MASTER

Identification and control of the EMPAct CVT

Meesters, J.A.C.

Award date:
2006

Link to publication
Identification and control of the EMPAct CVT

J.A.C. Meesters

DCT 2006.078

Master’s thesis

Committee:  ir. T.W.G.L. Klaassen
dr. P.A. Veenhuizen
dr. ir. A.A.H.Damen
prof. dr. ir. M. Steinbuch

Coach:  ir. T.W.G.L. Klaassen

Supervisor:  prof. dr. ir. M. Steinbuch

Technische Universiteit Eindhoven
Department Mechanical Engineering
Dynamics and Control Technology Group

Eindhoven, June, 2006
Chapter 1

Preface

Part of the curriculum of Mechanical Engineering at the TU/e is a traineeship of about three months. I did mine abroad in Mexico in the end of 2003. Before going there I really had the opinion that people living in poverty did so because they chose to themselves. After my stay in Mexico I was unable to defend this statement. Sure, people who claim to be poor in the Netherlands choose not to educate and evolve themselves, but this is generally not the case in 'developing' countries.

After my traineeship, I went to my professor for an assignment for a master project, at the exact date of the 29th of January, 2004. Normally a Master Project takes about 9 months. Looking at the date on the title page of this report, there is only one conclusion left. Something went wrong.

During the full length of this project I have had motivational problems. I first thought it had something to do with the subject of the project, being in the automotive atmosphere while I personally never had, don’t have, and most probably never will have interest in the automotive world. At the time I took the assignment, I thought it was a good opportunity to apply my knowledge of control technology to an experimental setup that did not belong to the family of robots that don’t have any function other than being educational for students.

Somewhere along the way, there was an opportunity to get a position as a control technology expert in a company that builds pick and place machines. I took a shot and was the lucky guy, thereby obtaining a fulltime job while I was not yet graduated. Although in reality this job had nothing to do with control technology, it did bring me to Taiwan for the period of a month. As Mexico showed me that poverty is not always (even mostly not) a choice, Taiwan showed me a glimpse at the art of suppression. After half a year I finally quit this job because it did not contain any of the aspects the recruiters had promised me. Moreover I discovered that it was almost impossible for me to graduate while having a fulltime job.

Upon returning at the TU/e, I expected to be highly motivated to finally finish my master project, but this was apparently not the case. In fact I was so highly unmotivated that from time to time I just did not show up for periods that occasionally lasted up to two weeks. During these periods I was mostly at home or in the library reading certain books, articles, and internet sites that provided some more information on topics like poverty, history and suppression. Being a control technology student, I am particularly baffled by the countless
occasions during modern history in which western governments (including our own) have been able to use the media as a controller for keeping large masses of people quiet. If there is a constant in modern history, it has to be the fact that by the proper use of misinformation, certain persons can get away with just about anything. Misinformation is a rather tricky concept. It does not contain false facts, because then, if the errors would come out, the information would not be taken serious anymore (for instance the fact that Iraq was producing weapons of mass destruction). On the contrary, the concept of misinformation exists in giving the kind of information that does not matter over, and over, and over again. In the end the mass that consumes the misinformation, thinks it is informed, but actually does not know anything. And, above all, does not want to know more because it thinks it already knows what is going on.

While viewing upon society as a linear system (which is a rather big assumption), any control technology student should know that there is a limit to the amount of disturbance the small group can attenuate before stability issues start playing a role. Trying to attenuate too much disturbance will inevitably cause instability. History has shown over and over that large amounts of people always reach a certain point in which they just can’t stand the suppression anymore, and start organizing a revolution.

Looking at history, it almost seems normal that large groups of people are suppressed by a small group of people for the welfare of the latter group. There is (at least) one difference between today’s suppression and that of the past. Although there almost always has been a group in between the small ruling and the large suffering group, this group in between (which I’m a member of) is nowadays thinking it is helping the large suffering group by donating money to several institutes. In reality however, the in-between group has as role model the small group, and the donation of less than one promille of the yearly national budget to development projects is mostly consumed by people who make large amounts of money on giving small amounts of aid to people in need of it. The general idea that poverty and suppression, that most of the inhabitants of this world have to deal with every day, is an unsolvable problem, and that it is therefore probably best to live your own life and try to obtain a certain amount of welfare and happiness for yourself and your family, is not only preventing up to thousands of people from really helping the suffering group, but in fact it is also actively keeping the small group in charge, and the large group in suffering. In combination with the possibility to obtain any kind of information one desires, be it in the library, internet, newspaper, radio, even sometimes television, it is a remarkable piece of control technology that is used by the small group to control the minds of the many. It might even be capable to turn the linear system (society) into a passive system, and hence remove the theorem that an attempt to attenuate too much disturbances will cause instability.

During the reading of various articles, books and internet sites, I must have come up with one hundred excuses for why my master project was not progressing. My advisor accepted them all, while probably thinking I was just lazy, which might be partly true. For his support and giving me the ability to do my own things, while he himself is under a large amount of time pressure, I’m very thankful. Thank you Tim. I also would like to take this opportunity to thank my professor Maarten Steinbuch. I remember that after being completely absent from the project for more than 2 weeks, he called me in to report for a motivational discussion. Instead of giving me a burnout, he first listened to me, then gave his opinion, and even handed me some suggestions. If there is any reason at all to sketch the above issue, then it would
have to be an attempt to explain to them one of the things that have been occupying my mind lately, and therefore was probably the true reason for the lack of my motivation.

Besides these two persons, I’m also thankful to some other persons. Among them are Bram Veenhuizen, Erwin Meinders, and the students that were (a year (or a bit more) ago) and are a member of the automotive group. Thanks to Bram for inviting me to the meetings and generating discussions about automotive topics, and thanks to Erwin for convincing me to quit my fulltime job with which I was not happy at all. I’m thankful to the students because of a nice working atmosphere and because their knowledge of automotive technology is much broader than mine, so I learned a lot from them.

The remainder of this report is about the identification and control of a CVT. With a CVT it is possible to operate the Internal Combustion Engine of a car in a more efficient working point. Nowadays several CVT’s are sold with a stepped shifting strategy. In this way the efficiency increase is lost, but the product apparently sells better. One day prior to the writing of this preface, there was a spokesman of the company that holds the patent for the pushbelt that is used in the CVT, claiming on national television that Dutch society should not worry that the production of the pushbelt will go to China in the nearby future, because the production process is too complex and therefore needs to stay in the Netherlands. This television appearance was made during the period when there was a possibility of the closing of NedCar, and therefore a general fear for loss of jobs in the automotive industry. Already more than one year ago the Dutch institute for engineers (the KIVI) published in an article in their magazine that at least one million pushbelts were produced in South-East Asia each year (by the same company the aforementioned spokesman works for (with the help of a partner)), and that they attempted to increase production.
## Contents

1 Preface .......................................................... 1

2 Introduction ..................................................... 3
   2.1 The CVT principle ........................................ 4
   2.2 The EMPAct CVT ......................................... 8

3 Simulation Model ............................................... 10
   3.1 Geometries and definitions .............................. 10
   3.2 Belt models .............................................. 12
      3.2.1 Quantification of the belt forces ............... 17
   3.3 Multi-body modelling ................................... 22
   3.4 Equilibrium at the pulleys ............................ 24

4 Identification .................................................. 27
   4.1 Step response estimation ............................... 27
      4.1.1 Results ............................................ 30
   4.2 Noise excitation estimation ............................ 34
   4.3 Comparison .............................................. 34

5 Control .......................................................... 36
   5.1 Sequential Loop Shaping ............................... 38
   5.2 Controller design for the simulation model .......... 40
   5.3 Controller validation ................................... 47

6 Implementation ................................................ 52
   6.1 The experimental setup ................................ 52
   6.2 Primary identification .................................. 54
   6.3 Secondary identification ............................... 57

7 Conclusions and recommendations ......................... 60
   7.1 Conclusions .............................................. 60
   7.2 Recommendations ....................................... 61

A Step response identification .................................. 64

B QFT Principles ................................................ 67
Chapter 2

Introduction

In automotive applications, transmissions are used to match the Internal Combustion Engine (ICE) to road characteristics (uphill, downhill, wind, no wind, etc.). It is well known that although ICE’s are rather powerful, their working points are rather restrictive. If a certain engine, for example one with 50 [kW] of power, has to accelerate a car of say 1000 [kg] from standstill, some gears have to be placed in between the wheels and the ICE, because else the ICE cannot generate enough power in the low rpm (revolutions per minute) region to accelerate the car. The idea is then to let the engine operate at a much higher rpm, than the wheels. So the wheels demand a certain high-torque energy from the ICE, but the transmission makes it possible to convert this high-torque power to a high-frequency power, which the ICE has enough of.

In cars equipped with stepped manual or automatic transmissions, only a certain (4, 5 or sometimes 6) number of transmission ratios are available. Considering road and traffic characteristics, the load can contain all kinds of combinations between high torque/low speed and low torque/high speed. Figure 2.1 shows the characteristics of a commercially available ICE. The numbers inside this figure refer to the amount of fuel that is necessary to generate an amount of energy [g/kWh]. The S refers to the Sweet spot of the engine, which is the point where the highest efficiency can be achieved. Clearly, there are some high and low efficiency regions. Because cars with manual transmissions only have a limited number of gears, not all regions can be reached, which results in non optimal usage of the ICE.

The idea behind a Continuously Variable Transmission is to replace the limited number of gears in a stepped transmission, by a continuous set of gears. In this way, at least theoretically, every region of Figure 2.1 can be reached. This would mean that a significant amount of energy could be saved, (or created from a performance viewpoint) with respect to the manually (or automatically) stepped transmission. There is however one big disadvantage up till now. The CVT, as presented in the next section, uses a belt that is captured by two pulley sets. One pulley of each of these pulley sets can translate, the other pulley is fixed (see ahead at Figure 2.3 for a clarifying picture). By pushing one of the pulley sets closer to another, and pulling the other set out of each other, or vice versa, it is possible to create a continuous number of transmission ratios (as opposed to 4, 5, or 6 in a stepped transmission). The problem is that moving the pulleys costs a considerable amount of energy. Forces necessary to move the pulleys are in the order of 10 [kN], depending on the operating conditions. Based on these forces, oil pumps are placed to actuate the pulleys hydraulically. Suppose the car is cruising down the highway and an emergency break has to be enforced, then the hydraulic
2.1 The CVT principle

One way of transmitting power between two parallel shafts is by means of two sets of gears. The transmission efficiency $\eta_t$ of this type of transmission is usually very high (over 95 %). Transmission efficiency is defined as the output power $P_s$ divided by the input power $P_p$, or

$$\eta_t = \frac{P_s}{P_p} = \frac{T_s\omega_s}{T_p\omega_p}$$

(2.1)

where $T_s$ and $T_p$ are the torques at the output shaft and input shaft of the transmission, and $\omega_s$ and $\omega_p$ the rotational velocities of the output shaft and the input shaft. With this type of
transmission, almost no power is lost because no power is needed to operate the transmission. Therefore, the main power losses in this transmission are due to bearing losses. There are, however, also two main disadvantages with this type of transmission. First, there are only a limited number of transmission ratios. Second, during shifting, the power transmission from input to output is lost. The second disadvantage is clear, but to explain the first one, Figure 2.1 can be used. In this figure, it can be seen that the efficiency of the ICE (Internal Combustion Engine) is highly dependant on the amount of rpm (revolutions per minute) it is running at. To obtain the highest possible efficiency of the ICE, it is desirable to stay in the region marked with S in Figure 2.1, the Sweet spot. With a geartype transmission, it is not possible for the ICE to maintain in the Sweet spot working area. This means that, although the input to output efficiency of the transmission itself ($\eta_t$) is very high, the overall transmission efficiency (the amount of power that is available at the wheels of a car, divided by the amount of power that is theoretically available from the gasoline in the ICE, $\eta_o$), leaves room for improvement. For a schematic drawing that illustrates the difference between these two efficiencies, see Figure 2.2.

By using a CVT, the efficiency of the transmission itself, $\eta_t$, is lower (at least below 95 %), but due to the fact that power can be transmitted during shifting, and above all due to the fact that the Sweet spot can be reached at most times, the overall efficiency $\eta_o$, can be higher. At least theoretically. Theoretically, because up till now, commercially available CVT’s require relative high actuation forces $F_{act}$, (see Figure 2.2), in order to transmit the torque from engine to wheels, which cause the overall efficiency to be even lower than that of the fixed gear transmission. Note that for the manual fixed gear transmission, the actuation forces $F_{act}$ are generated by the person who is driving the car. In order to explain why these actuation forces in the CVT are so high that they significantly reduce the overall efficiency, the working principle of the CVT is described.

Figure 2.3 represents the part of the CVT that is called the variator. The variator consists of two pulley sets and a belt. One pulley of each of the pulley sets is able to translate. The belt is wrapped around both pulley sets. The purpose of the variator is to transmit power from the primary shaft to the secondary shaft. During the transmission of this power, it might be necessary to change the transmission ratio. This can be done by moving the pulleys of the primary side closer, while moving the pulleys from the secondary side away from each other, or vice versa. In the first case, the belt that is wrapped around the pulleys, will run at a larger radius on the primary side, than on the secondary side. This situation is referred to as OverDrive (OD). When the opposite is true, the primary pulleys are moving from each other and the secondary pulleys are coming closer to each other, the belt is running on a larger radius secondary, than at the primary side. This situation is referred to as Low. Both situations are depicted in Figure 2.3. The belt that is wrapped around the pulleys consists
of several hundred metal elements with rough edges, and several metal bands (rings). A cut open belt is depicted in Figure 2.4. The power is transmitted from the primary to the secondary side, through the belt, by means of friction. The amount of friction between the sides of the elements and the pulleys determines whether the power is transmitted successfully or not. Obviously, the moving of the pulleys requires a force. When this actuation force is applied, the bands in the belt will be stretched out a little bit, generating a tensile force $S$. The resulting force from this tension in the bands, is directed to the center of the pulley and acts on the elements. Both actuation force and resulting tensile force are schematically drawn in Figure 2.5. In this figure, the contact force between pulley and element is $dN_2$, and the contact force between the bands and the elements is $dN_1$. So in order to translate the primary pulley in the direction of $x_p$, a force $F_{act}$ is applied at the pulley, which results in a tensile force in the bands and a contact force $dN_2$ on the element. The resulting tensile force in the bands at its turn has a contact force with the elements $dN_1$. Hence, there is an equilibrium. The magnitude of $dN_1$ determines the friction force between the bands and the elements, while the magnitude of $dN_2$ determines the friction force between the pulley and the elements. If this friction force is not large enough, the elements will start to slip around the
pulleys, and part of the transmitted power is lost. In order to be able to prevent slip between elements and pulleys, there is a secondary actuation force at the secondary pulley which can make $dN_2$ higher if necessary. If too much slip occurs for a large amount of time, the sides of the elements will become less rough, resulting in permanent damage of the transmission. Slip of the elements around the pulley occurs more easily when the load at the secondary side of the variator is large. In order to prevent slip, a certain actuation force is necessary. When the load torque increases, the actuation force has to increase also to prevent slip. In practice the force that is actually applied to prevent slip, is even a factor 1.3 higher than the minimum necessary force to prevent slip. This is not improving the transmission efficiency. In order to reduce the actuation forces and thus improving the transmission efficiency, research has been done. When the secondary actuation force is high (to prevent slip), the primary actuation force is also high in order to control the ratio. By reducing the secondary actuation force, also the primary actuation force can be reduced, resulting in a significant increase of efficiency. In [1] and [18] it is shown that small amounts of slip do not cause damage to the transmission, but definitely do increase the efficiency. In fact, [1] shows that there are certain amounts of slip that guarantee a maximum efficiency. The measured relation between efficiency $\eta$ and slip is also depicted in Figure 2.6.

