4.1: Greedy Subspace Clustering](image)

Algorithm 3 Greedy Subspace Clustering

1: **Initialize:** \( N_s = 0, n_o = N_p, X_o = X, N = \{1, \ldots, N_p\} \);
2: **while** \( n_o > n - 1 \) **do**
3: \( N_s = N_s + 1 \);
4: solve (3.6) by Algorithm 1 or Algorithm 2 with samples \( X_o \);
5: \( J \subset N \) is the union of \( j \) with \( s_j = 1 \), then \( x_j \in S_{N_s} \);
6: \( X_o = X_o \setminus \{x_j, j \in J\} \);
7: \( N = N \setminus J, n_o = n_o - \text{cardinality}(J) \);
8: **end while**

4.4 A Non-Greedy Approach to Clustering

In this section we present the main result of this chapter, a moments based convex optimization approach to solving Problem 3. The main idea is to recast the problem first into an equivalent nonlinear feasibility problem which can then be relaxed to a sequence of convex optimization problems using the results outlined in Chapter 2. For simplicity we will first consider that case with no
outliers and without any prior information except the number of subspaces. Handling of these cases will be covered in Section 4.4.4 and Section 4.4.5 right after the description of the basic algorithm and supporting theory.

4.4.1 Clustering as a Nonlinear Feasibility Problem

It can be easily shown that by introducing a set of binary variables \( s_{i,j} \) that indicate whether \( x_j \) belongs to \( S_i \) (i.e. \( s_{i,j} = 1 \iff x_j \in S_i \)) , and auxiliary variables \( e_{i,j} \), Problem 3 is equivalent to Problem 4.

Problem 4. Determine the feasibility of the following set of nonlinear equations:

\[
\begin{align*}
\forall_{i=1}^{N_s} \forall_{j=1}^{N_p} : & \quad r_i^T x_j - e_{i,j} = 0 \\
\forall_{i=1}^{N_s} \forall_{j=1}^{N_p} : & \quad |s_{i,j} e_{i,j}| \leq \epsilon s_{i,j} \\
\forall_{i=1}^{N_s} : & \quad \|r_i\|_2 = 1 \\
& \quad r_1(1) \geq r_2(1) \geq \ldots \geq r_{N_s}(1) \geq 0 \\
\forall_{i=1}^{N_s} \forall_{j=1}^{N_p} : & \quad s_{i,j}^2 = s_{i,j} \\
\forall_{j=1}^{N_p} : & \quad \sum_{i=1}^{N_s} s_{i,j} = 1 \\
\end{align*}
\]  

Here the first two constraints are equivalent to \( |r_i^T x_j| \leq \epsilon \) if \( s_{i,j} = 1 \); the fourth means to eliminate the symmetry of solutions; the fifth imposes that \( s_{i,j} \in \{0, 1\} \) and the sixth forces each point \( x_i \) to be assigned to exactly one subspace. Next, we will show that Problem 4 can be recast as a polynomial optimization problem.

Lemma 2. The following two statements are equivalent:

(i) Problem 4 is feasible.

(ii) \( p^* = 0 \), where \( p^* \) represents the optimal value of the objective function in the following polynomial optimization problem:

\[
p^* = \min_{r_i, s_{i,j}, e_{i,j}} \sum_{j=1}^{N_p} \sum_{i=1}^{N_s} (r_i^T x_j - e_{i,j})^2 \\
\text{subject to} \\
\forall_{i=1}^{N_s} : & \quad \|r_i\|_2 = 1 \\
& \quad r_1(1) \geq r_2(1) \geq \ldots \geq r_{N_s}(1) \geq 0 \\
\forall_{i=1}^{N_s} \forall_{j=1}^{N_p} : & \quad |s_{i,j} e_{i,j}| \leq \epsilon s_{i,j} \\
\forall_{i=1}^{N_s} \forall_{j=1}^{N_p} : & \quad s_{i,j}^2 = s_{i,j} \\
\forall_{j=1}^{N_p} : & \quad \sum_{i=1}^{N_s} s_{i,j} = 1
\]
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

Proof. The proof is immediate due to the fact that the objective in (4.3) is a sum of squares. Hence
\[ p^* \geq 0 \text{ and } p^* = 0 \iff \forall_{i=1}^{N_s} \forall_{j=1}^{N_p} : r_i^T x_j - e_{i,j} = 0 \] holds for (4.3) if and only if (4.2) is feasible.

4.4.2 A Sequence of Convex Relaxations

From the results in Chapter 2, it follows that the problem in the form of (4.3) can be relaxed to a sequence of convex problems of the form:

\[
p^*_m,N = \min_{\mathbf{m}_{2N}} \quad \mathbf{c}^T \mathbf{m}_{2N} \\
\text{subject to} \\
\mathbf{M}_N(\mathbf{m}_{2N}) \succeq 0 \\
\mathbf{L}_{N-1}(g_k \mathbf{m}_{2N}) \succeq 0, \forall g_k
\]

where \( \mathbf{c}^T \mathbf{m}_{2N} \) denotes the expectation of the objective function in (4.3). \( \mathbf{m}_{2N} \) denotes the expectation of all the monomials of \( \mathbf{v} = \{\forall_{i=1}^{N_s} \forall_{j=1}^{N_p} : r_i, s_{i,j}, e_{i,j}\} \) up to order 2\( N \), and \( \mathbf{c} \) is the vector of the coefficients corresponding to the monomials. \( \mathbf{M}_N \) and \( \mathbf{L}_N \) are the moments and localizing matrices associated with a truncated moments sequence \( \mathbf{m}_{2N} \) of order up to 2\( N \). \( g_k \) denotes all the constraints in (4.3). As pointed out by Theorem 3.4 in [25], \( p^*_m,N \) provides a lower bound to \( p^* \) in (4.3), that is, \( p^*_m,N \leq p^*_m,N+1 \leq \cdots \leq p^* \). The following theorem provides further analysis on the conditions under which the equality holds.

Theorem 6. (i) If Problem 4 is feasible, then \( p^*_m,N = p^* = 0 \) at \( N = 1 \);
(ii) Problem 4 is infeasible if and only if there exists some \( N \) such that \( p^*_m,N > 0 \).

Proof. (i) From Lemma 2 if Problem 4 is feasible then \( p^* = 0 \). Thus \( p - p^* = \sum_{j=1}^{N_p} \sum_{i=1}^{N_s} (r_i^T x_j - e_{i,j})^2 = \text{sum of squares} \). The fact that \( p^*_m,N = p^* = 0 \) at \( N = 1 \) follows now from Theorem 2 in Chapter 2.

(ii) \( p^* \geq p^*_m,N > 0 \Rightarrow \) Problem 4 is infeasible follows from the fact that \( p^*_m,N \leq p^* \). Conversely, from (monotonic) convergence of the sequence of relaxations (Theorem 4.2 in [25]) it follows that if \( p^* > 0 \) then there exists some \( N \) such that \( p^*_m,N > 0 \). \( \square \)
4.4.3 Computational Complexity Analysis

Both the objective \( p \) and the constraints \( K \) in (4.3) can be partitioned into \( N_p \) partitions as

\[
p = \sum_{j=0}^{N_p} p_j \quad \text{with} \quad p_0 = 0, \quad \forall j = 1, \ldots, N_p, \quad \forall j \in K_j = \bigcap_{j=1}^{N_p} K_j
\]

\[
K = \bigcap_{j=1}^{N_p} K_j \quad \text{with} \quad K_0 : \begin{cases}
   \forall i = 1 : \|r_i\|_2 = 1 \\
   r_1(1) \geq r_2(1) \geq \ldots \geq r_{N_s}(1) \geq 0 \\
   \forall j = 1, \ldots, N_p : \begin{cases}
      \forall i = 1 : |s_{i,j} e_{i,j}| \leq \epsilon_{i,j} \\
      \forall i = 1 : s_i^2 = s_{i,j} \\
      \sum_{i=1}^{N_s} s_{i,j} = 1
   \end{cases}
\end{cases}
\]

(4.5)

it is easy to check that for \( \forall j = 0, 1, \ldots, N_p, \ p_j \) and \( K_j \) contain variables in the subsets \( v_j \), where \( v_0 = \{\forall i = 1 : r_i\} \), and for \( j = 1, \ldots, N_p, \ v = \{\forall i = 1 : r_i, s_{i,j}, e_{i,j}\} \). The partitions of variables satisfy \( \bigcup_{j=0}^{N_p} v_j = v \) and

\[
v_{j+1} \cap \bigcup_{l=1}^{j} v_l = \{\forall i = 1 : r_i\} = v_0, \ \text{for} \ \forall j = 1, \ldots, N_p - 1
\]

meaning the problem exhibits the running intersection property and hence it can be simplified to solving

\[
p^*_{m,N} = \min_{m_{j,2N}} \sum_{j=1}^{N_p} c_j^T m_{j,2N}
\]

subject to

\[
\forall j = 1 : M_{j,N}(m_{j,2N}) \succeq 0 \\
\forall j = 1 : L_{j,N-1}(g_{jk}, m_{j,2N}) \succeq 0, \ \forall g_{jk} \in K_j
\]

(4.6)

where \( m_{j,2N} \) and \( M_{j,2N} \) denote the moments sequence of variables in \( v_j \) of order up to \( 2N \) and the associated moment matrix of order \( N \), \( c_j^T m_{j,2N} \) is the expectation of \( p_j \), and \( L_{j,N-1} \) is the localizing matrix related to constraints \( g_{jk} \) in \( K_j \). Instead of considering a single moment matrix, \( M_{N} \), of size \( \binom{N_s(n+2N_p)+N}{N_s(n+2N_p)} \) in (4.4), in (4.6) we consider \( N_p \) smaller moment matrices, \( M_{j,N} \), of size \( \binom{N_s(n+2)+N}{N_s(n+2)} \), substantially reducing the dimension of moments sequence from \( \mathcal{O}([N_s(n+2N_p)]^{2N}) \) to \( \mathcal{O}([N_s(n+2)]^{2N}) \).

Further, if Problem 4 is feasible, then the relaxation (4.6) corresponding to \( N = 1 \) is
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

exact. However, solving this relaxation only yields the value of \( p^* \) and the corresponding (truncated) moments sequence of the variables of interest (the \( r_i \)). In general, reconstructing a minimizer from the corresponding moments sequence is far from trivial (see for instance [49] for a description of a tool to accomplish this). To avoid this difficulty, in the sequel we will exploit the fact that moment matrices associated with atomic measures having a single atom is rank-1, since in this case the moments simply correspond to the powers of the variables, evaluated at the location of the atom. This observation motivates the next result, involving a modification of (4.6) that enforces low rank solutions.

**Theorem 7.** Assume that (4.2) is feasible and let \( m_{j,2}^* \) denote an optimal solution to the following problem

\[
p_{m,1}^* = \min_{m_{j,2}} \sum_{j=1}^{N_p} c_j^T m_{j,2}
\]

subject to

\[
\forall_{N_p}^{j=0} : M_{j,1}(m_{j,2}) \succeq 0, M_{j,1}(1,1) = 1
\]

(4.7)

\[
\forall_{N_p}^{j=0} : L_{j,0}(g_{j_k} m_{j,2}) \succeq 0, \forall g_{j_k} \in K_j
\]

\[
\forall_{j=1}^{N_p} : M_{j,1}(1 : nN_s + 1, 1 : nN_s + 1) = M_{0,1}
\]

\[
\text{rank}\{M_{0,1}\} = 1
\]

Then the vectors, \( r_i^* = m_{j,2}^*(r_i) \), \( \forall i = 1, \ldots, N_s \), obtained from the first order moment variables are feasible solutions to Problem 3.

**Proof.** Since \( \text{rank}\{M_{0,1}^*\} = 1 \), it follows that \( M_{0,1}^* \) has an associated atomic measure \( \mu_{r^*} = \delta(\{r_i^*\}_{i=1}^{N_s}) \). Consider now the following optimization problem

\[
p_{\mu}^* = \min_{\text{supp}(\mu)=K} \sum_{j=1}^{N_p} \sum_{i=1}^{N_s} (r_i^T x_j - e_{i,j})^2
\]

(4.8)

where \( K \) denotes the semi-algebraic set defined by the constraints in (4.3), and

\[
p_{\mu_{r^*}}^* = \min_{\mu_r} \sum_{j=1}^{N_p} \sum_{i=1}^{N_s} (r_i^T x_j - e_{i,j})^2
\]

(4.9)

where \( \mu_r \) denotes the subset of measures \( \mu \) supported in \( K \) and having marginal distributions with respect to \( r_k \) given by \( \delta(r_k^*) \). If \( \text{Problem 4} \) is feasible, or equivalently \( p^* = 0 \) in (4.3), it follows from Theorem 2 in Chapter 2, equation (4.7), and the fact that \( N = 1 \), that \( p_{m,1}^* = p_{\mu_{r^*}}^* = 0 \). Let \( \mu_{r^*} \) denote
an optimal measure for (4.9). From the fact that $p^*_{\mu_r} = 0$ it follows that $E_{\mu_r}(r^T_ix_j - e_{i,j})^2 = 0$ holds for $\forall i, j$, or equivalently $(r^*^T_ix_j - e_{i,j})^2 = 0$ almost everywhere in $K$, which is equivalent to feasibility of (4.2).

Since the problem above involves a rank constraint, it is no longer convex. However, a convex relaxation that seeks to obtain low rank solutions can be obtained by iteratively minimizing the re-weighted nuclear norm of $M_{0,1}$ (see for instance [50] for a discussion of the convergence properties of this relaxation), leading to Algorithm 4 outlined below, where we enforce $p^*_{m,1} = 0$ by setting each term in the sum of squares to 0.

**Algorithm 4 Moments-Based Subspace Clustering**

1: Initialize: $k = 0$, $W^{(0)} = I$

2: repeat

3: Solve

\[
\begin{align*}
\text{minimize}_{m_j,2} & \quad \text{Trace}(W^{(k)}M_{0,1}) \\
\text{subject to} & \quad c_j^Tm_{j,2} = 0 \\
& \quad \forall N_{p_j}^1: M_{j,1}(m_{j,2}) \succeq 0, M_{0,1}(1, 1) = 1 \\
& \quad \forall N_{p_j}^0: L_{j,0}(g_{jk}m_{j,2}) \succeq 0, \forall g_{jk} \in K_j \\
& \quad \forall j = 1: M_{j,1}(1 : nN_s + 1, 1 : nN_s + 1) = M_{0,1}
\end{align*}
\]

4: Update

\[
W^{(k+1)} = \left[ M_{0,1}^{(k)} + \sigma_2(M_{0,1}^{(k)}) \right]^{-1} \\
k = k + 1
\]

5: until rank$\{M_r\} = 1$.

### 4.4.4 Handling Outliers

We return now to the issue of handling outliers, defined as those points that lie beyond a given distance $\epsilon$ from every subspace in the arrangement (e.g. such that $\min_i |r_i^T x| > \epsilon$). They can be handled by simply relaxing the requirement that each point must be assigned to a certain subspace,
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

leading to the following constrained polynomial optimization problem:

\[
\begin{align*}
\text{maximize} & \quad \sum_{j=1}^{N_p} \sum_{i=1}^{N_s} s_{i,j} \\
\text{subject to} & \quad \forall N_s \sum_{i=1}^{N_s} s_{i} = 1 \\
& \quad \forall N_p \sum_{j=1}^{N_p} r_i - e_{i,j} = 0 \\
& \quad \forall N_s \sum_{i=1}^{N_s} |s_{i,j} e_{i,j}| \leq \epsilon s_{i,j} \\
& \quad \forall N_s \sum_{i=1}^{N_s} ||r_i||_2 = 1 \\
& \quad r_1(1) \geq r_2(1) \geq \ldots \geq r_{N_s}(1) \geq 0 \\
& \quad \forall N_p \sum_{j=1}^{N_p} s_{i,j}^2 = s_{i,j} \\
& \quad \forall N_p \sum_{i=1}^{N_p} \sum_{i=1}^{N_s} s_{i,j} \leq 1
\end{align*}
\]

which seeks to maximize the number of inliers.

Since (4.11) exhibits a sparsity pattern similar to that in (4.3), it can be solved by a modified version of Algorithm 4, where the equality constraint \( \sum_{i=1}^{N_s} s_{i,j} = 1 \) is replaced by \( \sum_{i=1}^{N_p} s_{i,j} \leq 1 \) (to accommodate the case where \( x_j \) is an outlier), and the objective function is replaced by

\[
\tilde{p} = \sum_{i=1}^{N_s} \sum_{j=1}^{N_p} (1 - m_j(s_{i,j})) + \lambda \operatorname{Tr}(W^{(k)}M_{0,1}),
\]

where \( \lambda > 0 \) is a parameter. Thus, this objective function penalizes both the number of outliers and the rank of \( M_{0,1} \).

**Remark 1.** A bound \( N_o \) on the number of outliers can be handled via a constraint of the form \( \sum_{i=1}^{N_s} \sum_{j=1}^{N_p} s_{i,j} \geq N_p - N_o \). However, since this constraint subverts the sparsity of the problem, the formulation (4.11) is preferable.

The left subfigure in Figure 1.2 shows the results of a toy synthetic example, consisting of inliers \( \{ x_j \in S_i : |r_i^T x_j| \leq 0.1 \} \), corrupted by several outliers \( \{ x_j : |r_i^T x_j| > 0.1, \forall N_s \} \). Solving the moments-based relaxation corresponding to (4.11) with \( \epsilon = 0.10 \) leads to perfect identification of both the subspaces and outliers. On the other hand, solving the same example using (4.3), with \( \epsilon = 0.30 \) (minimum value of \( \epsilon \) needed for feasibility) leads to the results shown in the right subfigure in Figure 1.2, where the subspaces are substantially skewed by the outliers.

### 4.4.5 Handling Additional A-Prior Information

In many scenarios of practical interest, a-priori information on the labels of some sample points is available. For instance, in surveillance videos, it is easy to identify points lying on buildings,
and hence definitely background, and often points belonging to moving vehicles and pedestrians. Similarly, in many situations, information is available about the relative size of the target, and thus (assuming uniform sampling) on the relative frequency of points in the corresponding subspace. As shown below, this additional information can be easily incorporated into our formulation by simply adding suitable polynomial constraints on the variables $s_{i,j}$. Specifically:

(i) $f\%$ of $\mathbf{X}$ belongs to $\mathcal{S}_i \iff \sum_{j=1}^{N_p} s_{i,j} = 0.01 f N_p$;

(ii) $x_m, x_n$ belong to the same subspace $\iff s_{i,m} = s_{i,n}, \forall i = 1, \cdots, N_s$;

(iii) $x_m, x_n$ belong to different subspaces $\iff s_{i,m} + s_{i,n} \leq 1, \forall i = 1, \cdots, N_s$.

The advantages of being able to exploit this information will be illustrated in Section 4.5.2.

### 4.4.6 Subspace Recovery Conditions

For the proposed constrained polynomial optimization framework, Theorem 7 guarantees that the correct clustering is recovered for the cases of “clean” and “clean+outlier” data. Further for the cases with intense noise for inliers, “noisy data” and “noisy data+outliers”, the proposed approach guarantees to detect the outliers and fit the inliers to subspaces within the given noise bound $\epsilon$, which is consistent with the given prior information. We also provide strong optimality certificates for the relaxed SDP based algorithm: obtaining a solution with rank $\{M_{0,1}\} = 1$ certifies that the correct clustering has been found.

On the other hand, to the best of our knowledge, the exact recovery properties of LRR and SSC are guaranteed only under some preconditions on either the arrangement of subspaces or the distribution of the data, as summarized in Table 4.1 (based on references [60–66]). Thus, we believe that our algorithm, which does not require these, provides recovery guarantees at least as good as LRR and SCC, with the additional advantage of being able to certify optimality of a given solution.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>LRR</th>
<th>SSC</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>subspaces</td>
<td>independent</td>
<td>independent/disjoint</td>
<td>none</td>
</tr>
<tr>
<td>clean data</td>
<td>exact</td>
<td>exact/conditional</td>
<td>exact</td>
</tr>
<tr>
<td>clean+outliers</td>
<td>conditional</td>
<td>$x$</td>
<td>exact</td>
</tr>
<tr>
<td>noisy</td>
<td>$x$</td>
<td>conditional [66]</td>
<td>consistent</td>
</tr>
<tr>
<td>noisy +outliers</td>
<td>$x$</td>
<td>$x$</td>
<td>consistent</td>
</tr>
</tbody>
</table>

Table 4.1: Recovery Conditions (“$x$” denotes no certificates)
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

4.5 Experiments

In this section, we demonstrate the advantage of the proposed method using both synthetic data and three computer vision motivated applications: texture segmentation, motion segmentation and planar segmentation by homography learning.

4.5.1 Synthetic Static Data

We first investigated the performance of the moments-based subspace clustering algorithm on synthetic static data as the number of subspaces, their dimensions and noise levels changed. For each set of experiment, we randomly generated 20 instances with sample points lying on a union of multiple linear subspaces corrupted by noise normal to the corresponding subspace and with uniform random magnitudes in $[0.8\epsilon, \epsilon]$. A comparison of the performance of Algorithm 4, implemented in Matlab using CVX [67], against existing state-of-the-art methods is summarized in Table 4.2. As shown there, in all cases the proposed method outperformed the others in terms of the worst-case fitting error of the identified subspaces, given by:

$$
\text{err}_f = \max_{j \in \{1, \ldots, N_p\}} \min_{i \in \{1, \ldots, N_s\}} |r_i^T x_j|
$$

subject to $||r_i||_2 = 1, \forall i = 1, \ldots, N_s$

where $r_i$’s are the normals to the subspaces found by each algorithm. For algorithms that cannot obtain $r_i$ directly, like SSC and LRR, we calculated $r_i$ by fitting a subspace to each cluster of points produced by the algorithm.

Next, the effect of outliers was investigated, by randomly corrupting 4 to 6 points with noise of magnitude $3\epsilon$ (the distribution of the data is shown in Figure 1.2. Using Algorithm 4 modified to solve (4.11) instead of (4.3) led to the results shown in the last row of Table 4.2. As illustrated there, the proposed algorithm can detect the outliers precisely and fit the inliers to subspaces within the given error bound.

---

1In order to provide a meaningful comparison, for each set of experiments, the adjustable parameters of each of the algorithms listed in the table were experimentally tuned to minimize the misclassification rate.
Table 4.2: Performance Comparison for Different Synthetic Data Scenarios \( n \): dimension of data, \( d_i \): dimension of each subspace, \( N_i \): number of samples on each subspace, \( \epsilon \): error bound, \( \mu \) and \( \sigma \): mean and standard deviation of \( \text{err}_i \), (*) : experiments with outliers

<table>
<thead>
<tr>
<th>( n )</th>
<th>( d_i )</th>
<th>( N_i )</th>
<th>( \epsilon )</th>
<th>Algorithm</th>
<th>Ozay’s</th>
<th>GPCA</th>
<th>SSC</th>
<th>LRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>[2 2]</td>
<td>[50 50]</td>
<td>0.10</td>
<td>( \mu = 0.0992 )</td>
<td>( \mu = 0.1015 )</td>
<td>( \mu = 0.1234 )</td>
<td>( \mu = 0.2465 )</td>
<td>( \mu = 0.3349 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma = 0.0003 )</td>
<td>( \sigma = 0.0034 )</td>
<td>( \sigma = 0.0230 )</td>
<td>( \sigma = 0.0705 )</td>
<td>( \sigma = 0.1565 )</td>
</tr>
<tr>
<td>3</td>
<td>[2 2]</td>
<td>[50 50]</td>
<td>0.15</td>
<td>( \mu = 0.1485 )</td>
<td>( \mu = 0.1608 )</td>
<td>( \mu = 0.2980 )</td>
<td>( \mu = 0.3224 )</td>
<td>( \mu = 0.4054 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma = 0.0008 )</td>
<td>( \sigma = 0.0140 )</td>
<td>( \sigma = 0.1696 )</td>
<td>( \sigma = 0.0831 )</td>
<td>( \sigma = 0.1417 )</td>
</tr>
<tr>
<td>3</td>
<td>[2 2]</td>
<td>[50 50]</td>
<td>0.20</td>
<td>( \mu = 0.1978 )</td>
<td>( \mu = 0.2020 )</td>
<td>( \mu = 0.4495 )</td>
<td>( \mu = 0.3793 )</td>
<td>( \mu = 0.4635 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma = 0.0010 )</td>
<td>( \sigma = 0.0143 )</td>
<td>( \sigma = 0.1459 )</td>
<td>( \sigma = 0.1112 )</td>
<td>( \sigma = 0.1083 )</td>
</tr>
<tr>
<td>4</td>
<td>[3 3]</td>
<td>[50 50]</td>
<td>0.15</td>
<td>( \mu = 0.1481 )</td>
<td>( \mu = 0.2058 )</td>
<td>( \mu = 0.3229 )</td>
<td>( \mu = 0.3495 )</td>
<td>( \mu = 0.3622 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma = 0.0013 )</td>
<td>( \sigma = 0.1029 )</td>
<td>( \sigma = 0.1430 )</td>
<td>( \sigma = 0.0588 )</td>
<td>( \sigma = 0.0885 )</td>
</tr>
<tr>
<td>2</td>
<td>[1 1 1]</td>
<td>[40 40 40]</td>
<td>0.10</td>
<td>( \mu = 0.0993 )</td>
<td>( \mu = 0.2190 )</td>
<td>( \mu = 0.4943 )</td>
<td>( \mu = 0.1649 )</td>
<td>( \mu = 0.1021 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma = 0.0003 )</td>
<td>( \sigma = 0.0212 )</td>
<td>( \sigma = 0.0173 )</td>
<td>( \sigma = 0.0452 )</td>
<td>( \sigma = 0.0071 )</td>
</tr>
<tr>
<td>2(*)</td>
<td>[1 1 1]</td>
<td>[40 40 40]</td>
<td>0.10</td>
<td>( \mu = 0.0996 )</td>
<td>( \mu = 0.4601 )</td>
<td>( \mu = 0.6362 )</td>
<td>( \mu = 0.2986 )</td>
<td>( \mu = 0.2223 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma = 0.0003 )</td>
<td>( \sigma = 0.1355 )</td>
<td>( \sigma = 0.1219 )</td>
<td>( \sigma = 0.1086 )</td>
<td>( \sigma = 0.0461 )</td>
</tr>
</tbody>
</table>

4.5.2 Motion Segmentation

The purpose of this example is to illustrate the advantages of exploiting priors in situations with relatively sparse data, using a non-trivial problem arising in computer vision: motion segmentation. The data used, shown in Figure 4.2, consists of (cropped) frames from the ‘cars9’ sequence in the Hopkins 155 database \([3]\). The goal is to segment the car, the van and the background. Features that make this a challenging task are the fact that the motion of the van across the entire sequence is small (and hence it may be confused with the background), and that the car and van have parallel motion. To simulate a sparse, noisy sampling scenario, we selected 20 points from the background \( (S_1) \), and 10 points each from the car \( (S_2) \) and van \( (S_3) \), respectively. The observations were generated by artificially corrupting the clean data from the original Hopkins 155 dataset by random noise uniformly distributed on \([-0.23, 0.23]\), resulting in a signal-to-noise ratio as approximately \( 3.27 : 1 \). This is not an uncommon situation; for example, in surveillance applications where the cameras are located on high buildings, the foreground may turn out to be too small to collect sufficient observations and the observations of the foreground may be severely contaminated by noise. Proceeding as in \([3]\), the data was first projected to points in \( \mathbb{R}^{5 \times 40} \) and the segmentation performed in this subspace. In order to show the advantages of incorporating priors, we performed the experiment in 5 stages, successively increasing the amount of a-priori information as follows:

* **Stage 0:** no prior information;

* **Stage 1:** given \( \{x_1, x_2\} \in S_1 \) and \( \{x_{31}, x_{32}\} \in S_3 \);

* **Stage 2:** given \( \{x_1, x_2\} \in S_1 \) and \( \{x_{31}, x_{32}, x_{33}\} \in S_3 \);

* **Stage 3:** given \( \{x_1, x_2\} \in S_1 \) and \( \{x_{31}, x_{32}, x_{33}, x_{34}\} \in S_3 \);
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

Stage 4: given \( \{x_1, x_2 \} \in S_1 \) and \( \{x_{31}, x_{32}, x_{33}, x_{34}, x_{35} \} \in S_3 \).

Figure 4.2: Sample Frames from the Hopkins 155 ‘cars9’ Dataset (Frame 1, Frame 41, Frame 61)

Figure 4.3 shows the resulting misclassification error rate as a function of the a-priori information. For comparison, we also indicate there the results obtained using existing methods, which cannot incorporate the priors. As shown there, the ability to exploit priors results in a substantial improvement, even when the amount of information is relatively modest.

![Graph showing misclassification error rate as a function of stage.](image)

Figure 4.3: Performance Comparison for the Motion Segmentation Example for Different Sets of Priors

4.5.3 Planar Segmentation

In this section we illustrate the advantages of taking into account prior information in subspace clustering by applying our algorithm to planar segmentation by homography estimation, which is an important problem in image registration and computation of camera motion [40]. Given the homogeneous coordinates of \( N_p \) correspondences from two images \( \{(p_j, p'_j)\}_{j=1}^{N_p} \in \)
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

\( \mathbb{R}^3 \), assuming that these \( N_p \) points are on the same plane, let \( p_j = \begin{bmatrix} x_j & y_j & 1 \end{bmatrix}^T \) and \( p'_j = \begin{bmatrix} x'_j & y'_j & 1 \end{bmatrix}^T \), and let \( H \in \mathbb{R}^{3 \times 3} \) denote the homography. Then \( h = \text{vectorize}(H^T) \) satisfies

\[
\begin{bmatrix}
  x_j^T \\
  x_{j+N_p}^T
\end{bmatrix} h = 0,
\]

with \( x_j^T = \begin{bmatrix} p_j^T & 0_{1 \times 3} & -x'_j p_j^T \end{bmatrix} \), \( x_{j+N_p}^T = \begin{bmatrix} 0_{1 \times 3} & p'_j & -y'_j p'_j \end{bmatrix} \) (4.13)

meaning \( h \) lies in the null space of the matrix \( X = \begin{bmatrix} x_1 & x_2 & \cdots & x_{2N_p} \end{bmatrix}^T \).

Now assume that \( \{ (p_j, p'_j) \}_{j=1}^{N_p} \) are distributed on \( N_s \) different planes (shown as the purple dots in Figure 4.4). In this case (4.13) no longer holds for all \( j = 1, \cdots, N_p \) with a single \( h \). Instead the planes can be segmented by clustering \( \{ x_j \}_{j=1}^{2N_p} \) to \( N_s \) subspaces \( S_i \), characterized by different normal vectors \( h_i, i = 1, \cdots, N_s \) (shown as the blue dots and red dots in Figure 4.4). Thus subspace clustering techniques can be applied to planar segmentation. Specifically, we can formulate this problem as seeking a feasible solution to:

\[
\begin{align*}
&\| h_i \|_2^2 = 1, \quad \forall i = 1, \cdots, N_s \quad (4.14a) \\
&| s_{i,j} h_i^T x_j | \leq \epsilon_{s_{i,j}}, \quad s_{i,j}^2 = s_{i,j}, \quad \forall i = 1, \cdots, N_s \quad \forall j = 1, \cdots, N_p \quad (4.14b) \\
&\Sigma_{i=1}^{N_s} s_{i,j} = 1, \quad \forall j = 1, \cdots, N_p \quad (4.14c) \\
&s_{i,j} = s_{i,j+N_p}, \quad \forall i = 1, \cdots, N_s \quad \forall j = 1, \cdots, N_p \quad (4.14d)
\end{align*}
\]

where (4.14a)-(4.14c) define a subspace clustering problem similar to Problem 4 and (4.14d) represents the prior information that \( x_j \) and \( x_{j+N_p} \) should be in the same subspace since they are derived from a single correspondence \( (p_j, p'_j) \) as in (4.13), which cannot be enforced by the existing subspace clustering methods.

Figure 4.4: Images for Planar Segmentation: Merton I

We tested on three pairs of real images Merton I, Merton II, and Wadham, given by VGG,
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

University of Oxford. Given each pair of images, firstly VLfeat toolbox [58] was used to obtain the SIFT features of two images, and correspondences were defined by those pairs of points whose $\ell_2$ norm are less than 5. Among these correspondences, we randomly generated 20 instances with 30 correspondences on each plane and $N_s = 2$. $\epsilon$ was determined by calculating a ground truth homography matrix for each plane, $H_1, H_2$ and $\epsilon = \max\{\text{erf}_1, \text{erf}_2\}$, $\text{erf}_i = \max_{j \in S_i} |x_j^{T} \text{vec}(H_i^T)|$, $i = 1, 2$. Performance was evaluated by the misclassification rate $\text{err}_1$ among $2N_p$ samples.

![Figure 4.5: Average Performance of LRR and SSC over 20 Instances](image)

**Analysis.** As reported in Table 4.3, our proposed approach outperformed GPCA and the denoised GPCA proposed in [10]. For LRR and SSC given by

(LRR):\[\minimize \|Z\|_* + \lambda \|E\|_{2,1}, \text{ s.t. } X = XZ + E\]

(SSC):\[\minimize \|C\|_1 + 0.5\tau \|E\|_F^2, \text{ s.t. } X = XC + E, \text{diag}(C) = 0,\]

we plotted the performance for $\lambda \in [10^{-4}, 1]$ and $\tau \in [10^{-3}, 10]$ in Figure 4.5 from which we can
see that the range of $\lambda$ ($\tau$) for LRR (SSC) to be competitive with the proposed approach in terms of classification accuracy, is quite small, roughly $\lambda^* \in [0.003, 0.01]$, $\tau^* \in [0.4, 0.7]$. Thus, in this case, LRR and SSC were quite sensitive to the parameters, as shown in Figure 4.5. In addition, such small values of $\lambda^*$ (or $\tau^*$) placed virtually no penalty on the noise terms. As a result, they yielded solutions where the denoised data $X - E$ fitted poorly the actual data. For example, for $\lambda = 0.01$, the misclassification rate of LRR for Merton I was 3.33%. However, as shown in Figure 4.6, LRR produced a solution where where $X$ and $E$ had roughly the same scale. A similar situation (also shown in Figure 4.6) arised with SSC when $\tau = 0.55$, the value yielding the lowest misclassification rate (5%). In contrast, the proposed method did not have to be tuned and yielded an accurate estimate of the homography parameters.

The specific reason why the existing methods performed worse is the structure of the samples for (4.14). It is easy to show that

$$
\begin{bmatrix}
  x_1 & x_2 & \cdots & x_{N_p}
\end{bmatrix} = \begin{bmatrix} * \\
  0_{3 \times N_p} \end{bmatrix}, \text{ and } \begin{bmatrix}
  x_{N_p+1} & x_{N_p+2} & \cdots & x_{2N_p}
\end{bmatrix} = \begin{bmatrix} 0_{3 \times N_p} \\
  * \end{bmatrix}
$$

where $*$ denotes the nonzero entries. Thus, without any prior information, the existing methods are likely to cluster $\{x_j\}_{j=1}^{N_p}$ to a subspace, and $\{x_j\}_{j=N_p+1}^{2N_p}$ to the second subspace. On the other hand, by enforcing the prior information that $x_j$ and $x_{N_p+j}$, for $j = 1, \ldots, N_p$, belong to the same subspace, the proposed approach has an extremely low misclassification rate. The running time of each method, averaged over 60 instances, is summarized in Table 4.4.
CHAPTER 4. SUBSPACE CLUSTERING WITH PRIORS

Table 4.3: Average Performance for Planar Segmentation (%)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>GPCA</th>
<th>[10]</th>
<th>Algorithm 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merton I</td>
<td>50.00 (±0)</td>
<td>49.12 (±3.91)</td>
<td>4.29 (±3.45)</td>
</tr>
<tr>
<td>Merton II</td>
<td>46.04 (±10.99)</td>
<td>49.58 (±0.66)</td>
<td>2.33 (±2.85)</td>
</tr>
<tr>
<td>Wadham</td>
<td>45.69 (±7.02)</td>
<td>49.74 (±0.92)</td>
<td>1.33 (±2.19)</td>
</tr>
</tbody>
</table>

Table 4.4: Running Time (sec)

<table>
<thead>
<tr>
<th></th>
<th>GPCA</th>
<th>[10]</th>
<th>LRR</th>
<th>SSC</th>
<th>Algorithm 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1421</td>
<td>730.94</td>
<td>0.3553</td>
<td>0.1904</td>
<td>683.74</td>
</tr>
</tbody>
</table>

4.6 Conclusions

In this chapter we propose a new approach to the problem of identifying an arrangement of subspaces from noisy samples, potentially corrupted by outliers. The main idea is to recast the problem into a constrained polynomial optimization problem, which in turn can be solved by solving convex semi-definite programs resulted from its lowest order moments relaxation. A salient feature of the proposed approach is its ability to exploit available a-priori information on the percentage of outliers, relative number of points in each subspace and partial labelings. These advantages were illustrated with several examples comparing the performance of the proposed method vis-à-vis existing ones. The main drawback of the proposed method stems from the need to solve semi-definite programs. However, exploiting the underlying sparse structure of the problem allows for imposing the semi-definite constraints only on a collection of smaller matrices, leading to an algorithm whose complexity scales linearly with the number of data points. Research is currently underway seeking to explore the use of Frank-Wolfe algorithms [68] and second order cone programming [69] to further reduce the computational burden.
Chapter 5

Moments-Based Semi-Supervised Support Vector Machine

5.1 Motivation

SVM is a well-known two-class classification technique since it can give a global optimal solution to the supervised learning problem. However, in many real application scenarios, labeled data are scarce for manual labeling is often slow, expensive and error-prone, while abundant amounts of unlabeled data can be cheaply and automatically collected. From this perspective, research on learning with consideration of information given by unlabeled data is of great importance. Using both labeled and unlabeled data for the purpose of learning is called semi-supervised learning (see [70][71] for reference).

