Using Wave Variables in Time Delayed Force Reflecting Teleoperation

by

Günter Niemeyer

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1996

Author .....
Department of Aeronautics and Astronautics
June 18, 1996

Certified by .....
Professor Jean-Jacques E. Slotine, Thesis Supervisor
Associate Prof. of Mechanical Engineering and Information Sciences

Certified by .....
Professor Wallace E. Vander Velde
Professor Emeritus of Aeronautics and Astronautics

Certified by .....
Doctor J. Kenneth Salisbury Jr.
Principal Research Scientist

Accepted by .....
Professor Jaime Peraire
Chair, Department Graduate Committee

Using Wave Variables in Time Delayed Force Reflecting Teleoperation

by

Günter Niemeyer

Submitted to the Department of Aeronautics and Astronautics
on June 18, 1996, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Adding force feedback to a teleoperation system provides the operator with increased awareness and can considerably improve his or her ability to perform complex tasks, particularly when visual information is limited. However, force reflecting systems have traditionally had severe stability problems when executed in the presence of transmission delays.

Recent work shows that stability can actually be guaranteed for all delays simply by having the communications mimic natural wave carriers, such as flexible beams or electrical transmission lines. The practical applicability of such an approach, however, is conditioned by transient characteristics and general system performance. It is the goal of this thesis to explore these issues, especially for large delays up to several seconds, and to establish a force feedback teleoperator that allows reliable and predictable operation.

To support such efforts, we develop the concept of wave variables based on the physically motivated theory of passivity. Wave variables encode information in a unique manner and automatically provide robustness to delays. After examining reflections and impedance matching, we propose a new configuration for the teleoperator, which includes a tuning mechanism to trade-off the clarity of force reflection with the speed of operation. We further explore the new wave domain by designing controllers directly from this point of view. This ranges from simple replacements of position controllers to the inclusion of predictive elements using a priori information about the system.

Delays impose fundamental limitations on the achievable performance of any technique. Thus, we pursue the goal of a virtual tool rather than telepresence, trying to give the teleoperator a simple appearance rather than hiding dynamics completely.

The developments are verified experimentally on a three degree of freedom system.

Thesis Supervisor: Professor Jean-Jacques E. Slotine
Title: Associate Professor of Mechanical Engineering and Information Sciences

Acknowledgments

First and foremost I would like to express my gratitude to Prof. Jean-Jacques E. Slotine. As an advisor he always made time, had lots of suggestions, and was great at pointing out the big picture or finding that fundamental flaw. As a friend he sparked many a discussion. And as both he was patient and forgiving, letting me get away with more than I probably should, and supportive of anything I did.

I am also very grateful to the other members of my committee, Dr. Kenneth Salibury and Prof. Wallace Vander Velde. They always provided encouragement and sage advise and never failed to accommodate my often suboptimal scheduling. Thank you.

And then there are the many friends I have made over the years that make the MIT experience truly worthwhile. From the current generation of Jesse Hong, Akhil Madhani, Craig Latimer, Brian Anthony, to the previous graduates including Brian Eberman, John Morrell, Rob Sanner, Hyun Seok Yang, Nobuhiko Mukai, to the guys who knew me way back when, Andre Sharon, Ted Clancey, Weiping Li, Joe Deck, and Charles Oppenheimer. Each and every one was always ready and willing to lend an ear, a helping hand, or just plain support. And, of course, life would not have been the same without the volleyball crowd who always made sure I remembered the lighter side of things.

Finally, but most definitely not least, I wish to thank my family, both immediate and not so immediate. Without the suggestions and support of several uncles, and without the strong support of my parents and my sister, I probably would never have come here.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Telerobotics encompasses a large body of technology and has enjoyed a long and rich history (Sheridan, 1989, 1992). It provides a merging point between human operations and autonomous machinery and combines developments from many fields of study, such as mechanical and electrical engineering, cognitive sciences, control theory, machine design, and many others. Its applications range from handling of hazardous material, undersea exploration, and space station maintenance to micro surgery and virtual reality.

Among these many applications, several involve large distances justifying the prefix 'tele'. Others may involve slow or restrictive data transfers. Both such situations can result in substantial delays between the time a command is issued by the operator and the time it is executed by the remote robot. The same effect appears in feedback signals from the remote cite.

Such delayed action and feedback can pose serious problems to a teleoperator. In mild cases, it interferes with and corrupts the normal perception of the remote world. In severe cases, it may drive a system to wild oscillations and eventually even unstable. It is such problems that we wish to address with this work.

1.1 Force Reflecting Time Delayed Teleoperation

The general terms of telerobotics and teleoperation describe systems involving two distant, yet coupled robots. A human operator moves and controls the local robot, also known as the input or master joystick. Motion commands are measured on this device and transmitted to the remote location. The second robot, usually called the slave, executes these commands and tries to track the input device. Other types of

commands may also be added, such as voice, keyboard, or mouse inputs.

Feedback may be provided to the operator in many different forms, including visual, audio, or tactile. Of particular interest is force feedback, which presents contact information to the human as forces on the input joystick. Thus both robots share not only their motion but also their forces. Under ideal conditions, this gives the operator a sensation equal to actually touching the remote environment. It can considerably improve the ability to perform complex tasks interacting with the environment, for example assembly or surface following and inspection.

Such goals are often extended with the concept of telepresence, in which case the operator is provided with sufficient information to create the perception of being located at the remote location. Virtual reality falls along very similar lines, except the physical remote setup is simulated in a purely artificial world.

In all cases the existence of time delays hinders the normal operation and distorts the perceived reality (Sheridan, 1993). In many cases the human will temporarily suspend inputs, in effect disconnecting any closed feedback loops, and wait until his senses have settled before resuming. Unfortunately, the combination of force feedback and time delay poses more serious threats. By displaying the contact forces on the very input device, the system creates an internal and uninteruptable feedback loop separate of the human. Though the operator can often help stabilize this system, the tendency to instability can become overwhelming and at best is a distraction from the task at hand.

