Institute for MECHANICS
In cooperation with SIEMENS

TU Graz

# Inverse force estimation for a railway vehicle system

A thesis by

**Thomas David**

submitted to

Institute for Mechanics
Faculty of Mechanical Engineering and Economic Sciences
**Graz University of Technology**

in partial fulfillment of the
requirements for the degree of

**Master of Science**
in Mechanical Engineering and Business Economics

*Supervisor:* ***Univ.-Prof. Dr.-Ing. habil. Katrin Ellermann***
*First Advisor:* ***Dipl.-Ing. Mathias Jesussek***
*Second Advisor:* ***Dipl.-Ing. Dr. Gerald Grabner***

Graz, January 2014

## Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, January 29, 2014 .....

(signature)

# Abstract

In a railway vehicle system information about the applied forces in the wheel-rail contact area is important as these forces have significant influences on running safety, track loading and ride characteristics. In this work a new method for the inverse estimation of lateral and vertical wheel-rail contact forces is presented.

Different methods and approaches of inverse input estimation are chosen and applied to a simple model of a two-mass system. After comparison and discussion of the different results a method based on a linear Kalman filter is selected to identify wheel-rail contact forces for a railway vehicle system.

The multibody simulation of a railway vehicle with all its complexity represents the starting point of the main task of this work and is performed by the software package SIMPACK. With generated measurement data an estimation of unknown input forces can be achieved by extension of the system with a form filter.

Comparisons of the obtained results with a SIMPACK simulation in time domain and frequency domain show high correlation of the estimated and simulated forces, especially in vertical direction.

# Kurzfassung

Die genaue Messung von Kräften im Rad-Schiene-Kontakt, welche einen signifikanten Einfluss auf die Fahrstabilität, die Sicherheit sowie den Fahrkomfort von Schienenfahrzeugen haben, ist mit einem großen Zeit- und Kostenaufwand verbunden. Diese Arbeit liefert eine Berechnungsmethode zur Abschätzung der auftretenden Kräfte in lateraler und vertikaler Richtung anhand einfach zu bestimmender Beschleunigungsdaten. Da man in der Mehrkörpersimulation für gewöhnlich die Anregungskräfte kennt, stellt dies ein inverses Problem dar, das in berechnungstechnischer Hinsicht zusätzliche Herausforderungen mit sich bringt.

Mithilfe der Simulationssoftware SIMPACK werden im Zuge einer Mehrkörpersimulation Beschleunigungsdaten erstellt. In MATLAB wird nachfolgend die Berechnung der auftretenden Rad-Schiene-Kräfte durch Erweiterung des Systems mit einem Formfilter und anschließender Zustandsschätzung eines Kalman-Filters durchgeführt.

Die erhaltenen Ergebnisse werden schließlich mit den generierten Messdaten verglichen und weisen gute Übereinstimmung auf. Vor allem in vertikaler Richtung ist die Abweichung der invers ermittelten Kräfte von den simulierten Kräften gering.

# Acknowledgments

First of all, I would like to thank professor Katrin Ellermann, head of the *Institute for Mechanics* at *Graz University of Technology*, for supervising this thesis. I also want to thank *Siemens AG Österreich* for the cooperation with the Institute for Mechanics for providing the opportunity to write this thesis. Especially Gerald Grabner at the bogie competence center in Graz and Mathias Jesussek at the Instiute for Mechanics have to be named at this place since they worked out the discussed problem.

I want to show my enormous gratitude to my first advisor Mathias Jesussek who supported me at any time during the development process of this thesis. He always tried to find a solution for any kind of problem and his effort in time was significant. Moreover, his knowledge about mechanics, system theory and programming helped me a lot during my work.

In addition a big thank goes to Gerald Grabner for the technical input in terms of railway vehicle dynamics. Due to him I learned a lot in this field. Furthermore he enabled me to work at the bogie competence center in Graz where I gained an insight into his job of a development engineer.

Finally, I would like to give thanks to my family for supporting me during my years of study and especially during the time where I was writing this thesis. Without the financial opportunities and the mental backing I would not have been able to finish my studies at this point.

# Contents


**1 Introduction** **8**

**2 Linear systems and signals** **10**

2.1 Linear systems theory . . . . . 10
2.1.1 Causality, linearity and time invariance . . . . . 10
2.1.2 State-space representation . . . . . 11
2.1.3 Discretization . . . . . 12
2.1.4 Stability . . . . . 14
2.1.5 Controllability and Observability . . . . . 16
2.1.6 Invertibility . . . . . 16
2.2 Probability theory . . . . . 17
2.2.1 Probability and random variables . . . . . 17
2.2.2 Stochastic processes . . . . . 21
2.3 Time domain and frequency domain analysis . . . . . 23
2.4 Filtering . . . . . 24

**3 Inverse input estimation** **25**

3.1 Characteristics of inverse problems . . . . . 25
3.2 Matrix inversion solutions . . . . . 26
3.2.1 Pseudoinverse of matrices . . . . . 26
3.2.2 Least squares solution . . . . . 27
3.2.3 Singular value decomposition . . . . . 28
3.3 One-dimensional two-mass system . . . . . 29
3.3.1 Mechanical model . . . . . 29
3.3.2 Input estimation problem . . . . . 30
3.4 Kalman filter . . . . . 31
3.4.1 The Kalman filter implementation . . . . . 32
3.4.2 Kalman filter application to the two-mass system . . . . . 33
3.5 Parity space approach . . . . . 34
3.5.1 The parity space approach algorithm . . . . . 35
3.5.2 Parity space approach applied to the two-mass system . . . . . 35
3.6 Regularized least squares solution . . . . . 36
3.6.1 Determination of the RLSS . . . . . 36
3.6.2 RLSS of the two-mass system . . . . . 37
3.7 Inverse dynamics method . . . . . 38
3.7.1 The inverse dynamics method algorithm . . . . . 38
3.7.2 Inverse dynamics method for the two-mass system . . . . . 40

3.8 Results for the two-mass system . . . 40
3.8.1 Generation of a random input signal . . . 41
3.8.2 Model and method parameters . . . 41
3.8.3 Discussion of results . . . 42

**4 Multibody simulation of a railway vehicle system 46**
4.1 Modeling of a railway vehicle system . . . 46
4.1.1 Mechanical Model . . . 46
4.1.2 Equations of motion . . . 48
4.1.3 Linearization and state-space formulation . . . 50
4.1.4 Modeling of the wheel-rail contact . . . 51
4.2 Input by excitation . . . 55
4.2.1 Track irregularities . . . 55
4.2.2 Attainment of the input-force-matrix . . . 57
4.2.3 Wheel-rail contact force calculation . . . 61

**5 Input force estimation for a railway vehicle system 63**
5.1 Inverse problem formulation . . . 63
5.1.1 Problem statement . . . 64
5.1.2 Estimation procedure . . . 64
5.2 Measurement data from simulation . . . 65
5.3 System extension with a form filter . . . 66
5.4 Kalman filter with unknown input . . . 69
5.5 Force calculation . . . 70
5.6 Comparison and validation of results . . . 71

**6 Conclusion 78**

**A Simple linear regression 80**

**B Model parameters 81**

**List of Figures 83**

**List of Tables 85**

**Bibliography 86**

# 1 Introduction

In railway vehicle dynamics, the compound wheel-rail system has to fulfill several fundamental functions, such as carrying, guiding and power transmission. Thus, forces occurring in the wheel-rail contact areas have main influences on running safety, track loading and ride characteristics of the vehicle. Furthermore, the durability of the vehicle depends on wheel-rail contact forces. Therefore the knowledge of these forces is essential.

Currently, wheel-rail contact forces are determined based on data from instrumented wheelsets or track-side measurement devices. The contact forces are computed from the measured strains as e.g. can be seen in [25]. Wheelsets containing measurement equipment are custom-made and track-side measurement points have to be installed at many locations on the track in order to obtain useful results. However, these methods are expensive. Moreover, it is practically impossible to install instrumented wheelsets during the long lasting regular operation of a train. Therefore, a great demand of alternative approaches exists. At present various methods based on the inverse problem of using acceleration data in order to determine unknown wheel-rail contact forces are proposed. In this work, an approach with use of a linear Kalman filter for state estimation based on given acceleration data is introduced.

At the beginning of this thesis the basic theories of system dynamics and signal processing are reviewed. The linear systems theory is described in detail, since the method will be applied on a linearized model of a railway vehicle system. In addition the characteristics of signals and some analyzing methods are presented.

This is followed by a description of inverse problems where four methods for the inverse identification of unknown input are presented. For better understanding the here regarded approaches are applied to a simple model of a one-dimensional two-mass system. It is shown that the unknown excitation can be estimated with knowledge about the system characteristics and acceleration data through application of the presented methods. The method with best estimation results is then selected for the identification of unknown input forces of a railway vehicle system.

The mechanical concepts of multibody simulation with application to railway vehicles are introduced before the inverse problem solving task for the railway vehicle system can be discussed. Random track irregularities have main influences on the excitation of a railway vehicle system and lead to unknown input forces. Therefore a brief introduction about the characteristics of these irregularities is given.

A main problem in terms of modeling of a railway vehicle is given by the wheel-rail contact, since the input forces are transmitted through the contact patches. Thus, several approaches for the mathematical description of the wheel-rail contact geometry are presented leading to modeling approaches for the contact forces.

After the complex system of a railway vehicle system is characterized and the mechanical model can be described, the main part of this work is given. First, the inverse force iden-

tification problem of a railway vehicle system based on a SIMPACK simulation [30] has to be formulated properly. In this case the identification of unknown wheel-rail contact forces on basis of acceleration measurements with assumption that the system characteristics are known is the task to be accomplished. Therefore the SIMPACK model is linearized before the system characteristics are used in order to be implemented in the estimation process. The system also has to be extended by a special filter element to obtain more accurate estimation results. On the basis of spectral characteristics of assumed random track irregularities a description of the filter can be given. After that, estimates of the wheel-rail contact forces with application of a Kalman filter are calculated from acceleration data. With the presented identification method forces in lateral and vertical direction acting on the single wheels of the vehicle can be estimated.

Finally the obtained results are compared and discussed with use of different analyses and measures. The identification method is developed in order to analyze measured accelerations of a train. When testing the technique, not only acceleration data is needed but also contact force data for comparison with the estimated results. In the present work, a SIMPACK model is used to generate this data. Since the results are promising, the application of the presented method to a multibody simulation model of a railway vehicle system can be regarded as a suitable approach to solve the given inverse problem.

# 2 Linear systems and signals

The multi-body simulation of a railway vehicle system requires some basic knowledge about the theory of systems and signals. This chapter gives a brief introduction in the characterization of systems, followed by a review of the probability theory which is essential for the work with signals. A further distinction in terms of signal description in time and frequency domain is made and finally some basics and methods of filtering are presented.

## 2.1 Linear systems theory

A mathematical description of a process, or a physical system, is needed in order to control the process or to obtain informations about it. The definitions and mathematical models of systems introduced here can be found in [18], [26] and [29]. Figure 2.1 illustrates the definition of a system in a block diagram: a dynamical system transforms an input signal $\underline{u}(t)$ into the output signal $\underline{y}(t)$.

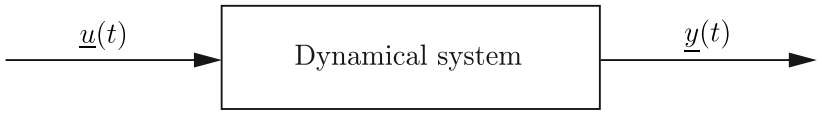



Figure 2.1: Block diagram of a dynamical system [18]

The modeling task begins with a characterization of the system or the process. Therefore some basic characteristics are presented next.

### 2.1.1 Causality, linearity and time invariance

The term of *causality* has to be introduced while describing systems. A system is causal if the output at time $t$ is a response because of an input at time $\bar{t} \leq t$. Causality has to be fulfilled for correct modeling and specification of systems at any time $t$. The definition for the causality with two inputs $\underline{u_1}(t)$ and $\underline{u_2}(t)$, that are equal until $t = T$, is given by:

$$\underline{u_1}(t) = \underline{u_2}(t) \quad \mapsto \quad \underline{y_1}(t) = \underline{y_2}(t) \quad \text{for } 0 \leq t \leq T \tag{2.1}$$

where the arrow "$\mapsto$" denotes that the system transforms the input $\underline{u}(t)$ into the output $\underline{y}(t)$. This means that the outputs $\underline{y_1}$ and $\underline{y_2}$ only differ if $t > T$ and the system can be regarded as causal for $0 \leq t \leq T$. [18]

If the output is proportional to the input the behavior of the system is linear. *Linearity*

is defined with two principles: additivity[1] (or superposition principle) and homogeneity[2]. That means a system is linear if:

$$\underline{u_1}(t) \quad \mapsto \quad \underline{y_1}(t) \quad \text{and} \quad \underline{u_2}(t) \quad \mapsto \quad \underline{y_2}(t) \tag{2.2}$$

$$\underline{u}(t) = a\underline{u_1}(t) + b\underline{u_2}(t) \quad \mapsto \quad \underline{y}(t) = a\underline{y_1}(t) + b\underline{y_2}(t) \ . \tag{2.3}$$

However, most real processes are nonlinear. Anyhow, since the mathematical tools for description, estimation and control are more accessible for linear systems, nonlinear systems are approximated as linear commonly.

*Time invariance* is another important property of a system and can be assumed if the response of the system to a given input is not changing by variation of time. That implies that the only change in the response is a time shift, which has to be equal to the input time shift $\tau$.

$$\underline{u_1}(t) = \underline{u_2}(t-\tau) \quad \mapsto \quad \underline{y_1}(t) = \underline{y_2}(t-\tau) \quad \text{for } \tau > 0 \tag{2.4}$$

Many time-variant systems can be regarded as time-invariant if a deviation of parameters can be neglected or is to slow to influence the system in the regarded period.

### 2.1.2 State-space representation

The transformation of the input $\underline{u}(t)$ into the output $\underline{y}(t)$ can be described by differential equations. Regarding a linear and time-invariant (LTI) system, shown in Figure 2.2, the corresponding mathematical model is a linear differential equation of $n$-th order containing the physical characteristics of the system.

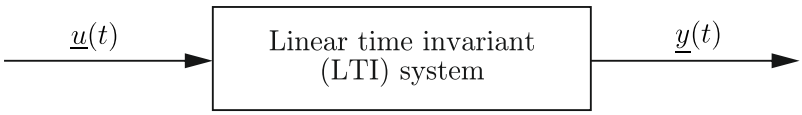



Figure 2.2: LTI system

An equivalent description is the so-called state-space representation of a system. This model consists of $n$ differential equations of first-order and is very popular for control applications. Therefore the differential equation of $n$-th-order has to be transformed by introducing the state vector $\underline{x}$, which is $n$-dimensional in general.

A continuous, linear and time-invariant system can be represented by a *state equation*, given by equation (2.5), and an *output equation* , given by equation (2.6), as follows:

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \tag{2.5}$$

$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \tag{2.6}$$

where $\underline{x}$ is the state vector with its time derivative $\dot{\underline{x}}$. The $m$ inputs of the system are collected in the input vector $\underline{u}$ as well as the output vector $\underline{y}$ contains of all $l$ outputs.

[1] Additivity: $f(a+b) = f(a) + f(b)$
[2] Homogeneity: $\lambda g(a) = g(\lambda a)$, for all $\lambda$

The $(n \times n)$ - matrix $\mathbf{A}$ is called the *system matrix* and accordingly $\mathbf{B}$ is the *input matrix* with dimension $(n \times m)$ and $\mathbf{C}$ is the *output matrix* with dimension $(l \times n)$. The so-called *feedthrough matrix* $\mathbf{D}$ has a dimension of $(l \times m)$ and often is zero if there is no direct feedthrough in the system. The system matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$ are constant in the time-invariant case.

Figure 2.3 shows a block diagram of the state-space model from equations (2.5) and (2.6). The diagram consists of an integration block where $\underline{x_0}$ represents the initial state $\underline{x_0} = \underline{x}(t = 0)$. It follows that for a given initial state a solution for the output $\underline{y}$ can be derived for every time $t \geq 0$ assuming that the present and future input $\underline{u}$ is given.

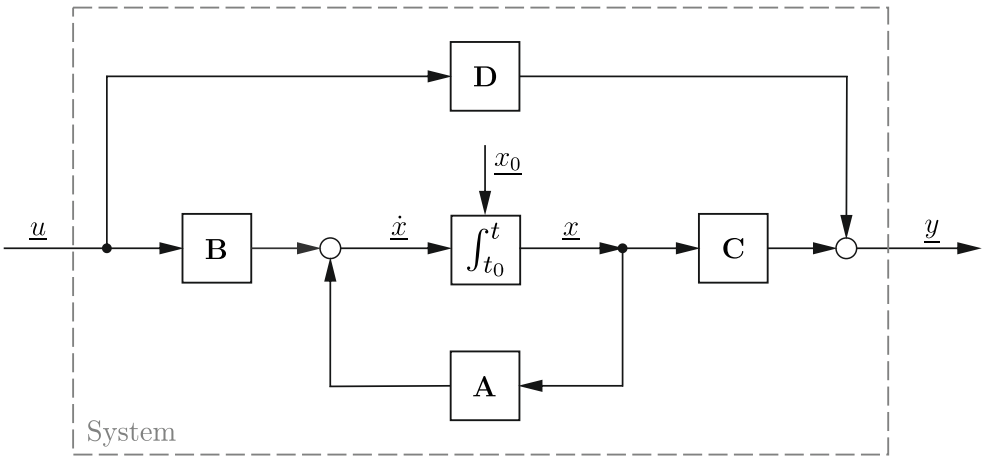



Figure 2.3: Block diagram of the state-space model [18]

### 2.1.3 Discretization

Since state estimation and control algorithms are usually implemented in digital electronics a discretization of the system has to be performed. Instead of a continuous signal, which is uninterrupted, a discrete signal is a time series consisting of a sequence of values. A discrete form $x_k$ of a continuous signal $x(t)$, as a sine function i.e., can be obtained by sampling with a sampling step size $\Delta t$.

$$x(t) = sin(\omega t + \varphi) \tag{2.7}$$

$$x_k = sin(\omega t_k + \varphi) = sin(\omega k \Delta t + \varphi) \tag{2.8}$$

The continuous sine function of equation (2.7), with the frequency $\omega$ and the phase $\varphi$ , is a function of $t$ and exists for all values of the independent variable $t$, whereas the associated discrete function, given by equation (2.8), only exists for the discrete time $t_k = k\Delta t$ with $k \in \{0, 1, 2, \dots\}$.

The discretization of signals can implicate unwanted effects as time or phase shifts, that have to be avoided or taken into account. Another effect is the effect of periodicity, which

implies a loss of the original period $T_0$ of a signal, if a periodic signal[3] is sampled with a noninteger sampling interval $T_0/\Delta t$.

If the sample rate is chosen to low the effect of *aliasing* can occur. Figure 2.4 shows a signal sampled with different intervals $\Delta t$ and illustrates the aliasing effect as a consequence of undersampling. As a result the period $T_1$ is higher than the original period $T_0$ according to the aliased frequency in this example.

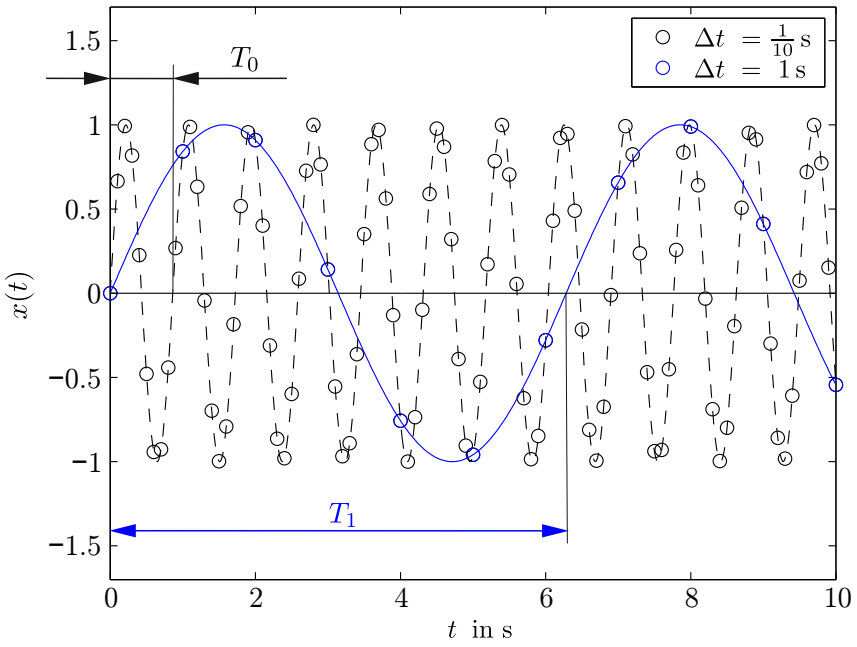



Figure 2.4: Aliasing effect

To avoid wrong reconstructions of a signal because of discrete values the sampling frequency $f_s$ should be at least more than twice the highest frequency $f_{max}$ of a signal. Therefore a sampling theorem including the so-called *Nyquist frequency* $f_N$ has been introduced:

$$f_s = \frac{1}{\Delta t} > f_N = \frac{2}{T_0} = 2f_{max} \ . \tag{2.9}$$

That means the Nyquist frequency $f_N$ is the highest frequency for which a reconstruction of the original signal is possible. In technical applications analog anti-aliasing filters are often implemented before digitization to remove frequency components that are higher than twice the sampling frequency.

Considering the discretization of systems not only single signals have to be sampled also the state-space model of the system has to be discretized. The discrete form of the linear state-space model from equation (2.5) and (2.6) is the following:

$$\underline{x_{k+1}} = \mathbf{A_d}\underline{x_k} + \mathbf{B_d}\underline{u_k} \tag{2.10}$$

$$\underline{y_k} = \mathbf{C_d}\underline{x_k} + \mathbf{D_d}\underline{u_k} \ . \tag{2.11}$$

[3]Definition from [26]: *"A periodic signal is a repetitive sequence of values with a well-defined timescale or period $T = p \cdot \Delta t$, so that $\underline{x_k} = \underline{x_{k+p}}$."*

The matrix index **d** denotes that the discrete-time system matrices are not equal to the corresponding matrices of the continuous-time system. This requires a transformation with the discretization step size $\Delta t$, which is derived from the solution of the continuous state equation (2.5). If $A$ is invertible the discretization can be derived according to [29] by the following:

$$\mathbf{A_d} = e^{\mathbf{A}\Delta t}, \qquad \mathbf{B_d} = \mathbf{A_d}[\mathbf{I} - e^{-\mathbf{A}\Delta t}]\mathbf{A^{-1}B} \ . \tag{2.12}$$

The dicretized output matrices $\mathbf{C_d}$ and $\mathbf{D_d}$ are equal to their continuous forms $\mathbf{C}$ and $\mathbf{D}$.

### 2.1.4 Stability

A dynamical system is also characterized by its *stability*. In general, system stability implies the transformation of a bounded input into a bounded output signal (BIBO stability).

If only the internal condition of a system is considered, the two types marginal stability (or Lyapunov stability) and asymptotic stability can be given as they are defined in [29]:

"***Definition 1:*** *A linear continuous-time, time-invariant system is marginally stable if the state* $\underline{x}(t)$ *is bounded for all* $t$ *and for all bounded initial states* $\underline{x_0} = \underline{x}(t = 0)$."

"***Definition 2:*** *A linear continuous-time, time-invariant system is asymptotically stable if, for all bounded initial states* $\underline{x_0}$*,*

$$\lim_{t\to\infty} \underline{x}(t) = 0 \ ." \tag{2.13}$$

It can been seen from Definition 1 and Definition 2 that a system is marginally stable if it is asymptotically stable. The two definitions are types of the so-called internal stability since they only deal with the state of a system without considering any output.

A linear, continuous-time system without any input is described by the following state-space model:

$$\dot{\underline{x}} = \mathbf{A}\underline{x} \tag{2.14}$$

$$\underline{y} = \mathbf{C}\underline{x} \ . \tag{2.15}$$

The solution of equation (2.14) is given by:

$$\underline{x}(t) = e^{\mathbf{A}t}\underline{x_0} \tag{2.16}$$

and is leading to two stability theorems from [29]:

"***Theorem 1:*** *A linear continuous-time, time-invariant system is marginally stable if and only if*

$$\lim_{t\to\infty} e^{\mathbf{A}t} \leq \mathbf{M} < \infty \tag{2.17}$$

*for some constant matrix* $\mathbf{M}$*. This is just a way of saying that the matrix exponential does not increase without bound."* [4]

[4] The left term in equation (2.17) is also a matrix what means that the *less than or equal to* - relation can be interpreted either as ($\mathbf{M} - \lim_{t\to\infty} e^{\mathbf{A}t}$) has to be a positive semidefinite matrix or that every element

"***Theorem 2:*** *A linear continuous-time, time-invariant system is asymptotically stable if and only if*

$$\lim_{t\to\infty} e^{\mathbf{A}t} = \mathbf{0} \text{ ."} \tag{2.18}$$

Introducing the Jordan form[5] $\hat{\mathbf{A}}$ of the matrix $\mathbf{A}$ two other theorems can be found in [29] relating the boundedness of $e^{\mathbf{A}t}$ with the eigenvalues of $\mathbf{A}$:

"***Theorem 3:*** *A linear continuous-time, time-invariant system is marginally stable if and only if one of the following conditions holds.*

1. *All of the eigenvalues of* $\mathbf{A}$ *have negative real parts.*
2. *All of the eigenvalues of* $\mathbf{A}$ *have negative or zero real parts, and those with real parts equal to zero have a geometric multiplicity equal to their algebraic multiplicity. That is, the Jordan blocks that are associated with the eigenvalues that have real parts equal to zero are first order."*

"***Theorem 4:*** *A linear continuous-time, time-invariant system is asymptotically stable if and only if all of the eigenvalues of* $\mathbf{A}$ *have negative real parts."*

In same manner the internal stability for a discrete system can be defined. Consider a linear, discrete-time system without any input. The state-space model is given by the following:

$$\underline{x_{k+1}} = \mathbf{A_d}\underline{x_k} \tag{2.19}$$

$$\underline{y_k} = \mathbf{C_d}\underline{x_k} \text{ .} \tag{2.20}$$

The stability definitions for discrete-time system found in [29] are analogous to the continuous case:

"***Definition 3:*** *A linear discrete-time, time-invariant system is marginally stable if the state* $\underline{x_k}$ *is bounded for all* $k$ *and for all bounded initial states* $\underline{x_0}$*."*

"***Definition 4:*** *A linear discrete-time, time-invariant system is asymptotically stable if, for all bounded initial states* $\underline{x_0}$*,*

$$\lim_{k\to\infty} \underline{x_k} = 0 \text{ ."} \tag{2.21}$$

Analogously to the continuous derivation from above the solution of the discrete-time state equation (2.19) can be found:

$$\underline{x_k} = {\mathbf{A_d}}^k \underline{x_0} \text{ .} \tag{2.22}$$

Now, the following theorems about marginal and asymptotic stability from [29] can be stated for discrete-time systems:

---

of $\lim_{t\to\infty} e^{\mathbf{A}t}$ is less than or equal to the corresponding elements of $\mathbf{M}$. The given theorem holds for both interpretations.

[5] The Jordan form $\hat{\mathbf{A}}$ can be defined with $e^{\mathbf{A}t} = \mathbf{Q}e^{\hat{\mathbf{A}}t}\mathbf{Q}^{-1}$, where the columns of $\mathbf{Q}$ contains the eigenvectors of $\mathbf{A}$ to the corresponding eigenvalues $\lambda_i$. The Jordan matrix consists of diagonal entries called Jordan blocks. The number of Jordan blocks depends on the geometric multiplicity of the relating eigenvalue.

"***Theorem 5:*** *A linear discrete-time, time-invariant system is marginally stable if and only if*

$$\lim_{k\to\infty} \mathbf{A_d}^k \leq \mathbf{M} < \infty \tag{2.23}$$

*for some constant matrix* $\mathbf{M}$*. This is just a way of saying that the matrix exponential does not increase without bound."*

"***Theorem 6:*** *A linear discrete-time, time-invariant system is asymptotically stable if and only if*

$$\lim_{k\to\infty} \mathbf{A_d}^k = \mathbf{0} \text{ ."} \tag{2.24}$$

"***Theorem 7:*** *A linear discrete-time, time-invariant system is marginally stable if and only if one of the following conditions holds.*

1. *All of the eigenvalues of* $\mathbf{A_d}$ *have magnitude less than or equal to one.*
2. *All of the eigenvalues of* $\mathbf{A_d}$ *have magnitude less than one or equal to one, and those with magnitude equal to one have a geometric multiplicity equal to their algebraic multiplicity. That is, the Jordan blocks that are associated with the eigenvalues that have magnitude equal to one are first order."*

"***Theorem 8:*** *A linear discrete-time, time-invariant system is asymptotically stable if and only if all of the eigenvalues of* $\mathbf{A_d}$ *have magnitude less than one."*

The above definitions and theorems only consider zero-input systems. According to [18] the stability of systems with inputs is defined as the following:

*A linear system is BIBO stable if for the initial condition* $\underline{x_0} = \underline{0}$ *and for any bounded input* $|u(t)| < u_{max}$ *for all* $t > 0$ *the output stays bounded* ($|y(t)| < y_{max}$) *for all* $t > 0$.

Moreover, a system is also BIBO stable if it is asymptotic stable. Finally, LTI systems can be characterized in terms of stability as described above, which is elementary for further investigations.

### 2.1.5 Controllability and Observability

For the design process of control algorithms the concepts of *controllability* and *observability* are essential. If a signal has to be controlled it is necessary to know how well the system and its states are controllable, that means how well a system can be driven at a desired state. The possibility of observation of the initial conditions or states after measurements can be described by the observability instead.

The characteristics of controllability and observability are closely related to the *stabilizability* and the *detectability* of a system. These two concepts are slightly weaker in their definitions, what implies that the states of a system can be stabilizable even if they are not controllable and detectable if they are not observable respectively.