In recent literature on machine learning, semi-supervised SVMs (S$^3$VM) have been developed based on the idea proposed in [72]: to solve the standard SVM problem while treating the unknown labels as extra optimization variables. It combines the powerful regularization of SVMs with a direct implementation of cluster assumption, which states that the decision boundary should not cross high density regions. However, the nonconvexity of the feasible set makes the extension from standard SVM to S$^3$VM nontrivial. Therefore, substantial efforts have been directed towards developing computationally efficient algorithms, leading to a number of techniques based either on heuristics or relaxations, for instance, local combinatorial search [73], gradient descent [72], continuation techniques [74], convex-concave procedures [75], deterministic annealing [76], quasi newton [77] and branch and bound [78]. All these methods except branch-and-bound algorithm can
just guarantee a local optimal solution, and the combinatorial nature of brand and bound restricts the associated algorithm to be applicable to small-scale problems.

The other category of methods of semi-supervised learning is based on graph theory, such as spectral methods [79], random walks [80], and graph mincuts [81]. The major idea of graph-based methods is to design a classifying function which is sufficiently smooth with respect to the intrinsic structure revealed by both the labeled and unlabeled points [82], [83]. However, for datasets of large scale, the computation intensiveness caused by the high dimension of the graph-based regularization term makes these methods inefficient and even impractical. To alleviate this problem, some efforts have been distributed to approximating the graph regularization matrix by a lower rank matrix, such as the Nyström method [84], Prototype Vector Machine [85] and AchorGraph [86]. However, one common and nontrivial problem for graph based methods is how to pick the proper graphic model to measure the intrinsic structure of the data.

Compared to the subspace clustering problem in Chapter 4, instead of fitting data on hyperplanes (equality constraints), SVM is keeping data beyond the gap between the decision hyperplanes (inequality constraints). Motivated by the framework of polynomial optimization based subspace clustering, in this chapter, we formulate $S^3$VM problem as a constrained polynomial optimization problem by treating the unknown labels for the unlabeled data as binary variables. A moments-based SDP relaxation is proposed for which sufficient conditions to guarantee its equivalence to the the original nonconvex problem are derived. In order to tailor the proposed framework to large-scale problem, an computationally efficient algorithm based on the Augmented Lagrangian Method (ALM) [87] is proposed at the cost of sacrificing global optimality guarantees. For ease of reading, starting with analyzing $S^3$VM for cases where the samples are separable in finite dimensional feature space, we extend the results to kernel $S^3$VM to handle the cases where the feature space is too large or even infinite. Finally the theoretical results are verified by both synthetic data and public datasets which are popular in testing the performance of algorithms solving $S^3$VM.

5.2 Problem Statement

This chapter is aimed to exploit a moments-based algorithm for extending the standard SVM to $S^3$VM which utilizes the unlabeled data to improve the generalization performance of SVM in the case when the labeled data is insufficient. Formally, it is stated as

**Problem 5.** Given:
CHAPTER 5. MOMENTS-BASED SEMI-SUPERVISED SUPPORT VECTOR MACHINE

- a training set consisting of \( l \) labeled samples \( \{x_i, t_i\}_{i=1}^l, x_i \in \mathbb{R}^n, t_i \in \{-1, 1\} \);
- a working set consisting of \( u \) unlabeled samples \( \{x_i^*\}_{i=1}^u \).

Determine a decision hyperplane

\[
y(x) = w^T \phi(x) + b = 0
\] (5.1)

where \( \phi(x) \) denotes a known feature-space transformation, such that all the samples are far away from the decision hyperplane.

Mathematically, Problem 5 is to find the optimal solution to

\[
p^* = \min_{w,b,\xi,\xi^*,t^*} \frac{1}{2} ||w||_2^2 + C \sum_{i=1}^{l} \xi_i + C^* \sum_{i=1}^{u} \xi_i^* \] (5.2a)

subject to

\[
\forall_{i=1}^{l}: t_i[w^T \phi(x_i) + b] \geq 1 - \xi_i, \quad \xi_i \geq 0 \] (5.2b)

\[
\forall_{i=1}^{u}: t_i^*[w^T \phi(x_i^*) + b] \geq 1 - \xi_i^*, \quad \xi_i^* \geq 0 \] (5.2c)

\[
\forall_{i=1}^{u}: t_i^{*2} = 1 \] (5.2d)

\[
\frac{1}{u} \sum_{i=1}^{u} \max\{t_i^*, 0\} = r \] (5.2e)

where \( \xi = [\xi_1, \ldots, \xi_l]^T, \xi^* = [\xi_1^*, \ldots, \xi_u^*]^T, t^* = [t_1^*, \ldots, t_u^*]^T \). \( C > 0 \) and \( C^* > 0 \) are fixed parameters, reflecting our confidence in the labels of the labeled data and the low density separation assumption, respectively. \( t_i^* \) represents the labels of the unlabeled data. (5.2f) is the balancing constraint in case that all the data falls in one class. \( r \) is the ratio of positive samples among the unlabeled set.
CHAPTER 5. MOMENTS-BASED SEMI-SUPERVISED SUPPORT VECTOR MACHINE

5.3 Moments-Based S³VM

The nonconvexity in (5.2) follows the bilinear constraint (5.2d) and the integer constraint (5.2e). By the affine transformation \( \tilde{t}_i^* = \frac{t_i^{*+1}}{2} \in \{0, 1\} \), (5.2) is equivalently transformed into

\[
p^* = \min_{w, b, \xi, \xi^*, \tilde{t}^*} \frac{1}{2}||w||^2_2 + C \sum_{i=1}^{l} \xi_i + C^* \sum_{i=1}^{u} \xi_i^*
\]

subject to

\[
\forall t_i : t_i[w^T \phi(x_i) + b] \geq 1 - \xi_i, \xi_i \geq 0
\]

\[
\forall u_i = 1 : (2\tilde{t}_i^* - 1)[w^T \phi(x_i^*) + b] \geq 1 - \xi_i^*, \xi_i^* \geq 0
\]

\[
\forall u_i = 1 : \tilde{t}_i^* = \tilde{t}_i
\]

\[
\frac{1}{u} \sum_{i=1}^{u} \tilde{t}_i^* = r
\]

In this section we propose a novel convex optimization approach to (5.3) based on the results in moments theory introduced in Section 2.2.

5.3.1 A Sequence of Convex Relaxations

Since both of the objective function and the constraints in (5.3) are in the form of polynomials, (5.3) can be relaxed into a convergent sequence of convex semi-definite programming problem of the form:

\[
p^*_{m,N} = \min_{m_{2N}} c^T m_{2N}
\]

subject to

\[
M_N(m_{2N}) \succeq 0
\]

\[
L_{N-1}(g_k m_{2N}) \succeq 0, \forall g_k
\]

where \( m_{2N} \) denotes the moment sequence of \( v = \{w, b, \xi, \xi^*, \tilde{t}^*\} \) of order up to \( 2N \), \( c \) is the vector associated to the monomials in the polynomial describing the objective function in (5.3a), and \( M_N \) and \( L_{N-1} \) are the moment matrix and localizing matrix corresponding to constraints (5.3c)-(5.3f).
Considering the partitioning of the objective and the constraints in the form of

\[
p = \sum_{i=1}^{l+1} p_i \quad \text{with} \quad \begin{cases} p_i = \frac{1}{2(l+u)} \|w\|^2 + C\xi_i, \ i = 1, \ldots, l \\
p_{l+1} = \frac{u}{2(l+u)} \|w\|^2 + C^* \sum_{j=1}^{u} \xi^*_j \end{cases}
\]

\[
K = \cap_{i=1}^{l+1} K_i \quad \text{with} \quad \begin{cases} K_i = t_i [w^T\phi(x_i) + b] \geq 1 - \xi_i, \ \xi_i \geq 0, \ i = 1, \ldots, l \\
K_{l+1} = \{ \forall u \sum_{j=1}^{u} \tilde{t}^*_j = r \}
\end{cases}
\]

where the variables associated with each partition are in the sets of

\[
v_i = \begin{cases} \{w, b, \xi_i\}, \ i = 1, \ldots, l; \\
\{w, b, \xi^*_1, \ldots, \xi^*_u, \tilde{t}^*_1, \ldots, \tilde{t}^*_u\}, i = l + 1.
\end{cases}
\]

Obviously, the partitions of variables satisfy

\[
v_{j+1} \cap \bigcup_{k=1}^{j} v_k = \{w, b\} \subseteq v_1, \ \text{for} \ j = 1, \ldots, l
\]

which means (5.3) exhibits the running intersection property, hence, the moments-based relaxation (5.4) can be simplified as

\[
p^*_{m,N} = \min_{m_{1,2N}} \sum_{i=1}^{l+1} c_i^T m_{i,2N} \sum_{i=1}^{l+1} c_i^T m_{i,2N}
\]

subject to

\[
\forall i = 1, \ldots, u : M_{i,N}(m_{i,2N}) \succeq 0 \\
\forall i = 1, \ldots, u : L_{i,N-1}(g_i, m_{i,2N}) \succeq 0, \ \forall g_i \in K_i
\]

where \( m_{i,2N} \) and \( M_{i,N} \) denote the moments sequence of variables in \( I_i \) of order up to \( 2N \) and the associated moment matrix of order \( N \), \( c_i^T m_{i,2N} \) is the expected value of \( p_i \), and \( L_{i,2N} \) is the localizing matrix corresponding to the constraints in \( K_i \).

Following Theorem 3.4 in [25], \( p^*_N \) gives a lower bound of \( p^* \), and \( p^*_N \leq p^*_{N+1} \leq \cdots \leq p^* \). Theoretically, as \( N \uparrow \infty \), \( p^*_N \uparrow p^* \). So far, we have not found a finite \( N \) theoretically such that the relaxation at \( N \) is exact. In the sequel we will focus on developing a numerically effective algorithm.
CHAPTER 5. MOMENTS-BASED SEMI-SUPERVISED SUPPORT VECTOR MACHINE

for solving (5.3).

5.3.2 Reweighted \( \ell_1 \) Relaxation

Instead of considering a single moment matrix, \( M_N \) of size \( \binom{n + 1 + l + 2u + N}{N} \) in (5.4), in (5.8) we consider \( l \) moment matrices \( M_i,N \) of size \( \binom{n + 2 + N}{N} \) and one moment matrix \( M_{l+1,N} \) of size \( \binom{n + 1 + 2u + N}{N} \), reducing the computational complexity. However, if the unlabeled data is of large amount, then the computation burden of (5.8) is still heavy, especially when the order of moment matrix \( N \) increases. To obtain an optimizer such that the \( p^*_m,N \) approaches \( p^* \), instead of by increasing \( N \), we resort to finding solutions to

\[
p^*_{m,1} = \min_{m_{i,2}} \sum_{i=1}^{l+1} c_i^T m_{i,2}
\]

subject to

\[
\forall i = 1, \ldots, l+1: M_{i,1}(m_{i,2}) \geq 0
\]

\[
\forall i = 1, \ldots, l+1: L_{i,0}(g_{i_k}m_{i,2}) \geq 0, \forall g_{i_k} \in K_i
\]

\[
\forall i = 1, \ldots, l+1: \text{rank}\{M_{i,1}\} = 1
\]

(5.9)

Theorem 8. \( p^*_{m,1} = p^* \).

Proof. (1) Assuming that the optimizer to (5.3) is \( \{w_o, b_o, \xi_o, \xi^*_o, \hat{t}^*_o\} \), \( v_i = [w_o^T, b_o, \xi_o]^T \), for \( i = 1, \ldots, l \), and \( v_{l+1} = [w_o^T, b_o, \xi_o^T, \hat{t}_o^T]^T \), then the moments sequence of \( v_i \) of order up to 2 is a feasible solution to \( m_{i,2} \) in (5.9), and correspondingly \( M_{i,1} = v_i v_i^T \) for \( i = 1, \ldots, l + 1 \). Therefore, \( p^* \geq p^*_{m,1} \).

(2) Conversely, assuming that the optimizer to (5.9) is \( \{m_{i,2}\}_{i=1}^{l+1} \), then the first order moments of \( \{w, b, \xi, \xi^*, \hat{t}^*\} \) is a feasible solution to (5.3). Therefore, \( p^*_{m,1} \geq p^* \).

Combining (1) and (2), we have \( p^*_{m,1} = p^* \). \( \square \)

Although in the formulation (5.9) moments up to order 2 are needed, however, then rank-1 constraint makes it nontrivial. To circumvent this difficulty, next through the analysis of the structure of the moment matrix, we seek for a simpler way to solve (5.9).
CHAPTER 5. MOMENTS-BASED SEMI-SUPERVISED SUPPORT VECTOR MACHINE

Theorem 9. Assume that for the problem

\[
\tilde{p}_{m,1}^* = \min_{\tilde{m}_{i,2}} \sum_{i=1}^{l+u} \tilde{c}_i^T \tilde{m}_{i,2}
\]

subject to

\[
\begin{align*}
\forall_{i=1}^l: & \quad t_i [m(w^T)\phi(x_i) + m(b)] \geq 1 - m(\xi_i), \quad m(\xi_i) \geq 0 \\
\forall_{i=1}^u: & \quad 2[m(w\tilde{t}_i^*)^T \phi(x_i^*) + m(b\tilde{t}_i^*)] - [m(w^T)\phi(x_i) + m(b)] \geq 1 - m(\xi_i^*), \quad m(\xi_i^*) \geq 0 \\
\forall_{i=1}^l: & \quad M_{i,1}(\tilde{m}_{i,2}) \geq 0 \\
\forall_{i=1}^u: & \quad m(\tilde{t}_i^*) = m(\tilde{t}_i^*)^2 \\
\frac{1}{u} \sum_{i=1}^u m(\tilde{t}_i^*) = r
\end{align*}
\]

(5.10)

where \(\tilde{m}_{i,2}\) and \(\tilde{M}_{i,1}\) are the moments sequence of order up to 2 and the 1st moment of variables in \(\tilde{l}_i\) respectively, and \(\tilde{c}_i^T \tilde{m}_{i,2}\) is the expected value of \(\tilde{p}_i\), with

\[
\tilde{l}_i = \begin{cases} 
\{w, b, \xi_i\}, & \text{for } i = 1, \ldots, l; \\
\{w, b, \tilde{\xi}_{i-l}^*, \tilde{t}_{i-l}^*\}, & \text{for } i = l + 1, \ldots, l + u.
\end{cases}
\]

(5.11)

and

\[
\tilde{p}_i = \begin{cases} 
\frac{1}{2(l+u)}\|w\|_2^2 + C\xi_i, & \text{for } i = 1, \ldots, l; \\
\frac{1}{2(l+u)}\|w\|_2^2 + C^*\tilde{\xi}_{i-l}^*, & \text{for } i = l + 1, \ldots, l + u.
\end{cases}
\]

(5.12)

Then \(\tilde{p}_{m,1}^* = p_{m,1}^*\).

Proof. From the constraint \(m(\tilde{t}_i^*) = m(\tilde{t}_i^*)^2\) in (5.10), we have \(m(\tilde{t}_i^*) \in \{0, 1\}\).

For an unlabeled sample, \(x_i^*\), the 1st order moment matrix of \(\{w, b, \xi_i^*, \tilde{t}_i^*\}\) is

\[
\tilde{M}_{i+l,l} = \begin{bmatrix}
1 & m(w)^T & m(b) & m(\xi_i^*) & m(\tilde{t}_i^*) \\
m(w) & m(w^Tw) & m(wb) & m(w\xi_i^*) & m(w\tilde{t}_i^*) \\
m(b) & m(wb)^T & m(b^2) & m(b\xi_i^*) & m(b\tilde{t}_i^*) \\
m(\xi_i^*) & m(w\xi_i^*)^T & m(b\xi_i^*) & m(\xi_i^*2) & m(\xi_i^*\tilde{t}_i^*) \\
m(\tilde{t}_i^*) & m(w\tilde{t}_i^*)^T & m(b\tilde{t}_i^*) & m(\xi_i^*\tilde{t}_i^*) & m(\tilde{t}_i^*)
\end{bmatrix} \succeq 0
\]

(5.13)

Case 1: If \(m(\tilde{t}_i^*) = 0\), then following the Schur complements for nonstrict inequalities [88], \(m(w\tilde{t}_i^*) = 0\) and \(m(b\tilde{t}_i^*) = m(\xi_i^*\tilde{t}_i^*) = 0\) hold. Then the 1st order localizing matrix corresponding to (5.3d) is reduced to

\[
-m(w)^T \phi(x_i^*) + m(b) \geq 1 - m(\xi_i^*), \quad \text{when } m(\tilde{t}_i^*) = 0.
\]

(5.14)
CHAPTER 5. MOMENTS-BASED SEMI-SUPERVISED SUPPORT VECTOR MACHINE

**Case 2:** If \(m(\tilde{t}_m) = 1\), the positive semi-definiteness of the principal minor

\[
\begin{bmatrix}
1 & m(b) & m(\tilde{t}_m) \\
m(b) & m(b^2) & m(b\tilde{t}_m^*) \\
m(\tilde{t}_m) & m(b\tilde{t}_m^*) & m(\tilde{t}_m^*)
\end{bmatrix} =
\begin{bmatrix}
1 & m(b) & 1 \\
m(b) & m(b^2) & m(b\tilde{t}_m^*) \\
1 & m(b\tilde{t}_m^*) & 1
\end{bmatrix}
\] (5.15)

of \(M_{i+l,1}^*\) is equivalent to

\[
\begin{align*}
m(b^2) & \geq m(b)^2 \\
m(b^2) & \geq m(b\tilde{t}_m^*)^2 \\
m(b^2) - m(b\tilde{t}_m^*)^2 - m(b)[m(b) - m(b\tilde{t}_m^*)] + m(b)m(b\tilde{t}_m^*) - m(b^2) \\
&= -m(b\tilde{t}_m^*)^2 + 2m(b)m(b\tilde{t}_m^*) - m(b^2) \\
&= -[m(b) - m(b\tilde{t}_m^*)]^2 \geq 0
\end{align*}
\] (5.16)

from which we have \(m(b) = m(b\tilde{t}_m^*)\). Similarly, we also have \(m(\xi_u) = m(\tilde{t}_u)\) and \(m(w) = m(w\tilde{t}_m^*)\). Then the 1st order localizing matrix corresponding to (5.3d) is reduced to

\[
m(w)^T\phi(x_i^*) + m(b) \geq 1 - m(\xi_u), \text{ when } m(\tilde{t}_m^*) = 1.
\] (5.17)

In both cases, the 2nd order moments in (5.10) have been eliminated, and then \(m(||w||_2^2) = \sum_j m(w_j^2)\) is the only 2nd order moments in (5.10). From the positive semi-definiteness of the principal minor

\[
\begin{bmatrix}
1 & m(w_j) \\
m(w_j) & m(w_j^2)
\end{bmatrix}
\] (5.18)

we have \(m(w_j^2) \geq m(w_j)^2\), and in (5.10), \(m(w_j^2)\) is minimized, then for the optimal solution to (5.10), \(m(w_j^2) = m(w_j)^2\) holds.

Let \([m(w)^T, m(b), m(\xi_1), \ldots, m(\xi_l), m(\xi_u), m(\tilde{t}_1), \ldots, m(\tilde{t}_u)^T\] denote the first order moments among the optimal solution to (5.10), then the moments sequence of \(v_i = [m(w)^T, m(b), m(\xi_1)]^T\) for \(i = 1, \ldots, l\) and \(v_{l+1} = [m(w)^T, m(b), m(\xi_u), m(\tilde{t}_1), \ldots, m(\tilde{t}_u)]^T\), is a feasible solution to (5.9), therefore, \(\tilde{p}_{m,1} \geq p_{m,1}^*\) holds.
Conversely, in (5.9), necessary conditions to guarantee that
\[
\begin{align*}
\text{rank}\{M_{l+1,1}\} &= 1 \\
m(t^*_i) &= m(t^{**}_i), \text{ for } i = 1, \ldots, u
\end{align*}
\]
(5.19)
holds include
\[
\begin{align*}
\text{rank}\{\tilde{M}_{i+l,1}\} &= 1, \text{ for } i = 1, \ldots, u \\
m(t^*_i) &= m(t^{**}_i)^2, \text{ for } i = 1, \ldots, u
\end{align*}
\]
(5.20)
then the constraints in (5.10) are a subset of constraints in (5.9), therefore, \(\tilde{p}_{m,1} \leq p^*_{m,1}\) holds.

In summary, \(\tilde{p}_{m,1} = p^*_{m,1}\) holds.

For the unlabeled data, compared to the partition in (5.9), where one single moment matrix \(M_{l+1,1}\) of size \((n + 2u + 2) \times (n + 2u + 2)\) is used, in (5.10) \(u\) moment matrices \(\tilde{M}_{i+l,1}\) of size \((n + 4) \times (n + 4)\) are used. However, the difficulty in (5.10) lies in the nonconvex constraint \(m(t^*_i) = m(t^{**}_i)^2\), meaning \(m(t^*_i) \in \{0, 1\}\). We notice that the balancing constraint \(\frac{1}{u} \sum_{i=1}^{u} m(t^*_i) = r\) fixes the summation of \(m(t^*_i)\), then the binary solution is also the sparse solution. Combining the reweighted \(\ell_1\) relaxation technique [89], we give a relaxed algorithm for solving (5.10) as follows.

**Algorithm 5** Moments-Based \(S^3VM\)

1: Initialize: \(\delta > 0, \gamma > 0, k = 0, w_i^{(0)} = 1, i = 1, \ldots, u\);
2: repeat
3: Solving
\[
\{m(t^*_i)^{(k)}\} = \arg \min_{m_{i,2}} \sum_{i=1}^{l+u} \tilde{c}_i T \tilde{m}_{i,2} + \gamma \sum_{i=1}^{u} w_i^{(k)} m(t^*_i)
\]
subject to
\[
\begin{align*}
\forall t_i & : t_i \left[ m(w^T) \phi(x_i) + m(b) \right] \geq 1 - m(\xi_i), \quad m(\xi_i) \geq 0 \\
\forall t_i & : 2\left[ m(w t^*_i)^T \phi(x^*_i) + m(b t^*_i) \right] - \left[ m(w^T) \phi(x_i) + m(b) \right] \\
& \geq 1 - m(\xi^*_i), \quad m(\xi^*_i) \geq 0 \\
\forall \tilde{M}_{i+l,1} & : \tilde{M}_{i+l,1}(\tilde{m}_{i,2}) \succeq 0 \\
\frac{1}{u} \sum_{i=1}^{u} m(t^*_i) &= r
\end{align*}
\]
(5.21)
4: Updating
\[
\forall t_i : w_i^{(k+1)} = 1/(m(t^*_i)^{(k)} + \delta) \\
k = k + 1
\]
5: until Converge
In Algorithm 5, $\delta$ is a small positive real number to avoid the numerical issue when $w_i^{(k)}$ is too small, and $\gamma$ is a parameter balancing the two parts of the objective function.

### 5.4 ALM Algorithm for S$^3$VM

Although Algorithm 5 proposed above is efficient for problems of small scale, it could easily become computationally intractable for problems of large scale, due to the fact that the size of the moment matrix increases with the dimension of the feature, and that the number of SDP constraints increases with the number of samples. To tailor the proposed framework to large-scaled datasets, in this section, we develop an computationally more efficient ALM type algorithm for solving the nonconvex problem (5.2).

First of all, by introducing nonnegative slack variables $\{za_i\}_{i=1}^l$, $\{zb_i\}_{i=1}^u$, the inequality constraints (5.3c) and (5.3d) are equivalent to the equality constraints of the form

$$\forall_i^{l} : \quad t_i[w^T \phi(x_i) + b] - 1 + \xi_i - za_i = 0, \quad \xi_i \geq 0,$$

$$\forall_i^{u} : \quad (2\tilde{t}_i^* - 1)[w^T \phi(x_i^*) + b] - 1 + \xi_i^* - zb_i = 0, \quad \xi_i^* \geq 0.$$

Then the augmented lagrangian of Problem (5.2) is given by

$$L(w, b, \{\xi_i\}_{i=1}^l, \{\xi_i^*\}_{i=1}^u, \{za_i\}_{i=1}^l, \{zb_i\}_{i=1}^u; \{\alpha_i \geq 0\}_{i=1}^l, \{\beta_i \geq 0\}_{i=1}^u)$$

$$= \frac{1}{2}||w||^2 + C\sum_{i=1}^l \xi_i + C^*\sum_{i=1}^u \xi_i^* \cdots$$

$$- \sum_{i=1}^l \alpha_i\{t_i[w^T \phi(x_i) + b] - 1 + \xi_i - za_i\} \cdots$$

$$+ \frac{\mu}{2}||t_i[w^T \phi(x_i) + b] - 1 + \xi_i - za_i||^2 \cdots$$

$$- \sum_{i=1}^u \beta_i\{(2\tilde{t}_i^* - 1)[w^T \phi(x_i^*) + b] - 1 + \xi_i^* - zb_i\} \cdots$$

$$+ \frac{\mu}{2}||(2\tilde{t}_i^* - 1)[w^T \phi(x_i^*) + b] - 1 + \xi_i^* - zb_i||^2$$  (5.22)

subject to

$$\forall_i^{l} : \quad \xi_i \geq 0, \quad za_i \geq 0$$

$$\forall_i^{u} : \quad \xi_i^* \geq 0, \quad zb_i \geq 0$$

$$\forall_i^{u} : \quad \tilde{t}_i^2 = \tilde{t}_i^*$$

$$\frac{1}{u} \sum_{i=1}^u \tilde{t}_i^* = r$$

In (5.22), the objective function is differentiable with respect to all variables. The binary constraint on $\tilde{t}_i^*$ and the bilinear terms $\tilde{t}_i^* w$ and $\tilde{t}_i^* b$ make the problem nonconvex. Next we adopt the
CHAPTER 5. MOMENTS-BASED SEMI-SUPERVISED SUPPORT VECTOR MACHINE

ALM algorithm to determine the local optimal solution by alternatively updating primal variables and dual variables.

By adding the square terms of the dual variables $\frac{1}{\mu} \alpha_i^2, \frac{1}{\mu} \beta_i^2$ to $L$, which do not affect the minimizer of the primal variables, to complete the squares, leads to the more compact form:

$$L(w, b, \{\xi_i\}_{i=1}^l, \{\xi_i^*\}_{i=1}^u, \{z_{a_i}\}_{i=1}^l, \{z_{b_i}\}_{i=1}^u; \{\alpha_i\}_{i=1}^l, \{\beta_i\}_{i=1}^u)$$

$$= \frac{1}{2}||w||_2^2 + C \sum_{i=1}^l \xi_i + C^* \sum_{i=1}^u \xi_i^* \cdots$$

$$+ \sum_{i=1}^l \frac{\mu}{2} ||t_i[w^T \phi(x_i) + b] - 1 + \xi_i - za_i - \frac{\alpha_i}{\mu}||^2_2 \cdots$$

$$+ \sum_{i=1}^u \frac{\mu}{2} ||(2\tilde{t}_i - 1)[w^T \phi(x_i^*) + b] - 1 + \xi_i^* - zb_i - \frac{\beta_i}{\mu}||^2_2 \quad (5.23)$$

Firstly, the primal variables are updated in two groups: continuous variables $w, \{\xi_i\}_{i=1}^l$, $\{\xi_i^*\}_{i=1}^u, \{z_{a_i}\}_{i=1}^l, \{z_{b_i}\}_{i=1}^u$, and binary variables $\{\tilde{t}_i\}_{i=1}^l, \{\tilde{t}_i^*\}_{i=1}^u$. With fixed $\mu^{(k)}, \tilde{t}_i^{(k)}, \alpha_i^{(k)}, \beta_i^{(k)}$, Problem (5.22) is a linearly constrained quadratic programming problem, and $w^{(k+1)}, b^{(k+1)}, \{\xi_i^{(k+1)}\}_{i=1}^l, \{\xi_i^{(k+1)}\}_{i=1}^u, \{z_{a_i}^{(k+1)}\}_{i=1}^l, \{z_{b_i}^{(k+1)}\}_{i=1}^u, \{\alpha_i^{(k)}\}_{i=1}^l, \{\beta_i^{(k)}\}_{i=1}^u$, integrating the constraint $\tilde{t}_i^{2} = \tilde{t}_i^*$ into the objective function (5.23), the optimization problem with respect to $\{\tilde{t}_i^*\}_{i=1}^u$ is equivalent to

$$\text{minimize} \quad \sum_{i=1}^u c_i \tilde{t}_i^*$$

subject to $\frac{1}{\mu} \sum_{i=1}^u \tilde{t}_i^* = r, \quad \tilde{t}_i^* \in \{0, 1\} \quad (5.25)$

with $c_i = 4[w^{(k+1)T} \phi(x_i^*) + b^{(k+1)}](-1 + \xi_i^{(k+1)} - z_{a_i}^{(k+1)} - \frac{\alpha_i^{(k)}}{\mu})$. Obviously (5.25) is a linear integer programming, whose optimal solution is, $\tilde{t}_i^{(k+1)} = 1$ corresponding to the $u \times r$ smallest coefficients $c_i$, while others are 0.
CHAPTER 5. MOMENTS-BASED SEMI-SUPERVISED SUPPORT VECTOR MACHINE

Secondly, for fixed primal variables, Problem (5.22) is a linearly constrained quadratic programming problem with respect to the dual variables, and they can be updated by

$$\forall_{i=1}^{l} : a_i^{(k+1)} = a_i^{(k)} - \mu^{(k)} \left\{ t_i \left[ w^{(k+1)^T} \phi(x_i) + b^{(k+1)} \right] - 1 + \eta_i^{(k+1)} - z a_i^{(k+1)} \right\}$$

$$\forall_{i=1}^{u} : \beta_i^{(k+1)} = \beta_i^{(k)} \left\{ (2\tilde{r}_i^{(k+1)} - 1) \left[ w^{(k+1)^T} \phi(x_i^*) + b^{(k+1)} \right] - 1 + \xi_i^{(k+1)} - z b_i^{(k+1)} \right\}$$

(5.26)

In summary, the ALM algorithm for S3VM is summarized as Algorithm 6. Practically, if there is a large number of examples, in each iteration, solving the optimization problem (5.24) is time consuming, so we try to solve it in an inexact way, that is, updating $w^{(k+1)}$, $b^{(k+1)}$, $\eta_i^{(k+1)}$, $\xi_i^{(k+1)}$, $za^{(k+1)}$, $zb^{(k+1)}$ as in (5.27) at the cost of slower convergence.

$$P = I + \mu^{(k)} \left[ \sum_{i=1}^{l} \phi(x_i) \phi(x_i)^T + \sum_{i=1}^{u} \phi(x_i^*) \phi(x_i^*)^T \right]$$

$$w^{(k+1)} = -\mu^{(k)} P^{-1} \left\{ \sum_{i=1}^{l} t_i b^{(k)} - 1 + \xi_i^{(k)} - z a_i^{(k)} - \frac{a_i^{(k)}}{\mu^{(k)}} \right\} t_i \phi(x_i)$$

$$+ \sum_{i=1}^{u} \left\{ (2\tilde{r}_i^{(k)} - 1) b_i^{(k)} - 1 + \xi_i^{(k)} - z a_i^{(k)} - \frac{a_i^{(k)}}{\mu^{(k)}} \right\} (2\tilde{r}_i^{(k)} - 1) \phi(x_i^*) \right\}$$

$$b^{k+1} = -\frac{1}{l+u} \left\{ \sum_{i=1}^{l} t_i w^{(k+1)^T} \phi(x_i) - 1 + \xi_i^{(k)} - z a_i^{(k)} - \frac{a_i^{(k)}}{\mu^{(k)}} \right\}$$

$$+ \sum_{i=1}^{u} \left(2\tilde{r}_i^{(k)} - 1\right) \left\{ (2\tilde{r}_i^{(k)} - 1) w^{(k+1)^T} \phi(x_i^*) - 1 + \xi_i^{(k)} - z b_i^{(k)} - \frac{\beta_i^{(k)}}{\mu^{(k)}} \right\} \right\}$$

(5.27)

$$\xi_i^{(k+1)} = \mathcal{T} \left( 1 + za_i^{(k)} + \frac{a_i^{(k)}}{\mu^{(k)}} - \frac{C}{\mu^{(k)}} \right)$$

$$\xi_i^{* (k+1)} = \mathcal{T} \left( 1 + zb_i^{(k)} + \frac{\beta_i^{(k)}}{\mu^{(k)}} - \frac{C^*}{\mu^{(k)}} \right) - (2\tilde{r}_i^{(k)} - 1) \left[ w^{(k+1)^T} \phi(x_i^*) + b^{(k+1)} \right]$$

$$za_i^{(k+1)} = \mathcal{T} \left( t_i \left[ w^{(k+1)^T} \phi(x_i) + b^{(k+1)} \right] - 1 + \xi_i^{(k+1)} - \frac{a_i^{(k)}}{\mu^{(k)}} \right)$$

$$zb_i^{(k+1)} = \mathcal{T} \left( (2\tilde{r}_i^{(k)} - 1) \left[ w^{(k+1)^T} \phi(x_i^*) + b^{(k+1)} \right] - 1 + \xi_i^{* (k+1)} - \frac{\beta_i^{(k)}}{\mu^{(k)}} \right)$$

where $\mathcal{T}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$

5.5 Extension: Kernel SVM

All the previous analysis is based on the assumption that $\phi(x)$ is of finite dimension, but in some cases, finite dimensional $\phi(x)$ are still not separable with satisfying misclassification rate. To deal with data of this type, the "kernel trick" is proposed by first mapping the data $x$ into a Hilbert
CHAPTER 5. MOMENTS-BASED SEMI-SUPERVISED SUPPORT VECTOR MACHINE

