Non-Linear System Identification Using Compressed Sensing

by

Manjish Arvind Naik

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved November 2011 by the
Graduate Supervisory Committee:

Douglas Cochran, Chair
Narayan Kovvali
Matthias Kawski
Rodrigo Platte

ARIZONA STATE UNIVERSITY

December 2011

ABSTRACT

This thesis describes an approach to system identification based on compressive sensing and demonstrates its efficacy on a challenging classical benchmark single-input, multiple output (SIMO) mechanical system consisting of an inverted pendulum on a cart. Due to its inherent non-linearity and unstable behavior, very few techniques currently exist that are capable of identifying this system. The challenge in identification also lies in the coupled behavior of the system and in the difficulty of obtaining the full-range dynamics.

The differential equations describing the system dynamics are determined from measurements of the system's input-output behavior. These equations are assumed to consist of the superposition, with unknown weights, of a small number of terms drawn from a large library of nonlinear terms. Under this assumption, compressed sensing allows the constituent library elements and their corresponding weights to be identified by decomposing a time-series signal of the system's outputs into a sparse superposition of corresponding time-series signals produced by the library components.

The most popular techniques for non-linear system identification entail the use of ANN's (Artificial Neural Networks), which require a large number of measurements of the input and output data at high sampling frequencies. The method developed in this project requires very few samples and the accuracy of reconstruction is extremely high. Furthermore, this method yields the Ordinary Differential Equation (ODE) of the system explicitly. This is in contrast to some ANN approaches that produce only a trained network which might lose fidelity with change of initial conditions or if facing an input that wasn't used during its training.

i

This technique is expected to be of value in system identification of complex dynamic systems encountered in diverse fields such as Biology, Computation, Statistics, Mechanics and Electrical Engineering.

This work is dedicated to my brother, Aashish.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The methodology of building a mathematical model of a dynamical system from the measurements of its observable outputs, in response to a particular stimulus, constitutes the basic concept of System Identification. System Identification aims at devising a mathematical relationship between such sets of inputs and outputs.

System Identification has gained a considerable amount of importance due to the fact that devising laws to bound the behavior of systems is the first step in studying their properties. Once we can establish the behavior of the system we can use this information to either control the system in order to produce a desired effect or to predict future system response to specific inputs. Due to this, control systems like say reactor temperature control of a chemical process depend significantly on this mathematical understanding of the plant temperature model. Hence the first task in designing controllers for a system (i.e. plant) is to deduce the mathematical equations governing the plant dynamics. Then we use this to either obtain the transfer function (Laplace analysis, in the linear case) or the state space model and we appropriately design the controller like PID, state feedback or observer based controller for this model.

System identification for nonlinear systems is well recognized to be a challenging problem, and the most effective methods usually rely on strong assumptions; e.g., that the system is operating near equilibrium. To identify such systems, researchers often try function approximations using mathematical series like Volterra or Wiener [2]-[4].

In recent years, compressed sensing has gained a significant amount of research interest [5]-[10]. The Shannon-Nyquist Sampling Theorem suggests that to accurately reconstruct a signal, it should be sampled at a frequency which is at least twice its bandwidth. In contrast, compressed sensing performs accurate reconstruction of signals which are sparse in a known basis, from samples collected at sub-Nyquist frequencies. Random or deterministic linear measurements of such sparse signals are taken and later optimization techniques are used to reconstruct the original signal from these sets of measurements. The sampling rate required and the success and accuracy of the reconstruction depend significantly on the level of sparsity of the signal in that particular basis.

This project describes a method for nonlinear system identification that uses compressive sensing in a way that builds upon an idea introduced by Wang et al. [28]. A library of functions of the inputs and outputs of the system are generated using a power series expansion, and then the *Basis Pursuit* technique is used to obtain the correct weights of these functions in the system equation.

The method is demonstrated with an inverted pendulum system operating well away from its equilibrium point. This well-studied dynamic system has always been of interest to researchers working in the field of system identification [34]. The main control objective of this system is to invert the pendulum completely and make it stand at 180 degrees from its initial rest position by moving the cart in a horizontal plane. In order to design such a controller the system needs to be accurately identified first, which in itself is a big challenge.

The approach described is shown to be effective in accurately identifying the system from among a rich set of candidate system models.

1.1   Organization of Thesis

The remainder of the thesis is organized as follows. Chapter 2 talks in detail about the fundamentals of System Identification and its mathematical treatment. Chapter 3 gives the basic concepts of Compressed Sensing and talks in detail about the mathematics involved. Chapter 4 describes the method for system identification, briefly as given in [28] and also with the extensions introduced here to enable its application to coupled nonlinear systems. Chapter 5 describes the application of this extended method to the SIMO inverted pendulum system and summarizes and discusses the results obtained. Chapter 6 concludes the thesis and, as an after-note, outlines areas for future work.

Chapter 2

SYSTEM IDENTIFICATION

This chapter establishes a fundamental understanding of system

identification and explains the various processes associated with it. The first

section talks about the basic concepts of system identification and defines the

main terms involved. The next section gives a rigorous mathematical treatment

of these concepts. A compilation of various techniques of system identification

is given in [29], which also gives a good theoretical background of the subject.

2.1   Dynamic Systems

Dynamical systems are ubiquitous in both natural and engineered

systems. A significant fraction of engineered dynamical systems are

mechanical systems such as cars, elevators or the various machines that

make our lives easier. These dynamic systems can be affected using external

stimuli which may be manipulated by a design engineer. Such a stimulus is

called the *input* to the system and is said to "drive the system". The system

responds to such stimuli based on internal parameters that define its behavior.

Such a response is often the signal of interest, called the *output* of the system,

and its observability is the key to identifying the system behavior. Also there

might be other factors like *disturbances* which cannot be estimated or

manipulated by the design engineer but might still exist and produce certain

effects on the outputs. An example would be that of a car, where the push of a

gas pedal (flow of fuel to the engine) might be the *input*, which produces the

displacement of the car as the *output*, subject to the friction of the road or air

resistance, which can be considered as *disturbances* in this system.

Establishing these three types of variables of a dynamical system, our

next step is in understanding how these variables relate to each other. Such a

relationship is called a *model* of the system. A model may take several forms, but for the purpose of this thesis we are concerned with the classical cores of dynamic continuous-time systems modeled by a system of Ordinary Differential Equations or discrete systems modeled by a system of Difference Equations. Furthermore, we are particularly interested in non-linear, multi-input, multi-output (MIMO) dynamical systems since they offer many challenges when it comes to their model identification. To demonstrate our methodology, we will consider mechanical systems belonging to this category since it is very hard to model such systems-especially when they are made up of many smaller flexible members [30].

2.2   Modeling of Systems

When it comes to mathematical modeling there are primarily two techniques of determining the system model. The first technique deals with deriving the system equations based on certain energy balance or mass balance equations. This technique is known as *modeling* and the relationships for such equations are based on well-known formulae derived from earlier empirical works. The second technique known as *system identification* is based on experimentation, where the input and output data of the system are measured and the model is inferred using data fitting tools. Hence it is extremely vital to experiment suitably on the system to gain maximum useful information about it in order to generate accurate identified models. This design of experiments is performed by the design engineer who determines suitable choices for the input signals, the sampling time and the initial conditions of the system. Once the experimental data is obtained, a class of models is selected based on engineering intuition, physical insight and *a priori* knowledge. The identification scheme where parameter adjustment is used for

5

data fitting, but these parameters don't have any real physical significance is known as *black box identification*. The case where parameters are known to represent certain physical quantities is known as *grey box identification*. The best fitting model is selected by comparing the output of the identified model and the actual system for the same input.