### 2.1.6 Invertibility

Another characteristic of a system is its *invertibility*, which is given if there exists any transformation yielding the input from the output. The inverse problem is a task where

the output is known and an unknown input has to be detected. More about the theory of inverse identification is presented later in this work since an inverse estimation method is applied to a railway vehicle system.

## 2.2 Probability theory

Since some identification methods described within this work deal with random or stochastic processes the fundamentals of probability theory are introduced in this section. First, basic concepts of probability and random variables are discussed. This is followed by a review of the concept of stochastic processes. The given definitions can be found in [19] and [29].

### 2.2.1 Probability and random variables

The concepts of *probability* base on random experiments with possible outcomes or *events*. The probability $P(A)$ describes the number of times event $A$ occurs compared to the total number of outcomes $\Omega$. That means that $P(A)$ is greater than or equal to zero for any event $A \in \Omega$ and $P(\Omega)$ is equal to one. Also the condition $P(A \cup B) = P(A) + P(B)$ is satisfied for any events $A, B \in \Omega$ if $A \cap B = 0$. [6]

Considering an event $A$ occurs given the fact that an event $B$ occurred is leading to the concept of *conditional probability*. The mathematical definition is:

$$P(A|B) = \frac{P(A,B)}{P(B)} \tag{2.25}$$

assuming that $P(B) > 0$. $P(A,B)$ is the so-called joint probability of $A$ and $B$. From equation (2.25) it follows that:

$$P(A,B) = P(A|B)P(B) \quad \text{or} \quad P(B,A) = P(B|A)P(A) \ . \tag{2.26}$$

Equating the two expressions from above is leading to the well-known *Bayes' Rule*:

$$\begin{aligned} P(A|B)P(B) &= P(B|A)P(A) \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \ . \end{aligned} \tag{2.27}$$

Two events $A$ and $B$ can either be dependent or independent if the occurrence of one effects on the probability of the occurrence of the other. The mathematical definition of *independence* in concepts of probability is:

$$P(A,B) = P(A)P(B) \ . \tag{2.28}$$

With equation (2.25) it can be seen that $P(A|B) = P(B)$ if $P(A) \neq 0$ and vice versa $P(B|A) = P(A)$ if $P(B) \neq 0$ presuming independence of the events $A$ and $B$.

A quantity whose exact value is uncertain but some statistical information is available is

[6] $A \cap B$ means "both $A$ and $B$" and $A \cup B$ stands for "$A$ or $B$ or both"

called a *random variable (RV)*. RVs can either be continuous or discrete depending on the set of values they can take. The outcome of an experiment never can be a RV since a RV only is described by its probabilities and can not be equal to a specific value.

The most general way of describing the probabilities of a RV is its *probability distribution function*. With a nonrandom independent variable $x$ the probability distribution function $F_X(x)$ of the RV $X$ is defined as:

$$F_X(x) = P(X \leq x) \ . \tag{2.29}$$

The function from equation (2.29) describes the probability of the event that the RV $X$ takes a value less than or equal to $x$ and has the following properties:

$$\begin{aligned} F_X(x) &\in [0,1] \qquad \text{for } -\infty < x < \infty \\ F_X(-\infty) &= 0 \\ F_X(\infty) &= 1 \\ F_X(x_1) &\leq F_X(x_2) \qquad \text{if} \quad x_1 \leq \times_2 \ . \end{aligned} \tag{2.30}$$

The derivative of $F_X(x)$ is called *probability density function* and is more appropriate for many observations describing the probability distribution of a RV $X$.

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{2.31}$$

Some properties of the probability density function $f_X(x)$ are given by equation (2.32):

$$\begin{aligned} f_X(x) &\geq 0 \quad \text{for } -\infty < x < \infty \\ \int_{-\infty}^{\infty} f_X(x)dx &= 1 \ . \end{aligned} \tag{2.32}$$

With both functions $F_X(x)$ and $f_X(x)$ the probability of an RV $X$ can be expressed with the following relation:

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx \ . \tag{2.33}$$

Usually random processes involve more than one RV. Therefore the joint probability distribution function $F_{XY}(x,y)$ and the joint probability density function $f_{XY}(x,y)$ are defined for two RVs $X$ and $Y$.

$$F_{XY}(x,y) = P(X \leq x, Y \leq y) \tag{2.34}$$

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} \tag{2.35}$$

The average of the expected value of a RV over a large number of experiments is called the *expectation value, mean or average of the RV X*. Its definition for a discrete RV $X$ is given by the following:

$$\bar{X} = E[X] = \frac{1}{N} \sum_{i=1}^{m} A_i n_i \tag{2.36}$$

where $A_i$ is the outcome that occurs $n_i$ times of an experiment that have been run $N$ times with $m$ different outcomes. For a continuous RV $X$ the expectation $E[X]$ is defined as:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx \ . \tag{2.37}$$

The variability of a RV is quantified by its *variance* $\sigma_X^2$. This measure denotes how much the RV varies from its mean and is defined as:

$$\begin{aligned} \sigma_X^2 &= E[(X - \bar{X})^2] \\ &= \int_{-\infty}^{\infty} (x - \bar{X})^2 f_X(x) dx \ . \end{aligned} \tag{2.38}$$

The square root of the variance $\sigma_X^2$ is called the *standard deviation* $\sigma_X$.

At this point the term of *moments* of RVs has to be introduced. The expected value of the $i$-th power of $X$, $X = E[X^i]$, is called the *$i$-th moment of $X$* and in addition when subtracting $\bar{X}$ before taking powers the so-called *central moment of $X$* can be obtained. Since $X = E[(X - \bar{X})^i]$ is the $i$-th central moment of $X$ it can be seen that the first central moment of an RV $X$ is always zero whereas the first moment of $X$ is its mean $\bar{X}$. From equation (2.38) it can be seen that the second central moment of a RV $X$ is equal to its variance $\sigma_X^2$.

Recalling the definition for statistical independence in equation (2.28) two RVs are said to be independent if the following relation is satisfied:

$$P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y) \ . \tag{2.39}$$

The probability density function with the following form:

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{\left[\frac{-(x-\bar{x})^2}{2\sigma_X^2}\right]} \tag{2.40}$$

is called a *Gaussian* or *normal* distribution and shown in Figure 2.5. It can be described by the *central limit theorem* that the sum of $N$ independent RVs converges towards a Gaussian distribution even if the individual distributions are not Gaussian. Many RVs in nature seem to be a so-called *Gaussian RV*, a RV with Gaussian distribution, but in fact are a sum of many individual and independent RVs.

To quantify the dependence of two RVs another measure is introduced here. The *covariance* $C_{XY}$ of $X$ and $Y$ is defined as:

$$\begin{aligned} C_{XY} &= E[(X - \bar{X})(Y - \bar{Y})] \\ &= E[XY] - \bar{X}\bar{Y} \end{aligned} \tag{2.41}$$

and is needed to estimate the so-called *correlation coefficient* $\rho_{XY}$ which is a normalized measure. In equation (2.41) the term $E[XY]$ is called the *correlation* $R_{XY}$ of $X$ and $Y$. The definition of the correlation coefficient $\rho_{XY}$ is given by:

$$\rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y} \ . \tag{2.42}$$

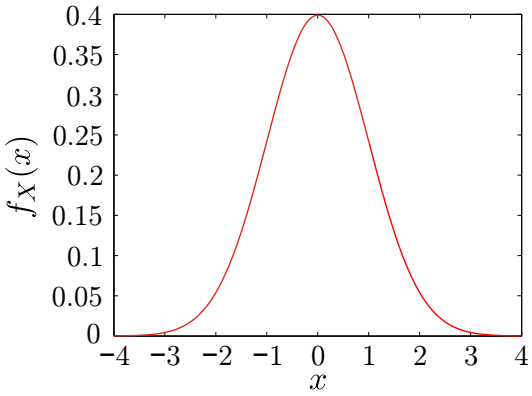



Figure 2.5: Probability density function of a Gaussian RV for $\sigma_X = 1$

It can be shown that two RVs $X$ and $Y$ are independent if $\rho_{XY} = 0$. Independence implies uncorrelatedness while the converse is not necessarily true.

For the derivations of the definitions from above scalar RVs are considered. When talking about multivariate statistics RVs are assumed to be vectors. Hence, if $\underline{X}$ is an $n$-element RV and $\underline{Y}$ is an $m$-element RV their correlation $R_{\underline{XY}}$ and covariance $C_{\underline{XY}}$ are given by the following matrices:

$$\begin{aligned} R_{\underline{XY}} &= E[\underline{X}\underline{Y}^T] \\ &= \begin{bmatrix} E[X_1Y_1] & \dots & E[X_1, Y_m] \\ \vdots & & \vdots \\ E[X_nY_1] & \dots & E[X_n, Y_m] \end{bmatrix} \end{aligned} \tag{2.43}$$

$$\begin{aligned} C_{\underline{XY}} &= E[(\underline{X} - \bar{\underline{X}})(\underline{Y} - \bar{\underline{Y}})^T] \\ &= E[\underline{X}\underline{Y}^T] - \bar{\underline{X}}\bar{\underline{Y}}^T \ . \end{aligned} \tag{2.44}$$

The correlation of a $n$-element RV $X$ with itself is called *autocorrelation* and defined as:

$$\begin{aligned} R_{\underline{X}} &= E[\underline{X}\underline{X}^T] \\ &= \begin{bmatrix} E[X_1]^2 & \dots & E[X_1, X_n]^2 \\ \vdots & & \vdots \\ E[X_n, X_1] & \dots & E[X_n^2] \end{bmatrix} \ . \end{aligned} \tag{2.45}$$

In analogous manner the *autocovariance* $C_{\underline{X}}$ can be introduced with equation (2.46) and

finally completes the set of statistical measures of RVs needed within this work.

$$\begin{aligned} C_{\underline{X}} &= E[(\underline{X} - \underline{\bar{X}})(\underline{X} - \underline{\bar{X}})^T] \\ &= \begin{bmatrix} E[(X_1 - \bar{X}_1)^2] & \dots & E[(X_1 - \bar{X}_1)(X_n - \bar{X}_n)] \\ \vdots & & \vdots \\ E[(X_n - \bar{X}_n)(X_1 - \bar{X}_1)] & \dots & E[(X_n - \bar{X}_n)^2] \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{n1} & \dots & \sigma_n^2 \end{bmatrix} \end{aligned} \tag{2.46}$$

### 2.2.2 Stochastic processes

A generalization of the concept of an RV is introduced with *stochastic* or *random processes*. The main idea of this approach is to consider the repeat of a random experiments many times. This leads to the definition of a stochastic process $X(t)$ as a RV that changes with time. Random processes are either continuous or discrete in correspondence to the characteristics of the RVs.

As a consequence of the time dependence the probability distribution and density functions also are functions of time. This leads to a further distinction of stochastic process into *non-stationary* and *stationary* processes if all marginal and joint density functions are changing with time or not. A special case of a stationary process is the so-called *ergodic process* which describes a stochastic process where every realization has the same statistical properties as the whole ensemble[7]. If the mean value $E[\underline{X}(t)] = \underline{\bar{x}}$ is independent of the choice of time $t$ and the correlation of two RVs $E[\underline{X}(t_1)\underline{X}(t_2)]$ only depends on the time difference $t_2 - t_1$, then the process is called *wide-sense stationary*. The assumption of wide-sense stationarity is usually made in practical applications since it can be made even if the probability density function changes with time.

Analogous to the previous section probability measures can be extended for stochastic processes. Considering the values $\underline{X}(t_1)$ and $\underline{X}(t_2)$, or $\underline{Y}(t_2)$, taken at time $t_1$ and $t_2$ as two RVs the corresponding definitions are given as follows:

- autocorrelation of a stochastic process:

$$R_{\underline{X}}(t_1, t_2) = E[\underline{X}(t_1)\underline{X}(t_2)^T] \; . \tag{2.47}$$

- autocovariance of a stochastic process:

$$C_{\underline{X}}(t_1, t_2) = E[\big(\underline{X}(t_1) - \underline{\bar{X}}(t_1)\big)\,\big(\underline{X}(t_2) - \underline{\bar{X}}(t_2)\big)^T] \; . \tag{2.48}$$

- cross correlation of a stochastic process:

$$R_{\underline{XY}}(t_1, t_2) = E[\underline{X}(t_1)\underline{Y}^T(t_2)] \; . \tag{2.49}$$

[7] An ensemble is a defined set of multiple possible realizations.

- cross covariance of a stochastic process:

$$C_{\underline{X}\underline{Y}}(t_1, t_2) = E[\left(\underline{X}(t_1) - \bar{\underline{X}}(t_1)\right)\left(\underline{Y}(t_2) - \bar{\underline{Y}}(t_2)\right)^T] \ . \tag{2.50}$$

The transformation from time to frequency domain often simplifies the analysis of linear systems. For deterministic signals[8] an important frequency domain concept is the *power spectral density* which also can be extended to stochastic processes. For a wide-sense stationary stochastic process $\underline{X}(t)$ the definition of the power spectral density $S_{\underline{X}}$ is equal to its definition for deterministic signals and given as the Fourier transform (see section 2.3) of the autocorrelation:

$$S_{\underline{X}}(\omega) = \mathcal{F}|R_{\underline{X}}(\tau)| = \int_{-\infty}^{\infty} R_{\underline{X}}(\tau) e^{-j\omega\tau} d\tau \ . \tag{2.51}$$

Conversely the inverse Fourier transform of the power spectrum yields the autocorrelation.

$$R_{\underline{X}}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\underline{X}}(\omega) e^{j\omega\tau} d\omega \tag{2.52}$$

Equations (2.51) and (2.52) are defined as the *Wiener-Khinchine Theorem* linking the time domain and the frequency domain for stochastic processes. The power spectral density is proportional to the total power $P_{\underline{X}}$ and equal to the mean-square value of a process. Equation (2.53) gives the definition of the total power:

$$P_{\underline{X}} = E[x^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\underline{X}}(\omega) d\omega \ . \tag{2.53}$$

Talking about power spectra of stochastic processes leads to the introduction of the term *noise*. Random processes can be classified either as *white noise* or *colored noise* depending on their power spectrum. If the power spectral density of a process is constant, i.e. $R_{\underline{X}}(0)$, over all frequencies the process is called white noise. The definition can then be written as

$$S_{\underline{X}}(\omega) = const. \ = R_{\underline{X}}(0) \ . \tag{2.54}$$

This means that the RV $\underline{X}(t_1)$ is independent from $\underline{X}(t_2)$ for all $t_1 \neq t_2$. On the other hand colored noise is given if dependency can be detected. From equation (2.52) it can be seen that the autocorrelation function of a continuous-time white noise process becomes the following:

$$R_{\underline{X}}(\tau) = R_{\underline{X}}(0)\delta(\tau) \tag{2.55}$$

where $\delta(\tau)$ is the unit impulse function[9]. This also can be determined in the autocorrelation of a discrete-time white noise process given in equation (2.56).

$$R_{\underline{X}}(k) = \begin{cases} \sigma^2_{\underline{X}} & \text{if} \quad k = 0 \\ 0 & \text{if} \quad k \neq 0 \end{cases} \tag{2.56}$$

[8] deterministic $\neq$ stochastic

[9] The function $\delta(\tau)$ is defined to be zero everywhere except at $\tau = 0$, its area is 1 while the width is 0 and the height is $\infty$.

The definition from above shows that there is no correlation of a white noise process with itself except at the initial time.

## 2.3 Time domain and frequency domain analysis

Besides the computation of mean values and mean-square values a signal analysis in time domain can be performed by decomposition of a signal into scaled and shifted impulses.

For system identification the so-called *impulse response* $h(t)$ is introduced and defined as the output signal generated by a linear time-invariant system when the input is an impulse. For a given input signal $u(t)$ the output $y(t)$ of a stable and causal system can be estimated as follows[10]:

$$y(t) \equiv \int_{-\infty}^{\infty} u(t)h(t-\tau)d\tau = \int_{-\infty}^{\infty} u(t-\tau)h(\tau)d\tau \ . \tag{2.57}$$

The output signal $y(t)$ is a so-called *convolution* between the input and the impulse response. Assuming that the input is a stochastic process $U(t)$ equation (2.57) becomes:

$$Y(t) = \int_{0}^{\infty} U(t-\tau)h(\tau)d\tau \tag{2.58}$$

with the stochastic output process $Y(t)$. In addition the expectation or mean value of the output $\bar{Y}$ , given by equation (2.59), provides more information about the characteristics of the output $Y(t)$.

$$\bar{Y} = E[Y(t)] = E\left[\int_{0}^{\infty} U(t-\tau)h(\tau)d\tau\right] \tag{2.59}$$

In the same way the autocorrelation of the output as well as the cross correlation between input and output can be expressed with the impulse response for further investigations in time domain.

The classical conversion method from time domain into frequency domain is provided by the *Fourier transform* which is defined as the following:

$$F(\omega) = \mathcal{F}[f(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{2.60}$$

where a decomposition of the time history $f(t)$ into its frequency components is performed. The inverse relationship is given by:

$$f(t) = \mathcal{F}^{-1}[F(\omega)] = \int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega \ . \tag{2.61}$$

As mentioned in the previous section, the power spectral density is a useful measure for signal analysis in frequency domain since it indicates how the signal power of a stochastic process is distributed over frequency. According to equation (2.51) its estimation requires a Fourier transform of the autocorrelation of the process.

[10] To simplify the notation the signals are assumed to be scalars for the given derivations.

## 2.4 Filtering

For noise control and other signal processing operations dynamical systems are implemented for passing or rejecting certain frequency components of a signal. These systems are called *filters* and are characterized by their different forms and applications. Only a brief introduction to a complex field in the theory of signals is given here.

Filters can be categorized in different ways as they may be linear or nonlinear, discrete-time or continuous-time, time-invariant or time-variant and many more classifications can be found. Categories concerning bandforms are *low-pass*, *high-pass*, *band-pass* and *all-pass*.

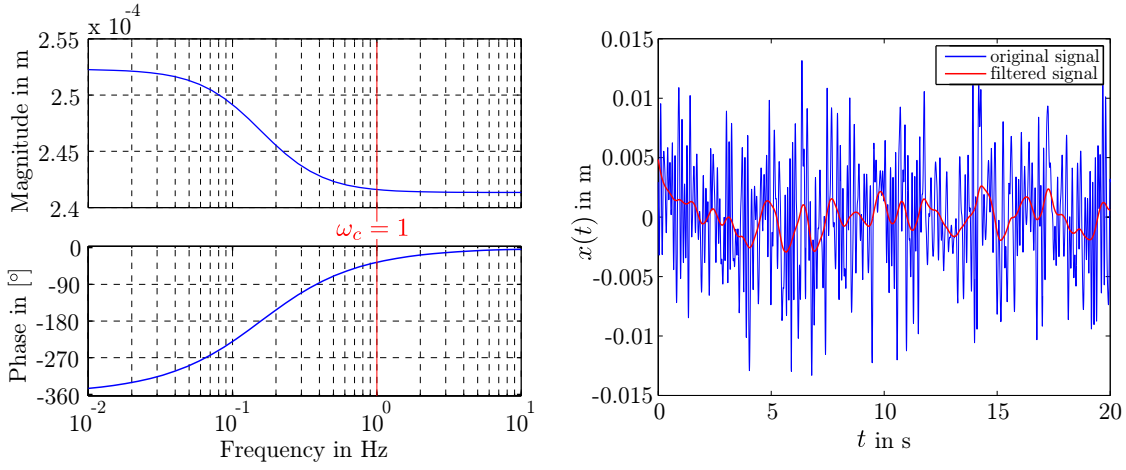
A *low-pass* filter element is implemented if only frequency components below a corresponding frequency, the so-called *cutoff frequency*, should pass. Frequency components above the selected cutoff frequency are attenuated or rejected depending on the polynomial order of the filter transfer functions.

Figure 2.6a shows the bode diagram representing the amplitude and phase response of a second-order low-pass filter with a cutoff frequency $\omega_c = 1\,\mathrm{Hz}$. The filtered result due to the application of this filter to a high frequency signal is given by Figure 2.6b.

The so-called *high-pass* filter is the opposite to the low-pass filter since it attenuates frequency components below the cutoff frequency. High-frequency components pass instead.

*Band-pass* filters are a combination of the two filters from above and only frequency components in a particular frequency band are able to pass the filtering element. A selected band of frequencies also can be rejected if a so-called *band-stop* filter is implemented. A very narrow band-stop filter is the so-called *notch filter*.

Depending on the filter design the filtered signal is time and/or phase shifted. A filter applied only for phase control is the *all-pass* filter. If a series of filters is implemented this filter can be used to correct the phase shift since its magnitude response is 1 across the spectrum.




(a) Bode diagram

(b) Original and low-pass filtered signal

Figure 2.6: Second-order low-pass filter

# 3 Inverse input estimation

The input estimation of a system where only output information is available is the task to be accomplished in this work. This states a so-called *inverse problem* which comes along with several difficulties as there might exist more than one solution, no appropriate existing model or instabilities in the solutions.

First, inverse problems are characterized and some mathematical background about the inversion of matrices, especially to handle difficulties like ill-condition or non-invertibility, is given. For better understanding a one-dimensional two-mass system is chosen for characterization, validation and comparison of the methods which are introduced in the following sections. The implementation of a *Kalman filter*, a *parity space approach*, a *regularized least squares solution* and an *inverse dynamics method* are presented in order to solve the inverse problem of the unknown input estimation. In the end of this chapter the results and differences of the given methods are discussed.

## 3.1 Characteristics of inverse problems

Considering a dynamical system, as shown in Figure 2.1, an engineering task can either be a forward problem or an inverse problem. On the one hand a forward problem can be given by a known input that is transformed by a system with known characteristics into an unknown output, this is called a *convolution*. On the other hand a forward problem can be the *system design* on the basis of a given input and predefined output.

The inverse problem instead is defined as a problem where the output is known and either the system characteristics or the input is unknown. These two tasks are called *system identification* and *deconvolution*. According to that, the inverse problem to be solved within this work is a deconvolution since an unknown input has to be estimated for a given system with measurement data.

The mathematical description of an inverse problem can be given for a linear system as the following:

$$\underline{y} = \mathbf{C}\underline{x} \qquad \textit{forward problem} \tag{3.1}$$

$$\underline{x} = \mathbf{C}^{-1}\underline{y} \qquad \textit{inverse problem} \tag{3.2}$$

with the state vector $\underline{x}$, the output vector $\underline{y}$ and the state or output matrix $\mathbf{C}$.

Generally, problems can be classified referring to the characteristics of the available information. A problem is said to be *underdetermined* if the numbers of unknowns $n$ exceeds the number of measurements (equations) $m$. Problems are regarded as *overdetermined* if $m > n$ and even determined if $m = n$. This has to be taken with care since a problem actually can be underdetermined as a result of interrelated measurements. If the amount of available information is not the same for different unknowns an inverse problem is clas-

sified as *mixed-determined*, since some unknown quantities may be overdetermined and some appear to be underdetermined.

This leads to the introduction of the terms *rank deficiency* and *condition number* of matrices. The rank of the state matrix, $rank(\mathbf{C})$, is defined as the number of linearly independent rows or columns and represents the number of independent measurements. The rank also can be misleading if a problem is so-called *ill-conditioned* as described in [26] with the following matrices:

$$\mathbf{C_1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{C_2} = \begin{bmatrix} 1 & 0 \\ 1 & 10^{-6} \end{bmatrix}, \quad \mathbf{C_3} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

with $rank(\mathbf{C_1}) = 2$, $rank(\mathbf{C_2}) = 2$ and $rank(\mathbf{C_3}) = 1$. While $\mathbf{C_1}$ is a regular[11] square matrix, $\mathbf{C_2}$ represents an ill-conditioned matrix which can be assessed by its large condition number $\kappa_{\mathbf{C_2}}$. The matrix $\mathbf{C_3}$ is a singular matrix and hence noninvertible. The definition of the condition number $\kappa$, given in equation (3.3), is the ratio between the maximum and minimum absolute value of the eigenvalues or singular values $\lambda$:

$$\kappa = \frac{\max|\lambda|}{\min|\lambda|} \ . \tag{3.3}$$

For $\mathbf{C_1}$,$\mathbf{C_2}$ and $\mathbf{C_3}$ the condition numbers are $\kappa_{\mathbf{C_1}} = 2.6$, $\kappa_{\mathbf{C_2}} = 2 \cdot 10^6$ and $\kappa_{\mathbf{C_3}} = \infty$ characterizing the invertibility of the matrices. If the condition number is very large the problem is ill-conditioned which results in high influences of numerical inaccuracies to the solution.

Another characterization of inverse problems is the consistency of a system. Inconsistent systems have data or model errors which means that there is no solution that can satisfy all the data.

## 3.2 Matrix inversion solutions

Different methods for the solution of inverse problems exist. In this section some basic concepts of matrix inversion for ill-conditioned problems which can be found in [26] are presented.

### 3.2.1 Pseudoinverse of matrices

For a regular square matrix an inverse always can be found. In contrast, the inversion of singular matrices is not possible. However, matrices with similar characteristics as an inverse can always be constructed and since these matrices are no normal inverse they are called *pseudoinverse* or *generalized inverse*.

Assuming that the output matrix $\mathbf{C}$ in equations (3.1) and (3.2) has full column rank, that means $rank(\mathbf{C}) = n$ (with $\underline{x} = [x_1, \ldots, x_n]^T$), it is possible that no solution $\underline{x}$ satisfies equation (3.1) exactly. The reason therefor is that a noisy output vector can easily lie

[11] regular $\neq$ singular, that means invertible.

outside the range[12] of $\mathbf{C}$ so that the dimension of the range of $\mathbf{C}$ ($= rank(\mathbf{C})$) is smaller than $m$ (with $\underline{y} = [y_1, \ldots, y_m]^T$). In this case $\mathbf{C}$ is noninvertible since $n < m$ and the equivalent inverse problem given in equation (3.2) can not be solved by matrix inversion. However, an estimated solution $\underline{\hat{x}}$ can be obtained if a pseudoinverse $\mathbf{C}^+$ of $\mathbf{C}$ is used instead:

$$\underline{\hat{x}} = \mathbf{C}^+ \underline{y} \; . \tag{3.4}$$

It can be shown that the pseudoinverse $\mathbf{C}^+$ is not the normal inverse of $\mathbf{C}$ if the products $\mathbf{C}\mathbf{C}^+$ and $\mathbf{C}^+\mathbf{C}$ are not equal to the identity matrix $\mathbf{I}$. According to the forward problem, given by equation (3.1), an estimated solution $\underline{\hat{y}}$ of the output $\underline{y}$ can be computed if the input parameters were $\underline{\hat{x}}$. The solution can be obtained by replacing $\underline{\hat{x}}$ with equation (3.4) as the following:

$$\underline{\hat{y}} = \mathbf{C}\underline{\hat{x}} = \mathbf{C}\mathbf{C}^+ \underline{y} = \mathbf{R}_D \underline{y} \tag{3.5}$$

with the so-called *data resolution matrix* $\mathbf{R}_D = \mathbf{C}\mathbf{C}^+$. If $\mathbf{R}_D$ is equal to the identity matrix $\mathbf{I}$ the problem can be resolved without any error between $\underline{\hat{y}}$ and $\underline{y}$. Similarly, a *model resolution matrix* $\mathbf{R}_M = \mathbf{C}^+\mathbf{C}$ can be introduced by solving the inverse problem from equation (3.2) and replacing the measured data $\underline{y}$ with $\mathbf{C}\underline{x}$:

$$\underline{\hat{x}} = \mathbf{C}^+ \underline{y} = \mathbf{C}^+\mathbf{C}\underline{x} = \mathbf{R}_M \underline{x} \; . \tag{3.6}$$

From equation (3.6) it can be seen that $\mathbf{R}_M = \mathbf{I}$ means that the estimated states $\underline{\hat{x}}$ are equal to the real states $\underline{x}$ of the system.

As a result, the resolution matrices $\mathbf{R}_D$ and $\mathbf{R}_M$ represent a measure for the appropriateness of the pseudoinverse for the given problem. In the following two sections the computation of the pseudoinverse $\mathbf{C}^+$ of $\mathbf{C}$ is presented by two different methods.

### 3.2.2 Least squares solution

Considering a linear system, as given in equations (3.1) and (3.2), for which an approximate solution $\underline{x}$ has to be found, the 2-norm of the difference between $\mathbf{C}\underline{x}$ and $\underline{y}$, $\|\mathbf{C}\underline{x} - \underline{y}\|_2$, is a measure of the quality of the solution.

A measure of the error between $\mathbf{C}\underline{x}$ and $\underline{y}$ can be introduced as a residual vector $\underline{e}$ with

$$\underline{e} = \underline{y} - \mathbf{C}\underline{x} \; . \tag{3.7}$$

By minimizing the 2-norm of the residuals, $\|\underline{e}\|_2$, a so called *least squares solution (LSS)* can be obtained. To find the minimum of $\sqrt{\underline{e}^T\underline{e}}$ an objective function $\Gamma = \underline{e}^T\underline{e}$ has to be minimized by setting its derivative equal to zero. This can be performed as follows:

$$\begin{aligned} \Gamma &= \underline{e}^T\underline{e} \\ &= (\underline{y} - \mathbf{C}\underline{x})^T(\underline{y} - \mathbf{C}\underline{x}) \\ &= \underline{y}^T\underline{y} - \underline{y}\,\mathbf{C}\underline{x} - \underline{x}^T\mathbf{C}^T\underline{y} + \underline{x}^T\mathbf{C}^T\mathbf{C}\underline{x} \end{aligned} \tag{3.8}$$

[12] The set of all possible linear combinations of the column vectors of a matrix is the *range* of the matrix and is also called the *column space* of the matrix.

$$\frac{\partial \Gamma}{\partial \underline{x}} \stackrel{!}{=} \underline{0} = \underline{0} - \underline{y}\,\mathbf{C} - \mathbf{C}^T\underline{y} + 2\,\mathbf{C}^T\mathbf{C}\underline{x}\ . \tag{3.9}$$

After rearranging equation (3.9) the least squares solution for $\underline{x}$ becomes the following:

$$\underline{x} = (\mathbf{C}^T\mathbf{C})^{-1}\,\mathbf{C}^T\underline{y}\ . \tag{3.10}$$

Therefore the square matrix $\mathbf{C}^T\mathbf{C}$ has to be invertible, which is fulfilled since $rank(\mathbf{C}) = n$. After substitution equation (3.10) can be written equally to equation (3.4) as:

$$\underline{x} = \mathbf{C}^+\underline{y} \tag{3.11}$$

with the generalized inverse $\mathbf{C}^+ = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$.