Algorithm 6 ALM Algorithm for $S^3$VM

1. $\rho > 1, k = 0, \mu^{(0)} > 0$;
2. Initialize:
3. $w^{(0)}, b^{(0)}, \{\xi_i^{(0)}\}_{i=1}^l, \{\xi_i^{(0)}\}_{i=1}^u$, are derived from running regular SVM on the labeled data;
4. $\forall i = 1 : \tilde{t}_i^{(0)} = \frac{\text{sign}[(w^{(0)})^T \phi(x_i^*) + b^{(0)}] + 1}{2}$;
5. $\forall i = 1 : \xi_i^{(0)} = \max\{0, 1 - (2\tilde{t}_i^{(0)} - 1)[w^{(0)}]^T \phi(x_i^*) + b^{(0)}]\};$
6. $\forall i = 1 : z_\alpha_i^{(0)} = 0, \alpha_i^{(0)} = 0$;
7. $\forall i = 1 : z_\beta_i^{(0)} = 0, \beta_i^{(0)} = 0$.
8. \textbf{repeat}
9. 1.- The primal variables $w, b, \{\xi_i\}_{i=1}^l, \{\xi_i\}_{i=1}^u, \{za_i\}_{i=1}^l, \{zb_i\}_{i=1}^u$, are updated by solving Problem (5.24);
10. 2.- The primal variables $\{\tilde{t}_i\}_{i=1}^u$ are updated by solving (5.25);
11. 3.- The dual variables $\{\alpha_i\}_{i=1}^l, \{\beta_i\}_{i=1}^u$ are updated by (5.26);
12. 4.- $\mu^{(k+1)} = \rho \mu^{(k)}, k = k + 1$.
13. \textbf{until} converge

space through $\phi(x)$, which may be of infinite dimension. From the Representer Theorem [90], the decision hyperplane to Problem 5 can be expressed as

$$y(x) = w^T \phi(x) + b = \sum_{i=1}^{l+u} \gamma_i \phi(x_i), \phi(x) > + b \quad (5.28)$$

From which we know instead of the infinite dimensional $\phi(x)$, the inner product $\phi(x_i), \phi(x_j) >$ is needed. Thus, define a kernel matrix $K \in \mathbb{R}^{(l+u) \times (l+u)}$ by

$$K = \begin{bmatrix}
< \phi(x_1), \phi(x_1) > & \ldots & < \phi(x_1), \phi(x_l) > & < \phi(x_1), \phi(x_*^+) > & \ldots & < \phi(x_1), \phi(x_*^+) > \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
< \phi(x_l), \phi(x_1) > & \ldots & < \phi(x_l), \phi(x_l) > & < \phi(x_l), \phi(x_*^+) > & \ldots & < \phi(x_l), \phi(x_*^+) > \\
< \phi(x_*^+), \phi(x_1) > & \ldots & < \phi(x_*^+), \phi(x_l) > & < \phi(x_*^+), \phi(x_*^+) > & \ldots & < \phi(x_*^+), \phi(x_*^+) > \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
< \phi(x_1), \phi(x_*^+) > & \ldots & < \phi(x_1), \phi(x_*^+) > & < \phi(x_1), \phi(x_*^+) > & \ldots & < \phi(x_1), \phi(x_*^+) > \\
\end{bmatrix}$$

Substituting (5.28) for $y(x)$ in (5.3), and letting $\Gamma = \text{diag} \{\gamma_1, \gamma_2, \ldots, \gamma_{l+u}\}$, we obtain
the kernel $S^3$VM in the form of

$$p^* = \underset{\Gamma, b, \xi, \tilde{t}}{\text{minimize}} \quad \frac{1}{2} \Gamma^T \mathbf{K} \Gamma + C \sum_{i=1}^{l} \xi_i + C^* \sum_{i=1}^{u} \xi_i^*$$

subject to

$$\forall l_i = 1 : \quad t_i [\Gamma^T \mathbf{K}(:,i) + b] \geq 1 - \xi_i, \quad \xi_i \geq 0$$
$$\forall u_i = 1 : \quad (2 \tilde{t}_i^* - 1) [\Gamma^T \mathbf{K}(:,l + i) + b] \geq 1 - \xi_i^*, \quad \xi_i^* \geq 0$$
$$\forall u_i = 1 : \quad \tilde{t}_i^* \geq \tilde{t}_i$$
$$\frac{1}{u} \sum_{i=1}^{u} \tilde{t}_i^* = r$$

(5.29)

The algorithms proposed in Section 5.3 and Section 5.4 can be extended to solve the nonlinear $S^3$VM with minor modification. However, one of the main drawbacks of kernel-based classifier is the curse of dimensionality, that is, the dimension of $\Gamma$ is equal to the number of the examples $l + u$. To overcome this weakness, several methods, such as Nyström method \[84] and the method of randomly picking landmarks \[91], have been proposed to replace the original kernel matrix $\mathbf{K}$ by a reduced-rank approximation $\tilde{\mathbf{K}}$ in the form of

$$\tilde{\mathbf{K}} = \mathbf{Q} \mathbf{K}_m \mathbf{Q}^T$$

(5.30)

where $\mathbf{Q} \in \mathcal{R}^{(l+u) \times m}$, $\mathbf{K}_m \in \mathcal{R}^{m \times m}$, and $m \ll l + u$. For a rank-$m$ kernel matrix $\mathbf{K}$, we can pick $\mathbf{Q} = \mathbf{U}_1$, and $\mathbf{K}_m = \Sigma_1$, where $\mathbf{U}_1$ denotes the singular vectors corresponding to the nonzero singular values of $\mathbf{K}$ and $\mathbf{K} = \mathbf{U}_1 \Sigma_1 \mathbf{U}_1^T$. Using this trick, by introducing new variable $\theta = \mathbf{Q}^T \Gamma$, the problem of kernel $S^3$VM is formulated as

$$p^* = \underset{\theta, b, \{\xi_i\}_{i=1}^{l}, \{\xi_i^*\}_{i=1}^{u}, \{\tilde{t}_i^*\}_{i=1}^{u}}{\text{minimize}} \quad \frac{1}{2} \theta^T \mathbf{K}_m \theta + C \sum_{i=1}^{l} \xi_i + C^* \sum_{i=1}^{u} \xi_i^*$$

subject to

$$\forall l_i = 1 : \quad t_i [\mathbf{Q}(i,:) \mathbf{K}_m \theta + b] \geq 1 - \xi_i, \quad \xi_i \geq 0$$
$$\forall u_i = 1 : \quad (2 \tilde{t}_i^* - 1) [\mathbf{Q}(l + i,:) \mathbf{K}_m \theta + b] \geq 1 - \xi_i^*, \quad \xi_i^* \geq 0$$
$$\forall u_i = 1 : \quad \tilde{t}_i^* \geq \tilde{t}_i$$
$$\frac{1}{u} \sum_{i=1}^{u} \tilde{t}_i^* = r$$

(5.31)

where the number of variables is reduced by $(l + u - m)$. Hitherto, we can use Algorithm 5 and Algorithm 6 to solve Problem (5.31).
5.6 Experiments

In this section, we illustrate several experiments to evaluate the quality of the proposed algorithms. Additional comparisons are made based on results reported on [92]. In this part, we apply Algorithm 5 and Algorithm 6 on several well-known datasets described in Table 5.1 and compare their performances with those of some typical existing algorithms.

Table 5.1: Datasets used in Section 5.6, \(c\): the number of classes, \(d\): the dimensionality of the data, \(l\): the number of labeled examples, and \(n\): the number of all the data in the dataset.

<table>
<thead>
<tr>
<th>DATA SET</th>
<th>(c)</th>
<th>(d)</th>
<th>(l)</th>
<th>(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G50C</td>
<td>2</td>
<td>50</td>
<td>50</td>
<td>550</td>
</tr>
<tr>
<td>TEXT1</td>
<td>2</td>
<td>7511</td>
<td>50</td>
<td>1946</td>
</tr>
<tr>
<td>WEBKB (PAGE)</td>
<td>2</td>
<td>3000</td>
<td>12</td>
<td>1051</td>
</tr>
<tr>
<td>WEBKB (LINK)</td>
<td>2</td>
<td>1840</td>
<td>12</td>
<td>1051</td>
</tr>
<tr>
<td>WEBKB (PAGE+LINK)</td>
<td>2</td>
<td>4840</td>
<td>12</td>
<td>1051</td>
</tr>
</tbody>
</table>

Protocols: G50C are generated from two standard normal multi-variate Gaussians and the means are located in 50-dimensional space such that the Bayes error is 5%. TEXT1 are the classes mac and mswindows of the Newsgroup20 dataset preprocessed as in [80]. The WebKB dataset is a subset of web documents of the computer science department of four universities made available by the CMU text-learning group. Each document belongs to either category course or category noncourse.

Results are summarized in Table 5.2. SVM\(_n\) (SVM\(_l\)) denotes the supervised SVM using all the data (only the labeled data), which are implemented by the software libsvm [93]. ALG 6 denotes the algorithm we proposed, and LapSVM denotes the graph based method introduced in [83]. From Table 5.2 we can tell the proposed ALM based algorithm shows competing performance to the state-of-the-art methods although only local optimality can be guaranteed.

Table 5.2: Experimental Results in Section 5.6 in ACCURACY (%): mean (std)

<table>
<thead>
<tr>
<th>DATA SET</th>
<th>G50C</th>
<th>TEXT1</th>
<th>PAGE</th>
<th>LINK</th>
<th>PAGE LINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM(_n)</td>
<td>100 (0.0)</td>
<td>100 (0.0)</td>
<td>99.52 (0.0)</td>
<td>97.8 (0.0)</td>
<td>100 (0.0)</td>
</tr>
<tr>
<td>SVM(_l)</td>
<td>86.6 (2.8)</td>
<td>81.2 (5.6)</td>
<td>80.1 (7.3)</td>
<td>76.3 (11.4)</td>
<td>80.7 (8.4)</td>
</tr>
<tr>
<td>ALGORITHM 6</td>
<td>95.0 (0.5)</td>
<td>94.1 (0.9)</td>
<td>94.9 (0.7)</td>
<td>94.0 (0.7)</td>
<td>97.1 (0.5)</td>
</tr>
<tr>
<td>LAP SVM</td>
<td>88.0 (2.2)</td>
<td>89.6 (1.1)</td>
<td>89.1 (1.2)</td>
<td>82.8 (9.0)</td>
<td>93.6 (0.9)</td>
</tr>
</tbody>
</table>
5.7 Conclusions

In this chapter, we proposed two methods of solving the nonconvex problems in $S^3$VM. Theoretically by moments based algorithm, the global optimal solution can be obtained. Due to the computation complexity resulted from the increasing order of moments matrices, taking advantage of the bijection between $\{+1, -1\}$ and $\{0, 1\}$, the reweighed $l_1$-norm heuristic is adopted to realize faster convergence to a reasonable local optimum solution. Comparing to the relaxation based on the moments theory, ALM based relaxation is much easier to be implemented, and it is one way to achieve a local optimum. Experiments show that ALM based algorithm present a competitive performance in solving $S^3$VM problems.

Research is currently under way showing theoretical support to the conjecture that the moments based algorithm (without further relaxation by reweighed $l_1$-norm heuristic) achieves the global optimum of the original nonconvex problem at finite relaxation order $N$, hopefully $N = 2$ as shown by some numerical experiments we have done so far.
Chapter 6

EIV Identification of LTI Systems

6.1 Motivation

In previous chapters, we have considered the model fitting problem for static data, where samples are assumed independent from each other. In spite of the successful application of the modeling technique for static data, this assumption is inappropriate for many other cases where the data are generated by a temporal sequence, for instance, the input/output signals generated by a running physical system. Under these circumstances, samples for model fitting are not independent from each other any more, and hence we call them dynamical data. The interdependence between samples makes the model fitting more difficult especially when the data is corrupted by measurement noise. In next few chapters, we will consider the model fitting problem for dynamical data. Firstly in Chapter 6 we will consider the simplest case, fitting a single linear time invariant (LTI) model, and next in Chapter 7 we will consider the problem of fitting multiple LTI models, named switched ARX model.

LTI systems identification has been the subject of extensive studies over the last decades, leading to a rich variety of methodologies either in time or in the frequency domain presented in [94], [95], [96] and references therein. Under the assumption that the structure of the model is known, if the input/output signals are exactly measurable, the identification problem is simply to find a solution to a set of linear equations. However, in many scenarios of practical interest, the measurements of the input/output signals are often contaminated by noise, making the identification problem nontrivial. The difficulty of the problem differs in terms of the prior assumptions on the noise and the ways noise enters the model.

One branch of research is based on the assumption that the noise is originated from
a stochastic process, resulting in many algorithms based on least squares methods [94]. While stochastic noise description is a well-established notion and has been extraordinarily effective in modeling many natural phenomena, deterministic, bounded noise description offers a logically sound and practically meaningful alternative, especially in situations where the assumption of randomness may be questionable. For instance, measurement errors may be caused by inaccuracy of measuring devices or defects in certain system parts or components. Such quantities are usually specified and hence known to be within certain range of accuracy. Thus, the other branch of research is based on the assumption of deterministic, bounded noise, leading to the development of set-membership techniques, for which there is a consistent set for the model parameter [95].

Roughly speaking, there are two kinds of model descriptions according to the ways noise enters the system. One is the output error or the equation error structure, where a single error term, named process noise, is added to the difference equation taking into account all possible sources of uncertainty affecting the output. The other is referred to as the Error-In-Variables (EIV) structure, where both input and output signals are affected by measurement noise. The main contributions on the identification of LTI systems in EIV structure where the measurements are corrupted by stochastic noise can be found in the survey paper [96]. On the other hand, for models with EIV structure, one research direction is to determine the parameter uncertainty intervals (PUI). As pointed out in [97], the exact feasible set of the model parameters for the EIV model structure is nonconvex, the use of either polytopic or ellipsoidal outer approximation is suggested. For instance, in [98], a outer-bounding interval approximation is developed based on results in [99], which is very conservative since it ignores the dynamic correlation between different regression equations. To circumvent this weakness, in [100] a tighter outer-bounding interval is computed resorting to the moments based polynomial optimization techniques [101], [102]. Contrary to finding a feasible set for the model parameter, the other research direction is to find a single model consistent with the given experimental data represented by [103] and [104], both of which are looking for a low-rank regression matrix by solving semi-definite programs (SDP), whose nonzero eigenvector associated with the zero eigenvalue is the model parameter. However, with the increasing of the measurement sequence, the size of SDP constraint augments, increasing the computational complexity. To circumvent this weakness, in this chapter we propose a new framework of identifying a model consistent with the noisy temporal input/output sequences via moments based polynomial optimization techniques. By exploiting the sparse structure exhibited by the problem, for a fixed model, the size of the SDP constraint in our formulation is constant rather than changing as the length of the temporal sequence changes.
6.2 Problem Statement

Consider the single-input single-output (SISO) linear time invariant (LTI) system shown in Figure 6.1 where the Auto-Regressive (AR) model, the Auto-Regressive with eXogenuous input (ARX) model and the Error-In-Variables (EIV) model are given by the following difference equations respectively.

\[ \begin{align*}
\xi_t + a_1 \xi_{t-1} + \cdots + a_n \xi_{t-n} &= b_0 x_t + b_1 x_{t-1} + \cdots + b_n x_{t-n_b} \\
\xi_t + a_1 \xi_{t-1} + \cdots + a_n \xi_{t-n} &= b_0 x_t + b_1 x_{t-1} + \cdots + b_n x_{t-n_b} + e_t \\
\xi_t + a_1 \xi_{t-1} + \cdots + a_n \xi_{t-n} &= b_0 x_t + b_1 x_{t-1} + \cdots + b_n x_{t-n_b} + e_t \\
u_t &= x_t + w_t \\
y_t &= \xi_t + \eta_t
\end{align*} \]

where \( \bar{r} = [-b_{n_b}, -b_{n_b-1}, \ldots, -b_0, a_n, \ldots, a_1, a_0]^T \) is the parameter defining the models. The complexity of identifying \( \bar{r} \) depends on different setups of the models.

- **AR model.** Assume the clean input/output data \( \{x_t, \xi_t\}_{t=0}^{T-1} \) are available, then the model parameter \( \bar{r} \) is determined by the equations

\[ X \begin{bmatrix} \bar{r} \\ 1 \end{bmatrix} = 0, \quad \text{where} \quad X = \begin{bmatrix} H_x \\ H_\xi \end{bmatrix}^T. \quad (6.1) \]

- **ARX model.** In ARX models, the system is perturbed by process noise \( e_t \). Assume \( e_t \) is \( \ell_\infty \)-norm bounded, i.e., \( \|e_t\|_\infty \leq \epsilon \), and that the input/output data \( \{x_t, y_t\}_{t=0}^{T-1} \) are available,
then the feasible set of $\bar{r}$ is

\[
\left\| X \begin{bmatrix} \bar{r} \\ 1 \end{bmatrix} \right\|_\infty \leq \epsilon, \text{ where } X \doteq \begin{bmatrix} H_x \\ H_y \end{bmatrix}^T.
\] (6.2)

- **EIV model.** In EIV models, the input $x_t$ and the output $\xi_t$ to the AR model is not available, but their measurements $\{u_t, y_t\}_{t=0}^{T-1}$ are available, and the measurement noise $w_t, \eta_t$ are unknown but $\ell_\infty$-norm bounded, i.e., $\|w_t\|_\infty \leq \epsilon_1$ and $\|\eta_t\|_\infty \leq \epsilon_2$, then the feasible set of $\bar{r}$ is

\[
\left\{ \begin{bmatrix} H_u - H_w \\ H_y - H_\eta \end{bmatrix}^T \begin{bmatrix} \bar{r} \\ 1 \end{bmatrix} = 0 \\
\|w_{0:T-1}\|_\infty \leq \epsilon_1, \|\eta_{0:T-1}\|_\infty \leq \epsilon_2 \right\}
\] (6.3a)

\[
\begin{bmatrix} \eta_0 & \eta_1 & \cdots & \eta_{T-n_a-1} \\ \eta_1 & \eta_2 & \cdots & \eta_{T-n_a} \\ \vdots & \vdots & \ddots & \vdots \\ \eta_{n_a} & \eta_{n_a+1} & \cdots & \eta_{T-1} \end{bmatrix}, \quad \begin{bmatrix} w_{n_a-n_b} & w_{n_a-n_b+1} & \cdots & w_{T-n_b-1} \\ w_{n_a-n_b+1} & w_{n_a-n_b+2} & \cdots & w_{T-n_b} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_a} & w_{n_a+1} & \cdots & w_{T-1} \end{bmatrix}
\]

Obviously, in the case of AR model or ARX model, if we consider each row of $X$ as a sample in $\mathbb{R}^n$, with $n = 2 + n_a + n_b$, then (6.1) and (6.2) are nothing but the linear model fitting problem discussed in Chapter 3. On the other hand, in the case of EIV model, due to the bilinear equality (6.3a), the feasible set of $\bar{r}$ is nonconvex but semi-algebraic. In this chapter, we will focus on the identification of the EIV models, which is formally stated as:

**Problem 6.** Given

- the measurements of the input and output sequence, $\{u_t, y_t\}_{t=0}^{T-1}$, generated by the EIV model as shown in Figure 6.1;

- the bound on the measurement noise, $\epsilon_1$ and $\epsilon_2$.

Determine the model parameter $\bar{r}$ such that (6.3) holds.
CHAPTER 6. EIV IDENTIFICATION OF LTI SYSTEMS

6.3 A Convex Approach to EIV Identification of LTI Systems

6.3.1 A Constrained POP Formulation

Let \( r = \begin{bmatrix} r^T & 1 \end{bmatrix}^T \), then Problem 6 can be solved by finding a feasible solution to

\[
\begin{bmatrix}
H_u^T - H_w^T & H_y^T - H_\xi^T
\end{bmatrix} r = 0
\tag{6.4a}
\]

\[
r^T r = 1
\tag{6.4b}
\]

\[
\|w_{0:T-1}\|_\infty \leq \epsilon_1
\tag{6.4c}
\]

\[
\|\eta_{0:T-1}\|_\infty \leq \epsilon_2
\tag{6.4d}
\]

of which the nonconvexity comes from the bilinear equalities (6.4a) and (6.4b). Because of the structure of the Hankel matrices \( H_w \) and \( H_\eta \), the measurement noise \( w_t \) (or \( \eta_t \)) enters the \( n_a + 1 \) (or \( n_b + 1 \)) regression equations in (6.4a).

6.3.2 Gramian Based Rank Minimization

As pointed out in [103] and [104], (6.4) is equivalent to finding noise sequences \( w_{0:T-1} \) and \( \eta_{0:T-1} \) in their feasible set defined by (6.4c)-(6.4d) such that the matrix \( H = \begin{bmatrix} H_u^T - H_w^T & H_y^T - H_\xi^T \end{bmatrix} \) is rank deficient. Both [103] and [104] solve the convex problem iteratively involving the positive semi-definiteness constraint of the matrix \( H \) resulting from the the rank minimization algorithm proposed in [50]. However, the size of \( H \in \mathbb{R}^{n \times (T-n_a)} \) increases with the length of the temporal sequence \( T \).

Lemma 3. Given a matrix \( A \in \mathbb{R}^{m \times n} \), Let \( B = AA^T \) and \( C = A^T A \in \mathbb{R}^{n \times n} \), then \( \text{rank}\{A\} = \text{rank}\{B\} = \text{rank}\{C\} = r \), \( r \leq m \), \( r \leq n \).

Proof. Assume the singular vector decomposition (SVD) of \( A \) is, \( A = USV^T \), with \( UU^T = VV^T = I \), and \( \sigma_i \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \) are the nonzero singular values of \( A \), then from \( B = AA^T = USS^T U^T = UAU^T \), where \( A \in \mathbb{R}^{m \times m} \) and \( \Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_m\} \), with \( \lambda_i = \sigma_i^2 > 0 \) for \( i = 1, \ldots, r \) and \( \lambda_i = 0 \) for \( i = r + 1, \ldots, m \), thus, \( B \) has \( r \) nonzero eigenvalues, i.e. \( \text{rank}\{B\} = r \). In the similar way we could show \( \text{rank}\{C\} = r \). \( \square \)

From Lemma 3 we know the Gramian matrix \( G \doteq HH^T \in \mathbb{R}^{n \times n} \) has the same rank as \( H \). Thus, by finding a rank deficient \( G \), we can obtain a rank deficient \( H \). Besides the size of \( G \) is
independent on the length of the sequence, in the sequel we will show the rank minimization of \( G \) can be reduced to a constrained quadratical program.

Based on [50], the rank minimization problem can be relaxed into solving the following convex SDP iteratively:

\[
\begin{align*}
\text{minimize} & \quad \eta_{0:T-1}, w_{0:T-1} \text{ subject to } \|w_{0:T-1}\|_\infty \leq \epsilon_1, \|\eta_{0:T}\|_\infty \leq \epsilon_2 \\
\text{subject to} & \quad \text{Trace}(W^{(k)}G) \\
\end{align*}
\]

(6.5)

where the weight matrix \( W \) is initialized by \( W^{(0)} = I \), and updated by

\[
W^{(k+1)} = (G^{(k)} + \sigma_n(G^{(k)}))^{-1},
\]

(6.6)

where \( \sigma_n(\bullet) \) denotes the smallest singular value of \( \bullet \).

**Theorem 10.** Problem (6.5) can be reformulated as a Constrained Quadratic Programming.

**Proof.** Since \( G \succeq 0 \), the way updating \( W^{(k)} \) guarantees that \( W^{(k)} \succeq 0 \), for any \( k \), thus, it can be factorized as \( W^{(k)} = M^{(k)T}M^{(k)} \). One way of factorizing is based on SVD of \( W^{(k)} \), supposing \( W^{(k)} = USU^T \), where \( S = \text{diag}(s_1^2, s_2^2, \ldots, s_n^2) \), \( s_i \geq 0 \), for \( \forall i = 1, \ldots, n \), \( U = [u_1 \ u_2 \ \ldots \ u_n] \), \( u_i^T u_j = \delta(i,j) \), for \( \forall i, j = 1, \ldots, n \), then

\[
M^{(k)} = S^{-1/2}U^T = [s_1 u_1 \ s_2 u_2 \ \ldots \ s_n u_n]^T \in \mathbb{R}^{n \times n}
\]

(6.7)

Then the objective function in (6.5) can be

\[
\begin{align*}
\text{Trace}(W^{(k)}G) \\
= \text{Trace}(M^{(k)T}M^{(k)}HH^T) \\
= \text{Trace}(M^{(k)T}HH^T M^{(k)T}) \\
= \text{Trace}( \begin{bmatrix}
    s_1 u_1^T H \\
    s_2 u_2^T H \\
    \vdots \\
    s_n u_n^T H
\end{bmatrix}
\begin{bmatrix}
    s_1 H^T u_1 & s_2 H^T u_2 & \ldots & s_n H^T u_n
\end{bmatrix}) \\
= \sum_{j=1}^n s_j^2 \|H^T u_j\|_2^2
\end{align*}
\]

(6.8)
H^T u_j \in \mathbb{R}^{T-n_a} is an affine function of the unknown variable \( v = \begin{bmatrix} w_{0:T-1} & \eta_{0:T-1} \end{bmatrix}^T \in \mathbb{R}^{2T} \), and it can be written as
\[
H^T u_j = P_j v + r_j, \quad \forall j = 1, \ldots, n,
\]
(6.9)
where \( P_j \in \mathbb{R}^{(T-n_a) \times 2T}, r_j \in \mathbb{R}^{T-n_a} \) are constant. Thus substituting (6.9) into (6.8), we can obtain that
\[
\text{Trace}(W^{(k)}G) = \sum_{j=1}^n s_j^2 \| P_j v + r_j \|_2^2 = v^T Qv + 2\ell^T v + r
\]
(6.10)
where
\[
\begin{cases}
Q = \sum_{j=1}^n s_j^2 P_j^T P_j \\
\ell = \sum_{j=1}^n s_j^2 P_j^T r_j \\
r = \sum_{j=1}^n s_j^2 r_j^T r_j
\end{cases}
\]
(6.11)
Obviously, \( Q \succeq 0 \) holds, therefore, we can find a low rank G by solving convex constrained quadratic programming, leading to Algorithm 7.

Algorithm 7 Rank Minimization by Constrained QPs
1: Initialize \( k = 0, W^{(k)} = M^{(k)} = I \);
2: repeat
3: Calculate \( Q, \ell, r \) by (6.11);
4: Solve the constrained QP
\[
\begin{align*}
\mathbf{v}^{(k)} &= \arg \min_{v} v^T Qv + 2\ell^T v + r \\
\text{subject to} \quad &\|w_{0:T-1}\|_\infty \leq \epsilon_1, \|\eta_{0:T}\|_\infty \leq \epsilon_2
\end{align*}
\]
(6.12)
5: Build the Gramian Matrix \( G^{(k)} \) by \( \mathbf{v}^{(k)} \);
6: Update \( W^{(k+1)} \) and \( M^{(k+1)} \) by (6.6) and (6.7);
7: \( k = k + 1 \);
8: until \( \sigma_n(G^{(k)}) \to 0 \).

6.3.3 Moments Based Convex Relaxation

Although the Gramian based approach introduced above is easy to implement, however, comparing to seeking for a rank deficient Hankel matrix H, looking for a rank deficient Gramian matrix \( G = HH^T \) may become more difficult due to the numerical issue that the condition number of \( G \) is much larger than that of \( H \) since the singular value of \( H \) is the square root of that of \( G \). On the other hand, as we notice, in spite of its nonconvexity, (6.4) is a semi-algebraic set determined
by a set of quadratic (in)equalities, an alternative approach to solve (6.4) is proposed by using the moments based polynomial optimization techniques to relax it into a sequence of convex SDP problems of the form:

$$p_N^* = \min_{m_{2N}} 0$$
subject to
$$M_N(m_{2N}) \succeq 0$$
$$L_{k,N-1}(g_k m_{2N}) \succeq 0, \forall k = 1 \ldots d$$

(6.13)

where \(m_{2N}\) denotes the moment sequence of the variables \(v = \{r^T, w_0, w_1, \ldots, w_{T-1}, \eta_0, \eta_1, \ldots, \eta_{T-1}\}\) of order up to \(2N\), \(M_N\) and \(L_{k,N-1}\) denote the \(N\)-th order moment matrix and \((N - 1)\)-th order localizing matrix associated with the constraint \(g_k\) that defines the feasibility set in (6.4), respectively. \(d\) represents the number of constraints in (6.4).

By partitioning \(v\) into \(T - n_a + 1\) sets

\[v_t = \begin{cases} \{r, w_t, \eta_t\}, & w_t = [w_{t-n_a}, \ldots, w_t]^T, \eta_t = [\eta_{t-n_a}, \ldots, \eta_t]^T, \text{ for } \forall t = n_a, \ldots, T - 1 \\ \{r\}, & \text{for } t = T \end{cases} \]

(6.14)

and partitioning the feasible set \(K\) in (6.4) into \(T - n_a + 1\) sets,

\[K = \bigcap_{t=n_a}^T K_t, \quad \forall t = n_a + 1 \ldots T - 1 : \begin{cases} u_{t-n_a:t} - w_{t-n_a:t}, y_{t-n_a:t} - \eta_{t-n_a:t} & r = 0 \\ \|w_{t-n_a:t}\|_\infty \leq \epsilon_1 \\ \|\eta_{t-n_a:t}\|_\infty \leq \epsilon_2 \\ r^T r = 1 \end{cases} \]

we can observe that for \(t = n_a, \ldots, T\), each set \(K_t\) consists constraints associated with only variables in \(v_t\), and the partitions of variables satisfy

\[\bigcup_{t=n_a}^T v_t = v\]
\[v_t \cap (\bigcup_{t'=1}^{t-1} v_{t'}) = \{r, w_{t-n_a:t-1}, \eta_{t-n_a:t-1}\} \subseteq v_{t-1}, \text{ for } \forall t = n_a + 1, \ldots, T - 1\]
\[v_t \cap (\bigcup_{t'=1}^{t-1} v_{t'}) = \{r\} \subseteq v_{t-1}, \text{ for } t = T.\]

Thus the problem in (6.4) exhibits the running intersection property, and according to Section 2.2.3
CHAPTER 6. EIV IDENTIFICATION OF LTI SYSTEMS

the nonconvex problem can be relaxed into a sequence of reduced SDP problems of the form

\[
p^*_{m,N} = \min_{m_{t,2N}} 0
\]

subject to

\[
\begin{align*}
\forall T_{t=n_a} : & \quad M_{t,N}(m_{t,2N}) \succeq 0 \\
\forall T_{t=n_a} : & \quad L_{t,N-1}(g_{tk} m_{t,2N}) \succeq 0, \quad \forall g_{tk} \in K_t
\end{align*}
\]

(6.15)

where \( N \) is the relaxation order, \( m_{t,2N} \) denotes the moments sequence of variables in \( \nu_t \) of order up to \( 2N \), \( M_{t,N} \) and \( L_{t,N-1} \) represent the truncated moment matrix and localizing matrix of \( \nu_t \).

Comparing to the SDP relaxation in (6.13), for the same relaxation order \( N \), instead of considering a single moment matrix of of size \( \frac{n + 2T + N}{N} \) in (6.13), in (6.15), \( T - n_a \) moment matrices of size \( \frac{2n + N}{N} \) and one moment matrix of size \( \frac{n + N}{N} \) are considered, reducing the computational complexity dramatically especially for cases where the length of the temporal sequence \( T \) is much larger than the order of the model \( n_a \) (\( n = 2(n_a + 1) \)).

**Remark 2.** From Theorem 3.1 in [102] we know for the reduced SDP problem (6.15), as the relaxation order \( N \) increases, the feasible set defined by the convex constraints in (6.15) approaches the original nonconvex set in (6.4) from outside. However, on the one hand, with \( N \) increasing, the exponential increasing of the size of the moment matrix would make (6.15) computational intractable. On the other hand, we need to extract the feasible solution to (6.4) from the moment matrix which is nontrivial when the moment matrix is of high rank. Surprisingly we find that by enforcing a low rank constraint on a small matrix, by solving only the first order moment relaxation, we can extract a feasible solution to the original problem (6.4).

**Lemma 4.** For a \( 3 \times 3 \) symmetric matrix \( M \) denoted by

\[
M = \begin{bmatrix}
M_{11} & M_{12} \\
M_{21} & M_{22}
\end{bmatrix} = \begin{bmatrix}
1 & m_1 & m_2 \\
m_1 & m_{11} & m_{12} \\
m_2 & m_{12} & m_{22}
\end{bmatrix},
\]

(6.16)

if \( M \succeq 0 \) and \( \text{rank}\{M_{11}\} = 1 \), then \( m_{12} = m_1 m_2 \) holds.

**Proof.** Since \( \text{rank}\{M_{11}\} = 1 \), then \( m_{11} = m_1^2 \). From the positive semidefiniteness of \( M \), \( \det(M) \geq \)
0 should hold, that is,

\[
\det(M) = m_{11}m_{22} - m_{12}^2 - m_1(m_1m_{22} - m_2m_{12}) + m_2(m_1m_{12} - m_2m_{11}) \\
= m_1^2m_{22} - m_{12}^2m_{22} + 2m_1m_2m_{12} - m_2^2m_1^2 \\
= -(m_{12} - m_1m_2)^2 \geq 0,
\]

therefore, \(\det(M) = 0\), and \(m_{12} = m_1m_2\).

**Theorem 11.** The first order reduced SDP relaxation given by

\[
\{m^*_t, 2\}_{t=n_a}^T = \begin{array}{c}
\text{minimize} \\
0 \\
\text{subject to} \\
\forall t = n_a: \ M_{t,1}(m_t, 1) \succeq 0 \\
\forall t = n_a: \ L_{t,0}(g_k m_t, 1) \succeq 0, \forall g_k \in K_t 
\end{array}
\]

is exact if the first order moment matrix \(M_{T,1}^*\) built by \(m_t^*\) is rank-1. Further, in this case the corresponding optimal first order moments sequence is a feasible solution to the original nonconvex problem \(6.4\).

**Proof.** The first order moment matrix \(M_{t,1}\) is built in the form of

\[
M_{t,1} = m \left( \begin{bmatrix} 1 & v_t^T \end{bmatrix} \begin{bmatrix} 1 & v_t^T \\ r & w_i^T & \eta_t^T \end{bmatrix} \right) = \begin{cases} m \left( \begin{bmatrix} r \\
1 \ \ w_i^T \\
\eta_t \\ 1 \ \ r^T \\
1 \ \ r^T \\ 1 \\
\eta_t \\ 1 \ \ r^T \end{bmatrix} \right), & \text{for } \forall t = n_a \\
m \left( \begin{bmatrix} 1 & r \\
1 \ \ r^T \end{bmatrix} \right), & \text{for } t = T. 
\end{cases}
\]

Thus, \(M_{T,1}\) is a principal minor of \(M_{t,1}\) for \(t = n_a, \ldots, T - 1\). For \(t = n_a, \ldots, T - 1\), extract a \(3 \times 3\) submatrix, \(M_{t,1}^{(ij)}\),

\[
M_{t,1}^{(ij)} = m \left( \begin{bmatrix} 1 & r(i) \\
1 \ \ r(i) \ | \ w_j \end{bmatrix} \right)^T = \begin{bmatrix} 1 & m(r(i)) & m(w_j) \\
m(r(i)) & m(r(i)^2) & m(r(i)w_j) \\
w_j & w_j^2 & m(w_j^2) 
\end{bmatrix}
\]
from the interaction between rows and columns associated with \( \{1, r(i), w_j\} \), where \( i = 1, \ldots, n \), \( j = t - n_a, \ldots, t \), and \( r(i) \) denotes the \( j \)-th entry of \( r \). Since \( M_{t,1}^{ij} \) is a principal minor of \( M_{t,1}^* \) which is positive semi-definite, then \( M_{t,1}^{ij} \succeq 0 \) holds. Besides, if \( M_{t,1}^* T \) is rank-1, then the upper left block of \( M_{t,1}^{ij} \) is also rank-1. Then according to Lemma 4, we have

\[
m(r(i)w_j) = m(r(i))m(w_j), \quad \forall i = 1, \ldots, n, \quad \forall j = t - n_a, \ldots, t. \tag{6.19}
\]

Similarly we can also prove that

\[
m(r(i)\eta_j) = m(r(i))m(\eta_j), \quad \forall i = 1, \ldots, n, \quad \forall j = t - n_a, \ldots, t. \tag{6.20}
\]

Substituting (6.19) and (6.20) into the localizing matrix (which is a scalar) associated with (6.4a), for each \( t = n_a, \ldots, T - 1 \),

\[
\begin{bmatrix}
    u_{t-n_a:t} & y_{t-n_a:t}
\end{bmatrix} m(r) + m\left( \begin{bmatrix}
    w_{t-n_a:t} & \eta_{t-n_a:t}
\end{bmatrix} r \right) = 0
\]

\[
\Leftrightarrow \begin{bmatrix}
    u_{t-n_a:t} & y_{t-n_a:t}
\end{bmatrix} m(r) + m\left( \begin{bmatrix}
    w_{t-n_a:t} & \eta_{t-n_a:t}
\end{bmatrix} \right) m(r) = 0 \tag{6.21}
\]

Also the positive semi-definiteness of the localizing matrices associated with (6.4b)-(6.4d) guarantees

\[
m(r^T r) = m(r)^T m(r) = 1, \quad -\epsilon_1 \leq m(w_{0:T-1}) \leq \epsilon_1, \quad \text{and} \quad -\epsilon_2 \leq m(\eta_{0:T-1}) \leq \epsilon_2. \tag{6.22}
\]

Combining (6.21) and (6.22), we can conclude that the first order moments sequence is a feasible solution to the nonconvex problem (6.4).