On basis of this research, slip control is implemented with a hydraulically actuated CVT. In [11] it is shown that slip control is possible, does indeed increase efficiency, and does not cause any damage. There is one problem however. When the load torque at the secondary side is small, the required secondary actuation force is also relatively low. In [11], a CK2 transmission is used, that uses a hydraulic actuation system for the movement of the pulleys. This hydraulically actuated CK2 transmission will be used as a reference in the following chapters. The hydraulic system provides a minimum pressure of 6 bar that is always active on the pulleys. When small loads are applied, this pressure is too high compared to the required force that is needed for slip control. In this situation it is not always possible to obtain those slip levels that produce maximum efficiency. Another problem is that the hydraulic actuation is linked to the ICE speed, hence at highway speed energy is waisted as a result of a large oil flow (at a minimum pressure of 6 [bar]). Also, having a hydraulic system, the stiffness of the actuation is not very high which leads to a rather low obtainable bandwidth of the controlled system. This is an undesired effect considering disturbance rejection.
In order to solve this problem, a CVT was designed, which uses an ElectroMechanical Pulley ACTuation (EMPACT) system. Instead of a hydraulic pump, this system uses electric servomotors to supply the actuation forces. On top of that the actuation forces are decoupled from the ICE speed. Using an electromechanical actuation a higher stiffness can be expected, so that besides the higher efficiency a higher bandwidth can be obtained. The variator stays exactly the same as in the hydraulic situation. Only the way in which the actuation forces are generated and applied to the pulleys is different.

### 2.2 The EMPAct CVT

For the EMPAct CVT, the actuation forces are generated by two electric servomotors. The EMPAct system uses four epicyclic gears in combination with two spindles to transmit the torques from the servomotors to the pulleys. A schematic drawing of an epicyclic gear is depicted in Figure 2.7. As can be seen from Figure 2.7, an epicyclic gear consists of a sun gear, a ring gear, a carrier and three planet gears. The planets are mounted on the carrier.
A schematic drawing of the EMPAct CVT is given in Figure 2.8. In this figure one servomotor is depicted as $M_p$ (the primary) and the other as $M_s$ (the secondary). Besides the two servomotors, two epicyclic gear sets can be distinguished. One epicyclic gear set at the primary side and one set at the secondary side. Each set (both at the primary and at the secondary side) has a left and a right epicyclic gear. At both the primary and the secondary side, the carriers of the left and right gear are coupled. Furthermore, the ring of the right epicyclic gear at the primary side is fixed to the transmission housing. The ring of the left epicyclic gear on the primary side is directly meshed with the ring of the right epicyclic gear on the secondary side. The primary servomotor $M_p$ is connected with the ring of the left primary epicyclic gear through a worm gear. By applying a torque with the primary servomotor, a difference in rotational velocity between the left and right primary sun gears is introduced. Because the left primary ring gear is meshed with the right secondary ring gear, a torque from the primary servomotor also generates a difference between the rotational velocities of the sun gears on the secondary side. This difference between sun gears is used on both sides in a spindle mechanism to generate a translation of the moveable pulleys. So by applying a torque with the primary servomotor, it is possible to move both moveable pulleys, and hence it is possible to shift.

However, the displacement of the primary movable pulley is not equal to the displacement of the secondary moveable pulley, and therefore there is a secondary servomotor to compensate for this effect. Another function of this secondary servomotor is to adjust the secondary clamping force, and thereby the slip in the variator. The secondary servomotor is coupled to the left secondary ring gear by means of a chain connection and a planetary reduction. To support the clamping forces, there are thrust bearings between the sun gears.
Chapter 3

Simulation Model

In this chapter the simulation model is discussed. The simulation model is a multi-body model, in combination with a model of the force distribution in the belt. The latter is based on a Coulomb friction model, see [2].

3.1 Geometries and definitions

In this section definitions and geometries of the variator are introduced. Figure 3.1 shows a schematic drawing of the variator. In this figure $R_p$ is the running radius of the belt at the primary pulley, and $R_s$ the running radius of the belt at the secondary pulley. Furthermore, $a$ is the distance between the two pulleysets which is 168 [mm] for this transmission. The length of the belt can be expressed as

$$L = 2acos(b) + R_p(\pi + 2b) + R_s(\pi - 2b)$$

(3.1)

with

$$b = \arcsin\left(\frac{R_p - R_s}{a}\right)$$

(3.2)

where $b$ is an angle that will be used further on in the distribution of belt forces. With this angle $b$, the angles at which the belt is wrapped around the pulleys can be expressed as

$$b_p = \pi + b$$

(3.3)

$$b_s = \pi - b$$

(3.4)

As can be seen from Eq. (3.1), the length of the belt depends on the belt’s running radii. For the running radii, also a relation exists with the translation of the pulleys, see Figure 3.2. $\theta$ is called the pulley wedge angle. In equation form this gives

$$x_p = 2\tan(\theta)(R_p - R_{\text{min}})$$

(3.5)

$$x_s = 2\tan(\theta)(R_s - R_{\text{min}})$$

(3.6)

where $x_p$ is the translation of the primary moveable pulley and $x_s$ the translation of the secondary moveable pulley. The constant $R_{\text{min}}$ represents the minimal radius that can be
CHAPTER 3. SIMULATION MODEL

Figure 3.1: Geometry of the variater.

Figure 3.2: Translation of the primary movable pulley.
obtained when the pulleys are at maximum distance from each other. Note that the moveable pulley on the secondary side is on the opposite side of the primary one and by definition a positive translation $x_s$ is in the negative direction of $x_p$.

Obviously, a very important quantity in power transmissions is the transmission ratio. In this work, ratio is defined in two ways. There is a geometric ratio $r_g$, and a speed ratio $r_s$. The geometric ratio is defined as

$$r_g = \frac{R_p}{R_s}$$  \hspace{1cm} (3.7)

whereas the speed ratio is defined as

$$r_s = \frac{\omega_s}{\omega_p}$$  \hspace{1cm} (3.8)

When operating the CVT, $r_g$ will obtain all kinds of values between $r_g = 0.43$ and $r_g = 2.25$. The situation in which $r_g \approx 0.43$ is called LOW, the situation in which $r_g \approx 1.0$ is called Medium, and when $r_g \approx 2.25$ the situation is called OverDrive (OD). Figure 2.3 clarifies the situations LOW and OD. Note that in Figure 2.3 only the variator is depicted. The difference between $r_g$ and $r_s$ accounts for the following definition of slip

$$\nu = 1 - \frac{r_s}{r_g}$$  \hspace{1cm} (3.9)

One of the metal elements of the belt (see Figure 2.4) is schematically drawn in Figure 3.3. With the use of this figure, some geometric properties can be defined. As can be seen, there are two radii given. First, $R_c$ represents the distance from the center of rotation of the pulleys to the rocking edge of the element. The rocking edge is the line around which the element tilts when turning around the pulleys. Second, $R_b$ represents the distance from the center of rotation of the pulleys to the centerline of the metal bands (the bands are situated on top of the shoulders of the elements). The distance between both radii is denoted as $\Delta R$. The distance between the rocking edge and the shoulder is given by $h$. Obviously the side angle of the element has to be the same angle as the wedge angle of the pulleys ($\theta = 11^\circ$).

### 3.2 Belt models

Depending on the way the variator is operated, several force distributions in the belt can be found. At all times, tensile forces act on the bands, while compressive forces act on the elements. In general, there are four different force models, the Low-Positive-Compression model
(LPC), the Low-Negative-Compression model (LNC), the OverDrive-Positive-Compression model (ODPC), and finally the OverDrive-Negative-Compression model (ODNC). The LPC model is depicted in Figure 3.5.

**Tension Forces**

As can be seen from this figure, the tensile force in the lower belt part is larger than the tensile force in the upper belt part. The same tensile force distribution can be found for the LNC model, see Figure 3.6. For the ODPC and the ODNC models however, the tensile force distribution is the opposite. In these two models, the largest tensile force is in the upper belt part, whereas the lower tensile force is in the lower belt part.

The reason for this difference of tensile force distribution is the influence of friction. Indeed had there been no friction at all, then the tensile forces would have been the same in the upper and lower part of the belt, independent of the transmission ratio. Due to the difference between $R_c$ and $R_b$, the bands should have a higher circumferential velocity than the elements at both pulleys, in order to maintain the same rotational velocity as the pulley. As can be seen from Figure 3.4, the circumferential velocities $v_c$ and $v_b$ of both the elements and bands are the highest at the smaller pulley radius.

Therefore it is more likely that the bands will start slipping with respect to the elements (which have the same rotational velocity as the pulley) at the smaller pulley radius. This slipping effect introduces a circumferential velocity difference between the bands and the elements, hence a friction force is generated. This friction force is opposite to the direction of rotation and therefore the tensile force will be reduced (from $S_1$ to $S_0$) along the pulley that is running at the lowest pulley ratio. Consequently, for the sake of continuity, the opposite mechanism has to take place at the pulley that is running at the larger pulley radius, hence at that pulley, the tensile force is built up (from $S_0$ to $S_1$) in the direction of rotation (see Figure 3.5). In principle, the decline from $S_1$ to $S_0$ takes place along the full angle of wrap of
the pulley that is running at the smallest pulley ratio. The building up from $S_0$ to $S_1$ takes place at the same magnitude of angle, but now at the pulley that is running at the largest pulley ratio. This means that at the pulley that is running at the largest pulley ratio, there is a part of the angle of wrap where the tensile force stays constant ($S_0$). This angle is called an idle tension arc, and is of magnitude $4b$. See the following subsections for the exact evolution of angles and forces. Concluding, for $r_g < 1.0$, the highest tensile force is in the lower belt part, while for $r_g \geq 1.0$, the highest tensile force is in the upper belt part.

Compression forces

This still leaves the distribution of compressive forces. In general there are two possibilities, i.e. the compressive forces act in the upper belt part, or in the lower belt part. In normal operation (not too high clamping forces), the compressive forces act in the upper belt part for positive input torques $T_p$, and in the lower belt part for negative input torques $T_p$. However, if the friction between bands and elements is larger than the friction between the elements and the pulleys (this is the case if very high clamping forces are used in comparison with the applied input torque $T_p$), the elements are stacked in the other belt part (the lower belt part for positive input torque, and the upper belt part for negative input torque), creating a compressive force in this other part of the belt. For the exact angles at which the compressive force is built up, and declined at both pulleys, see the following subsections, where each belt force model will be discussed briefly.

LPC model

The LPC model (see Figure 3.5) is used for positive input torques $T_p$, and ratios $r_g < 1$ with normal clamping forces. In this situation, both the compressive force and the difference in tensile force between the upper and lower part, transmit the torque from the primary to the secondary pulley. At the primary pulley, the tensile force declines using the full angle of wrap $b_p$. At this pulley also the compressive force builds up, starting at angle $y = b_p/2 - b_1$ with $Q = 0$, and ending at $y = b_p/2$ with $Q = Q_{\text{max}}$. Note that at a part of the primary angle of wrap the tensile force is already declining, while the compressive force is not yet building up.

At the secondary pulley, there is first an idle angle in which both forces stay constant for a length of $y = 4b$, then at $y = b_s/2 - b_p$, the tensile force starts to increase from $S_0$ to $S_1$. The final value of $S_1$ is achieved at $y = b_s/2$. The compressive force starts to decline, at angle $y = b_s/2 - b_2$, where it still has the value $Q_{\text{max}}$, and ends at angle $y = b_s/2$ with $Q = 0$. Also here, at a part of the secondary angle of wrap the tensile force already starts to build up, while the compressive force is not yet declining.

LNC model

The LNC model (see Figure 3.6) is used for two situations. The first one is when a negative torque $T_p$ is transmitted from the primary to the secondary pulley, for ratios $r_g < 1$. In this case, the tensile force is actually transmitting a positive torque, instead of a negative one, thereby reducing the efficiency of the transmission of the belt. The other situation is when a positive torque $T_p$ is transmitted from primary to secondary pulley, for ratios $r_g < 1$, but with relatively high clamping forces. The clamping forces are now that high, that the
friction between the bands and elements is higher than the friction between the elements and the pulley. In this case, the friction between elements and pulleys drags the elements around the pulley, and stacks them in the lower belt part, causing the compressive force to act at the lower belt part. The compressive force now actually transmits a negative torque, decreasing efficiency. The absolute values of $Q$, $S_0$ and $S_1$ differ per situation, but the distribution of forces is exactly the same, therefore, for both situations the LNC model can be used.

At the primary pulley, the full angle of wrap is used to decrease the tensile force from $S_1$ at $y = -b_p/2$, to $S_0$ at $y = b_p/2$. The compressive force starts to decrease from $Q_{\text{max}}$ at $y = b_p/2 - b_1$, to 0 at $y = b_p/2$. There is a part of the primary angle of wrap at which the tensile force is already decreasing, while the compressive force is still at maximum value. At the secondary pulley, there is first an idle angle of $y = 4b$ at which $S_0$ stays constant, and there is no compressive force. Then, at $y = b_s/2 - b_p$, the tensile force starts to increase to its final value of $S_1$ at $y = b_s/2$. The compressive force starts to increase from $y = b_s/2 - b_2$, to its maximum value at $y = b_s/2$. 
ODPC model

The ODPC model (see Figure 3.7) is used for two situations. The first situation is when a positive torque $T_p$ is transmitted from the primary to the secondary pulley, when $r_g \geq 1$. In this case, the tensile force actually transmits a negative torque, and the compressive force transmits a positive torque, reducing efficiency. In the other situation, a negative torque $T_p$ is transmitted from primary to secondary pulley, when $r_g \geq 1$, in combination with relatively high clamping forces. Due to these high clamping forces, the elements are stacked in the upper belt part, creating a compressive force there, instead of in the lower belt part (during negative $T_p$). In this situation, the compressive force actually transmits a positive torque, while the tensile force transmits a negative torque, again decreasing efficiency.