### 3.2.3 Singular value decomposition

For ill-conditioned and/or rank-deficient systems the *singular value decomposition (SVD)* can be a useful method to find a pseudoinverse of a matrix. In addition, a solution for inverse problems can be estimated similarly to above (compare equation (3.4)).

Applying the SVD the $m \times n$ matrix $\mathbf{C}$ from equation (3.1) and (3.2) is factored into

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{3.12}$$

where according to [3]

- $\mathbf{U}$ is a $m \times m$ orthogonal matrix with columns that are the eigenvectors $\underline{u}$ of $\mathbf{C}\mathbf{C}^T$ spanning the data space, $\mathbb{R}^m$.
- $\mathbf{V}$ is a $n \times n$ orthogonal matrix with columns that are the eigenvectors $\underline{v}$ of $\mathbf{C}^T\mathbf{C}$ spanning the model space, $\mathbb{R}^n$.
- $\mathbf{S}$ is a $m \times n$ diagonal matrix with nonnegative diagonal elements in descending order called *singular values*.

Some singular values may be zero so that $\mathbf{S}$ can be written as:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{3.13}$$

where $\mathbf{S}_p$ is a $p \times p$ diagonal matrix composed of those $p$ singular values that are nonzero. This leads to a compact form of equation (3.12):

$$\mathbf{C} = \mathbf{U}_p\mathbf{S}_p\mathbf{V}_p^T\ . \tag{3.14}$$

Since $\mathbf{U}_p$ and $\mathbf{V}_p$ are orthogonal matrices their inverse is equal to their transpose and a generalized inverse of $\mathbf{C}$ can be written as:

$$\mathbf{C}^+ = \mathbf{V}_p\mathbf{S}_p^{-1}\mathbf{U}_p^T\ . \tag{3.15}$$

The matrix $\mathbf{C}^+$ from equation (3.15) is called the *Moore-Penrose-Pseudoinverse* of $\mathbf{C}$.

An inverse solution of equation (3.1) can be estimated with the use of the generalized inverse from above in equation (3.11). However, the generalized inverse solution may include terms involving column vectors in $\mathbf{V}_p$ with very small nonzero singular values for ill-conditioned problems. This causes high sensitivity to small amounts of noise in the data and can lead to an instability of the solution which can be measured with the above introduced condition number $\kappa$ (see equation (3.3)).

## 3.3 One-dimensional two-mass system

In this section a simple dynamical system is presented for better understanding of the application of inverse input estimation methods. Therefore a mechanical model is described first by its equations of motion and transformed into a state-space model for the statement of the inverse input estimation problem.

### 3.3.1 Mechanical model

Figure 3.1 shows the mechanical model of a two-mass system with free movement in vertical direction. It consists of two rigid bodies characterized by the masses $m_1$ and $m_2$, two springs with stiffnesses $c_1$ and $c_2$ and a damper with damping constant $d_1$.

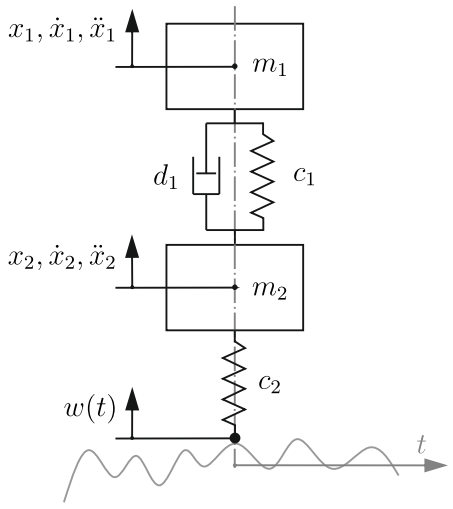



Figure 3.1: One-dimensional two-mass system

The model is described by the following equations of motions according to Euler's first law of motion[13]:

$$\ddot{x}_1 + \frac{d_1}{m_1}(\dot{x}_1 - \dot{x}_2) + \frac{c_1}{m_1}(x_1 - x_2) = 0 \tag{3.16}$$

$$\ddot{x}_2 - \frac{d_1}{m_2}(\dot{x}_1 - \dot{x}_2) - \frac{c_1}{m_2}(x_1 - x_2) + \frac{c_2}{m_2}x_2 = \frac{c_2}{m_2}w \tag{3.17}$$

[13]The definitions of Euler's equations of motions for rigid bodies can be found e.g. in [15] and [23].

where $x_1$, $x_2$ and the time derivatives $\dot{x}_1$, $\dot{x}_2$, $\ddot{x}_1$ and $\ddot{x}_2$ are describing the motion of the center of the masses $m_1$ and $m_2$. The term on the right side of equation (3.17) denotes an input force as a result of the excitation $w = w(t)$. [14]

Equations (3.16) and (3.17) are two second order differential equations. A transformation into a system of differential equations of first order can be performed by introducing the state-vector $\underline{x}$. The state-space model can then be written as the following:

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{G}w \tag{3.18}$$

with

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dot{x}_1 \\ \dot{x}_2 \end{bmatrix}, \qquad \dot{\underline{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{c_1}{m_1} & \frac{c_1}{m_1} & -\frac{d_1}{m_1} & \frac{d_1}{m_1} \\ \frac{c_1}{m_2} & -\frac{(c_1+c_2)}{m_2} & \frac{d_1}{m_2} & -\frac{d_1}{m_2} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{c_2}{m_2} \end{bmatrix}.$$

At this point no classification of the stated problems in terms of knowledge of the input $w$ has been made. Since the characteristics of the system are assumed to be known an inverse problem can be formulated with the goal of estimation of the input $w$ if any output can be measured.

### 3.3.2 Input estimation problem

For many applications a LTI system (see section 2.1) is characterized by a known input $\underline{u}(t)$ generating the output $\underline{y}(t)$ as a system response. If the system consists of unknown inputs an additional input quantity $\underline{w}(t)$ has to be introduced. A block diagram representation for a system with known and unknown input is shown in Figure 3.2.

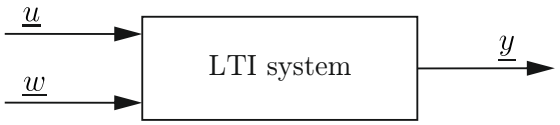



Figure 3.2: LTI system with known and unknown input

The corresponding state-space model of the system is given by:

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} + \mathbf{G}\underline{w} \tag{3.19}$$

$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} + \mathbf{H}\underline{w} \tag{3.20}$$

with the state vector $\underline{x}$, the input vectors $\underline{u}$, $\underline{w}$ and the system matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{G}$ and $\mathbf{H}$.

The two-mass system from above (see Figure 3.1) only has a single input $w(t)$ which is assumed to be unknown[15]. This means that no known input affects the system and $\underline{u}(t)$ can be set to zero which simplifies the state-space model from equations (3.19) and (3.20).

[14] The excitation term is named "$w$" instead of "$u$" since "$u$" usually represents a known input signal. Within this work the input of the given system is regarded as unknown and therefore named "$w$".

[15] Note that the vector $\underline{w}$ becomes a scalar $w$ as there is only a single excitation $w(t)$.

Another assumption is made concerning the output $\underline{y}(t)$ of the system in this example: The output defined to be the acceleration of the mass $m_1$, that is $y = \ddot{x}_1$, and as it is only one quantity it becomes a scalar. Therefore the output matrix $\mathbf{C}$ is given by:

$$\mathbf{C} = \begin{bmatrix} -\frac{c_1}{m_1} & \frac{c_1}{m_1} & -\frac{d_1}{m_1} & \frac{d_1}{m_1} \end{bmatrix} .$$

The described system has no direct feedthrough and as a result the system matrix $\mathbf{H}$ is a $(4 \times 1)$ - zero matrix. Finally, the state-space model of the two-mass system can be written as follows:

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\cancelto{0}{\underline{u}} + \mathbf{G}w \qquad \Rightarrow \qquad \dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{G}w \tag{3.21}$$

$$y = \mathbf{C}\underline{x} + \mathbf{D}\cancelto{0}{u} + \cancelto{\mathbf{0}}{\mathbf{H}}w \qquad \Rightarrow \qquad y = \mathbf{C}\underline{x} \; . \tag{3.22}$$

Since later in this work the obtained output signals are measurement data of the response of a railway vehicle system it is required to use a discrete form of the state-space model (see section 2.1.3). With a given time step $k \in \{0, 1, 2, \ldots, n\}$ the discretization of equations (3.21) and (3.22) leads to:

$$\underline{x_{k+1}} = \mathbf{A_d}\underline{x_k} + \mathbf{G_d}w_k \tag{3.23}$$

$$y_k = \mathbf{C_d}\underline{x_k} \tag{3.24}$$

with the discrete-time state matrices $\mathbf{A_d}$, $\mathbf{G_d}$ and $\mathbf{C_d}$. According to [29] the conversion from continuous-time to discrete-time can be computed if $\mathbf{A}$ is invertible, that is the case here, as follows:

$$\begin{aligned} \mathbf{A_d} &= e^{\mathbf{A}\Delta t} \\ \mathbf{G_d} &= \mathbf{A_d}[\mathbf{I} - e^{-\mathbf{A}\Delta t}]\mathbf{A}^{-1}\mathbf{G} \\ \mathbf{C_d} &= \mathbf{C} \end{aligned} \tag{3.25}$$

where the discretization step size $\Delta t$ is defined as $\Delta t = t_k - t_{k-1}$ with the discrete time point $t_k$.

The inverse problem to be solved is the estimation of the unknown excitation $w$ of the two-mass system described above with knowledge about the system characteristics, given by the system matrices, and the acceleration data of the mass $m_1$ representing the output $y$.

## 3.4 Kalman filter

For the inverse input estimation a method using a *Kalman filter*[16] is introduced. This filter computes state estimates of a system and is, according to [31], a so-called optimal filter as its purpose is to minimize the spread of the estimate-error probability density. The filter propagates the mean and the covariances of the state taking into account system dynamics and inputs as well as it incorporates measurements and measurement error statistics.

[16] The filter is named after R.E. Kalman, who presented in 1960 the well-known filtering method in [14].

An introduction of the discrete Kalman filter algorithm, which is implemented later, is given in [35]. The following definitions and derivations needed for the filter implementation can also be found in [29]. Later, the filtering method is applied to estimate the unknown input of the two-mass system from section 3.3.

### 3.4.1 The Kalman filter implementation

The aim of the application of a Kalman filter element is to minimize the steady-state error covariance by constructing a state estimate $\underline{\hat{x}}$. Figure 3.3 shows the block diagram of a system with an integrated Kalman filter.

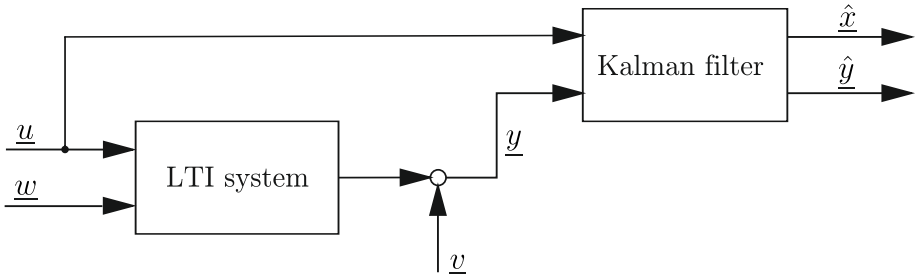



Figure 3.3: LTI system with a Kalman filter

For this system the discrete-time state-space model is given by:

$$\underline{x_{k+1}} = \mathbf{A_d}\underline{x_k} + \mathbf{B_d}\underline{u_k} + \mathbf{G_d}\underline{w_k} \tag{3.26}$$

$$\underline{y_k} = \mathbf{C_d}\underline{x_k} + \mathbf{D_d}\underline{u_k} + \mathbf{H_d}\underline{w_k} + \underline{v_k} \tag{3.27}$$

with the state-vector $\underline{x}$, the known input $\underline{u}$ and the output $\underline{y}$. The terms $\underline{w}$ and $\underline{v}$ represent *process noise* and *measurement noise*.

In the context of Kalman filtering it is assumed that the noise processes $\underline{w}$ and $\underline{v}$ in equations (3.26) and (3.27) are white, uncorrelated and zero-mean. Hence the covariance matrices (see section 2.2) are known and defined as:

$$E[\underline{w}\underline{w}^T] = \bar{\mathbf{Q}}, \quad E[\underline{v}\underline{v}^T] = \mathbf{R}, \quad E[\underline{v}\underline{w}^T] = \mathbf{0}\ . \tag{3.28}$$

The amount of available information, e.g. measurements $\underline{y}$, affects the state estimation and depends on the problem to be solved. Therefore a distinction in the notation for an estimated state $\underline{\hat{x}}$ is made as the following:

$$\underline{\hat{x}_{k|k-1}} = E[\underline{x_k}|\underline{y_1}, \ldots, \underline{y_{k-1}}] \qquad \textit{a priori estimate}$$

$$\underline{\hat{x}_{k|k}} = E[\underline{x_k}|\underline{y_1}, \ldots, \underline{y_k}] \qquad \textit{a posteriori estimate}\ .$$

That means if all measurements up to and including time $k$ are available the estimate is called an *a posteriori estimate* and if all measurements before time $k$ is available the estimate is called an *a priori estimate*.

The relationship between real and estimated states can be described by the covariance

of the estimation error $\underline{x_k} - \underline{\hat{x}_k}$ which is defined as:

$$\mathbf{P_k} = E[(\underline{x_k} - \underline{\hat{x}_k})(\underline{x_k} - \underline{\hat{x}_k})^T] \ . \tag{3.29}$$

At the beginning of the estimation process an initialization has to be performed by declaring an initial estimate $\underline{\hat{x}_0}$ of the initial state $\underline{x_0} = \underline{x}(t = t_0 = 0)$:

$$\underline{\hat{x}_{0|0}} = E[\underline{x_0}] \tag{3.30}$$

$$\mathbf{P_{0|0}} = E[(\underline{x_0} - \underline{\hat{x}_{0|0}})(\underline{x_0} - \underline{\hat{x}_{0|0}})^T] \ . \tag{3.31}$$

The following equations describe the Kalman filter algorithm and have to be calculated for each time step $k = 1, \ldots, n$ after the initialization:

$$\mathbf{P_{k|k-1}} = \mathbf{A_d P_{k-1|k-1} A_d}^{-\mathbf{1}} + \mathbf{Q} \tag{3.32}$$

$$\underline{\hat{x}_{k|k-1}} = \mathbf{A_d}\underline{\hat{x}_{k-1|k-1}} + \mathbf{B_d}\underline{u_{k-1}} \ . \tag{3.33}$$

These equations are called *time-update equations* and describe a *prediction step*. For the covariance $\mathbf{P_k}$ the prediction is given by equation (3.32) with $\mathbf{Q} = \mathbf{G_d}\bar{\mathbf{Q}}\mathbf{G_d^T}$ and the a priori state estimate $\underline{\hat{x}_{k|k-1}}$ is predicted according to (3.33). The next task is given by the so-called *measurement-update equations* or *update step*, where the measurements $\underline{y_k}$ are considered to derive the a posteriori state estimate $\underline{\hat{x}_{k|k}}$:

$$\begin{aligned} \mathbf{L_k} &= \mathbf{P_{k|k-1}C_d}^{\mathbf{T}}(\mathbf{C_d P_{k|k-1} C_d}^{\mathbf{T}} + \mathbf{R})^{-\mathbf{1}} \\ &= \mathbf{P_{k|k}C_d}^{\mathbf{T}}\mathbf{R}^{-\mathbf{1}} \end{aligned} \tag{3.34}$$

$$\begin{aligned} \mathbf{P_{k|k}} &= (\mathbf{I} - \mathbf{L_k C_d})\mathbf{P_{k|k-1}}(\mathbf{I} - \mathbf{L_k C_d})^{\mathbf{T}} + \mathbf{L_k R L_k}^{\mathbf{T}} \\ &= \cdots = (\mathbf{I} - \mathbf{L_k C_d})\mathbf{P_{k|k-1}} \end{aligned} \tag{3.35}$$

$$\underline{\hat{x}_{k|k}} = \underline{\hat{x}_{k|k-1}} + \mathbf{L_k}(\underline{y_k} - \mathbf{C_d}\underline{\hat{x}_{k|k-1}}) \ . \tag{3.36}$$

At this point the so-called *Kalman filter gain* $\mathbf{L_k}$ has to be introduced and is defined by equation (3.34). The final state estimation is given by equation (3.36) and contents the Kalman filter gain $\mathbf{L_k}$ for weighting the estimated measurement error term $(\underline{y_k} - \mathbf{C_d}\underline{\hat{x}_{k|k-1}})$.

With the presented algorithm states can be estimated on the basis of measurements and system characteristics which can be applied for inverse problem solving.

### 3.4.2 Kalman filter application to the two-mass system

Recalling the two-mass system from section 3.3 the Kalman filter is applied here to solve the inverse problem of the unknown input estimation.

If the unknown input $w(t)$ is regarded as process noise the Kalman filter can be implemented as shown in Figure 3.4. This system does not contain any known input $u(t)$ and leads to the approach to derive the unknown input, which is assumed to occur as process noise, through using a Kalman filter algorithm.

Similar to the derivation of equations (3.21) and (3.22) the discrete state-space model

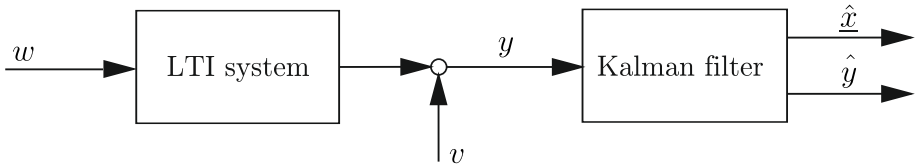



Figure 3.4: State-space model with only unknown input

from equations (3.26) and (3.26) can be simplified for the given system as follows:

$$\underline{x_{k+1}} = \mathbf{A_d}\underline{x_k} + \mathbf{B_d}\cancelto{0}{u_k} + \mathbf{G_d}w_k \qquad \Rightarrow \qquad \underline{x_{k+1}} = \mathbf{A_d}\underline{x_k} + \mathbf{G_d}w_k \tag{3.37}$$

$$y_k = \mathbf{C_d}\underline{x_k} + \mathbf{D_d}\cancelto{0}{u_k} + \cancelto{\mathbf{0}}{\mathbf{H_d}}w_k + v_k \qquad \Rightarrow \qquad y_k = \mathbf{C_d}\underline{x_k} + v_k \tag{3.38}$$

with the state-vector $\underline{x}$, the unknown input $w$, the known output $y$ and the measurement noise $v$ where the covariance data is given for. The state matrices $\mathbf{A_d}$, $\mathbf{G_d}$ and $\mathbf{C_d}$ already have been derived in section 3.3.

According to the given system the estimator equations (3.33) and (3.36) of the Kalman filter algorithm for the two-mass system can be written as:

$$\underline{\hat{x}_{k|k-1}} = \mathbf{A_d}\underline{\hat{x}_{k-1|k-1}} + \mathbf{B_d}\cancelto{0}{u_{k-1}} \Rightarrow \quad \underline{\hat{x}_{k|k-1}} = \mathbf{A_d}\underline{\hat{x}_{k-1|k-1}} \tag{3.39}$$

$$\underline{\hat{x}_{k|k}} = \underline{\hat{x}_{k|k-1}} + \mathbf{L_k}(y_k - \mathbf{C_d}\underline{\hat{x}_{k|k-1}}) \ . \tag{3.40}$$

The calculation of a state estimate $\underline{\hat{x}}$ can be performed by the application of the recursive estimation algorithm which is given by equations (3.32) - (3.36). Therefore equation (3.33) has to be substituted by equation (3.39).

After the state estimates have been derived an estimated input can be computed. Rearranging of equation (3.23) leads to the following:

$$w_k = \mathbf{G_d^{-1}}[\underline{x_{k+1}} - \mathbf{A_d}\underline{x_k}] \ . \tag{3.41}$$

Finally the unknown input $w$ can be estimated by solving equation (3.41) using the state estimate $\underline{\hat{x}_k}$ instead of $\underline{x_k}$. The results of the performed input estimation are presented and discussed in section 3.8.

## 3.5 Parity space approach

Another method is given by the *parity space approach* which can be found in [7] where two methods for stochastic fault detection are compared. In this section the general diagnosis algorithm is introduced first, followed by the application to the two-mass system from section 3.3.

### 3.5.1 The parity space approach algorithm

Originally, the parity space approach is applied for fault detection and isolation in a stochastic setting. Within this work, the algorithm is introduced in order to detect an unknown input based on the following discrete-time state-space model for a linear system:

$$\underline{x_{k+1}} = \mathbf{A_d}\underline{x_k} + \mathbf{B_{u,d}}\underline{u_k} + \mathbf{B_{w,d}}\underline{w_k} + \mathbf{B_{v,d}}\underline{v_k} \tag{3.42}$$

$$\underline{y_k} = \mathbf{C_d}\underline{x_k} + \mathbf{D_{u,d}}\underline{u_k} + \mathbf{D_{w,d}}\underline{w_k} + \underline{e_k} \tag{3.43}$$

with the known input $\underline{u}$ and an unknown fault input $\underline{w}$. The occurring terms $\underline{v}$ and $\underline{e}$ are process and measurement noise which also can be found in the model of the Kalman filter introduced above.

To formulate the diagnosis task a sliding window is applied to a recursive problem. The sliding window $L$ is generated by stacking signal values to define signal vectors $\mathbf{Y_k} = [\underline{y_{k-L+1}}^T, \ldots, \underline{y_k}^T]^T$ for all signals $s \in \{\underline{u}, \underline{w}, \underline{v}\}$.

Also the so-called *Hankel matrices* $\mathbf{H_s}$ for all signals $s$ and an observability matrix $\mathbf{O}$ have to be built as the following:

$$\mathbf{H_s} = \begin{bmatrix} \mathbf{D_{s,d}} & 0 & 0 & \ldots & 0 \\ \mathbf{C_d B_s} & \mathbf{D_{s,d}} & 0 & \ldots & 0 \\ \mathbf{C_d A_d B_s} & \mathbf{C_d B_s} & \mathbf{D_{s,d}} & \ldots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \mathbf{C_d A_d^{L-2} B_s} & \ldots & \ldots & \mathbf{C_d B_s} & \mathbf{D_{s,d}} \end{bmatrix}, \quad \mathbf{O} = \begin{bmatrix} \mathbf{C_d} \\ \mathbf{C_d A_d} \\ \vdots \\ \mathbf{C_d A_d^{L-1}} \end{bmatrix}. \tag{3.44}$$

Considering the matrices from above equations (3.42) and (3.43) can be rewritten. This is leading to the central expression of the parity space approach:

$$\mathbf{Y_k} - \mathbf{H_u U_k} = \mathbf{O}\underline{x_{k-L+1}} + \mathbf{H_W W_k} + \mathbf{H_v V_k} + \mathbf{E_k}\ . \tag{3.45}$$

For the solution of an inverse problem equation (3.45) can be rearranged as needed and computed after an initialization has been made. The process is explained more detailed in the following section where the method is applied exemplary to the two-mass system introduced in section 3.3.

### 3.5.2 Parity space approach applied to the two-mass system

For the application of the parity space approach to the state-space model of the two-mass system, given in equations (3.23) and (3.24), some assumptions are made. The system is excited by an unknown quantity $\underline{w}$ called fault input in the system presented above (see equations (3.42) and (3.43)). Additionally, no occurrence of noise is expected and since not any known input $\underline{u}$ is applied to the system equation (3.45) can be written as:

$$\mathbf{Y_k} = \mathbf{O}\underline{x_{k-L+1}} + \mathbf{H_W W_k} \tag{3.46}$$

or in matrix form:

$$\mathbf{Y_k} = \underbrace{\begin{bmatrix} \mathbf{C_d} \\ \mathbf{C_d A_d} \\ \mathbf{C_d A_d^2} \\ \vdots \\ \mathbf{C_d A_d^{L-1}} \end{bmatrix}}_{:=\mathbf{O}} \underline{x_{k-L+1}} + \underbrace{\begin{bmatrix} \cancel{\mathbf{D_w}}^{\mathbf{0}} & 0 & 0 & \dots & 0 \\ \mathbf{C_d G_d} & 0 & 0 & \dots & 0 \\ \mathbf{C_d A G_d} & \mathbf{C G_d} & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \mathbf{C_d A^{L-2} G_d} & \dots & \dots & \mathbf{C_d G_d} & 0 \end{bmatrix}}_{:=\mathbf{H_w}} \mathbf{W_k} \ . \tag{3.47}$$

The goal is to rearrange equation (3.46) to obtain the following expression:

$$\mathbf{W_k} = \mathbf{H_w^{-1}}(\mathbf{Y_k} - \mathbf{O}\underline{\hat{x}_{k-L+1}}) \tag{3.48}$$

with $\mathbf{W_k} = [\hat{w}_{k-L+1}, \dots, \hat{w}_k]^T$. It can be seen from equation (3.47) that the Hankel matrix $\mathbf{H_w}$ is noninvertible since it contains a zero column and is therefore not of full rank, respectively is singular. Alternatively, a pseudoinverse $\mathbf{H_w^+}$ has to be built. This can be achieved by computation of a least squares solution e.g., see equation (3.9), to obtain a matrix which has the same size as $\mathbf{H_w}$ and satisfies the conditions of an inverse.

Equation (3.48) gives a recursive solution for the unknown input $w$. After an initialization and computation of the estimate $\hat{w}_k$ the state estimates $\underline{\hat{x}_{k-L+1}}$ can be computed for each time step $k = 1, \dots, n$ with:

$$\underline{\hat{x}_{k-L+1}} = \mathbf{A_d}\underline{\hat{x}_{k-L}} + \mathbf{G_d}\hat{w}_{k-L} \ . \tag{3.49}$$

An algorithm for the unknown input identification has been derived using a sliding window considering the influences of prior states and inputs to the present state of the system. For the purpose of method validation the estimation results for the two-mass system are discussed in section 3.8.

## 3.6 Regularized least squares solution

In section 3.2 matrix inversion methods for inverse problem solving were presented. The here introduced *regularized least squares solution* can be found in [26] and is an extension of a least squares solution. The aim is to include available information about the solution during the inversion stage in order to solve mixed- and underdetermined problems. The method is applied to the two-mass system from section 3.3 to estimate the unknown input.

### 3.6.1 Determination of the RLSS

As the least squares solution (LSS), which is described in section 3.2.2, applies to overdetermined problems and many inverse problems are mixed-determined (see section 3.1), an additional term is introduced in the objective function leading to the so-called regularized least squares solution (RLSS). The RLSS tries to include a priori information about the solution during the inversion stage to prevent noise in the solution.

Considering an inverse problem, as given by equation (3.2), the expanded objective func-

tion $\Gamma$ according to a residual vector $\underline{e}$ (see equation (3.7)) is the following:

$$\begin{aligned}\Gamma &= \underline{e}^T\underline{e} + \lambda\left[(\mathbf{R_r}\underline{x})^T\mathbf{R_r}\underline{x}\right] \\ &= (\underline{y} - \mathbf{C}\underline{x})^T(\underline{y} - \mathbf{C}\underline{x}) + \lambda\left[(\mathbf{R_r}\underline{x})^T\mathbf{R_r}\underline{x}\right] \\ &= \underline{y}^T\underline{y} - \underline{y}\,\mathbf{C}\underline{x} - \underline{x}^T\mathbf{C}^T\underline{y} + \underline{x}^T\mathbf{C}^T\mathbf{C}\underline{x} + \lambda\left[\underline{x}^T{\mathbf{R_r}}^T\mathbf{R_r}\underline{x}\right] .\end{aligned} \tag{3.50}$$

Equation (3.50) includes, compared to equation (3.9), an additional regularization criterion $\lambda[(\mathbf{R_r}\underline{x})^T\mathbf{R_r}\underline{x}]$ with a regularization matrix $\mathbf{R_R}$ and a nonnegative coefficient $\lambda$ which provides weighting of the regularization term. Setting the partial derivative of $\Gamma$ equal to zero in order to minimize the objective function leads to:

$$\frac{\partial\Gamma}{\partial\underline{x}} \stackrel{!}{=} \underline{0} = \underline{0} - 2\,\mathbf{C}^T\underline{y} + 2\,\mathbf{C}^T\mathbf{C}\underline{x} + 2\,\lambda\,\mathbf{R_r}{\mathbf{R_r}}^T\mathbf{R_r}\;\underline{x} \tag{3.51}$$

$$= -\mathbf{C}^T\underline{y} + (\mathbf{C}^T\mathbf{C} + \lambda{\mathbf{R_r}}^T\mathbf{R_r})\;\underline{x} . \tag{3.52}$$

After inversion of the term $(\mathbf{C}^T\mathbf{C} + \lambda{\mathbf{R_r}}^T\mathbf{R_r})$ and rearranging of equation (3.52) an estimate $\hat{\underline{x}}$ for $\underline{x}$ can be obtained representing a regularized least squares solution of the inverse problem:

$$\hat{\underline{x}} = (\mathbf{C}^T\mathbf{C} + \lambda\,{\mathbf{R_r}}^T\mathbf{R_r})^{-1}\,\mathbf{C}^T\underline{y} . \tag{3.53}$$

The regularization matrix $\mathbf{R_r}$ has to be constructed with different criteria depending on the stated problem with respect to the a priori knowledge of $\underline{x}$. If the local variation of $\underline{x}$ can be approximated with a straight line e.g., that means $x = a + b\cdot p$, the second derivative $\frac{d^2x}{dp^2}$ is minimized and leads to a finite difference of $x_{i+1} - 2x_i + x_{i-1}$ which can be used to compute the rows of $\mathbf{R_r}$. In this case a row would have the form of $[0\ldots0\;1\;-2\;1\;0\ldots0]$.[26]

Another value to be chosen is the regularization coefficient $\lambda$ which depends on the characteristics of the given problem, the quality of the data as well as the adequacy of the considered model. The optimal value of $\lambda$ can either be found by following the evolution of the residual or by observation of the evolution of the solution for different values. After these recursive tasks have been performed the right level for $\lambda$ can be selected to return a physical meaningful solution and to adequately justify the data during the estimation process.