In spite of the inspiring result in Theorem 11, if we solve (6.18) directly, it is uncommon for us to get a rank-1 moment matrix \( M^* T \) for free, for instance, choosing the moments sequence \( m_{t,2} \) in the way that the second order moments of \( r \) are nonzero and summing to 1 and all the others are equal to 0, then obviously, \( m_{t,2} \) is a feasible solution to (6.18). However, the lowest rank of the corresponding moment matrices \( M_{t,1}(m_{t,2}) \) is 2 rather than 1. Thus, we still need to add the rank-1 constraint explicitly which is nonconvex. In Algorithm 8 we apply the reweighted nuclear norm minimization to achieve rank-1 solution iteratively.
CHAPTER 6. EIV IDENTIFICATION OF LTI SYSTEMS

Algorithm 8 Moments Based EIV Identification of LTI Systems

1: Initialize: \( k = 0, W^{(1)} = I \in \mathbb{R}^{(n+1) \times (n+1)}, 0 < \delta \ll 1; \)
2: repeat
3: \( k = k + 1; \)
4: Solving
   \[
   M^{(k)}_{T,1} = \arg \min_{m_{t,2}} \text{trace}(W^{(k)}M^{(k)}_{T,1})
   \]
   subject to
   \[
   \forall_{t=m_1}^{T}: M_{t,N}(m_{t,2N}) \succeq 0
   \]
   \[
   \forall_{t=m_1}^{T}: L_{t,N-1}(g_t, m_{t,2N}) \succeq 0, \forall g_t \in K_t
   \]
5: Updating
   \[
   W^{(k+1)} = [M^{(k)}_{T,1} + \sigma_2(M^{(k)}_{T,1})^{-1}]^{-1}
   \]
6: until \( \sigma_2(M^{(k)}_{T,1}) < \delta \sigma_1(M^{(k)}_{T,1}). \)

6.4 Extensions

In this section we explore the flexibility of the proposed framework in taking into account the degradation of the quality of data, which is not uncommon in real applications, for instance, the missing data or outliers caused by the fault of the sensors or the occlusions in multi-object tracking.

6.4.1 Handling Missing Data

In the presence of missing data, the identification problem can be stated as

Problem 7. Given

- the measurements of the input and output sequence, \( \{u_t, y_t\}_{t=0}^{T-1} \), generated by the EIV model as shown in Figure 6.1, \( u_t \) at \( t \in T_m^{(u)} \) and \( y_t \) at \( t \in T_m^{(y)} \) are missing;
- the bound on the measurement noise, \( \epsilon_1 \) and \( \epsilon_2. \)

Determine the model parameter \( r \), complete the missing measurements, and find a noise sequence consistent with the given prior information.
Without of generality, assume the measurements at these time instances are 0. Then Problem 7 can be reformulated mathematically as finding a feasible solution \( \{r, w_{0:T-1}, \eta_{0:T-1}\} \) to

\[
\begin{bmatrix}
H_u^T - H_w^T & H_y^T - H_\xi^T
\end{bmatrix} r = 0
\]

(6.25a)

\( r^T r = 1 \)

(6.25b)

\( \|w_t\|_\infty \leq \epsilon_1, \forall t \in \{0 : T - 1\} \setminus T_m(u) \)

(6.25c)

\( \|\eta_t\|_\infty \leq \epsilon_2, \forall t \in \{0 : T - 1\} \setminus T_m(y) \)

(6.25d)

which can be solved by the same approaches to solving (6.4) proposed in Section 6.3 without the constraints on the noise sequence when the measurements are missing.

### 6.4.2 Handling Outliers

In the presence of outliers, the identification problem is more complicated than the case with missing data due to it is unknown when the outliers happen. We can apply the similar technique as in Chapter 3, introducing a binary variable \( s_{t,u} \) (or \( s_{t,y} \)) for the measurement \( u_t \) (or \( y_t \)) at each time instant to indicate whether it is an outlier or not. Minimizing the number of outliers, the identification problem can be recast as an optimization problem of the form

\[
p^* = \min_{r, w_t, \eta_t, s_{t,u}, s_{t,y}} \sum_{t=0}^{T-1} s_{t,u} + \sum_{t=0}^{T-1} s_{t,y}
\]

subject to

\[
\begin{bmatrix}
H_u^T - H_w^T & H_y^T - H_\xi^T
\end{bmatrix} r = 0
\]

\( r^T r = 1 \)

(6.26)

\( \|(1 - s_{t,u})w_t\|_\infty \leq (1 - s_{t,u})\epsilon_1, s_{t,u}^2 = s_{t,u}, \forall t = 0, \ldots, T-1 \)

\( \|(1 - s_{t,y})\eta_t\|_\infty \leq (1 - s_{t,y})\epsilon_2, s_{t,y}^2 = s_{t,y}, \forall t = 0, \ldots, T-1 \)

Obviously the problem above is nonconvex due to the last two constraints. However, it is also a second order polynomial constrained problem which can be solved via the moments based relaxations as in Section 6.3.3.

(6.26) also exhibits running intersection property, which is easy to confirm by partitioning the variables \( v = \{r^T, w_0, w_1, \ldots, w_{T-1}, \eta_0, \eta_1, \ldots, \eta_{T-1}, s_{0,u}, \ldots, s_{T-1,u}, s_{0,y}, \ldots, s_{T-1,y}\} \) as

\[
v_t = \begin{cases} 
\{r, w_t, \eta_t, s_{t,u}, s_{t,y}\}, & \text{for } t = n_a, \ldots, T - 1 \\
\{r\}, & \text{for } t = T
\end{cases}
\]

(6.27)
where \( w_t = [w_{t-na}, \ldots, w_t]^T \), \( \eta_t = [\eta_{t-na}, \ldots, \eta_t]^T \), \( s_{t,u} = [s_{t-na,u}, \ldots, s_{t,u}]^T \), and \( s_{t,y} = [s_{t-na,y}, \ldots, s_{t,y}]^T \), and partitioning the objective function \( p \) and the constraint set \( K \) in the way

\[
p = \sum_{t=na}^{T-1} p_t, \quad p_t = \begin{cases} 
\sum_{t-na}^t (s_{t,u} + s_{t,y}), & \text{for } t = na \\
0, & \text{for } t = T
\end{cases}
\]

\[
K = \bigcap_{t=na}^{T-1} K_t, \quad \forall_{t=na}^{T-1} : K_t = \left\{ \begin{bmatrix} u_{t-na:t} - w_{t-na:t} \\
y_{t-na:t} - \eta_{t-na:t} \end{bmatrix} \right\} r = 0
\]

\[
\| (1 - s_{k,u})w_k \|_\infty \leq (1 - s_{k,u})\epsilon_1, s_{k,u}^2 = s_{k,u}, \forall_{k=na}^t
\]

\[
\| (1 - s_{k,y}\eta_k \|_\infty \leq (1 - s_{k,y})\epsilon_2, s_{k,y}^2 = s_{k,y}, \forall_{k=na}^t
\]

Due to the nonconvexity introduced in the last two constraints in (6.26), the optimality condition for its first order moment relaxation is more restrictive than in Theorem 11.

**Theorem 12.** The first order reduced SDP relaxation for (6.26) is exact if the first order moment matrix \( M_{T,1}^* \) built for the model parameter \( x \) is rank-1 and the first order moments \( m(s_{t,u})^*, m(s_{t,y})^* \) for the indicator variables are binary.

**Proof.** The proof is similar to the proof of Theorem 11. The binary condition on \( m(s_{t,u})^* \) and \( m(s_{t,y})^* \) is to guarantee the rank-1 condition of the moment matrix associated with \( s_{t,u} \) and \( s_{t,y} \), which is required in applying Lemma 5 to prove

\[
\left\{ \begin{align*}
\| m(w_k) - m(s_{k,u}w_k) \|_\infty & \leq (1 - m(s_{k,u}))\epsilon_1 \\
\| m(\eta_k) - m(s_{k,y}\eta_k) \|_\infty & \leq (1 - m(s_{k,y}))\epsilon_2
\end{align*} \right.
\]

can guarantee

\[
\left\{ \begin{align*}
\| m(w_k) - m(s_{k,u})m(w_k) \|_\infty & \leq (1 - m(s_{k,u}))\epsilon_1 \\
\| m(\eta_k) - m(s_{k,y})m(\eta_k) \|_\infty & \leq (1 - m(s_{k,y}))\epsilon_2
\end{align*} \right.
\]

\[\square\]
6.5 Experiments

Consider a strictly proper second order LTI system with the true parameter vector $\hat{r} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}^T = \begin{bmatrix} -1.6 & 0.9 \\ 0.1 & 0.1 \end{bmatrix}^T$. Measurements of the input/output sequence $\{u_t, y_t\}_{t=0}^{T-1}$ ($T = 100$) are available, the input $x_t$ is uniformly distributed between $[-1, +1]$ and the measurements of the input and the output are corrupted by random noise $w_t \in [-\epsilon_1, +\epsilon_1]$ and $\eta_t \in [-\epsilon_2, +\epsilon_2]$ respectively. Different noise bounds $\epsilon_1, \epsilon_2$ would generate data of different the signal-to-noise ratios

$$\text{SNR}_u = 10 \log_{10} \left\{ \frac{\sum_{t=0}^{T-1} u_t^2}{\sum_{t=0}^{T-1} w_t^2} \right\}$$

and

$$\text{SNR}_y = 10 \log_{10} \left\{ \frac{\sum_{t=0}^{T-1} y_t^2}{\sum_{t=0}^{T-1} \eta_t^2} \right\}.$$ 

In Table 6.1, we record the estimates of model parameters, $\hat{r} = \begin{bmatrix} \hat{a}_1 & \hat{a}_2 \\ \hat{b}_1 & \hat{b}_2 \end{bmatrix}$, under different experimental settings.

To substantiate the capacity of the proposed framework to take care of the missing data, using the same dataset as in Table 6.1, we randomly mask 10% and 20% of the measurements of the output as missing data, the system is identified by solving (6.25). As shown in Table 6.2 and Table 6.3, it is seen that the proposed approach can handle cases with missing data, and with the missing data taking up more percentage, the estimation error of the model parameter becomes larger.

### Table 6.1: Estimates of Model Parameters under Different Noise Bounds

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\epsilon_1$</th>
<th>$\epsilon_2$</th>
<th>$\text{SNR}_u$</th>
<th>$\text{SNR}_y$</th>
<th>Estimation Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0</td>
<td>0.20</td>
<td>17.2121</td>
<td>-1.5870</td>
<td>0.8884</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
<td>0.20</td>
<td>17.2121</td>
<td>-1.5650</td>
<td>0.8704</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.15</td>
<td>20.2351</td>
<td>-1.5858</td>
<td>0.8849</td>
</tr>
<tr>
<td>100</td>
<td>0.075</td>
<td>0.15</td>
<td>20.2351</td>
<td>-1.5562</td>
<td>0.8617</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.10</td>
<td>24.0155</td>
<td>-1.5961</td>
<td>0.8962</td>
</tr>
<tr>
<td>100</td>
<td>0.05</td>
<td>0.10</td>
<td>24.0155</td>
<td>-1.6060</td>
<td>0.9037</td>
</tr>
</tbody>
</table>

### Table 6.2: Estimates of Model Parameters with 10% Missing Data

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\epsilon_1$</th>
<th>$\epsilon_2$</th>
<th>$\text{SNR}_u$</th>
<th>$\text{SNR}_y$</th>
<th>Estimation Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0</td>
<td>0.20</td>
<td>17.2121</td>
<td>-1.5850</td>
<td>0.8757</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.15</td>
<td>20.2351</td>
<td>-1.5348</td>
<td>0.8372</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.10</td>
<td>24.0155</td>
<td>-1.5874</td>
<td>0.8867</td>
</tr>
</tbody>
</table>

### Table 6.3: Estimates of Model Parameters with 20% Missing Data

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\epsilon_1$</th>
<th>$\epsilon_2$</th>
<th>$\text{SNR}_u$</th>
<th>$\text{SNR}_y$</th>
<th>Estimation Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0</td>
<td>0.20</td>
<td>17.2121</td>
<td>-1.5282</td>
<td>0.8008</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.15</td>
<td>20.2351</td>
<td>-1.5348</td>
<td>0.8330</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.10</td>
<td>24.0155</td>
<td>-1.5007</td>
<td>0.8035</td>
</tr>
</tbody>
</table>
6.6 Conclusions

In this chapter we consider the identification of LTI systems when the measurements of both input and output are corrupted by $\ell_\infty$-norm bounded noise. Although it is more complicated than the case where noise enters the system in error-in-equation form, it can be also solved by moments based relaxation due to the fact that it can be reformulated as a constrained polynomial optimization problem. By exploiting the sparsity pattern inherited by the problem and low rank solution to the moment matrix associated with the model parameters, we show the optimality condition for the first order moment relaxation. In addition, we also show the flexibility of the proposed framework to handle missing data and outliers.
Chapter 7

EIV Identification of SARX Systems

7.1 Motivation

Although the LTI model can be used to model a wide variety of physical systems and it has obtained extensive research effects, its simple structure is deemed to its failure and incapability in modeling complex systems arising in our real life, for instance, biological cell growth and division, multi-agent systems like swarms of micro-air vehicles, promoting the development of nonlinear systems. Known to be universal approximators [105], switched linear models, dynamical systems with interacting continuous-time dynamics and discrete-event dynamics, are used as surrogates for many complex nonlinear systems, and identification of switched systems has been the subject of intense research during the past few years. It is well known that in the presence of noise, the problem is NP-hard. Thus, most of the research efforts have concentrated on finding computationally efficient methods (see for instance, [106] for an excellent survey of earlier work and [15–21] for more recent results). Roughly speaking these methods above either attempt to find good solutions to a nonconvex problem (for instance, [19] reduces the problem to optimization of a difference of convex functions), or resort to convex surrogates as in [15–18]. While these methods have been empirically shown to work well in many scenarios, those based on nonconvex optimization can get trapped in local minima, and there are no a priori bounds on the gap between the solution obtained using convex relaxations and the ground truth.

Similar to the analysis of LTI models, there are two types of noise contaminating the model: process noise and measurement noise. For the case with process noise, the identification problem

\footnote{While there are conditions under which these relaxations are known to be exact, typically these conditions do not hold for identification problems due to the underlying Hankel-like structure of the data.}
can be solved in the way as the switched linear model fitting for static data proposed in Chapter 4. In contrast, the case with measurement noise, rather than process noise is substantially harder, and fewer results are available. Most of the existing Error-In-Variables (EIV) methods (taking care of measurement noise) are aimed at linear systems (see [96,100] and references therein). To our best knowledge, for switched systems, only the work in [21] extends the idea of [20] to cases with measurement noise. However, since this method is based on rank-minimization, there is no guarantee that it will find a suitable model even if one exists. In addition, as an extension of GPCA, it also inherits all the weakness of GPCA when applied to noisy data.

Motivated by these weaknesses, in this chapter, we propose an alternative framework for identifying affine switched systems from data corrupted by measurement noise. Rather than exploiting the hybrid decoupling constraint as in previous work [5,20,21], we apply the trick of binary indicator variables \( s_{i,j} \) introduced in Chapter 4. Although the proposed method in Chapter 4 can be used to identify SARX models from data corrupted by process noise, its extension to cases with measurement noise is not trivial, which will be shown in this chapter.

In the first part of this chapter, we propose the main result, i.e. the moments-based EIV method of identifying SARX models from noisy input/output data with complete topology, that is, submodels are allowed to switch to each other arbitrarily fast. In the second part, we exploit the capability of the proposed framework to handling cases of incomplete topology, such as the prohibition of transitions from some subset of submodels to another subset of submodels, or the requirement of minimum number of switches.

### 7.2 Problem Statement

This chapter is aimed to identify SARX models from noisy input/output data such that certain priors satisfy. For the ease of reading, here we just consider the single-input single-output case, and the proposed approach can be extended to multi-input multi-output systems easily.

In the case that the experimental data are noiseless, the SARX model

\[
y_t = -\sum_{k=1}^{n_a} a_k(\sigma_t)y_{t-k} + \sum_{k=1}^{n_b} b_k(\sigma_t)u_{t-k}, \quad \sigma_t \in \mathbb{N}_{N_a}
\]  

(7.1)
CHAPTER 7. EIV IDENTIFICATION OF SARX SYSTEMS

consists of \( N_s \) linear autoregressive submodels of the form

\[
y_t = - \sum_{k=1}^{n_a} a_k y_{t-k} + \sum_{k=1}^{n_b} b_k u_{t-k} \quad (G_i)
\]

where \( u_t \) and \( y_t \) denotes the input and the output signals respectively and the discrete mode variable \( \sigma_t \) denotes the submodel \( G_{\sigma_t} \) is active at time instant \( t \).

On the other hand, in the presence of noise, according to the different ways the noise enters the model, SARX models can be written as:

- **Process Noise** \( e_t \). The equation error structure:

\[
y_t = - \sum_{k=1}^{n_a} a_k(\sigma_t)y_{t-k} + \sum_{k=1}^{n_b} b_k(\sigma_t)u_{t-k} + e_t, \quad \sigma_t \in \mathbb{N}_{N_s}; \tag{7.2}
\]

- **Measurement Noise** \( w_t, \eta_t \). The EIV structure:

\[
y_t - \eta_t = - \sum_{k=1}^{n_a} a_k(\sigma_t)(y_{t-k} - \eta_{t-k}) + \sum_{k=1}^{n_b} b_k(\sigma_t)(u_{t-k} - w_{t-k}), \quad \sigma_t \in \mathbb{N}_{N_s}. \tag{7.3}
\]

Let \( r_i = \begin{bmatrix} -b_{n_b}(i) & \ldots & -a_1(i) \end{bmatrix}^T \in \mathbb{R}^{1+n_a+n_b} \) denote the model parameter for \( G_i \), and write the input/output data in a compact form as \( x_t = \begin{bmatrix} u_{t-n_b} & \ldots & u_{t-2} & u_{t-1} & y_{t-n_a} & \ldots & y_{t-1} & y_t \end{bmatrix}^T \in \mathbb{R}^{1+n_a+n_b} \), then the SARX model in equation error structure (7.2) can be written as

\[
x_t^T r_{\sigma_t} - e_t = 0. \tag{7.4}
\]

For the EIV structure, letting \( w_t = \begin{bmatrix} w_{t-n_b} & \ldots & w_{t-2} & w_{t-1} \end{bmatrix}^T \) and \( \eta_t = \begin{bmatrix} \eta_{t-n_a} & \eta_{t-1} & \eta_t \end{bmatrix}^T \), (7.3) is reformulated as

\[
x_t^T r_{\sigma_t} - \begin{bmatrix} w_t^T & \eta_t^T \end{bmatrix} r_{\sigma_t} = 0. \tag{7.5}
\]

Obviously, the identification of SARX model is equivalent to fitting samples \( \{x_t\} \) to a subspace arrangement characterized by the normal vectors \( \{r_i\}_{i=1}^{N_s} \). Thus, it can be solved by the method proposed in Chapter 4. However, for the EIV structure, the identification becomes more complicated due to the bilinear term and the measurement noise variables \( w_t, \eta_t \) affect more than one regression equations. Thus, in the sequel, we will focus on solving the identification problem of
CHAPTER 7. EIV IDENTIFICATION OF SARX SYSTEMS

SARX models with EIV structure, which is formally stated as:

**Problem 8.** Given:

- A set of input/output data \( \{u_t, y_t\}_{t=t_0}^T \) generated by the SARX model of the form \( (7.3) \);

- A-priori information consisting of (i) the discrete mode \( \sigma_t \) can switch arbitrarily fast among \( N_s \) submodels, that is, \( \sigma_t \in \mathbb{N}_{N_s} \), (ii) a bound \( \epsilon \) on the measurement noise \( \eta_t \), (iii) additional information, such as, \( N_{f_i} \), the relative frequency of each submodel, (v) point wise co-occurrence information, and constraints on the switch sequence.

Establish whether the observed data is consistent with the a priori assumptions and, if so, recover the parameters of each submodel, that is, find a set of coefficients \( \{a_{k=1}^{n_a}(i), b_{k=1}^{n_b}(i)\} \) associated with submodel \( G_i \), \( \forall i = 1, \ldots, N_s \).

### 7.3 A Convex Approach to Identification of SARX Models

In this section, we present the main result of this chapter: a moments based convex optimization approach to **Problem 8**. The main idea is to first recast the problem into a semi-algebraic feasibility form, and then solve using the results outlined in Section 2.2. For simplicity, we will first develop the main idea without considering additional prior information listed in (iii) in the statement of **Problem 8**. These cases will be covered in Section 7.3.4.

#### 7.3.1 A Semi-Algebraic Optimization Reformulation

Without loss of generality, let \( r_i = [a_1(i) \cdots a_{n_a}(i) \ b_1(i) \cdots b_{n_b}(i)]^T \) and \( x_t = [-y_{t-1} \cdots -y_{t-n_a} \ u_{t-1} \cdots u_{t-n_b}]^T \) denote the unknown parameters associated with the submodel \( G_i \) and the known continuous state (or regression vector) at time instant \( t \), respectively. By introducing a set of binary variables \( s_{i,t} \) that indicates whether the submodel \( G_i \) is active at time instant \( t \) (\( \sigma_t = i \iff s_{i,t} = 1 \)), it is easy to show that **Problem 8** is equivalent to:
CHAPTER 7. EIV IDENTIFICATION OF SARX SYSTEMS

Problem 9. Determine feasibility of the following set of nonlinear equations and inequalities:

\[
\forall_{T=0+na} \forall_{i=1}^N: s_{i,t}[y_t - \eta_t - r_i^T x_t + \sum_{k=1}^{na} a_k(i)\eta_{t-k} - \sum_{k=1}^{nb} b_k(i)w_{t-k}] = 0 \tag{7.6a}
\]

\[
\forall_{T=0} : \|w_t\|_\infty \leq \epsilon_1, ||\eta_t||_\infty \leq \epsilon_2 \tag{7.6b}
\]

\[
\forall_{T=0+na} \forall_{i=1}^N: s_{i,t}^2 = s_{i,t} \tag{7.6c}
\]

\[
\forall_{T=0+na} \forall_{i=1}^N: \sum_{i=1}^N s_{i,t} = 1 \tag{7.6d}
\]

\[
r_1(1) \geq r_2(1) \geq \cdots \geq r_Ns(1) \tag{7.6e}
\]

Here the last constraint means to eliminate the symmetry of solutions, and constraints (7.6c)-(7.6d) on \(s_{i,t}\) guarantee that only one submodel is active at time instant \(t\). Next, we show that the feasibility of Problem 9 can be recast as a polynomial optimization problem.

Lemma 5. The following two statements are equivalent: (i) Problem 9 is feasible.

(ii) \(p^* = 0\), where \(p^*\) represents the optimal value of the objective function in the following polynomial optimization problem:

\[
p^* = \min_{r, s, r, \eta} \sum_{t=0+na} \sum_{i=1}^N s_{i,t}[y_t - \eta_t - r_i^T x_t + \sum_{k=1}^{na} a_k(i)\eta_{t-k} - \sum_{k=1}^{nb} b_k(i)w_{t-k}]^2 \tag{7.7}
\]

subject to

\[
r_1(1) \geq r_2(1) \geq \cdots \geq r_Ns(1) \tag{7.7}
\]

\[
\forall_{T=0} : \|w_t\|_\infty \leq \epsilon_1, ||\eta_t||_\infty \leq \epsilon_2 \tag{7.7}
\]

\[
\forall_{T=0+na} \forall_{i=1}^N: s_{i,t}^2 = s_{i,t} \tag{7.7}
\]

\[
\forall_{T=0+na} \forall_{i=1}^N: \sum_{i=1}^N s_{i,t} = 1 \tag{7.7}
\]

Proof. The proof follows from the fact that the objective function of (7.7) is a sum of squares, thus \(p^* \geq 0\), and \(p^* = 0\) in (7.7) is equivalent to the feasibility of (7.6) in Problem 9.

7.3.2 A Sequence of Convex Relaxations

From the results in Section 2.2 it follows that the nonconvex problem above can be relaxed into a sequence of (monotonically convergent) convex problems of the form:

\[
p^*_{m,N} = \min_{m_{2N}} c^T m_{2N} \tag{7.8}
\]

subject to

\[
M_N(m_{2N}) \succeq 0 \tag{7.8}
\]

\[
L_{N-1}(g_k m_{2N}) \succeq 0, \forall g_k \text{ in constraints of (7.7)} \tag{7.8}
\]

93
CHAPTER 7. EIV IDENTIFICATION OF SARX SYSTEMS

where \( m_{2N} \) denotes the moment sequence of \( v = \{\forall_{i=1}^{N_x} \forall_{t=0}^{T} : r_i, s_{i,t}, e_{i,t}, w_t, \eta_t\} \) up to order \( 2N \), \( c \) is the vector of the coefficients corresponding to the monomials in the polynomial describing the objective in (7.7), and \( M_N \) and \( L_{N-1} \), consisting of moment sequence up to order \( 2N \), are the moment matrix and localizing matrix corresponding to the constraint in (7.7). Following Theorem 3.4 in [25], \( p^*_{m,N} \) provides a lower bound to \( p^* \), and the bound gets tighter with the increasing of \( N \), that is, \( p^*_{m,N} \leq p^*_{m,N+1} \leq \cdots \leq p^* \). The following theorem gives further analysis on the conditions under which the equality holds.

**Theorem 13.** (i) For Problem 9 if (7.6) is feasible, then \( p^*_{m,N} = p^* = 0 \) at \( \forall N \geq 3 \) for (7.7).

(ii) Problem 9 is infeasible if and only if there exists some \( N \) such that \( p^*_{m,N} > 0 \).

**Proof.** (i) From Lemma 5 if (7.6) is feasible then \( p^* = 0 \) in (7.7). Thus \( p(x) - p^* = s_i^2[y_t - \eta_t - r_i^T x_t + \sum_{k=1}^{n_a} a_k(i)\eta_{t-k}]^2 = \) sum of squares. The fact that \( p^*_{m,N} = p^* = 0 \) at \( N = 3 \) follows now from Theorem 2 in Chapter 2.

(ii) \( p^* \geq p^*_{m,N} > 0 \Rightarrow \) Problem 9 is infeasible follows from the fact that \( p^*_{m,N} \leq p^* \). Conversely, from (monotonic) convergence of the sequence of relaxations (Theorem 4.2 in [25]) it follows that if \( p^* > 0 \) then there exists some \( N \) such that \( p^*_{m,N} > 0 \).

7.3.3 Computational Complexity Analysis

Although it is convenient for us to derive some sufficient (or necessary) condition for the equivalence between the convex formulation (7.8) and the nonconvex formulation (7.7), the lowest relaxation order for (7.7) is 3, which restricts its application to problems of limited scale due to its computational complexity. On the other hand, the lowest relaxation order of (7.6) is 2, thus, under the assumption that (7.6) is feasible, we prefer to solve the convex relaxation of (7.6).

Beside, as noticed, the constraints \( K \) in (7.6) can be partitioned into \( L = T - t_0 - n_a + 2 \) subsets as

\[
K = \bigcap_{t=n_a+t_0-1}^{T} K_t, \text{ where } \\
K_{n_a+t_0-1} : \begin{cases} 
|w_k| \leq \epsilon_1, \forall_{k=t-n_a}^{t-1} \\
|\eta_k| \leq \epsilon_2, \forall_{k=t-n_a}^{t-1} \\
\forall_{t=n_a+t_0}^{T-n_a-1} : K_t : 
\end{cases}
\]

\[
s_{i,t} \geq s_{i,t}, \forall_{i=1}^{N_x} \\
\sum_{i=1}^{N_x} s_{i,t} = 1 \\
s_{i,t}[y_t - \eta_t - r_i^T x_t + \sum_{k=1}^{n_a} a_k(i)\eta_{t-k} - \sum_{k=1}^{n_b} b_k(i)w_{t-k}] = 0, \forall_{i=1}^{N_x}
\]
CHAPTER 7. EIV IDENTIFICATION OF SARX SYSTEMS

with each partition \( K_t \) for \( t = n_a + t_0, \ldots, T \) associated with a time instant \( t \), it is easy to check that for \( \forall t = t_0 + n_a - 1, \ldots, T \), \( p_t \) and \( K_t \) contain variables in the subsets

\[
v_t = \begin{cases} \{ \forall i=1 : r_i \} \in \mathbb{R}^{(n_a+n_b)N_s} \text{, for } t = t_0 + n_a - 1; \\ \{ \forall i=1 : r_i, \forall k=t-n_b : w_i, \forall k=t-n_a : \eta_i, \forall i=1 : s_i \} \in \mathbb{R}^{(n_a+n_b)(N_s+1)+N_s+1} \text{, for } \forall t = t_0 + n_a. \end{cases}
\]

It is easy to show that \( \bigcup_{t=t_0+n_a-1}^{T} v_t = v \), and

\[
v_{t+1} \cap \bigcup_{t=t_0+n_a-1}^{t} v_t = \begin{cases} \{ \forall i=1 : r_i \} \subseteq v_t \text{, for } t = t_0 + n_a - 1 \\ \{ \forall i=1 : r_i, \forall k=t-n_b+1 : w_i, \forall k=t-n_a+1 : \eta_i \} \subseteq v_t \text{, for } \forall t = t_0 + n_a \end{cases}
\]

hold, meaning the problem exhibits the running intersection property and hence Problem \( \mathcal{V} \) can be solved by solving a sequence of convex problem of the form

\[
p_{m,N}^* = \min_{m_{t,2N}} \quad 0 \\
\text{subject to} \\
\forall t = t_0 + n_a - 1 : \quad M_{t,N}(m_{t,2N}) \succeq 0 \\
\forall t = t_0 + n_a - 1 : \quad L_{t,N-1}(g_{t_k} m_{t,2N}) \succeq 0, \forall g_{t_k} \in K_t \\
\forall t = t_0 + n_a : \quad L_{t,N-2}(h_{t_k} m_{t,2N}) \succeq 0, \forall h_{t_k} \in K_t
\]

(7.9)

where \( m_{t,2N} \) and \( M_{t,2N} \) denote the moments sequence of variables in \( v_t \) of order up to \( 2N \) and the associated moment matrix of order \( N \), and \( L_{t,N-1}, L_{t,N-2} \) are the localizing matrices related to constraints in \( K_t \). Instead of considering a single moment matrix, \( M_{N} \), of size

\[
(L-1+n_a+n_b)(N_s+1)+L-2+N \\
N
\]

in (7.8), in (7.9) we consider \( L \) smaller moment matrices, \( M_{t,N} \), of size at most

\[
((n_a+n_b)(N_s+1)+N_s+1+N) \\
N
\]

substantially reducing the computational burden when the length of the temporal sequence is long (that is, \( L = T - t_0 - n_a + 1 \) is large).

Problem \( \mathcal{V} \) requires not only to check the feasibility of Problem \( \mathcal{V} \), but also to extract a feasible solution to (7.6). However, solving the relaxation only yields the value of \( p^* \) and the corresponding (truncated) moments sequence of the variables of interest (the \( r_i \)'s). In general, reconstructing a minimizer from the corresponding moments sequence is far from trivial (see for
instance [39] for a description of a tool to accomplish this). To circumvent this difficulty, next, we will exploit the fact that moment matrices associated with atomic measures having a single atom are of rank 1, since in this case the moments simply correspond to the powers of the variables, evaluated at the location of the atom. This observation motivates the next result, involving a modification of \((7.9)\) that enforces low rank and sparse solutions.

**Theorem 14.** Assume that Problem \((9)\) is feasible and let \(m_{t,4}^*\) denote an optimal solution to the following problem

\[
\begin{aligned}
\min_{m_{t,4}} \quad & 0 \\
\text{subject to} \quad & \forall t = t_0 + n_a - 1: \ M_{t,2}(m_{t,4}) \succeq 0 \\
& \forall t = t_0 + n_a - 1: \ L_{t,1}(g_t, m_{t,4}) \succeq 0, \ \forall g_t \in K_t \\
& \forall t = t_0 + n_a: \ L_{t,N-2}(h_t, m_{t,4}) \succeq 0, \ \forall h_t \in K_t \\
& \text{rank}\{M_{t_0+n_a-1,2}\} = 1 \\
& \forall i = 1, \forall t = t_0 + n_a: \ m(s_{i,t}) \in \{0, 1\} \quad (7.10e)
\end{aligned}
\]

where \(m(\bullet)\) denotes the moment of \(\bullet\). Then the vectors, \(r_i^* = m_{t,4}^*(r_i), \ \forall i = 1, \ldots, N_s\), obtained from the first order moment variables are feasible solutions to Problem \(8\).

**Proof.** Let \(r = [\begin{array}{c} r_1^T \\ \vdots \\ r_{N_s}^T \end{array}]^T, s_t = [\begin{array}{c} s_{1,t} \\ \vdots \\ s_{N_s,t} \end{array}]^T\), and \(e_t = [\begin{array}{c} w_{t-n_a} \\ \vdots \\ w_{t-1} \\ \eta_{t-n_a} \\ \vdots \\ \eta_{t-1} \\ \eta_t \end{array}]^T\), for \(t = t_0 + n_a, \ldots, T\). Without loss of generality, denote any entry of \(r, s_t, e_t\) by \(r(i), s_t(j), \text{and} e_t(k)\), respectively, for \(i = 1, \ldots, N_s(n_a + n_b), j = 1, \ldots, N_s, k = 1, \ldots, n_a + n_b + 1\). Then from the positive semi-definiteness of \(M_{t,2}\), we know that its principal minors

\[
M_{t,2}^{(ij,k)} = \begin{bmatrix}
1 & r(i) \\
1 & s_t(j)e_t(k)
\end{bmatrix}
\begin{bmatrix}
1 & r(i) \\
1 & s_t(j)e_t(k)
\end{bmatrix}
\begin{bmatrix}
0 & m(s_t(j)e_t(k)) & m(r(i)e_t(k)) \\
m(r(i)) & m(r(i)^2) & m(r(i)s_t(j)e_t(k)) \\
m(s_t(j)e_t(k)) & m(r(i)s_t(j)e_t(k)) & m(s_t(j)^2e_t(k)^2)
\end{bmatrix}
\begin{bmatrix}
0 & m(r(i)e_t(k)) \\
m(r(i)) & m(r(i)^2) & m(r(i)s_t(j)e_t(k)) \\
m(s_t(j)e_t(k)) & m(r(i)s_t(j)e_t(k)) & m(s_t(j)^2e_t(k)^2)
\end{bmatrix}
\geq 0
\]

96
and

\[
M_{t,2}^{(ij)} = m \begin{pmatrix}
1 \\
\frac{r(i)}{s_t(j)}
\end{pmatrix}
\begin{pmatrix}
1 & r(i) & s_t(j) \\
\end{pmatrix}
\]

\[
= \begin{bmatrix}
1 & m(r(i)) & m(s_t(j)) \\
m(r(i)) & m(r(i)^2) & m(r(i)s_t(j)) \\
m(s_t(j)) & m(r(i)s_t(j)) & m(s_t(j)^2)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
M_{t,2}^{(i)} & m(s_t(j)) & m(s_t(j)^2) \\
m(s_t(j)) & m(r(i)s_t(j)) & m(s_t(j)^2)
\end{bmatrix} \succeq 0
\]

From the partitioning of the variables in (7.3.3), we can see that the rank-1 constraint in (7.10) is enforced on the moment matrix of the model parameters \(r\). As a principal minor of the rank-1 matrix \(M_{t_0+n_a-1,2}\), \(M_{t,2}^{(i)}\) is also rank-1, thus, following Lemma 4 in Chapter 6, we can obtain

\[
\begin{align*}
\{ & m(r(i)s_t(j)e_t(k)) = m(r(i)m(s_t(j)e_t(k)), \\
& m(r(i)s_t(j)) = m(r(i)m(s_t(j))
\}
\]

(7.11)

(7.12)

Combining the sparse constraint enforced on \(m(s_t(j))\) in (7.10e) and the localizing matrix corresponding to (7.6c), that is, \(m(s_t(j)) = m(s_t(j)^2)\), it is easy to show that the submatrix of \(M_{t,2}\),

\[
M_{t,2}^{(j)} = m \begin{pmatrix}
1 \\
\frac{s_t(j)}{e_t(k)}
\end{pmatrix}
\begin{pmatrix}
1 & s_t(j) \\
\end{pmatrix}
\]

\[
= \begin{bmatrix}
1 & m(s_t(j)) \\
m(s_t(j)) & m(s_t(j)^2)
\end{bmatrix}
\]

is rank-1 for each \(i = 1, \ldots, N_s\), \(t = t_0 + n_a, \ldots, T\).