In fact many people believe that force reflection and time delay are mutually exclusive for these very reasons. However, recent work (Anderson and Spong, 1989b; Niemeyer and Slotine, 1991b) has demonstrated that force reflection can indeed be included in time delayed systems in a stable fashion. This is based on the physically motivated theory of passivity and essentially mimics natural systems which are both bilateral and delayed, yet stable. Examples are electrical transmission lines and flexible mechanical structures.

It is on these stability results that we wish to build. While stability must be guaranteed, it is not sufficient in itself. After all, simply cutting the force feedback loop assures stability, but also prevents any meaningful performance. Instead, the practical applicability is determined by transient characteristics and general system performance. It is the goal of this thesis to explore these issues, especially for large delays up to several seconds, and to establish a force feedback teleoperator that allows reliable and predictable operation.

To support such efforts, we develop the concept of wave variables based on the physically motivated theory of passivity. Wave variables encode information in a

unique manner, which automatically provides robustness to delays. They also present a new and different point of view for both analyzing and designing passive and delayed systems.

After examining wave reflections and impedance matching, we propose a new configuration for the teleoperator. This includes a tuning mechanism to trade-off the clarity of force reflection with the speed of operation. We further explore the new wave domain and continue to improve the system by designing controllers entirely in terms of wave variables. This ranges from wave filters and simple replacements of position controllers to the inclusion of predictive elements using a priori information about the system.

1.2 Virtual Tool Design Goal

Clearly the time delays impose fundamental limitations on the achievable performance, regardless of the individual technique. In particular, the reaction to an unknown disturbance or an environment contact cannot take effect in a time less than the total round trip delay $2T$. In effect, the closed loop bandwidth will be limited by the same time constant.

As such, it is all the more important to present the operator with a simple and predictable system. Any unexpected behavior would lead to distractions and further complicate the situation. Instead the operator should be free to focus his or her entire concentration on the task at hand and not worry about the details of the telerobots. The appearance of the system may even be compared to a video game, which is simple to use and has just enough features to complete the desired work.

To account for the limitations as well as these objectives, we pursue the goal of a virtual tool. This tries to modify the entire dynamics of the teleoperator system into a simple and well known form. It is in contrast to the notion of telepresence, which hopes to hide all dynamics completely and is not achievable for delayed systems.

Such a virtual tool shows both basic inertial and stiffness characteristics. The magnitude of these effects is determined by the delay time. For example, a small delay would make the system appear much like a light and stiff screwdriver.

As the delay time increases the tool becomes softer and/or heavier, so that the bandwidth of the overall operation is reduced appropriately. A large delay can thus be turned into either a heavy drill or a soft sponge. Both provide a reduced performance but with different emphases. Indeed such a heavy versus soft trade off appears naturally as a tuning parameter throughout the following work.

In practice this tuning may even occur in real time. If no contact is made, for example during a visual inspection, the virtual tool should be tuned to a soft sponge making quick motions easy. Then, in close proximity to a desired contact location, the tool should be adjusted to a heavy and stiff appearance, allowing good force clarity under slow speeds.

In general it is important to recognize the delay induced limitations and reduce the performance goals accordingly. Fortunately, the wave variable approach does this automatically. We are left with the task of creating a simple behavior within the given bounds.

1.3 Time Scaling and Practical Limits

While we are able to guarantee stability for any amount of time delay, teleoperation ceases to be useful above some maximum practical value. Obviously a ten minute delay will not permit operation in any normal sense. In the following we try to estimate and justify what this limit might be.

Consider first an undelayed system. The human operator is the slowest element with a typical reaction time on the order of 0.1 seconds. Any task is therefore executed at speeds based on this reaction time and the command signals typically contain only low frequency components below a few Hertz.

Now introduce a delay of 1 second. In effect the total reaction time has grown to 1.1 seconds, or increased eleven fold (Minsky, 1992). To allow the same reactions as before, the speed of operation should be reduced accordingly. Equivalently, the total time to complete a task should also increase eleven fold.

To better understand the effects, consider the added delay as a time scaling by a factor eleven. And compare this to classic dimensional scaling. For example, doubling the dimension of an object increases its inertia by a factor eight. So in the simplest setting, mass is proportional to the cube of size ($kg \sim m^3$). But to operate at the same force levels, this increased object must be moved slower. From Newton's equation $F = ma$, mass must be inversely proportional to acceleration and thus directly proportional to the square of time ($kg \sim s^2$).

And so we argue that the effects of increased dimension, being increased mass and slower operation, are very much the same as the effects of added delay. In essence, time scaling is similar to space scaling.

Meanwhile, handling of oversized objects at low speeds is performed in numerous scenarios. Consider astronauts retrieving large satellites. Clearly the inertias are very

high compared to the forces the astronauts can apply and the slow speed is evident when watching the orbital work. The same is true for navigating large ships, such as supertankers. Any course corrections or other changes take long periods of time to achieve. And similar problems appear controlling hot air balloons. In all cases the increased physical dimensions also increase the temporal requirements.

This leads us to believe that operating a time delayed system is as difficult as handling large and heavy objects. If astronauts and supertanker pilots can fulfill their jobs, then so should teleoperators. Indeed some later results show that a delay can behave very much like an inertial element. Of course, this emphasizes that the operations will not appear exactly like their undelayed counterparts and may require some adjustments or learning. And this is also consistent with the notion of a virtual tool discussed above.

In practical terms, a one second delay implies an eleven fold slowdown and is still within acceptable ranges. But a ten second delay would force a more than 100 fold slowdown and exceeds this limit. For example, a task, which ordinarily takes one minute, would require nearly two hours. We conclude that operations with delays above a few seconds are not useful. At the other extreme, in a properly designed system, delays below the human reaction time should be hardly noticeable. And indeed our experimental findings support these conclusions.