The presented method for stabilizing the solution of inverse problems in a regularized least squares sense is also called *Tikhonov regularization*. An introduction with examples can be found [3].

### 3.6.2 RLSS of the two-mass system

A RLSS can be estimated for the two-mass system with the unknown input $w$, shown in Figure 3.1, and the state-space model is described with equation (3.23) and (3.24).

The unknown input can be obtained by solving equation (3.41) using a state estimate $\hat{\underline{x}}$. According to equation (3.53) the estimation of the state vector can be performed applying a RLSS method for the two-mass system as follows:

$$\hat{\underline{x}} = (\mathbf{C}^T\mathbf{C} + \lambda\,{\mathbf{R_r}}^T\mathbf{R_r}\mathbf{R})^{-1}\,\mathbf{C}^T y \tag{3.54}$$

with the output or data matrix $\mathbf{C}$ and the given output $y$. The regularization matrix $\mathbf{R_r}$ and the regularization coefficient $\lambda$ for the considered system has to be defined next as described above.

After all, the unknown input $w$ of the two-mass system can be computed based on a state estimate $\underline{\hat{x}}$ obtained from the RLSS. The results of the performed computation can be found in section 3.8.

## 3.7 Inverse dynamics method

The last estimation method presented in this work is the *inverse dynamics method* and can be found in [37] and [38] where it is applied to a railway vehicle system to identify wheel-rail contact forces. First the mathematical derivation of the algorithm is given which is followed by the solution of the inverse problem of the unknown input estimation for the two-mass system from section 3.3.

### 3.7.1 The inverse dynamics method algorithm

Considering a discrete-time LTI system (see section 2.1.3) the inverse problem of unknown input estimation can be given by the following state-space model:

$$\underline{x_{k+1}} = \mathbf{A_d}\underline{x_k} + \mathbf{G_d}\underline{w_k} \tag{3.55}$$

$$\underline{y_k} = \mathbf{C_d}\underline{x_k} \tag{3.56}$$

with the state vector $\underline{x}$, the output vector $\underline{y}$ the unknown input $\underline{w}$ and the system matrices $\mathbf{A_d}$, $\mathbf{C_d}$ and $\mathbf{G_d}$. The least-squares error of the output including a weighting term $\lambda_\mathbf{1}$ can be written as the following:

$$E = \sum_{k=1}^{n} (\underline{y_k} - \underline{\hat{y}_k})^T \lambda_\mathbf{1} (\underline{y_k} - \underline{\hat{y}_k}) \tag{3.57}$$

with $n$ measurements $\underline{y_i} \in \{\underline{y_1}, \ldots, \underline{y_n}\}$ and its estimates $\underline{\hat{y}_i} \in \{\underline{\hat{y}_1}, \ldots, \underline{\hat{y}_n}\}$. If the problem is ill-conditioned, as is the case very often because of noisy data, a smoothing term can be added referring to the Tikhonov regularization (see section 3.6). Equation (3.57) then becomes to:

$$E = \sum_{k=1}^{n} (\underline{y_k} - \underline{\hat{y}_k})^T \lambda_\mathbf{1} (\underline{y_k} - \underline{\hat{y}_k}) + \underline{w_k}^T \lambda_\mathbf{2} \underline{w_k} \tag{3.58}$$

with the input $\underline{w}$ and the regularization matrix $\lambda_\mathbf{2}$ which provides the flexibility of weighting the input term. The matrices $\lambda_\mathbf{1}$ and $\lambda_\mathbf{2}$ are symmetric and positive definite[17] matrices.

Similarly to the RLSS from section 3.6 the objective is to minimize the least-squares error to obtain a solution of the inverse problem. The definition of the minimization of $E$

[17] A symmetric square matrix $\mathbf{A}$ is positive definite if all of its eigenvalues are greater than zero.

for any initial state $\underline{x}$ is given by:

$$\mathbf{W_k}(\underline{x}) = \min_{\underline{w_k}} E_k(\underline{x}, \underline{w_k}) \ . \tag{3.59}$$

The application of the so-called *Bellman principle of optimality*[18] is leading to a recursive expression which selects $\underline{w_{k-1}}$ to minimize $E_k$:

$$\begin{aligned} W_{k-1}(\underline{x}) = \min_{\underline{w_{k-1}}} [&(\mathbf{C_d}\underline{x_{k-1}} - \underline{\hat{y}_{k-1}})^T \lambda_\mathbf{1} (\mathbf{C_d}\underline{x_{k-1}} - \underline{\hat{y}_{k-1}}) + \underline{w_{k-1}}^T \lambda_\mathbf{2} \underline{w_{k-1}} \\ &+ \mathbf{W_k}(\mathbf{A_d}\underline{x_{k-1}} + \mathbf{G_d}\underline{w_{k-1}})] \ . \end{aligned} \tag{3.60}$$

By starting at the end of the process where $k = n$ and working backward towards $k = 1$ a solution can be obtained. At the end point the minimum is derived by choosing $\underline{w_n} = \underline{0}$ which simplifies equation (3.60) as the following:

$$\begin{aligned} W_n(\underline{x}) &= \min_{\underline{w_n}} [(\mathbf{C_d}\underline{x_n} - \underline{\hat{y}_n})^T \lambda_\mathbf{1} (\mathbf{C_d}\underline{x_n} - \underline{\hat{y}_n})] \\ &= \underline{x_n}^T (\underbrace{\mathbf{C_d}^T \lambda_\mathbf{1} \mathbf{C_d}}_{:=\mathbf{R_n}} \underline{x_n}) + \underline{x_n}^T \underbrace{(-2\, \mathbf{C_d}^T \lambda_\mathbf{1} \underline{\hat{y}_n})}_{:=\mathbf{S_n}} + \underbrace{\underline{\hat{y}_n} \lambda_\mathbf{1} \underline{\hat{y}_n}}_{:=q_n} \\ &= \underline{x_n}^T \mathbf{R_n} \underline{x_n} + \underline{x_n}^T \mathbf{S_n} + q_n \ . \end{aligned} \tag{3.61}$$

Since equation (3.61) shows that $W_n$ is quadratic in $\underline{x_n}$ all of the $W_k$ have to be quadratic in $\underline{x_k}$ which leads to:

$$W_k(\underline{x}) = \underline{x_k}^T \mathbf{R_k} \underline{x_k} + \underline{x_k}^T \mathbf{S_k} + q_k \tag{3.62}$$

with $\mathbf{R_k} = \mathbf{R_n} = \mathbf{C_d}^T \lambda_\mathbf{1} \mathbf{C_d}$ and $\mathbf{S_k} = -2\, \mathbf{C_d}^T \lambda_\mathbf{1} \underline{\hat{y}_k}$. This expression can be substituted into equation (3.60) and becomes the following:

$$\begin{aligned} W_{k-1} = \min_{\underline{w_{k-1}}} [&(\mathbf{C_d}\underline{x_{k-1}} - \underline{\hat{y}_{k-1}})^T \lambda_\mathbf{1} (\mathbf{C_d}\underline{x_{k-1}} - \underline{\hat{y}_{k-1}}) + \underline{w_{k-1}}^T \lambda_\mathbf{2} \underline{w_{k-1}} \\ &+ (\underline{x_k}^T \mathbf{R_k} \underline{x_k} + \underline{x_k}^T \mathbf{S_k} + q_k)(\mathbf{A_d}\underline{x_{k-1}} + \mathbf{G_d}\underline{w_{k-1}})] \ . \end{aligned} \tag{3.63}$$

The next step is the minimization of equation (3.63) by setting the derivative $\frac{dW_{k-1}}{dw_{k-1}}$ equal to zero. This operation returns the following:

$$(2\lambda_\mathbf{2} + 2\, \mathbf{G_d}^T \mathbf{R_k} \mathbf{G_d}) \underline{\hat{w}_{k-1}} = -\mathbf{G_d}^T \mathbf{S_k} - 2\, \mathbf{G_d}^T \mathbf{R_k} \mathbf{A_d} \underline{\hat{x}_{k-1}} \tag{3.64}$$

where $\underline{\hat{w}}$ is the estimated, or optimal, input term and $\underline{\hat{x}}$ represents an estimated state. For reasons of simplification the following matrices are introduced:

$$\mathbf{V_k} = (2\lambda_\mathbf{2} + 2\, \mathbf{G_d}^T \mathbf{R_k} \mathbf{G_d})^{-1} \tag{3.65}$$

$$\mathbf{H} = 2\, \mathbf{G_d}^T \mathbf{R_k} \ . \tag{3.66}$$

[18] The optimization principle is a dynamic programming method named after R. Bellman and described in [4].

After rearranging equation (3.64) and substitution of the matrices from above a recursive expression for estimation of the unknown input can be derived and is given by:

$$\underline{\hat{w}_{k-1}} = -\mathbf{V_k}\mathbf{G_d}^T\mathbf{S_k} - \mathbf{V_k}\mathbf{H}\mathbf{A_d}\underline{\hat{x}_{k-1}} \ . \tag{3.67}$$

At the beginning of the estimation an initialization has to be made by choosing the initial state $\underline{\hat{x}_0}$ and computation of the first estimate $\underline{\hat{w}_0}$ according to equation (3.67). The other estimates of $\underline{\hat{x}}$ can then be calculated for each time step $k = 1, \ldots, n$ using the input estimates $\underline{\hat{w}}$ from equation (3.67) as follows:

$$\underline{\hat{x}_k} = \mathbf{A_d}\underline{\hat{x}_{k-1}} + \mathbf{G_d}\underline{\hat{w}_{k-1}} \ . \tag{3.68}$$

After performing the recursive algorithm for every time step an estimate $\underline{\hat{w}}$ of the unknown input $\underline{w}$ is obtained. The introduced procedure of the inverse dynamics method is applied for the two-mass system in the next section.

### 3.7.2 Inverse dynamics method for the two-mass system

The application of the inverse dynamics method to the two-mass system presented in section 3.3 is based on the recursive computation of equations (3.67) and (3.68). The state-space model of the here regarded system is given by equations (3.23) and (3.24) with the unknown excitation $w$ and the known output data $y$.

Since $y$ and $w$ are scalar quantities, the inverse dynamics algorithm for this system contains of scalar weighting terms $\lambda_1$ and $\lambda_2$. Before the estimation process can be started, an initialization has to be performed by choosing a value for the initial state $\underline{\hat{x}_0}$. Now the algorithm can be implemented and computed for each time step $k = 1, \ldots, n$ as the following:

$$\hat{w}_{k-1} = -V_k\mathbf{G_d}^T\mathbf{S_k} - V_k\mathbf{H}\mathbf{A_d}\underline{\hat{x}_{k-1}} \tag{3.69}$$

$$\underline{\hat{x}_k} = \mathbf{A_d}\underline{\hat{x}_{k-1}} + \mathbf{G_d}\hat{w}_{k-1} \tag{3.70}$$

where the matrices, which were introduced in equations (3.65) and (3.66), are given according to the inverse-problem of the two-mass system by:

$$V_k = (2\lambda_2 + 2\,\mathbf{G_d}^T\mathbf{R_k}\mathbf{G_d})^{-1} \tag{3.71}$$

$$\mathbf{H} = 2\,\mathbf{G_d}^T\mathbf{R_k} \tag{3.72}$$

with $\mathbf{R_k} = \mathbf{C_d}^T\lambda_1\mathbf{C_d}$ and $\mathbf{S_k} = -2\,\mathbf{C_d}^T\lambda_1\hat{y}_k$.

The simulation results are discussed in the next section and compared with the estimated inputs which were obtained with other inverse identification methods.

## 3.8 Results for the two-mass system

The application of the methods from above yields different results for the unknown input estimation for the one-dimensional two-mass system, which was introduced in section 3.3 and is shown in Figure 3.1.

For the given inverse problem output data representing the acceleration $\ddot{x}_1$ of mass $m_1$ is needed. The data is obtained by measuring the response of the system to a generated excitation, which is assumed to be unknown for further estimations. Consequently, the generation of the input signal has to be investigated before the results can be compared and discussed.

### 3.8.1 Generation of a random input signal

For the later introduced railway vehicle system unknown input forces due to random track irregularities have to be estimated. Referring to that a signal with similar stochastic characteristics is applied to the excitation of the two-mass system.

The signal can be generated as described in [6] by summation of sinusoids with $k$ different frequencies as the following:

$$w(t) = \sum_{i=1}^{k} A(\omega_i) \cdot \sin(\omega_i t + \varphi_i) \tag{3.73}$$

with the frequencies $\omega_i = 2\pi f_i$, the amplitudes $A(\omega_i) = \sqrt{S(\omega_i)\Delta\omega} \cdot \sqrt{2}$ and the phase shifts $\varphi_i = 2\pi r_i$, where $r_i$ is a random number[19] between 0 and 1. The amplitude depends on $S(\omega_i)$, the power spectral density (PSD) of the corresponding frequency $\omega_i$.

For white noise the PSD is constant, $S(\omega_i) = R_0 = \sigma^2$ (see section 2.2.2, equations (2.54) and (2.56)), which results for every frequency $\omega_i$ in equal amplitudes $A(\omega_i) = \bar{A} = \sqrt{\sigma^2\Delta\omega} \cdot \sqrt{2}$.

Figure 3.5a shows a generated signal with a frequency range from 0.5 Hz to 10 Hz. The signal is normal or Gaussian distributed with zero-mean and has a variance of $\sigma^2 = 2 \cdot 10^{-6}\,\mathrm{m}^2$. From the PSD analysis[20], shown in Figure 3.5b, it can be seen that the signal has similar characteristics to white noise, meaning that the PSD is constant over the frequency range, which allows the use of the term *pseudorandom signal*.

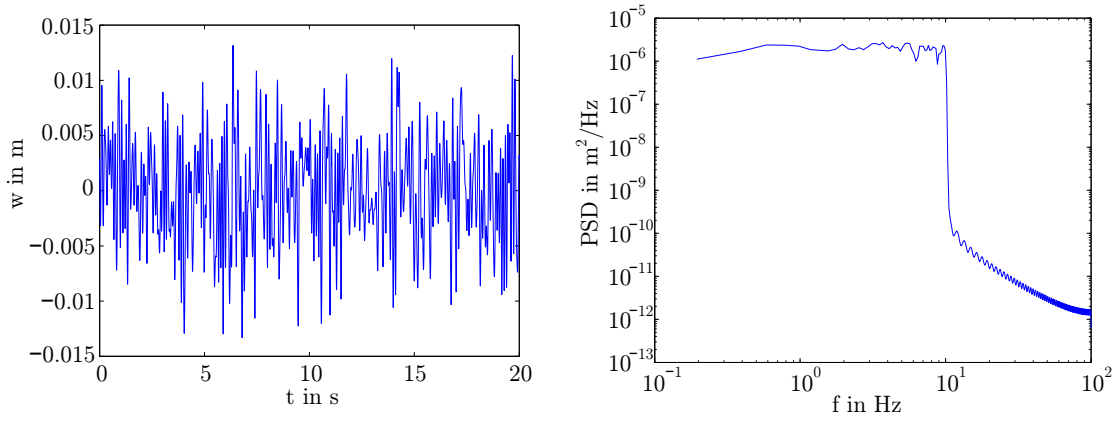
The signal shown in Figure 3.5a is applied to the two-mass system and results in a response which is used for the inverse estimation method. The results of the input which was estimated through application of inverse methods and the generated input from above are discussed in section 3.8.3

### 3.8.2 Model and method parameters

For the input estimation of the two-mass system introduced in section 3.3 model parameters have to be defined. The later presented results of the inverse problem solution have been obtained with the values shown in Table 3.1.

[19] The random number $r_i$ can be generated by the MATLAB function `rand`: *"rand(n) returns an n-by-n matrix containing pseudorandom values drawn from the standard uniform distribution on the open interval (0,1)."* [20]

[20] Here a special form of the PSD function is computed using the algorithm of Peter D. Welch. [36] This method uses a fast fourier transform for estimation of the power spectra and divides the signal into overlapped sections. After computation of modified periodograms and averaging them the estimate of the PSD is obtained. The MATLAB function `pwelch` returns the PSD on the basis of this algorithm. [20].




(a) Generated input signal

(b) PSD of generated input signal

Figure 3.5: Generated pseudorandom input signal

In addition the methods presented in sections 3.4 - 3.7 contain different parameters that can be varied. Table 3.2 gives an overview of the here chosen values.

It has to be mentioned at this point that no further investigations in terms of parameter optimization have been made as the aim of this task was to compare the implementation and appropriateness of the single methods. The comparison is given in the next section.

| Parameter | Value | Unit |
|---|---|---|
| $m_1$ | 1 | kg |
| $m_2$ | 1 | kg |
| $c_1$ | 30 | N/m |
| $c_2$ | 60 | N/m |
| $d_1$ | 0.5 | Ns/m |

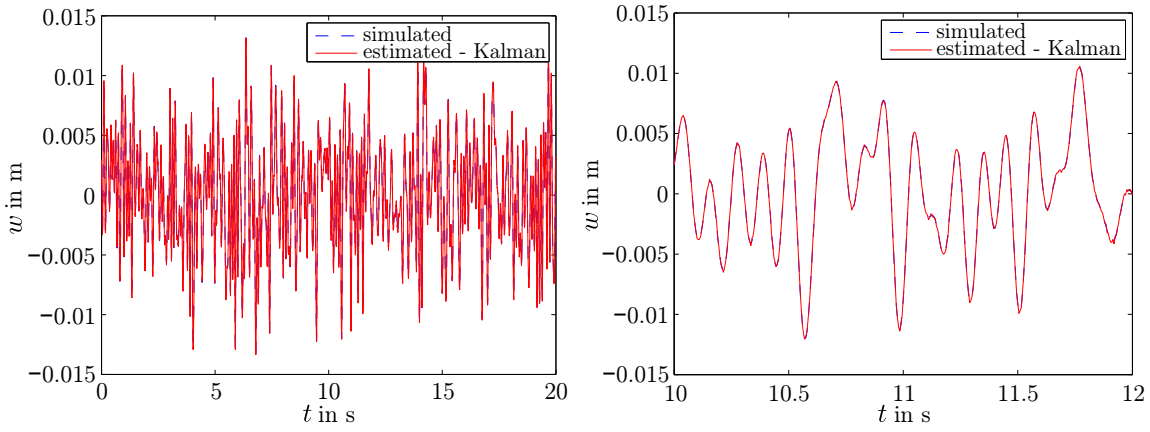
Table 3.1: Model parameters of the two-mass system

| Method | Parameter | Value |
|---|---|---|
| Kalman filter | $\mathbf{Q} = Q$ | 1 |
| | $\mathbf{R} = R$ | 0.0001 |
| Parity space approach | $L$ | 20 |
| RLSS | $\lambda$ | 5 |
| | $\mathbf{R_r}$ | $\begin{bmatrix} 3 & -1 & 0 & 0 \\ -3 & 3 & -1 & 0 \\ 1 & -3 & 3 & -1 \\ 0 & 1 & -3 & 3 \end{bmatrix}$ |
| Inverse dynamics method | $\lambda_1$ | 1 |
| | $\lambda_2$ | 1 |

Table 3.2: Parameters for the different estimation methods

### 3.8.3 Discussion of results

The first method presented in section 3.4 is using a linear Kalman filter algorithm for state estimation. The result of the identified input is given in Figures 3.6a and 3.6b where significant correlation can be detected.

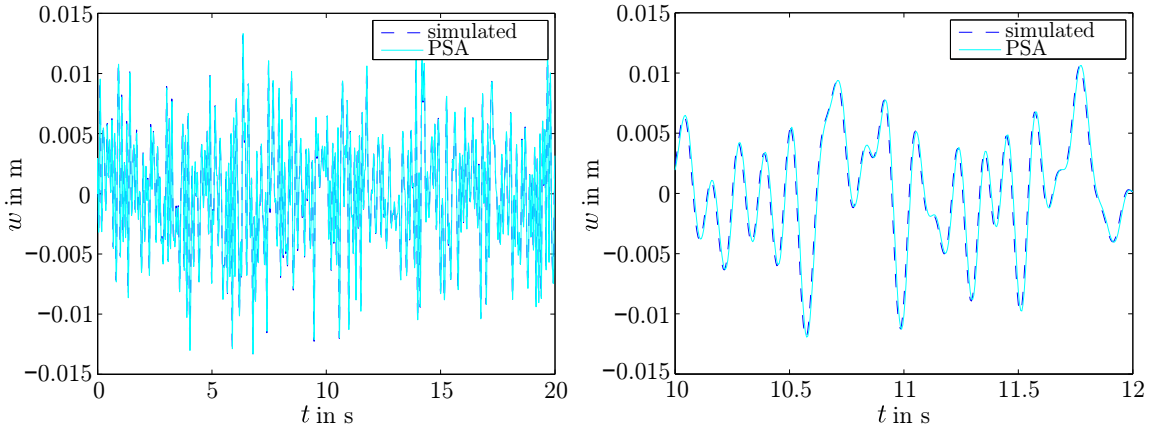



(a) Kalman filter estimation

(b) Kalman filter estimation, enlarged

Figure 3.6: Results for the Kalman filter method

The parity space approach, introduced in section 3.5, also provides good estimation of the unknown input as can be seen in Figures 3.7a and 3.7b.



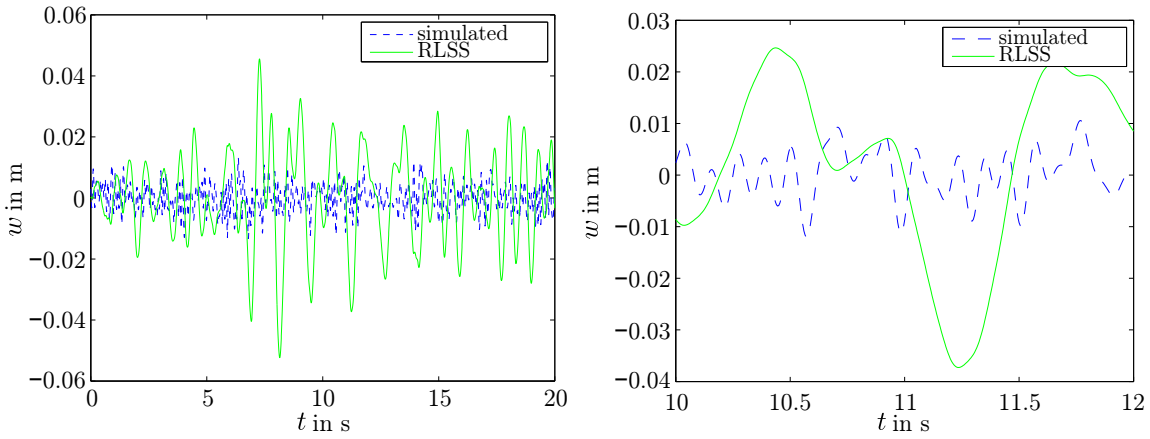

(a) PSA estimation

(b) PSA estimation, enlarged

Figure 3.7: Results for the parity space approach (PSA) method

The results of the regularized least squares solution (RLSS) according to section 3.6 of the given inverse problem are shown in Figures 3.8a and 3.8b. It can be seen that the estimation does not match with the simulated signal. Especially the amplitudes are to high but also higher frequency components are not identified with this method. The regularization coefficient $\lambda$ and the regularization matrix $\mathbf{R_r}$ are not chosen adequately enough for the derivation of correlating results. An optimization of the parameters may could result in a higher correlation. However, since another method already shows satisfying results and will be chosen for further estimations within this work, an optimization is not performed here.

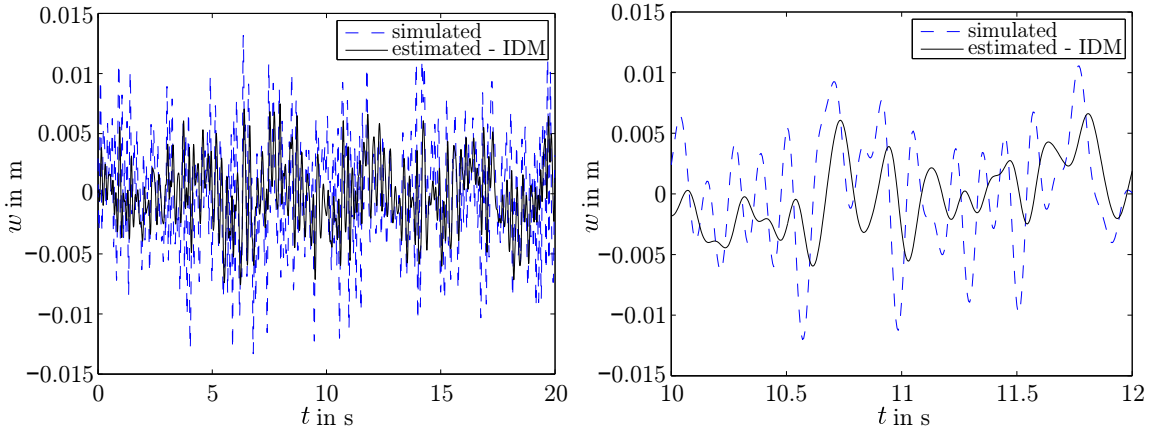
Better results can be obtained with the inverse dynamics method (IDM) which was introduced in section 3.7. The diagrams in Figures 3.9a and 3.9b show the comparison of the simulated and the estimated signal where a notable correlation can be detected. Similar




(a) Estimation of RLSS (b) Estimation of RLSS, enlarged

Figure 3.8: Regularized least squares solution (RLSS) results

to above, the estimation result might be improved if the weighting and regularization parameters are chosen optimally.



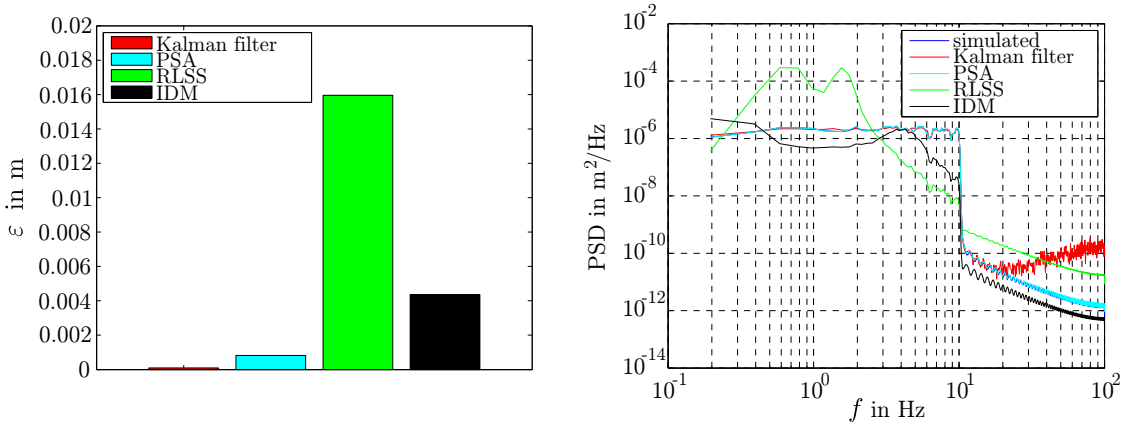

(a) Inverse dynamics method estimation (b) Inverse dynamics method estimation, enlarged

Figure 3.9: Results for the inverse dynamics method (IDM)

To quantify the deviation of the estimated signals to the simulated input the root-mean-square error can be calculated as the following:

$$\varepsilon = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(w_i - \hat{w}_i)^2}\ . \tag{3.74}$$

Figure 3.10a compares the root-mean-square errors of estimated results and shows that the Kalman filtering method returns the best results followed by the parity space approach. The highest error can be detected for the RLSS. It has to be mentioned at this point, that this error measure does not take phase shifts into account. Hence, a comparison of estimated and simulated signals in frequency domain should be considered additionally to be able to make a reliable assessment.




(a) Root-mean-square error errors

(b) Power spectral densities

Figure 3.10: Comparison of inverse estimation methods

A comparison of the correlations in frequency domain is given by Figure 3.10b where the PSDs of the single methods are illustrated. It can be detected that the spectral characteristics of the input signals estimated through the Kalman filter algorithm and the parity space approach are almost equal to the PSD of the simulated signal. As well as in time domain the RLSS result shows the highest deviation in frequency domain. The PSD for the signal obtained by the inverse dynamics method indicates that the amplitudes for frequencies higher than $\sim 0.5\,\mathrm{Hz}$ are lower than for the simulated signal.

Concluding the investigations from above it can be said that the Kalman filter method provides the best estimation results. With the parity space approach almost as good estimates have been achieved which confirms the suitability of this method for inverse problem solving tasks. Further optimization of the regularization parameters for the inverse dynamics method and the RLSS is required to obtain satisfying and stable results.

Since the unknown input estimation for a simple model can be performed successfully - at least for two methods - the application of an inverse method to a more complex model, which is introduced in the following chapter, can be attempted. In this work the method using a Kalman filter for state estimation is selected to identify unknown wheel-rail contact forces of a railway vehicle system.

# 4 Multibody simulation of a railway vehicle system

This chapter reviews the general procedure of multi-body simulation applied to the complex system of a sample railway vehicle. At the beginning, the mathematical modeling in terms of mechanics and the wheel-rail contact theory is presented. Since a railway vehicle system is excited by irregularities of the rail track the second part of the chapter introduces track characteristics and the concept of input force implementation.