Following \(M_{t,2} \succeq 0\), its principal minor

\[
M_{t,2}^{(jk)} = m \begin{pmatrix}
1 \\
\frac{s_t(j)}{e_t(k)}
\end{pmatrix}
\begin{pmatrix}
1 & s_t(j) & e_t(k) \\
\end{pmatrix}
\]

\[
= \begin{bmatrix}
1 & m(s_t(j)) & m(e_t(k)) \\
m(s_t(j)) & m(s_t(j)e_t(k)) & m(e_t(k)^2)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
M_{t,2}^{(j)} & m(e_t(k)) & m(s_t(j)e_t(k)) \\
m(e_t(k)) & m(s_t(j)e_t(k)) & m(e_t(k)^2)
\end{bmatrix} \succeq 0,
\]

97
and $\mathbf{M}^{(j)}_{i,2}$ is rank-1, then similarly as the previous derivation, we have

$$m(s_t(j)e_t(k)) = m(s_t(j))m(e_t(k)).$$  \hspace{1cm} (7.13)

Combining (7.11) and (7.13), it can be obtained that

$$m(r(i)s_t(j)e_t(k)) = m(r(i))m(s_t(j))m(e_t(k)).$$ \hspace{1cm} (7.14)

Now inserting (7.12), (7.13) and (7.14) into the localizing matrix corresponding to (7.6a) (which is a linear equation for relaxation order 2), we get

$$m(s_{i,t}[y_t - \eta_t - r_{i}^T x_t + \sum_{k=1}^{n_a} a_k(i)\eta_{t-k} - \sum_{k=1}^{n_b} b_k(i)w_{t-k}]) = 0$$
$$\Leftrightarrow m(s_{i,t})y_t - m(s_{i,t})m(\eta_t) - m(s_{i,t})m(r_i)^T x_t$$
$$+ \sum_{k=1}^{n_a} m(s_{i,t})m(a_k(i))m(\eta_{t-k}) - \sum_{k=1}^{n_b} m(s_{i,t})m(b_k(i))m(w_{t-k}) = 0,$$

meaning the first order moments sequence is a feasible solution to (7.6a). In addition, by the definition of the localizing matrix and (7.10e), we know that the first order moments sequence satisfies (7.6b)-(7.6e). Thus, the first order moments sequence to (7.10) is a feasible solution to the original nonconvex problem (7.6) and Problem 8.

In (7.10), both the low rank constraint (7.10d) and the sparsity constraint (7.10e) are nonconvex, by resorting to the reweighted nuclear minimization in [50] and reweighted $\ell_1$ norm minimization in [89], we come up with Algorithm 9 achieving a low rank and sparse solution to (7.10) iteratively.

### 7.3.4 Handling Additional A-Priori Information

In this section, we illustrate the ability of the proposed framework to incorporate additional prior information. For instance, in many biological applications, certain transitions are inhibited and the relative frequency of each metabolic stage is known. Similarly, some of the data may be annotated manually so that it is known that at two given time instants the same submodel is active. As shown below, this additional information can be easily incorporated into our formulation by simply adding suitable linear constraints on the variables $s_{i,t}$. Specifically:

(i) submodel $G_i$ is active for $f\%$ of the time $\iff \sum_{t=t_0+n_a}^{T} s_{i,t} = 0.01f(T + 1 - t_0 - n_a)$;
Algorithm 9 Moments-Based SARX Identification

1. Initialize: \( k = 0, 0 < \delta \ll 1, W^{(0)} = I, w_{i,t}^{(0)} = 1, \) for \( \forall_{i=1}^{N_s} \forall_{t=0+n_a}^{T} \).

2. repeat

3. Solve

\[
\{ M_{t_0+n_a-1,2}^{(k)}, m(s_{i,t})^{(k)} \} = \arg \min_{\lambda \sum_{i=1}^{N_s} \sum_{t=t_0+n_a}^{T} w_{i,t}^{(k)} m(s_{i,t})}
\]

subject to

\[
\sum_{i=1}^{N_s} \sum_{t=t_0+n_a}^{T} w_{i,t}^{(k)} W_{t}^{(k)} M_{t_0+n_a-1,2}^{(k)} \cdots
+ \lambda \sum_{i=1}^{N_s} \sum_{t=t_0+n_a}^{T} w_{i,t}^{(k)} m(s_{i,t})
\]

\[
\forall_{t=t_0+n_a-1}^{T}: M_{t,2}(m_{j,4}) \geq 0 \quad \forall_{t=t_0+n_a-1}^{T}: L_{t,1}(g_{t_k} m_{t,4}) \geq 0, \forall g_{t_k} \in K_t
\]

\[
L_{t,0}(h_{t_k} m_{t,4}) \geq 0, \forall h_{t_k} \in K_t
\]

4. Update

\[
W^{(k+1)} = \left[ M_{t_0+n_a-1,2}^{(k)} + \sigma_2(M_{t_0+n_a-1,2}) \right]^{-1}, W^{(k+1)} = W^{(k+1)}/\|W^{(k+1)}\|_2
\]

\[
w_{i,t}^{(k+1)} = \left[ s_{i,t}^{(k)} + \delta \right]^{-1}, w_{i,t}^{(k+1)} = w_{i,t}^{(k+1)} / \max \{ w_{i,t}^{(k+1)} \}
\]

5. until \( \sigma_2(M_{t_0+n_a-1,2}) \to 0 \) and \( m(s_{i,t})^{(k)} - (m(s_{i,t})^{(k)})^2 \to 0 \), for \( \forall_{i=1}^{N_s} \forall_{t=t_0+n_a}^{T} \).

(ii) the same submodel is active at time instants \( m \) and \( n \iff s_{i,m} = s_{i,n}, \forall i = 1, \ldots, N_s \);

(iii) different submodels are active at time instants \( m \) and \( n \iff s_{i,m} s_{i,n} = 0, \forall i = 1, \ldots, N_s \);

(iv) submodel \( i \) cannot be switched to submodel \( j \iff s_{i,t} s_{j,t+1} = 0, \forall t.

7.4 Extension: Identification with Minimum Number of Switches

In the past few years, an interesting application of switched system identification to data segmentation has emerged, aiming to segment signals according to the parameters of the underlying system. Thus, in this scenario, in order to obtain the largest possible segments, it is of interest to identify an SARX model that explains the observed data with the minimum number of switches. Examples of application of this framework to video or signal segmentation can be found in [17] and [107], where it is assumed that the data was corrupted by processing noise. Indeed, to the best of our knowledge, the problem of SARX identification with minimum number of switches in the presence of measurement noise has not been addressed yet. As we briefly show in the sequel, this scenario can be accommodated by our framework with minor modifications. The key observation is to note that, for two consecutive time instants, \( \sum_{i=1}^{N_s} s_{i,t} s_{i,t+1} \in \{0, 1\} \), and \( \sum_{i=1}^{N_s} s_{i,t} s_{i,t+1} = 1 \).
CHAPTER 7. EIV IDENTIFICATION OF SARX SYSTEMS

holds if and only if the same submodel $G_i$ is active. Thus, the total number of switches is given by $\sum_{t=0+n_a}^{T-1} \sum_{i=1}^{N_s} s_{i,t} s_{i,t+1}$. It follows that identification with minimum switches can be achieved by solving

$$
\begin{align*}
\text{minimize} & \quad \sum_{t=0+n_a}^{T-1} \sum_{i=1}^{N_s} s_{i,t} s_{i,t+1} \\
\text{subject to} & \quad \text{constraints in } (7.6)
\end{align*}
$$

(7.16)

Note that the running intersection property still holds, albeit in this case the variables are partitioned into $T - t_0 - n_a$ sets, each containing the variables associated with two consecutive time instants $\{\forall_{i=1}^{N_s} : r_i, \forall_{i=t-n_a}^{T} : \eta_k, s_i, s_{i+1}\}$, leading to increased computational complexity (due to fewer semi-definite constraints but of larger size). An alternative formulation that avoids this difficulty, but at the price of potentially introducing more submodels is given by

$$
\begin{align*}
p^*_s &= \text{minimize} \sum_{t=0+n_a}^{T-1} s_t \\
\text{subject to} & \quad \forall_{t=0+n_a}^{T} : y_t - \eta_t - r_t^T x_t + \sum_{k=1}^{n_a} a_{k,t} \eta_{t-k} - \sum_{k=1}^{n_b} b_k w_{t-k} = 0 \\
\forall_{t=0}^{T} : & \quad \|w_t\|_\infty \leq \epsilon_1, \|\eta_t\|_\infty \leq \epsilon_2 \\
\forall_{t=0+n_a}^{T-1} : & \quad (1 - s_t)(r_t - r_{t+1}) = 0 \\
\forall_{t=0+n_a}^{T-1} : & \quad s^2_t = s_t
\end{align*}
$$

(7.17)

where $r_t = [a_{1,t}, \cdots, a_{n_a,t}, b_{1,t}, \cdots, b_{n_b,t}]^T$ represents the unknown coefficients of the ARX model at time instant $t$, and no bounds are imposed on the number of submodels.

Obviously, both the objective and constraints in (7.17) can be partitioned into $T - t_0 - n_a$ partitions as

$$
p_s = \sum_{t=0+n_a}^{T-1} p_{s,t} \text{ with } p_{s,t} = s_t, \quad K_s = \bigcap_{t=0+n_a}^{T-1} K_{s,t} \text{ with }
$$

$$
\begin{align*}
& \forall_{t_0+n_a}^t : y_t - \eta_t - r_t^T x_t + \sum_{k=1}^{n_a} a_{k,t} \eta_{t-k} - \sum_{k=1}^{n_b} b_k w_{t-k} = 0 \\
& \forall_{t_0}^t : \|w_t\|_\infty \leq \epsilon_1 \\
& \forall_{t_0+n_a}^{t+1} : \|\eta_t\|_\infty \leq \epsilon_2 \\
& (1 - s_t)(r_t - r_{t+1}) = 0 \\
& s^2_t = s_t
\end{align*}
$$

with each partition $\{p_{s,t}, K_{s,t}\}$ associated with two consecutive time instants, containing variables in
CHAPTER 7. EIV IDENTIFICATION OF SARX SYSTEMS

the subset $v_t = \{s_t, r_t, r_{t+1}, v^t_{k=t-n_b} : w_k, v^{t+1}_{k=t-n_a} : \eta_k \}$. Due to the fact that

$$v_{t+1} \cap \bigcup_{t=t_0+n_a}^t v_t = \{r_{t+1}, v^{t+1}_{k=t-n_a+1} : \eta_k \} \subseteq v_t, \text{ for } \forall t = t_0 + n_a, \ldots, T - 2$$

the running intersection property holds. Thus, the nonconvex problem (7.17) can be relaxed into a sequence of convex problems in the form of

$$p^*_{m,N} = \min_{m_{t,2N}} \sum_{t=t_0+n_a}^{T-1} m(s_t)$$

subject to

$$\forall t = t_0+n_a : M_{t,N}(m_{t,2N}) \succeq 0$$

$$\forall t = t_0+n_a : L_{t,N-1}(g_k m_{t,2N}) \succeq 0, \forall g_k \in K_{s,t}$$

(7.18)

where $M_{t,N}$ and $L_{t,N-1}$ denote the moment matrix and localizing matrix associated with the moments sequence of variables in the subset $v_t$ of order up to $2N$. Since the minimum number of switches is equivalent to the sparse solution of the binary variables $s_t$, therefore, using the reweighted heuristic proposed in [89], it leads to Algorithm 10 outlined below. Consistent numerical examples show that when we use moment relaxation of order $N = 2$, Algorithm 10 can provide us a sparse solution to $s_t$ and a rank-1 moment matrix associated with $r_t$, but there has not been any theoretical support yet.

Algorithm 10 Moments-Based Segmentation

1: Initialize: $k = 0, 0 < \delta \ll 1, w_t^{(0)} = 1, \forall t = t_0 + n_a, \ldots, T - 1$
2: repeat
3: Solve

$$\{m(s_t)^{(k)}\} = \arg \min_{m_{t,2N}} \sum_{t=t_0+n_a}^{T-1} w_t^{(k)} m(s_t)$$

subject to

$$\forall t = t_0+n_a : M_{t,2}(m_{j,4}) \succeq 0$$

$$\forall t = t_0+n_a : L_{t,2}(g_k m_{t,4}) \succeq 0, \forall g_k \in K_{s,t}$$

(7.19)

4: Update

$$w_t^{(k+1)} = \left[ m(s_t)^{(k)} + \delta \right]^{-1}$$

$$k = k + 1$$

5: until $m(s_t)^{(k)} - (m(s_t)^{(k)})^2 \to 0, \forall t = t_0 + n_a$. 

101
7.5 Experiments

In this section, we illustrate the advantages of the proposed approach using several academic examples.

7.5.1 Identification of WARX Models of Equation Error Structure

This example is aimed to show that the switched linear model fitting algorithm proposed in Chapter 4 can be used to identify SARX models with equation error structure. The 30 × 40 image shown in Figure 7.1 was generated by the following switched-coefficient difference equation:

\[
y(i, j) = a_{1,1}(\sigma_{i,j})y(i-1, j-1) + a_{0,1}(\sigma_{i,j})y(i, j-1) + a_{1,0}(\sigma_{i,j})y(i-1, j) + c_{0,0}(\sigma_{i,j})u(i, j) + \eta(i, j)
\]

with

\[
\sigma_{i,j} = \begin{cases} 
1, & \text{if } (i - 15)^2 + (j - 30)^2 < 60 \text{ holds} \\
2, & \text{otherwise}
\end{cases}
\]

where \(a_{1,1}(1) = 0.3\), \(a_{0,1}(1) = -0.3\), \(a_{1,0}(1) = -0.3\), \(a_{1,1}(2) = -0.5\), \(a_{0,1}(2) = 0.5\), \(a_{1,0}(2) = 0.7\), and \(c_{0,0}(1) = c_{0,0}(2) = 1\). The boundary conditions and the input \(u\) were generated randomly from a zero-mean, unit-variance Gaussian distribution. The output \(y\) was corrupted by random processing noise, \(\eta\), uniformly distributed in \([-0.2, 0.2]\). The inputs and outputs comprise 1200 sample points of dimension 5, \([y(i-1, j-1), y(i, j-1), y(i-1, j), u(i, j), y(i, j)]^T\). Running Algorithm 4 we obtain the coefficients vectors for two submodels given by

\[
r_1 = \begin{bmatrix} 0.3157 & -0.2727 & -0.2790 & 0.9636 & -1.0000 \end{bmatrix}
\]

\[
r_2 = \begin{bmatrix} -0.4990 & 0.4990 & 0.7003 & 0.9894 & -1.0000 \end{bmatrix}
\]

The labeling error rate, \(err_1\), defined as

\[
err_1 = \frac{\text{Number of Points with Different Labels from the Ground Truth}}{\text{Total Number of Sample Points } N_p} \times 100\% = 1.67\%. \quad (7.20)
\]
7.5.2 General Identification of WARX Models of EIV Structure

For this example, the data are generated by the SARX model (7.3) defined by \( n_a = 2, n_b = 1 \) and \( N_s = 2 \), and \( \sigma_t \in \{1, 2\} \). We take 60 measurements with 4 switches, the input signal \( u_t \) is generated randomly in the range of \(-2\) to \(2\), and the measurement noise \( \eta_t \) is of the uniform random magnitude within \([0.8\epsilon, \epsilon]\), \( \epsilon = 0.20 \). The signal-to-noise ratio is 22.00dB. The ground truth and the identified results of the parameters associated with the model are given in Table 7.1, and the identified discrete mode \( \sigma_t \) is plot in Figure 7.2, which is compatible with the ground truth.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Ground Truth</th>
<th>( \epsilon = 0.20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1(1) )</td>
<td>1.4000</td>
<td>1.4000</td>
</tr>
<tr>
<td>( a_2(1) )</td>
<td>-0.7400</td>
<td>-0.7396</td>
</tr>
<tr>
<td>( b_1(1) )</td>
<td>1.0000</td>
<td>0.9992</td>
</tr>
<tr>
<td>( a_1(2) )</td>
<td>-0.5000</td>
<td>-0.5082</td>
</tr>
<tr>
<td>( a_2(2) )</td>
<td>-0.0600</td>
<td>-0.0682</td>
</tr>
<tr>
<td>( b_1(2) )</td>
<td>1.0000</td>
<td>0.9963</td>
</tr>
</tbody>
</table>

Figure 7.1: Textured Image, Ground Truth of Labels, and Labels Obtained by Algorithm

Figure 7.2: Identified Active Mode \( \sigma_t \) vs. Ground Truth for Example

Table 7.1: Results for Example 7.5.2

103
7.5.3 Additional A-Priori Information

For this example, the data are generated by the same SARX model as in Example 7.5.2, with the additional constraint that transitions from $G_2$ to $G_1$ are forbidden. The ground truth and the results running Algorithm 9 with and without the additional constraint are shown in Table 7.2 and Figure 7.3. Note that as illustrated in Figure 7.3, if the constraint $s_{2,t}^t s_{1,t+1} = 0$ is not enforced, the solution found may be not compatible with the a-priori information.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Ground Truth</th>
<th>without Constraints</th>
<th>with Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1(1)$</td>
<td>-0.5000</td>
<td>-0.5016</td>
<td>-0.4153</td>
</tr>
<tr>
<td>$a_2(1)$</td>
<td>-0.0600</td>
<td>-0.0445</td>
<td>-0.0151</td>
</tr>
<tr>
<td>$b_1(1)$</td>
<td>1.0000</td>
<td>1.0394</td>
<td>0.9230</td>
</tr>
<tr>
<td>$a_1(2)$</td>
<td>1.4000</td>
<td>1.3667</td>
<td>1.1702</td>
</tr>
<tr>
<td>$a_2(2)$</td>
<td>-0.7400</td>
<td>-0.6936</td>
<td>-0.6706</td>
</tr>
<tr>
<td>$b_1(2)$</td>
<td>1.0000</td>
<td>0.9564</td>
<td>0.8314</td>
</tr>
</tbody>
</table>

Figure 7.3: Identified Active Mode $\sigma_t$ vs. Ground Truth for Example 7.5.3

7.5.4 Segmentation of ARX Models

For this example, the data is generated by the SARX model (7.3) with $n_a = 2$, $n_b = 1$, $N_s = 3$, driven by an input $u$ uniformly distributed in $[-2, 2]$. The parameters for each submodel are $a_1(1) = -1.8$, $a_2(1) = -0.81$, $b_1(1) = 1.5$, $a_1(2) = -0.5$, $a_2(2) = -0.06$, $b_1(2) = 1$, an $a_1(3) = 1.4$, $a_2(3) = -0.74$, $b_1(3) = 1$. The measurements are corrupted by $\eta_t$ with magnitudes uniformly distributed in $[0.8\epsilon, \epsilon]$, $\epsilon = 0.20$, and the discrete mode $\sigma_t$ is $\sigma_t = 1$, $t \in [2, 13] \cup [50, 61]$; $\sigma_t = 2$, $t \in [14, 25] \cup [38, 49]$; and $\sigma_t = 3$, $t \in [26, 37]$. The signal-to-noise ratio is 24.10dB. As shown in Figure 7.4, Algorithm 10 yields the correct segmentation. Further, the parameters of the identified models are quite smooth and close to the ground truth.
CHAPTER 7. EIV IDENTIFICATION OF SARX SYSTEMS

Figure 7.4: Detection of Switches $s_t$ and Parameters $r_t$ for Example 7.5.4

7.6 Conclusions

In this chapter, we propose a novel general framework of switched linear model fitting from noisy dynamical data, i.e., identifying SWAX models from experimental temporal sequence in the presence of noise. The main idea is to recast the problem into a nonlinear polynomial optimization form, and then relax into a sequence of convex semi-definite programs. For the case where the noise affects the model in equation error structure, the problem can be solved in the same way for static data discussed in Chapter [4]. For the case where the noise affects the model in error-in-variables fashion, the problem becomes more difficult due to the fact that the noise term affects more than one regression equality, thus, a minimum relaxation order $N = 2$ is required, increasing the computational complexity. Combining the convex surrogates for low rank constraint and sparsity constraint, we provide an algorithm to obtain a single atomic measure, whose first order moments sequence is the exact solution to the original nonconvex problem. A salient feature of the proposed approach is its ability to handle available a-priori information such as constraints on switch sequence, relative frequency of each submodel or existence of pre-annotated data. Finally we also develop a simple algorithm to segment the EIV temporal sequence with minimum number of switches by moments-based convex relaxation, but so far there is no theoretical proof that Algorithm [10] could guarantee to find a feasible solution to (7.18). In the future, research effort will focus on verifying this observation or showing it is wrong by more numerical experiments and theoretical derivation.
Chapter 8

Identification of LPV Systems with LFT Parametric Dependence

8.1 Motivation

In Chapter 6 and Chapter 7, we considered the identification of systems which can be modeled as a finite number of linear regression submodels. In contrast, Linear Parameter Varying (LPV) systems, defined as linear systems where either the matrices in state equations or coefficients in input-output difference equations depend on time varying parameters (also called scheduling variables) which are available for measurement, can be considered as a new type of switched system switching among an infinite number of submodels, and the continuous rather than discrete scheduling variables govern the switching. LPV systems are interesting in their own, since they provide a formalization of “gain scheduling” ideas commonly used in engineering, and as tractable descriptions of nonlinear dynamics. While there is now a large body of tools for synthesizing controllers for LPV systems, one of the key issues that needs to be addressed before these tools can be applied to practical scenarios, is the development of identification methods capable of extracting the appropriate models from experimental data.

Several identification methods have been proposed for both input-output and state space models (see [108], [109], [110], [111], [112], [113], [114], [115], [116] and references therein). Some of these techniques, such as the subspace identification methods for state-space descriptions of affine LPV systems proposed in [111] and [117], assume that the scheduling parameter enters the plant affinely. In [111] the identification is performed in one global experiment in which both the control
CHAPTER 8. IDENTIFICATION OF LPV SYSTEMS WITH LFT PARAMETRIC DEPENDENCE

input and the scheduling variables are persistently excited simultaneously, local experiments are adapted in [17], that is, several LTI models are obtained independently by subspace identification method from data collected under different constant scheduling variables but persistently exciting control input, and the LPV model is achieved by interpolating these LTI models.

The case of non-affine parameter dependence has been addressed mainly in the context of LPV systems described by input-output difference equations. In this case, a standard approach to overcome the difficulty caused by a nonlinear dependence in the scheduling variables is to parameterize the LPV model as an affine combination of a set of basis functions known a priori [13, 15]. To mitigate the structural bias caused by inaccurate basis selection and the effects of over-parametrization, [13] and [14] proposed a Least-Squares Support Vector Machines (LS-SVM) framework, integrating a mechanism for automatically selecting basis. While in most of the papers mentioned above, the measurements are assumed to be corrupted by stochastic noise, [15], [16] consider set-membership identification for cases with deterministic bounded measurement noise. By applying Carathéodory-Fejér interpolation theory, [15] recasts the problem as linear matrix inequality feasibility problems. Finally, [16] considers the case where measurements of both the outputs and the scheduling variables are corrupted by bounded noise, leading to a nonconvex polynomial optimization problem, solved using moments-based convex relaxations [18].

The main drawback of the methods mentioned above is that the resulting models are not in a form that can be readily used for controller synthesis. Indeed, while in principle existing synthesis methods can handle general parameter dependencies, from a practical standpoint, this requires gridding the parameter space and solving a large number of linear matrix inequalities (see for instance [19]). On the other hand, the case where the parameter enters the plant in a feedback form leads to very efficient, “self-scheduled” controllers that can be synthesized solving a problem with roughly the same complexity as in the LTI case [20]. Unfortunately, in contrast to the wealth of results in identification of affine dependent LPV systems, there are just a few results on identification of LPV systems under LFT parametric dependence. In [10] the identification of SISO LFT-LPV systems with a single scheduling variable is reduced to recursive least squares under the assumption of full state measurements. In [10] a prediction error method is presented to minimize the cost function using gradient and Hessian based nonlinear optimization algorithms, which are sensitive to initial conditions.

Motivated by these difficulties and weaknesses, in this chapter we propose a novel framework to identify SISO LPV systems with LFT dependence on a single scheduling variable from experimental data. Our main result shows that, surprisingly in view of the LFT structure, if the
CHAPTER 8. IDENTIFICATION OF LPV SYSTEMS WITH LFT PARAMETRIC DEPENDENCE

scheduling parameter can be suitably manipulated during the experiments, then, in the noiseless case the identification problem is convex, while in the presence of noise tractable convex relaxations can be obtained. Based on these theoretical results, in the second portion of the chapter we develop a computationally efficient identification algorithm, based on recent results on atomic norm minimization [32]. These results are illustrated with a simple example showing the effectiveness of the proposed approach.

8.2 Problem Statement

Consider the SISO LPV system shown in Figure 8.1, in which the “forward” part is a LTI model with a minimal state space realization:

$$\begin{bmatrix} x_{k+1} \\ s_k \\ y_k \end{bmatrix} = \begin{bmatrix} A & b_1 & b_2 \\ c_1 & d_{11} & d_{12} \\ c_2 & d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} x_k \\ r_k \\ u_k \end{bmatrix} \tag{8.1}$$

and the “feedback” part is represented by

$$r_k = \rho_k s_k, \tag{8.2}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}$, $y_k \in \mathbb{R}$ and $\rho_k \in \mathbb{R}$ denote the state, input, output and the scheduling variable, respectively. The measurements of the output is corrupted by additive noise, i.e., $\tilde{y}_k = y_k + \eta_k$.

Eliminating $r_k$ and $s_k$ which are not available for measurements, the LPV system above can be described as the linear time varying system of the form

$$\begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix} = \begin{bmatrix} \bar{A}_k & \bar{b}_k \\ \bar{c}_k & \bar{d}_k \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix} \tag{8.3}$$
where the state matrices
\[
\begin{bmatrix}
\bar{A}_k & \bar{b}_k \\
\bar{c}_k & \bar{d}_k
\end{bmatrix} = \begin{bmatrix} A & b_2 \\ c_2 & d_{22} \end{bmatrix} + \frac{\rho_k}{1 - d_{11} \rho_k} \begin{bmatrix} b_1 \\ d_{21} \end{bmatrix} \begin{bmatrix} c_1 \\ d_{12} \end{bmatrix}
\]
are LFT parametric dependent. In the sequel, we will make the following assumptions:

A.1- The feasible set for the scheduling variable is of the form
\[
\mathcal{P}(\rho_{\min}, \rho_{\max}) = \{\rho \in \mathbb{R} : \rho_{\min} \leq \rho \leq \rho_{\max}\};
\]

A.2- While performing experiments, the scheduling variable \(\rho\) can be set to any value in \(\mathcal{P}(\rho_{\min}, \rho_{\max})\);

A.3- \(0 \in \mathcal{P}(\rho_{\min}, \rho_{\max})\);

A.4- \(1 - d_{11} \rho \neq 0\) for all \(\rho \in \mathcal{P}(\rho_{\min}, \rho_{\max})\);

A.5- A bound on the measurement noise \(\eta_k\) is available, e.g. \(|\eta_k| \leq \epsilon\), where \(\epsilon\) is known a priori.

In this context, the problem considered in this chapter can be stated formally as:

**Problem 10.** Given experimental data \(\{u_k, \tilde{y}_k, \rho_k\}\), identify a causal stable LPV system with the structure given above and such that its trajectories interpolate the experimental data, or show that none exist.\(^1\)

**Remark 3.** Assumption A.5 is standard in set membership identification. Assumption A.4 guarantees that the system is well posed for all possible parameter trajectories. Finally, Assumption A.3 is made for notational simplicity and it does not entail any loss of generality if A1 holds. This follows from the fact that one can always define a new parameter \(\tilde{\rho} = \rho - \rho_o, \rho_o = \frac{\rho_{\min} + \rho_{\max}}{2}\), having a feasible set
\[
\tilde{\mathcal{P}} = \{\tilde{\rho} \in \mathbb{R} : \tilde{\rho}_{\min} \leq \tilde{\rho} \leq \tilde{\rho}_{\max}\}, 0 \in \tilde{\mathcal{P}},
\]
where \(\tilde{\rho}_{\min} = \frac{\rho_{\min} - \rho_{\max}}{2}\) and \(\tilde{\rho}_{\max} = \frac{\rho_{\max} - \rho_{\min}}{2}\). Furthermore, the LPV system in Figure 8.1 with \(\rho\) as the scheduling variable is equivalent to the new LPV system with \(\tilde{\rho}\) as the scheduling variable, as shown in Figure 8.2.\(^1\)

\(^1\)In this case the a-priori information about the noise and structure of the system are invalidated by the experiments.
CHAPTER 8. IDENTIFICATION OF LPV SYSTEMS WITH LFT PARAMETRIC DEPENDENCE

8.3 LPV Identification via Convex Optimization

In this section we present the main result of this chapter: a convex optimization approach to control oriented identification of LPV systems with LFT dependence on the parameter. In order to develop these ideas, we begin by considering the simpler, noiseless case and then extend the approach to noisy scenarios.

8.3.1 Identification with noiseless measurements

Surprisingly, in the noiseless case, Problem [10] is convex, in spite of the feedback structure. Specifically, we have the following result.

**Theorem 15.** Under assumptions A1-A5, in the noiseless case (e.g. \( \epsilon = 0 \)), Problem [10] can be solved via convex optimization.

**Proof.** The proof is constructive, based on performing several experiments to sequentially identify the different components of the system. Consider again the LPV system in Figure [8.1]. Let

\[
\begin{bmatrix}
S(z) \\
Y(z)
\end{bmatrix} =
\begin{bmatrix}
P_{11}(z) & P_{12}(z) \\
P_{21}(z) & P_{22}(z)
\end{bmatrix}
\begin{bmatrix}
R(z) \\
U(z)
\end{bmatrix}
\] (8.7)
CHAPTER 8. IDENTIFICATION OF LPV SYSTEMS WITH LFT PARAMETRIC DEPENDENCE

and

\[
\begin{align*}
    s &= T_{11}^N r + T_{12}^N u \\
    y &= T_{21}^N r + T_{22}^N u
\end{align*}
\]

(8.8a)  (8.8b)

declare the transfer function description of the “forward” part of the system in the z and time domain respectively, where \( T_{ij}^N \) denotes the toeplitz matrix of the \( N \)-length truncated impulse response \( h_{ij} \) associated with \( P_{ij}(z) \), for \( i, j = 1, 2 \).

**Step 1: Identification of** \( P_{22}(z) \). This can be easily accomplished by performing experiments while setting \( \rho_k = 0 \). Since in this case \( r_k = 0 \), the input/output mapping simply reduces to

\[
Y(z) = P_{22}(z)U(z)
\]

(8.9a)

\[
y^{(1)} = T_{u^{(1)}} h_{22}
\]

(8.9b)

from where \( P_{22}(z) \) can be identified, using for instance the atoms based approach outlined in Section 2.4.

**Step 2: Identification of** \( P_{11}(z) \) and \( P_{12}(z)P_{21}(z) \). To identify these transfer functions, we perform experiments where \( \rho_k \) is kept at a nonzero constant \( \rho \). In this case the internal signals satisfy \( r = \rho s \), leading to:

\[
y^{(2)} - T_{22}^N u^{(2)} = \rho T_{21}^N s = \rho T_{21}^N (T_{11}^N r + T_{12}^N u^{(2)})
\]

\[
= \rho T_{11}^N T_{21}^N r + \rho T_{21}^N T_{12}^N u^{(2)}
\]

\[
= \rho T_{11}^N (y^{(2)} - T_{22}^N u^{(2)}) + \rho T_{21}^N T_{12}^N u^{(2)}.
\]

(8.10)

Define the auxiliary signal \( \bar{r} = y^{(2)} - T_{22}^N u^{(2)} \). Since \( T_{22}^N \) is known from Step 1, it follows that \( \bar{r} \) is also known. Rewriting (8.10) in terms of \( \bar{r} \) yields:

\[
\bar{r} = \rho T_{11}^N \bar{r} + \rho T_{21}^N T_{12}^N u^{(2)} = \rho T_{11}T_{h_{12}} + \rho T_{u^{(2)}h_{12,21}}
\]

(8.11)

where \( h_{12,21} \) denotes the \( N \)-length truncated impulse response of the system represented by the transfer function \( P_{12}(z)P_{21}(z) \), i.e., \( h_{12,21} = h_{12} * h_{21} \). Note that (8.11) consists of \( N \) equations in \( 2N \) unknowns. Thus, we need to collect at least two sets of data \( \{u_1^{(2)}, y_1^{(2)}, \rho_1\}, \{u_2^{(2)}, y_2^{(2)}, \rho_2\} \), by persistently exciting the system with a control input while holding \( \rho \) at different constants. Using data from these experiments, \( P_{11}(z) \) (\( h_{11} \)) and the product \( P_{12}(z)P_{21}(z) \) (\( h_{12,21} \)) can be determined by
solving

\[
\begin{align*}
\bar{r}_1 &= y_1^{(2)} - T_{22}^N u_1^{(2)} \\
\bar{r}_2 &= y_2^{(2)} - T_{22}^N u_2^{(2)} \\
\bar{r}_1 &= \rho_1 T_{r_1} h_{11} + \rho_1 T_{u_1}^{(2)} h_{12,21} \\
\bar{r}_2 &= \rho_2 T_{r_2} h_{11} + \rho_2 T_{u_2}^{(2)} h_{12,21}.
\end{align*}
\]

(8.12)

Step 3: Identification of \( P_{12}(z) \) and \( P_{21}(z) \). Now consider an experiment where both the scheduling variable \( \rho \) and the input \( u \) are persistently excited simultaneously for \( N \) time instants. In this case

\[
r = \Lambda s
\]

(8.13)

where \( \Lambda = \text{diag}[\rho_0, \rho_1, \cdots, \rho_{N-1}] \). Combining (8.8) and (8.13) yields:

\[
\begin{align*}
\begin{cases}
\bar{r} = (I - \Lambda T_{11}^N)^{-1} \Lambda T_{12}^N u^{(3)} \\
y^{(3)} = T_{21}^N (I - \Lambda T_{11}^N)^{-1} \Lambda T_{12}^N u^{(3)} + T_{22}^N u^{(3)}
\end{cases}
\end{align*}
\]

(8.14a, 8.14b)

Multiplying both sides of (8.14b) by \( T_{12}^N \) leads to:

\[
T_{12}^N y^{(3)} = T_{12}^N T_{21}^N (I - \Lambda T_{11}^N)^{-1} \Lambda T_{12}^N u^{(3)} + T_{12}^N T_{22}^N u^{(3)}.
\]

(8.15)

from where it follows that (up to a constant) \( h_{12} \) is the nonzero vector in the null space of \( T_{y^{(3)}} - T_{12}^N T_{21}^N (I - \Lambda T_{11}^N)^{-1} \Lambda T_{u^{(3)}} - T_{u} \), where we have defined \( \bar{u} = T_{22}^N u^{(3)} = T_{u^{(3)}} h_{22} \). Finally, \( h_{21} \) can be directly determined by solving the deconvolution

\[
h_{12,21} = T_{12}^N h_{21}.
\]

(8.16)

\[
\Box
\]

8.3.2 The noisy measurements case

We consider now the more realistic case where only noisy measurements \( \tilde{y} = y + \eta \) of the output are available. In the presence of noise, equations (8.9b), (8.12), (8.14) and (8.16) have now the form:
\[ \hat{y}^{(1)} = T_u^{(1)} h_{22} + \eta^{(1)} \] (8.17a)

\[ \tilde{r}_i = \hat{y}^{(2)}_i - T_{22}^N u^{(2)}_i \] (8.17b)

\[ \tilde{r}_i - (I - \rho_i T_{11}^N) \eta^{(2)}_i = \rho_i T_{r_i} h_{11} + \rho_i T_{u_i} h_{12,21} \] (8.17c)

\[ T_{12}^N (\hat{y}^{(3)} - \eta^{(3)}) = T_{12}^N T_{21}^N (I - \Delta T_{11}^N)^{-1} \Delta T_{12}^N u^{(3)} + T_{12}^N T_{22}^N u^{(3)} \] (8.17d)

Thus, the problem is no longer convex, due to the bilinear terms involving the product of the Toeplitz matrices of the system and noise. However, as we show in the next section, tractable convex relaxations are readily available.