At the primary pulley, there is first an idle angle of magnitude $y = 4b$ in which the tensile force stays $S_0$, and where there is no compressive force. Then, at angle $y = b_p/2 - b_s$ the tensile force starts to build up to its final value $S_1$ at angle $y = b_p/2$. The compressive force starts building up from zero at $y = b_p/2 - b_1$, to $Q_{max}$ at $y = b_p/2$. There is a part of the angle of wrap at which the tensile force is already increasing, while the compressive force is still zero. At the secondary pulley, the full angle of wrap $b_s$ is used to decrease the tensile force from $S_1$ at $y = -b_s/2$ to $S_0$ at $y = b_s/2$. The compressive force starts to decrease from angle $y = b_s/2 - b_2$, and obtains the value zero at $y = b_s/2$. Again there is a part of the angle of wrap at which the tensile force already starts to decrease, while the compressive force is still at its maximum value.

ODNC model

The ODNC model (see Figure 3.8) is used for transmitting a negative torque $T_p$ from the primary to the secondary pulley, and $r_g \geq 1$, while clamping forces are normal. The elements are stacked in the lower belt part, creating a compressive force there. The highest tensile force is in the upper part of the belt. This means that both the difference in tensile forces between the upper and lower part of the belt, as well as the compressive force transmit a negative torque.
Figure 3.8: Distribution of belt forces for the ODNC model.

At the primary pulley there is an idle arc of length \( y = 4b \) at which the tensile force stays \( S_0 \), and the compressive force stays \( Q_{\text{max}} \). Then at angle \( y = b_p/2 - b_s \), the tensile force starts to build up to finally end at \( y = b_p/2 \) with \( S_1 \). The compressive force starts decreasing at \( y = b_p/2 - b_1 \), to finally end at \( y = b_p/2 \) with zero. There is a part of the angle of wrap \( b_p \), at which the tensile force already starts building up while the compressive force is still at maximum value. At the secondary pulley, the full angle of wrap is used to decrease the tensile force from \( S_1 \) at \( y = -b_s/2 \), to \( S_0 \) at \( y = b_s/2 \). The compressive force starts to build up from zero at \( y = b_s/2 - b_2 \) to \( Q_{\text{max}} \) at \( y = b_s/2 \). Again, there is a part of the angle of wrap \( b_s \), at which the tensile force already is decreasing, while the compressive force is still zero.

### 3.2.1 Quantification of the belt forces

Now that the four different force models are known, the forces can be quantified. This is done by using two infinitesimal small pieces of band and element, see Figure 3.9. In the following, both packs of bands are assumed to form one band. In this figure, the coordinate \( y \) runs along the pulley, similar to \( y \) in the figures of the belt models.

**Tension forces**

From Figure 3.9(a), in radial direction, the equilibrium

\[
dS = -dW_1 \tag{3.10}
\]

can be obtained. In tangential direction, the equilibrium is

\[
(S - B)dy = dN_1 \tag{3.11}
\]

In these equations, \( S \) is the tensile force, \( W_1 \) is the friction force between band and element, \( N_1 \) is the normal force between bands and elements and \( B \) is the centrifugal force per unit angle of band, defined by \( \rho_b v_b^2 \), where \( v_b \) is the circumferential velocity for the piece of band and \( \rho_b \) the bands mass per unit length. The friction force and normal force are coupled through

\[
dW_1 = \chi_1 s_1 dN_1 \tag{3.12}
\]
CHAPTER 3. SIMULATION MODEL

(a) Infinitesimal part of band.  
(b) Infinitesimal part of element.

Figure 3.9: Infinitesimal small pieces of band and element and the forces acting on them.

Table 3.1: Boundary Conditions for all models at both pulleys

<table>
<thead>
<tr>
<th>Model</th>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$BC_1$</th>
<th>$BC_2$</th>
<th>$BC_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPC-prim</td>
<td>1</td>
<td>1</td>
<td>$S(b_p/2) = S_0$</td>
<td>$Q(b_p/2 - b_1) = 0$</td>
<td>$Q(b_p/2) = Q_{max}$</td>
</tr>
<tr>
<td>LPC-sec</td>
<td>-1</td>
<td>-1</td>
<td>$S(b_s/2 - b_p) = S_0$</td>
<td>$Q(b_s/2) = 0$</td>
<td>$Q(b_s/2 - b_2) = Q_{max}$</td>
</tr>
<tr>
<td>LNC-prim</td>
<td>1</td>
<td>-1</td>
<td>$S(b_p/2) = S_0$</td>
<td>$Q(b_p/2) = 0$</td>
<td>$Q(b_p/2 - b_1) = Q_{max}$</td>
</tr>
<tr>
<td>LNC-sec</td>
<td>-1</td>
<td>1</td>
<td>$S(b_p/2 - b_p) = S_0$</td>
<td>$Q(b_p/2 - b_2) = 0$</td>
<td>$Q(b_s/2) = Q_{max}$</td>
</tr>
<tr>
<td>ODPC-prim</td>
<td>-1</td>
<td>1</td>
<td>$S(b_p/2 - b_p) = S_0$</td>
<td>$Q(b_p/2 - b_1) = 0$</td>
<td>$Q(b_p/2) = Q_{max}$</td>
</tr>
<tr>
<td>ODPC-sec</td>
<td>1</td>
<td>-1</td>
<td>$S(b_s/2) = S_0$</td>
<td>$Q(b_s/2) = 0$</td>
<td>$Q(b_s/2 - b_2) = Q_{max}$</td>
</tr>
<tr>
<td>ODNC-prim</td>
<td>-1</td>
<td>-1</td>
<td>$S(b_p/2 - b_p) = S_0$</td>
<td>$Q(b_p/2) = 0$</td>
<td>$Q(b_p/2 - b_1) = Q_{max}$</td>
</tr>
<tr>
<td>ODNC-sec</td>
<td>1</td>
<td>1</td>
<td>$S(b_s/2) = S_0$</td>
<td>$Q(b_s/2 - b_2) = 0$</td>
<td>$Q(b_s/2) = Q_{max}$</td>
</tr>
</tbody>
</table>

where $s_1 = \text{sign}(v_c - v_b)$, $\chi_1$ is a friction coefficient and $v_c$ is the circumferential velocity of the element. If there is a relative motion between element and band, $\chi_1 = \mu_1$ (Coulomb friction), else $\chi_1 < \mu_1$. Here, $\mu_1$ is the Coulomb friction parameter that determines the amount of friction between the bands and elements. Combining these last three equations gives the following differential equation

$$dS = -\chi_1 s_1 (S - B) dy$$ (3.13)

While deriving the force models as described above (LPC, LNC, ODPC and ODNC), the assumption was already made that there is an idle arc ($4b$) at the pulley that is running at the larger pulley radius. This justifies that $\chi_1 = \mu_1$ at both full wrapped angles, except the idle $4b$-part. Now $S$ is the only variable depending on $y$, the differential equation (3.13) can be solved, resulting in

$$S(y) = K_s e^{-\mu_1 s_1 y} + B$$ (3.14)

This solution is only usable if it is accompanied by some boundary conditions. These boundary conditions can be derived from the figures belonging to the force models. From these models the angles at which $S(y) = S_0$ are determined. These boundary conditions are summed in column 4 ($BC_1$) of table 3.1.
CHAPTER 3. SIMULATION MODEL

By using these boundary conditions, the integration constant $K_S$ of Eq. (3.14) can be calculated. The integration constants $K_S$ for all models on both the primary and secondary pulley can be found in column 2 ($K_S$) of Table 3.2.

Next an expression for $S_0$ is needed. If that expression is available, the distribution of the tensile force is complete. In order to obtain such an expression, Hookes Law is used

$$\sigma = E\epsilon$$  \hspace{1cm} (3.15)

where $\sigma$ is the tensile stress, $E$ the Young’s modulus, and $\epsilon$ the strain. Strain is defined as $\epsilon = \frac{\Delta L}{L}$, with $L$ the length of the bands, and $\Delta L$ the elongation of the bands. Expanding this expression for a tensile force that is not constant but varying along the length of the bands, the following integral expression can be obtained

$$L - L_0 = \frac{1}{EA} \int_y S(y)dy$$  \hspace{1cm} (3.16)

Here, $L_0$ is the initial length of the bands and $A$ the cross section of the belt. Solving the integral equation eventually leads to an expression for $S_0$, thereby completing the distribution of the tensile force in the bands.

### Compression forces

For the distribution of compressive forces, Figure 3.9(b) can be used. Here an infinitesimal small part of an element and the forces that act on it are shown. Taking again the tangential equilibrium, results in

$$2dW_{2,tan} + dD = 0$$  \hspace{1cm} (3.17)

Here, $dW_{2,tan}$ is the tangential component of the friction between element and pulley. The force $D$ is defined as $D = S - Q$. Note that the friction force between pulley and element act on both sides of the element, hence the factor 2. Taking the equilibrium in radial direction, results in

$$(B + C - D)dy + 2dN_2 \sin \theta + 2dW_{2,rad} \cos \theta = 0$$  \hspace{1cm} (3.18)

Here, $dW_{2,rad}$ is the radial component of the friction force between element and pulley. $C$ is the centrifugal force, defined as $C = \rho_c v_c^2$, $v_c$ is the longitudinal velocity for the piece of

<table>
<thead>
<tr>
<th>Model</th>
<th>Integration Constant $K_S$</th>
<th>Integration Constant $K_Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPC-prim</td>
<td>$(S_0 - B)e^{(\mu_1 b_y/2)}$</td>
<td>$(C - K_S)e^{(\mu_1 (b_y/2-b_1))}e^{(\mu_3 (b_y/2-b_1))}$</td>
</tr>
<tr>
<td>LPC-sec</td>
<td>$(S_0 - B)e^{(-\mu_1 (b_y/2-b_2))}$</td>
<td>$(C - K_S)e^{(\mu_3 b_y/2)}$</td>
</tr>
<tr>
<td>LNC-prim</td>
<td>$(S_0 - B)e^{(\mu_1 b_y/2)}$</td>
<td>$(C - K_S)e^{(\mu_1 (b_y/2-b_1))}e^{(\mu_3 (b_y/2-b_1))}$</td>
</tr>
<tr>
<td>LNC-sec</td>
<td>$(S_0 - B)e^{(-\mu_1 (b_y/2-b_2))}$</td>
<td>$(C - K_S)e^{(\mu_3 b_y/2)}$</td>
</tr>
<tr>
<td>ODPC-prim</td>
<td>$(S_0 - B)e^{(-\mu_1 (b_y/2-b_1))}$</td>
<td>$(C - K_S)e^{(\mu_3 b_y/2)}$</td>
</tr>
<tr>
<td>ODPC-sec</td>
<td>$(S_0 - B)e^{(\mu_1 b_y/2)}$</td>
<td>$(C - K_S)e^{(\mu_3 b_y/2)}$</td>
</tr>
<tr>
<td>ODNC-prim</td>
<td>$(S_0 - B)e^{(-\mu_1 (b_y/2-b_1))}$</td>
<td>$(C - K_S)e^{(\mu_3 b_y/2)}$</td>
</tr>
<tr>
<td>ODNC-sec</td>
<td>$(S_0 - B)e^{(\mu_1 b_y/2)}$</td>
<td>$(C - K_S)e^{(\mu_3 b_y/2)}$</td>
</tr>
</tbody>
</table>

Table 3.2: Integration constants for all models at both pulleys.
element and \( \rho_c \) the element’s mass per unit length. The contact force between pulley and element is \( dN_2 \). Using again Figure 3.9(b), it can be seen that

\[
dW_{2, \text{rad}} = dW_2 \sin \gamma
\]

\[
dW_{2, \text{tan}} = dW_2 \cos \gamma
\]

where \( \gamma \) is the angle as shown in Figure 3.9(b), indicating the difference between radial and tangential friction between pulley and element. Again a Coulomb friction model is used to model the friction force

\[
dW_2 = \chi_2 s_2 dN_2
\]

with \( \chi_2 = \mu_2 \) if there is a relative motion between the pulley and the element (else \( \chi_2 < \mu_2 \)). Here, \( \mu_2 \) is the Coulomb friction parameter that determines the amount of friction between the elements and pulleys. \( s_2 \) depends on the difference in velocity between the pulley and element and is either 1 or -1 \( (s_2 = \text{sign}(v_p - v_c), \) where \( v_p \) is the circumferential velocity of the pulley). If these equations are combined, the following differential equation is obtained

\[
1 \frac{dD}{dy} + D = B + C
\]

where, \( \chi_3 \) is defined as

\[
\chi_3 = \frac{\chi_2 s_2 \cos \gamma}{\sin \theta + \chi_2 s_2 \sin \gamma \cos \theta}
\]

It is assumed that \( \gamma = 0 \), meaning that the friction between pulley and element is only present in tangential direction. Using the fact that during a relative motion between pulley and element, \( \chi_2 = \mu_2 \), and the assumption that \( \gamma = 0 \), the parameter \( \chi_3 \) can be approximated by \( \mu_3 \), defined as follows

\[
\mu_3 = \frac{\mu_2 s_2}{\sin \theta}
\]

Now, with the expression of \( \mu_3 \) the differential equation for \( D(y) \) can be solved. Note that this solution is only valid on those arcs where the compressive force is changing.

\[
D(y) = -K_Q e^{-\mu_3 y} + B + C
\]

Again a boundary condition is needed in order to be able to quantify the integration constant \( K_Q \). These boundary conditions are given in column 5 (\( BC_2 \)) of table 3.1. The resulting integration constant \( K_Q \) are given in column 3 (\( K_Q \)) of table 3.2. The boundary conditions depend on the parameters \( b_1 \) and \( b_2 \). These angles are not yet derived. So extra equations are needed in order to obtain the values of these angles. These equations can be found by the momentum equilibrium at each pulley. See Figure 3.10 for an equilibrium at the primary pulley for the LPC model. With the use of this momentum equation, an expression for \( Q_{\text{max}} \) can be obtained

\[
Q_{\text{max}} = \frac{T_p - (S_1 - S_0)(R_p + \Delta R)}{R_p}
\]

This leads to an expression for \( b_1 \) and \( b_2 \). However, no exact solution of these expressions
exists, so it has to be approximated. The approximation can be done either by numerically iterating for \( b_1 \) and \( b_2 \), or by approximating the expression containing \( b_1 \) and \( b_2 \), such that the expression is solvable.