## 4.1 Modeling of a railway vehicle system

The modeling task starts with a definition of the mechanical model where the parts and components of the railway vehicle are characterized and is described in [17], [22] and [28]. This is followed by a mathematical description to be able to state the equations of motion which are linearized in the next step. In section 4.1.4 a major problem in railway vehicle dynamics, the correct modeling of the wheel-rail contact geometry and the occurring contact forces, is introduced.

### 4.1.1 Mechanical Model

A railway vehicle system is set up by several components and parts. For execution of a multibody simulation the system can be reduced considering the components with influence to the general dynamic behavior only. Figure 4.1 shows the bodies of the regarded railway vehicle system which are linked by joints and kinematic constraints that allow certain relative motions and restrict others.

The *carbody* is assumed to be a single body and is attached to two *bogies*. A bogie connects the carbody with the rails and has to fulfill several functions as it has to provide stability of the train on straight and curved tracks at any speed. The assembly of an axle and two wheels is called *wheelset* and every bogie contains two of them.[21] Depending on the type, a wheelset usually carries brake discs and/or a gear box, what is neglected within this work. The wheelsets are mounted on the *bogie frame* suspended by several spring and damping elements. Similarly, the frame is attached spring-loaded to the so-called *bogie bolster* which is connected to the carbody. The *primary suspension* is defined as the connection between the wheelsets and the bogie frame whereas the connection between bolsters and bogie frames is the *secondary suspension*. The bodies of a railway vehicle system can be regarded either as rigid or flexible if the modal deformation of bodies is

[21] This is the case for the regarded railway vehicle system. In reality, also bogies that contain three wheelsets exist.

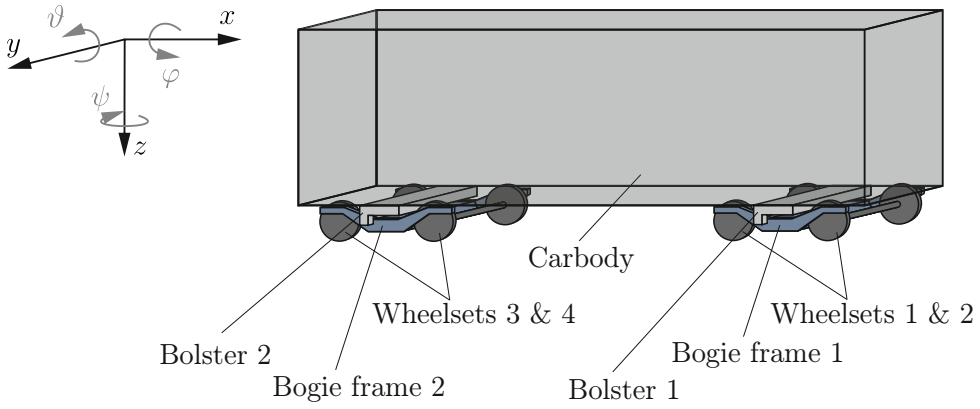



Figure 4.1: Railway vehicle system with 2 bogies

considered. In this work all bodies are assumed to be rigid, since structural dynamics is of minor importance to the investigated problem.

In railway vehicle dynamics the coordinate system is usually chosen as it is illustrated in Figure 4.1. The driving direction is along the *longitudinal* axis $x$ while the direction along the $y$-axis is called the *lateral* direction. Accordingly, the *vertical* direction is along the $z$-axes which is positive counted downwards in general. The rotations about the Cartesian coordinates $x, y$ and $z$ are described by the angles $\varphi, \vartheta$ and $\psi$ characterizing a *roll*-motion ($\varphi$) about $x$, a *pitch*-motion ($\vartheta$) about $y$ and a *yaw*-motion ($\psi$) about $z$.

In general, time-dependent properties of a simulation model are called states. Considering a dynamical system (see chapter 2.1.2) its mechanical motions can be described with knowledge about the states and their change in time. The states are defined measures to describe the kinematical behavior of the bodies and represent the so-called *generalized coordinates* of the body movements which will be used later to formulate the equations of motion.

As the bodies of a multibody system may not move freely due to geometric properties, so-called *geometric constraints* are introduced according to [15] in order to restrain the abilities of movement. Geometric constraints are taken into account for the simulation task as so-called *algebraic states* and lead to *differential-algebraic equations* (DAE)[22].

Regarding constrained dynamic systems the term *degrees of freedom* (DOF) has to be introduced describing the number of independent coordinates which are required to describe the configuration of a system. In the three-dimensional case the total number of DOFs $n_{DOF}$ of a model or system can be obtained by:

$$n_{DOF} = 6 \cdot n_b - n_c \tag{4.1}$$

and depends on the number of bodies $n_b$ and on the number of algebraic constraints

[22] A differential-algebraic set of equations is a mixed system of differential and algebraic equations and usually can not be solved with an ordinary numerical solving algorithm for ordinary differential equations (ODE).

$n_c$. Since the motion of a body can be described by six independent coordinates (three translations and three rotations) the number of DOFs of a unconstrained moving body is six. Each algebraic constraint removes one DOF of the system.

With the given definitions of the system components and state properties from above the equations of motion can be formulated. This is presented in the following section.

### 4.1.2 Equations of motion

As an appropriate mechanical model has been defined, the next task of the multibody simulation process can be performed by formulating the equations of motion of the railway vehicle system.

Before a general formulation of the equations of motion can be given, the state definitions used by the multibody simulation software SIMPACK are introduced, since the data used within this work base on a SIMPACK simulation and the obtained results are compared with it later. The different types of states according to [30] are the following[23]:

- *Joint states* $\underline{s_j}$, $\underline{\dot{s}_j}$ describe the kinematical behavior of bodies and represent the generalized coordinates of the bodies.
- *Flexible body states* $\underline{s_{fb}}$, $\underline{\dot{s}_{fb}}$ have to be considered if the modal deformation of bodies are allowed. If all bodies are assumed to be rigid these states are neglected.
- *Dynamic states* $\underline{s_{dyn}}$ are used to describe the internal behavior of particular elements. Considering two bodies connected by a spring and a damper in series e.g. (see Figure 4.2) it can be seen that the length of the spring/damper cannot described by the joint states of the two bodies ($s_{j_1}$ and $s_{j_2}$). Therefore a third state ($s_{dyn}$) has to be defined for the exact description of the spring/damper length.
- *Algebraic states* $\underline{s_{alg}}$ are either used to describe the dynamic behavior of elements (e.g. force values) or the non-dynamic behavior of elements (e.g. contact point positions of the wheel-rail contact).
- *Constraint states* $\underline{s_\lambda}$ are special kinds of algebraic states containing forces and torques which are applied due to constraints in directions of the restrained kinematic DOFs (e.g. contact forces in the wheel-rail contact point).

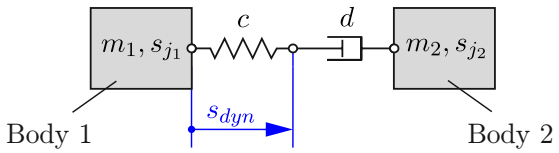



Figure 4.2: Serial spring-damper element

The formulation of the dynamic equations of motion requires a definition of the generalized coordinate system, that is the coordinate system where the generalized coordinates

[23] Note that the nomenclature differs from the original definitions which can be found in [30].

are referred to. In developing general-purpose multibody system algorithms the so-called *absolute Cartesian coordinates* with a fixed origin are widely used. However, in railway dynamics it might be advantageous to use other sets of coordinates in order to derive the equations of motion.

Figure 4.3 shows an alternative coordinate system which is moving along a specified trajectory $g$ that is e.g. following the motion of a body. The displacement of a body can then be uniquely described using the six so-called *trajectory coordinates*:

- the arc length coordinate $s$,
- the lateral displacement $y_t$ relative to the trajectory,
- the vertical displacement $z_t$ relative to the trajectory,
- the roll-angle $\varphi_t$ about $x_t$,
- the pitch-angle $\vartheta_t$ about $y_t$ and
- the yaw-angle $\psi_t$ about $z_t$.

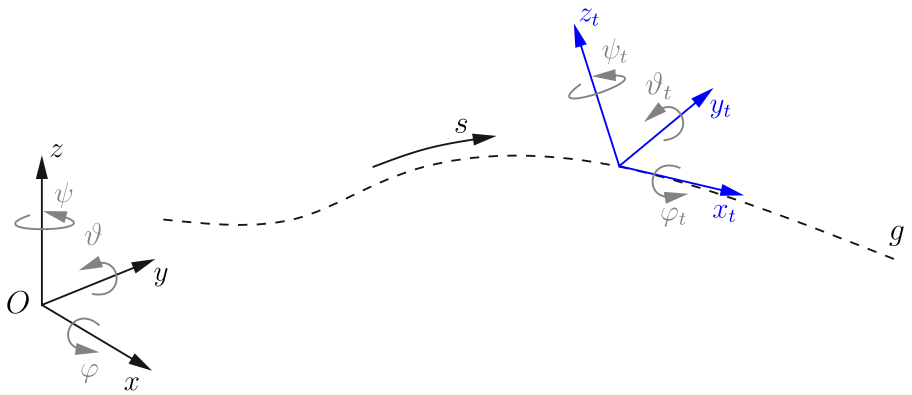



Figure 4.3: Absolute and trajectory coordinates

With a proper coordinate transformation the relationship between any two sets of coordinates can always be achieved. This makes clear that every coordinate system can be used to formulate the kinematic and dynamic equations needed for the multibody simulation process. An advantage of the use of trajectory coordinates is the simpler formulation of some railway vehicle constraints and of some forcing functions. However, it can make the implementation in general-purpose multibody system algorithms more difficult.

Constraint functions $C_i \in \{C_1, \ldots, C_{n_c}\}$ describe the mechanical joints or specified motion trajectories. They build a set of $n_c$ algebraic equations which is added to the *nonlinear Newton-Euler equations of motion* to obtain a complete mathematical description of the dynamical system. The nonlinear DAE set can be written as follows:

$$\underline{C}(\underline{q}, t) = \underline{0} \tag{4.2}$$

$$\mathbf{M_q}\ddot{\underline{q}} = \underline{F_q} - \mathbf{C_q}^T \underline{\lambda} \tag{4.3}$$

where $\underline{C}$ is the vector of the constraint functions, $\underline{q}$ is the vector of the generalized coordinates referring to the chosen coordinate system (and its second time derivative $\underline{\ddot{q}}$), $\mathbf{M_q}$ is system mass matrix, $\underline{F_q}$ is the vector of external forces and inertia forces which are quadratic in velocity (that are centrifugal and Coriolis forces), $\mathbf{C_q}$ is the Jacobian matrix of the constraint equations and $\underline{\lambda}$ is the vector containing the so-called *Lagrange multipliers*. The index $\mathbf{q}$ (or $q$) denotes that a proper formulation of the matrices and vectors according to the generalized coordinate system has to be used. Forces invoked by constraints are given by the term $-\mathbf{C_q}^T\underline{\lambda}$ in equation (4.3) and are e.g. contact forces in the wheel-rail contact point.

For computational purpose a matrix form of the nonlinear equations of motion is aspired. Therefore the second time derivative of the constraint functions, given by equation (4.2), has to be introduced as follows:

$$\underline{\ddot{C}}(\underline{q},t) = \mathbf{C_q}\underline{\ddot{q}} - \underline{Q} = \underline{0} \tag{4.4}$$

where a vector $\underline{Q}$ can be derived. The nonlinear equations of motion from above can then be written in matrix form as follows:

$$\begin{bmatrix} \mathbf{M_q} & \mathbf{C_q}^T \\ \mathbf{C_q} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \underline{\ddot{q}} \\ \underline{\lambda} \end{bmatrix} = \begin{bmatrix} \underline{F_q} \\ \underline{Q} \end{bmatrix} \tag{4.5}$$

with

$$\mathbf{M_q} = \begin{bmatrix} \mathbf{M_{q_1}} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{M_{q_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{M_{q_{n_b}}} \end{bmatrix}, \quad \mathbf{C_q} = \begin{bmatrix} \frac{\partial C_1}{\partial \underline{q_1}} & \dots & \frac{\partial C_1}{\partial \underline{q_{n_b}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial C_{n_c}}{\partial \underline{q_1}} & \dots & \frac{\partial C_{n_c}}{\partial \underline{q_{n_b}}} \end{bmatrix},$$

$$\underline{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_{n_c} \end{bmatrix}, \quad \underline{q} = \begin{bmatrix} \underline{q_1} \\ \underline{q_2} \\ \vdots \\ \underline{q_{n_b}} \end{bmatrix}, \quad \underline{\lambda} = \begin{bmatrix} \underline{\lambda_1} \\ \underline{\lambda_2} \\ \vdots \\ \underline{\lambda_{n_c}} \end{bmatrix}, \quad \underline{F_q} = \begin{bmatrix} \underline{F_{q_1}} \\ \underline{F_{q_2}} \\ \vdots \\ \underline{F_{q_{n_b}}} \end{bmatrix}, \quad \underline{Q} = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_{n_c} \end{bmatrix}. \tag{4.6}$$

Recalling the state definitions used in SIMPACK the generalized coordinates vector $\underline{q}$ is composed by the joint states $\underline{s_j}$, the flexible body states $\underline{s_{fb}}$ and the dynamic states $\underline{s_{dyn}}$. Accordingly, the vector $\underline{\lambda}$ consists of the algebraic states $\underline{s_{alg}}$ and constraint forces and torques $\underline{s_\lambda}$, which are described by the constraint functions $\underline{C_i}$.

### 4.1.3 Linearization and state-space formulation

A linearization of the nonlinear equations of motion is performed by removing all of the constraints and replacing them by spring and damper elements. If all bodies are assumed to be rigid no flexible body states have to be defined. The well-known formulation of the *linear Newton-Euler equations of motion* is given by the following differential equation of second-order:

$$\mathbf{M}\underline{\ddot{q}}(t) + \mathbf{K_1}\underline{\dot{q}}(t) + \mathbf{K_2}\underline{q}(t) = \mathbf{H}\underline{f}(t) = \underline{F}(t) \tag{4.7}$$

where the vector $\underline{q}(t) = (q_1(t), \ldots, q_n(t))^T$ represents $n$ generalized coordinates, $\mathbf{M}$ is the mass matrix, $\mathbf{K_1}$ is the damping matrix, $\mathbf{K_2}$ is the stiffness matrix and the term $\mathbf{H}\underline{f}(t)$ results in a vector $\underline{F}(t)$ containing external forces. As described above, the matrices and vectors have to be defined with reference to the generalized coordinate system, that means $\mathbf{M} = \mathbf{M_q}$ and so on. For reasons of simplification the indices $\mathbf{q}$ and $q$ are suppressed here and in further derivations.

To obtain a state-space representation of the dynamical system as a set of differential equations of first-order the state vector $\underline{x}$ has to be introduced. Before that, equation (4.7) is rearranged as follows:

$$\underline{\ddot{q}}(t) = -\mathbf{M^{-1}K_1}\underline{\dot{q}}(t) - \mathbf{M^{-1}K_1}\underline{q}(t) + \mathbf{M^{-1}H}\underline{f}(t). \tag{4.8}$$

The conversion of equation (4.8) into the state-space can be achieved with $\underline{x} = [\underline{q}\ \underline{\dot{q}}]^T$, $\underline{\dot{x}} = [\underline{\dot{q}}\ \underline{\ddot{q}}]^T$ and $\underline{u} = \underline{f}$ yielding the well known representation for a linear system (see section 2.1.2):

$$\underline{\dot{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \tag{4.9}$$

$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \tag{4.10}$$

where $\underline{x}$ is the state vector, $\underline{u}$ is the input vector containing known input quantities, $\underline{y}$ is the output vector and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$ are the corresponding system matrices. Equation (4.9) represents a set of ordinary differential equations of first-order for which a solution usually can be found with ordinary numerical solving algorithms.

The matrices $\mathbf{A}$ and $\mathbf{B}$ can be obtained directly from equation (4.8) and are given by:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M^{-1}K_2} & -\mathbf{M^{-1}K_1} \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{M^{-1}H} \end{bmatrix}. \tag{4.11}$$

By contrast, the system output matrix $\mathbf{C}$ and the feedthrough matrix $\mathbf{D}$ are unknown since the output vector $\underline{y}$ have not been defined at this point.

According to the state definitions from SIMPACK the state vector $\underline{x}$ of the linearized dynamical system is composed as follows:

$$\underline{x}(t) = \begin{bmatrix} \underline{s_j} \\ \underline{s_{dyn}} \\ \underline{\dot{s}_j} \end{bmatrix}, \qquad \underline{\dot{x}}(t) = \begin{bmatrix} \underline{\dot{s}_j} \\ \underline{\dot{s}_{dyn}} \\ \underline{\ddot{s}_j} \end{bmatrix} \tag{4.12}$$

where $\underline{s_j}$ are the joint states and $\underline{s_{dyn}}$ are the dynamical states (see section 4.1.2). It can be seen from above that the dynamical states influence the system's behavior only in position and velocity. This is because the elements using dynamical states are defined to be without mass and therefore their inertia forces become zero what decreases the order of their equations of motion by one.

### 4.1.4 Modeling of the wheel-rail contact

The interaction of vehicle and rail is an important element that distinguishes railway vehicles from other multibody system applications. The supporting and guiding forces of

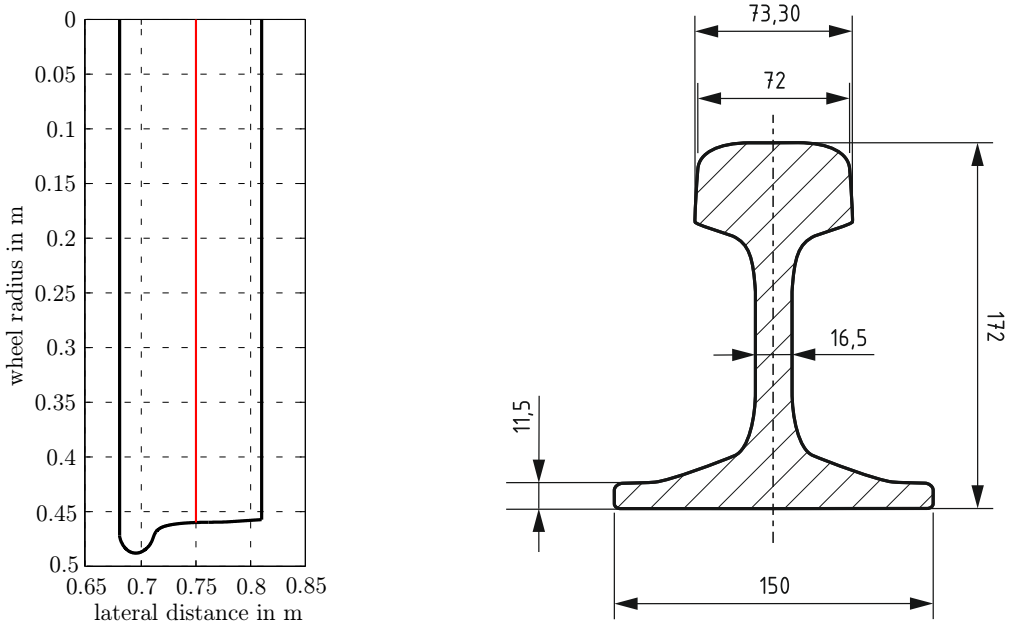
a railway vehicle must be transmitted through the contact patches between the wheels and the rails.

Before the forces can be modeled, a description of the wheel and rail profile geometries has to be introduced to be able to determine the locations of the contact points. The formulation of the contact patch geometry as well as the force estimation is a complex field in railway vehicle dynamics. This chapter gives a brief overview of the implemented approaches.

### Profile geometry

The geometries of the wheel profiles and the rail profiles have significant influences on the driving stability of a vehicle, the traveling comfort and the wear of the single parts.

As described in [17], the most common profile pair used in central Europe is the UIC60E1 profile for the rails and the S1002 profile for the wheels, see Figure 4.4.




(a) Wheel profile S1002

(b) Rail profile UIC60E1

Figure 4.4: Wheel and rail profile geometries

It can be seen in figure 4.4a that the wheel profile is conical which supports the guidance of the vehicle due to its kinematics as the following: If an unrestrained wheelset with conical profiles moves laterally from the center of a track the outer wheel is rolling on a larger radius than the inner one. This results in a yaw motion about the vertical axis as the wheels have same rotational speed and as a result the wheelset will then naturally tend back to the center of the track. Considering a curve the rolling radius difference then matches the yaw velocity needed for the curve, since the wheelset will move outwards. This leads to the introduction of the term *conicity* which is the difference in rolling radii between the wheels over a given lateral shift $y$ of the wheelset.

Since most profiles are not purely conical, either by design or through wear in service, the term *equivalent conicity* is applied and defined in EN 15302 [2] as follows:
*"By definition, the equivalent conicity is equal to the tangent of the cone angle of a wheelset with coned wheels whose lateral movement has the same kinematic wavelength as the given wheelset (but only on tangent track and on very large-radius curves)."*

According to [8] the wheelset will in fact tend to overshoot its equilibrium position due to the conicity which effects the so-called *kinematic oscillation*. Klingel observed in [16] a sinusoidal motion and described the angular frequency $\omega$ by

$$\omega = v\sqrt{\frac{\lambda_c}{r_0 b}} \tag{4.13}$$

where $v$ is the velocity of the wheelset, $\lambda_c$ is the equivalent conicity, $r_0$ is the wheel radius when there is no lateral displacement and $b$ is half the track gauge. The formula makes clear that the greater the conicity the greater the frequency. Due to that, the oscillation can cause instability in the motion of the wheelset if the so-called *critical speed* is reached. The instability is known also known as *hunting* and is limited by flange[24] contacts but can lead to damage of track.

In most countries the rails are canted inwards in order to match the conicity of the wheel to direct the normal force along the web of the rail when the wheelset is in central position. The cant angle depends on the railway organization and in Germany and Austria e.g. the rails are canted with a ratio of 1:40.

The wheel profile S1002, as shown in figure 4.4a, is a so-called *worn profile* since it is derived from measurements of worn profiles.

With knowledge about the geometric characteristics the next task to be accomplished in terms of wheel-rail contact force simulation is given by the contact patch description.

### Contact geometry and contact forces

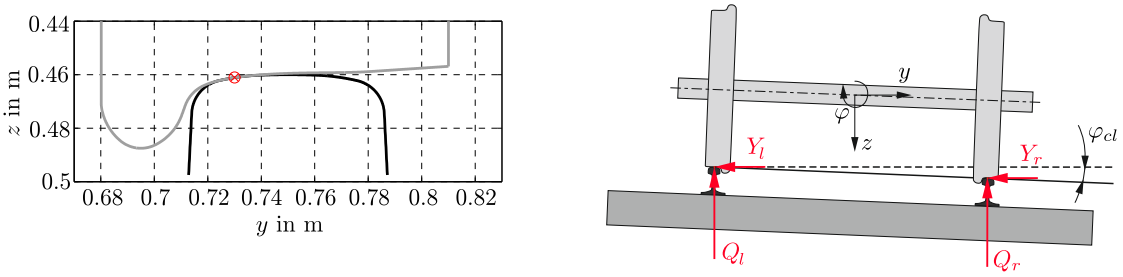
One of the fundamental problems that must be addressed in railway vehicle dynamics is the formulation of the contact forces describing the interaction between the wheels and the rails. Therefore, a formulation of the wheel-rail contact geometry has to be introduced. A sample geometry configuration of the contact between a wheel profile (S1002) and a rail profile (UIC60E1) is shown in Figure 4.5a.

In the contact search process the identification of location, size and shape of the contact patch is the first step. In SIMPACK the result of the contact search is the number of contact patches and their locations which are converted into an equivalent ellipse needed for the normal force calculation. [33]

The *Hertzian theory of contact* between two elastic bodies is commonly used to give an approximation of the resulting normal force on an elliptical contact patch and can be found in [11]. The normal force in the contact area can be calculated with the assumption that the pressure distribution is a semi-ellipsoid.

In SIMPACK an elastic contact formulation is used. That means that the normal force in the contact patch is calculated by equivalent springs and dampers. For the calculation

[24] The flange is the inner part of the wheel that keep the wheels from running off the rails.




(a) Contact point

(b) Vertical and lateral contact forces

Figure 4.5: Wheel-rail contact

the user can choose between the Hertian method and a linear method. Both methods use an equivalent penetration of the wheels. [30]

The acting force in the contact patch results from a normal and tangential component and is usually split into a lateral force $Y$ and vertical force $Q$ (see Figure 4.5b). The ratio $Y/Q$ is often used to assess the safety against derailment.

The static component in lateral direction of the normal force helps to center the wheelset if the wheelset is displaced since the force is directed towards the track center. This effect is called the *gravitational stiffness*.

In reality some points on the surfaces in the contact region may slip while others may stick when the two bodies, wheel and rail, move relative to each other due to the elasticity of the bodies. As a result differences between the tangential strains of the bodies occur in the contact area and leads to a small apparent slip. This slipping effect is called *creepage* and leads to tangential creep forces and creep spin moments which have a significant effect on the driving characteristics of railway vehicles.

Thus, a good model of the contact can be achieved by dividing the contact area in a region where the particles of the bodies do not slide to each other, the adhesion region, and a region where the particles are sliding, the slip region. This was first discovered by Carter and can be found in his work. [5]

The tangential forces due to creepage in the contact area are usually split into longitudinal and lateral components. Kalker studied the relationship between creepages and creep forces carefully and proposed well-established methods and algorithms. [12]

In all three directions in which relative motion can occur creepage can be found. According to [8] the creepages are defined as follows:

$$\gamma_x = \frac{v_x' - v_x}{v} \tag{4.14}$$

$$\gamma_y = \frac{v_y' - v_y}{v} \tag{4.15}$$

$$\omega_z = \frac{\Omega_z' - \Omega_z}{v} \tag{4.16}$$

where $v_x$ and $v_y$ are the actual wheel velocities in $x$- and $y$-direction and $\Omega_z$ is the angular velocity about the vertical direction. The terms $v_x'$, $v_y'$ and $\Omega_z'$ represent the pure rolling

velocities[25] and $v$ is the forward velocity of the wheelset.

A very popular approach is the *simplified theory* of Kalker which is used in the algorithm FASTSIM to calculate the tangential forces in the contact and is described in [13]. It has some assumptions in common to Hertz as it assumes the contact area to be elliptic and flat. In FASTSIM, the contact surface is separated into a grid of parallel strips and the creep forces can then be calculated by computing internal creepages for every element. The relationship between creepage and creep forces is rather complicated and can be found in [12].

Kalker also proposed a *linear theory* where the creep forces are linearly dependent on the creepages as follows [8]:

$$F_x = -f_{11}\gamma_x \tag{4.17}$$

$$F_y = -f_{22}\gamma_y + f_{23}\omega_z \tag{4.18}$$

$$M_z = -f_{23}\gamma_y - f_{33}\omega_z \tag{4.19}$$

where $\gamma_x$, $\gamma_y$ and $\omega_z$ are the longitudinal, lateral and spin creepages and $f_{11}$, $f_{22}$, $f_{23}$ and $f_{33}$ are the linear creep coefficients depending on the Young's modulus, the Poisson ration and the ellipse semi-axis. These coefficients are constants and can be calculated from formulas which were approximated in [12].

## 4.2 Input by excitation

The excitation of a railway vehicle results mainly from geometry irregularities of the rails or the wheels. With the assumption that the wheels are not worn and perfectly shaped the only irregularities are given by the rail track. Input forces due to rail track irregularities build the unknown quantities to be identified within this work. A possible method of modeling of the input is presented in the end of this chapter.

### 4.2.1 Track irregularities

Track irregularities are present even on new tracks. These irregularities are mostly random and occur in different directions. Figure 4.6 illustrates the different types of track irregularities.

Generally a track can be described according to [17] by four irregularities with the following definitions:

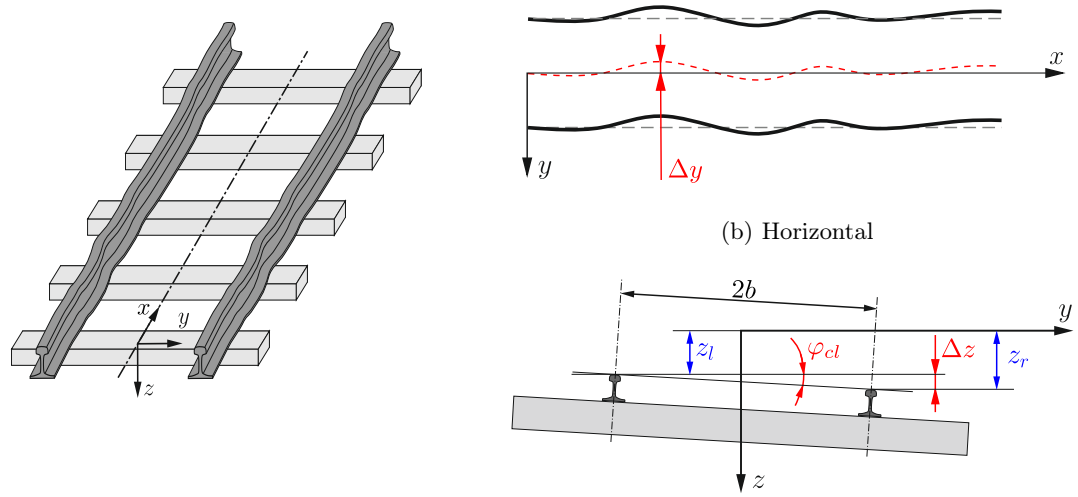
$$\text{lateral:} \qquad \Delta y = \frac{1}{2}(y_r + y_l) \tag{4.20}$$

$$\text{vertical:} \qquad \Delta z = \frac{1}{2}(z_r + z_l) \tag{4.21}$$

$$\text{crosslevel:} \qquad \varphi_{cl} = \frac{1}{2b}(z_r - z_l) \tag{4.22}$$

$$\text{gauge:} \qquad \Delta y_G = y_r - y_l \tag{4.23}$$

[25] The pure rolling velocity is the velocity when no creep occurs and thus both bodies have the same forward velocity.




(a) Rail track (exaggerated illustration)

(b) Horizontal

(c) Vertical and crosslevel

Figure 4.6: Track irregularities

where $y_l$ and $z_l$ describes the position of the left rail and $y_r$ and $z_r$ describe the position of the right rail. The term $b$ is half the track base.