### 8.4 Proposed Algorithms

In this section, we propose two convex relaxations to solve the problem formulated in Section 8.3.2.

#### 8.4.1 A Single-Step Algorithm

Equations (8.17), together with the constraint \( \| \eta \|_\infty \leq \epsilon \) form a set of polynomial equations that can be solved using a convergent sequence of convex relaxations, obtained using moments (or sum-of-squares) arguments \[24, 118\]. A potential difficulty when pursuing this approach stems from the fact that, ideally, the impulse responses \( h_{ij} \) should be chosen such that the overall resulting system has low order (e.g. the matrix \( A \) in (8.1) has the lowest possible dimension). We propose to accomplish this by enforcing sparsity of the set of the atoms representing \( h_{11}, h_{12}, h_{21}, \) and \( h_{22} \) via a group LASSO argument \[121\], leading to the following regularized version of (8.17):

\[
\begin{align*}
\text{minimize} & \quad \| [c_{11}, c_{12}, c_{21}, c_{22}, \eta] \|_{2,1} \\
\text{subject to} & \quad \text{eqn. (8.17)}, \quad \| \eta \|_k \leq \epsilon, \text{ and} \\
& \quad h_{ij} = A c_{ij}, \forall i, j = 1, 2.
\end{align*}
\]

Since the problem above is still semi-algebraic, it can be solved using the techniques in \[24, 118\].
8.4.2 An Efficient Multi-Step Relaxation

Although theoretically the algorithm proposed in Section 8.4.1 can find a low order system in a single step, it is only practical for moderately sized problems, due to the computational complexity entailed in solving the corresponding sequence of SDP relaxations. In this section we present a computationally efficient alternative, based on the introduction of fictitious noise in (8.17) and softening of the noise bound constraint, that allow for solving the problem via a multi-step procedure similar to the one presented in Section 8.3.1.

\textit{Step 1:} Estimate $h_{22}$ by solving
\[
\hat{h}_{22} = \arg \min_{h_{22}} \|h_{22}\|_A \quad \text{subject to} \quad \|\tilde{y}^{(1)} - T_{u(1)}h_{22}\|_\infty \leq \epsilon \tag{8.19}
\]

\textit{Step 2:} To eliminate the bilinear terms in (8.17c), replace the measurement noise $\eta_i^{(2)}$ with the fictitious noise $\tilde{\eta}_i^{(2)} = (I - \rho_i T_{11}^N)\eta_i^{(2)}$, relaxing (8.17c) to
\[
\tilde{\eta}_i^{(2)} = b - B_1 h_{11} - B_2 h_{12,21},
\]
with $\tilde{\eta}_i^{(2)} = \begin{bmatrix} \tilde{\eta}_1^{(1)} \\ \tilde{\eta}_1^{(2)} \end{bmatrix}$, $b = \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \end{bmatrix}$, $B_1 = \begin{bmatrix} \rho_1 T_{r_1} \\ \rho_2 T_{r_2} \end{bmatrix}$, and $B_2 = \begin{bmatrix} \rho_1 T_{u(1)}^{(1)} \\ \rho_2 T_{u(2)}^{(2)} \end{bmatrix}$. Then estimate $h_{11}$ and $h_{12,21}$ from the equation above by solving the convex problem
\[
\{\hat{h}_{11}, \hat{h}_{12,21}\} = \arg \min_{h_{11}, h_{12,21}} f(h_{11}, h_{12,21}) \quad \text{subject to} \quad f(h_{11}, h_{12,21}) = \frac{1}{2}\|b - B_1 h_{11} - B_2 h_{12,21}\|^2_2 \tag{8.21}
\]
\[
\|h_{11}\|_A \leq \tau_1 \text{ and } h_{12,21}\|_A \leq \tau_2,
\]

The problem above can be solved by modifying the original Franke Wolfe algorithm proposed in [32] to take into account the fact that the problem above involves two different atomic norm constraints, a feature handled by simply optimizing $h_{11}$ and $h_{12,21}$ alternatively as shown in Algorithm [11].

\textit{Step 3:} As in Step 2, introducing the fictitious noise $\tilde{\eta}_i^{(3)} = T_{12}^N \eta_i^{(3)}$ allows for relaxing (8.17d) to
\[
\tilde{\eta}_i^{(3)} = Bh_{12},
\]
\[\text{Recall that the number of variables growths combinatorially with the relaxation order.}\]
CHAPTER 8. IDENTIFICATION OF LPV SYSTEMS WITH LFT PARAMETRIC DEPENDENCE

with \( B = (T_{y(3)} - T_{12}T_{21}^N (I - \Lambda T_{11}^N)^{-1} \Lambda T_{u(3)} - T_u), \) from where \( h_{12} \) can be estimated by solving

\[
\hat{h}_{12} = \arg \min_{h_{12}} \| Bh_{12} \|_2^2 \\
\text{subject to } \| h_{12} \|_A \leq \tau_{12}, \ h_{12}(m) = 1
\]

(8.23)

where \( h_{12}(m) \) denotes the \( m \)-th entry of \( h_{12} \), and the constraint \( h_{12}(m) = 1 \) is added to remove gain ambiguity entailed in the problem\(^3\). Handling this extra equality constraint requires modifying the basic algorithm in [32] as follows:

\[
\hat{x} = \arg \min_{x} \| B(:,m) + Bx \|_2^2 \\
\text{subject to } \| x \|_A \leq \tau_{12}
\]

(8.24)

where \( B(:,m) \) denotes the \( m \)-th column of \( B \), and \( \hat{A} \) denotes the atom set resulted from setting the \( m \)-th entry of the impulse responses associated with all atoms in \( A \) to be zero. It is easy to show the \( m \)-th entry of \( \hat{x} \) is 0, and \( \hat{h}_{12} = \hat{x} + e_m \), where \( e_m \) is the unit vector with the \( m \)-th entry being 1.

**Step 4:** Estimate \( h_{21} \) by using standard atomic norm optimization to solve

\[
\hat{h}_{21} = \arg \min_{h_{21}} \frac{1}{2} \| h_{12,21} - T_{12} h_{21} \|_2^2 \\
\text{subject to } \| h_{21} \|_A \leq \tau_{21}
\]

(8.25)

**Step 5:** (Optional) Use balanced truncations [122] to the MIMO system obtained combining \( P_{11}(z), P_{12}(z), P_{21}(z), \) and \( P_{22}(z) \), to further reduce the order of the system.

**Remark 4.** When compared against the moments based approach of Section 8.4.1, the relaxation above has the advantage of substantially lower computational complexity, at the price of potentially introducing conservatism when replacing the actual noise with a fictitious one. Note however that the fictitious noise sequences introduced in steps 2 and 3 above satisfy: \( \| \eta_i^{(2)} \|_\infty \leq \| (I - \rho_i T_{11}^N)^{-1} \|_1 \| \tilde{\eta}_i^{(2)} \|_\infty \) and \( \| \eta_i^{(3)} \|_\infty \leq \| (T_{12}^N)^{-1} \|_1 \| \tilde{\eta}_i^{(3)} \|_\infty \). Since both \( I - \rho_i T_{11}^N \) and \( T_{12}^N \) are non singular, it follows that the relaxation above simply amounts to allowing for increased error when interpolating that experimental data (in the sense that if the relaxed problem is feasible for some \( \tilde{\eta}_i \), then the original one is feasible, with noise levels not exceeding the ones shown above).

\(^3\)Note that if the pair \( \{T_{12}, T_{21}\} \) solves Problem 1, so does \( \{\beta T_{12}, \frac{1}{\beta} T_{21}\} \) for any \( \beta \neq 0 \)
Algorithm 11 Atomic Norm Minimization Based LPV-LFT Systems Identification

1. **Initialization:** \( t \leftarrow 0, h_{11}^{(0)} \leftarrow \tau_1 a_{0,1}, h_{12,21}^{(0)} \leftarrow \tau_2 a_{0,2} \), and \( a_{0,2} \) are the impulse responses associated with randomly picked atoms in \( A \) defined by \((2.17)\).
2. **repeat**
3. Randomly generate samples \( p_t, 1, \ldots, n_s \) uniformly distributed in \( \mathbb{D}_p \), constructing the finite set of atom \( A_t = \{ a_1, \ldots, a_{n_s} \} \);
4. Update \( h_{11} \) by

\[
\begin{align*}
\alpha_{t,1} & \leftarrow \arg \min_{a \in [0,1]} f(h_{11}^{(t)} + \alpha(\tau a_{1}^{(t)} - h_{11}^{(t)}), h_{12,21}^{(t)}) ; \\
\alpha_{t,1} & \leftarrow \arg \min_{a \in [0,1]} f(h_{11}^{(t)} + \alpha(\tau a_{1}^{(t)} - h_{11}^{(t)})) ; \\
\end{align*}
\]

5. Update \( h_{12,21} \) by

\[
\begin{align*}
\alpha_{t,2} & \leftarrow \arg \min_{a \in [0,1]} f(h_{11}^{(t)} + \alpha(\tau a_{2}^{(t)} - h_{12,21}^{(t)})) ; \\
\alpha_{t,2} & \leftarrow \arg \min_{a \in [0,1]} f(h_{11}^{(t)} + \alpha(\tau a_{2}^{(t)} - h_{12,21}^{(t)})) ; \\
\end{align*}
\]

6. \( t \leftarrow t + 1 ; \)
7. **until** \( \nabla h_{11} f(h_{11}^{(t)} , h_{12,21}^{(t)}) = 0 \) and \( \nabla h_{11,21} f(h_{11}^{(t)} , h_{12,21}^{(t)}) = 0 \).

8.5 Experiments

In this section we illustrate the features of the proposed framework through a simple numerical example. The SISO LPV system used to generate data has the structure shown in Figure 8.1 with the “forward” part described by the state space realization

\[
\begin{bmatrix}
x_{k+1} \\
s_k \\
y_k
\end{bmatrix} =
\begin{bmatrix}
0.000 & 1.000 & 0.000 & 1.073 \\
-0.100 & 0.700 & 0.816 & 1.075 \\
0.524 & -0.625 & -0.500 & 0.500 \\
0.443 & 0.060 & 0.000 & 0.500
\end{bmatrix}
\begin{bmatrix}
x_k \\
r_k \\
u_k
\end{bmatrix},
\]

and the “feedback” part represented by

\[
r_k = \rho_k s_k, \quad \rho_k \in [-1, 1].
\]

We performed two sets of experiments: in the first experiment, the system \( G_1 \) was identified
CHAPTER 8. IDENTIFICATION OF LPV SYSTEMS WITH LFT PARAMETRIC DEPENDENCE

from clean experimental data by directly solving the equations (8.9b), (8.12), (8.14) and (8.16) as described in 8.3.1 in the second experiment, the output was corrupted by measurement noise in the form of

$$\tilde{y}_k = y_k + \eta_k, \ |\eta_k| \leq 0.15,$$

leading to a signal-to-noise ratio of 21.34dB. In this case, we applied the multi-step algorithm based on atoms described in 8.4.2 to identify a model $G_2$.

The experimental data used for identification is shown in Figure 8.3. Data within the time intervals $[1, 40]$, $[41, 360]$ and $[361, 400]$ were used in Step 1, Step 2, and Step 3, respectively. Equations (8.9b), (8.12), (8.14), (8.16) and (8.17) were solved using data within the time interval $[1 + (i - 1) \times N, i \times N]$, with $\forall i = 1, 2, \ldots, 10$ and $N = 40$. At $k = 1 + (i - 1) \times N$, the states of the system $x_k$ were reset to zero. Figure 8.3 shows the outputs of the identified models $G_1$ and $G_2$. As expected, $G_1$ exactly interpolated the experimental data, while $G_2$ did so within the a-priori noise bounds. In addition, running on an iMac with 3.4 GHz Intel Core i7, 32GB RAM, in each step the greedy algorithm terminated in $45 \sim 90$ iterations, with each iteration taking around 1.5 sec.

![Figure 8.3: Data for Identification](image)

Next, the identified systems were validated with experiments involving additional data, not used in the identification. This additional data was generated by exciting the system for 400 steps with a unit-magnitude pseudorandom binary input sequence $\{u_k\}$, while the scheduling variable $\rho$
followed a sinusoidal trajectory $\rho_k = \sin(2\pi k/100)$. The output sequences generated by the two identified models are shown in Figure 8.4. As illustrated there, the model identified from clean data, $G_1$, recovered the output virtually without error (the maximum interpolation error was $5.88 \times 10^{-9}$), supporting our theoretical analysis. As expected, the model identified from noisy data, $G_2$ resulted in a larger (but still moderate) prediction error. Finally, to measure the accuracy of the identified models, we considered the variance-accounted-for criterion (VAF) [123]

$$VAF = \max \left\{ 1 - \frac{\text{var}(y - \hat{y})}{\text{var}(y)}, 0 \right\} \times 100\%,$$

where $y$ is the output sequence and $\hat{y}$ is its estimate. The operator $\text{var}(\cdot)$ denotes the variance of the argument. For this experiment, the VAFs resulted from $G_1$ and $G_2$ are 100% and 99.39% respectively, proving the effectiveness of the proposed approach.

8.6 Conclusions

Many problems of practical interest involve designing controllers for parameter varying systems where the parameter enters the plant in a feedback form. While the design of controllers for such systems is well understood, the problem of identifying models that can be directly used
CHAPTER 8. IDENTIFICATION OF LPV SYSTEMS WITH LFT PARAMETRIC DEPENDENCE

for control is far less developed. Indeed, most of the techniques available in the literature for LPV identification consider affinely parametrized models, which entail a substantial computational burden when used for controller design. Motivated by this difficulty, in this chapter, we considered the identification problem for LPV systems in LFT form and proved that efficient, convex optimization based solutions can be obtained, provided that the scheduling parameter can be suitably manipulated. These results were illustrated with a simple example showing the ability of the proposed method to generate models that capture the LFT dependence. An issue that was not addressed in this chapter is that of obtaining the worst case bounds on the identification error. Research is currently under way seeking to develop these bounds by exploiting ideas from Information Based Complexity [124].
Chapter 9

Model (In)validation of SARX Systems with Unknown Switches

9.1 Motivation

As pointed out in Chapter 7, due to NP-hard nature of the problem of identifying SARX models, a large portion of research efforts in this area has been directed towards developing computationally efficient algorithms, either formulating convex problems resorting to heuristics and relaxation techniques (see for instance [18]) or seeking for solutions to nonconvex formulation as in [16, 19, 21] or the method we proposed in Chapter 7. Despite their empirical success in solving some problems, the equivalence between the original problem and its relaxed version or the performance of the local optimizer is not easy to prove, especially in cases with noisy data. Thus, a key step before using the resulting models, is to (in)validate them using additional experimental data.

In the case where the mode variable can be directly measured, the problem is closely related to that of validating Linear Parameter Varying (LPV) models and can be solved by techniques similar to those proposed in [125, 126]. However, in many practical situations, the mode variable is not directly available. Examples include, amongst others, fault detection problems and general data segmentation problems arising in the context of systems biology and video analytics. The case where the discrete mode variable must be inferred from the measured, noisy data, is considerably less developed. In [127], necessary and sufficient conditions for SARX models to be (in)validated by extra experimental data are proposed, based upon the idea of recasting the problem into one of checking emptiness of a semi-algebraic set. In turn, as shown there, this condition can be checked by
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

solving a sequence of convex optimization problems involving matrices of increasing size until either an optimal positive objective is found or the rank of certain matrices formed using the solution ceased to increase. The former case provides an invalidation certificate, while the latter indicates that the observed experimental data is consistent with the model under consideration. While this approach has been shown to be effective with relatively small sized problems, its computational complexity increases with the problem size, and there is no upper bound on the worst-case computational complexity. In addition, in its present form, it cannot accommodate constraints on the trajectory of the mode variable, that is, the underlying graph is assumed to be complete.

To circumvent these difficulties, in this chapter we present an alternative framework for model (in)validation of SARX models. Contrary to previous work based on exploiting the so-called hybrid decoupling constraints, the present approach relies on the similar technique in Chapter 4, introducing a set of sparse binary variables $s_{i,t}$ to indicate whether the $i$-th submodel is active at time instant $t$. In this context, the problem reduces to seeking solutions to a set of nonlinear inequalities such that $s_{i,t} \in \{0, 1\}$, or showing that no such solutions exist. Further, topological constraints on the underlying graph (e.g., sets of admissible transitions) translate into simple linear constraints on these binary variables.

Since the problem above is NP-hard, in Section 9.3.2, we develop a linear programming based relaxation by using a (weighted) $\ell_1$ norm as a surrogate for cardinality \cite{89, 128, 129}. This relaxation is computationally efficient and can handle large scale problems. However, in some cases it may fail to establish whether or not the model is (in)validated by the experimental data. In order to handle these cases, in Section 9.3.3, we develop an alternative convex relaxation, based on the solution of a constrained polynomial optimization problem using moments-based techniques introduced in Section 2.2. The main result of this section shows that this relaxation is guaranteed to provide a necessary and sufficient (in)validation certificate, using moment matrices of order at most $T - n_a + 1$, where $T$ and $n_a$ denote the number of experimental data and the order of the model, respectively. In Section 9.4, we show how to incorporate the constraints on the switching sequence in our proposed framework. Finally in Section 9.5, these results are illustrated using both academic examples and a nontrivial problem arising in computer vision: detecting abnormal activities using video data.

\footnote{Due to its special structure, the problem does not satisfy, in general, the restricted isometry property. Thus, there is no guarantee that minimizing the $\ell_1$ norm will indeed find the sparsest solution.}
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

9.2 Problem Statement

In this chapter we consider multi-input multi-output (MIMO) SARX model of the form:

\[
\begin{align*}
\xi_t &= \sum_{k=1}^{n_a} A_k(\sigma_t) \xi_{t-k} + \sum_{k=1}^{n_c} C_k(\sigma_t) u_{t-k} + f(\sigma_t) \\
y_t &= \xi_t + \eta_t
\end{align*}
\] (9.1)

where \(u_t \in \mathbb{R}^{n_u}, y_t \in \mathbb{R}^{n_y}\) and \(\sigma_t \in \mathbb{N}_{N_s} = \{1, \ldots, N_s\}\) denote the input, the measured output corrupted by the noise \(\eta_t \in \mathbb{R}^{n_y}\), and the discrete mode variable, respectively. Since in this paper we do not make any dwell-time assumptions, the mode signal \(\sigma_t\) can switch arbitrarily amongst subsets of the \(N_s\) subsystems \(G_i\), associated with the coefficient matrices \(\{\forall_{k=1}^{n_a} A_k(i), \forall_{k=1}^{n_c} C_k(i), f(i)\}\). In this context, the model invalidation problem of interest, illustrated in Figure 9.1, can be formally stated as follows:

![Figure 9.1: Problem Setup](image)

Figure 9.1: Problem Setup (The coefficient matrices of the submodels \(G_i\) and a bound on the noise are known a priori. The experimental data consists of input/output measurements, \(u_t\) and \(y_t\). The mode signal \(\sigma_t\) and noise sequence \(\eta_t\) are unknown.)

**Problem 11.** Given a nominal SARX model of the form (9.1) together with its \(N_s\) submodels \(G_1, \ldots, G_{N_s}\), an a priori bound \(\epsilon\) on noise, and experimental data \(\{u_t, y_t\}_{t=0}^{T-1}\), determine whether or not the a priori information and the a posteriori experimental data are consistent, i.e. whether the consistency set

\[
\mathcal{T}(\eta, \sigma) = \{||\eta_t||_\infty \leq \epsilon, \sigma_t \in \mathbb{N}_{N_s} \text{ subject to } (9.1) \forall t \in [0, T - 1]\}
\]

is nonempty.

**Remark 5.** The formulation above does not place constraints on the admissible transitions of the mode variable \(\sigma_t\) (equivalently, the underlying graph is assumed to be complete). In section 9.4 we will briefly indicate how to incorporate constraints on the set of admissible transitions.
9.3 Convex (In)Validation Certificates

In this section we present the main result of this paper: convex necessary and sufficient conditions for a MIMO SARX model to be (in)validated by experimental data. Firstly, we show that the problem can be reduced to a Mixed Integer Linear Program (MILP), which is in principle NP-hard. To circumvent this difficulty, we carry out our research in two directions. On the one hand, we relax the problem to Linear Program (LP) based on sparsification technique, which is of much lower computational complexity and provides a sufficient (in)validation certificate. On the other hand, we reformulate the problem to SDP problems based on moments theory based polynomial optimization mentioned in Section 2.2, which leads to convex sufficient and necessary certificates for the (in)validation problem considered in this paper.

9.3.1 A MILP reformulation

For notational simplicity, rewrite equation (9.1) as

\[
\sum_{k=1}^{n_a} A_k(\sigma_t)(y_{t-k} - \eta_{t-k}) - (y_t - \eta_t) \sum_{k=1}^{n_c} C_k(\sigma_t)u_{t-k} + f(\sigma_t) = g_{\sigma_t,t} - h_{\sigma_t,\eta_{t:t-n_a}} \tag{9.2}
\]

where we define \(h_{\sigma_t} = \begin{bmatrix} -I & A_1(\sigma_t) & \ldots & A_{n_a}(\sigma_t) \end{bmatrix}\), \(g_{\sigma_t,t} = -y_t + \sum_{k=1}^{n_a} A_k(\sigma_t)y_{t-k} + \sum_{k=1}^{n_c} C_k(\sigma_t)u_{t-k} + f(\sigma_t)\), and \(\eta_{t:t-n_a} = \begin{bmatrix} \eta_t^T & \ldots & \eta_{T-n_a}^T \end{bmatrix}^T\). Note that since the mode variable \(\sigma_t\) is not available for measurements, the actual subsystem \(G_{\sigma_t}\) that is active at any given time \(t\) is unknown. However, regardless of the value of \(\sigma_t\), the set of submodels \(\{G_{\sigma_i}\}_{i=1}^{N_s}\) given as part of \textit{a priori} information is not invalidated by the experimental data if and only if Eq. (9.2) holds true for some \(\sigma_t \in \{1, \ldots, N_s\}\) for each \(t \in \{0, \ldots, T - 1\}\). Clearly, this condition is equivalent to the existence of a set of binary variables \(s_{i,t}\) and an admissible noise sequence \(\eta_t\) such that

\[
\forall_{i=1}^{N_s} \forall_{t=n_a}^{T-1} : s_{i,t} (g_{i,t} - h_i \eta_{t:t-n_a}) = 0 \tag{9.3a}
\]

\[
\forall_{i=1}^{N_s} \forall_{t=n_a}^{T-1} : s_{i,t} \in \{0, 1\} \tag{9.3b}
\]

\[
\forall_{t=n_a}^{T-1} : \sum_{i=1}^{N_s} s_{i,t} = 1 \tag{9.3c}
\]

\[
\forall_{t=n_a}^{T-1} : \|\eta_t\|_{\infty} \leq \epsilon \tag{9.3d}
\]
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

The constraint (9.3a) is nonconvex. However, by defining the auxiliary variables \( \eta_{i,t:n_a} \) to substitute \( s_{i,t} \), it is easy to reformulate the feasibility of (9.3) equivalently to the feasibility of a MILP problem, as claimed in the following Theorem.

**Theorem 16.** The feasibility of (9.3) is equivalent to a MILP feasibility problem of the form

\[
\forall i \in \mathbb{N}_s, \forall t = n_a : s_{i,t} g_{i,t} - h_{i,t} = 0 \tag{9.4a}
\]

\[
\forall i \in \mathbb{N}_s, \forall t = n_a : s_{i,t} \in \{0, 1\} \tag{9.4b}
\]

\[
\forall t = n_a : \sum_{i=1}^{N_s} s_{i,t} = 1 \tag{9.4c}
\]

\[
\forall t = n_a : \| \eta_{i,t:n_a} \|_{\infty} \leq s_{i,t} \tag{9.4d}
\]

\[
\forall t = n_a : \sum_{i=1}^{N_s} \eta_{i,t:n_a} = \eta_{t:n_a} \tag{9.4e}
\]

**Proof.** Suppose \( \{s_{i,t}^*, \eta_{i,t}^*\} \) is a feasible solution to (9.3). By constructing \( \eta_{i,t:n_a}^* = s_{i,t}^* \eta_{i,t:n_a}^* \), obviously (9.4a)-(9.4c) are feasible at \( \{s_{i,t}^*, \eta_{i,t}^*, \eta_{i,t:n_a}^*\} \). Besides,

\[
\| \eta_{i,t:n_a}^* \|_{\infty} = s_{i,t}^* \| \eta_{i,t:n_a}^* \|_{\infty} \leq s_{i,t}^* \tag{9.5}
\]

and

\[
\sum_{i=1}^{N_s} \eta_{i,t:n_a}^* = \sum_{i=1}^{N_s} s_{i,t}^* \eta_{i,t:n_a}^* = \eta_{t:n_a}^* \sum_{i=1}^{N_s} s_{i,t}^* = \eta_{t:n_a}^* \tag{9.6}
\]

mean (9.4d) and (9.4c) also hold at \( \{s_{i,t}^*, \eta_{i,t}^*, \eta_{i,t:n_a}^*\} \). Thus, \( \{s_{i,t}^*, \eta_{i,t}^*, \eta_{i,t:n_a}^*\} \) is a feasible solution to (9.4).

Reversely, suppose \( \{s_{i,t}^*, \eta_{i,t}^*, \eta_{i,t:n_a}^*\} \) is a feasible solution to (9.4). From (9.4b) and (9.4c), we know for each \( t \), there exists a unique \( \sigma_t \in \mathbb{N}_s \) such that \( s_{i,t}^* = 1 \) for \( i = \sigma_t \), and \( s_{i,t}^* = 0 \), for \( \forall i \in \mathbb{N}_s \setminus \{\sigma_t\} \). Then from (9.4d) and

\[
0 \leq \| \eta_{i,t:n_a}^* \|_{\infty} \leq s_{i,t}^* \tag{9.7}
\]

we get

\[
\eta_{i,t:n_a}^* = 0 = s_{i,t}^* \eta_{i,t:n_a}^*, \forall i \in \mathbb{N}_s \setminus \{\sigma_t\} \tag{9.8}
\]
Further from (9.4e), it is derived that
\[
\eta^*_{t:t-na} = \eta^*_{\sigma_t:t-na} + \sum_{i \in \mathbb{N}_a \setminus \{\sigma_t\}} \eta^*_{i:t-na} = \eta^*_{\sigma_{t-1}:t-na} = s_{\sigma_t,t}^* \eta^*_{t:t-na}.
\] (9.6)
Combining (9.5) and (9.6), we have
\[
\eta^*_{i,t-na} = s_{i,t}^* \eta^*_{t:t-na}, \forall i \in \mathbb{N}_a,
\]
thus,
\[
s_{i,t}^* (g_{i,t} - h_i \eta^*_{t:t-na}) = s_{i,t}^* g_{i,t} - h_i \eta^*_{i:t-na} = 0
\]
holds for \(\forall i \in \mathbb{N}_a\), meaning (9.3a) is feasible at \(\{s_{i,t}^*, \eta^*_{t:t-na}\}\). Also from
\[
\|\eta^*_{t:t-na}\|_\infty = \|\sum_{i=1}^{N_a} \eta^*_{i:t:t-na}\|_\infty \leq \sum_{i=1}^{N_a} \|\eta^*_{i:t:t-na}\|_\infty \leq \sum_{i=1}^{N_a} s_{i,t}^* \epsilon = \epsilon,
\]
we know (9.3d) is feasible at \(\eta^*\). Therefore, \(\{s_{i,t}^*, \eta^*_{t:t-na}\}\) is a feasible solution to (9.3).

In summary, the feasibility of (9.3) is equivalent to the feasibility of (9.4). \(\Box\)

Note that the problem above is a MILP feasibility problem that can be solved with existing tools, for instance, CPLEX [130] and Gurobi [131]. However, while this approach works well for small to medium sized problems, its poor scaling properties make it impractical for larger sized ones. As we show next, this difficulty can be circumvented by exploiting recent results in sparsification and polynomial optimization to derive convex certificates that can be computed in polynomial time.

9.3.2 Sparsification Based Certificates

In this section we derive sparsification based sufficient conditions for the model (9.1) to be (in)validated by the experimental data. The starting point is the following result relating the feasibility of (9.4) to the cardinality of the solution to an LP problem.

**Lemma 6.** Writing the variables in a compact form as \(\eta = \begin{bmatrix} \eta_0^T & \cdots & \eta_{T-1}^T \end{bmatrix}^T\), and \(s = \begin{bmatrix} s_{na} & s_{na+1} \\
\vdots & \ddots & \vdots \\
\vdots & \cdots & s_{T-1} \\
s_{1,t} & \cdots & s_{N_a,t} \end{bmatrix}\) for \(\forall t = na, \ldots, T-1\), then Eq. (9.4) is feasible if and only
if there exists an optimal solution to the following problem:

\[
p^* = \text{minimize } \|s\|_0, \text{ subject to } \\
\forall_{i=1}^{N_s} \forall_{t=n_a}^{T-1} s_{i,t} \mathbf{g}_{i,t} - h_i \mathbf{\eta}_{i,t:t-n_a} = 0 \\
\forall_{i=1}^{N_s} \forall_{t=n_a}^{T-1} 0 \leq s_{i,t} \leq 1 \\
\forall_{t=n_a}^{T-1} s_{i,t} = \sum_{i=1}^{N_s} s_{i,t} = 1 \\
\forall_{i=1}^{N_s} \forall_{t=n_a}^{T-1} \|\mathbf{\eta}_{i,t:t-n_a}\|_\infty \leq s_{i,t} \epsilon \\
\forall_{t=n_a}^{T-1} \sum_{i=1}^{N_s} \mathbf{\eta}_{i,t:t-n_a} = \mathbf{\eta}_{t:t-n_a}
\]

with \(p^* = T - n_a\).

**Proof.** (Sufficiency) Suppose \(s^*\) and \(\eta^*\) is an optimal solution to (9.7) with

\[
p^* = \|s^*\|_0 = T - n_a.
\]

As a feasible solution to (9.7), from (9.7c) and (9.7d), \(s^*\) and \(\eta^*\) should satisfy

\[
\begin{align*}
\|s_t^*\|_0 &\geq \|s_t^*\|_1 = 1, \forall_{t=n_a}^{T-1} \\
\|s^*\|_0 &\geq \sum_{t=n_a}^{T-1} \|s_t^*\|_0 \geq \sum_{t=n_a}^{T-1} \|s_t^*\|_1 = T - n_a
\end{align*}
\]

Combining (9.8) with (9.9),

\[
\|s_t^*\|_0 = \|s_t^*\|_1 = 1, \forall_{t=n_a}^{T-1}
\]

holds, meaning this solution has exactly one \(\sigma_t \in \mathbb{N}_{N_s}\) at each \(t = n_a, \ldots, T - 1\), satisfying \(s^*_{\sigma_t,t} = 1\) and \(s^*_{i,t} = 0\) for \(\forall i \in \mathbb{N}_{N_s} \setminus \{\sigma_t\}\), equivalently, it follows \(s^*_{i,t}\) are binary. Thus, \(s^*\) and \(\eta^*\) is also a feasible solution to (9.4).

(Necessity) Suppose \(s^*\) and \(\eta^*\) is a feasible solution to (9.4). Since the feasible set for (9.4) is a subset of that for (9.7), then \(s^*\) and \(\eta^*\) is also a feasible solution to (9.7), and it follows \(p^* \leq \|s^*\|_0 = T - n_a\). On the other hand, following (9.9) we know \(p^* \geq T - n_a\). Thus, \(s^*\) and \(\eta^*\) is the optimal solution to (9.7) with \(p^* = \|s^*\|_0 = T - n_a\). 

From analysis above, it follows that a feasible solution to the MILP problem (9.4) is exactly the sparse solution to (9.7) with cardinality equal to \(T - n_a\). Furthermore, if and only if the
cardinality of the optimal solution to (9.7), $s^*$, is $T - n_a$, the model (9.1) is invalidated. However, (9.7) is nonconvex due to the $\ell_0$-norm in the objective function. Using the standard $\ell_1$ norm based relaxation of cardinality, combined with the re-weighted heuristic proposed in [89] we could relax (9.7) into a sequence of LPs, leading to sufficient convex certificates for (in)validation given by Algorithm 12 outlined below:

Algorithm 12 Sparsification Based (In)Validation Certificates

1: Initialize: $k = 0$, $w^{(0)}_{i,t} = 1$, for $\forall N_s i = 1, \forall T - 1 t = n_a$, a (small) regularization constant $\delta$, and the maximum number of iterations $k_{\text{max}}$;

2: repeat

3: solve

$$\{s^{(k)}, \eta^{(k)}\} = \text{arg minimize} \sum_{i=1}^{N_s} \sum_{t=n_a}^{T-1} w^{(k)}_{i,t} s_{i,t}$$ \hspace{1cm} (9.10) \\
subject to (9.7b) - (9.7f) (9.10)

4: update

$$w^{(k+1)}_{i,t} = (s^{(k)}_{i,t} + \delta)^{-1}, \forall i = 1, \forall T - 1 t = n_a$$

5: until convergence or $k = k_{\text{max}} + 1$ or the problem is infeasible.

Remark 6. In Algorithm 12 if (9.10) is infeasible, then the experimental data invalidates the model (9.1). On the other hand, if (9.10) is feasible and it converges to $s^*$ which satisfies $s^*_{i,t} \in \{0, 1\}$, or equivalently, $\|s^*\|_0 = T - n_a$, then the experimental data is consistent with the model (9.1) and the a priori information.

Algorithm 12 provides a computationally efficient way obtaining convex (in)validation certificates. Note however that these conditions are only sufficient, since on the one hand the exact recovery of sparse signals from $\ell_1$ minimization relies on the properties of the coefficients matrices such as restricted isometry property [132], the analysis of which is nontrivial for (9.7) due to the data corrupted by $\ell_\infty$-norm bounded noise, and on the other hand, the performance of the reweighted $\ell_1$ minimization varies with the choice of the weight function.

In consequence, Algorithm 12 cannot elucidate the situation when the relaxation above provides a feasible solution $s_o$ with non-integer elements. Such a situation can arise both in the case of valid models (when some data points can be explained by more than one model due to noise), or invalid ones (when one of the models can be obtained as a linear combination of the others). In principle, one can attempt to address this case proceeding in the spirit of MILP, by setting the variables with non-integer values to all possible combinations in $\{0, 1\}$, and, for each combination,
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

solving (9.10) with respect to the rest of the variables, in an attempt to find an integer solution or establish that such a solution does not exist. However, this approach becomes quickly intractable when the number of non-integer elements in \( s_o \) is not small. To circumvent this difficulty, next we use results from the theory of moments to obtain a convex necessary and sufficient (in)validation certificate.

9.3.3 Moments Based Certificates

In this section we explore the convex necessary and sufficient certificate for Problem 11 by appealing to the moments based optimization technique introduced in Section 2.2.