The various steps of the identification process and their inputs and outputs are made clear through the flow charts in Fig. 2.1. *A priori* knowledge, intuition, physical laws and empirical formulae form the basis for selection of the class of models. Fig. 2.2 shows how the selected model from the set of data models is validated. We see that the criterion for validation is the error between the plant output and the selected model output. Various error metrics might be used such as the Mean Square Error (MSE) or the Absolute Error (AE). An acceptable bound on this error is application specific and depends on the level of accuracy expected. In some applications an error to signal ratio of -60 dB may be considered small enough and the corresponding identified model may be considered a good estimate. If the model doesn't satisfy the desired error bound, we feed the next model from the set and again check for the error condition. We can either stop the process after getting the first model that satisfies the error condition or we can continue to run all the models in the set and then select the one with the least error.
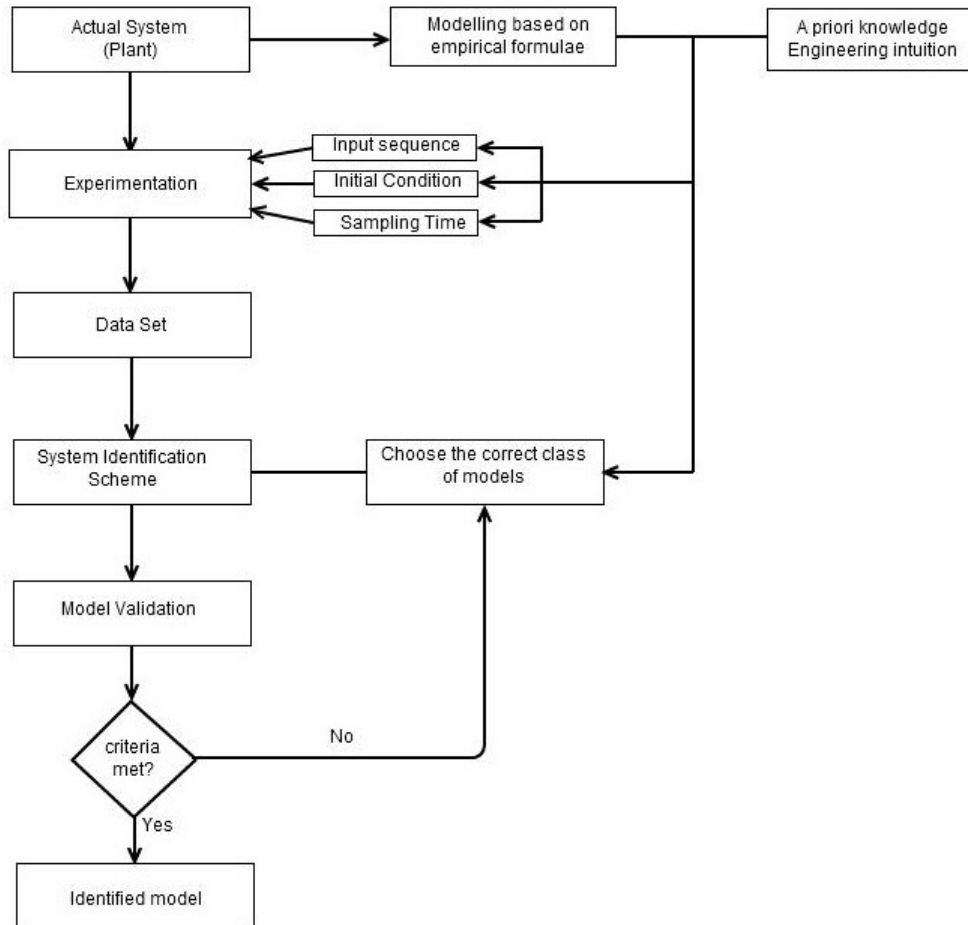
Figure 2.1: Flow chart illustrating the system identification process.
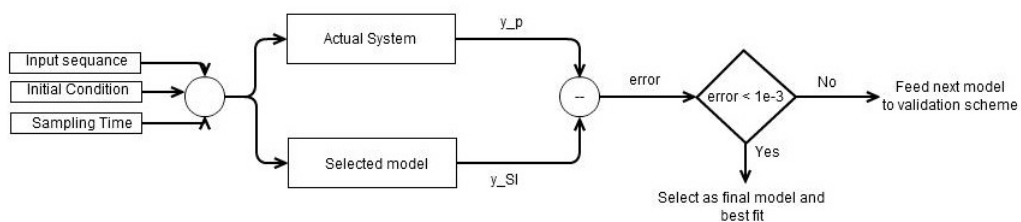
Figure 2.2: Flow chart for validation process of the models from selected set.

2.3   Mathematical Treatment of System Identification

The prevalent method in representing dynamic systems in control

engineering, is by using vector differential or vector difference equations. For

example, we express a m input, p output and n state time invariant and memoryless system as:

$$\dot{x}(t) = \phi[x(t), u(t)]$$
$$y(t) = \psi[x(t)]$$

(2.1)

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$.

In this expression, $x_i(t)$ are the n states of the system, $u_i(t)$ are the m inputs of the system and $y_i(t)$ are the p outputs of the system. The functions ϕ and ψ are static non-linear maps defined as $\phi : \mathbb{R}^m \to \mathbb{R}^n$ and $\psi : \mathbb{R}^n \to \mathbb{R}^p$. The equations (2.1) constitute the *Input-State-Output* representation of the system. The corresponding representation for discrete time systems is shown in Equation (2.2).

$$x(k+1) = \phi[x(k), u(k)]$$
$$y(k+1) = \psi[x(k)]$$

(2.2)

All these representations are for a generic class of systems. Linear Time Invariant (LTI) systems are special case of Equations (2.1) and (2.2) where the maps ϕ and ψ are linear. The non-linear mappings ϕ and ψ get transformed into the matrices A, B and C, where A is the *State Matrix*, B is the *Input matrix* and C is the *Output Matrix*. These are of dimensions $n \times n$, $n \times m$ and $p \times n$ respectively for a n state, m input and p output system. The system *State Space* representation now becomes:

$$\dot{x} = Ax + Bu$$
$$y = Cx$$

(2.3)

System Identification theory for such LTI systems is highly developed. The most common methods employ parametric estimation techniques commonly using the Least Squares Estimator as described in [31]. In case of grey-box modeling, these methods assume a structure of the system (e.g., an nth order polynomial) and then estimate the relative weights or the value of the parameters of the terms in the structure. While for black box modeling, no structural information is available, still certain properties about the system function can often be assumed; e.g., it is an analytic function. But it is still very difficult to perform analysis and identification of non-linear systems as expressed in equation (2.1). For our discussions, we will be concentrating more on such systems.

<center>*System Characterization and Identification*</center>

System characterization deals with the mathematical operator P from an input space U to an output space Y_o and the objective is to determine the class $\wp$ to which P belongs. While on the other hand given a $P \in \wp$, *System Identification* deals with determination of a class $\hat{\wp}$ and an element $\hat{P} \in \hat{\wp}$ so that $\hat{P}$ approximates P in some desired sense.

In dynamical systems P is defined by using the input-output pairs $u(t), y(t), t \in [0, T]$. The objective is to determine $\hat{P}$ such that:

$$||\hat{y} - y||_{l_2} = ||\hat{P}(u) - P(u)||_{l_2} \leq \epsilon, u \in U \qquad (2.4)$$

i.e., the difference between the identified output and the real system output for the same input should be within a small tolerance limit $0 < \epsilon < 1$. This difference is also termed as the *identification error* $e = \hat{y} - y$. Please note that though the above equation specifies an l_2 norm, any suitably defined norm (denoted by $||\cdot||$) can be used.