At this point, the distribution of both the compressive and tensile force along the pulleys are known. Integrating the distributions along both pulleys leads to an expression for \( F^p \), the force between the belt and the primary pulley, and \( F^s \), the force between the belt and the secondary pulley. The equilibrium of forces that act on the side of the element that is in contact with the pulley gives

\[
dF = dN_2 \cos \theta - dW_{2, rad} \sin \theta
\]

After some substitution and simplification, this leads to

\[
\frac{dF}{dy} = \frac{D(y) - B - C}{2 \tan \theta}
\]

Assuming \( B \) and \( C \) to be constant along the angles of wrap, the expression for \( F^p \) and \( F^s \) result in

\[
F^p = \int_{b_p}^{b_p + \Delta R} -\frac{K_Q e^{-\mu_3 y}}{2 \tan \theta} dy
\]

\[
F^s = \int_{b_s}^{b_s + \Delta R} -\frac{K_S e^{-\mu_3 y}}{2 \tan \theta} dy
\]

Although the centrifugal forces \( B \) and \( C \) do not appear explicitly in the expressions for \( F^p \) and \( F^s \), they are incorporated in the integration constants \( K_Q \) and \( K_S \).

To end this section, the calculation procedure is summarized. Note that the main goal of the distribution of forces as described above, is to end up with an expression for the forces \( F^p \) and \( F^s \). It all starts with the actual positions of the pulleys. From the position of the pulleys, an expression for \( S_0 \) can be found, as well as an expression for the angles of wrap \( b_p \) and \( b_s \). Also from the position of the pulleys the geometric ratio \( r_g \) can be calculated. Together with the input torque \( T_p \) and the geometric ratio, the force model can be chosen (LPC, LNC,
ODPC or ODNC). With the use of the rotational velocities of both pulleys, the centrifugal forces \( B \) and \( C \) can be determined. Next the angles \( b_1 \) and \( b_2 \) can be approximated, and the forces can be integrated along the angles of wrap, which terminates the calculation. Making an equilibrium of all the forces that act on the pulleys (see further on), the new positions of the pulleys are determined, and the calculation procedure continues from the beginning.

### 3.3 Multi-body modelling

Although the mechanism to transform the actuation torques from the servomotors to the translation of the moveable pulleys is already described in the introduction, the kinematics of this mechanism are described in this section. As stated in the introduction, the mechanism consists of four epicyclic gears. The kinematic relations of an epicyclic gear are

\[
\dot{\theta}_{\text{sun}} = (z + 1) \dot{\theta}_{\text{carrier}} - z \dot{\theta}_{\text{ring}} \tag{3.31}
\]

\[
\dot{\theta}_{\text{sun}} = \dot{\theta}_{\text{ring}} - (z - 1) \dot{\theta}_{\text{planet}} \tag{3.32}
\]

where \( z \) is the ratio of the radius of the ring gear and the radius of the sun gear. In these and the following equations, \( \dot{\theta} \) is used to describe the rotational velocity of a body and \( \dot{x} \) to describe the translational velocity of a body. Moreover, each body has one superscript and two subscripts. The superscript is used to denote if a body belongs to the primary or the secondary side. The subscript contains two letters. The first letter refers to the type of body (\( s \) for sun, \( r \) for ring, \( c \) for carrier, and \( p \) for planet). The second letter refers to which epicyclic gear the body belongs to (\( l \) for left, \( r \) for right). Because the carriers on the primary side are coupled, as well as the carriers on the secondary side, these bodies do not have a second letter. Figure 3.11 shows a schematic drawing of the coupled epicyclic sets and clarifies the naming of the gears. Using this notation in combination with the above kinematic relations for epicyclic gears, this results in

\[
(z + 1) \dot{\theta}_c^p = \dot{\theta}_{sr}^p - z \dot{\theta}_{rl}^p \tag{3.33}
\]

\[
\dot{\theta}_{sl}^p = \dot{\theta}_{sr}^p - z \dot{\theta}_{rl}^p \tag{3.34}
\]

\[
(z - 1) \dot{\theta}_{pr}^p = -z \dot{\theta}_{sr}^p \tag{3.35}
\]

\[
(z - 1) \dot{\theta}_{pl}^p = -z \dot{\theta}_{sr}^p + 2z \dot{\theta}_{rm}^p \tag{3.36}
\]

\[
(z + 1) \dot{\theta}_s^p = \dot{\theta}_{sl}^p + z \dot{\theta}_{rl}^p \tag{3.37}
\]

\[
\dot{\theta}_{sr}^p = z \dot{\theta}_{rl}^p + \dot{\theta}_{sl}^p + z \dot{\theta}_{rl}^p \tag{3.38}
\]

\[
(z - 1) \dot{\theta}_{pr}^p = -2z \dot{\theta}_{rl}^p - \dot{\theta}_{sl}^p - z \dot{\theta}_{rl}^p \tag{3.39}
\]

\[
(z - 1) \dot{\theta}_{pl}^s = - \dot{\theta}_{sl}^s + z \dot{\theta}_{rl}^s \tag{3.40}
\]

Note that the sun gear of the primary right epicyclic gear is connected to the input shaft of the CVT, and that the secondary left sun gear is connected to the output shaft of the CVT. The kinematic relation for a screw is given by

\[
\dot{x}_{\text{out}} - \dot{x}_{\text{in}} = s(\dot{\theta}_{\text{out}} - \dot{\theta}_{\text{in}}) \tag{3.41}
\]

where \( s \) is the pitch of the screw, \( x_{\text{out}} \) is the screw output position (i.e. the moveable pulley position) and \( x_{\text{in}} \) the screw input position. Using this relation with the relations for the epicyclic sets, this results in

\[
\dot{x}_{\text{in}}^p = \dot{x}_{\text{out}}^p - s(\dot{\theta}_{sl}^p - \dot{\theta}_{sr}^p) = sz \dot{\theta}_{rl}^p + \dot{x}_{\text{out}}^p \tag{3.42}
\]
The hydraulically actuated CK2 CVT uses a sensor to determine the position of the primary pulley. In the EMPact system no such sensor is available. The geometric ratio \( r_g \) is therefore estimated with the use of the above relations between the pulley positions (\( x_{out}^p \) and \( x_{out}^s \)) and the rotation of the servomotors (\( \theta_p \) and \( \theta_s \)). As will be seen in the following chapters, the parameter \( \theta_p \) is used for the control of the geometric ratio \( r_g \).

Figure 3.12 shows a schematic drawing of the EMPAct system. This figure shows that the primary servomotor is connected to the primary left ring gear by means of a worm gear. This connection is modelled as a rotational stiffness between the servo and the worm. The secondary servomotor is connected to the secondary left ring gear by means of a planetary reduction and a chain. The chain is modelled as a translational stiffness. The thrust bearings that have to support the clamping forces (at both the primary and secondary side) are situated between the sun gears and are modelled as a translational stiffness. The driveshaft and the Torque Converter (TC) are modelled as torsional springs. The TC can be modelled as a rotational stiffness, because in all simulations the TC is locked up. Friction is implemented only in a viscous form (no Coulomb offset) and only in the screws and in the rotation of the servomotors. All other bodies are not influenced by friction and are considered rigid within the frequency band of interest (below 100 \([\text{Hz}]\)).
3.4 Equilibrium at the pulleys

The force and torque equilibria at the pulleys are depicted in Figure 3.13. Besides the forces from the belt to the pulleys \( (F^p_p\text{ and } F^s_s) \), also the forces from the screws to the pulleys \( (F^p_{act}\text{ and } F^s_{act}) \) are shown. Resulting forces from shifting are not covered by the belt forces \( (F^p_p\text{ and } F^s_s) \). To incorporate shifting forces, a Shafai model is used which can be seen as a damping relative to the translation of the pulleys. The Shafai damping is equally divided over both pulleys, resulting in the forces \( F^p_d \) and \( F^s_d \) of Figure 3.13. More explicitly the Shafai model can be described as

\[
F^p_d = \frac{b}{2} p_p
\]

\[
F^s_d = \frac{b}{2} s_s
\]

where \( b \) is the Shafai damping constant. The resulting force equilibrium is

\[
F^{p,s}_{act} = F^{p,s} + F^p_d
\]

In [1] it is shown that the transmittable torque through the belt from the primary to the secondary pulleys and vice versa is dependant on the clamping force \( F_s \), and the traction coefficient \( \mu \). The amount of transmittable torque through the belt at both pulleys is expressed as

\[
T^s_{belt} = \frac{2\mu(\nu)F_s R_s}{\cos (\theta)}
\]
CHAPTER 3. SIMULATION MODEL

Figure 3.13: Equilibrium of torque and force on the movable pulleys.

Figure 3.14: The traction coefficient as a function of slip and ratio.

\[ T^p_{\text{belt}} = r_g T^s_{\text{belt}} \]  \hspace{1cm} (3.48)

The traction coefficient \( \mu \) is slip (\( \nu \)) and ratio (\( r_g \)) dependant. Figure 3.14 shows measured data of the traction coefficient that is used in the simulation model. As can be seen from this figure, the slope of the traction coefficient is the steepest when the variator is in Medium, then just a bit less steep for OD, and significantly less steep for Low. Clearly, the traction curve can be divided into two regions. The first region is where the traction coefficient increases for increasing slip, the second region is where the traction coefficient does not increase anymore for increasing slip. If the traction coefficient decreases for increasing slip, then this has a destabilizing effect. The instability can physically be described by the following. The reason for the slip to increase is generally because of an increasing load torque. So when the load torque increases, it increases the slip in the variator, which at its turn increases the traction coefficient (if the slope of \( \mu \) is positive), and with a higher traction coefficient more torque can be transmitted. If too much slip is present in the variator (when the slope of \( \mu(\nu) \) is not increasing anymore), and the load torque is increasing, the traction coefficient does not increase anymore (it even decreases slightly), and the amount of torque that can be transmitted by the belt is lower than it is required to be, so slip will increase more and more.
Note that the point at which the slope of $\mu(\nu)$ changes is different for different $r_g$. 
Chapter 4

Identification

Although the simulation model is very elaborated and reflects reality at least in a qualitative way, it is not suited for controller design purposes. In general, controller design can be divided into linear and nonlinear. Here the linear approach is chosen. Therefore the idea is to take the CVT into several equilibrium points and linearize around these distinct equilibrium points, each time estimating a linear model. In the end this will result in a large set of linear plants. Several controller design approaches exist that can cope with a set of plants, interpreting them as one nominal plant and a set of plants that represents the uncertainty around the nominal one. With $H_\infty$, for instance, one makes a separate filter to catch all the deviations from the nominal plant. $H_\infty$ however needs a state-space description of the plants, whereas for Quantitative Feedback Theory (QFT) only frequency response functions are needed. As a starting point it is preferable to use the QFT-approach of [19] for the reason that SISO linear systems techniques can then be used. Normally an identification with the use of noise injection would be used to obtain the frequency response functions, but this takes extremely much simulation time. In this chapter a minimal MIMO-realization is estimated with the use of step responses. The end result of the identification should be a frequency response function to describe the input-output behavior between four variables. These input-output relations are: (see Figure 4.1) from the primary servomotor torque $T_{p_m}$ to the rotation of the primary servomotor $\theta_p$, from the primary servomotor torque $T_{p_m}$ to the slip $\nu$, from the secondary servomotor torque $T_{s_m}$ to the primary servomotor rotation $\theta_p$, and finally from the secondary servomotor torque $T_{s_m}$ to the slip $\nu$.

4.1 Step response estimation

The estimation on basis of the step-responses makes use of the Markov-parameters of the system, and is widely described in literature, see for example [9], [12], [3], [15] and [13]. A summary of the method can be found in appendix A. In this appendix the method is described how to obtain an estimation for the four matrices $F,G,H$, and $K$ that define the dynamical system $H$

$$\begin{align*}
z_{k+1} &= Fz_k + Gu_k \\
y_k &= Hz_k + Ku_k
\end{align*}$$

where $z$ is the state-vector of dimension $n$, $u$ is the input vector, $y$ is the output vector, and $k$ is the time step. The impulse response of this dynamical system at time step $k = 0$ is an
estimation for the matrix $K$. For $k > 0$, the relation between the other system matrices and the impulse response is

$$D_i = HF^{i-1}G$$

where $D_i$ is the impulse response at time steps $k > 0$. The parameters $D_i$ are defined as the Markov parameters. With these Markov parameters a Hankel matrix $H_E$ can be constructed as is shown in appendix A. From this Hankel matrix, a singular value decomposition can be made in order to throw away those states in the state-vector that contribute little in the input-output behavior.

$$H_E = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

Here, $\Sigma_1$ is the matrix that contains the largest singular values. With this singular value decomposition the system matrices $F$, $G$, and $H$ can be constructed as follows

$$F = \Sigma_1^{-\frac{1}{2}}U_1^TH_AV_1\Sigma_1^{-\frac{1}{2}}$$

$$G = \Sigma_1^{-\frac{1}{2}}U_1^TH_B$$

$$H = H_CV_1\Sigma_1^{-\frac{1}{2}}$$

Here, $H_A$, $H_B$, and $H_C$ are parts of the Hankel matrix $H_E$, see appendix A. In this way it is possible to construct the system matrices by using the system’s impulse response. From a practical viewpoint it is more useful to use step responses instead of impulse responses for the estimation. Furthermore, the estimation has to be extended to the MIMO case.

To extend the method to step responses, note that for strictly-proper discrete systems the step function is defined as $u_k = 1, k = 0, 1, 2, \ldots$ and $u_k = 0, k = -1, -2, \ldots$. Then the step response satisfies $S_k = 0, k = 0, -1, -2, \ldots$ and

$$S_k = \sum_{i=1}^{k} D_i \quad k = 1, 2, \ldots$$
Consequently, the matrices $H_A$, $H_B$, $H_C$, and $H_E$ have to be adapted to (see [9])

$$
H_E = \begin{bmatrix}
S_1 & S_2 & S_3 & S_4 & \cdots & S_{k_2} \\
S_2 - S_1 & S_3 - S_1 & S_4 - S_1 & \cdots & \cdots & S_{k_2+1} - S_1 \\
S_3 - S_2 & S_4 - S_2 & \cdots & \cdots & S_{k_2+2} - S_2 \\
S_4 - S_3 & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
S_{k_1} - S_{k_1-1} & -S_{k_1+1} - S_{k_1-1} & S_{k_1+2} - S_{k_1-1} & \cdots & \cdots & S_{k_1+k_2-1} - S_{k_1-1}
\end{bmatrix}
$$

$$
H_A = \begin{bmatrix}
S_2 - S_1 & S_3 - S_1 & S_4 - S_1 & \cdots & \cdots & S_{k_2+1} - S_1 \\
S_3 - S_2 & S_4 - S_2 & S_5 - S_2 & \cdots & \cdots & S_{k_2+2} - S_2 \\
S_4 - S_3 & S_5 - S_3 & \cdots & \cdots & S_{k_2+3} - S_3 \\
S_5 - S_4 & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
S_{k_1+1} - S_{k_1} & S_{k_1+2} - S_{k_1} & S_{k_1+3} - S_{k_1} & \cdots & \cdots & S_{k_1+k_2} - S_{k_1}
\end{bmatrix}
$$

$$
H_C = \begin{bmatrix}
S_1 & S_2 & S_3 & S_4 & \cdots & S_{k_2}
\end{bmatrix}
$$

$$
H_B = \begin{bmatrix}
S_1 \\
S_2 - S_1 \\
S_3 - S_2 \\
S_4 - S_3 \\
\vdots \\
S_{k_1} - S_{k_1-1}
\end{bmatrix}
$$

(4.6)

The identification is performed in a closed-loop environment. In this way it is possible to determine a Sensitivity ($S(z)$) and a ProcessSensitivity ($PS(z)$). From $S(z)$ and $PS(z)$, the plant ($P(z)$) is calculated via $P(z) = PS(z)S^{-1}(z)$. In the MIMO case every $S_i$ in (4.6) is constructed from two separate experiments. First a step function is applied to the primary disturbance input $d_p$, see Figure 4.1, and step responses of four outputs ($y_{11} = T_m^p$, $y_{21} = T_m^a$, $y_{31} = \theta_p$, $y_{41} = \nu$), are measured. Then a step function is applied to the secondary disturbance input, $d_s$, and again the four outputs ($y_{12} = T_m^p$, $y_{22} = T_m^a$, $y_{32} = \theta_p$, $y_{42} = \nu$) are measured. Now $S_k$ can be constructed according to

$$
S_k = \begin{bmatrix}
y_{11_k} & y_{12_k} \\
y_{21_k} & y_{22_k} \\
y_{31_k} & y_{32_k} \\
y_{41_k} & y_{42_k}
\end{bmatrix}
$$

(4.7)

To obtain proper gain estimation of the system, the outputs must be normalized to a unit
CHAPTER 4. IDENTIFICATION

Figure 4.2: Normalized step responses due to a step on the primary servo at \( \nu = 1.0\% \).