Consider the nonconvex polynomial optimization problem in the form of

\[
p^* = \min_{s, \eta} \mathcal{P}(s, \eta) = \sum_{i=1}^{N_s} \sum_{t=n_a}^{T-1} s_{i,t}^2 \| g_{i,t} - h_i \eta_{t-n_a} \|_2^2
\]

where

\[
K = \begin{cases} 
  s_{i,t} = s_{i,t}^2, \quad \forall_{i=1}^{N_s} \forall_{t=n_a}^{T-1} \\
  \sum_{i=1}^{N_s} s_{i,t} = 1, \quad \forall_{t=n_a}^{T-1} \\
  \| \eta_{t-n_a} \|_\infty \leq e_{T-n_a}^{T-1}
\end{cases}
\]

and its \( N \)-th \( (N \geq 2) \) order moments based relaxation

\[
p^*_N = \min_{m} \sum_{i=1}^{N_s} \sum_{t=n_a}^{T-1} l_{i,t} \]

subject to

\[
M_N(m) \succeq 0 \\
L_{N-1}(g_k m) \succeq 0, \forall_{k=1}^{d}
\]

where each \( l_{i,t} \) is the linear functional of moments \( m \) defined as \( l_{i,t}(m) = E \left\{ s_{i,t}^2 \| g_{i,t} - h_i \eta_{t-n_a} \|_2^2 \right\} \). \( E \) denotes expectation and where \( M_N \) and \( L_{N-1} \) are the moments and localizing matrices associated with a truncated moments sequence containing terms up to order \( 2N \), as outlined in section 2.2 and \( d \) is the number of constraints in (9.11).

Theorem 17. The relaxation (9.12) is exact at \( N = T - n_a + 1 \), i.e. \( p^*_N = p^* \), for \( \forall N \geq T - n_a + 1 \).

Proof. Firstly, it is obvious that \( p^*_N \leq p^* \). Next we will show that \( p - p^*_N \) has a representation of the
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

form

\[ p - \sum_{t=n_0}^{T-1} N_s \sum_{i=1}^{N_t} \beta_{ii}(\epsilon^2 - \eta_{ii}^2) \]

for some SoS polynomials \( p_o \) of degree at most 2\((T - n_0 + 1)\) and \( \beta_{ii} \) of degree at most 2\((T - n_0)\).

Let \( I \equiv \{ i_{n_0}, i_{n_0+1} \ldots i_{T-1} \} \), \( i_t \in \{ 1, \ldots, N_s \} \) denote a generic multi-index and define the \( N_s^{T-n_0} \) variables \( s_I \equiv s_{i_{n_0}, n_0} \ldots s_{i_{T-1}, T-1} \). From this definition it follows that \( \sum_I s_I^2 = \prod_{t=n_0}^{T-1} (\sum_{i=1}^{N_s} s_{i,t}^2) = 1 \). In terms of these variables, the objective in (9.11) can be rewritten as:

\[ p = \sum_I s_I^2 p_I = \sum_I s_I^2 \left[ (g_{i_1,n_0} - h_{i_1} \eta_{n_0,0})^2 + \ldots + (g_{i_T,T} - h_{i_T} \eta_{T,T-n_0})^2 \right] \quad (9.13) \]

where each term inside the brackets \( p_I \) corresponds to the costs of the trajectory \( I \), and hence \( p_I \geq p^* \).

Direct application of Lemma[1] to \( p_I \) shows that

\[ p_I - p^* = p_{o,I} + \sum_{t=0}^{T-1} \sum_{i=1}^{N_y} \lambda_{it,I}(\epsilon^2 - \eta_i(i))^2 \quad (9.14) \]

where \( p_{o,I} \) is some second order SoS polynomial of variables \( \eta_i \) and \( \lambda_{t,I} \equiv \left[ \lambda_{1t,I} \ldots \lambda_{n_{yt,I}} \right]_T \in \mathbb{R}^{n_y}, \lambda_{it,I} \geq 0 \). Furthermore, it follows that

\[ p - p_N^* = (p^* - p_N^*) + p - p^* = (p^* - p_N^*) + \sum_I s_I^2(p_I - p^*) \]

\[ = (p^* - p_N^*) + \sum_I s_I^2[p_{o,I} + \sum_{t=0}^{T-1} \sum_{i=1}^{N_y} \lambda_{it,I}(\epsilon^2 - \eta_i(i))^2] \]

\[ = p_o + \sum_{t=0}^{T-1} \sum_{i=1}^{N_y} \beta_{ii}(\epsilon^2 - \eta_i(i))^2 \quad (9.15) \]

where \( p_o \equiv (p^* - p_N^*) + \sum_I s_I^2 p_{o,I} \) is a SoS polynomial of order 2\((T - n_0 + 1)\), and \( \beta_{ii} \equiv \sum_I \lambda_{it,I} s_I^2 \) is a SoS polynomial of order 2\((T - n_0)\).

Therefore, following from Theorem[2] the relaxation at \( N = T - n_0 + 1 \) is exact and \( p_{T-n_0+1} = p^* \). Due to the fact that \( p^* \geq \cdots \geq p_{N+1}^* \geq p_N^* \), we have \( p_N^* = p^* \), for \( \forall N \geq T - n_0 + 1 \).

In the sequel, we give the convex necessary and sufficient certificate for the case when the model in Problem[11] is not invalidated.

**Theorem 18.** The following statements are equivalent:
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

(i) Problem (9.3) is feasible, i.e. the model in Problem 11 is not invalidated by the experimental data;

(ii) \( p^* = 0 \), where \( p^* \) denotes the optimal objective in (9.11);

(iii) \( p^*_{T-n_a+1} = 0 \), where \( p^*_{T-n_a+1} \) denotes the optimal objective in (9.12) at \( N = T - n_a + 1 \).

Proof. (i) \( \Rightarrow \) (ii) follows the fact that \( p^* \geq 0 \) since the objective in (9.11) is a SoS polynomial. Further, the feasible solution to (9.3) is also a feasible solution (9.11) with \( p = 0 \geq p^* \). Thus, \( p^* = 0 \).

(ii) \( \Rightarrow \) (i) follows that \( p^* = 0 \) is equivalent to that there exist binary \( s^* \) and \( \eta^*_t \in [-\epsilon, \epsilon] \) such that \( s^*_{i,t}(g_{i,t} - h_i \eta^*_{t-n_a}) = 0 \), for \( \forall i = 1, \ldots, N \), \( \forall t = n_a, \ldots, T - 1 \), thus \( s^* \) and \( \eta^* \) is a feasible solution to (9.3).

(ii) \( \Leftrightarrow \) (iii) follows directly from Theorem 17.

Similarly, the convex necessary and sufficient certificate for the case when the model in Problem 11 is invalidated is shown in the following theorem.

**Theorem 19.** The following statements are equivalent:

(iv) Problem (9.3) is infeasible, i.e. the model in Problem 11 is invalidated by the experimental data;

(v) \( p^* > 0 \), where \( p^* \) denotes the optimal objective in (9.11);

(vi) \( p^*_{T-n_a+1} > 0 \), where \( p^*_{T-n_a+1} \) denotes the optimal objective in (9.12) at \( N = T - n_a + 1 \).

Proof. (iv) \( \Rightarrow \) (v) and (v) \( \Rightarrow \) (iv) follow the transposition property in propositional logic of (ii) \( \Rightarrow \) (i) and (i) \( \Rightarrow \) (ii), respectively.

(v) \( \Leftrightarrow \) (vi) follows directly from Theorem 17.

From Theorems 1, 17, 18, 19 and the fact that \( p^*_N \) is monotonically nondecreasing in terms of \( N \) (see for instance [25]), it follows that Problem 11 can be recast as a (finite) sequence of convex SDP optimization problems, as outlined in Algorithm 13 below.

**Remark 7.** It is worth emphasizing that, while Theorem 17 provides a guaranteed stopping condition for the algorithm above, in practice this condition has proved to be conservative. Indeed, consistent numerical experience supports the conjecture that a much smaller relaxation order \( N \) need to be considered. However, at the present time no formal proof of this fact is available.
Algorithm 13 Moments Based (In)Validation Certificates
1: Initialize: \( N = 2 \);
2: repeat
3: solve (9.12);
4: update \( N = N + 1 \).
5: until
6: If \( p_N^* > 0 \) the model is invalid, otherwise the experimental data is consistent with the \( a \) \( p \)r\( i \)\( o \)r\( i \) assume\( s \). 

The number of variables in (9.11), \( v = [s^T, \eta^T]^T \) is equal to \( L = N_s(T - n_a) + Tn_y \), then the number of the associated variable \( m \) up to order 2\( N \) in (9.12), \( S^{2N}_L \), grows exponentially as \( N \) increases. The computation complexity will turn too high to be tractable. However, by partitioning both the objective \( p(s, \eta) \) and the feasible set \( K \), we can rewrite (9.11) into

\[
p^* = \min_{s, \eta \in K} \sum_{t=n_a}^{T-1} p_t
\]

where \( p_t = \sum_{i=1}^{N_s} s_{i,t}^2 \| g_{i,t} - h_{i,\eta_{t:t-n_a}} \|_2^2 \),

\[
K = \bigcup_{t=n_a}^{T-1} K_t, \text{ and }
\]

\[
\forall_{t=n_a}^{T-1} : K_t = \begin{cases} 
  s_{i,t} = s_{i,t}^2, \forall_{i=1}^{N_s} \\
  \sum_{i=1}^{N_s} s_{i,t} = 1 \\
  \| \eta_{t:t-n_a} \|_{\infty} \leq \epsilon 
\end{cases}
\]

(9.16)

Obviously, each partition \( p_t, K_t \) is just related to variables \( v_t = [s_t^T, \eta_{t,t-n_a}^T]^T \in \mathbb{R}^l \), with \( l = N_s + (n_a + 1)n_y \), and

\[
v_{t+1} \cap \bigcup_{i=n_a}^{t} v_i = \{ \eta_{t:t-n_a+1} \} \in v_t, \forall t = n_a, \ldots, T - 2,
\]

holds, thus the running intersection property holds, and the optimal solution of (9.11) can be obtained.
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

by solving the convergent hierarchy of reduced SDP problem of the form

\[
\hat{p}_N^* = \min_m \sum_{t=n_a}^{T-1} l_t
\]

subject to

\[
M_N(m, v_t) \succeq 0, \forall T-1 t = n_a
\]

\[
L_{N-1}(g_{k,t} m, v_t) \succeq 0, \forall g_{k,t} \in K_t, \forall T-1 t = n_a.
\]

(9.17)

In contrast to (9.12), for relaxation order at \(N\), which involves one SDP constraint of size \(S_N^L\) and \(d\) SDP constraints of size \(S_N^{N-1}\), in the reduced SDP problem (9.17), the number of variable is \((T-n_a) S^2_N\), and it contains \(T-n_a\) SDP constraints of size \(S^N_1\) and \(d\) SDP constraints of size \(S^{N-1}_1\). When \(T\) is large (i.e., the experimental data is a long sequence), \(L \approx (T-n_a) l\), and \(S^{2N}_L \gg S^{2N}_1\), thus (9.17) will reduce the computational complexity dramatically.

**9.4 Incorporating Structural Constraints**

In many situations of practical interest, \(a\)-\textit{priori} constraints are available on the admissible transitions of the mode variable. Incorporating these constraints can rule out models that may otherwise be deemed valid. For instance, in biological systems applications, only certain transitions can take place from a given metabolic stage. Similarly, when dealing with human activities, only some sequences of activities are “normal”: for instance, depending on the context, a sequence such as \(\{\text{run} \rightarrow \text{walk} \rightarrow \text{stop}\}\) could be normal, while \(\{\text{walk} \rightarrow \text{stop} \rightarrow \text{run}\}\) may indicate an abnormal event. As we briefly discuss below, one of the advantages of the proposed framework is that structural constraints on the underlying graph can be easily accommodated. Specifically, assume that the transition from \(\sigma_t = i\) to \(\sigma_{t+\tau} = j\) is not admissible (that is, if at time instant \(t\), submodel \(G_i\) is active, then \(G_j\) cannot be active \(\tau\) stages later). Clearly, such constraints are easily incorporated in our framework by simply adding constraints of the form:

\[
s_{i,t} + s_{j,t+\tau} \leq 1, \text{ or } s_{i,t}s_{j,t+\tau} = 0
\]

(9.18)

Similarly, situations where the feasible set \(\mathcal{I}_t\) of the discrete mode variable at time instant \(t\) constrains its mode variable at time instant \(t + \tau\) to some set \(\mathcal{I}_{t+\tau}\) simply reduce to:

\[
\sum_{i \in \mathcal{I}_t} s_{i,t} = \sum_{j \in \mathcal{I}_{t+\tau}} s_{j,t+\tau}
\]

(9.19)
that is, \( \sigma_{t+\tau} \in \mathcal{I}_{t+\tau} \) if and only if \( \sigma_t \in \mathcal{I}_t \).

A point worth noting is that these additional constraints have relatively little effect on the overall computational complexity when using the sparsification based approach proposed in section 9.3.2. On the other hand, they directly impact the running intersection property for the moments based approach in section 9.3.3, since these constraints couple the binary variables associated with more than two consecutive time instants. Thus, the variables should be partitioned into larger sets than \( v_t \), and moments and localizing matrices required are enlarged. For the problem (9.11) with extra constraint (9.18) or (9.19), the variables can be partitioned into \( \tilde{v}_t = \bigcup_{j=t}^{t+\tau} v_t, \) for \( t = n_a, \ldots, T - 1 - \tau \), and the objective and the constrains are partitioned into

\[
p = \sum_{t=n_a}^{T-1-\tau} \tilde{p}_t, \quad \tilde{K} = \bigcup_{t=n_a}^{T-1-\tau} \tilde{K}_t,
\]

where

\[
\tilde{p}_t = \begin{cases}
\sum_{i=1}^{N_s} s_{i,t}^2 \| g_{i,t} - h_i \eta_{t:n_a} \|_2^2, & \forall t = n_a \ldots T - 2 - \tau \\
\sum_{j=T-1-\tau}^{T-1} \sum_{i=1}^{N_s} s_{i,j}^2 \| g_{i,j} - h_i \eta_{j:j-n_a} \|_2^2, & t = T - 1 - \tau.
\end{cases}
\]

\[
\tilde{K}_t = \begin{cases}
s_{i,j} = s_{i,j}^2, & \forall i = 1 \ldots n_a \ldots T - 1 - \tau \\
\sum_{i=1}^{N_s} s_{i,j} = 1, & \forall t = n_a \ldots T - 1 - \tau. \\
\| \eta_t : t:n_a \|_\infty \leq \epsilon
\end{cases}
\]

9.5 Experiments

In this section we illustrate the effectiveness of the proposed method both using academic examples and a computer vision application.

9.5.1 Academic Examples

The goal of these examples is to compare the performance of the MILP formulation (9.4) and its sparsification-based and moments-based relaxations introduced in this paper against the method in [127]. In all cases, the MILP (9.4) and the LPs (9.10) in sparsification-based relaxation were solved by Gurobi [131], and the SDP problems (9.17) resulted from moments-based relaxation were solved using cvx [133]. All the experiments were run on iMac with 32GB memory and 3.4GHz Intel Core i7 Processor.

A.1 Example without constraints on the switching sequence
In this example we consider data generated by a system that switched randomly between the following three subsystems:

\[
\begin{align*}
\xi_t &= 0.2\xi_{t-1} + 0.24\xi_{t-2} + 2u_{t-1} \\
\xi_t &= 1.7\xi_{t-1} - 0.72\xi_{t-2} + 0.5u_{t-1} \\
\xi_t &= -1.4\xi_{t-1} - 0.53\xi_{t-2} + u_{t-1} \\
\xi_t &= -0.5\xi_{t-1} - 0.06\xi_{t-2} - u_{t-1}
\end{align*}
\]

and the measurement equation:

\[y_t = \xi_t + \eta_t\]

with unknown uniformly distributed random noise \(\|\eta_t\|_\infty \leq \epsilon, \epsilon = 0.2\), and uniformly distributed random input \(\|u_t\|_\infty \leq 1\). We collect data \(\{u_t, y_t\}_{t=1}^{102}\) with different signal-to-noise ratios (SNR). Here we carry out experiments under three different settings, and the results are summarized in Table 9.1 to Table 9.3 respectively. In the rows labeled “results from SDP using Algorithm 13” and “results from SDP from [127]”, we show both the relaxation order \(N\) and the value of the objective function in the corresponding polynomial optimization problem \(p^*_N\): a value \(p^*_N > 0\) certifies that the model is invalid, while a value \(p^*_N \approx 0\), together with a flat-extension, indicates that the data observed so far is consistent with the a-priori assumptions on the submodels and noise bound.

In the first two experiments we assume the prior knowledge on the noise bound is \(\epsilon = 0.2\) which is consistent with ground truth. In the first experiment we test the different approaches by attempting to validate the prior knowledge on the model, and in the second experiment we attempt to invalidate the prior knowledge on the model which is just a subset of \(\{G_i\}\)s while the data is generated by a different subset. The results are shown in Table 9.1 and Table 9.2. In the third experiment we test on the same data used in experiment 1, attempting to invalidate the prior knowledge on the noise bound \(\epsilon\), the results of which are summarized in Table 9.3.

As expected, the MILP and LPs resulted from sparsification based relaxation run considerably faster than the SDPs resulted from moments-based approaches. However, as illustrated in Table 9.1 in some cases MILP and LPs fail to provide a conclusive certificate.

Regarding the moments-based approaches, both the one proposed in this paper and the one proposed in [127] have comparable performance. In the tables we just show the results using the lowest order relaxations, for “not invalidated” cases, according to Algorithm 13 we need to solve the
CHAPTER 9. MODELS (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

SDP problems with increasingly higher order of relaxations until the stopping condition is satisfied, which may lead to computationally intractable problems. However, consistent experimental results enlighten us that for the moments-based relaxation proposed in this paper, we could use the moments variables for the binary variables $s_{i,t}$ obtained from a lower order relaxation (for cases here $N = 2$) to initialize the weight $w_{i,t}^{(0)}$ in Algorithm 12 and by running Algorithm 12 we could obtain a binary solution to $s_{i,t}$ and a sequence of $\eta_t$ to denoise the measurements $y_t$.

<table>
<thead>
<tr>
<th>Models</th>
<th>Actual</th>
<th>A Priori Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>SNR</td>
<td>$G_1, G_2, G_3, G_4$ ($\epsilon = 0.2$)</td>
</tr>
<tr>
<td>MILP</td>
<td>Results</td>
<td>feasible, $s_{i,t} \in {0, 1}$</td>
</tr>
<tr>
<td></td>
<td>Interpretation</td>
<td>not invalidated</td>
</tr>
<tr>
<td></td>
<td>Time (sec.)</td>
<td>0.0721</td>
</tr>
<tr>
<td></td>
<td>Results</td>
<td>feasible, $s_{i,t} \in {0, 1}$</td>
</tr>
<tr>
<td></td>
<td>Interpretation</td>
<td>no decision</td>
</tr>
<tr>
<td></td>
<td>Time (sec.)</td>
<td>0.2415</td>
</tr>
<tr>
<td>LP using Alg. 12</td>
<td>Results</td>
<td>$2 \times 4.8777e-07$</td>
</tr>
<tr>
<td></td>
<td>Interpretation</td>
<td>not invalidated</td>
</tr>
<tr>
<td></td>
<td>Time (sec.)</td>
<td>13.4313</td>
</tr>
<tr>
<td>SDP using Alg. 13</td>
<td>Results</td>
<td>$4 \times 5.6407e-09$</td>
</tr>
<tr>
<td></td>
<td>Interpretation</td>
<td>not invalidated</td>
</tr>
<tr>
<td></td>
<td>Time (sec.)</td>
<td>19.0838</td>
</tr>
</tbody>
</table>

| SDP from [127] | Results | $2 \times 5.7845$ | $2 \times 1.2114$ |
|                | Interpretation | invalidated | invalidated |
|                | Time (sec.) | 9.0006 | 9.0013 |
| SDP from [127] | Results | $2 \times 10.6028$ | $2 \times 0.9110$ |
|                | Interpretation | invalidated | invalidated |
|                | Time (sec.) | 5.8081 | 9.3809 |

Table 9.1: Invalidation Results for Example A.1 (Part 1) (both the actual noise bound and the prior knowledge on it are $\epsilon = 0.2$)

Table 9.2: Invalidation Results for Example A.1 (Part 2) (both the actual noise bound and the prior noise bound to initialize the weight $w_{i,t}^{(0)}$ in Algorithm 12)

Finally, as shown in Table 9.2, in some cases the approach in [127] runs faster than our SDP formulation, this is because with the same relaxation order, our formulation involves more variables $s_{i,t}$. In Table 9.4 we provide the formulations to calculate the number of variables and the size of LMIs (associated with each time instant) as functions of number of submodels $N_a$ and the order of ARX models $n_a$. In Figure 9.2 we plot the contour of the difference between the size of the LMIs from [127] and that from (9.17) (with the lowest relaxation orders). Obviously we can observe

135
Table 9.3: Invalidation Results for Example A.1 (Part 3)

<table>
<thead>
<tr>
<th>Models</th>
<th>Actual</th>
<th>MILP</th>
<th>LP using Alg. 12</th>
<th>SDP using Alg. 13</th>
<th>SDP from [127]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$G_1, G_2, G_3, G_4 (\epsilon = 0.2)$</td>
<td>$G_1, G_2, G_3, G_4 (\epsilon = 0.2)$</td>
<td>$G_1, G_2, G_3, G_4 (\epsilon = 0.1)$</td>
<td>$G_1, G_2, G_3, G_4 (\epsilon = 0.1)$</td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>SNR</td>
<td>Interpretation</td>
<td>Time (sec.)</td>
<td>Interpretation</td>
<td>Time (sec.)</td>
</tr>
<tr>
<td></td>
<td>39.1072</td>
<td>infeasible</td>
<td>0.0624</td>
<td>infeasible</td>
<td>0.0936</td>
</tr>
<tr>
<td></td>
<td>27.0712</td>
<td>invalidated</td>
<td>0.0718</td>
<td>invalidated</td>
<td>0.0964</td>
</tr>
<tr>
<td></td>
<td>19.5676</td>
<td>invalidated</td>
<td>0.0724</td>
<td>invalidated</td>
<td>0.0964</td>
</tr>
</tbody>
</table>

that except for a few specific cases where either $N_s$ or $n_a$ is not larger than 2, the size of LMIs in (9.17) is smaller than that in [127].

Table 9.4: Complexity Analysis for Moments-Based Formulations ($n_y = 1$)

<table>
<thead>
<tr>
<th>Formulations</th>
<th>SDP in [127]</th>
<th>SDP in (9.17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smallest degree of polynomials</td>
<td>$2N_s$</td>
<td>$4$</td>
</tr>
<tr>
<td>Lowest order of relaxation</td>
<td>$N_s$</td>
<td>$2$</td>
</tr>
<tr>
<td>No. of variables in polynomials</td>
<td>$n_a + 1$</td>
<td>$n_a + 1 + N_s$</td>
</tr>
<tr>
<td>No. of moments variables</td>
<td>$\begin{pmatrix} \frac{2N_s}{n_a + 1 + 2N_s} \ \frac{4}{n_a + N_s + 5} \end{pmatrix}$</td>
<td>$\begin{pmatrix} \frac{2}{n_a + N_s + 3} \end{pmatrix}$</td>
</tr>
<tr>
<td>size of LMI</td>
<td>$\begin{pmatrix} \frac{N_s}{n_a + 1 + N_s} \end{pmatrix}$</td>
<td>$\begin{pmatrix} \frac{2}{n_a + N_s + 3} \end{pmatrix}$</td>
</tr>
</tbody>
</table>

A.2 Example with constraints on the switching sequence

Next, we use a simple example to illustrate the ability of the proposed methods to handle constraints on the admissible transitions. The data was generated using the same ARX submodels as in A.1, for ease of comparison, we test on the same data as in experiment 1 (with results in Table 9.1), which is generated by arbitrary switching among $\{G_i\}_{i=1}^4$ containing switching from $G_1$ to $G_2$. When invalidating, we impose the constraint that the underlying graph has the topology shown in Fig. 9.3, that is, one-step transitions from $G_1$ to $G_2$ are not admissible. Following the idea outlined in Section 9.4, this translates into a constraint of the form

$$s_{1,t} + s_{2,t+1} \leq 1, \forall t.$$

(9.20)
The results obtained running both Algorithms 12 and 13 are summarized in Table 9.5.

Comparing the results in Table 9.5 with those in Table 9.1, we can see if this constraint is not enforced.
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

(as in experiment 1), then the model is deemed to be valid.

Table 9.5: Invalidation under Structural Constraints (the same data as in Table 9.1)

<table>
<thead>
<tr>
<th>Models</th>
<th>Actual</th>
<th>$G_1$, $G_2$, $G_3$, $G_4$, with $G_1 \rightarrow G_2$ ($\epsilon = 0.2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>SNR</td>
<td>39.1072 27.0712 19.5676</td>
</tr>
<tr>
<td>MILP with Constraint (9.20)</td>
<td>Results</td>
<td>infeasible infeasible infeasible</td>
</tr>
<tr>
<td></td>
<td>Interpretation</td>
<td>invalidated invalidated invalidated</td>
</tr>
<tr>
<td></td>
<td>Time (sec.)</td>
<td>0.0982 0.0781 0.0790</td>
</tr>
<tr>
<td>LP using Alg. 12 with Constraint (9.20)</td>
<td>Results</td>
<td>2 2 2 9755 2 1.0854 2 1.0040</td>
</tr>
<tr>
<td></td>
<td>Interpretation</td>
<td>invalidated invalidated invalidated</td>
</tr>
<tr>
<td></td>
<td>Time (sec.)</td>
<td>2 296.1764 2 348.8198 2 288.8557</td>
</tr>
<tr>
<td>SDP using Alg. 13 with Constraint (9.20)</td>
<td>Results</td>
<td>2 2 2 9755 2 1.0854 2 1.0040</td>
</tr>
<tr>
<td></td>
<td>Interpretation</td>
<td>invalidated invalidated invalidated</td>
</tr>
<tr>
<td></td>
<td>Time (sec.)</td>
<td>2 296.1764 2 348.8198 2 288.8557</td>
</tr>
</tbody>
</table>

9.5.2 Contextually Abnormal Activity Recognition

In this section we apply the proposed model invalidation framework to a non-trivial problem arising in computer vision: activity monitoring. Following the framework proposed in [127], we model a sequence of activities as the output of switched ARX system, where each submodel corresponds to a certain activity. As shown in [127], contextually abnormal activities can be detected by simply invalidating the data record against the set of trajectories that can be obtained by switching among a set of known, normal activities, whose associated models can be obtained using identification tools. In this section, we further develop this framework and show that using the algorithms proposed here allows for not only recognizing abnormal activities, but also abnormal activity sequences. The model obtained for the activities \{walk, stop, run\}, given below for ease of reference, are:

$$
\begin{pmatrix}
  x_t \\
  y_t
\end{pmatrix} = \begin{pmatrix}
  0.9571 & 1.7629 \\
  -0.0095 & -0.2260
\end{pmatrix} \begin{pmatrix}
  x_{t-1} \\
  y_{t-1}
\end{pmatrix} + \begin{pmatrix}
  -0.4897 \\
  0.4718
\end{pmatrix} u_{t-1} \quad (G_1 : \text{walk})
$$

$$
\begin{pmatrix}
  x_t \\
  y_t
\end{pmatrix} = \begin{pmatrix}
  1 & 0 \\
  0 & 1
\end{pmatrix} \begin{pmatrix}
  x_{t-1} \\
  y_{t-1}
\end{pmatrix} \quad (G_2 : \text{stop})
$$

$$
\begin{pmatrix}
  x_t \\
  y_t
\end{pmatrix} = \begin{pmatrix}
  0.9614 & 0.9154 \\
  -0.4002 & 1.2487
\end{pmatrix} \begin{pmatrix}
  x_{t-1} \\
  y_{t-1}
\end{pmatrix} \quad (G_3 : \text{run})
$$

where $(x_t, y_t)$ denote the coordinates, at time instant $t$, of the center of the person’s mass, and $u_{t-1} = 1$. Sample frames from the video sequences involving combinations of these activities are shown in Figure 9.4. As in [127], background subtraction was used to locate the person, and $(x_t, y_t)$
CHAPTER 9. MODEL (IN)VALIDATION OF SARX SYSTEMS WITH UNKNOWN SWITCHES

were estimated using the silhouettes and normalized to be within the range \([0, 1]\).

The results by solving MILP and using Algorithms 12 and 13 to sequences involving combinations of these activities are summarized in Table 9.6. For the purpose of this example, we assumed that transitions from \textit{stop} to \textit{run} were not admissible. Hence, a sequence of the form \{\textit{run} \rightarrow \textit{walk} \rightarrow \textit{stop}\} should be labeled normal, while \{\textit{walk} \rightarrow \textit{stop} \rightarrow \textit{run}\} should be flagged as contextually abnormal, mathematically,

\[ s_{1,t} + s_{2,t+1} \leq 1, \forall t. \] (9.21)

For comparison, we also recorded the results by solving the problem without enforcing the transition constraint (9.21). As shown in Table 9.6, when the actual data containing transitions not allowed, only the methods which can handle extra constraints like (9.21) could provide correct results.

Figure 9.4: Detecting Contextually Abnormal Activities in Video Sequences (For the purpose of this example we assumed that transitions from \textit{wait} to \textit{running} are not allowed. Hence the top sequence (\textit{walk} \rightarrow \textit{stop} \rightarrow \textit{run}) should be labeled abnormal, while the bottom one (\textit{stop} \rightarrow \textit{walk} \rightarrow \textit{run}) should not be flagged.)

9.6 Conclusions

In this chapter we considered the model (in)validation problem for SARX systems with unknown mode signal. Given a nominal model, a bound on the measurement noise and experimental input/output data, we provided a necessary and sufficient condition that certifies the existence/nonexistence of admissible noise and switching sequences such that the resulting output sequence interpolates the given experimental data within the noise bound. The main result of this paper shows that these certificates can be computed by solving convex optimization problems.
Moreover, these problems can be easily modified to incorporate constraints on the topology of the underlying graph. In many practical cases, a validation/(in)validation certificate can be obtained by simply solving a sequence of LP problems as in Algorithm 12. Cases where this approach fails to ascertain the validity of the model can be addressed, at the price of additional computational complexity, by using a moments-based reformulation containing moments of order up to \(2(T - n_a + 1)\), where \(T\) denotes the number of experimental data points and \(n_a\) denotes the order of the ARX submodels.

The effectiveness of the proposed method was illustrated using both academic examples and a non-trivial problem arising in computer vision: activity monitoring. An interesting feature borne out by consistent numerical experience is the fact that, the resulting moments variables of the binary variables \(s_{i,t}\) by solving lower order moments relaxation (for instance, the lowest \(N = 2\)) could serve as an efficient warm-start for the initialization of the weight \(w_{i,t}^{(0)}\) in the sparsfication based algorithm. When using the moments-based approach, flat extensions (and hence exact solutions) were achieved numerically for relaxations involving only moments of order up to 4 (e.g. \(N = 2\)). However, no formal proof of this fact is currently available.
Chapter 10

Robust Superstable Controller Design for LTI Systems

10.1 Motivation

Many practical scenarios in control community involves designing controllers to the model identified from a combination of experimental data and some a-priori information to achieve some performance requirements such as stability. Classically, this problem has been addressed by a multi-step procedure. Firstly, a suitable plant and associated worst-case identification bounds are obtained using control-oriented identification methods (see for instance [122]; [134]; [95] and references therein). If additional experimental data is available, then the uncertainty bounds can be refined through a model (in)validation step for instance ([135], [136]) and methods mentioned in Chapter 9. Finally, a robust controller is synthesized that attempts to stabilize the family of plants defined by the identified nominal model and uncertainty bounds. While this multi-step approach usually works in practice, it can be overly conservative (due to the use of upper bounds in the identification and (in)validation steps). Thus, there is no guarantee that a suitable controller can be found. In addition, the entailed computational complexity is far from trivial.

Data-driven controller tuning tries to lump these steps and present a direct “data-to-control” algorithm ([137],[140]). The basic idea of data-driven approaches is to choose a controller which minimizes a quadratic cost function measuring the gap between a given reference model and the closed-loop system. The main challenges in data-driven approaches consist of (1) the closed-loop stability is not guaranteed; (2) if the experimental data are corrupted by noise, then the statistic
properties of the controller (such as unbiasedness) are affected by the finite number of data record and the statistic properties of the noise.

Motivated by these difficulties, in this chapter we propose a new framework for directly identifying controllers from experimental data corrupted by \( \ell_{\infty} \)-norm bounded process noise, without explicitly identifying the plant. This is accomplished by requiring the controller to super-stabilize \(^{[33],[34]}\) all possible plants which are consistent with the experimental data. While super-stability is a stronger requirement than Schur stability, it has the advantage of having a direct connection with time-domain, peak-to-peak performance, allowing for designing a controller with guaranteed worst case performance over the entire consistency set. Further, as the order of the controller increases, performance of these controllers monotonically converges to that of the optimal \( \ell^1 \) controller.

The main result of this chapter shows that the problem of synthesizing a robustly super-stabilizing controller directly from the experimental data, albeit in principle non-convex, can be recast into a convex Linear Program by exploiting results from Robust Optimization. \(^{[35,37,141]}\). A potential difficulty with this approach is that the number of constraints required to handle the robust optimization in a convex setting grows exponentially with the problem size, although problems involving relatively low order plants are well within the capabilities of modern solvers such as Gorubi \(^{[142]}\) and CPLEX \(^{[130]}\). In order to handle larger sized problems, in the last portion of this chapter we introduce several LP based relaxations that substantially reduce the computational complexity at the price of some decrease in performance. These results are illustrated with a simple example.

10.2 Problem Statement

Consider a single input single output (SISO) strictly proper plant characterized by the transfer function\(^1\)

\[
P(\lambda) = \frac{P_p(\lambda)}{D_p(\lambda)} = \frac{\sum_{k=1}^{n_b} b_k(p) \lambda^k}{1 + \sum_{k=1}^{n_a} a_k(p) \lambda^k}, n_a \geq n_b \tag{10.1}
\]

Given experimental input/output data \( \{u_t, y_t\}, \forall t = 0, \ldots, T \), where \( u_t \) and \( y_t \) denote a known input and the system output corrupted by additive process noise, respectively, e.g.

\[
y_t = -\sum_{k=1}^{n_a} a_k(p) y_{t-k} + \sum_{k=1}^{n_b} b_k(p) u_{t-k} + \eta_t, \quad |\eta_t| \leq \epsilon \tag{10.2}
\]

\(^1\)Keeping with the superstability literature here we use the \( \lambda \) transform, where \( \lambda \equiv z^{-1} \) is the unit delay operator.
then the consistency set is defined as the set of all plants of the form (10.1) that could have generated the observed data. In the case of $\ell^\infty$ bounded noise, this set is a polyhedron of the form

$$
\mathcal{P}(\theta_p) = \{ \theta_p \in \mathbb{R}^{n_a+n_b} : |y_t - x_t^T \theta_p| \leq \epsilon, \forall t=1-n_a \}
$$

(10.3)

where $x_t = [ -y_{t-1} \cdots -y_{t-n_a} u_{t-1} \cdots u_{t-n_b}]^T$ and $\theta_p = [a^{(p)}_1 \cdots a^{(p)}_{n_a} b^{(p)}_1 \cdots b^{(p)}_{n_b}]^T$ denotes the unknown parameters of the plant. Here, $D \in \mathbb{R}^{2T(1-n_a) \times (n_a+n_b)}$ and $q \in \mathbb{R}^{2(T+1-n_a)}$ are known constants.

Consider the feedback control system shown in Figure 10.1, where $P$ is an unknown plant. The goal of this chapter is to synthesize a controller guaranteed to stabilize all plants compatible with the a-priori information and the experimental data in (10.3), without having to identify the plant itself. Formally, this problem can be stated as:

**Problem 12.** Given a set of experimental data $\{u_t, y_t\}_{t=0}^{T} \times$ collected from a plant $P$ of the form (10.2) with known order $n_a$ and $n_b$, and a bound $\epsilon$ on $\ell^\infty$ norm of the noise, design a controller $C$ (determined by unknown $\theta_c = [a^{(c)}_1, \cdots, a^{(c)}_{n_c}, b^{(c)}_0, \cdots, b^{(c)}_{n_c}]]^T$) with the transfer function

$$
C(\lambda) = \frac{N_c(\lambda)}{D_c(\lambda)} = \frac{\sum_{k=0}^{n_c} b^{(c)}_k \lambda^k}{1 + \sum_{k=1}^{n_c} a^{(c)}_k \lambda^k}
$$

(10.4)

such that the resulting closed-loop system is superstable for any plant in the consistency set $\mathcal{P}(\theta_p)$, and the $(n_c + n_a)$-equalized performance level $\mu$ is minimized.