<center>9</center>

The ability to approximate the plant arises due to the following theorem:

Theorem 2.3.1. (Weierstrass Theorem) *Let $C([a,b])$ be a set of continuous valued functions defined on the interval $[a,b]$ with the norm of $f \in C([a,b])$ defined by*

$$||f|| = \sup_t \{|f(t)|\colon t \in [a,b]\} \tag{2.5}$$

For any $f \in C[a,b]$ and any $\epsilon > 0$, there is a polynomial $p(t) = a_0 + a_1 t + \ldots + a_l t^d$, such that $||(f-p)|| < \epsilon$.

Naturally the Weierstrass theorem and its extensions for high dimensional data find wide applications in approximation of nonlinear functions $f : \mathbb{R}^n \to \mathbb{R}^m$ using polynomials. A generalization of this theorem is the Stone-Weierstrass theorem [32].

Theorem 2.3.1. (Stone Weierstrass Theorem) *Let U be a compact metric space. If $\wp$ is sub-algebra of $C(U, \mathbb{R})$ which contains the constant functions and separates points of U, then $\wp$ is dense in $C(U, \mathbb{R})$.*

Corollary. *If P is the plant to be identified and if $P \in \wp$, where $\wp$ is a space of continuous, bounded, time-varying and causal functions, then if $\hat{P}$ satisfies the Stone Weierstrass theorem, then a model $\hat{P} \in \hat{\wp}$ can be chosen which approximates $P \in \wp$. [33]*

Using the Stone Weierstrass Theorem it has been shown in [1] and [3] that a large class of non-linear functions under certain conditions can be represented by a corresponding series such as Volterra series or Wiener series. In spite of this, very few of these techniques have found wide application in identification of a large class of practical non-linear systems.

Following from these two theorems now we define the actual problem of system identification as follows:

The input and output of a time-invariant, causal discrete time plant are $u(\cdot)$ and $y_p(\cdot)$, respectively, where $u(\cdot)$ is a uniformly bounded function of time. The objective is to construct a suitable identification model which, when subjected to same input $u(k)$ as the plant produces an output $\hat{y}_p(k)$ which approximates the plant output $y_p(k)$ in some desired sense as (2.4).

Chapter 3

COMPRESSED SENSING

Over the past few years a new technique called *Compressed Sensing*
has emerged for reconstructing signals that are sparse in a known basis, from
relatively fewer measurements or samples than suggested by the
Shannon-Nyquist Sampling Theorem. This technique has been applied in
various fields such as image processing [13]-[16], medical imaging [17]-[20],
computational biology[21]-[23], audio and speech processing [24]-[26] and
many others.

3.1   Introduction

The well-known Shannon/Nyquist Sampling Theorem says that in order
to exactly recover (in the sense of L^2) an arbitrary real-valued,
continuous-time, band-limited signal from uniformly spaced samples, the
samples need to be taken at a frequency which is greater than or equal to
twice the signal's bandwidth. This critical sampling frequency is often called
the *Nyquist Frequency*.

But often, it's impractical to obtain these samples of the signal due to
many reasons. Primarily, construction of some sensors, like the ones used in
medical imaging, might require certain level of sophistication, resulting in very
high costs. Also, in other cases, the signal of interest might have some high
frequency components in it and the sensor will have to be capable of extremely
fast sampling, in order to prevent aliasing. This puts a major constraint on the
sensor's capacities and the designer might find it very difficult to cope with this
requirement. In addition to this, the sensor might generate a large number of
samples and will require a large storage space for the sampled data. A
common example of this scenario is evident from the large file sizes of videos

recorded using high resolution cameras. Also, the transmission of such high frequency sampled signals might require expensive communication equipment. This problem is often encountered in online video streaming and the resolution of such videos has to be scaled down to prevent huge buffering durations.

In order to deal with these difficulties, the signal being sampled needs to be compressed and the most concise representation is sought while staying within the maximum resulting distortion bound. The two main concepts that enable such concise representations are: *Sparsity* and *Compressibility*. *Sparsity* implies that the information rate of a continuous time signal is much less than the actual bandwidth [7] or that most amplitude coefficients of a discrete-time signal are zero and only a few have non-zero amplitudes. *Compressibility* implies the ability of representing sparse signals concisely when they are represented in a proper basis, Ψ or that most amplitude coefficients of a discrete-time signal are very small or negligible and only a few have large amplitudes.

A technique underpinning most modern compression algorithms is transform coding, an optimal variant of which is called the *Karhunen-Loeve Transform* (KLT). For a signal, such as an image, sparse in the Fourier basis, with only a few non-zero coefficients, the KLT identifies these main contributors and discards the others, maintaining the essence of the signal. This process is also called *Sparse Approximation* and forms the basis of compression standards like JPEG, JPEG2000 or MP3. But transform coding requires tremendous amount of computation and storage as it computes all the coefficients and then decides which of them to discard.

Taking inspiration from this idea of sparse approximation, a new ideology has emerged, known as *Compressed Sensing*. Compressed Sensing is highly effective for signals which are sparse in a particular basis and can

accurately reconstruct them using far fewer linear measurements than suggested by the Sampling Theorem. Since Compressed Sensing uses fewer measurements than unknowns, the linear system describing the sensing process, is highly under-determined and will have many or none solutions. But the fact that the signal is known to be sparse in that particular basis gives reason to choose only the sparsest from amongst the infinitely many possible solutions.

The effects produced by the use of compressed sensing in various fields have given promising and inspiring results. As mentioned in [27] the speed of Magnetic Resonance Imaging has gone up by a factor of seven, still managing to preserve the diagnostic quality of the image. Computer Aided Tomography (CAT) also has benefited from the use of Compressed Sensing as accurate scans are obtained in far fewer measurements, reducing the risk of over-exposure to radiation [10].

3.2   Mathematical Treatment

Suppose we have a discrete time signal X having N samples. We represent such a signal in form of a $N \times 1$ vector. Suppose this signal is sparse in an orthonormal basis Ψ, then we can represent the signal, X as a form of weighted vector of this representation basis as:

$$X = \sum_{i=1}^{N} s_i \psi_i \text{ or } X = \Psi s \tag{3.1}$$

where Ψ is an $N \times N$ matrix whose columns are the basis vectors and s is an $N \times 1$ vector of weighting coefficients. The sparsity assumption implies most of the weights, s_i will be zero.

Now we take M linear measurements of such a signal by using an $M \times N$ sensing matrix $\{\Phi_{j=1}^{M}\}$ with Φ_j as rows. Using this we get the $M \times 1$ measurement vector, Y.

14

$$Y = \Phi X = \Phi \Psi s = \Theta s \tag{3.2}$$

In order to obtain the accurately reconstructed signal from the measurement vector Y, the following objectives must be met:

1) Design Φ for a K-sparse signal. (K must be generally considered smaller than M).

2) Design a reconstruction algorithm to solve $Y = \Theta s$ in order to get the optimally sparse s.

Designing the Sensing Matrix, Φ

The fundamental property of compressive sensing matrix Φ was introduced by Candés in [10] and independently by Donoho in [11]. Candés called it the *Uniform Uncertainty Principle* (UUP); while Donoho called it as the *Restricted Isometry Property* (RIP). This property implies that the $M \times N$ sensing matrix Φ satisfies a "Restricted Isometry Condition". This idea was later refined in [12].

Property 3.2.1. The Restricted Isometry Property

Let $\Phi_A, A \subset 1, ..., N$; having cardinality, K be the $M \times K$ sub-matrix of Φ, obtained by extracting the columns of Φ corresponding to the indices in A; then [12] defines the K-restricted isometry constant δ_K of Φ_A which is the smallest quantity such that:

$$(1 - \delta_K)||X||_{l_2}^2 \leq ||\Phi_A X||_{l_2}^2 \leq (1 + \delta_K)||X||_{l_2}^2 \tag{3.3}$$

This implies that if Φ satisfies the RIP, then it approximately preserves the Euclidean length of every K-sparse signal. Equivalently, all subsets of K columns taken from Φ are nearly orthogonal. Since the Euclidean lengths are

preserved, it suggests that accurate reconstruction of the sparse signal is possible with high probability.