1.4 Teleoperator Setup and Assumptions

Our work focuses on force reflecting teleoperators in the presence of time delays. The traditional approach and layout of such a system is shown in Figure 1-1. At the local site, the operator interacts with an actuated joystick, allowing him to both issue commands to the system and receive force information from the system. The motion commands are measured and transmitted to the remote location, where they are used control the slave manipulator. Simultaneously, the forces used by the telerobot are returned and applied to the master.

The time delay is embedded within the communications. This may be caused by large distances, for example in orbiting telerobots, or by slow data transfer rates, as with underwater acoustic links. Also computational delays needed to compress or decompress images and transfer data via a complex network are acceptable. Or these methods can be used simply to provide robustness against the possibility of delays.

Additional comments and assumptions on the system are:

- The environment is unknown but passive. The simple lack of energy sources

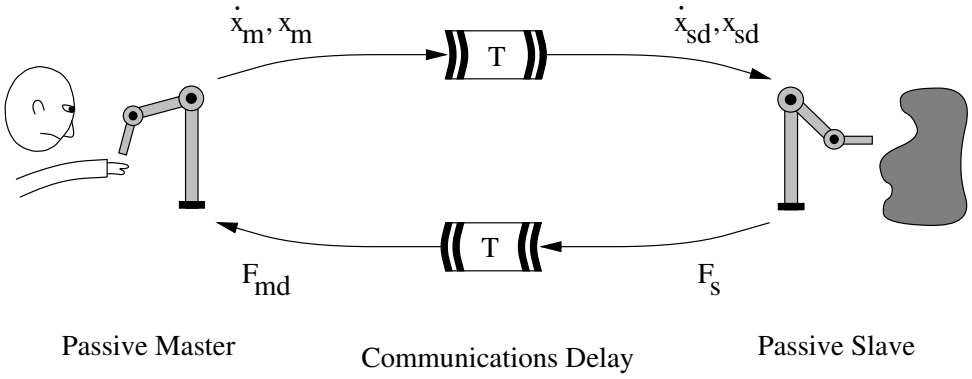


Figure 1-1: In a traditional teleoperator system, the position and velocity are measured at the master joystick, transmitted to the slave, and used as desired commands. Meanwhile the force is measured at the slave, fed back and applied to the operator. The system is asymmetric and subject to stability problems.

can verify this. It is the most general assumption that can be made on the environment and may lead to problems only if the telerobot is to interact with other controlled and powered machinery.

- The system should appear passive to both the environment and the human operator. This provides good robustness to unknown contacts and is even necessary if we allow all possible passive environments. Also, while humans are not passive themselves, they typical interact well with passive systems.

- Both the joystick and the telerobot are fully actuated and have position and velocity measurements. They need not have the same kinematics, instead their cartesian endpoint should be controlled. The nonlinear model including gravity compensation can be included.

- The mechanisms are dynamically clean and backdrivable. This implies that any external interaction forces will result in motion if left uncompensated. Or equivalently, the actuator forces provide a good representation of any contact forces. If this is not the case, as with highly geared robots, force sensors may be used locally at the mechanism to hide the unwanted dynamics and make the system effectively appear backdrivable.

- Both local and remote sites have computers that can implement the necessary algorithms and control the mechanical hardware. At the basic level, the algorithms are extremely simple and do not require much CPU power. However, they do require computation on either side of the time delay.

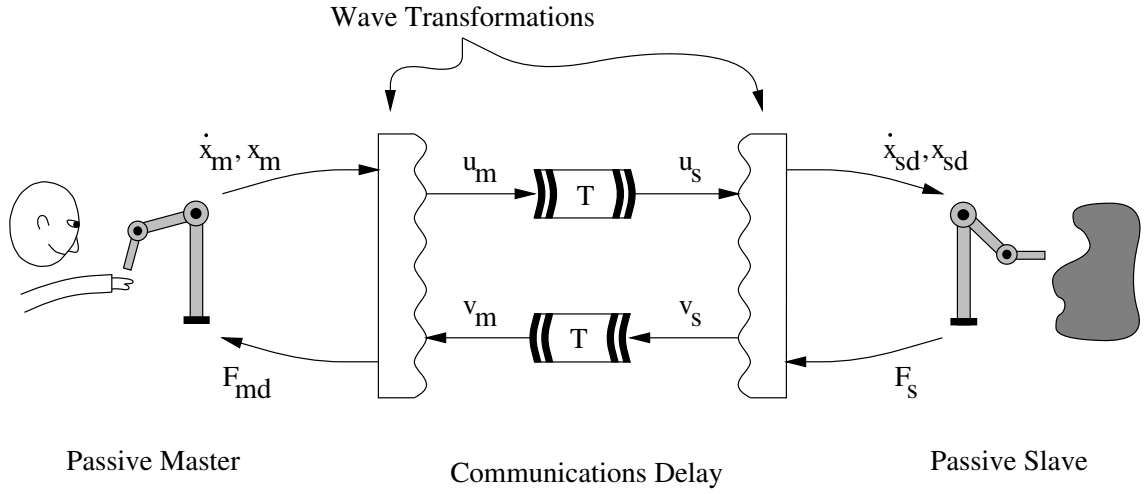


Figure 1-2: The wave based teleoperator inserts transformations at both sides of the communications, encoding the data into wave variables. Transmitting these representations in place of the standard variables, the system is stabilized regardless of the time delay.

This may be relaxed on the slave side, if sufficient natural damping is present. This can also leads to extensions beyond teleoperation where the delay cannot be isolated from the other dynamics.

- The amount of the delay need not be known, though such knowledge may help to tune some of the advanced schemes. It is generally assumed to remain constant, but adjustments are possible to account for time varying delays.

1.5 Wave Based Approach

While traditional teleoperators work well without delays, they are prone to stability problems even for small delays. We attack this problem with the introduction of wave variables, which are particular combinations of velocity and force signals. They present a modification to the theory of passivity, making them applicable to nonlinear systems. And by encoding information in this unique manner, they create robustness to arbitrary delays.

Transforming all information into wave variable form and transmitting this data in place of the standard quantities stabilizes the system regardless of delay. The new system has added wave transformations at both local and remote site and is depicted in Figure 1-2. These transformations are algebraic and simple to implement. And for

small delays below the human reaction time the changes are transparent.