Since misalignments of the track gauge are changing the wheel-rail contact geometry and do not have influence on the input forces directly, they are neglected for further investigations.

As stated above the irregularities are random and therefore are commonly characterized by their power spectral densities (PSDs). These spectral characteristics are obtained from exemplary track measurements and depend on the regarded track, which is described more detailed in [6].

According to ERRI B176 the polynomials of the PSDs for track irregularities in lateral ($y$), vertical ($z$) and crosslevel ($\varphi$) direction are the following:

$$S_y(\Omega) = \frac{b_{0y}}{0.00028855 + 0.6803895\Omega^2 + \Omega^4} \tag{4.24}$$

$$S_z(\Omega) = \frac{b_{0z}}{0.00028855 + 0.6803895\Omega^2 + \Omega^4} \tag{4.25}$$

$$S_\varphi(\Omega) = \frac{b_{2\varphi}\Omega^2}{5.535659 \cdot 10^{-5} + 0.1308172\Omega^2 + 0.8722335\Omega^4 + \Omega^6} \tag{4.26}$$

where $\Omega$ is the spatial frequency in $^{\text{rad}}/_{\text{m}}$. The units of the PSDs are $^{\text{m}^2}/_{(\text{rad/m})}$ in lateral and vertical direction and $^{\text{rad}^2}/_{(\text{rad/m})}$ in crosslevel direction. With $\Omega = 2\pi f$ and $f = \frac{1}{L}$ the spatial wave length $L$ of the rails can be obtained. [30] For a vehicle traveling with the velocity $V$ the angular frequency $\omega$ in $^{\text{rad}}/_{\text{s}}$ is defined by $\omega = V\Omega$. [10] The numerator coefficients for low and high track irregularities are shown in Table 4.1.

The rails are mounted on *sleepers*, which are nowadays usually made out of concrete. The regarded SIMPACK model also includes four sleeper elements for a more realistic modeling of the so-called *elastic foundation* of the rail track. [30]

The sleeper components can also be used to describe an unknown input term as described

| Coefficient | low irregularity | high irregularity |
|---|---|---|
| $b_{0y}$ | $1.440846 \cdot 10^{-7}$ | $4.164787 \cdot 10^{-7}$ |
| $b_{0z}$ | $2.741619 \cdot 10^{-7}$ | $7.343623 \cdot 10^{-7}$ |
| $b_{2\varphi}$ | $4.87399 \cdot 10^{-7}$ | $1.305533 \cdot 10^{-6}$ |

Table 4.1: Coefficients for low and high irregularities

in the following section.

### 4.2.2 Attainment of the input-force-matrix

The track irregularities lead to input forces which have to be modeled in order to formulate the inverse problem of unknown input estimation.

A front view of a simplified bogie is given in Figure 4.7 representing the model from the simulation software SIMPACK. For the later performed force estimation an input matrix has to be given to be able to identify unknown input signals. Since it is not possible in SIMPACK (Release 9.5) to declare random track irregularities $\Delta y$, $\Delta z$ and $\varphi_{cl}$ as input elements, which would provide the opportunity to export an input matrix, an alternative approach has to be found.

Assuming the occurrence of a perfect track without any track irregularities, the input can be regarded as excitations of the sleepers which are transmitted through the wheel-rail contact to the vehicle. In case of an elastic foundation, a sleeper element has three degrees of freedom and its motion can be described by the three generalized coordinates $y_{Sl}$, $z_{Sl}$ and $\varphi_{Sl}$.

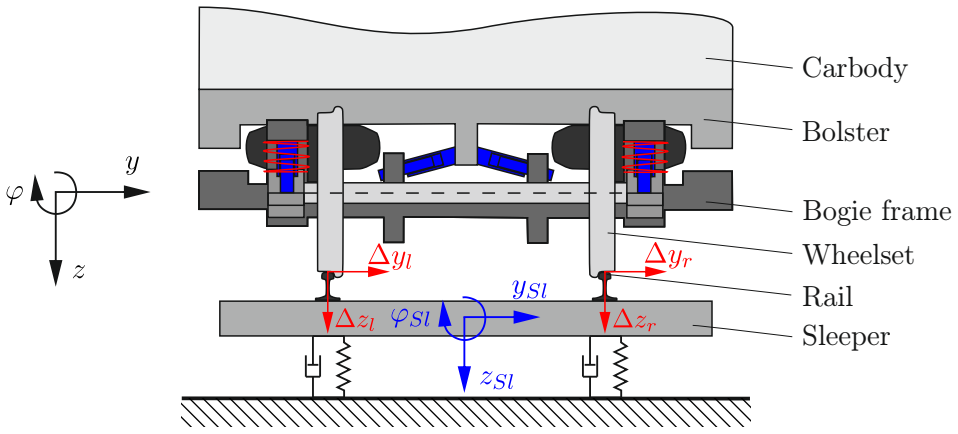
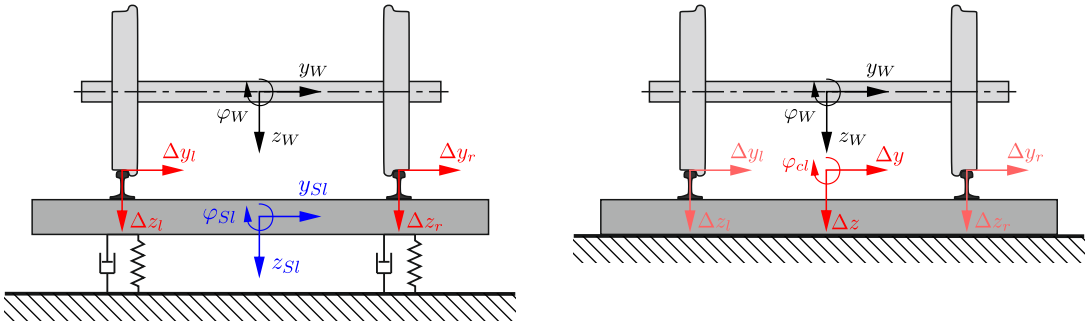



Figure 4.7: Front view of a bogie (simplified)

The state vector $\underline{x_{Sim}}$ of the whole railway vehicle system consists of $n$ generalized coordinates (and its derivatives) including the sleeper coordinates. To model the sleeper coordinates as input excitations the system has to be transformed. Therefore the SIMPACK model which is assembled by $n_b$ bodies has to be reduced by the components that belong to the sleepers.

Figure 4.8 shows the change in the model due to the reduction process. The generalized coordinates of the sleeper elements can then be used to describe the input because of the

track irregularities $\Delta y$, $\Delta z$ and $\varphi_{cl}$. Of course this approach comes along with the loss of the characteristics of an elastic foundation, but as can be seen later in this work, their influence for the estimation of input forces is not very high for the regarded model.




(a) Model before reduction

(b) Model after reduction

Figure 4.8: Reduction of the SIMPACK model

In further computations a formulation of the input forces can be found with the eliminated elements. Equation (4.27) gives the state-space representation of the linearized SIMPACK model without known input (see equation (4.9)):

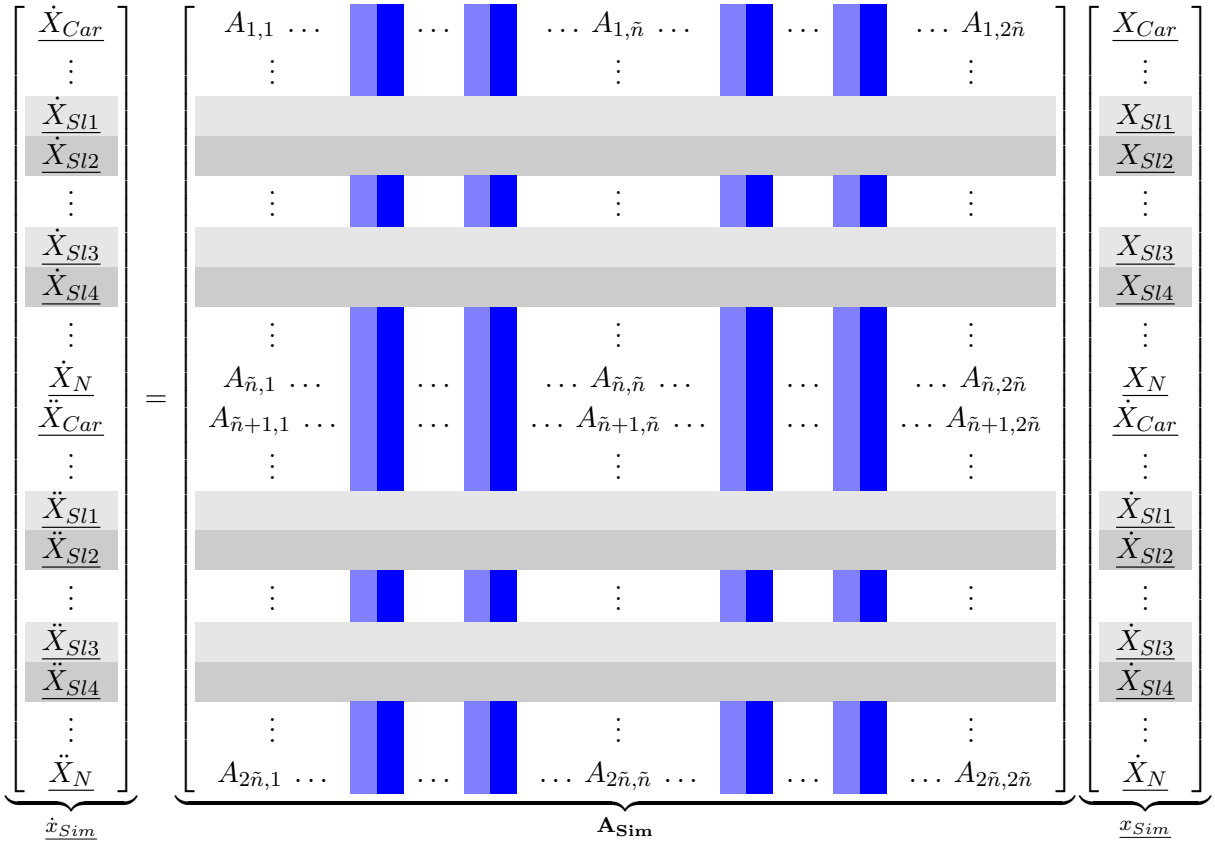
$$\underline{\dot{x}_{Sim}} = \mathbf{A_{Sim}}\, \underline{x_{Sim}} \tag{4.27}$$

with the state vector $\underline{x_{Sim}}$, its derivative $\underline{\dot{x}_{Sim}}$ and the system matrix $\mathbf{A_{Sim}}$. The aim is to rearrange the equation from above to obtain an additional term $\mathbf{G}\underline{w}$ representing the input force as shown in equation (4.28).

$$\Rightarrow \underline{\dot{x}} = \mathbf{A}\underline{x} + \mathbf{G}\underline{w} \tag{4.28}$$

Therefore the state vector $\underline{x_{Sim}}$ and the system matrix $\mathbf{A_{Sim}}$ have to be reduced by elimination of components which belong to the sleepers. The procedure is given by equations (4.29)-(4.31).

First, the rows of the sleeper components (marked in gray) have to be selected and eliminated. As a result, the input matrix $\mathbf{G}$ can be built with the columns of the sleeper components since they represent the influence of the sleepers on all the other states (marked in blue). The reduction of $\underline{x_{Sim}}$ and $\mathbf{A_{Sim}}$ can be achieved through elimination of these entries.



$$\dot{x}_{Sim} = \mathbf{A_{Sim}} \, x_{Sim} \tag{4.29}$$

$$\Rightarrow \quad \dot{\underline{x}} = \underline{A} \, x_{Sim} \tag{4.30}$$

Equation (4.31) is the matrix form of equation (4.28) and represents the new state-space model of the reduced system.

$$\Leftrightarrow \underbrace{\begin{bmatrix} \underline{\dot{X}_{Car}} \\ \vdots \\ \vdots \\ \vdots \\ \underline{\dot{X}_N} \\ \underline{\ddot{X}_{Car}} \\ \vdots \\ \vdots \\ \vdots \\ \underline{\ddot{X}_N} \end{bmatrix}}_{\underline{\dot{x}}} = \underbrace{\begin{bmatrix} A_{1,1} & \dots & \dots & \dots & A_{1,n} & \dots & \dots & \dots & A_{1,2n} \\ \vdots & & & & \vdots & & & & \vdots \\ \vdots & & & & \vdots & & & & \vdots \\ \vdots & & & & \vdots & & & & \vdots \\ A_{n,1} & \dots & \dots & \dots & A_{n,n} & \dots & \dots & \dots & A_{n,2n} \\ A_{n+1,1} & \dots & \dots & \dots & A_{n+1,n} & \dots & \dots & \dots & A_{n+1,2n} \\ \vdots & & & & \vdots & & & & \vdots \\ \vdots & & & & \vdots & & & & \vdots \\ \vdots & & & & \vdots & & & & \vdots \\ A_{2n,1} & \dots & \dots & \dots & A_{2n,n} & \dots & \dots & \dots & A_{2n,2n} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \underline{X_{Car}} \\ \vdots \\ \vdots \\ \vdots \\ \underline{X_N} \\ \underline{\dot{X}_{Car}} \\ \vdots \\ \vdots \\ \vdots \\ \underline{\dot{X}_N} \end{bmatrix}}_{\underline{x}} + \underbrace{\begin{bmatrix} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \underline{X_{Sl1}} \\ \underline{X_{Sl2}} \\ \underline{X_{Sl3}} \\ \underline{X_{Sl4}} \\ \underline{\dot{X}_{Sl1}} \\ \underline{\dot{X}_{Sl2}} \\ \underline{\dot{X}_{Sl3}} \\ \underline{\dot{X}_{Sl4}} \end{bmatrix}}_{\underline{w}} \tag{4.31}$$

with the state vectors containing the generalized coordinates:

$$\underline{X_{Car}} = \begin{bmatrix} y_{Car} \\ z_{Car} \\ \varphi_{Car} \\ \psi_{Car} \\ \vartheta_{Car} \end{bmatrix}, \quad \underline{X_{Sl1}} = \begin{bmatrix} y_{Sl1} \\ z_{Sl1} \\ \varphi_{Sl1} \end{bmatrix}, \quad \underline{X_{Sl2}} = \begin{bmatrix} y_{Sl2} \\ z_{Sl2} \\ \varphi_{Sl2} \end{bmatrix},$$

$$\underline{X_{Sl3}} = \begin{bmatrix} y_{Sl3} \\ z_{Sl3} \\ \varphi_{Sl3} \end{bmatrix}, \quad \underline{X_{Sl4}} = \begin{bmatrix} y_{Sl4} \\ z_{Sl4} \\ \varphi_{Sl4} \end{bmatrix}, \quad \underline{X_N} = \begin{bmatrix} x_{1N} \\ \vdots \\ x_{iN} \end{bmatrix}.$$

The indices have the following meanings:

| | | | |
|---|---|---|---|
| $Car$ | Carbody | $N$ | Random body |
| $Sl1$ | Front sleeper of front bogie | $iN$ | Number of generalized coordinates of body $N$ |
| $Sl2$ | Rear sleeper of front bogie | $\tilde{n}$ | Total number of generalized coordinates of the system |
| $Sl3$ | Front sleeper of rear bogie | $n$ | Reduced number of generalized coordinates of the system |
| $Sl3$ | rear sleeper of rear bogie | | |

As there are four sleepers with three generalized coordinates each ($n_s = 12$), the reduced system has $n = \tilde{n} - n_s$ generalized coordinates who form the reduced state vector $\underline{x}$. The state-space matrix $\mathbf{A}$ has less elements than $\mathbf{A_{Sim}}$ since the sleeper components who build the input matrix $\mathbf{G}$ were eliminated. The input vector $\underline{w}$ is composed by the generalized coordinates of the four sleepers. Table 4.2 shows the dimensions of the single elements of the state-space model before and after the reduction of the sleeper elements.

| Vector/Matrix | Dimension | |
|---|---|---|
| $\underline{\dot{x}_{Sim}}, \underline{x_{Sim}}$ | $[2\tilde{n} \times 1]$ | |
| $\mathbf{A_{Sim}}$ | $[2\tilde{n} \times 2\tilde{n}]$ | |
| $\underline{\dot{x}}, \underline{x}$ | $[(2(\tilde{n} - n_s) \times 1]$ | or $[2n \times 1]$ |
| $\mathbf{A}$ | $[(2(\tilde{n} - n_s) \times (2(\tilde{n} - n_s)]$ | or $[2n \times 2n]$ |
| $\mathbf{G}$ | $[(2(\tilde{n} - n_s) \times 2n_s]$ | or $[2n \times n_s]$ |
| $\underline{w}$ | $[n_s \times 1]$ | |

Table 4.2: Dimensions of vectors and matrices before and after reduction

The input vector is named $\underline{w}$ since it is regarded as unknown within this work. In chapter 5 an inverse problem is formulated with the objective of estimation of these particular inputs.

### 4.2.3 Wheel-rail contact force calculation

The wheel-rail contact forces of the regarded railway vehicle, shown in Figure 4.1, occur in eight contact patches, since the model contains four wheelsets. Figure 4.9 shows the front view of a bogie and the applied wheel-rail contact forces on the left and right wheel of a wheelset.

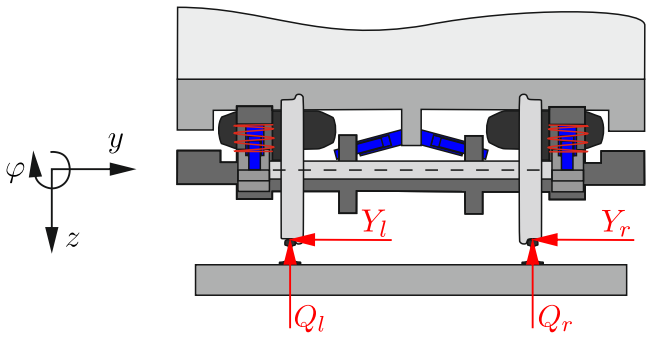



Figure 4.9: Applied contact forces

The forces can be written in a vector $\underline{F}$ where $Y_{i1}$ and $Y_{i2}$ represent the lateral forces for the right and the left wheel and $Q_{i1}$ and $Q_{i2}$ are the vertical forces for the different sides of the $i$-th wheelset respectively ($i = 1, 2, 3, 4$). If only the resultant forces for each wheelset are requested a summation of the individual forces can be performed. The output force

vectors $\underline{F}$ and $\underline{F_\Sigma}$ including all four wheelsets can then be given as follows:

$$\underline{F} = \begin{bmatrix} Y_{11} \\ Q_{11} \\ Y_{12} \\ Q_{12} \\ \vdots \\ Y_{42} \\ Q_{42} \end{bmatrix} \quad \text{or} \quad \underline{F_\Sigma} = \begin{bmatrix} Y_1 \\ Q_1 \\ Y_2 \\ Q_2 \\ Y_3 \\ Q_3 \\ Y_4 \\ Q_4 \end{bmatrix} = \begin{bmatrix} Y_{11} + Y_{12} \\ Q_{11} + Q_{12} \\ Y_{21} + Y_{22} \\ Q_{21} + Q_{22} \\ Y_{31} + Y_{32} \\ Q_{31} + Q_{32} \\ Y_{41} + Y_{42} \\ Q_{41} + Q_{42} \end{bmatrix} \tag{4.32}$$

where the first index of the forces denotes the wheelset and the second index stands for the location of the wheel[26].

An expression for lateral and vertical contact forces as output vector in state-space representation can be given analogous to equation (4.10) by the following:

$$\underline{F} = \bar{\mathbf{C}}_\mathbf{F} \underline{x_{Sim}} \tag{4.33}$$

$$= \mathbf{C}_\mathbf{F} \underline{x} + \mathbf{D}_\mathbf{F} \underline{w}. \tag{4.34}$$

where $\underline{x_{Sim}}$ is the state vector and $\bar{\mathbf{C}}_\mathbf{F}$ is the force output matrix. As well as the linearized state-space matrices this matrix can be obtained from SIMPACK through linearization about the initial value. Similarly to above, the matrix $\bar{\mathbf{C}}_\mathbf{F}$ has to be split into $\mathbf{C}_\mathbf{F}$ and $\mathbf{D}_\mathbf{F}$ according to the reduced model, where $\underline{x}$ is the state vector and $\underline{w}$ is the unknown input vector.

With equation (4.34) the contact forces can be calculated directly if the states $\underline{x}$ and the input $\underline{w}$ are known. Later in this work, the wheel-rail forces can be derived using this expression since the state and input estimates are obtained through application of an inverse estimation approach as described in chapter 3.

[26] $Y_{11}$ e.g. represents the lateral force occurring at the right wheel of the first wheelset, since the first 1 stands for the first wheelset and the second 1 stands for the right wheel.

# 5 Input force estimation for a railway vehicle system

The main part of this work is presented in this chapter. First, the inverse input estimation problem for a railway vehicle system is formulated. Measurement data provided from a simulation are the known signals needed for the inverse identification method and obtained from accelerometers. Before the Kalman filter can be applied in order to obtain state and input estimates the system has to be extended by a form filter. In the end, the results for the estimated wheel-rail contact forces are compared with the simulated forces.

## 5.1 Inverse problem formulation

The wheel-rail contact forces that occur in the contact between wheels and rails of a railway vehicle system, as described in chapter 4, have significant influences on running safety and track loading as well as on the durability of vehicle components. Therefore, their identification is of strong interest.

Due to random track irregularities the prediction of these forces is not trivial. For testing, vehicles are equipped with devices to compute the forces from strain measurements of strain gauge bridges which are mounted on wheel discs and/or the rail track. Therefore, instrumented wheelsets are applied which are more or less custom-made. Track-side measurement points have to be installed at many locations on the track in order to attain useful results. More about the testing method can be found in [25].

However, these methods are expensive. Moreover, it is practically impossible to install instrumented wheelsets during the long lasting regular operation of a train. Therefore, a great demand for alternative approaches exists. But even though testing procedures are often time consuming and expensive, some measurement data is irreplaceable. At present various methods based on the inverse problem of using acceleration data in order to determine unknown wheel-rail contact forces are proposed. Uhl presents in [32] an inverse method for dynamic load estimation using measurements of system responses. In [37] and [38] an inverse dynamics method is introduced by Zhu et al. in order to identify lateral wheel-rail forces. These methods make use of the Thikonov regularization and the Bellman principle of optimality to minimize the objective function for the estimation of applied forces due to excitation, which is common for the inverse identification problem. For the application of these methods no expensive instrumented wheelsets are needed since accelerometers can be mounted more or less easily on every common vehicle, and sometimes are already installed by default.

In this work, a new approach is introduced also using measurement data from accelerometers. Hence, the inverse problem to be solved is the estimation of unknown wheel-rail contact forces with knowledge of acceleration data.

### 5.1.1 Problem statement

According to chapter 4.2 the linear state-space model of the regarded railway vehicle system is given by:

$$\dot{\underline{x}} = \mathbf{A}\,\underline{x} + \mathbf{G}\,\underline{w} \tag{5.1}$$

$$\underline{y} = \mathbf{C}\,\underline{x} + \mathbf{D}\,\underline{w} \tag{5.2}$$

with the state vector $\underline{x}$, the input vector $\underline{w}$, the output vector $\underline{y}$ and the system matrices $\mathbf{A}$, $\mathbf{G}$, $\mathbf{C}$ and $\mathbf{D}$.

For the inverse problem described above the input vector $\underline{w}$ is assumed to be unknown whereas the output vector $\underline{y}$ contains the measured accelerations. The system matrices can be obtained from the simulation in SIMPACK where they are derived from the linearized equations of motion (see sections 4.1.3 and 4.2.2).

The unknown input represents the sleeper displacements and sleeper velocities. Therefore a formulation for the calculation of the wheel-rail contact forces has to be introduced separately. The expression derived in section 4.2.3 is the following:

$$\underline{F} = \mathbf{C_F}\underline{x} + \mathbf{D_F}\underline{w}\ . \tag{5.3}$$

where the output vector $\underline{F}$ contains the wheel-rail contact forces. The force-output matrices $\mathbf{C_F}$ and $\mathbf{D_F}$ can be obtained in similar manner as the system matrices $\mathbf{A}$, $\mathbf{G}$, $\mathbf{C}$ and $\mathbf{D}$ from the simulation in SIMPACK.

The stated inverse problem for the railway vehicle system is according to section 3.1 a deconvolution with the aim to estimate unknown forces for a given system with measurement data.

### 5.1.2 Estimation procedure

For the inverse problem from above the estimation procedure can be split into five steps. The five steps are defined as follows and are described more detailed in the next sections:

**Step 1:** Simulation of the railway vehicle system and generation of measurement data.

**Step 2:** Extension of the linearized system with a form filter.

**Step 3:** State and input estimation with a Kalman filter on basis of measurement data.

**Step 4:** Force calculation with estimated states and inputs.

**Step 5:** Comparison of estimated and simulated results for the input forces.

It has to be mentioned at this point that the same procedure could be applied with real measurements and results. This would result in exchanging the simulated data and results used in steps 1 and 5 by real measurements. Within this work measurements are generated with a simulation in SIMPACK.

## 5.2 Measurement data from simulation

As stated before, the measurements are acquired from a SIMPACK simulation instead of an actual vehicle since the main goal was the implementation of an inverse method. The usage of real measurement data comes along with several additional problems and phenomena which wanted to be prevented in this study to avoid a growth of complexity for the considered system.

For the discussed model the positions of the sensors are shown in Figure 5.1. A symmetrical configuration is chosen, where eight accelerometers are installed in the axle boxes of the wheelsets, respectively two on the bogie frames and two in the center between the bolsters and the carbody. For better identification of rotations about the $z$-axis (yaw-motion), those sensors mounted on the wheelsets provide data in all three directions whereas the others only measure the $y$- and $z$-component of the accelerations.

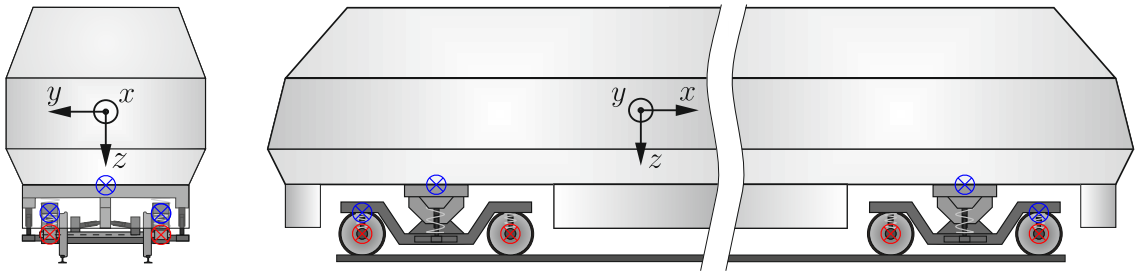



Figure 5.1: Positions of sensors on the railway vehicle

In the SIMPACK simulation the vehicle is traveling with a constant speed of $44\,\mathrm{m/s}$ ($= 158.4\,\mathrm{km/h}$). This reduces the number of DOFs by one.[27] The number of degrees of freedom $n$ of the regarded model is then 41 which can also be seen in Table 5.1. In the SIMPACK model the rail track foundation is elastic and therefore four sleeper elements are implemented. These bodies are used later to obtain the input-matrix for the estimation process as described in section 4.2.2.

| Body | DOFs |
|---|---|
| Carbody | 5 |
| Bogie frame, front | 6 |
| Bogie frame, rear | 6 |
| Wheelset 1 | 6 |
| Wheelset 2 | 6 |
| Wheelset 3 | 6 |
| Wheelset 4 | 6 |
| $n_b = 7$ | $n = 41$ |

| Body | DOFs |
|---|---|
| Sleeper 1 | 3 |
| Sleeper 2 | 3 |
| Sleeper 3 | 3 |
| Sleeper 4 | 3 |
| $n_{sb} = 4$ | $n_s = 12$ |

Table 5.1: DOFs of the simulation model

The model contains 34 additional so-called dynamic states ($n_{dyn} = 34$) describing special

[27] In this case the forward motion of the carbody ($x$-direction) is constrained due to the constant velocity.

force elements as serial spring-dampers e.g. (see section 4.1.2, $s_{dyn}$). This finally results in a state vector with 140 state entries ($= (n + n_s) \cdot 2 + n_{dyn}$) for the simulation model.[28]

The vehicle is driving on a straight track, where high track irregularities (ERRI B176 high, according to [30]) are assumed to be given. This is important in order to select the correct form filter parameters, given in Table 4.1, required for the system extension which is introduced next.

## 5.3 System extension with a form filter

With background knowledge about the spectral characteristics of the unknown input the state-space model representation in equations (5.1) and (5.2) can be extended with a form filter containing information about the random track irregularities, which were introduced in chapter 4.2.1. This is necessary because of the fact that the Kalman filter algorithm, which will be applied for state estimation, works best with white noise.

Figure 5.2 shows the block diagram of the extended system, where $\underline{w}$ represents the unknown input for the railway vehicle system and is equal to the output $\underline{y_{FF}}$ of the form filter element. The extension of the model allows the implementation of the Kalman filter algorithm for the given state-space model even when the unknown input is not white noise, which is the case for the assumed excitation by random track irregularities.