4.1.1 Results

As was already explained, the system will first be taken into several equilibrium points, at which the system is linearized. In order to bring, and keep, the system around these equilibrium points the \( \theta_p\)-loop is closed, thereby stabilizing \( r_g \) (\( r_g \) can be estimated by using \( \theta_p \) and \( \theta_s \), see Chapter 5). The control of the slip-loop is not so easy and therefore in the beginning no controller is applied for the slip-loop. The first plants are estimated with an open loop torque from \( T_s \) to maintain a certain slip-level. Next, based on the estimated plants, a controller is designed for both the \( \theta_p\)-loop and the slip-loop. The resulting step responses (with use of the new controllers) stay closer to the equilibrium points, hence a better estimation is obtained. When different equilibrium points are identified, it is often necessary to redesign the slip controller to obtain satisfactory step responses. Based on the dynamical analysis it is expected that the plants depend most significantly on \( \nu \) and \( r_g \). Therefore the equilibrium points at which the linearization is applied are a combination of these variables. In order to obtain more smooth templates for the QFT control design (see appendix B and chapter 5), a rather fine grid of equilibrium points is used.

\[
\begin{align*}
\nu & \in [0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0] \\
r_g & \in [0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25]
\end{align*}
\] (4.9) (4.10)

In Figure 4.2, the normalized step responses to a step of \( d_p = 10^{-4}[Nm] \) applied at the primary servo are shown. As can be seen, both \( \theta_p \) and \( \nu \) are plotted as a function of time and at different ratio. The nominal slip level is \( \nu = 1.0\% \), and the nominal ratios are \( r_g = 2.0 \) (OD), \( r_g = 1.0 \) (Medium) and \( r_g = 0.5 \) (Low). The response of \( \theta_p \) is almost setpoint independent. The response of \( \nu \) however does show some strange effects. For the ratios Medium and OD, a positive step applied at the primary servo results in a negative response in \( \nu \). For Low however, a positive primary step results in a positive response in \( \nu \). Applying
a negative primary step yields exactly the opposite effect (although this is not depicted in the figure). Furthermore it can be seen that the measurement time of 2 seconds is a bit too short for the response of $\theta_p$ to converge back to the equilibrium point.

Figure 4.3 shows the responses to a step at the secondary servo. Both plots in this figure show that the response has the smallest amplitude at Low, then OD, and the largest at Medium. Note that the controller used for this simulation influences the responses. Computing the Hankel Matrix according to Eq. (4.6), with a sampling frequency of $F_s = 1000$ [Hz], $k_1 = 1500$ and $k_2 = 1500$, a Singular Value Decomposition (SVD) can be made of this matrix. The first 100 singular values are shown in Figure 4.4. The Singular Values are sorted from large to small, and a choice of 50 would be more than sufficient to incorporate the states that have significant effect on the output. Note that the plotted singular values are those of the system $\mathcal{H}$ of Eq. 4.8, and not of $P$ in Figure 4.1. $P$ can then be obtained by $P = PS \cdot S^{-1}$. When 50 is indeed chosen, the resulting error between simulated (normalized) step responses and the step responses generated with the estimated model (Process Sensitivity), is shown in Figure 4.5. These errors are within acceptable values. The Figures 4.2 to 4.5 all correspond to the parameter $\nu = 1.0\%$. But the same analysis is also done for different parameter combinations of both $\nu$ and $r_g$, see (4.10). This results in the Transfer Function Matrices (TFM) of Figure 4.6. Clearly, from this figure, some interesting conclusions can be drawn. First Figure 4.6(a),

![Normalized step responses](image)

(a) $d_s \rightarrow \theta_p$

![Normalized step responses](image)

(b) $d_s \rightarrow \nu$

Figure 4.3: Normalized step responses due to a step on the secondary servo at $\nu = 1.0\%$.

![Singular Value Decomposition](image)

Figure 4.4: Typical Singular Value Decomposition.
shows that for the Transfer Function (TF) from $T_p$ to $\theta_p$, hardly any variation is present. Just before 20 [Hz] a mass decoupling can be recognized. Besides this decoupling, a resonance can be seen in Figure 4.6(b) that occurs at approximately 94 [Hz]. This resonance seems to come from the (2,2)-element of the TFM (Figure 4.6(d)), and can be appointed to the stiffness of the chain at the secondary actuation side. After this resonance, the estimation is not too accurate anymore. As can be seen, several (1,2)-elements tend to become proper instead of strictly-proper, meaning that their amplitude doesn’t go to zero for infinite frequency. In the controller design only a small part of the frequency-band will be used, which certainly doesn’t include frequencies higher than 100 [Hz], so for now this estimation error is unimportant.

Looking at Figure 4.6(c) and 4.6(d), it can immediately be seen that the variation in these elements of the TFM is significantly larger, indicating that slip is more sensitive to changes in equilibrium points. From Figure 4.6(c) it can furthermore be seen that just before 30 [Hz] an anti-resonance is present. The zero of this anti-resonance is in some estimated plants minimum-phase, and in some estimated plants non-minimum-phase. This is not an estimation error. As can also be seen from Figure 4.3, in certain equilibrium points minimum-phase response exists, while in others non-minimum-phase response is present. The same distinction can be seen in Figure 4.6(d). This time however, there is no direct relation with the step-responses from $T_s$ to $\nu$ for this difference in phase behavior of the zero belonging to the anti-resonance around 70 [Hz]. Just after this anti-resonance a resonance at approximately 94 [Hz] is present, which is also present in the (1,2)-element. Furthermore a strongly damped resonance at approximately 7 [Hz] is present due to the flexibility of the driveshaft.
Figure 4.6: Bode plots from both inputs to both outputs.
4.2 Noise excitation estimation

As can be seen from Figure 4.6(b), the estimation based on step responses contains some errors in the high-frequency region. To be more confident that the estimations are accurate in the frequency-band of interest, noise is injected to measure the Frequency Response Function Matrix (FRFM) independently from the step-responses. Thereto noise is injected at the same point where the step is applied for the step response identification, see Figure 4.1. In this way the diagonal elements of the FRFM are estimated by dividing Process Sensitivity ($P_S(j\omega)$) by Sensitivity ($S(j\omega)$), while the cross-terms are estimated by direct division, that is

\[
FRFM = \begin{bmatrix}
\frac{\theta_p(j\omega)}{np(j\omega)} & \frac{\theta_p(j\omega)}{T_m(j\omega)} \\
\frac{T_m(j\omega)}{np(j\omega)} & \frac{T_m(j\omega)}{ns(j\omega)} \\
\frac{\nu(j\omega)}{np(j\omega)} & \frac{\nu(j\omega)}{T_m(j\omega)} \\
\end{bmatrix}
\]

(4.11)

Here, $np(j\omega)$ is the noise injected at the primary servo, $ns(j\omega)$ the noise injected at the secondary servo. All signals in Eq. (4.11) are transformed to the Fourier domain. Typical parameters of this transformation are: length of data blocks $nfft = 2048$ samples, a measuring time of $T = 20 \text{ [sec]}$, and a sample frequency of $f_s = 1000 \text{ [Hz]}$. During the step estimations, also a sampling frequency of $f_s = 1000 \text{ [Hz]}$ is used, but the measurement time was only 2 sec.

4.3 Comparison

In this section, the TFM’s based on step-response data are compared with the FRFM’s based on noise injected data. In Figure 4.7 a comparison between a step-response estimation and a noise estimation can be seen for $\nu = 0.4 \%$ and $r_g = 2.25 [-]$. Clearly, these estimations seem to correspond very well to each other. There is however a difference. In the (2,1)-element the step-response estimation calculates a minimum-phase zero, whereas the noise estimation calculates a non-minimum-phase zero for these parameters. A logical explanation can be that the zero is very weakly damped, hence the zero is very close to the imaginary axis (in a continuous time representation). Due to inaccuracies in the calculation procedure, the zero might have drifted from the left to the right half of the complex plane, or vice versa. Other parameter sets give similar comparisons, except for the fact that there is not always a difference in the phase-character of the (2,1)-element.
Figure 4.7: Differences in estimated Bode Plots from both inputs to both outputs.
In chapter 4, the identification of the EMPAct CVT is discussed. Available is now a set of TFM’s, depicted in Figure 4.6. The goal of the controller that is to be designed in this chapter, is to keep the system’s outputs close to their setpoints, despite the variations in the TFM, and despite disturbances and modeling errors. Due to the MIMO character of the system, several options to solve the control problem exist. The most logical way to get the best performance and robust stability out of the system is to use $H_\infty$ or $H_2$ techniques, see [5]. These techniques are also often used for SISO systems, but their potential is bigger for MIMO systems, because of the fact that they use the state space representation of systems, and end up with a full block controller. Although this approach is very tempting, it is decided to use another approach. The mechanical design was performed in such a way that the primary servomotor is used for ratio control, and the secondary servomotor for slip control. This is done for an energy saving goal. When cruising down the highway, for instance, it is not necessary to shift, and it is thus possible to keep the primary servomotor at rest, thereby saving energy, while using the secondary servomotor to maintain a certain slip level. Also actuation power can be led from the primary to the secondary pulleys or vice versa, by the meshing of the ring gears.

Alternatively, a Sequential Loop Shaping method can be used. By using this method, some performance is lost because of the full potential of MIMO controllers is not used, indeed leaving the non-diagonal elements zero. The ultimate goal of MIMO control would actually be to use these non-diagonal terms to use information from other loops for Sensitivity reduction in the loop of interest. The very interesting benefit obtained from keeping the non-diagonal terms of the controller zero, is that the theorems for SISO systems can be used. The robustness part of the design process is captured with Quantitative Feedback Theory (QFT). Excellent textbooks on QFT are [6] and [19]. Appendix B describes important aspects of QFT.

Variables that can be measured are the primary servomotor rotation $\theta_p$, the secondary servomotor rotation $\theta_s$, the rotational velocity of the input shaft of the CVT $\omega_p$, and the rotational velocity of the output shaft of the CVT $\omega_s$. The geometric ratio $r_g$ can be reconstructed by using $\theta_p$ and $\theta_s$. With this reconstructed $r_g$ the slip $\nu$ can be calculated in combination with the measurements of $\omega_p$ and $\omega_s$. The variables for control are $\theta_p$ and $\nu$. The aimed bandwidth of the $\theta_p$-loop is around 1 [Hz]. A significant higher bandwidth may result into discomfort during the driving of the car (too aggressive shifting actions). The desired bandwidth of the slip-loop is around 10 [Hz]. In [11] the control of slip in a hydraulically actuated CK2 CVT
is realized with a bandwidth below 3 [Hz]. The controller applied in [11] is tested through experiments on the road, and is found to be satisfactory.

Before designing the controllers of both loops, first some effort is done to diagonalize the plant, to proceed with the diagonal controller design. The most idealistic diagonalizing action would be to pre-or postmultiply the plant with its inverse. Then a unit matrix is obtained. Unfortunately it is not possible to take the inverse of a strictly proper plant. Another approach is to define a cost-function and try to diagonalize the plant as much as possible, taking the cost function as a measure, see [7]. Due to the uncertainty in the plant and the high damping, these techniques obtain no good results. It is however still worthwhile to look at a measure of diagonal dominance. A lot of time can be saved if the 2 x 2 MIMO system can be interpreted as 2 SISO systems, even without a diagonalizing action. In literature, the Relative Gain Array (RGA) is often used as a measure for diagonal dominance in a MIMO plant. The RGA is defined as

\[ \Lambda(j\omega) = P(j\omega) \cdot (P(j\omega)^T)^{-1} \]  

(5.1)

Here, \(P(j\omega)\) obviously represents the FRFM as a function of frequency \(\omega\). Performing the multiplication (5.1), it can be shown that both the sum of each row, and the sum of each column should equal one. Plotting the absolute values of \(\Lambda\) therefore gives a good estimate for diagonal dominance. Figure 5.1 displays such a plot for all \(r_g\) at \(\nu = 1.0\%\). It can clearly be seen that for high frequencies, the plant is diagonal dominant because \(\Lambda\) has its non-diagonal terms equal to zero, and its diagonal terms equal to one. Ideally this would be the case over the whole frequency band, but as can be seen, before 30 Hz, there is no diagonal dominance. This is a very strong reason to not choose for the strategy to just design two SISO controllers and interpret the cross-terms as disturbances on the diagonal-terms. Instead, a Sequential Loop Shaping procedure (SLS) is followed, in which the full aspect of MIMO dynamics is respected. Figure 5.1 refers to \(\nu = 1.0\%\) and all ratios, but the RGA’s
for different slip-levels are of the same qualitative form.