For larger delays the updated communications create a new set of dynamics, which may cause oscillations and can interfere with the tasks at hand. We analyze these in the wave domain and develop the notion of wave reflections. They are countered by impedance matching, wave filtering and wave controllers. The teleoperator layout is updated accordingly.

Wave variables provide an implicit reduction in performance as the delay increases as well as a tuning mechanism to trade-off the clarity of force reflection with the speed of operation. This nicely fits the physical limitations imposed by the delay and our virtual tool design goals. They may also be used to include other classical performance enhancements into the system, such as predictors and sensory substitutions. Especially in cases with a priori knowledge of the environment, such methods can prove beneficial.

1.6 Experimentation

This work is demonstrated in various experiments, which we present and describe as they become relevant to the individual topics in the following chapters. At this point we briefly introduce the physical hardware and general settings common to all experiments.

We are fortunate to be able to use two light weight, cable driven, manipulators with very clean dynamics for both the master joystick and the slave robot. They were developed at the MIT Artificial Intelligence Laboratory and are known as 'Tool Handles'. Each robot has three degrees of freedom, which we use to control its endpoint position along the cartesian axes. Figure 1-3 shows a photograph of the setup.

While both arms share the same kinematic layout, they are installed at opposing orientations. As the system tracks their endpoint movements, the joint trajectories will differ substantially. Forward and inverse kinematics are implemented to support this behavior.

The link lengths are approximately half a meter in length, giving the manipulators a work space of several cubic feet. The fully extended arm can reach nearly one meter from its base. The single stage cable mechanisms implement a gear ratio of $14:1$, so that the motors can generate a force output at the endpoint of up to 15N. Position measurements are available from high resolution encoders and are filtered and differentiated numerically to compute velocity signals.

Figure 1-3: The experiments are performed using two robots with three degrees of freedom each. The master is shown on the left reaching up, while the slave is on the right interacting with objects on the floor. The controllers match their cartesian tip motion, so that the user may ignore the kinematics.

The cable drive provides an efficient force transmission between motors and links, resulting in the clean dynamics. This also implies that open loop force control provides good results and circumvents the need to explicit force sensors and local force controllers. We feel confident that the force applied by the motor is a very good measure of the interaction force at the endpoint.

The motors and other inertia are located close to the base of the robot. The effective inertia values along the cartesian axes range between 0.1kg and 0.2kg, which highlights the low mass of the manipulator. The position and velocity feedback gains, used in accordance with the individual controller layouts, are tuned to provide good behavior of a traditional system without delay. They are valued at approximately $K = 400$ and $B = 10$ respectively. All values are given in S.I. units.

Both manipulators are connected to a P.C. computer operating under linux with software written in C. The sampled data is time stamped to avoid servo rate uncertainties of the control loops, which execute at about 5KHz. Any delay is simulated by a large circular buffer. And while both robots are controlled by the same program, they remain effectively isolated as is required by teleoperation.

A typical set of experimental results is shown in Figure 1-4. The six plots present position and force measurements along all three cartesian axes. Within each diagram, the master values are given by the solid line while the slave follows the dashed line.

In this case, the task was executed in two distinct phases. First the joystick was moved across the horizontal plane formed by the x and y axes. The 1 second delay shows the characteristics of an inertial element. Notice the input forces during both acceleration and deceleration phases. Meanwhile the slave lags behind and overshoots at those times respectively. During the intermediate constant motion (for $t = 3 \ldots 4$), the forces are low and the position tracking is good.

The second portion of the experiment presses against a solid slave contact in the vertical z direction. The large deflection produces a constant torque, illustrating the spring like characteristics of the delay. The smaller deflections and forces along the horizontal plane stem from contact friction.

Also note that some experiments allow visual feedback. To avoid complex buffering of visual signals as well as cameras and monitors, we use the standard technique of doubling the delay in the forward transmission and eliminating it in the reverse direction. This is possible because both sites are completely isolated and autonomous. Absolute time has no relevance, instead only the total round trip delay time is important. And so the visual and mechanical information appears simultaneously to the operator, as is expected.