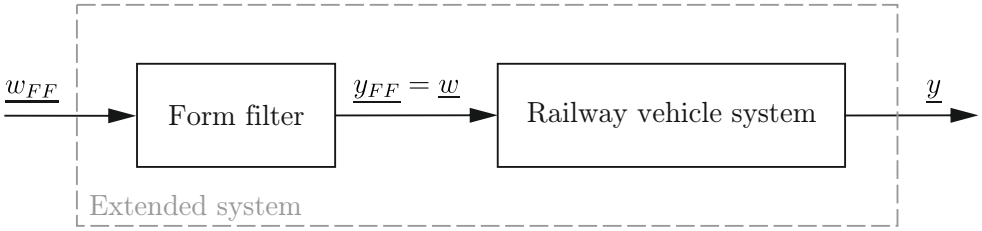



Figure 5.2: Model of form filter and system

With the PSD polynomials describing the spectral characteristics of the track irregularities from chapter 4.2.1 the form filter can be formulated. The PSDs $S(\omega)$ of the irregularities can be formulated according to [27] with the constant PSD $S_0$ of white noise as the following:

$$S(\omega) = S_0 \cdot |F(j\omega)|^2 \tag{5.4}$$

with the transfer function $F(j\omega)$ of the form filter. The transfer function $F(j\omega)$ of the form filter can be derived through factorization of the PSD polynomial since:

$$|F(j\omega)|^2 = F(j\omega) \cdot F^*(j\omega) = F_{ff}(j\omega) \cdot F^*(j\omega) \tag{5.5}$$

[28] The states of the bodies with masses are described by displacement and velocity (two first-order states) whereas the dynamical states only are described by a differential equation of first-order each (one first-order state).[30] See also sections 4.1.2 and 4.1.3.

where $F^*(j\omega)$ is the conjugate complex of the transfer function $F(j\omega) = F_{ff}(j\omega)$ of the form filter. The transformation with $s = j\omega$ leads to:

$$S(s) = S_0 \cdot F(-s) \cdot F(s) \tag{5.6}$$

$$F_{ff}(s) = F(s) \tag{5.7}$$

with the transfer function $F_{ff}(s)$ of the form filter. This formulation is necessary to derive a state-space model description of the filter.

With assumption of a standard normal distribution of white noise ($\sigma^2 = 1,\ \Rightarrow S_0 = 1$) the PSD polynomials from equations (4.24)-(4.26) can be written after substitution of the spatial frequency $\Omega$ with $s = j\omega$ ($\Rightarrow \Omega = \frac{s}{jV}$) as follows:

$$\begin{aligned} S_y(s) &= \frac{b_{0y}V^4}{0.00028855V^4 - 0.6803895V^2s^2 + s^4} = \cdots = \\ &= \underbrace{\frac{\sqrt{b_{0y}}V^2}{0.0169876V^2 - 0.8452Vs + s^2}}_{F_y(-s)} \cdot \underbrace{\frac{\sqrt{b_{0y}}V^2}{0.0169876V^2 + 0.8452Vs + s^2}}_{F_y(s)=F_{ff_y}(s)} \end{aligned} \tag{5.8}$$

$$\begin{aligned} S_z(s) &= \frac{b_{0z}V^4}{0.00028855V^4 - 0.6803895V^2s^2 + s^4} = \cdots = \\ &= \underbrace{\frac{\sqrt{b_{0z}}V^2}{0.0169876V^2 - 0.8452Vs + s^2}}_{F_z(-s)} \cdot \underbrace{\frac{\sqrt{b_{0z}}V^2}{0.0169876V^2 + 0.8452Vs + s^2}}_{F_z(s)=F_{ff_z}(s)} \end{aligned} \tag{5.9}$$

$$\begin{aligned} S_\varphi(s) &= \frac{b_{2\varphi}V^4s^2}{5.535659 \cdot 10^{-5}V^6 + 0.1308172V^4s^2 - 0.8722335V^2s^4 + s^6} = \cdots = \\ &= \underbrace{\frac{\sqrt{b_{2\varphi}}V^2s}{0.000744V^3 - 0.387184V^2s + 1.2832Vs^2 - s^3}}_{F_\varphi(-s)} \cdot \\ &\qquad \cdot \underbrace{\frac{\sqrt{b_{2\varphi}}V^2s}{0.000744V^3 + 0.387184V^2s + 1.2832Vs^2 + s^3}}_{F_\varphi(s)=F_{ff_\varphi}(s)} . \end{aligned} \tag{5.10}$$

The factorization of the PSD polynomials in equations (5.8), (5.9) and (5.10) lead to the transfer functions $F_{ff_y}(s)$, $F_{ff_z}(s)$ and $F_{ff_\varphi}(s)$. With these transfer functions a state-space model description for the form filter element can be derived and is given by:

$$\underline{\dot{x}_{ff_i}} = \mathbf{A_{ff_i}}\, \underline{x_{ff_i}} + \mathbf{B_{ff_i}}\, \underline{w_{ff_i}} \tag{5.11}$$

$$\underline{y_{ff_i}} = \mathbf{C_{ff_i}}\, \underline{x_{ff_i}} \tag{5.12}$$

with the matrices for the lateral, vertical and crosslevel direction:

$$
\begin{aligned}
\mathbf{A_{ff_y}} &= \begin{bmatrix} 0 & 1 \\ -0.01698676V^2 & -0.8452V \end{bmatrix}, \quad \mathbf{B_{ff_y}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{C_{ff_y}} = \begin{bmatrix} \sqrt{b_{0y}}V^2 & 0 \\ 0 & \sqrt{b_{0y}}V^2 \end{bmatrix}, \\
\mathbf{A_{ff_z}} &= \begin{bmatrix} 0 & 1 \\ -0.01698676V^2 & -0.8452V \end{bmatrix}, \quad \mathbf{B_{ff_z}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{C_{ff_z}} = \begin{bmatrix} \sqrt{b_{0z}}V^2 & 0 \\ 0 & \sqrt{b_{0z}}V^2 \end{bmatrix}, \\
\mathbf{A_{ff_\varphi}} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.000744V^3 & -0.387184V^2 & -1.2832V \end{bmatrix}, \quad \mathbf{B_{ff_\varphi}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \\
\mathbf{C_{ff_\varphi}} &= \begin{bmatrix} 0 & \sqrt{b_{2\varphi}}V^2 & 0 \\ 0 & 0 & \sqrt{b_{2\varphi}}V^2 \end{bmatrix}.
\end{aligned}
\tag{5.13}
$$

The combination of all three directions of the irregularities yields a set of equations with the following matrices:

$$
\mathbf{A_{ff}} = \begin{bmatrix} \mathbf{A_{ff_y}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A_{ff_z}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A_{ff_\varphi}} \end{bmatrix}, \mathbf{B_{ff}} = \begin{bmatrix} \mathbf{B_{ff_y}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B_{ff_z}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B_{ff_\varphi}} \end{bmatrix}, \mathbf{C_{ff}} = \begin{bmatrix} \mathbf{C_{ff_y}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_{ff_z}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C_{ff_\varphi}} \end{bmatrix}.
\tag{5.14}
$$

Since the railway vehicle consists of four independent wheelsets, where any correlation between their excitation is neglected, an appropriate description of the form filter has to be found. Hence the form filter system including all irregularities of the four wheelsets results in:

$$
\underline{\dot{x}_{FF}} = \mathbf{A_{FF}}\, \underline{x_{FF}} + \mathbf{B_{FF}}\, \underline{w_{FF}} \tag{5.15}
$$

$$
\underline{y_{FF}} = \mathbf{C_{FF}}\, \underline{x_{FF}} \tag{5.16}
$$

with the system matrices

$$
\begin{aligned}
\mathbf{A_{FF}} &= \begin{bmatrix} \mathbf{A_{ff}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A_{ff}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A_{ff}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A_{ff}} \end{bmatrix}, \quad \mathbf{B_{FF}} = \begin{bmatrix} \mathbf{B_{ff}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B_{ff}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B_{ff}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B_{ff}} \end{bmatrix}, \\
\mathbf{C_{FF}} &= \begin{bmatrix} \mathbf{C_{ff}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_{ff}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C_{ff}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C_{ff}} \end{bmatrix}.
\end{aligned}
\tag{5.17}
$$

The form filter is implemented through the combination with the state-space model of the railway vehicle system given by equations (5.1) and (5.2). The combination is performed by setting the unknown input of the railway vehicle system $\underline{w}$ equal to the output $\underline{y_{FF}}$ of the form filter described in equation (5.15) and (5.16). This results in a set of three

equations:

$$\dot{\underline{x}} = \mathbf{A}\,\underline{x} + \mathbf{GC_{FF}}\,\underline{x_{FF}} \tag{5.18}$$

$$\dot{\underline{x}}_{FF} = \mathbf{A_{FF}}\,\underline{x_{FF}} + \mathbf{B_{FF}}\,\underline{w_{FF}} \tag{5.19}$$

$$\underline{y} = \mathbf{C}\,\underline{x} + \mathbf{DC_{FF}}\,\underline{x_{FF}}\ . \tag{5.20}$$

The extended system can be rearranged by introducing the state vector $\underline{\tilde{x}}$ of the combined system as follows:

$$\underbrace{\begin{bmatrix}\dot{\underline{x}}\\ \dot{\underline{x}}_{FF}\end{bmatrix}}_{\dot{\underline{\tilde{x}}}} = \underbrace{\begin{bmatrix}\mathbf{A} & \mathbf{GC_{FF}}\\ \mathbf{0} & \mathbf{A_{FF}}\end{bmatrix}}_{\tilde{\mathbf{A}}} \underbrace{\begin{bmatrix}\underline{x}\\ \underline{x_{FF}}\end{bmatrix}}_{\underline{\tilde{x}}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{B_{FF}}\end{bmatrix}}_{\tilde{\mathbf{B}}} \underline{w_{FF}} \tag{5.21}$$

$$\underline{y} = \underbrace{\begin{bmatrix}\mathbf{C} & \mathbf{DC_{FF}}\end{bmatrix}}_{\tilde{\mathbf{C}}} \underbrace{\begin{bmatrix}\underline{x}\\ \underline{x_{FF}}\end{bmatrix}}_{\underline{\tilde{x}}}\ . \tag{5.22}$$

Finally, a state-space model description of the railway vehicle system extended with a form filter for random track irregularities is obtained and given by:

$$\dot{\underline{\tilde{x}}} = \tilde{\mathbf{A}}\,\underline{\tilde{x}} + \tilde{\mathbf{B}}\,\underline{w_{FF}} \tag{5.23}$$

$$\underline{y} = \tilde{\mathbf{C}}\,\underline{\tilde{x}}\ . \tag{5.24}$$

## 5.4 Kalman filter with unknown input

In section 3.4 the Kalman filter algorithm is described and applied for the unknown input estimation of a two-mass system. In similar manner this approach can be used to identify the states of the regarded railway vehicle system in order to estimate the unknown wheel-rail contact forces.

The discrete-time state-space model for the extended system from equations (5.23) and (5.24) is given by the following set of equations:

$$\underline{\tilde{x}_k} = \mathbf{A_d}\,\underline{\tilde{x}_{k-1}} + \mathbf{B_d}\,\underline{w_{FF_{k-1}}} \tag{5.25}$$

$$\underline{y_k} = \mathbf{C_d}\,\underline{\tilde{x}_k} \tag{5.26}$$

with the discrete-time state vector $\underline{\tilde{x}_k} = [\underline{x_k}\ \underline{x_{FF_k}}]^T$, the output vector $\underline{y_k}$ and the discrete-time system matrices $\mathbf{A_d}$, $\mathbf{B_d}$ and $\mathbf{C_d}$. The discretization with the time step size $\Delta t = t_k - t_{k-1}$ can be performed according to section 2.1.3 by:

$$\mathbf{A_d} = e^{\tilde{\mathbf{A}}\Delta t}, \quad \mathbf{B_d} = \mathbf{A_d}[\mathbf{I} - e^{-\tilde{\mathbf{A}}\Delta t}]\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\ . \tag{5.27}$$

The discrete-time output matrix $\mathbf{C_d}$ is equal to its continuous form $\mathbf{C}$.

Before the update equations can be evaluated the system has to be initialized. The initial

estimate $\underline{\hat{\tilde{x}}_0}$ of the initial state $\underline{\tilde{x}_0} = \underline{\tilde{x}}(t = t_0 = 0)$ is defined as:

$$\underline{\hat{\tilde{x}}_{0|0}} = E[\underline{\tilde{x}_0}] \tag{5.28}$$

$$\mathbf{P_{0|0}} = E[(\underline{\tilde{x}_0} - \underline{\hat{\tilde{x}}_{0|0}})(\underline{\tilde{x}_0} - \underline{\hat{\tilde{x}}_{0|0}})^T] \tag{5.29}$$

and set to be zero here assuming that the vehicle is at rest at the beginning of the simulation process.

Identically to section 3.4 the Kalman filter algorithm can be applied after the initialization for each time step $k = 1, \ldots, n$ and is defined by the following time-update equations according to the description of the extended system given in equation (5.25) and (5.26):

$$\mathbf{P_{k|k-1}} = \mathbf{A_d P_{k-1|k-1} A_d}^{\mathbf{-1}} + \mathbf{Q} \tag{5.30}$$

$$\underline{\hat{\tilde{x}}_{k|k-1}} = \mathbf{A_d}\underline{\hat{\tilde{x}}_{k-1|k-1}} \tag{5.31}$$

which represents a prediction step for the covariance $\mathbf{P_k}$ and the a priori state estimate $\underline{\hat{\tilde{x}}_{k|k-1}}$ with $\mathbf{Q} = \mathbf{G_d}\bar{\mathbf{Q}}\mathbf{G_d^T}$.

The measurement-update equations lead to the a posteriori state estimate $\underline{\hat{\tilde{x}}_{k|k}}$ and are defined as follows:

$$\begin{aligned} \mathbf{L_k} &= \mathbf{P_{k|k-1}C_d}^{\mathbf{T}}(\mathbf{C_d P_{k|k-1} C_d}^{\mathbf{T}} + \mathbf{R})^{-\mathbf{1}} \\ &= \mathbf{P_{k|k}C_d}^{\mathbf{T}}\mathbf{R}^{-\mathbf{1}} \end{aligned} \tag{5.32}$$

$$\begin{aligned} \mathbf{P_{k|k}} &= (\mathbf{I} - \mathbf{L_k C_d})\mathbf{P_{k|k-1}}(\mathbf{I} - \mathbf{L_k C_d})^{\mathbf{T}} + \mathbf{L_k R L_k}^{\mathbf{T}} \\ &= \cdots = (\mathbf{I} - \mathbf{L_k C_d})\mathbf{P_{k|k-1}} \end{aligned} \tag{5.33}$$

$$\underline{\hat{\tilde{x}}_{k|k}} = \underline{\hat{\tilde{x}}_{k|k-1}} + \mathbf{L_k}(\underline{y_k} - \mathbf{C_d}\underline{\hat{\tilde{x}}_{k|k-1}}) \ . \tag{5.34}$$

With equation (5.34) the final state estimation can be obtained and used for the force calculation which is presented next.

## 5.5 Force calculation

The final step of the estimation process is given by the calculation of the wheel-rail contact forces with the estimated states $\underline{\hat{x}_k}$ and the estimated inputs $\underline{\hat{w}_k}$ of the railway vehicle system. With use of equations (5.16) and (5.3) an estimation of the wheel-rail contact forces can be obtained as follows:

$$\underline{\hat{w}_k}(= \underline{\hat{y}_{FF_k}}) = \mathbf{C_{FF}}\,\underline{\hat{\tilde{x}}_k} \tag{5.35}$$

$$\underline{\hat{F}_k} = \mathbf{C_F}\underline{\hat{x}_k} + \mathbf{D_F}\underline{\hat{w}_k} \tag{5.36}$$

where $\underline{\hat{\tilde{x}}}$ is the estimated state of the extended system and $\underline{\hat{x}}$ is the estimated state vector of the railway vehicle system.

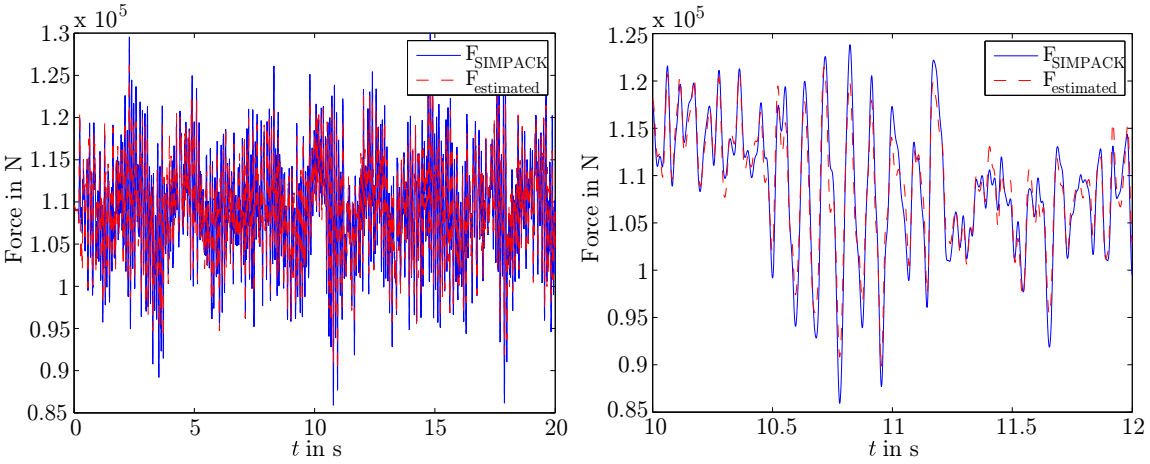
With the presented inverse identification method results for the wheel-rail contact forces acting in the contact patches can be derived and are compared with the simulated results in the following section.

## 5.6 Comparison and validation of results

The wheel-rail contact forces are evaluated according to EN 14363 [1] up to frequencies of 20 Hz. Therefore the contact force signals have to be low-pass filtered in advance. Here, a 4th order Butterworth low-pass filter is applied to all of the simulated and the estimated forces before they are compared with each other.

The simulation time is 100 s while the vehicle, as described above, is driving with a constant speed of 44 m/s on a straight track, where ERRI high track irregularities are assumed. The vehicle elements and chosen parameters are given in appendix B.

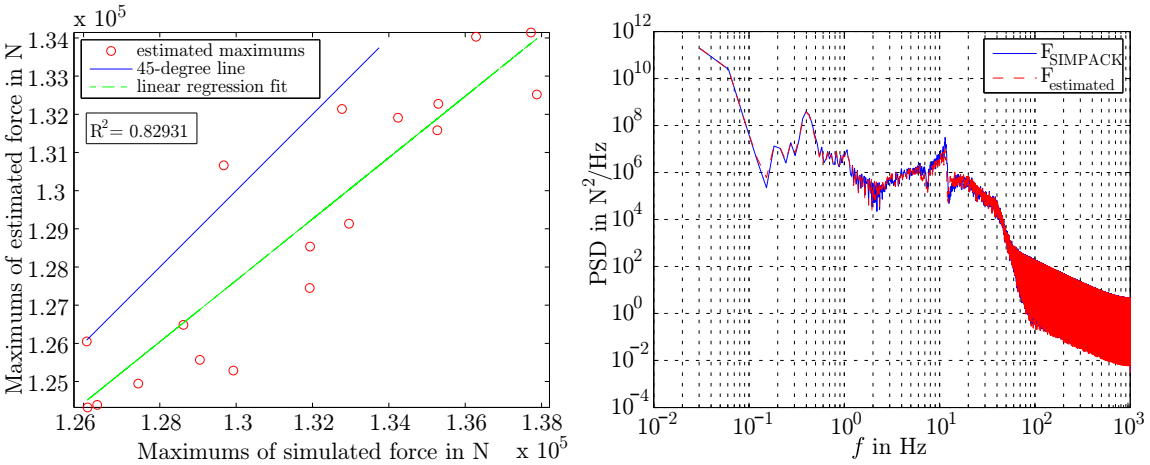
A comparison of estimated and simulated signals for the resultant vertical contact forces for the first wheelset (named *Wheelset 1* in Figure 4.1) in the time $t = 0 - 20$ s is shown in Figure 5.3a. An enlargement of the diagram is given in Figure 5.3b where a significant correlation can be identified.




(a) Vertical contact force

(b) Vertical contact force, enlarged

Figure 5.3: Resultant vertical contact forces for the first wheelset



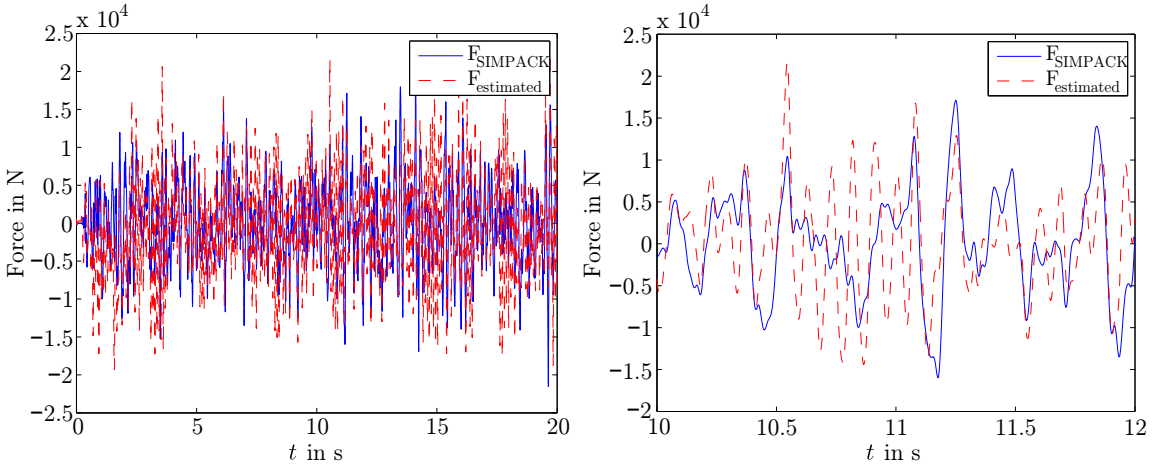

(a) Maximums of vertical contact force

(b) PSD of vertical contact force

Figure 5.4: Comparison for the first wheelset, vertical forces

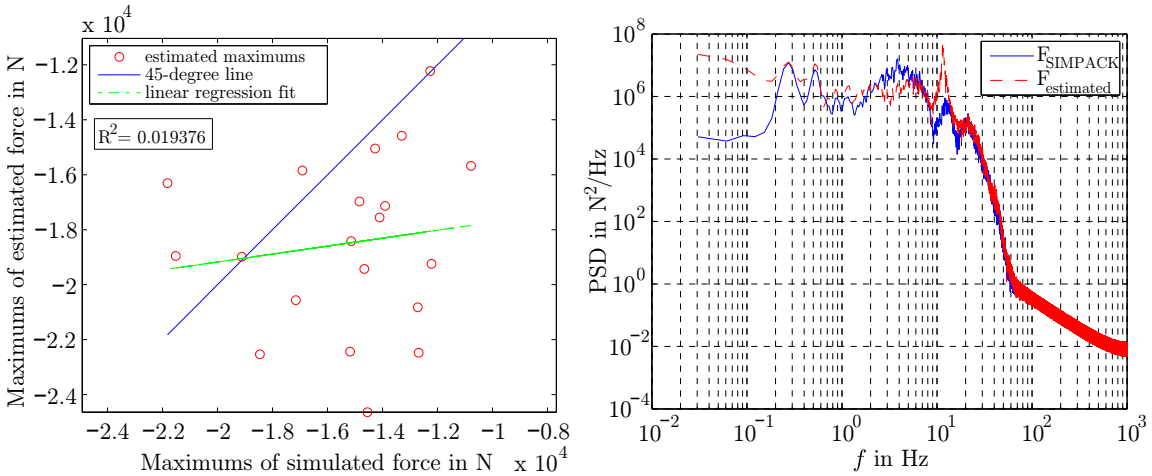
For the analysis in time domain an additional validation method is provided by comparison of the maximum forces of time segments, which are defined here with a constant length of 5 s (≙220 m of the corresponding rail track at a speed of 44 m/s). The linear regression fit of the maximums of the vertical contact forces occurring at the first wheelset is given in Figure 5.4a as well as the coefficient of determination $R^2$ of the maximums (see appendix A).

In terms of frequency domain analysis a PSD comparison of the simulated and the estimated forces is chosen. Figure 5.4b shows a good agreement of the forces for the first wheelset in frequency domain.




(a) Lateral contact force (b) Lateral contact force, enlarged

Figure 5.5: Resultant lateral contact forces for the first wheelset




(a) Maximums of lateral contact force (b) PSD of lateral contact force

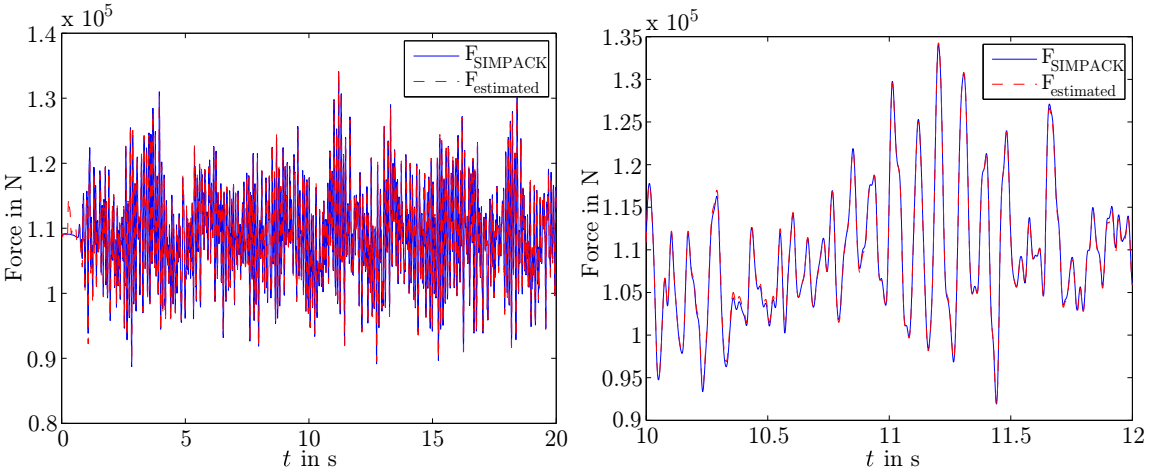
Figure 5.6: Comparison for the first wheelset, lateral forces

There are significant nonlinear effects in the SIMPACK simulation because of the contact point geometry and tangential force calculation, as described in section 4.1.4. This causes differences between the SIMPACK-model and the estimation-model, which is based on a

linearized model. Especially in lateral direction, the influence of non-linearity is significant, hence it is not unexpected that the estimated results are not as good as in vertical direction.

A comparison of the simulated and estimated contact forces in lateral direction for the first wheelset is given in Figures 5.5a and 5.5b.

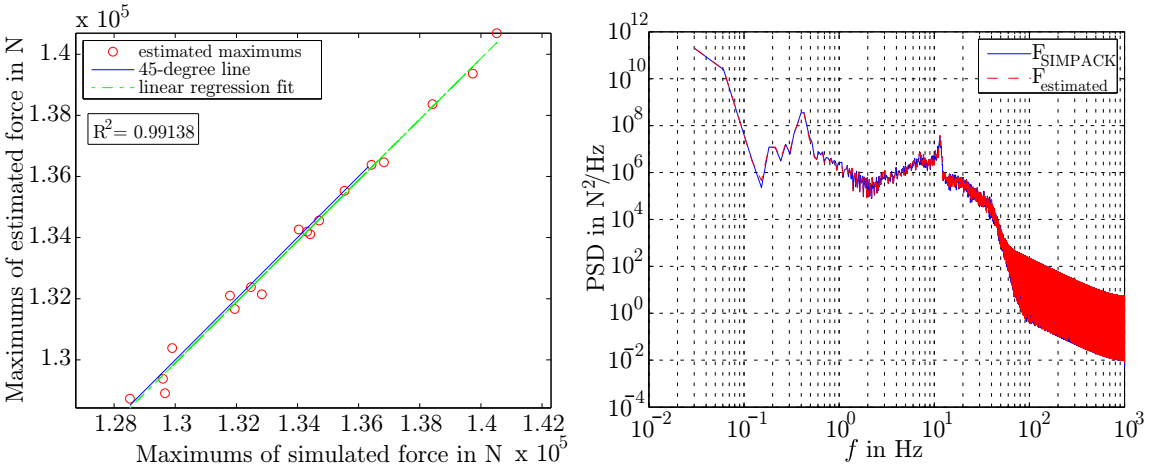
Differences of the resulting lateral contact force acting on the first wheelset in time domain can be especially observed through inspection of the maximum comparison, given in Figure 5.6a. The PSD analysis in Figure 5.6b compares the simulated and estimated lateral forces in frequency domain.




(a) Vertical contact force (b) Vertical contact force, enlarged

Figure 5.7: Resultant vertical contact forces for the fourth wheelset

The estimation results for the first wheelset are not as good as for the other wheelsets whereas the best estimation results are obtained for the fourth wheelset. A reliable reason for this observation cannot be given at this point and would require further investigations.