Designing the Reconstruction Algorithm

If RIP holds for the sensing matrix Φ, then the sparse reconstruction s of the signal X is obtained by solving the linear program [6]:

$$\min_{s \in \mathbb{R}^N} ||s||_{l_1}, \text{ subject to } \Theta s = Y \tag{3.4}$$

The reason why we use the l_1 norm is explained in [6], where the sparse vector s lies on a K-dimensional hyper plane aligned close to the axes due to the sparseness and most of the entries being zero. The l_2 ball intersects this hyper plane at a point far away from the axes and hence an optimal reconstruction is not obtained. But on the other hand, the l_1 ball has a diamond structure and it intersects this hyper plane at a point much closer to the axes and hence gives a better result. As the dimensions increases this l_1 structure becomes pointier and gives better and optimal reconstructions.

Chapter 4

METHODOLOGY OF SYSTEM IDENTIFICATION

This section deals with the actual methodology used to perform non-linear system identification using Compressed Sensing. We study the original technique proposed by Wang et al. in [28], then we identify some important drawbacks which make it difficult to apply the technique to complex dynamic systems, especially those that exhibit coupled behavior. After that we draw some inferences, and based on these inferences, we propose our modifications to the existing technique.

4.1   Problem Formulation

Many mechanical, electrical, chemical, biological, and other important systems are governed, or at least well modeled, by phenomena that cause their input and state variables to be related in constrained ways. In mechanical systems, for example, forces, positions, velocities, and accelerations are all related by ordinary differential equations of order no greater than two. It is also often the case that observation of, or prior knowledge about the behavior of, the system can suggest or impose relationships (possibly nonlinear) among the input and state variables. In such situations, it is possible to postulate a *library* of functions and derivatives of these variables that are candidates to appear in a differential equation describing the system dynamics.

Although constrained, such a library may be large. The key assumption that enables the use of compressive sensing ideas in this context is that the actual system is a weighted superposition of only a small number of the library elements. With this assumption, the system identification problem becomes that of identifying which library elements are present, and with what weights, in the system based on observations of its inputs and outputs. To achieve this,

17

corresponding time series segments from the system inputs and outputs and from all of the library elements are computed and form the entries of the sensing matrix Φ. The corresponding unknown weights are then calculated by sparse reconstruction using optimization techniques such as the *Basis Pursuit Method*.

4.2 The Method of Wang et al.

Consider a non-linear dynamic system which can be represented by the equation $\dot{x} = F(x)$ where x is the state vector containing the states $[x_1, x_2, x_3, \cdots, x_n]^T$ of the system. Wang et al. propose in [28] that the kth component of $F(x)$ can be written as a power series expansion as:

$$[F(x)]_k = \sum_{l_1=0}^{P} \sum_{l_2=0}^{P} \cdots \sum_{l_n=0}^{P} [a_j]_{l_1, l_2..l_n} x_1^{l_1} \cdot x_2^{l_2} \cdots \cdots x_n^{l_n} \qquad (4.1)$$

Where, P is the truncation of the power series and the coefficient vector $a = [a_j]_{l_1, l_2 \cdots l_n}$ is to be determined from the time series data. If the time series data for the states are available at time instants $t_1, t_2 \cdots t_w$ then we can write

$$[F(x(t))]_1 = g(t)a \qquad (4.2)$$

where

$$g(t) = [x_1(t)^0 x_2(t)^0 \cdots x_n(t)^0, x_1(t)^0 x_2(t)^0 \cdots x_n(t)^1,$$
$$\cdots, x_1(t)^{l_1} x_2(t)^{l_2} \cdots x_n(t)^{l_n})]$$

Therefore, now this can be written as a familiar system of linear equations as: $Y = \Phi \cdot X$. where $Y = [F(x(t))]_1 = [\dot{x}_1(t_1), \dot{x}_1(t_2), \dot{x}_1(t_w)]^T$ Therefore:

$$Y = \begin{bmatrix} \dot{x}_1(t_1) \\ \dot{x}_1(t_2) \\ \vdots \\ \dot{x}_1(t_w) \end{bmatrix} = \begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_w) \end{bmatrix} (a) \qquad (4.3)$$

18

Thus we construct the matrix Φ from the power series expansion where the columns denote the various terms of the power series and the rows denote the time series evaluation of these terms at the particular time instant: t_k.

4.3   The Modified Technique

Given this background we see two main drawbacks of such a scheme: Firstly in control applications, most dynamic systems, especially mechanical systems have a forcing input or excitation term u (e.g. the force which drives the cart in an inverted pendulum system) and the generalized system equation is represented as $\dot{x} = F(x, u)$. Secondly, for coupled systems, it is difficult to express $F(x, u)$ as in Equation 4.1. To explain this, we will consider the following Ordinary Differential Equations describing the behavior of the inverted pendulum system, the details of which will be covered in Chapter 5.

$$\ddot{y} = \frac{1.29F - 0.0645\dot{y} + 0.0257\dot{\theta}\cos\theta + 3.477\sin\theta\cos\theta + 0.1774\dot{\theta}^2\sin\theta}{1 - 0.3548\cos^2\theta} \quad (4.4)$$

$$\ddot{\theta} = \frac{-0.145\dot{\theta} - 19.6\sin\theta - 2.58F\cos\theta + 0.129\dot{y}\cos\theta - 0.3548\dot{\theta}^2\sin\theta\cos\theta}{1 - 0.3548\cos^2\theta}$$

$$(4.5)$$

Converting these two Ordinary Differential Equations to a standard representation $\dot{x} = F(x, u)$, the state vector $x = [x_1, x_2, x_3, x_4]^T = [y, \dot{y}, \theta, \dot{\theta}]^T$ is:

$$\dot{x}_1 = x_2 \quad (4.6)$$

$$\dot{x}_2 = \frac{1.29u - 0.0645x_2 + 0.0257x_4\cos x_3 + 3.477\sin x_3\cos x_3 - 0.1774x_4^2\sin x_3}{1 - \cos^2 x_3} \quad (4.7)$$

$$\dot{x}_3 = x_4 \quad (4.8)$$

$$\dot{x}_4 = \frac{-0.15x_4 - 19.6\sin x_3 - 2.58u\cos x_3 + 0.13x_2\cos x_3 - 0.35x_4^2\sin x_3\cos x_4}{1 - \cos^2 x_3} \quad (4.9)$$

Therefore, due to the denominator terms in $\dot{x}_2$ and $\dot{x}_4$, several non-linear terms will be needed to represent the state space equation using a power series.

19

Due to this, the assumption of the weights of library terms being a sparse vector will no longer hold true.

Now after establishing these two main drawbacks, it's evident that the methodology of Wang et al. needs considerable modification in order to be applied to reconstruct the differential equations of the Inverted Pendulum system. We now take a different approach to solve this problem. We first lay the groundwork by some preliminary inferences:

1) Since it's a mechanical system, the ODE will have maximum order of 2. This is a reasonable assumption considering the physics and understanding that main physical quantities involved are displacement, velocity and acceleration.

2) There will be two Ordinary Differential Equations to describe the system completely. This also is reasonable considering there are two motions in space: the motion of the cart and the motion of the pendulum.

3) Since it is a system exhibiting oscillatory behavior the power series expansion will have sinusoids of θ along with the polynomial terms.