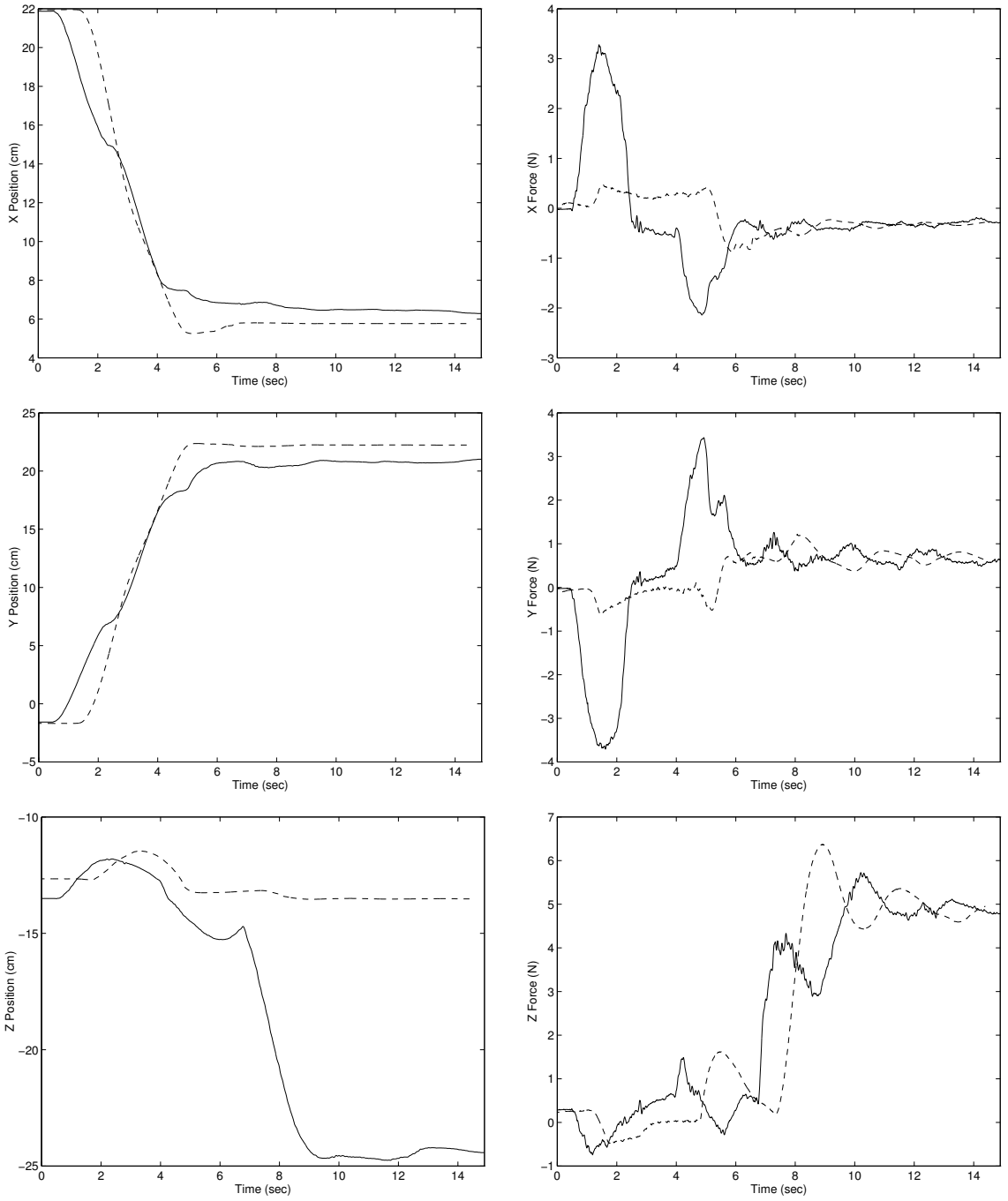


Figure 1-4: These typical experimental results show the wave based delayed teleoperator in two phases. First a motion is executed across the x-y plane. Then pressure is applied against contact in the z direction.

1.7 Outline

This document follows an approach of alternately developing tools and applying them to the teleoperation problem. We begin in Chapter 2 with the review and development of some basic tools used to analyze time delayed systems. While passivity forms the main element, attention is also paid to classical linear time invariant methods that can illustrate some of the basic system behaviors.

In Chapter 3 we examine the basic teleoperation problem and study its stability limits. Passivity leads us to find that the delayed communications creates power and drives the instabilities.

The main thrust of our work is developed in Chapter 4 in the form of wave variables. They are based on passivity and incorporate robustness to time delays. We present a complete description and begin to analyze some basic elements in the wave domain. Using these variables, we update the teleoperator in Chapter 5. This leads to a better understanding of the delay and its effects on teleoperation, and to an improved configuration.

We immerse ourselves completely in the wave domain in Chapter 6. It begins to design elements from scratch in wave space, ranging from simple a wave filter to more complex wave controllers. It also suggests additional applications in the form of wave junctions and predictors.

And finally we conclude in Chapter 7 with a brief summary, additional recommendations and comments.

Chapter 2

Background

Presented in this chapter are various tools we use to analyze time delayed systems. Passivity forms the main element and is the most general, handling both nonlinear systems and interactions with unknown environments. We will use it frequently and extend it with wave variables in Chapter 4.

However, we also examine several methods unique to linear time-invariant systems. While not as general, they illustrate some of the fundamental characteristics introduced by time delays. Sections 2.2.2 and 2.2.3 discuss root loci with respect to delay and gain, followed by a brief review of the Nyquist criterion in Section 2.2.5, and concluded in Section 2.2.6 detailing the use of rational approximations.

2.1 Passivity

The passivity formalism represents a mathematical description of the intuitive physical concepts of power and energy. It provides a simple and robust tool to analyze the stability of a system based solely on input-output properties. By avoiding internal state models, passivity can be applied to a large variety of systems, for example linear, nonlinear, continuous time, discrete time, distributed, or even noncasual systems. It also allows for connections between various subsystems while maintaining global stability properties. These generalities make it well suited for our application, especially regarding interactions with unmodeled environments and human operators.

Passivity has been presented in many ways, for instance see (Slotine and Li, 1991), (Desoer and Vidyasagar, 1975), or (Popov, 1973). The following summary is geared to the our purposes and assumes continuous, causal, possibly nonlinear systems.

2.1.1 Passive Systems

Intuitively, a system is passive if it absorbs more energy than it produces. To formalize this, define the "power input" P_{in}, which is positive when entering a system, as the scalar product between the input vector $\mathbf{x}$ and the output vector $\mathbf{y}$ of the system.

$$P_{in} = \mathbf{x}^T \mathbf{y} \tag{2.1}$$

Note that this power does not necessarily correspond to any physical power. Only for specific choices of input and output variables, such as velocity and force signals, is this the case.

The input power should either be stored or dissipated in the system. Define a lower bounded "energy storage" function $E_{store} \geq E_{min}$ and a nonnegative "power dissipation" function $P_{diss} \geq 0$, which again need not represent true physical quantities. Without loss of generality, we may set $E_{min} = 0$ as the energy storage function can always be shifted by an additive constant. A system is passive, if it obeys

$$P_{in} = \frac{\mathrm{d}}{\mathrm{d}t} E_{store} + P_{diss} \tag{2.2}$$

In other words, the system can not generate energy and provide only as much as was stored initially. This passivity condition is also often expressed in the integral form

$$\int_0^t P_{in} \, \mathrm{d}\tau = E_{store}(t) - E_{store}(0) + \int_0^t P_{diss} \, \mathrm{d}\tau \geq -E_{store}(0) \qquad \forall t \geq 0 \tag{2.3}$$

If the power dissipation is zero for all time, regardless of input or output, the system is also termed lossless. In contrast, if the power dissipation is positive during any power transfer, the system is dissipative.