![Figure 10.1: Closed-loop System](image)

**Remark 8.** The consistency set $\mathcal{P}(\theta_p)$ defined by (10.3) is the tightest set containing the actual plant. Thus, the proposed controller is guaranteed to be the least conservative amongst all controllers capable of super-stabilizing all plants compatible with the available information, including those...
CHAPTER 10. ROBUST SUPERSTABLE CONTROLLER DESIGN FOR LTI SYSTEMS

obtained pursuing a classical approach based on first identifying the plant and associated uncertainty bounds, followed by a robust controller synthesis step.

10.3 Convex Reformulation

In this section, we present the main contributions of this chapter: (i) theoretical results showing that Problem 12 can be reformulated as a convex optimization problem, and (ii) a synthesis algorithm that exploits this reformulation. To this effect, we first show in Section 10.3.1 that Problem 12 can be formulated as a constrained nonconvex optimization problem. Then, in Section 10.3.2 we use robust optimization techniques to obtain a convex counterpart.

10.3.1 Problem Reformulation

To reformulate Problem 12 as an optimization problem, begin by writing explicitly the closed-loop transfer function of the system in Figure 10.1 as:

\[ T(\lambda) = \frac{P(\lambda)}{1 + P(\lambda)C(\lambda)} = \frac{N_pD_c}{D_pD_c + N_pN_c} = \frac{\sum_{k=1}^{n} b_k \lambda^k}{1 + \sum_{k=1}^{n} a_k \lambda^k} \]  \hspace{1cm} (10.5)

where \( n = n_a + n_c \). Since (by scaling, if necessary), we can assume that the characteristic polynomials of the controller and plant are monic, it follows that the unknown coefficients \( \theta = [a^T \ b^T]^T \in \mathbb{R}^{2n+1} \) of \( T(\lambda) \) are bilinear functions of \( \theta_p \) and \( \theta_c \), of the form:

\[ a_k = \sum_{(i,j)\in I_{ka}} a_k^{(p)} a_j^{(c)} + \sum_{(i,j)\in I_{kb}} b_i^{(p)} b_j^{(c)}, \forall k = 1 \]

\[ b_k = \sum_{(i,j)\in I_k} b_i^{(p)} a_j^{(c)}, \forall k = 1 \]

\[ b_k = 0, \forall k = n_c + n_a + 1 \text{ or } k = 0 \]  \hspace{1cm} (10.6)

where \( I_{ka} = \{(i,j) \in \mathbb{N}^2 : 0 \leq i \leq n_a, 0 \leq j \leq n_c, i + j = k\} \), \( I_{kb} = \{(i,j) \in \mathbb{N}^2 : 0 \leq i \leq n_b, 0 \leq j \leq n_c, i + j = k\} \), and \( I_k = \{(i,j) \in \mathbb{N}^2 : 0 \leq i \leq n_b, 0 \leq j \leq n_c, i + j = k\} \). In compact form, the entries of \( \theta \) can be rewritten as:

\[ \theta_i = (Q_i \theta_p + f_i^{(c)})^T \theta_c + f_i^{(p)T} \theta_p + r_i, \forall i = 1 \]  \hspace{1cm} (10.7)

where \( Q_i \in \mathbb{R}^{(2n_c+1)(n_a+n_b)}, f_i^{(c)} \in \mathbb{R}^{2n_c+1}, f_i^{(p)} \in \mathbb{R}^{n_a+n_b}, \) and \( r_i \in \mathbb{R} \) are known constants. Using this notation, from Theorem 3 it follows that Problem 12 can be reformulated as a constrained
CHAPTER 10. ROBUST SUPERSTABLE CONTROLLER DESIGN FOR LTI SYSTEMS

optimization problem of the form

\[
\mu^* = \min_{\theta_c} \max_{\theta_p \in \mathcal{P}(\theta_p)} f(\theta_p, \theta_c) = \frac{||b||_1}{1-||a||_1}
\]

subject to
\[
1 - ||a||_1 > 0, \forall \theta_p \in \mathcal{P}(\theta_p)
\]

(10.8)

Theorem 20. If \( \bar{\theta}_p = \left[ \bar{a}_p^T, \bar{b}_p^T \right]^T \), with \( \bar{a}_p \in \mathbb{R}^{na} \) and \( \bar{b}_p = 0 \in \mathbb{R}^{nb} \), do not belong to the feasible set \( \mathcal{P}(\theta_p) \), the problem (10.8) is equivalent to

\[
\tilde{\mu}^* = \min_{\theta_c, \mu > 0} \mu
\]

subject to
\[
\mu ||a||_1 + ||b||_1 \leq \mu, \forall \theta_p \in \mathcal{P}(\theta_p)
\]

(10.9b)

Proof. (i) Problem (10.8) is equivalent to

\[
\mu^* = \min_{\theta_c, \mu > 0} \mu
\]

subject to
\[
\frac{||b||_1}{1-||a||_1} \leq \mu, \forall \theta_p \in \mathcal{P}(\theta_p)
\]

(10.10b)

(10.10c)

Due to the fact that any feasible solution \( (\theta_c, \mu > 0) \) to (10.10) is also a feasible solution to (10.9), we have \( \mu^* \geq \tilde{\mu}^* \).

(ii) Any feasible solution to (10.9b) \( (\theta_c, \mu > 0) \) is also feasible to

\[
||b||_1 \leq \mu(1-||a||_1), \forall \theta_p \in \mathcal{P}(\theta_p)
\]

(10.11)

From \( \bar{\theta}_p \not\in \mathcal{P}(\theta_p) \), \( ||b||_1 > 0 \) holds, and thus (10.10c) holds automatically. Furthermore, combining (10.11) and (10.10c), (10.10b) also satisfied, meaning any feasible solution to (10.9) is also a feasible solution to (10.10) and \( \tilde{\mu}^* \leq \mu^* \). In summary, \( \tilde{\mu}^* = \mu^* \) and (10.9) \( \leftrightarrow \) (10.10) \( \leftrightarrow \) (10.8). \( \square \)

Note that the assumption that \( \mathcal{P}(\theta_p) \) does not contain elements of the form \( \bar{\theta}_p = \left[ \bar{a}_p^T, 0^T \right]^T \) simply rules out the trivial case where the observed output can be explained solely by noise, and hence it is not restrictive.
10.3.2 Convex Reformulation via Robust Optimization

In principle, problem (10.9) is nonconvex due to the bilinear terms appearing in (10.6) and the product of \( \mu \) and the unknown coefficients \( a \) in (10.10). Nevertheless, as we show next, a convex reformulation can be obtained by firstly replacing the optimization over \( \mu \) by a binary search and then using robust optimization to obtain a convex equivalent to the resulting problem. The next results provide the theoretical justification for this approach.

Lemma 7. Given \( \mu > 0 \), the following two statements are equivalent.

(i) Equation (10.9b) is feasible.

(ii) The following set of constraints is feasible:

\[
\mu \left( \sum_{k=1}^{n} S_{i,k} a_k \right) + \sum_{k=0}^{n} S_{i,n+1+k} b_k \leq \mu, \forall \theta_p \in \mathcal{P}(\theta_p), \forall i = 1, \ldots, 2n+1
\]  

(10.12)

where \( S_{i,j} \in \{1, -1\} \) is the \((i, j)\)-th entry of \( S \). Each row of \( S \) denotes a specific sign assignment to \( \theta \), and \( S \) denotes all the possible combinations of signs of \( \theta \).

Proof. It follows directly the definition of the \( \ell_1 \) norm. \( \square \)

Next, we use Theorem 4 to obtain a convex equivalent to the constraints above. Using the explicit expressions for \( \{a_k, b_k\} \) given in (10.7), (10.12) can be rewritten as

\[
f_i(\theta_p, \theta_c) = (\tilde{Q}_i \theta_p + \tilde{f}_i^{(c)})^T \theta_c + \tilde{r}_i \leq 0, \forall \theta_p \in \mathcal{P}(\theta_p), \forall i = 1, \ldots, 2n+1
\]  

(10.13)

where

\[
\begin{align*}
\tilde{Q}_i & = \mu \left( \sum_{j=1}^{n} S_{i,j} Q_j \right) + \sum_{j=n+1}^{2n+1} S_{i,j} Q_j \\
\tilde{f}_i^{(c)} & = \mu \left( \sum_{j=1}^{n} S_{i,j} f_j^{(c)} \right) + \sum_{j=n+1}^{2n+1} S_{i,j} f_j^{(c)} \\
\tilde{f}_i^{(p)} & = \mu \left( \sum_{j=1}^{n} S_{i,j} f_j^{(p)} \right) + \sum_{j=n+1}^{2n+1} S_{i,j} f_j^{(p)} \\
\tilde{r}_i & = \mu \left( \sum_{j=1}^{n} S_{i,j} r_j \right) + \sum_{j=n+1}^{2n+1} S_{i,j} r_j - \mu
\end{align*}
\]  

(10.14)

The feasibility set for \( \theta_c \) is

\[
\mathcal{P}(\theta_c) = \{ \theta_c \in \mathbb{R}^{2n+1} : f_i(\theta_p, \theta_c) \leq 0, \forall \theta_p \in \mathcal{P}(\theta_p), \forall i = 1, \ldots, 2n+1 \}
\]  

(10.15)

Applying Theorem 4 to (10.13) for each \( i \) leads to the following result, showing that feasibility of (10.12) reduces to a Linear Programming problem:
Theorem 21. For a fixed $\mu > 0$, consider the following LP

$$\mathcal{P}(\theta_c, z_i) : \begin{cases} q^T z_i + \bar{r}_i^{(c)} \theta_c + \bar{r}_i \leq 0 \\ -D^T z_i = \bar{r}_i^{(p)} + Q_i^T \theta_c \\ z_i \geq 0 \\ \forall i = 1, \ldots, 2^{2n+1}, \end{cases}$$  \hspace{1cm} (10.16)$$

Then the set

$$\mathcal{P}_{LP}(\theta_c) = \text{Proj}_{\theta_c}(\mathcal{P}(\theta_c, z_i))$$  \hspace{1cm} (10.17)$$

is nonempty if and only if $\mathcal{P}(\theta_c)$ in (10.15) is nonempty, and hence (10.12) and (10.9b) are feasible.

Proof. Follows from combining Lemma 7 and Theorem 4. \hfill \Box

Remark 9. The result above shows that indeed, Problem 12 can be exactly solved by alternating between Linear Programming in $\theta_c$ and a binary search in $\mu$.

10.4 Convex Relaxation

Theorem 21 shows that Problem 12 can be recast into a tractable combination of LP and binary search. However, because of the combinatorial nature of the transformation used to handle $| \cdot |$ in a robust optimization context, the total dimension of the auxiliary variables $\{z_i\}$ in (10.16) is $2^{2n+1} \times 2(T + 1 - n_a)$, and the number of constraints also increases exponentially with the order of the closed-loop system. Thus, the approach above works well if the order of the plant and controller is not too large ($\sim 10$). In order to be able to handle larger sized problems, in this section, we develop several simpler ways of convexifying (10.9b), leading to simpler LPs at the cost of a smaller feasible set for $\theta_c$ than $\mathcal{P}_{LP}(\theta_c)$.

10.4.1 Relaxation 1

This relaxation is based on introducing a new variable $s \in \mathbb{R}^{(2n+1)}$ that is used to decouple and upper bound each term in (10.9b). The main idea is to consider the feasible set for $\theta_c$, $\mathcal{P}_s(\theta_c)$, determined by

$$\mathcal{P}_s(\theta_c) = \text{Proj}_{\theta_c}(\mathcal{P}(\theta_c, s_i))$$  \hspace{1cm} (10.18)$$
where $\mathcal{P}(\theta_c, s_i)$ denotes

\[
\begin{cases}
-s_i + \theta_i \leq 0, \forall \theta_p \in \mathcal{P}(\theta_p), \forall_{i=1}^{2n+1} \\
-s_i - \theta_i \leq 0, \forall \theta_p \in \mathcal{P}(\theta_p), \forall_{i=1}^{2n+1} \\
\mu \sum_{i=1}^{n} s_i + \sum_{i=1+n}^{2n+1} s_i \leq \mu.
\end{cases}
\]

(10.19a) (10.19b) (10.19c)

Since (10.19a)-(10.19b) is equivalent to

\[
s_i = \max_{\theta_p \in \mathcal{P}(\theta_p)} |\theta_i(\theta_p, \theta_c)|,
\]

(10.20)

then (10.19c) guarantees that

\[
\max_{\theta_p \in \mathcal{P}(\theta_p)} \mu |a|_1 + |b|_1 \\
\leq \mu \sum_{i=1}^{n} \max_{\theta_p \in \mathcal{P}(\theta_p)} |\theta_i| + \sum_{i=n+1}^{2n+1} \max_{\theta_p \in \mathcal{P}(\theta_p)} |\theta_i|
\]

\[
\leq \mu \sum_{i=1}^{n} s_i + \sum_{i=1+n}^{2n+1} s_i
\]

\[
\leq \mu.
\]

(10.21)

Therefore, $\mathcal{P}_s(\theta_c)$ is a subset of $\mathcal{P}(\theta_c)$. In addition, $\mathcal{P}_s(\theta_c) = \mathcal{P}(\theta_c)$ holds if and only if $\max_{\theta_p \in \mathcal{P}(\theta_p)} |\theta_i|$ is achieved at the same $\theta_p$ for all $i = 1, \ldots, 2n + 1$.

**Theorem 22.** For a fixed $\mu > 0$, if the set $\mathcal{P}_1(\theta_c) = \text{Proj}_{\theta_c}(\mathcal{P}_1(\theta_c, z_{i,1}, z_{i,2}, s_i))$ is nonempty, where $\mathcal{P}_1(\theta_c, z_{i,1}, z_{i,2}, s_i)$ is determined by the following LP:

\[
\mathcal{P}_1(\theta_c, z_{i,1}, z_{i,2}, s_i) : \begin{cases}
z_{i,1}^T q + f_{i}^{(c)} \theta_c + r_i - s_i \leq 0, \forall_{i=1}^{2n+1} \\
z_{i,2}^T q - f_{i}^{(c)} \theta_c - r_i - s_i \leq 0, \forall_{i=1}^{2n+1} \\
D^T z_{i,1} + Q_i^T \theta_c + f_{i}^{(p)} = 0, \forall_{i=1}^{2n+1} \\
-D^T z_{i,2} + Q_i^T \theta_c + f_{i}^{(p)} = 0, \forall_{i=1}^{2n+1} \\
z_{i,1} \geq 0, z_{i,2} \geq 0, \forall_{i=1}^{2n+1} \\
\mu \sum_{i=1}^{n} s_i + \sum_{i=1+n}^{2n+1} s_i \leq \mu
\end{cases}
\]

(10.22)

then (10.9b) is feasible.

**Proof.** It follows directly from applying Theorem 4 to (10.19a)-(10.19b).
10.4.2 Relaxation 2

Lemma 8. For a deterministic variable $\xi$, its absolute value can be obtained through the optimization problem

$$|\xi| = \minimize \xi^+ + \xi^- \quad \text{subject to} \quad \xi^+ - \xi^- = \xi, \xi^+ \geq 0, \xi^- \geq 0.$$  

From Lemma 8 we can derive upper bounds for $\theta_i$:

$$|\theta_i| \leq 2\theta_i^+ - \theta_i \quad \text{where} \quad \theta_i^+ \geq \max \left\{ 0, \theta_i \right\} \quad (10.23a)$$

$$|\theta_i| \leq \theta_i - 2\theta_i^- \quad \text{where} \quad \theta_i^- \geq \max \left\{ 0, -\theta_i \right\} \quad (10.23b)$$

(10.23a) leads to a tighter set than (10.9b) but in a simpler form $P(\theta_c, \theta_i^+)$:

$$\begin{cases} 
\theta_i \text{ is subject to } (10.7) \\
\theta_i^+ - \theta_i \geq 0, \forall \theta_p \in P(\theta_p), \forall_{i=1}^{2n+1} \\
\theta_i^+ \geq 0, \forall_{i=1}^{2n+1} \\
\mu \sum_{i=1}^{n}(2\theta_i^+ - \theta_i) + \sum_{i=n+1}^{2n+1}(2\theta_i^+ - \theta_i) \leq \mu, \forall \theta_p \in P(\theta_p) 
\end{cases} \quad (10.24a-b-c-d)$$

and similarly a tighter set $P(\theta_c, \theta_i^-)$ from (10.23b) is:

$$\begin{cases} 
\theta_i \text{ is subject to } (10.7) \\
\theta_i^- + \theta_i \geq 0, \forall \theta_p \in P(\theta_p), \forall_{i=1}^{2n+1} \\
\theta_i^- \geq 0 \\
\mu \sum_{i=1}^{n}(2\theta_i^- + \theta_i) + \sum_{i=n+1}^{2n+1}(2\theta_i^- + \theta_i) \leq \mu, \forall \theta_p \in P(\theta_p) 
\end{cases} \quad (10.25a-b-c-d)$$

which result in two sets: $P_+(\theta_c) = \text{Proj}_{\theta} P(\theta_c, \theta_i^+)$ and $P_-(\theta_c) = \text{Proj}_{\theta} P(\theta_c, \theta_i^-)$.

Let $I = \{j \in \mathbb{N} : 1 \leq j \leq 2n + 1\}$. From (10.24a)-(10.24c) we have

$$\theta_i^+ \geq \max \left\{ 0, \max_{\theta_p \in P(\theta_p)} \theta_i \right\}, \quad (10.26)$$

so for those $j \in I_n = \{j \in \mathbb{N} : 1 \leq j \leq 2n + 1, \max_{\theta_p \in P(\theta_p)} \theta_j \leq 0\}$, we get

$$\theta_j^+ = 0, 2\theta_j^+ - \theta_j = |\theta_j|, \forall j \in I_n, \quad (10.27)$$
and
\[
\begin{align*}
\theta^+_j &= \max_{\theta_p \in \mathcal{P}(\theta_p)} \theta_j, \\
2\theta^-_j - \theta_j &\geq \max_{\theta_p \in \mathcal{P}(\theta_p)} \theta_j, \forall j \in I - I_n.
\end{align*}
\] (10.28)

Similarly, from (10.25a)-(10.25c) we have
\[
\theta^-_j = \min \left\{ 0, \max_{\theta_p \in \mathcal{P}(\theta_p)} -\theta_j \right\},
\] (10.29)

so for those \( j \in I_p \doteq \{ j \in \mathbb{N} : 1 \leq j \leq 2n + 1, \min_{\theta_p \in \mathcal{P}(\theta_p)} \theta_j \geq 0 \} \), we get
\[
\theta^-_j = 0, 2\theta^-_j + \theta_j = |\theta_j|, \forall j \in I_p,
\] (10.30)

and
\[
\begin{align*}
\theta^-_j &= \max_{\theta_p \in \mathcal{P}(\theta_p)} -\theta_j, \\
2\theta^-_j + \theta_j &\geq \max_{\theta_p \in \mathcal{P}(\theta_p)} \theta_j, \forall j \in I - I_p.
\end{align*}
\] (10.31)

Thus, feasibility of (10.24d) implies feasibility of (10.9b), and feasibility of (10.25d) implies feasibility of (10.9b). These observations are summarized below.

**Lemma 9.** (10.24d) \( \iff \) (10.9b) and \( \mathcal{P}_+(\theta_c) = \mathcal{P}(\theta_c) \) hold, if \( I = I_n \); (10.25d) \( \iff \) (10.9b) and \( \mathcal{P}_-(\theta_c) = \mathcal{P}(\theta_c) \) hold, if \( I = I_p \).

On the other hand, generally there is no uniform relationship between \( \mathcal{P}_+(\theta_c) \) and \( \mathcal{P}_+(\theta_c) \) (or between \( \mathcal{P}_-(\theta_c) \) and \( \mathcal{P}_-(\theta_c) \)), since for \( \forall j \in I_n, s_j \geq \theta^+_j \), while for \( \forall j \in I - I_n, s_j \leq \theta^+_j \) (or for \( \forall j \in I_p, s_j \geq \theta^-_j \), while for \( \forall j \in I - I_p, s_j \leq \theta^-_j \)).

**Theorem 23.** For a fixed \( \mu > 0 \), if the set \( \mathcal{P}^-_2(\theta_c) \doteq \text{Proj}_{\theta_c}(\mathcal{P}^+_2(\theta_c, z_{i,+}, z, \theta^+_i)) \) is nonempty, where \( \mathcal{P}^-_2(\theta_c, z_{i,+}, z, \theta^+_i) \) is the feasible set to the following LP:

\[
\begin{align*}
\theta^+_i &\geq 0, \forall i = 1, 2n + 1, \\
\mathbf{f}_i^{(c)T} \theta_c - \theta^+_i + z^+_i + \mathbf{q} + r_i &\leq 0, \forall i = 1, 2n + 1, \\
\mathbf{Q}_i^{T} \theta_c + \mathbf{f}_i^{(p)} &= -\mathbf{D}^T z_{i,+}, \forall i = 1, 2n + 1, \\
2\mu \sum_{i=1}^{n} \theta^+_i + 2 \sum_{i=1}^{2n+1} \theta^+_i - \mu \sum_{i=1}^{n} [\mathbf{f}_i^{(c)T} \theta_c + r_i] - \sum_{i=n+1}^{2n+1} [\mathbf{f}_i^{(c)T} \theta_c + r_i] + z^T \mathbf{q} &\leq \mu, \\
\mu \sum_{i=1}^{n} [\mathbf{Q}_i^{T} \theta_c + \mathbf{f}_i^{(p)}] + \sum_{i=n+1}^{2n+1} [\mathbf{Q}_i^{T} \theta_c + \mathbf{f}_i^{(p)}] &= \mathbf{D}^T z, \\
z_{i,+} &\geq 0, \forall i = 1, 2n + 1, z \geq 0
\end{align*}
\] (10.32)
or if the set $\mathcal{P}_2^-(\theta_c) \equiv \text{Proj}_{\theta_c}(\mathcal{P}_2^-(\theta_c, z_{i,-}, z, \theta_i^-))$ is nonempty, where $\mathcal{P}_2^-(\theta_c, z_{i,-}, z, \theta_i^-)$ is the feasible set to the following LP:

$$
\begin{align*}
\theta_i^- &\geq 0, \forall i = 1, \ldots, n, \\
n\theta_i^- - z_{i,-}q - r_i &\leq 0, \forall i = 1, \ldots, n, \\
Q_i^T \theta_c + t_i(p) &\leq D^T z_{i,-}, \forall i = 1, \ldots, n, \\
2\mu &\sum_{i=1}^n \theta_i^- + \mu &\sum_{i=1}^n [f_{i}^{(c)}T \theta_c + r_i], \\
&\sum_{i=1}^n [Q_i^T \theta_c + t_i(p)] \leq &\sum_{i=1}^n [Q_i^T \theta_c + t_i(p)], \\
z_{i,-} &\geq 0, \forall i = 1, \ldots, n, z \geq 0
\end{align*}
$$

(10.33)

then (10.9b) is feasible.

Proof. It follows that by applying Theorem 4 (10.32) and (10.33) are the robust counterparts for (10.24) and (10.25) respectively.

10.4.3 Relaxation 3

This relaxation, motivated by the idea of Affine Adjustable Robust Counterpart (AARC) in [141], is based upon introducing the affine function

$$
s_a(\theta_p) = s^0 + Q_s \theta_p,
$$

(10.34)

where $s^0 \in \mathbb{R}^{2n+1}$ and $Q_s \in \mathbb{R}^{(2n+1) \times (n_a + n_b)}$ are coefficients to be optimized.

Letting $s_a(i)$, $s^0(i)$ and $Q_s(i, :)$ denote the $i$-th entry of $s_a$, the $i$-th entry of $s^0$ and the $i$-th row of $Q_s$ respectively, now consider the feasible set for $\theta_c$, $\mathcal{P}_{s_a}(\theta_c) \equiv \text{Proj}_{\theta_c}(\mathcal{P}(\theta_c, s^0, Q_s))$, with $\mathcal{P}(\theta_c, s^0, Q_s)$ denoting

$$
\begin{align*}
-s^0(i) - Q_s(i, :) \theta_p + \theta_i &\leq 0, \forall \theta_p \in \mathcal{P}(\theta_p), \forall i = 1, \ldots, n, \\
-s^0(i) - Q_s(i, :) \theta_p - \theta_i &\leq 0, \forall \theta_p \in \mathcal{P}(\theta_p), \forall i = 1, \ldots, n, \\
\mu &\sum_{i=1}^n s_a(i) + \sum_{i=1}^{2n+1} s_a(i) \leq \mu, \forall \theta_p \in \mathcal{P}(\theta_p).
\end{align*}
$$

(10.35a)-(10.35c)

Lemma 10. $\mathcal{P}_{s_a}(\theta_c) \subset \mathcal{P}_{s_a}(\theta_c) \subset \mathcal{P}(\theta_c)$.

Proof. (10.35a)-(10.35b) guarantee that

$$
s^0 + Q_s \theta_p \geq |\theta|, \forall \theta_p \in \mathcal{P}(\theta_p),
$$

(10.36)
therefore, if (10.35c) holds for some \( \mu \), so does (10.9b), and thus \( \mathcal{P}_{s_0}(\theta_c) \subset \mathcal{P}(\theta_c) \). Similarly, for any feasible solution \( (\theta_c, s) \) to (10.19), we can construct a feasible solution to (10.35), \((\theta_c, s^0 = s, Q_s = 0)\). Hence \( \mathcal{P}_{s}(\theta_c) \subset \mathcal{P}_{s_0}(\theta_c) \).

**Theorem 24.** For a fixed \( \mu > 0 \), if the set \( \mathcal{P}_3(\theta_c) = \text{Proj}_\theta(\mathcal{P}_3(\theta_c, z_{i,1}, z_{i,2}, z_a, s^0, Q_s)) \) is nonempty, where \( \mathcal{P}_3(\theta_c, z_{i,1}, z_{i,2}, z_a, s^0, Q_s) \) is the feasible set for the following LP:

\[
\begin{align*}
- &\begin{cases} 
  z_{i,1}^T q + f^{(c)}_i T \theta_c + r_i - s^0(i) \leq 0, & \forall i = 1,2, \ldots, n+1 \\
  D^T z_{i,1} + Q^T_i T \theta_c + f^{(p)}_i - Q_s(i,:) = 0, & \forall i = 1,2, \ldots, n+1 \\
  z_{i,2}^T q - f^{(c)}_i T \theta_c - r_i - s^0(i) \leq 0, & \forall i = 1,2, \ldots, n+1 \\
  -D^T z_{i,2} + Q^T_i T \theta_c + f^{(p)}_i + Q_s(i,:) = 0, & \forall i = 1,2, \ldots, n+1 \\
  \mu^T s^0 + z_a^T q \leq \mu \\
  Q_s^T h + D^T z_a = 0 \\
  z_{i,1} \geq 0, z_{i,2} \geq 0, & \forall i = 1,2, \ldots, n+1, z_a \geq 0
\end{cases} \\
\end{align*}
\]

where \( h = \begin{bmatrix} \mu & \ldots & \mu & 1 & \ldots & 1 \end{bmatrix}^T \), then (10.9b) is feasible. 

**Proof.** It follows that by applying Theorem 4, (10.37) is the robust counterpart to (10.35). 

In summary, the advantage of the relaxations introduced in Sections 10.4.1-10.4.3 is a substantial reduction on the number of constraints vis-à-vis the exact formulation (10.16) (from \( O(4^n) \) to \( O(4n) \)), at the cost of a suboptimal \( \mu \).

### 10.4.4 Proposed Algorithm

From section 10.3.2 it follows that if Problem 12 is feasible (e.g. the entire consistency set can be superstabilized by a \( n_c \)-order controller), then a solution that optimizes the \( n_a + n_c \) equalized performance level \( \mu \) can be found by a combination of binary search with respect to \( \mu \) and LP to check the feasibility of (10.9b) for fixed \( \mu \), leading to the following algorithm:

### 10.5 Experiments

In this section we illustrate the proposed approach with a simple example and compare the exact solution against the different relaxations in terms of the number of constraints and performance.
Algorithm 14 Superstable Controller design from Experimental Input/Output Data by RO

1: Initialize $\mu_{\text{min}} \leftarrow 0$, $\mu_{\text{max}} \leftarrow M$, $\mu_{f} = M$, a small number $0 < \delta << 1$ denoting the precision;

2: repeat
3: Fix $\mu$ at $\mu_{f} \leftarrow 0.5(\mu_{\text{min}} + \mu_{\text{max}})$;
4: Check the feasibility of LP (10.16) or LPs corresponding to RC (10.19), (10.24), (10.25), or (10.35);
5: if The above LP is feasible then
6: $\mu_{\text{max}} \leftarrow \mu_{f}$;
7: else
8: $\mu_{\text{min}} \leftarrow \mu_{f}$;
9: end if
10: until $(\mu_{\text{max}} - \mu_{\text{min}})/\mu_{\text{max}} < \delta$.

The experimental input/output data $\{u_t, y_t\}_{t=0}^{21}$ was generated from the following strictly proper 2nd-order plant

$$y_t = -\sum_{k=1}^{2} a_k^{(p)} y_{t-k} + \sum_{k=1}^{2} b_k^{(p)} u_{t-k} + \eta_t,$$

with true parameter values $a_1^{(p)} = 0.8$, $a_2^{(p)} = 1$, $b_1^{(p)} = 0.2$, and $b_2^{(p)} = 0.1$. The input $u_t$ was generated by PRBS generator integrated in MATLAB, and the process noise $\eta_t$ was uniformly distributed in $[-0.1, 0.1]$. The goal was to find a 2nd-order controller capable of super-stabilizing all the strictly proper second order plants consistent with the given data $\{u_t, y_t\}$, while minimizing the 4-equalized performance level $\mu$. Table 10.1 compares the exact solution (obtained solving (10.12) exactly) against the various relaxations. As shown there, (10.12) yields the smallest $\mu$ at the cost of the highest computational complexity, while the relaxation (10.19), (10.24), (10.25), (10.35), substantially reduce the computational complexity, at the price of a larger $\mu$. Among them, the relaxation based on AARC, (10.35), gives an approximation to results of (10.12) of the best quality.

Since the feasible set for the parameters for the plant $\mathcal{P}(\theta_p)$ is convex, we can compute the upper and lower bounds of $\theta_p$ by solving

$$\overline{\theta}_i^{(p)}(\theta_3^{(p)}) = \text{maximize}(\text{minimize}) \quad \theta_i$$

subject to $D\theta_p + q \geq 0$

for $i = 1, \ldots, n_a + n_b$. Then a simple outer herperractangular approximation for $\mathcal{P}(\theta_p)$ is

$$\hat{\mathcal{P}}(\theta_p) = \{\theta_p \in \mathbb{R}^{n_a+n_b} : D\theta_p + q \geq 0\},$$

(10.39)
where $\tilde{D} = \begin{bmatrix} I & -I \end{bmatrix}^T \in \mathbb{R}^{2(n_a + n_b) \times (n_a + n_b)}$, and $\tilde{q} = \begin{bmatrix} -\theta_1^{(p)} & \ldots & -\theta_{n_a+n_b}^{(p)} & \bar{y}_1^{(p)} & \ldots & \bar{y}_{n_a+n_b}^{(p)} \end{bmatrix}^T$.

Obviously, the size of $\tilde{D}$ is dependent on the number of parameters defined the plant and independent on the length of the experimental data. Thus, comparing to the equivalent LP (10.16) derived from $P(\theta^{p})$, (10.16) derived from $\tilde{P}(\theta^{p})$ is simpler at the cost of a larger $\mu$. As shown in Table 10.1, even the relaxations based on the original feasible set $P(\theta^{p})$ achieve smaller $\mu$ than the equivalent LP based on the approximation set $\tilde{P}(\theta^{p})$.

Table 10.1: Comparison between Different LP Formulations for Example 1

<table>
<thead>
<tr>
<th>feasible set of $\theta^{p}$</th>
<th>$P(\theta^{p})$</th>
<th>$\tilde{P}(\theta^{p})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPs</td>
<td>(10.16)</td>
<td>(10.16)</td>
</tr>
<tr>
<td># variables</td>
<td>4101</td>
<td>20485</td>
</tr>
<tr>
<td># constraints</td>
<td>2560</td>
<td>23040</td>
</tr>
<tr>
<td>Optimal $\mu$</td>
<td>2.6544</td>
<td>0.4936</td>
</tr>
</tbody>
</table>

To verify the theoretical results, we applied the controllers resulted from the LPs on the plant as in Figure 10.1. The disturbance $w_t$ was randomly generated between $-1$ and $1$, with the zero initial condition for both the plant and the controllers, the output $e_t$ of different closed-loop systems were shown in Figure 10.2, from which we could conclude each sequence of $e_t$ was within the bound given by its corresponding LP in Table 10.1.
CHAPTER 10. ROBUST SUPERSTABLE CONTROLLER DESIGN FOR LTI SYSTEMS

10.6 Conclusions

This chapter presents a framework for directly synthesizing a robust super-stabilizing controller from experimental data, avoiding an explicit plant identification step. As shown here, imposing superstability, combined with robust optimization arguments allows for reformulating the problem as a tractable LP. A potential difficulty with this approach is that the number of constraints grows exponentially with the problem size, potentially limiting this exact solution to moderately sized problems. In order to handle larger sized problems, the last portion of the chapter introduces several relaxations that substantially reduce the computational complexity at the price of reduced performance. While the results in the chapter apply only to LTI systems, research is currently under way seeking to extend the framework to the switched ARX case.
Chapter 11

Conclusion

In this dissertation, several problems arising in model fitting for a set of noisy experimental data have been considered. We proposed novel frameworks to reformulate these problems, theoretically derived conditions to guarantee the global optimality or consistency, and developed algorithms to solve them efficiently by recent developments in convex optimization. Specifically speaking, the main contributions of this dissertation contains:

- For cases of linear model and switched linear model where the mode variable is finite and discrete, we proposed a general formulation in the form of constrained polynomial optimization problem, which is flexible to incorporate extra prior information either on the model or the data distribution. Appealing to the theory of moments, these generally nonconvex problems could be relaxed into a sequence of monotonically convergent convex semi-definite programs, for which, the exact equivalence is quite difficult to validate: either to prove the positive semi-definiteness of infinite dimensional matrices or to show rank preserveness holds for a finite order relaxation. By exploiting the sparsity pattern exhibited by specific problems and resorting to rank minimization and $\ell_1$-norm minimization techniques, we developed computationally tractable algorithms to achieve optimal solutions to the original nonconvex problem via lower order relaxation.

- For cases of LPV model with LFT parametric dependence where the mode variable is continuous, we proposed a framework based on atomic norm minimization, where atoms are impulse response of all first order or second order LTI systems.

- Contrary to determining a single nominal model and a simple uncertainty set before design a robust controller, to reduce the conservativeness caused by the the convex outer approximation...
CHAPTER 11. CONCLUSION

of the uncertainty set, we established a controller stabilizing all the plants consistent with the feasibility set determined by the experimental data directly without identifying a model.

Besides the results we have obtained so far, there are several related topics worth further research efforts as follows:

- For problems of identifying linear model and switched linear model, all the theoretical results on extracting a consistent model for the data are based on the assumption that a rank-1 solution can be found, which is nonconvex, and relies on its convex relaxation. In spite of the efficiency of rank minimization algorithms, the equivalence between the nonconvex problem and its convex counterpart is established upon some conditions on the data distribution. However, consistent experiments show that for the problem of linear model fitting, without any rank constraint, a moments based relaxation of order 2 can result a low rank solution directly, but there is not theoretical proof. Further research can be carried on to provide some theoretical support to this observation or show it is not generally correct.

- Although the proposed framework for model fitting scales linearly in the number of data samples, which makes it promising for “Big” datasets, however, one impedance prevents it from wide application mainly comes from the poor scaling property of the standard interior-point SDP solver, especially if higher order relaxations are used. A potential solution is given by the first order methods like Alternating Direction Methods of Multipliers (ADMM) or Frank-Wolfe based algorithms rather than interior point ones. An alternative research direction is to research on relax the nonconvex into Second Order Cone Programs which can be solved efficiently by Gurobi rather than semi-definite problems.
Chapter 12

List of Publications

The following list includes all the papers published or submitted for publication by the author during her graduate studies.


CHAPTER 12. LIST OF PUBLICATIONS


Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