Considering these inferences we now develop our library of basis functions and the terms involved in the power series expansion. They are,

$$\Phi = [y, \dot{y}, \ddot{y}, \theta, \dot{\theta}, \ddot{\theta}, \sin\theta, \cos\theta, F] \tag{4.10}$$

where y is the displacement of the cart, θ is the angle between the pendulum and the vertical in degrees and F is the force applied to the cart or the excitation signal.

The power series now looks like:

$$\ddot{y} = \sum_{l_1=0}^{P}\sum_{l_2=0}^{P}\cdots\sum_{l_9=0}^{P}[a_{y_j}]_{l_1,l_2..l_9}y^{l_1}\dot{y}^{l_2}\ddot{y}^{l_3}\theta^{l_4}\dot{\theta}^{l_5}\ddot{\theta}^{l_6}\sin^{l_7}\theta\cos^{l_8}\theta F^{l_9} \tag{4.11}$$

20

$$\ddot{\theta} = \sum_{l_1=0}^{P}\sum_{l_2=0}^{P}\cdots\sum_{l_9=0}^{P}[a_{\theta j}]_{l_1,l_2..l_9}y^{l_1}\dot{y}^{l_2}\ddot{y}^{l_3}\theta^{l_4}\dot{\theta}^{l_5}\ddot{\theta}^{l_6}\sin^{l_7}\theta\cos^{l_8}\theta F^{l_9} \qquad (4.12)$$

We take the maximum power of the expansion, $P = 2$. Also it is important to note that when we construct the Φ matrix using these terms, we have to eliminate the column representing the term $\ddot{y}$ from equation (4.11) and the term $\ddot{\theta}$ from equation(4.12) for obvious reasons. After eliminating these columns, we obtain two sensing matrices: Φ_y used in equation(4.11) and Φ_θ used in equation (4.12). Thus we get the linear system of equations:

$Y_1 = \ddot{y} = \Phi_y a_y$ and $Y_2 = \ddot{\theta} = \Phi_\theta a_\theta$.

Now using compressed sensing we reconstruct the coefficient vectors a_y and a_θ. It is important to note that these coefficient vectors are sparse due to the fact that only a few terms amongst all the terms of the power series expansion in Φ will form the ODE for $\ddot{y}$ and $\ddot{\theta}$.

In order to get accurate reconstruction from these two systems of linear equations using compressed sensing, Φ_y and Φ_θ should satisfy the Restricted Isometry Property (RIP) [10]. We ensure this by normalizing the columns of Φ_y and Φ_θ by dividing each element in the column with the l_2 norm of the column as explained in [28]. Though it is combinatorially quite complex to verify that these normalized matrices satisfies the RIP, we can estimate its performance with great accuracy for a sufficient number of Monte Carlo trials. After conducting several combinatorial trials on the matrices it was found that,

$1 - \delta_k \leq \frac{||\Phi_y a_y||_{l_2}^2}{||a_y||_{l_2}^2} \leq 1 + \delta_k$ and $1 - \delta_k \leq \frac{||\Phi_\theta a_\theta||_{l_2}^2}{||a_\theta||_{l_2}^2} \leq 1 + \delta_k$.

Table 4.1: Monte Carlo trials for testing RIP of Φ_y and Φ_θ matrices

	$\dfrac{||\Phi_y a_y||^2_{l_2}}{||a_y||^2_{l_2}}$	$\dfrac{||\Phi_\theta a_\theta||^2_{l_2}}{||a_\theta||^2_{l_2}}$
Minimum	0.1901	0.7875
Mean	0.9988	1.0013
Maximum	1.9679	1.2137

To quantify these trials, the minimum, maximum and mean values are represented as shown in table 4.1. These trials show that the mean of the δ_k is very small and almost zero in both Φ_y and Φ_θ matrices. But some combination of K elements in the Φ_y matrix may lead to difficulty in reconstruction due to their large δ_k values as seen for the minimum and maximum case. But generally speaking, the Φ_y and Φ_θ matrices can satisfy the RIP with great probability for a small RIP constant δ_k.

We obtain the sparse reconstruction of these coefficient vectors using the Basis Pursuit technique as:

$$\min ||a_y||_{\ell_1} \text{ s.t. } \Phi_y a_y = \ddot{y}$$

$$\min ||a_\theta||_{\ell_1} \text{ s.t. } \Phi_\theta a_\theta = \ddot{\theta}$$

Chapter 5

THE INVERTED PENDULUM ON A CART

To illustrate and demonstrate the approach developed by this project, it has been applied to the system identification problem for an inverted pendulum on a cart. This system has always been of research interest due to its non-linear, unstable and non-minimum phase dynamics. It also is an under-actuated system having more degrees of freedom than control inputs and the full state is not always measurable, and so the identification objectives from such limited dynamics is always a challenge. This system is widely used in robotics, control theory, computer control and space rocket guidance systems.

The single inverted pendulum on a cart system has two frictional components: between the pendulum axis and the pivot joint and between the cart wheels and the track. It becomes very difficult to model these two components for the reasons stated in [34], adding to the complexity of the identification problem. This chapter describes this system and establishes the mathematical model for the system which is later used in simulation to obtain measured time-series input and output data.

5.1   Physical Model

This physical system is shown in Fig 5.1. This system consists of a rail on which a cart moves and a pendulum hinged on the top of the cart. The cart and pendulum rod are constrained to move within a vertical plane. The pendulum rod is free to rotate about its pivot joint on the cart. The cart acceleration induces a torque on the free-moving pendulum to swing it up. Since the pendulum is exactly centered above the cart, there are no sidelong forces on the rod and it remains balanced upright. But any small disturbance in

the motion of the cart for this balanced pendulum shifts it farther away from the upright position, indicating that the upright is an unstable equilibrium point. Also under no external force the oscillations decay and the rod comes to rest at the 0° position which is also the stable equilibrium, known as the *pendant* position. A force F is applied to the cart which is the *input* to the system.
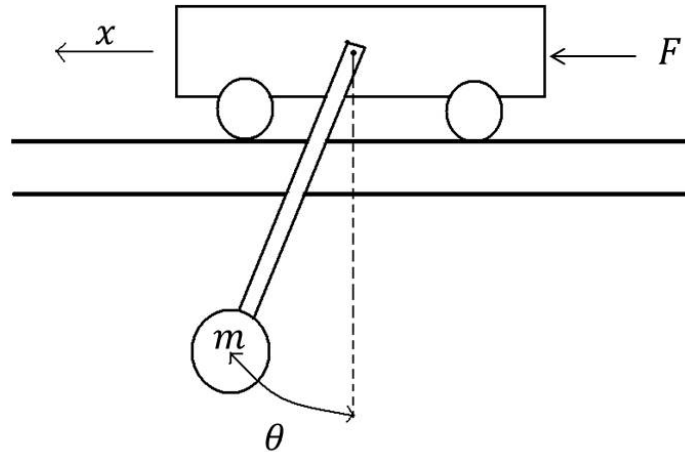

Figure 5.1: Physical model of the inverted pendulum on cart system.

Positive value of F causes the cart to move towards left and negative value to the right. This is due to the forward and reverse rotations of the motor driving the cart and suitable mechanical amplification of this motion is achieved by a gearing system. The angle θ of the pendulum rod with the vertical is measured and is one of the outputs of the system, while the displacement of the cart x is the other output. The initial conditions of this system are at zero torque where $\theta = 0$. Due to the Right Half Plane zeros of this system the typical *inverse response* is seen, where for the cart to move to the right, it must first move to the left and unbalance the pendulum in the correct direction. Also the cart rail has finite length, an additional constraint on the cart motion.