## 5.1 Sequential Loop Shaping

As mentioned above, the Sequential Loop Shaping procedure will be used to design a controller. In general [19] and [4] are completely followed from derivation to application in Matlab. In [19] a rather unique system representation is used compared to [8], [14], and [10]. This is necessary because of two facts. First, [19] wants to quantify disturbances, and considering the fact that also disturbances are directional in MIMO systems, special attention has to be paid to notations. Second, [19] only uses frequency bands of interest of the Frequency Response Function Matrices (FRFM’s) that define the set of plants (instead of a Transfer Function Matrix, or State-space matrices). Following [19] it can be seen from the MIMO system in Figure 5.2 that

\[(I + PG)y = d\]  \hspace{1cm} (5.2)

With

\[d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}\]  \hspace{1cm} (5.3)

representing disturbances acting on the output of \(\theta_p\) and \(\nu\) respectively. Premultiplying both sides by \(P^{-1}\) results in

\[(P^{-1} + G)y = P^{-1}d\]  \hspace{1cm} (5.4)

Note that taking the inverse of \(P\) is only possible because a certain frequency band of interest is taken, instead of frequencies tending towards infinity. If the inverse of the plant (2x2-system) is now defined as

\[P^{-1} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix}\]  \hspace{1cm} (5.5)

then Eq. (5.4) can be premultiplied at both the lefthand and righthand side with the same matrix to get

\[y_1 = \frac{\pi_{11}d_1 + \pi_{12}d_2 - \pi_{12}y_2}{\pi_{11} + g_1}\]  \hspace{1cm} (5.6)

\[y_2 = \frac{\pi_{21}d_1 + \pi_{22}d_2}{\pi_{22} + g_2}\]  \hspace{1cm} (5.7)
(a) First Loop

\[ G_1 \xrightarrow{\frac{1}{\pi_{11}}} y_1 \]  

(b) Second Loop

\[ G_2 \xrightarrow{\frac{1}{\pi_{22}}} y_2 \]  

Figure 5.3: Two SISO Loops representing the MIMO system.

with

\[ \pi_{21}^2 \equiv \pi_{21} - \frac{\pi_{21}\pi_{11}}{\pi_{11} + g_1} = \frac{\pi_{21}g_1}{\pi_{11} + g_1} \]  

(5.8)

\[ \pi_{22}^2 \equiv \pi_{22} - \frac{\pi_{21}\pi_{12}}{\pi_{11} + g_1} \]  

(5.9)

Note that Eq. (5.6) and (5.7) are only valid if the controller is diagonal, thus

\[ G = \begin{bmatrix} g_1 & 0 \\ 0 & g_2 \end{bmatrix} \]  

(5.10)

Equations (5.6) and (5.7), show that \( y_1 \) is dependant on \( y_2 \), whereas \( y_2 \) is not dependant on \( y_1 \). Of course, a transformation of coordinates can be applied, which turns this dependency around. The idea is now to choose a specification that determines how much both outputs may differ from their setpoints, under influence of all disturbances. These specifications should be a function of frequency. Choosing for instance

\[ |y_1| \leq e_1(w) \]

\[ |y_2| \leq e_2(w) \]  

(5.11)

where \( e_1(w) \) and \( e_2(w) \) are specifications as a function of frequency, the controller design can proceed. Replacing \( y_2 \) with \( e_2 \) in Eq. (5.6) gives

\[ y_1 = \frac{\pi_{11}d_1 + \pi_{12}d_2 - \pi_{12}e_2}{\pi_{11} + g_1} \]  

(5.12)

This gives a little over design, and hence is conservative (because \( y_2 \) is replaced by \( e_2 \), while there is no reason to assume that \( y_2 \) is not smaller than \( e_2 \)). Now that \( y_1 \) and \( y_2 \) are fully decoupled, two SISO systems that completely represent the MIMO system can be defined. These two SISO systems are depicted in Figure 5.3 as \( \frac{1}{\pi_{11}} \) and \( \frac{1}{\pi_{22}} \). Consequently, the disturbances \( d_1 \) and \( d_2 \) have to be transformed (into \( w_1 \) and \( w_2 \)). In the example in this section, a disturbance at the plant output is chosen, but all other possible disturbances can also be
defined and transformed in similar ways, see [19]. This is exactly the power of this method; every disturbance and the MIMO system can be transformed into the two SISO systems $\frac{1}{\pi_{11}}$ and $\frac{1}{\pi_{22}}$, and transformed disturbances. The transformed disturbances $w_1$ and $w_2$ are now

$$ w_1 = |\pi_{11}d_1 + \pi_{12}d_2| + |\pi_{12}|e_2 \quad (5.13) $$

$$ w_2 = \pi_{22}^2d_1 + \pi_{22}^2d_2 \quad (5.14) $$

For $w_1$ the worst-case scenario is taken. Consequently, a design constraint in the following form can now be made

$$ \frac{|w_1|}{\pi_{11} + g_1} \leq e_1(w) \quad (5.15) $$

$$ \frac{|w_2|}{\pi_{22}^2 + g_2} \leq e_2(w) \quad (5.16) $$

It can be seen in [4] that there is one specification-type in the QFT-Toolbox of Matlab, which exactly captures the specification as in Eq. (5.15) and Eq. (5.16). This specification-type is of the form

$$ \frac{|A_1 + B_1G|}{C_1 + D_1G} \leq W(w) \quad (5.17) $$

with $A_1, B_1, C_1$ and $D_1$ functions of frequency, and $G$ the controller to be designed. Obviously, $W(w)$ in Eq. (5.17) now equals $e_1(w)$ in Eq. (5.15) and is also a function of frequency $w$. In principle a closed-loop specifications in the form of Eq. (5.15) and Eq. (5.16) is now available, and therefore the robust controller design can proceed. In order to proceed however, $d_1$ and $d_2$ from Eq. (5.3) have to be available as a function of frequency. In general this is not trivial. In fact not a lot of information about $d_1$ and $d_2$ is available, let alone their amplitude as a function of frequency. It is known from earlier research however, that up to approximately 10 [Hz] these disturbances act, and therefore need to be attenuated. On basis of this information, it seems better to just design controllers in order to achieve a best possible Sensitivity and Process Sensitivity for both $\frac{1}{\pi_{11}}$ and $\frac{1}{\pi_{22}}$. This results in choosing a constant of 6 [dB] (so not a function of frequency) for the numeric value of $W(w)$ in 5.17.

### 5.2 Controller design for the simulation model

Proceeding with normal SISO controller design, the first loop to be closed is chosen as the slip-loop (the aimed bandwidth of the slip-loop is higher than the aimed bandwidth of the $\theta_p$-loop). Figures 5.4 and 5.5 show the corresponding $\frac{1}{\pi_{11}}$. Comparing Figure 5.4(a) with 5.4(b), it can be seen that although the difference in slip-level is small, already a large difference exists between the static gains. This static gain further increases for higher slip-levels. Unfortunately, the phase of both figures already has crossed the -180 [°]-line before 20 [Hz]. This phase-loss becomes even more dramatic for higher slip-levels, as can clearly be seen in Figure 5.5.

From these observations it seems almost impossible to achieve a bandwidth higher than 10 [Hz], especially for the higher slip-levels when every plant in the set (4.10) is taken into account. This, in principle, is also not necessary. The parameter set is reduced to define the
Figure 5.4: $\frac{1}{\pi_{11}}$ for $\nu = 0.4$ and $\nu = 0.6$ and all ratios.

Figure 5.5: $\frac{1}{\pi_{11}}$ for $\nu = 1.4$ and $\nu = 1.6$ and all ratios.
regions around maximum variator efficiency (see [1] and Figure 2.6). After all, the goal of the controller design is to design a controller that can keep the system in the vicinity of these optimal operating parameters, under influence of disturbances. In Figure 5.6, the plants that correspond to these optimal parameters are plotted. The optimal parameter set is roughly a combination of $\nu = 2.0 \%$ for Low, $\nu = 1.0 \%$ for Medium, and $\nu = 0.75 \%$ for OD. With this optimal parameter set the design problem is relaxed a little compared with the full parameter set, but still is very tough. Already after $5 \ [Hz]$, phase loss begins to act, and between 20 and 70 $[Hz]$ the phase is approximately $-240 \ [^\circ]$. As can be seen in Appendix B, it will be of interest to have a look at the templates that belong to the plants in Figure 5.6. Templates describe the variation in gain and phase of a plant, at one frequency point. A crucial issue in calculating the templates, is choosing the frequencies at which they are constructed. Obviously, constructing the templates for all frequencies would be redundant. Certainly it is interesting to get the templates around the aimed bandwidth. Furthermore, some characteristic points in the plant-set of Figure 5.6 are the region between 5 and 20 $[Hz]$, and between 70 and 95 $[Hz]$. These are the regions from which some problems can be expected concerning stability and robustness. First, the template belonging to frequency 1 $[Hz]$ is plotted in Figure 5.7. Obviously, this template corresponds well to Figure 5.6 at 1 $[Hz]$, looking at a phase variation between $-7$ and $-35 \ [^\circ]$ and a gain variation between $14$ and $39 \ [dB]$. From this template it can be seen that its contour is a bit rough. If the parameter set (4.10) would have been less dense, the contours of the templates would have been even more course. The extra time spent on estimating a lot of plants is thus worthwhile. There is a rather big difference in the uncertainty between the plants at different frequencies. Figure 5.8 shows the template at 20 $[Hz]$, which has a completely different shape. This observation asks for extreme caution in the further analysis. Of course it is impossible to plot all the templates and corresponding bounds (see Appendix B), but during the actual controller design, this is of importance, and hence is taken care of. As is explained in Appendix B and in [6] and [19], the next step in the design process is to calculate the bounds with use of the templates. The construction of the bounds can be imagined by moving the templates around the 6 $[dB]$ M-circle (see Appendix C), while just
CHAPTER 5. CONTROL

Figure 5.7: Template at 1 [Hz].

Figure 5.8: Templates at 20 [Hz].
touching the M-circle with only one point of the template. The bound is imaginary made by putting a pencil through that point on the template that represents the nominal plant. Considering the radius of the 6 [dB] M-circle, the sharp edges in the templates as shown in Figures 5.7 and 5.8 are blended during the creation of the bounds. In Figure 5.9(a), several bounds corresponding to 20 [Hz] are shown. These bounds correspond to different nominal plant situations. As can be seen, choosing different nominal plants effects the location of the bounds. The solid bound is more severe than the other three (its ‘radius’ around the critical point 0 [dB], −180 [°] is extending more to the 0 [°] phase-line, leaving less phase-margin). Looking at Figure 5.9(b), however, it can be seen that this more severe bound perfectly corresponds to a more loosened nominal plant (located more to the 0 [°] phase-line, having more phase-margin). This means that it does not matter what nominal plant is chosen because the distance between the plant and the bounds remains the same. For clarity of plotting however, it is wise to choose that nominal plant that has the least phase-margin. In that way, the bounds will be situated at the left upper side of the plant in the Nichols Chart at the critical point (0 [dB], -180 [°]). By taking this nominal plant, the controller that is depicted in Figure 5.10 is designed.
In this controller, several elements can be distinguished. First there is an integrator in combination with a zero at 2 [Hz]. Furthermore there is a 2nd order roll-off (or complex pole) at 80 [Hz] with a damping of 0.3. Then there is a spread out notch with a complex zero at 8 [Hz] and a damping of 0.3, in combination with a complex pole at 35 [Hz] and a damping of 0.7.

Applying this controller to the \( \frac{G_1}{\pi_{11}} \) loop, the resulting open-loop is plotted in the Bode-plot of Figure 5.11(a), and the resulting Sensitivity is plotted in Figure 5.11(b). It might be a bit difficult to conclude robust stability from these figures, especially from Figure 5.11(a), and therefore, also the Nichols plot of the open loop is plotted in Figure 5.12. Now it is clear that robust stability is achieved, but the overall bandwidth (and thus performance) is not too high (and also not the same for every plant). Figure 5.11(a) shows that several open loops already cross the 0 dB-line before 1 [Hz]. Figure 5.11(b) however shows that minor disturbance attenuation is achieved up to at least 8 [Hz]. For some plants this means that it takes the system a rather long time to attenuate disturbances. Better performance could be obtained by using a gain-scheduling approach. The plants are obtained from a simulation.
model however, and it is more interesting to use this gain-scheduling technique with the use of real hardware data from the test rig. Especially considering the fact that when using a gain-scheduling technique on a non-linear system, a lot of simulation work has to be done in order to make conclusions about stability.

Now that the controller $G_1$ is designed for the first loop, the controller $G_2$ for the second loop can be designed. Figure 5.13 shows the plants of $\frac{1}{\pi_{22}}$. Note that controller $G_1$ is incorporated in this plant. Comparing these plants with the plants of Figure 5.6, it can be seen that the controller design for this loop is much more relaxed (more phase and gain-margin). Also the uncertainty in this loop is much less than the uncertainty of the $\frac{1}{\pi_{11}}$-loop. Although a much higher bandwidth could have been obtained in this loop, it is chosen to go for a low bandwidth on purpose to improve driving comfort. It is not considered very comfortable if the transmission ratio is constantly adjusting rapidly due to a very active controller. The controller has to be capable though to smoothly damp out any disturbance. The final controller is shown in Figure 5.14. Applying this controller to the $\frac{1}{\pi_{22}}$-loop, results in the open-loop as depicted in Figure 5.15(a) and the Sensitivity as shown in Figure 5.15(b). For completeness also the
With these controllers, it is possible to verify the system response to a step disturbance input \((d_p, d_s)\) of Figure 4.1). During the identification, the applied step is in the order of \(1 \cdot 10^{-4} [Nm]\). This magnitude was chosen this small in order to stay as close to the system operating point (that is to be identified) as is possible. It is interesting to see if the concept of linearity still holds for significant higher input torques than the ones that are used to identify the linearized system. In the Figures 5.17 to 5.20, the system responses are shown to steps of 0.05 and 0.10 [Nm], both in positive and negative directions, and both at the primary disturbance input and the secondary disturbance input. The results that are plotted are for

5.3 Controller validation
the equilibrium point \( r_g = 0.5 \) with \( \nu = 1.9 \% \). Starting with Figure 5.17 (a), the response of \( \theta_p \) to a positive step at the primary disturbance input of magnitude 0.05 and 0.10 [Nm] can be seen. Clearly, \( \theta_2 \) is twice as big as \( \theta_1 \), thereby confirming linearity. Figure 5.17 (b), shows that the response of \( \theta_p \) is exactly the opposite of the response of \( \theta_p \) to a positive step. The responses in this figure are also still linear. Figure 5.18 (a), shows that the response of \( \theta_2 \) is exactly twice the response of \( \theta_1 \) for a positive step at the secondary disturbance input, but Figure 5.18 (b) shows that for negative steps at the secondary disturbance input, the response of \( \theta_p \) is not the exact opposite of the response of \( \theta_p \) to a positive step at the secondary disturbance input, which indicates nonlinear behavior. Despite this, the controller is still capable of attenuating the disturbance. Figure 5.19 shows the response of \( \nu \) to the steps at the primary input. Again a linear response is shown for both positive and negative steps. The response of \( \nu \) is not linear anymore when the steps are applied (both positive and negative) at the secondary disturbance input. Note that although the responses are not linear anymore, the controller is still capable of attenuating disturbances for \( \nu \) within a second, while the attenuation of the disturbance for \( \theta_p \) takes some more time, which is designed this way for driving comfort. Note also that it is generally difficult to judge if a step function on the secondary disturbance input of 0.10 [Nm] is representative for real-life disturbances. It is an interesting conclusion though that the responses of the model show that the concept of linearity is invalid for secondary disturbance steps larger than 0.10 [Nm]. Especially because, as will be seen in chapter 6, a step at the secondary disturbance input lower than 0.10 [Nm] shows no significant response in \( \theta_p \) and \( \nu \) on the prototype due to the effects of friction.