2.1.2 System Connections

One of the nicest features of passivity is its ability to connect two passive elements into a single passive unit. This occurs if the connection is made in either a feedback or parallel configuration (see Figure 2-1). These configurations are unique, because the total power input P_{in} is automatically divided into the individual elements.

$$P_{in} = P_{in,1} + P_{in,2} \tag{2.4}$$

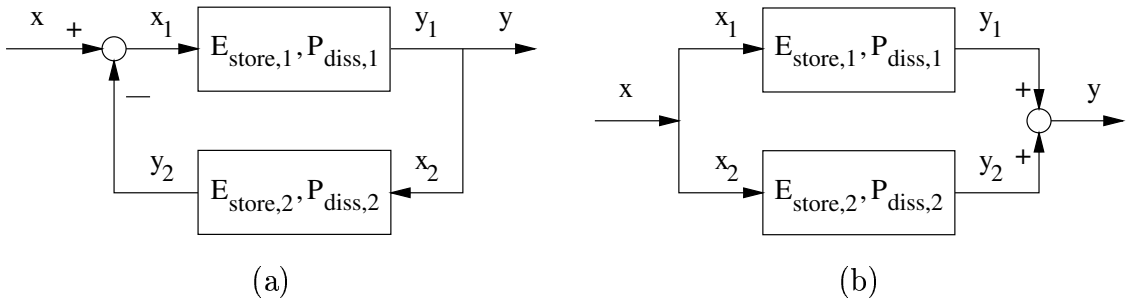


Figure 2-1: Both (a) feedback and (b) parallel connections of passive systems are themselves passive.

Substituting (2.1) we require that

$$\mathbf{x}^T\mathbf{y} = \mathbf{x}_1^T\mathbf{y}_1 + \mathbf{x}_2^T\mathbf{y}_2$$

which is quickly verified for both the feedback ($\mathbf{x}_1 = \mathbf{x} - \mathbf{y}_2$, $\mathbf{x}_2 = \mathbf{y}_1 = \mathbf{y}$) and parallel ($\mathbf{x} = \mathbf{x}_1 = \mathbf{x}_2$, $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$) layouts.

This additive property of the power input P_{in} is also extended to the stored energy E_{store} as well as the power dissipation P_{diss} functions.

$$E_{store} = E_{store,1} + E_{store,2} \qquad P_{diss} = P_{diss,1} + P_{diss,2} \tag{2.5}$$

Therefore, if both elements satisfy the passivity condition (2.2), so will their connection.

By induction, many passive elements can be combined in this fashion without loss of the global stability properties. In every case the total energy and power dissipation is summed over all individual elements. This also allows connection of a passive system to an arbitrary passive environment, as we will see later.

2.1.3 Cascading of Two-Port Elements

In the later sections we will construct the teleoperator from a large number of passive elements arranged in both feedback and parallel configurations. They will reach from the human operator (to be depicted at the extreme left) all the way to the remote environment (at the far right). Figure 2-2 sketches a three-element example.

To facilitate the development and simplify the notation of these cascaded loops and to aid the introduction of wave variables in Chapter 4, we proceed with some additional definitions. First, we define a major direction, along which positive power

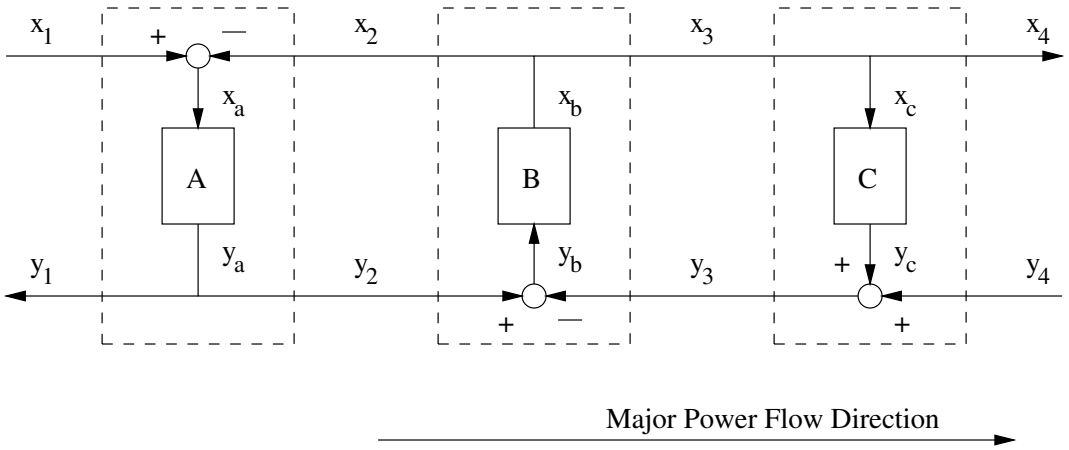


Figure 2-2: Introducing two-port elements allows simple cascading while hiding the specifics of feedback or parallel loops. Power is considered positive when flowing along the major direction from left to right.

will flow. Generally this direction will be from left to right, thus from the human operator to the remote environment.

Next we establish two-port elements, which provide a separate interaction port on both their left and right side. Positive power will enter a two-port element from the left and exit to the right. These elements are shown by the dashed lines in Figure 2-2.

The total power input P_{in} for a two-port element is given in two parts

$$P_{in} = \mathbf{x}_l^T \mathbf{y}_l - \mathbf{x}_r^T \mathbf{y}_r \tag{2.6}$$

where we have labeled the left port with 'l' and the right port with 'r'. Notice the power for the right port is negated as it is exiting the system. As before this power input is substituted in (2.2) or (2.3) to find explicit conditions for passivity.