The control objective is to swing up the pendulum to the upright position and stabilize it there by moving the cart back and forth on the rail.

24

The parameters of this system and their values in the test system are given in Table 5.1. The values of these parameters are selected based on the physical setup in [35]. These values are extremely important for the complete swing-up of the pendulum and if chosen incorrectly the inversion of pendulum won't be physically possible.The parameter values can be chosen by the designer of the system and may be different than the ones in Table 5.1, but for the sake of this project the values in Table 5.1 will be used.

5.2 Mathematical Model

After a study of the physical setup of the pendulum on a cart system, detailed mathematical analysis will be performed. The course notes [35] give an excellent step by step derivation of the mathematical modeling of this system and it will also be covered in this section. The application of the force F results in a horizontal movement described by

$$(m + M)\ddot{x} = F c_c \dot{y}$$

Now the cart position relative to the center-of-mass is given by

$$y = x - \frac{Lm}{m + M} \sin\theta$$

Table 5.1: Parameters of the pendulum system

Parameter	Symbol	Value
Mass of Cart	M	0.5 Kg
Mass of pendulum Bob	m	0.275 Kg
Length of Pendulum rod	L	0.5 m
Coefficient of friction between pendulum and pivot	c_p	0.01
Coefficient of friction between cart and track	c_c	0.05

From this, the equation for the cart motion is obtained as

$$(m + M)\ddot{y} = F - c_c\dot{y} - mL\ddot{\theta}\cos\theta + mL\dot{\theta}^2\sin\theta \qquad (5.1)$$

Now, the rotation of pendulum around its pivot is caused by gravity and the acceleration of the pivot point itself. From the tangential forces the following is obtained.

$$mL^2\ddot{\theta} = -c_p\dot{\theta} - mgL\sin\theta - mL\ddot{y}\cos\theta \qquad (5.2)$$

Equations 5.1 and 5.2 describe the model of the cart-and-pendulum motion. It should be noted how these two equations of motion are inter-dependant with each other and this shows the *Coupled* behavior of the system. This is evident from the fact that the angular acceleration $\ddot{\theta}$ of the pendulum is dependant on the horizontal acceleration of the cart $\ddot{y}$ and vice versa.

We now try to decouple these two motions in space and also substitute the values of the parameters of Table 5.1. The following is thus obtained:

$$\ddot{y} = 1.2903F - 0.0645\dot{y} - 0.1774\ddot{\theta} + 0.1774\dot{\theta}\cos\theta \qquad (5.3)$$

$$\ddot{\theta} = -0.1455\dot{\theta} - 19.6\sin\theta - 2\ddot{y}\cos\theta \qquad (5.4)$$

5.3   Simulating the System

We implement the Ordinary Differential Equations in SIMULINK to simulate the behavior of the actual physical system. The purpose of this simulation is to obtain the time-series input-output measurements of the system which will be later used in system identification.

Since this system is non-linear and unstable, we use a stabilizing controller initially to stabilize the system in a defined range and then use a

broad-band square signal, in an effort to excite all possible output levels of both outputs of the system. This approach is explained in detail in [36]. The
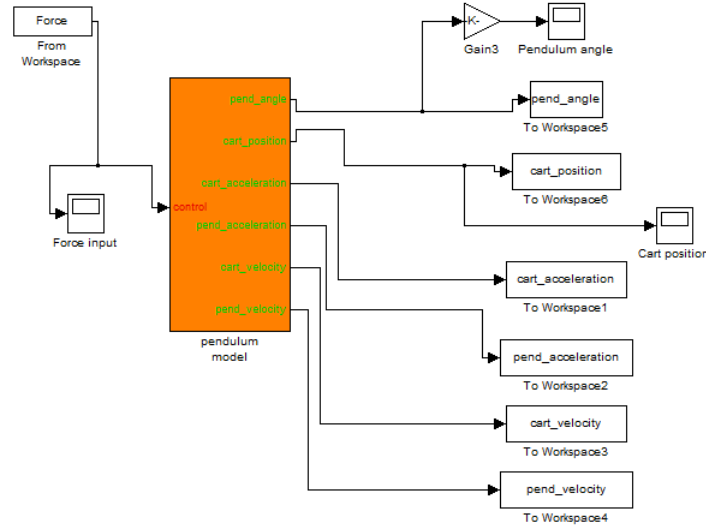

Figure 5.2: SIMULINK Block Diagram model of the inverted pendulum on cart system.

SIMULATION diagram is shown in figures 5.2 and 5.3. Figure 5.2 shows the input and output block diagram of the system. It is seen that the input from our stabilizing controller is fed into the block and the outputs of pendulum angle and cart position are taken out from the block. Also secondary outputs like velocity and acceleration are derived from the primary outputs. Figure 5.3 shows the internal working of the system block. It is basically a representation of the Ordinary Differential Equations derived in equations 5.1 and 5.2. The input (Force F) is shown in red and primary and secondary outputs are shown in green. Data is generated for 10,000 discrete samples taken at 0.1 KHz sampling frequency and the corresponding multi-square input sequence is as shown in Figure 5.4. The corresponding outputs generated is shown in Figure 5.5 and 5.6. Even though these readings are taken over a wide range of values for 10,000 samples, we will be selecting only a few out of these for our system

Figure 5.3: SIMULINK model of the inverted pendulum on cart system describing the ODEs of the system.

Figure 5.4: Force applied to cart sampled at 0.1 KHz.

identification scheme. As is seen from the excitation signal, a series of square waves both in negative and positive amplitudes is given as input to the system. Such an input represents small motions of the cart back and forth around a central point. Initially, the pendulum swings in the right direction, as the cart moves towards the left due to the so called *Inverse characteristics*. Then as it is about to attain its maximum swing point, the cart is moved in the opposite, i.e. the right direction. This helps the pendulum to gain momentum and swing up to a higher point on the right side. The reason for this gain in momentum is
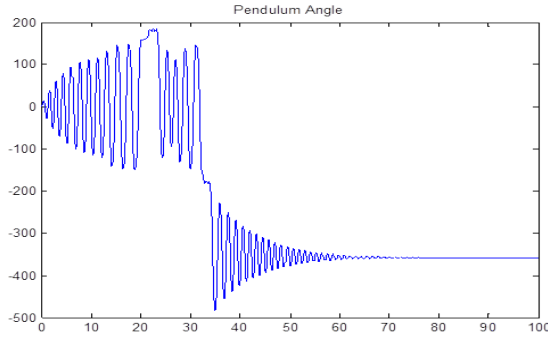
28

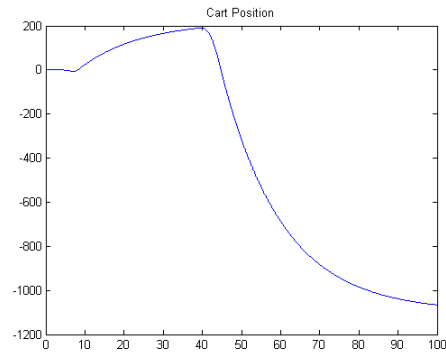Figure 5.5: Angle of pendulum with vertical in degrees.

Figure 5.6: Position of cart.

evident from the fact that as the pendulum is about to come down from its highest swing point, its potential energy is getting converted into kinetic energy and the horizontal component of gravitational force and the force due to cart motion add up. By moving the cart in such a way, we are able to swing the pendulum up completely to the upright position and get enough dynamics in order to perform suitable identification on the system. After it attains the upright position, no input force is given and oscillations are allowed to die down to the pendant position.