As already noted, it is not guaranteed that steps at either primary or secondary disturbance input represent real life disturbances. One real life disturbance though is a step function on the torque at the wheels. This represents a car whose wheels are from the ground, and thus accelerating quickly, while suddenly hitting the ground again. Actually this represents a possible worst case disturbance, whose occurrence in normal day driving is very limited.
Figure 5.18: Response of $\theta_p$ to a positive (a) and negative (b) step at the secondary disturbance input, around $r_g = 0.5$ and $\nu = 1.9\%$. The applied steps are $\pm 0.05 \ [Nm]$ and $\pm 0.10 \ [Nm]$.

Figure 5.19: Response of $\nu$ to a positive (a) and negative (b) step at the primary disturbance input, around $r_g = 0.5$ and $\nu = 1.9\%$. The applied steps are $\pm 0.05 \ [Nm]$ and $\pm 0.10 \ [Nm]$. 
(a) Secondary positive step.

(b) Secondary negative step.

Figure 5.20: Response of $\nu$ to a positive (a) and negative (b) step at the secondary disturbance input, around $r_g = 0.5$ and $\nu = 1.9\%$. The applied steps are $\pm 0.05 \text{ [Nm]}$ and $\pm 0.10 \text{ [Nm]}$.

Nevertheless, Figure 5.21 shows the response of $\theta_p$ (a) and $\nu$ (b), when a step of $-50 \text{ [Nm]}$ is applied at the wheels. Figure 5.21 (a) shows the response of $\theta_p$ to this step for two different equilibrium points. The response $\theta_1$ corresponds to $r_g = 0.5$ and $\nu = 1.9 \%$, while $\theta_2$ corresponds to $r_g = 2.0$ and $\nu = 0.8 \%$. Note that the deviations from the equilibrium points are shown, in order to be able to compare the responses of two different equilibrium points in one figure. Figure 5.21 (b) shows the response of $\nu$ to the step, at the same equilibrium points as the ones of Figure 5.21 (a). As can be seen, the maximum deviation from the equilibrium point of $\nu_1$ is approximately the same as the deviation of $\nu_2$. Nonetheless the controller is capable of attenuating the disturbance within a second.
Figure 5.21: Response of $\theta_p$ (a) and $\nu$ (b) to a step of 50 [Nm] at the wheels. $\theta_1$ and $\nu_1$ correspond to $r_g = 0.5$ and $\nu = 1.9$ [%], while $\theta_2$ and $\nu_2$ correspond to $r_g = 2.0$ and $\nu = 0.8$ [%]. For clarity, from all responses the nominal value is subtracted.
Chapter 6

Implementation

In this chapter, the idea of Sequential Loop Shaping in the QFT Framework will be extended to the experimental setup. First the experimental setup is described.

6.1 The experimental setup

As is shown in Figure 6.1 the experimental setup exists of two Siemens electrical motors, the EMPAct CVT with its two servomotors as actuators, a Torque Converter and a lubrication system. One Siemens electrical motor acts as an Internal Combustion Engine (ICE) and one acts as the vehicle load. The electrical motor that represents the ICE (called the primary and labelled $EM_I$ in Figure 6.1) is velocity controlled. The electrical motor that represents the load (called the secondary and labelled as $EM_{II}$ in Figure 6.1) is torque controlled and can generate an independent amount of torque and hence simulates a load torque. Both electrical motors have a maximum power of $70 \ [kW]$. The secondary motor (load) is connected with the secondary side of the CVT by means of manual transmission (MT). The primary motor is connected with the primary side of the CVT by means of the Torque Converter (TC). Note that there are two shafts between the outgoing shaft of $EM_{II}$ and the secondary shaft (that is the shaft at the secondary side of the CVT). Not shown in Figure 6.1 are the servomotors that are used for the actuation of the pulleys. In this setup, several variables can be measured, i.e. the rotation of both servomotors $\theta_p$ and $\theta_s$, the torque $T_1$ and $T_2$ and velocity $\omega_{EMI}$ and $\omega_{EMII}$ of both Siemens motors and the velocities of the ingoing and outgoing shafts of the CVT $\omega_p$ and $\omega_s$ can be measured by means of hall-sensors. Note that the velocity of the ingoing shaft $\omega_p$ equals the velocity of $EM_I$, if the torque converter is closed. The torque converter closes at a pressure of $6 \ [bar]$. These $6 \ [bar]$ are generated by means of an independently operating oil pump that is flow controlled. During all measurements, the torque converter is closed. Furthermore it is possible to measure the torque through the shaft that is connected with the outgoing shaft of $EM_{II}$. Although this last torque measurement is not used in this report, it can be useful to determine the efficiency of the CVT. During all measurements the sampling frequency is $1000 \ [Hz]$.

Figure 6.2 shows a picture taken from the side of $EM_I$ (a) and $EM_{II}$ (b). Furthermore, Figure 6.3 shows pictures of the EMPAct CVT, taken from the top (a) and side (b). Although it is rather difficult, from Figure 6.3 the actuation motors can be seen. These are the black units.
CHAPTER 6. IMPLEMENTATION

Figure 6.1: The test rig.

(a) Primary Siemens Motor
(b) Secondary Siemens Motor

Figure 6.2: Photos of the Siemens Motors.
6.2 Primary identification

In the simulation model no Coulomb friction is used (besides of course the Coulomb friction between bands and elements and between elements and pulleys) because that would introduce extra non-linear effects, while one of the purposes of the simulation model is to obtain a linear representation of the input-output behavior. In the experimental setup, Coulomb friction is unfortunately very dominantly present, especially in the screw mechanism that turns the relative rotation between the sun gears into a translation of the movable pulley. This Coulomb friction effect is nonlinear. In order to avoid this nonlinearity the spindles are at constant nonzero velocity during identification. However, due to the moving of the spindles, the transmission ratio $r_g$ will change. Therefore, a reference pattern for the rotation of the primary actuation motor $\theta_p$ is applied in the form of a sawtooth. The idea is to divide the system into several regions (as in the simulation model) and therefore the time-period or the velocity of the sawtooth cannot be too large.

A sawtooth pattern is chosen as is depicted in Figure 6.4. In Figure 6.4(a) the sawtooth of the resulting $\theta_p$ is shown, and in Figure 6.4(b) the corresponding $r_g$ is shown. As can be seen, a typical measurement has a duration of about 200 [sec], in which approximately 20 upward and downward shifting actions take place. The excitation signal is a combination of random noise and a chirp signal. Because the time of shifting with nonzero velocity is so short, the chirp signal is divided into separate measurements. Meaning that during one measurement (of approximately 200 seconds) a chirp with a frequency between 10 and 20 [Hz] is applied, while at another measurement the frequency of the chirp is between 20 and 40 [Hz], and so on. If all frequencies of the chirp signal would have to be combined into a single upshifting or downshifting action, not enough time would be available for the chirp signal to evolve with enough energy at each frequency. An example of the excitation signal is shown in Figure 6.5. The signal that is plotted here contains a chirp with frequencies between 0.1 and 2.5 [Hz]. Obviously this excitation signal is injected in the disturbance input, while $\theta_p$ is in a feedback loop.
Note once more that for the identification, only those parts of the measurement data are taken in which the velocity of $\theta_p$ is nonzero. If the Process Sensitivity $PS(j\omega)$ is then divided by the Sensitivity $S(j\omega)$, the plant $P(j\omega)$ is obtained. Figure 6.6 shows the importance of the chirp signal. The figure shows the resulting plant, when a chirp is applied with frequencies between 20 and 40 $[Hz]$. Obviously, in this region the estimation of the plant looks a lot smoother than in the other frequency regions. This figure shows three estimations for the $(1,1)$-element of the plant $P$. One is for upward shifting, one for downward shifting, and the third one is the average of the two. As can be seen, the difference between upshifting and downshifting is minimal. Note that both the upward shifting and downward shifting estimations are based on an average of approximately 20 estimated plants. Combining the different FRF measurements, the plant estimation of Figure 6.7 is obtained. The transmission ratio $r_g$ of this region is around Low, while the slip level is around 0.5 [%].
Figure 6.6: The positive effect of the chirp signal on the resulting plant estimation.

Figure 6.7: The resulting estimation of the (1,1)-element when combining the different measurements.
Unfortunately, the excitation at the primary side has no significant influence on the measurement of \( \nu \). This is mainly due to the disturbance on \( \nu \) as shown in Figure 6.8 and possible nonlinear effects. Shown in this figure is the measured slip as a function of time. Clearly, a disturbance can be recognized. The frequency of the main content of this disturbance is exactly the frequency at which the secondary shaft is running. The cause for this disturbance is mainly the unroundness of gears in the epicyclic set and a contact effect of the secondary shaft with its bearings. The data shown in Figure 6.8 is measured without a feedback loop of \( \nu \) and without shifting the CVT, so with an open-loop force at the secondary pulley. Whatever the reason is for this disturbance, it prevents an estimation of the non-diagonal element of the plant. The method for the \((1,1)\)-element of the plant can be repeated for different operating points, but because it seems at this point rather more difficult to obtain an estimation for the cross element, the \((2,2)\)-element is tried first.

### 6.3 Secondary identification

Applying a combination of chirp signals with random noise at the secondary disturbance input (like in the primary identification) yields no proper result. Again, the disturbance from the secondary shaft and possible nonlinear effects are dominant. Applying a step as was done in the simulation model, only yields a significant response if the step is more than 0.10 [Nm]. As can be seen in Chapter 5, a step of this size already leads to nonlinear effects (at least in the simulation model). Therefore a different approach is tried. Instead of a step, an impulse is applied. By using an impulse, it is possible to break through the Coulomb friction effect rapidly, thereby making this nonlinear effect less dominant, however, it is still present.

During the identification, the same procedure as with the simulation model is used. First an estimation is tried in an open-loop setup. Then a controller is designed on this estimated system, to proceed with an identification using this controller. Strangely, the estimated system based on the open-loop identification differs from the estimated system based on the closed-loop identification. A convergence to a certain estimation is fortunately obtained (at least in a very narrow frequency band), but this region is not very useful for controller design. To cancel out the dominantly present disturbance coming from the secondary shaft, the excitation impulse (applied at the secondary disturbance input) is repeated several times
(in the order of 50), and then the response is averaged. After averaging, the signal is filtered by using an offline notch filter. This signal conditioning does not guarantee that applying the same impulse gives the same impulse response in different measurements. In the end it comes down to engineering judgement to determine what impulse response represents the system the best. Figure 6.9 shows such a response. It is noted that although this response is representative (because it is measured more often), it is not linear, because applying an impulse that is twice as large, does not have a response that is twice as large. This coincides with the simulation model. The data that is plotted in Figure 6.9 represents the situation without feedback of slip, hence the open-loop identification.

When applying the algorithm of Chapter 4 to the measured data, the estimated (2,2)-element can be seen in Figure 6.10. In this figure, the solid line represents the estimation based on the open-loop identification, and the dashed line represents the estimation with closed-loop identification in the same operating point. Looking more closely at Figure 6.10, it can be seen that the estimations are more equal between 20 and 40 [Hz] than in any other frequency region. It should be noted though, that the notch filter with which the data is filtered has a notch frequency of 24 [Hz] (equal to the number of revolutions per second of the secondary shaft), so making use of the estimation in this frequency region is rather dangerous. In the
low frequency region, the phase of the closed-loop estimated (2,2)-element reads -100 [°], although the magnitude has a horizontal slope here. The phase of the open-loop estimated (2,2)-element goes to 0 [°] for low frequencies however. When applying a controller with an integrating action, the system is close to instability with a very low oscillating frequency, indicating that there is not sufficient phase-margin at that frequency. It seems therefore that the closed-loop measured (2,2)-element is more useable than the one that was measured open-loop. Furthermore, the controller should begin to roll off at a rather low frequency because else it will amplify the disturbance coming from the secondary shaft.

All together it can be concluded that identification of the (2,2)-element is a failure due to two main causes. The first one being the high friction forces in the spindles, and the second one being the disturbance coming from the secondary shaft. Considering the high forces that go through the spindles, there is no other way of solving the friction problem then modifying the actuation system. Other actuation designs have already been prototyped at the TU/e. The problem of the disturbance could possibly be solved mechanically.

Unfortunately, at this point it has to be concluded that controller design as it is applied to the simulation model, is not possible at the experimental setup. Together with the estimations of the simulation model, several measured estimations of the experimental setup and some safety factors preventing excessive slip, a controller can still be designed, although not as described in chapter 5. Validation is then the key to conclude satisfactory stability and performance properties.
Chapter 7

Conclusions and recommendations

7.1 Conclusions

A multi-body model of the EMPAct CVT is discussed. The contact between elements and bands and elements and pulleys is based on Coulomb friction. Although there is a form of viscous friction present at several bodies, none of the bodies contain Coulomb friction effects. The choice not to incorporate a Coulomb offset is made in order to prevent too much non-linear effects, because the simulation model is used to derive a linear approximation of the input-output behavior. The two inputs are the torque from the primary servomotor \( T_p \) and the torque from the secondary servomotor \( T_s \).

The two outputs that need to be controlled are the primary servomotor rotation \( \theta_p \) and the slip \( \nu \) in the variator. Because the detailed simulation model is not useful for controller design purposes, a minimal 2x2 MIMO representation is estimated, based on steps at the disturbance inputs as indicated in Figure 4.1. The resulting estimations are compared with the estimations obtained by applying noise at the disturbance inputs. Both methods obtain similar estimations. The gain and phase variation in the (2,2)-element of the estimated plant is significant. The linearization is performed at all equilibrium points that belong to the optimal efficiency set of [1]. This set is a combination of geometric ratio \( r_g \) and slip \( \nu \).

With the use of the estimated MIMO plants, a diagonal controller is designed, based on Sequential Loop Shaping, in a QFT framework. Despite large variation in the estimated MIMO plants, a satisfactory behavior concerning stability and performance is obtained with the use of a diagonal controller. Load disturbances are attenuated within 1 second for both ratio and slip control.