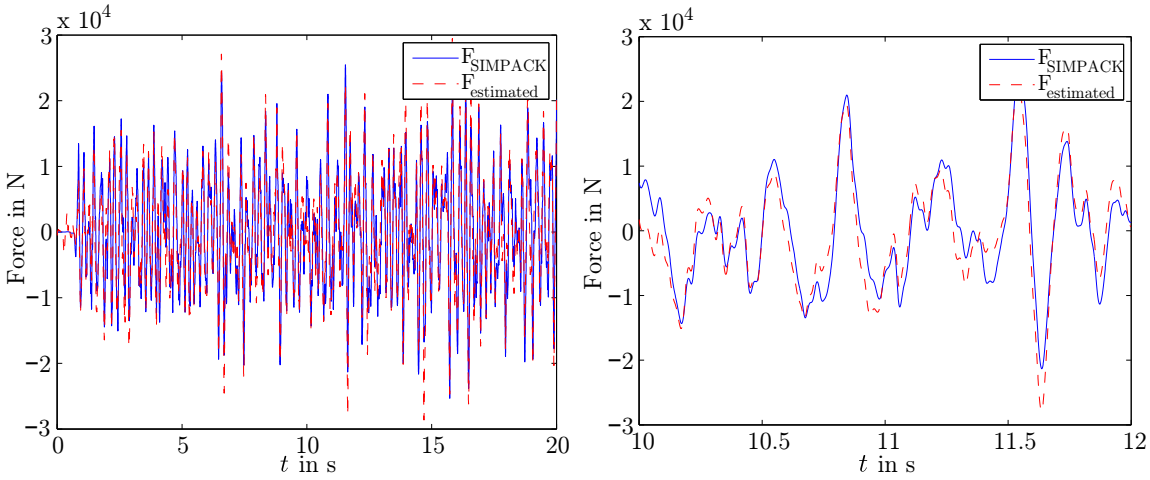

(a) Maximums of vertical contact force (b) PSD of vertical contact force

Figure 5.8: Comparison for the fourth wheelset, vertical forces

Figures 5.7a and 5.7b show the time domain comparison of the resultant force in vertical direction for the fourth wheelset. It can be seen that the correlation between the estimated

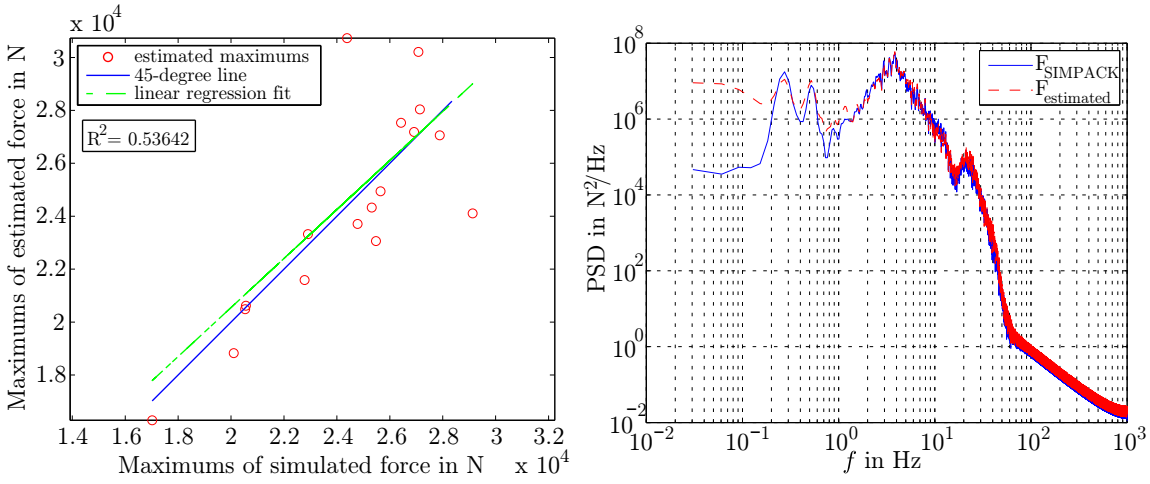
and simulated forces is better as for the first wheelset. This can also be detected by comparing the maximums of the forces, shown in Figure 5.8a, where the linear regression fit of the estimated and the simulated force maximums approves the considerable correlation in vertical direction. The PSD analysis of the vertical force for the fourth wheelset is given in Figure 5.8b and also shows good agreement of the results from simulation and estimation in frequency domain.

For the contact forces acting on the fourth wheelset in lateral direction the comparison of the simulated and estimated results is given in Figure 5.9a. Differences of the resultant lateral contact force acting on the fourth wheelset in time domain can be observed through inspection of the enlarged view, which is shown in Figure 5.9b.




(a) Lateral contact force (b) Lateral contact force, enlarged

Figure 5.9: Resultant lateral contact forces for the fourth wheelset



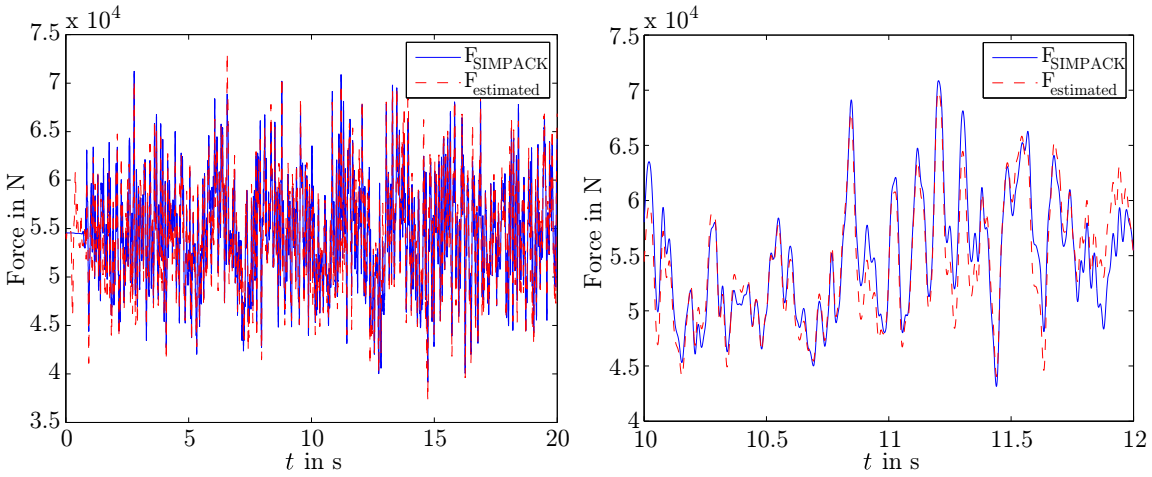

(a) Maximums of lateral contact force (b) PSD of lateral contact force

Figure 5.10: Comparison for the fourth wheelset, lateral forces

The force maximums in lateral direction, see Figure 5.10a, show not as good correlation of the estimated and simulated forces acting on the fourth wheelset as in vertical direction.

However, correlation is still given and thus the application of the estimation method also for the lateral direction can be regarded as suitable.

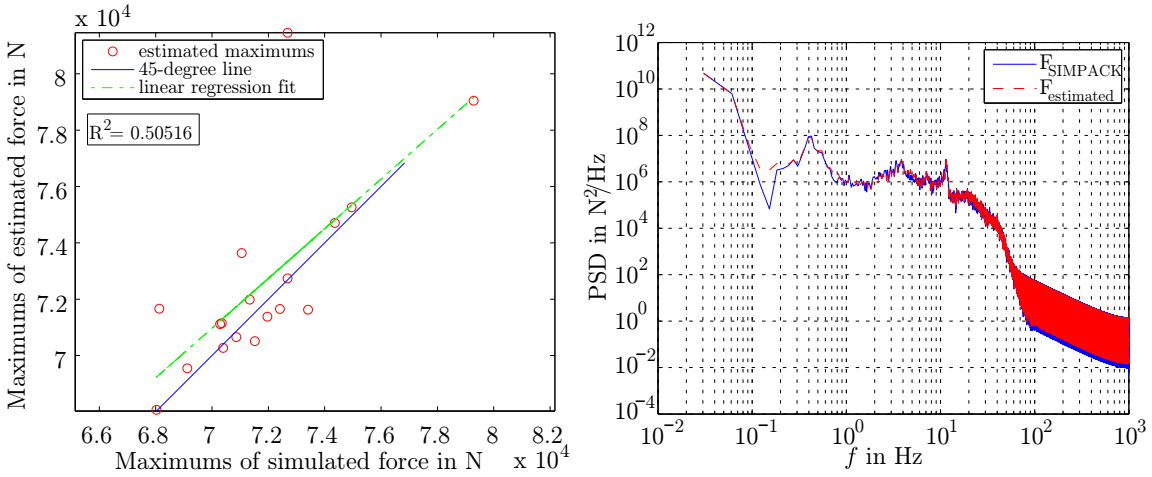
The presented method also provides a distinction between the forces applied on the left and right wheels of the wheelsets without restriction to resultant force estimation, as described in section 4.2.3. As an example, the estimated vertical wheel-rail contact force of the right wheel (direction of view is the positive $x$-direction) of the fourth wheelset is compared to the simulated result in Figures 5.11a and 5.11b.




(a) Vertical contact force

(b) Vertical contact force, enlarged

Figure 5.11: Vertical contact forces the right wheel of wheelset 4




(a) Maximums of vertical contact force

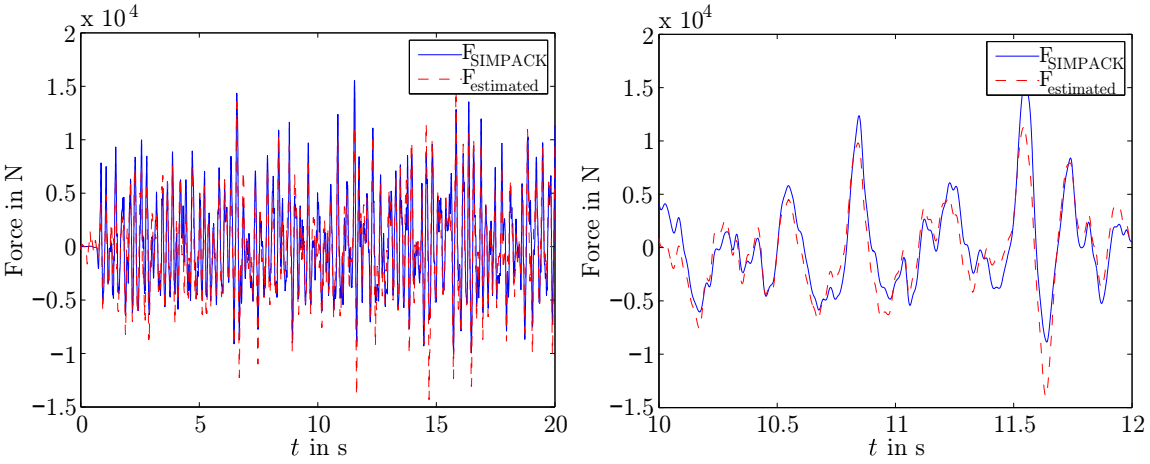
(b) PSD of vertical contact force

Figure 5.12: Comparison for the right wheel of wheelset 4, vertical forces

In Figure 5.12a the comparison of the force maximums is given and shows that the correlation is not as good as for the resultant forces (see Figure 5.8a). The PSD comparison of the single vertical force on the right side of the fourth wheelset is shown in Figure 5.12b and underlines the accordance of the estimated and the simulated force in frequency domain.

Accordingly, the time domain results of the estimated and simulated force acting in

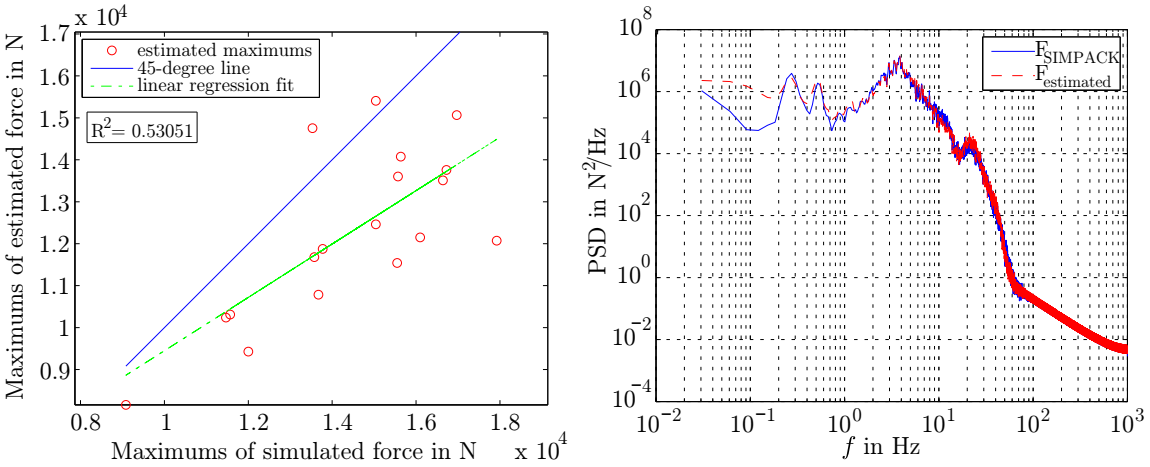
lateral direction on the right wheel of the fourth wheelset are given in Figures 5.13a and 5.13b.




(a) Lateral contact force (b) Lateral contact force, enlarged

Figure 5.13: Lateral contact forces for the right wheel of wheelset 4

The comparison of the maximums of the simulated and the estimated forces is shown in Figure 5.14a and underlines the assumptions from above that the correlation in lateral direction is not as good as in vertical direction. The frequency domain results, given by Figure 5.14b, show differences in the spectral characteristics for very low frequencies.




(a) Maximums of lateral contact force (b) PSD of lateral contact force

Figure 5.14: Comparison for the right wheel of wheelset 4, lateral forces

Similar to previous results, the estimation of forces in lateral direction on single sides of the wheelsets is not as accurate as in vertical direction because of the higher influence of non-linear factors. However, the results are still correlating.

In general, the estimation of resultant forces provide better results, which can be quantified by comparing the coefficients of determination considering all sampling points for contact forces acting on the right wheel in lateral ($R^2_{lat-r} = 0.82683$) and vertical ($R^2_{vert-r} = 0.96591$) direction with the values for their corresponding resultant forces ($R^2_{lat} = 0.85986$,

$R^2_{ver} = 0.99824$). Although the estimation is not as accurate for a single side, the correlation can be regarded as significant, especially in vertical direction.

# 6 Conclusion

An estimation method using a Kalman filter algorithm for the inverse determination of lateral and vertical wheel-rail contact forces was introduced. The method was applied on a railway vehicle system based on a model from a SIMPACK simulation.

It was shown that inverse problems can be classified as a deconvolution or a system identification problem. Within this work a deconvolution, the identification of inputs on the basis of output data with knowledge about the system, was the task to be accomplished. Consequently, four methods for the inverse identification of unknown input were presented. The regarded approaches were applied on a simple model of a one-dimensional two-mass system in order to estimate an unknown excitation quantity with knowledge about the system characteristics and acceleration data. Good estimation results were obtained from two methods which are a parity space approach and a method based on a Kalman filter algorithm. The best fitting estimates compared to a given signal were obtained by the application of the Kalman filter algorithm. As a result, this method was selected for the estimation of unknown input forces of the more complex system of a railway vehicle.

Before the inverse problem solving task for a railway vehicle system could have been discussed, the mechanical concepts of multibody simulation with application to railway vehicles were presented. Especially random track irregularities are important to consider for the multibody simulation as they excite the vehicle and lead to unknown input forces. Since the input forces are transmitted through the contact patches between wheels and rails, a main problem in terms of modeling of a railway vehicle is given by the wheel-rail contact. Hence, several approaches for the mathematical description of the wheel-rail contact geometry were given. After that, the modeling of the contact forces could have been introduced. Finally, all components needed for the mechanical modeling of a railway vehicle system were described.

The main goal of this work was the inverse force identification of a railway vehicle system based on a SIMPACK simulation. Hence, a statement of the given inverse problem had to be formulated as the identification of unknown wheel-rail contact forces on basis of acceleration measurements with assumption that the system characteristics are known. Acceleration data from the simulation instead of real measurement data were taken alternatively. Afterwards the SIMPACK model was linearized before the system matrices were exported in order to be implemented in the estimation process. The system was extended by a form filter element to obtain more accurate estimation results. The form filter was modeled on the basis of power spectral density polynomials of assumed random track irregularities. After state estimation by the use of a Kalman filter, wheel-rail contact forces have been calculated from generated acceleration data. The identification method is not restricted to the estimation of resultant forces acting on a wheelset. Moreover, a distinction between the forces occurring on the left and on the right wheel of the single wheelsets can be made.

The comparison of the results with the SIMPACK simulation show significant correlations and motivates further investigation of this approach. Especially in vertical direction good agreement between the estimated and the simulated forces can be detected. Although the correlation in lateral direction is not as good as in vertical direction, the applied method still can be regarded as suitable for the solution of the given inverse problem.

# A Simple linear regression

The *simple linear regression* model is a useful tool to quantify the dependence of two random variables $x$ and $y$. More about the characteristics and applications of this model can be found in [24] and [34]. This chapter presents the derivation of a measure called the *coefficient of determination* which is used in this work for the evaluation of estimation results.

A residual $r_i$ for $n$ data points $x_i$ and $y_i$ is given by the following:

$$r_i = y_i - \hat{y}_i \qquad \text{for} \quad i = 1, \ldots, n \tag{A.1}$$

with the mean function $\hat{y}_i$ which is defined as:

$$\hat{y}_i = \beta_0 + \beta_1 x_i \tag{A.2}$$

with the intercept $\beta_0$ and the slope $\beta_1$ which are unknown coefficients.

The discussed method chooses the parameters $\beta_0$ and $\beta_1$ to minimize the so-called *residual sum of squares* $SS_{resid}$. The derivation of the least-squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ can be found e.g. in [34]. The residual sum of squares is given by:

$$SS_{resid} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{A.3}$$

For the calculation of the *coefficient of determination* $R^2$ the so-called *regression sum of squares* $SS_{reg}$ and the *total sum of squares* $SS_{total}$ have to be introduced as follows:

$$SS_{reg} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{A.4}$$

$$SS_{total} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{A.5}$$

where $\bar{y}$ is the mean of $y$.

The coefficient of determination $R^2$ can then be obtained by:

$$R^2 = \frac{SS_{reg}}{SS_{total}} = 1 - \frac{SS_{resid}}{SS_{total}} \tag{A.6}$$

and represents the goodness of fit of the regression line $\hat{y}$ to the data $y$. $R^2$ takes a value between 0 and 1 where $R^2 = 1$ denotes a perfect fit.

# B Model parameters

A multibody simulation of the railway vehicle system was performed with SIMPACK in order to obtain acceleration data. The model can be split into bodies and force elements. The corresponding parameters used within this work are shown in Tables B.1 and B.2.

| Bodies | Parameter | Description | Value |
|---|---|---|---|
| 1 Carbody | $m_{Car}$ | body mass | 32000 kg |
| | $I_{xx_{Car}}$ | moment of inertia about $x$ | 56800 kgm$^2$ |
| | $I_{yy_{Car}}$ | moment of inertia about $y$ | 19700000 kgm$^2$ |
| | $I_{zz_{Car}}$ | moment of inertia about $z$ | 19700000 kgm$^2$ |
| 2 Bogie frames | $m_{Bogie}$ | body mass | 2615 kg |
| | $I_{xx_{Bogie}}$ | moment of inertia about $x$ | 1722 kgm$^2$ |
| | $I_{yy_{Bogie}}$ | moment of inertia about $y$ | 1476 kgm$^2$ |
| | $I_{zz_{Bogie}}$ | moment of inertia about $z$ | 3076 kgm$^2$ |
| 2 Bolsters (dummy) | $m_{Bolster}$ | body mass | $10^{-6}$ kg |
| | $I_{xx_{Bolster}}$ | moment of inertia about $x$ | $10^{-6}$ kgm$^2$ |
| | $I_{yy_{Bolster}}$ | moment of inertia about $y$ | $10^{-6}$ kgm$^2$ |
| | $I_{zz_{Bolster}}$ | moment of inertia about $z$ | $10^{-6}$ kgm$^2$ |
| 4 Wheelsets | $m_{WS}$ | body mass | 1813 kg |
| | $I_{xx_{WS}}$ | moment of inertia about $x$ | 1120 kgm$^2$ |
| | $I_{yy_{WS}}$ | moment of inertia about $y$ | 112 kgm$^2$ |
| | $I_{zz_{WS}}$ | moment of inertia about $z$ | 1120 kgm$^2$ |

Table B.1: Body parameters of the railway vehicle system simulated in SIMPACK

| Elements | Parameter | Description | Value |
|---|---|---|---|
| 2 Traction rods | $c_{pTR}$ | linear spring stiffness | 5000000 N/m |
| | $c_{sTR}$ | serial linear spring stiffness | 10000000 N/m |
| | $d_{sTR}$ | serial linear damping | 25000 Ns/m |
| 8 Primary springs | $c_{sx_{PS}}$ | serial linear spring stiffness | 60000000 N/m |
| | $c_{sy_{PS}}$ | serial linear spring stiffness | 7500000 N/m |
| | $c_{px_{PS}}$ | parallel linear spring stiffness | 31391000 N/m |
| | $c_{py_{PS}}$ | parallel linear spring stiffness | 3884000 N/m |
| | $c_{pz_{PS}}$ | parallel linear spring stiffness | 1220000 N/m |
| | $d_{sx_{PS}}$ | serial linear damping | 15000 Ns/m |
| | $d_{sy_{PS}}$ | serial linear damping | 2000 Ns/m |
| 8 Primary dampers | $c_{s_{PD}}$ | serial linear spring stiffness | 600000 N/m |
| | $d_{s_{PD}}$ | serial linear damping | 4000 Ns/m |
| 4 Secondary springs | $c_{x_{SS}}$ | linear shear stiffness | 160000 N/m |
| | $c_{y_{SS}}$ | linear shear stiffness | 160000 N/m |
| | $c_{z_{SS}}$ | linear vertical stiffness | 430000 N/m |
| | $c_{\varphi_{PS}}$ | linear roll bending stiffness | 10500 Nm/rad |
| | $c_{\psi_{PS}}$ | linear pitch bending stiffness | 10500 Nm/rad |
| 4 Secondary dampers (lateral) | $c_{s_{SD_y}}$ | serial linear spring stiffness | 6000000 N/m |
| | $d_{s_{SD_y}}$ | serial linear damping | 32000 Ns/m |
| 4 Secondary dampers (vertical) | $c_{s_{SD_z}}$ | serial linear spring stiffness | 6000000 N/m |
| | $d_{s_{SD_z}}$ | serial linear damping | 20000 Ns/m |
| 2 Anti-roll bars | $c_{\alpha_{AR}}$ | bending stiffness about x | 940000 Nm/rad |

Table B.2: Force elements of the railway vehicle system simulated in SIMPACK

# List of Figures


2.1 Block diagram of a dynamical system [18] . . . . . 10
2.2 LTI system . . . . . 11
2.3 Block diagram of the state-space model [18] . . . . . 12
2.4 Aliasing effect . . . . . 13
2.5 Probability density function of a Gaussian RV for $\sigma_X = 1$ . . . . . 20
2.6 Second-order low-pass filter . . . . . 24

3.1 One-dimensional two-mass system . . . . . 29
3.2 LTI system with known and unknown input . . . . . 30
3.3 LTI system with a Kalman filter . . . . . 32
3.4 State-space model with only unknown input . . . . . 34
3.5 Generated pseudorandom input signal . . . . . 42
3.6 Results for the Kalman filter method . . . . . 43
3.7 Results for the parity space approach (PSA) method . . . . . 43
3.8 Regularized least squares solution (RLSS) results . . . . . 44
3.9 Results for the inverse dynamics method (IDM) . . . . . 44
3.10 Comparison of inverse estimation methods . . . . . 45

4.1 Railway vehicle system with 2 bogies . . . . . 47
4.2 Serial spring-damper element . . . . . 48
4.3 Absolute and trajectory coordinates . . . . . 49
4.4 Wheel and rail profile geometries . . . . . 52
4.5 Wheel-rail contact . . . . . 54
4.6 Track irregularities . . . . . 56
4.7 Front view of a bogie (simplified) . . . . . 57
4.8 Reduction of the SIMPACK model . . . . . 58
4.9 Applied contact forces . . . . . 61

5.1 Positions of sensors on the railway vehicle . . . . . 65
5.2 Model of form filter and system . . . . . 66
5.3 Resultant vertical contact forces for the first wheelset . . . . . 71
5.4 Comparison for the first wheelset, vertical forces . . . . . 71
5.5 Resultant lateral contact forces for the first wheelset . . . . . 72
5.6 Comparison for the first wheelset, lateral forces . . . . . 72
5.7 Resultant vertical contact forces for the fourth wheelset . . . . . 73
5.8 Comparison for the fourth wheelset, vertical forces . . . . . 73
5.9 Resultant lateral contact forces for the fourth wheelset . . . . . 74
5.10 Comparison for the fourth wheelset, lateral forces . . . . . 74

5.11 Vertical contact forces the right wheel of wheelset 4 . . . . . 75
5.12 Comparison for the right wheel of wheelset 4, vertical forces . . . . . 75
5.13 Lateral contact forces for the right wheel of wheelset 4 . . . . . 76
5.14 Comparison for the right wheel of wheelset 4, lateral forces . . . . . 76

# List of Tables

3.1 Model parameters of the two-mass system . . . . . 42
3.2 Parameters for the different estimation methods . . . . . 42

4.1 Coefficients for low and high irregularities . . . . . 57
4.2 Dimensions of vectors and matrices before and after reduction . . . . . 61

5.1 DOFs of the simulation model . . . . . 65

B.1 Body parameters of the railway vehicle system simulated in SIMPACK . . . 81
B.2 Force elements of the railway vehicle system simulated in SIMPACK . . . . 82

# Bibliography


[1] EN 14363:2005, Railway applications - Testing for the acceptance of running characteristics of railway vehicles - Testing of running behaviour and stationary tests.

[2] EN 15302:2008, Railway applications - Method for determining the equivalent conicity.

[3] Richard C. Astor, Brian Borchers, and Clifford H. Thurber. *Parameter Estimation and Inverse problems*. Elsevier Academic Press, 2005.

[4] Richard Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, USA, 1 edition, 1957.

[5] F.W. Carter. On the action of a locomotive driving wheel. *Proc. R. Soc. Lond.*, A 112:151–157, 1926.

[6] Fritz Frederich. Die Gleislage - aus fahrzeugtechnischer Sicht. *ZEV-Glas. Ann. 108*, 12:355–362, 1984.

[7] Anna Hagenblad, Fredrik Gustafsson, and Inger Klein. A comparison of two methods for stochastic fault detection: The parity space approach and principal components analysis. In *13 th IFAC Symposium on system identification*, 2003.

[8] S. Iwnicki. Simulation of wheel-rail contact forces. *Fatigue & Fracture of Engineering Materials & Structures*, 26:887–900, 2003.

[9] Simon Iwnicki. *Handbook of Railway Vehicle Dynamics*. CRC Press, 2006.

[10] Mathias Jesussek and Katrin Ellermann. Fault detection and isolation for a nonlinear railway vehicle suspension with a hybrid extended kalman filter. *Vehicle System Dynamics*, 51:1489–1501, 2013.

[11] K.L. Johnson. *Contact mechanics*. Cambridge University Press, 1985.

[12] J. J. Kalker. The computation of three-dimensional rolling contact with dry friction. *International Journal for numerical methods in engineering*, 14:1293–1307, 1979.

[13] J.J. Kalker. A fast algorithm for the simplified theory of rolling contact. *Vehicle System Dynamics*, 11:1–13, 1989.

[14] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering*, 82 (Series D):35–45, 1960.

[15] Siegfried Kessel and Dirk Fröhling. *Technische Mechanik - Engineering Mechanics*. Vieweg+Teubner Verlag, 2012.

[16] J. Klingel. Über den Lauf von Eisenbahnwagen auf gerader Bahn. *Organ für die Fortschritte des Eisenbahnwesens*, Neue Folge 20:113–123, 1883.

[17] Klaus Knothe and Sebastian Stichel. *Schienenfahrzeugdynamik*. Springer Berlin Heidelberg, 2003.

[18] Jan Lunze. *Regelungstechnik 1*. Springer Berlin Heidelberg, 2013.

[19] Loren D. Lutes and Shahram Sarkani. *Random Vibrations*. Butterworth-Heinemann, 2004.

[20] MATLAB. *MATLAB Documentation - Release R2013a*. The MathWorks Inc., Natick, Massachusetts, 2013.

[21] Jesus Otero Yugat, Jordi Martinez Miralles, and Maria De los Santos. Analytical model of wheel-rail contact force due to the passage of a railway vehicle on a curved track. *Revista Facultad de Ingenieria Universidad de Antioquia*, pages 135 – 144, 12 2009.

[22] K. Popp, I. Kaiser, and H. Kruse. System dynamics of railway vehicles and tracks. *Archive of Applied Mechanics*, 72:949–961, 2003.

[23] A. Rao. *Dynamics of Particles and Rigid Bodies: A Systematic Approach*. Cambridge University Press, 2006.

[24] Alvin C. Rencher and G. Bruce Schaalje. *Linear Models in Statistics*. John Wiley & Sons, Inc., 2 edition, 2008.

[25] Hamed Ronasi and Jens C.O. Nielsen. Inverse identification of wheel-rail contact forces based on observation of wheel disc strains: an evaluation of three numerical algorithms. *Vehicle System Dynamics*, 51(1):74–90, 2013.

[26] J. Carlos Santamarina and Dante Fratta. *Discrete Signals and Inverse Problems*. John Wiley & Sons, Inc., 2005.

[27] Herbert Schlitt. *Systemtheorie für stochastische Prozesse*. Springer Berlin Heidelberg, 1992.

[28] Ahmed A. Shabana, Khaled E. Zaazaa, and Hiroyuki Sugiyama. *Railroad Vehicle Dynamics: A Computational Approach*. CRC Press, 2007.

[29] Dan Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. John Wiley & Sons, Inc., 2006.

[30] SIMPACK. *SIMPACK Documentation - Release 9.5*. SIMPACK AG, Germany, 2013.

[31] R.F. Stengel. *Optimal Control and Estimation*. Dover Books on Mathematics. Dover Publications, 2012.

[32] Tadeusz Uhl. The inverse identification problem and its technical application. *Archive of Applied Mechanics*, 77(5):325–337, 2007.

[33] Edwin A.H. Vollebregt, Christoph Weidemann, and Andreas Kienberger. Use of "contact" in multi-body vehicle dynamics and profile wear simulation: Initial results. In *22nd International Symposium on Dynamics of Vehicles on Roads and Tracks*, 2011.

[34] Sanford Weisberg. *Applied Linear Regression*. John Wiley & Sons, Inc., 2005.

[35] Greg Welch and Gary Bishop. An introduction to the kalman filter. Technical Report 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, 1999.

[36] Peter D. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73, 1967.

[37] Tao Zhu, Shoune Xiao, Guangwu Yang, Weihua Ma, and Zhixin Zhang. An inverse dynamics method for railway vehicle systems. *Transport*, iFirst:1–8, 2013.

[38] Tao Zhu, Shoune Xiao, Guangwu Yang, Weihua Ma, and Zhixin Zhang. The inverse identification theory and application to high-speed trains. *Advanced Science Letters*, 19:1582–1586, 2013.