5.4   Results

The technique was tested for the data set obtained from simulation, by varying two main criteria: number of samples used for reconstruction and the initial condition. These two results are summarized by Figures 5.7, 5.8 and 5.9. Figure 5.7 give the Mean Squared Error (MSE) between the original and reconstructed ODE plotted versus the *subrate* or the ratio of number of samples (M), to the number of unknowns (N) for the pendulum being initially at the 0^0 degree or the pendant position. Figures 5.8 and 5.9 show the MSE plotted against the subrate for varying initial conditions. Six such random initial conditions are chosen and the performance of the technique is validated for
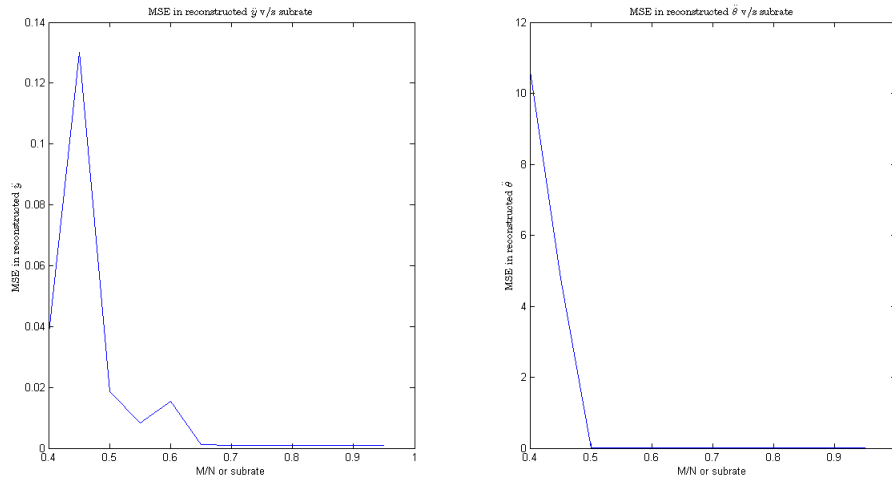
29

Figure 5.7: MSE v/s subrate

each case. Tables 5.2 and 5.3 show the reconstruction results. It should be

noted that the tables show only the values of non-zero reconstructed

coefficients. All other coefficients were found to be zero or extremely small and

ignored. It is seen that for different initial conditions, the technique is still able

to obtain the accurate reconstructed equations, which indicates that the

technique is highly robust in this respect. Also only 45% of the samples are

sufficient for accurate reconstruction, which is a very important advantage over

Neural Network system identification methods, which need large number of
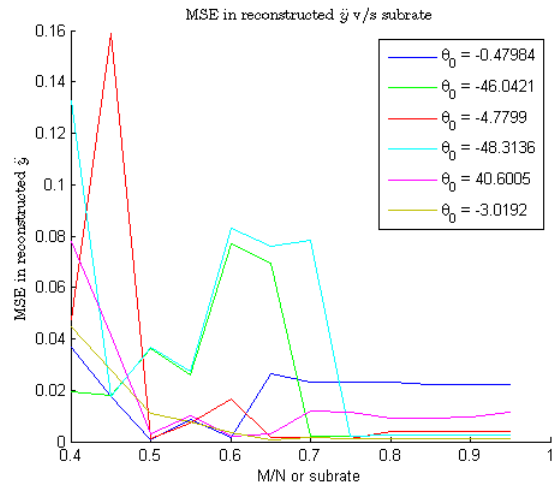
measurements for training.

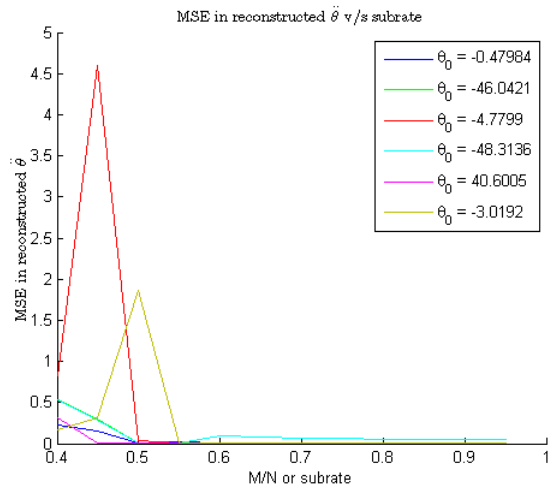Figure 5.8: MSE v/s subrate for varying initial conditions for $\ddot{y}$

Figure 5.9: MSE v/s subrate for varying initial conditions for $\ddot{\theta}$

Table 5.2: Reconstructed coefficients (zero initial conditions)

LHS	Term	Original Coefficients	Reconstructed Coefficients
	F	1.2903	1.113
$\ddot{y}$	$\dot{y}$	-0.0645	-0.0601
	$\ddot{\theta}$	-0.1774	-0.1701
	$\dot{\theta}\cos\theta$	0.1774	0.1381
	$\dot{\theta}$	-0.1455	-0.00973
$\ddot{\theta}$	$\sin\theta$	-19.6	-19.23
	$\ddot{y}\cos\theta$	-2	-1.905

Note. This experiment was performed for the pendulum to be at pendant position and an initial force of -1 units given to the cart. The number of samples needed for accurate reconstruction were 375 and the number of unknown terms or the size of a_y or a_θ was 768. Therefore 49% samples were needed as compared to the unknowns.

Table 5.3: Reconstructed coefficients (non-zero initial conditions)

LHS	Term	Original Coefficients	Reconstructed Coefficients
	F	1.2903	1.2481
$\ddot{y}$	$\dot{y}$	-0.0645	-0.0272
	$\ddot{\theta}$	-0.1774	-0.1739
	$\dot{\theta}\cos\theta$	0.1774	0.1691
	$\dot{\theta}$	-0.1455	-0.00130
$\ddot{\theta}$	$\sin\theta$	-19.6	-19.4021
	$\ddot{y}\cos\theta$	-2	-1.89

Note. This experiment was performed for the pendulum to be at 28° position and an initial force of -1 units given to the cart. The number of samples needed for accurate reconstruction were 350 and the number of terms to reconstruct or the size of a_y or a_θ was 768. Therefore 45% samples were needed as compared to the unknowns.

Chapter 6

CONCLUSION AND FUTURE SCOPE

This chapter summarizes and concludes the thesis and outlines several areas for research to further study and improve the technique developed herein.

6.1   Conclusion

System Identification is an extremely important aspect of control engineer, and inverted pendulum systems are longstanding benchmarks to analyze the efficacy of identification techniques. This thesis introduces a technique for system identification based on Compressed Sensing. Although based on a recent method of Wang et al., this technique provides significant extensions to the applicability of the approach and rectifies two main drawbacks of the method. The technique is applied to the identification of an inverted pendulum system. In order to do this, the pendulum and cart system is simulated and the excitation signal is constructed based on a stabilizing controller. The time-series input-output measurement data for this system is obtained from the simulation. The power series expansion is then formulated to include a library of functions and later compressed sensing approach is used to reconstruct the unknown weights of the power series expansion from the input-output data. This technique is highly robust considering that it can accurately reconstruct the ODE despite the different initial conditions of the system. It also can reconstruct the ODE using very few samples as compared to the unknowns to a great degree of accuracy.

This technique has a broad application potential, its success for the inverted pendulum problem is a promising start.

6.2 Future work

This work hopes to stimulate future research in certain similar areas. There are some issues which are being currently worked on. A few will be provided in this section.

Currently we are focusing on sinusoids of single frequency for our power series expansion, i.e. $\sin\theta$. But for its application to a broader class of systems, we should find a way to incorporate other frequencies as well and include $\sin\omega\theta$ in our basis functions. This will increase the number of terms to be reconstructed and will in turn increase computational complexity.