In essence, these definitions distinguish between power flow P along the main direction and power input P_{in} into an element. At any left port, power flow and input are in unison and no changes occur. But at any right port, they are opposed and a negation appears. We could reconcile the single-port (2.1) and two-port (2.6) definitions of the power input by recombining both ports.

$$P_{in} = \begin{bmatrix} \mathbf{x}_l^T & \mathbf{x}_r^T \end{bmatrix} \begin{bmatrix} \mathbf{y}_l \\ -\mathbf{y}_r \end{bmatrix} = \begin{bmatrix} \mathbf{x}_l^T & -\mathbf{x}_r^T \end{bmatrix} \begin{bmatrix} \mathbf{y}_l \\ \mathbf{y}_r \end{bmatrix} \tag{2.7}$$

Nevertheless, either the right port input or output vector would be negated.

The advantage of using the two-port elements and their power input formula (2.6) stems from this automatic negation. It implies that a simple cascaded sequence of passive two-port elements is also passive. Indeed, in Figure 2-2 the individual power inputs are

$$P_{in,A} = \mathbf{x}_1^T \mathbf{y}_1 - \mathbf{x}_2^T \mathbf{y}_2 \qquad P_{in,B} = \mathbf{x}_2^T \mathbf{y}_2 - \mathbf{x}_3^T \mathbf{y}_3 \qquad P_{in,C} = \mathbf{x}_3^T \mathbf{y}_3 - \mathbf{x}_4^T \mathbf{y}_4$$

which combine for the total power input

$$P_{in} = P_{in,A} + P_{in,B} + P_{in,C} = \mathbf{x}_1^T \mathbf{y}_1 - \mathbf{x}_4^T \mathbf{y}_4$$

The additive nature of energy storage and power dissipation functions persists as well. The summation junctions, which appear explicitly in the single-port feedback and feedforward configurations and which are needed there for the closure property (2.4), are now embedded inside the two-port elements.

Finally, we reiterate that the effect of these definitions is purely notational. They do not change the dynamics or the passivity of a system. For example, consider element B in Figure 2-2. The two-port power input is given by

$$P_{in,B} = \mathbf{x}_2^T \mathbf{y}_2 - \mathbf{x}_3^T \mathbf{y}_3$$

But substituting the internal junctions $\mathbf{y}_b = \mathbf{y}_2 - \mathbf{y}_3$ and $\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}_b$ we find

$$P_{in,B} = \mathbf{x}_b^T \mathbf{y}_b$$

which represents the underlying one-port power input. Consequently also neither the stored energy nor power dissipation functions are altered.

2.1.4 Positive LTI Systems

Within the context of linear time-invariant (LTI) systems, the condition for passivity can be expressed many other forms, providing many different tests. Given its input-output nature, we find it only fitting to examine passivity in the Laplace domain where the following result is well known. A causal LTI system is passive if and only if it is positive, meaning its transfer function or transfer matrix is positive real.

Definition of Positive Real. Consider the $m \times m$ transfer matrix $G(s)$ of a multi-input multi-output (MIMO) system. Assume $G(s)$ analytic everywhere, with the

exception of isolated singularities. It is positive real if the Hermitian matrix

$$G(s) + G^*(s) \quad \text{is positive semidefinite for all} \quad \text{Re}(s) \geq 0 \tag{2.8}$$

where G^* denotes the complex conjugate transpose of G. As transfer matrices of physical systems contain only real elements, the complex conjugation can be passed through to the argument $G^*(s) = G^T(\bar{s})$.

For a single-input single-output (SISO) system with the transfer function $G(s)$, the condition collapses to

$$\text{Re}\left(G(s)\right) \geq 0 \quad \text{for all} \quad \text{Re}(s) \geq 0 \tag{2.9}$$

This implicitly prohibits unstable, right half plane poles, as any complex function takes on positive and negative real components in the immediate neighborhood of any singularity.

Frequency Response Test. It is rarely practical to examine a transfer function in the entire right half plane, as suggested by the above definition. Fortunately, other tests are available. The most common analyses the frequency response. It finds a transfer function or matrix $G(s)$ is positive real if, and only if,

(a) $G(s)$ is stable, thus has no right half plane poles (2.10a)

(b) any poles of $G(s)$ on the imaginary axis are simple with real and nonnegative residues (2.10b)

(c) $\text{Re}\left(G(j\omega)\right) \geq 0 \quad \text{for all} \quad \omega \in \Re$ (2.10c)

For physical systems $G(j\omega)$ is symmetric and we can further limit condition (2.10c) to nonnegative frequencies $\omega \geq 0$. And for MIMO systems, it applies to the Hermitian $H(j\omega) + H^*(j\omega) \geq 0$.

This test can also be derived directly from the condition for passivity (2.3) via Parseval's theorem, which relates the power in time and frequency domains.

Properties of Positive Realness. Satisfying the above requirements has several noteworthy implications for a transfer function, including

- $G(s)$ is minimum phase, i.e. the right half plane contains neither poles nor zeros.
- The phase shift $\angle G(j\omega)$ is limited between $\pm 90°$

- The relative degree is -1, 0 or 1.
- The Nyquist plot of $G(j\omega)$ lies entirely within the closed right half plane.

Relation to State Space. Finally, using the Kalman-Yakubovich or positive real lemma, we can transfer these results to the state space domain. A transfer matrix $H(s)$ relating an input $\mathbf{u}$ to an output $\mathbf{y}$ and with a state representation

$$\begin{aligned} \dot{\mathbf{x}} &= A\mathbf{x} + B\mathbf{u} \\ \mathbf{y} &= C\mathbf{x} \end{aligned}$$

is positive real if there exist a symmetric positive definite matrix P and a symmetric positive semidefinite matrix Q, such that

$$\begin{aligned} A^T P + PA &= -Q \\ PB &= C^T \end{aligned}$$

2.1.5 Scattering Operators

Scattering operators link passivity to the small gain theorem. They were originally motivated by linear circuit theory, but have been extended to nonlinear systems. They also form the main inspiration for the introduction of wave variables in Chapter 4.