While applying the method that is used for the simulation model to the experimental setup, it becomes clear that the Coulomb friction in the screws of the EMPAct-CVT is not negligible. Furthermore, a dominantly present disturbance origins from the secondary shaft, resulting in disturbances in the slip measurement. These two problems are the main cause for the failure of the estimation of the (2,2)-element of the plant in practise. Furthermore, in the experimental setup, the non-diagonal elements are far less influential than in the simulation model. Having no proper estimation for the (2,2)-element, no controller can be designed by means of a loop-shaping method as is done for the simulation model.
7.2 Recommendations

The control scheme of the simulation model does not make use of gain-scheduling. Gain scheduling can improve the performance significantly, while maintaining stability. Upon further improving the performance, gain-scheduling might be implemented and tested.

The main problem of the extension from simulation model to experimental setup, is that the (2,2)-element cannot be estimated properly with the methods that are applied to the simulation model. In order to improve the estimation of the (2,2)-element the renewal of for instance the secondary epicyclic gears might be beneficial. Another way to improve the estimation of this term is to give more attention to the estimation of the friction in the screws. Although it is not described in this report, during the actual identification, several excitation signals are applied, i.e. steps, impulses and pulses. In combination with a reduction of the disturbance which acts at the secondary shaft, the application of pulses instead of impulses and steps might result in better system estimations.
Bibliography


62


Appendix A

Step response identification

The estimation on basis of the step-responses makes use of the Markov-parameters of the system, and is widely described in literature, see for example [9], [12], [3], [15] and [13]. The Markov parameters of a system can mathematically be expressed as a function of the state-space matrices that define a dynamical system. The continuous time dynamical system is defined by means of the state-space matrices as follows

\[
\begin{align*}
\frac{d}{dt} x(t) &= Ax(t) + Bu(t) \\
y(t) &= Cx(t) + Du(t)
\end{align*}
\]  

(A.1)

which is a continuous time representation. The discrete equivalent of this state-space representation is

\[
\begin{align*}
z_{k+1} &= Fz_k + Gu_k \\
y_k &= Hz_k + Ku_k
\end{align*}
\]  

(A.2)

It is well known that state-space representations of dynamical systems are not unique, see for instance [16]. A transfer function, however, gives a unique representation of a dynamical system. If the initial conditions of system (A.2) are all assumed zero, the transfer function can be written as a function of the state-space matrices

\[
TF(z) = H(zI_n - F)^{-1}G + K
\]

(A.3)

where \( I_n \) is the n-dimensional (dimension of the state-vector) unit-matrix. Using a formal power series expansion, the transfer function can be written as an infinite sum of Markov parameters

\[
H(zI_n - F)^{-1}G = z^{-1}HG + z^{-2}HFG + z^{-3}HF^2G + \ldots
\]

\[
= \sum_{i=1}^{\infty} D_i z^{-i}
\]

(A.4)

where

\[
D_i = HF^{i-1}G
\]

(A.5)

are the actual Markov parameters. These Markov parameters exactly equal the time elements
APPENDIX A. STEP RESPONSE IDENTIFICATION

of the unit impulse response of the system (A.2). Therefore a triple (F,G,H) that uniquely determines the Markov parameters can be obtained by simply measuring the impulse response. Furthermore it is noted that Markov parameters are insensitive to transformations of the system (A.2), so the Markov parameters are also insensitive to continuous to discrete transformations. Considering again the formal series development of the transfer function

\[ TF(z) = z^{-1}D_1 + z^{-2}D_2 + z^{-3}D_3 + \ldots \]  

(A.6)

the following matrices can be defined to represent this transfer function.

\[
H_E = \begin{bmatrix}
D_1 & D_2 & D_3 & D_4 & \cdots & D_{k_2} \\
D_2 & D_3 & D_4 & \cdots & \cdots & D_{k_2+1} \\
D_3 & D_4 & \cdots & \cdots & D_{k_2+2} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
D_{k_1} & D_{k_1+1} & D_{k_1+2} & \cdots & \cdots & D_{k_1+k_2-1}
\end{bmatrix}
\]

\[
H_A = \begin{bmatrix}
D_2 & D_3 & D_4 & D_5 & \cdots & D_{k_2+1} \\
D_3 & D_4 & D_5 & \cdots & \cdots & D_{k_2+2} \\
D_4 & D_5 & \cdots & \cdots & D_{k_2+3} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
D_{k_1+1} & D_{k_1+2} & D_{k_1+3} & \cdots & \cdots & D_{k_1+k_2}
\end{bmatrix}
\]

\[
H_C = \begin{bmatrix}
D_1 & D_2 & D_3 & D_4 & \cdots & D_{k_2}
\end{bmatrix}
\]

\[
H_B = \begin{bmatrix}
D_1 \\
D_2 \\
D_3 \\
D_4 \\
\vdots \\
D_{k_1}
\end{bmatrix}
\]

(A.7)

Obviously from a practical point of view, these matrices are truncated, instead of having infinite dimensions. Practical values of \( k_1 \) and \( k_2 \) that are used run up to 1500. Increasing further lead to very long calculation times. Together with these newly defined matrices (A.6) can be written in another way

\[
\begin{bmatrix} 0 & H_C \\ H_B & H_A - zH_E \end{bmatrix} \begin{bmatrix} u_k \\ x_{1,k} \\ x_{2,k} \\ x_{3,k} \\ \vdots \end{bmatrix} = \begin{bmatrix} y_k \\ 0 \\ 0 \\ \vdots \end{bmatrix}
\]  

(A.8)

The idea is now that only \( n \) columns and only \( n \) rows of these matrices contribute to the actual response of the system, the rest of them are linearly dependent on these \( n \) columns and rows. In theory you can therefore perform a rank test of these matrices and determine the dimension of the state vector. In practice there is always some noise present amongst
measurements, and this rank test will therefore fail because the other rows and columns will
not exactly be linear combinations of the first \( n \) ones. Instead of the rank test, a singular
value decomposition is therefore performed. As can be seen from (A.7), \( H_A, H_B \) and \( H_C \) are
just parts of \( H_E \), which is called the Hankel matrix in literature. Therefore it is sufficient to
analyze a singular value decomposition of the Hankel matrix, and truncate the other matrices
in consequent ways. In general, for every matrix there exists a singular value decomposition,
so no matter how bad the dynamics or measurements are, the decomposition will be possible.
The idea is to split the matrix into three parts. That is two unitary matrices, \( U \) and \( V \), and
one diagonal matrix in which the upper left corner exists of those singular values that are
significantly bigger than the ones in the lower right corner of the diagonal matrix. Indeed if
no noise is present, the lower right corner will only contain zeros.

\[
H_E = \begin{bmatrix}
U_1 & U_2 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
\Sigma_1 & 0 \\
0 & \Sigma_2
\end{bmatrix}
\begin{bmatrix}
V_1^T \\
V_2^T
\end{bmatrix}
\]  \hspace{1cm} \text{(A.9)}

The decomposition can be performed in a very cost effective way with for instance Matlab. It
is done is such a way that we end op with unitary matrices, that is \( U^TU = I_{k_1} \) and \( V^TV = I_{k_2} \).
In practice the choice of the number of singular values that are significant \( (\Sigma_1) \) is somewhat
subjective, but it is guaranteed that neglecting the lower right corner of the diagonal matrix
results in a small error. Suppose we know that the real system contains 50 singular values,
and we estimate it by, say, 30 singular values, then it can be proven that this leads to a
maximum distance between the estimated and real magnitude of the bode plot of the system
that is equal to twice the sum of the neglected singular values. Therefore we can make our
own choice wether it is worthwhile taking more singular values to be significant at the cost of
a higher order model. The triple \( (F, G, H) \) that represents the Markov parameters can now
be calculated as follows

\[
F = \Sigma_1^{-\frac{1}{2}}U_1^TH_AV_1\Sigma_1^{-\frac{1}{2}}
\]

\[
G = \Sigma_1^{-\frac{1}{2}}U_1^TH_B
\]

\[
H = H_C\Sigma_1^{-\frac{1}{2}}
\]  \hspace{1cm} \text{(A.10)}

The direct feedthrough matrix \( K \) can be estimated by taking the first time value of the
impulse response

\[
K = D_0
\]  \hspace{1cm} \text{(A.11)}
Appendix B

QFT Principles

In this appendix an attempt is made to explain the basics of Quantitative Feedback Theory by the use of a simple example, which has nothing to do with the rest of the work presented in this report. The example in this appendix contains the following SISO linear system

\[ P = \frac{k}{(s + a)(s + b)}, \quad k \in [1, 2, 5, 8, 10], a \in [1, 3, 5], b \in [20, 25, 30] \]  

(B.1)

A bode plot of this set of plants is given in Figure B.1(a). For convenience a Nichols Chart representation is also given in B.1(b), because the Nichols Chart will be the domain in which the design will take place. Clearly some gain and phase variation can be seen. These variations can be thought of as an uncertainty when we take one plant out of the whole set \( P \) as the nominal one, for instance the plant belonging to the parameters \( k = 1, a = 5, b = 30 \).

![Bode plot of \( P \)](image)

![Nichols plot of \( P \)](image)

Figure B.1: Representation of the set \( P \) as Bode plot and as Nichols plot.

Together with this nominal plant, and the gain and phase uncertainties, we can make templates. Templates are areas in the complex plane that represent the uncertainty at one frequency point. Figure B.2 represents such a template at 5 [rad/sec]. If we look at Figure B.1(a) we can see that at 5 [rad/sec] the maximum possible gain in set \( P \) is approximately...
Figure B.2: Template of $P$ at 5 [rad/sec].

$-20 \ [dB]$, while the minimum is approximately $-47 \ dB$. The maximum phase at 5 [rad/sec] is approximately $-55 \ [^\circ]$ while the minimum is approximately $-93 \ [^\circ]$. These values can be recognized directly from figure B.2. The importance of choosing the right nominal plant lies in the procedure of constructing bounds. A bound is a region in the complex plane where the open loop transfer function has to stay out of, in order to meet some specifications, see Appendix C. One such specification is robust stability, say in the form of

$$\left| \frac{PG}{1 + PG} \right| \leq 6dB \tag{B.2}$$

Where $P$ is the set of plants and $G$ is an appropriate controller such that (B.2) is fulfilled. In the Nichols Chart (B.2) would just be a circle to stay out of if we only had one plant, see C. However, we have several plants in the set $P$, and would like to maintain robust stability for all possible plants, using the same controller. To achieve this goal, we take a template, and put a pencil through that point on the template that represents the nominal plant. Next we walk around the 6 dB-circle with our template in such a way that the circle is only touched by one point on the template. While doing this, we keep on marking with the pencil that is pinned through the nominal plant inside the template. If we have made a full 360 degrees around the circle, the bound is finished.

If we now plot our nominal open loop transfer function, and keep an eye to the fact that the open loop at 5 rad/sec does not intersect this bound, we are given complete confidence that robust stability is maintained at 5 rad/sec. The next step is to compute the bounds at other frequencies. Depending on both the choice of nominal plant, and the frequency points, these bounds may look quite different, hence it is important to choose the right ones. Besides bounds for stability, also bounds for performance (for instance Sensitivity) can be calculated. Just suppose we would like to design a controller for the set $P$, aiming at the following two specifications, plus specification B.2
\[
\left| \frac{1}{1 + PG} \right| \leq W_2 \quad \text{(B.3)}
\]
\[
\left| \frac{P}{1 + PG} \right| \leq W_3 \quad \text{(B.4)}
\]

Specification B.3 represents an upper limit for sensitivity, while specification B.4 corresponds to a rejection of disturbance at the plant’s input. If we take for example \(W_2\) to be the function plotted in figure B.3, and \(W_3\) to be equal to say 0.01, then the bounds are shown (together with the nominal plant) in figure B.4(a).

![Figure B.3: Specification on sensitivity.](image)

As can be seen from figure B.4(a), the not all bounds are satisfied, therefore we have to design a controller. Taking for example the controller
\[
G = \frac{379\left(\frac{s}{42} + 1\right)}{\frac{s^2}{2477} + \frac{s}{247} + 1} \quad \text{(B.5)}
\]
gives us figure B.4(b), in which one can clearly see that all bounds are satisfied.
Figure B.4: All bounds on the Nichols Chart together with the nominal plant.
Appendix C

Complex Plane Analysis

In this appendix some complex-vector analysis will be used to show some characteristics of the Nichols Chart. The horizontal axis of the Nichols Chart reads phase, while the vertical axis reads magnitude, in $[dB]$. Normally the open-loop of a system is plot on this Chart, but due to some nice properties derived in this appendix, also closed loop behavior can be analyzed, at the same time. To see this, we analyse the following closed loop function

\[ T = \frac{L(j\omega)}{1 + L(j\omega)} \quad \text{(C.1)} \]

In this equation $L(j\omega)$ is the open loop of a feedback system, thus $L(j\omega) = G(j\omega)P(j\omega)$, with $G(j\omega)$ the Controller, and $P(j\omega)$ the plant. Every complex number contained in $T(j\omega)$, $(T(j\omega)$ is a non-linear function of frequency) can be represented by an exponential function with the use of Euler’s rule:

\[ x + jy = r(\cos(\theta) + j\sin(\theta)) = re^{j\theta} \quad \text{(C.2)} \]

This result will be used further on. We are interested in curves that represent constant closed loop magnitudes, or

\[ |T(j\omega)| = M \Rightarrow \left| \frac{L(j\omega)}{1 + L(j\omega)} \right| = M \Rightarrow \frac{|1 + |L|^{-1}e^{-j\theta}|}{1} = \frac{1}{M} \]

\[ \Rightarrow |L| + \cos(\theta) - i\sin(\theta) = \frac{|L|}{M} \Rightarrow \sqrt{(|L| + \cos(\theta))^2 + \sin^2(\theta)} = \frac{|L|}{M} \]

\[ \Rightarrow \left(1 - \frac{1}{M^2}\right)|L|^2 + 2|L|\cos(\theta) + 1 = 0 \]

The solution to the last equation is the following

\[ |L| = \frac{-\cos(\theta) \pm \sqrt{\cos^2(\theta) - (1 - \frac{1}{M^2})}}{1 - \frac{1}{M^2}}, \quad M \neq 1 \quad \text{(C.3)} \]
Because the solution is not defined for $M = 1$, we must add

$$|L| = \frac{-1}{2\cos(\theta)}, \quad M = 1 \quad (C.4)$$

So now we have a relation between the magnitude and the phase that the open loop $L(j\omega)$ should have, in order to fulfill the closed-loop specification $|T(j\omega)| = M$. These relations can be visualized on the Nichols Chart as curves, as is done in figure C.1. For $M > 1$ these curves are circles. The smallest circle in figure C.1 corresponds to $M = 6 \text{ dB}$, while the horizontal line at the bottom corresponds to $M = -20 \text{ dB}$. All curves in between correspond to $M$-values that lie between $-20$ and $6 \text{ dB}$. The curve corresponding to (C.4) can clearly be distinguished as the curve approaching infinite Open-Loop-Magnitude for $-90$ and $-270$ degrees.
Figure C.1: Constant closed-loop curves on the Nichols Chart.