Also, the columns of the Φ matrix may be highly correlated in some scenarios where two functions approximate each other in a given range of values. Example: θ and $\sin\theta$ are almost equal for small values of θ. In this case it becomes extremely hard for the optimizer to differentiate between the correlated columns and reconstruct the coefficients. A method to de-correlate the columns is essential.

We did some preliminary tests by subjecting the system to random Gaussian white noise and tried to quantify its effects on the accuracy of the reconstruction algorithm. This will serve to be another area of research and finding noise bounds within which the technique works accurately remains to be identified.

Also, in order to reduce the size of the library of non-linear terms generated by the power-series expansion, a pre-filtering method based on physical knowledge of the system can be invoked by restricting the total power of a non-linear term. After invoking such conditions, the number of unknown weights of the non-linear terms will be reduced significantly; consequently reducing the number of samples needed to accurately identify the system.

Furthermore, analytical studies need to be performed to understand the effects of the nature of excitation of the system, on the RIP of the Φ matrix. Some preliminary studies were performed in this respect; where an input (Force) sequence generated from a random Gaussian distribution mapped to an amplitude range of -5 to $+5$ units was used to stimulate the inverted pendulum system and the output measurements in a defined window of time were used to reconstruct the system equations to a good degree of accuracy.

REFERENCES

[1]   N. Wiener, *Nonlinear Problems in Random Theory*, New York Technology Press, M.I.T. and Wiley, 1958.

[2]   V. Volterra, *Theory of Functionals and Integrals and Integro-Differential Equations*, Dover, New York, 1959.

[3]   M. Schetzen, *The Volterra and Wiener Theories of Non-Linear Systems*, Wiley-Interscience, 1980.

[4]   D. Mansour, *Efficient Nonlinear System Identification*, IEEE, 1984.

[5]   E. Candés, *Compressive Sampling*, Int. Congress of Mathematics, 3, pp. 1433-1452, Madrid, Spain, 2006.

[6]   R. Baraniuk, *Compressive Sensing*, IEEE Signal Processing Magazine, 24(4), pp. 118-121, July 2007.

[7]   E. Candés and M. Wakin, *An Introduction to Compressive Sampling*, IEEE Signal Processing Magazine, 25(2), pp. 21 - 30, March 2008.

[8]   M. Fornasier and H. Rauhut, *Handbook of Mathematical Methods in Imaging*, (O. Scherzer Ed.), Springer, 2011.

[9]   M. Davenport, M. Duarte, Y. C. Eldar, and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2011.

[10] E. Candés, J. Romberg, and T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans. Inf. Theory 52, 489 (2006); Commun. Pure Appl. Math. 59, 1207 (2006).

[11] D. Donoho, *For Most Large Underdetermined Systems of Linear Equations the Minimal l_1-norm Solution is also the Sparsest Solution*, Commun. Pure Appl. Math. 59(6), June 2006.

[12] E. J. Candés and T. Tao, *Decoding by Linear Programming*, IEEE Trans. Inform. Theory 51 (2005), 4203-4215.

[13] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, *An Architecture for Compressive Imaging*, Int. Conf. on Image Processing (ICIP), Atlanta, Georgia, October 2006.

[14] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk, *A New Compressive Imaging Camera Architecture Using Optical-Domain Compression*, Computational Imaging IV at SPIE Electronic Imaging, San Jose, California, January 2006.

[15] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, *Compressive Imaging for Video Representation and Coding*, Proc. Picture Coding Symposium (PCS), Beijing, China, April 2006.

[16] J. Xu, J. Ma, D. Zhang, Y. Zhang, and S. Lin, *Compressive Video Sensing Based on User Attention Model*, 28th Picture Coding Symposium, PCS 2010, Dec. 8-10, 2010, Nagoya, Japan.

[17] M. Lustig, D. Donoho, and J. M. Pauly, *Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging*, Magnetic Resonance in Medicine, 58(6) pp. 1182 - 1195, December 2007.

[18] J. Trzasko, A. Manduca, and E. Borisch, *Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic L_0 Minimization*, IEEE Trans. Medical Imaging 28(1): 106-121, 2009.

[19] C. Qiu, W. Lu and N. Vaswani, *Real-time Dynamic MR Image Reconstruction using Kalman Filtered Compressed Sensing*, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Taipei, Taiwan, April 2009.

[20] B. Zhao, J. P. Haldar, C. Brinegar, and Z. P. Liang, *Low Rank Matrix Recovery for Real-Time Cardiac MRI*, IEEE Int. Symp. Biomed. Imag., pp. 996-999, April 2010.

[21] M. Sheikh, O. Milenkovic, and R. Baraniuk, *Compressed Sensing DNA Microarrays*, Rice ECE Department Technical Report TREE 0706, May 2007.

[22] M. Mohtashemi, H. Smith, F. Sutton, D. Walburger and J. Diggans, *Sparse Sensing DNA Microarray-Based Biosensor: Is It Feasible?*, 2010 IEEE Sensors and Applications.

[23] M. Sheikh, O. Milenkovic, and R. Baraniuk, *Designing Compressive Sensing DNA Microarrays*, IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), St. Thomas, U.S. Virgin Islands, December 2007.

[24] A. Griffin and P. Tsakalides, *Compressed Sensing of Audio Signals in a Wireless Sensor Network*, in Proc. 5th European Workshop on Wireless Sensor Networks (EWSN '08), Bologna, Italy, January 30 - February 1, 2008.

[25] Y. N. Lilis, D. Angelosante and G. B. Giannakis, *Sound Field Reproduction using the Lasso*, IEEE Trans. on Audio, Speech and Language Processing, 2010.

[26] A. Griffin, T. Hirvonen, C. Tzagkarakis, A. Mouchtaris, and P. Tsakalides, *Single-channel and Multi-channel Sinusoidal Audio Coding Using Compressed Sensing*, IEEE Trans. on Audio, Speech, and Language Processing.

[27] A. Fannjiang, P. Yan, and T. Strohmer. *Compressed Remote Sensing of Sparse Objects*, preprint, 2009.

[28] W. X. Wang, R. Yang, Y. C. Lai, V. Kovanis, and C. Grebogi, *Predicting Catastrophes in Nonlinear Dynamical Systems by Compressive Sensing*, Phys, Rev. Lett., 2011.

[29] L. Ljung, *System Identification-Theory for the User*, Prentice Hall, 1987.

[30] K. A. Nadsay, *A Two Stage Method for System Identification from the Time Series*, Ohio University, March 1998.

[31] L. Ljung, *Linear System Identification as Curve Fitting*, Spinger Lecture Notes on Control and Information, Vol 286: 203-215 2003, December 2003.

[32] K. S. Narendra and K. Parthasarthy, *Identification and Control of Dynamical Systems Using Neural Networks*, IEEE Transactions on Neural Networks, Vol 1. No. 1. March 1990.

[33] N. B. Haaser and J. A. Sullivan, *Real Analysis*, New York: Van Nostrand Reinhold, 1971.

[34] S. Brock, *Identification of the Parameters in Inverted Pendulum Model*, 7th International Workshop on Advanced Motion Control, 2002.

[35] K. Tsakalis, *Examples of System Models*, Class Notes, `http://tsakalis.faculty.asu.edu/notes/models.pdf.`

[36] S. Boonto and H. Werner, *Closed-Loop System Identification of LPV Input-Ouput Models- Application to an Arm-Driven Inverted Pendulum*, IEEE Conf. on Decision and Control, Caucun, Mexico, 2008.

[37] L. Gan, C. Ling, T. T. Do and T. D. Tran *Analysis of the Statistical Restricted Isometry Property for Deterministic Sensing Matrices Using Stein's Method*.

[38] E. R. M. Verschurueren, *Identification and Control of an inverted pendulum Using Neural Networks*, Eindhoven, December 1995.