Without developing the rigorous definitions needed to properly discuss operators, let us simply state the following results (Desoer and Vidyasagar, 1975). Assume the operator H maps the input vector $\mathbf{x}$ to the output vector $\mathbf{y}$ for some system. The scattering operator is then defined as

$$S = (H - 1)(1 + H)^{-1} \tag{2.11}$$

where it is assumed that the inverse of $(1 + H)$ is well defined. Rather than mapping input to output directly, the scattering operator relates their sum and difference

$$\mathbf{y} = H\mathbf{x} \qquad \Rightarrow \qquad (\mathbf{y} - \mathbf{x}) = S(\mathbf{y} + \mathbf{x})$$

Building on these definitions, a system is passive if and only if its scattering

operator has a norm less than or equal to unity. Within our context this leads to

$$||S||^2 = \max_{\mathbf{u}\neq 0} \frac{||S\mathbf{u}||_T{}^2}{||\mathbf{u}||_T{}^2} = \max_{\mathbf{u}\neq 0} \frac{\int_0^t (S\mathbf{u})^T (S\mathbf{u})\,\mathrm{d}\tau}{\int_0^t \mathbf{u}^T\mathbf{u}\,\mathrm{d}\tau} \leq 1 \qquad \forall t \geq 0 \tag{2.12}$$

For LTI systems, the scattering operator can be expressed in the Laplace domain, relating to the transfer function $H(s)$ by

$$S(s) = \frac{H(s)-1}{H(s)+1} \tag{2.13}$$

And its norm can be obtained from its frequency response, i.e. along the imaginary axis. Hence a system is passive if and only if

$$||S||^2 = \sup_{\omega} ||S(j\omega)||^2 = \sup_{\omega} \lambda_{\max}\left(\, S^*(j\omega)S(j\omega)\,\right) \leq 1 \tag{2.14}$$

where $\lambda_{\max}(.)$ is the largest eigenvalue. Indeed $||S(j\omega)||$ may be considered the power gain at the frequency ω, which of course must be below unity.

Finally, for single-input single-output LTI systems, this implies that the magnitude of $S(j\omega)$ remains below unity and its graph remain inside the unit circle in the complex plane. This is consistent with the previous condition (2.10c), as (2.13) maps the entire right half plane into the unit circle.

2.1.6 Remarks on Stability

Though intuitively passivity is often equated with stability, these concepts arise from different vantage points and show some subtle distinctions. Generally, the energy storage function E_{store} can be used as a Lyapunov-like function. It is bounded because the power dissipation P_{diss} forces a nonpositive derivative. This can then be used to deduce stability.

However, passivity in itself does not require any internal models or even the specification of states. And without knowledge of the states, the very notions of equilibrium points, convergence, and even stability are in question.

We first give a classic, though maybe pathological, example to illustrate this point. Then we comment on the assumptions made and the process needed to infer stability from passivity. These arguments are implied throughout the remaining chapters.

2.1.6.1 Classic Example

Consider a rolling ball or any other moving object on a frictionless table. The dynamics are governed by the single inertial element

$$m\ddot{x} = F \tag{2.15}$$

Clearly this system is passive. The energy storage function equals the kinetic energy

$$E_{store} = \tfrac{1}{2}m\dot{x}^2$$

and the dissipation function is zero, i.e. the system is lossless.

Most people would also consider the system stable. Without further input, the velocity will remain constant and thus bounded by its own initial value.

Problems, however, arise if we consider the position as an additional state, making the system second order. As the velocity remains constant, the position never settles to any steady state location. Instead it continues to grow without bound. Hence this second order system is *not* stable in the sense of Lyapunov.

Added Damping. This behavior does not depend on the lossless nature of the system. With quadratic friction, which is typical for aerodynamic drag, the example becomes

$$m\ddot{x} + b|\dot{x}|\dot{x} = F \tag{2.16}$$

The energy storage remains equal to the kinetic energy, but the power dissipation is now nonzero while the object is in motion.

$$P_{diss} = b|\dot{x}|\dot{x}^2$$

The system is dissipative.

From Lyapunov's direct method we know the velocity will asymptotically converge to zero as the kinetic energy is dissipated. A more detailed analysis shows that the velocity can be expressed as

$$\dot{x}(t) = \frac{\dot{x}_0}{1 + \lambda|\dot{x}_0|\, t}$$

where $\lambda = b/m$ is positive and $\dot{x}_0$ is the initial velocity. Again we see the asymptotic convergence. However, integrating this equation we find the position given as

$$x(t) = x_0 + \frac{\dot{x}_0}{\lambda|\dot{x}_0|}\,\ln(1 + \lambda|\dot{x}_0|\, t)$$

which still grows without bound. So even though the system is dissipative, the position is not bounded and the second order system is not stable.

Added Position Feedback. The problem, of course, is solved by adding position feedback to the system in the form of a spring or a spring-like element. This includes potential energy in the total storage and thus in the Lyapunov function, making it dependent on both position and velocity. Consequently, both position and velocity are bounded.

2.1.6.2 Arguments for Stability

Our work involves systems which interact both with unknown passive environments as well as human operators. To establish stability, we proceed in multiple steps. First we consider the system when isolated from the both environment and the human operator. Next we include interactions with the environment and finally allow human inputs.

Isolated Stability. When the system is isolated we have a complete system model and with it knowledge of the total stored energy E_{store} and power dissipation P_{diss} functions. We use the lower bounded energy E_{store} as a Lyapunov-like function. As there is no external contact and thus no external power flow, we can compute its derivative as

$$\frac{\mathrm{d}}{\mathrm{d}t} E_{store} = -P_{diss} \leq 0 \tag{2.17}$$

We must now assume that all states are associated with nonzero inertial or capacitive elements, i.e. with energy storage elements. This implies that they are represented in the total energy function, such that the bounded energy will also bound all states. Consequently the system is stable in the sense of Lyapunov. Naturally this excludes using states not associated with any energy storage, as shown in the above example.

To achieve asymptotic stability, we must examine the power dissipation in more detail. In particular we must check strict passivity and determine the state dependence of the dissipation. Using either the invariant set theorem or Barbalat's Lemma (Slotine and Li, 1991), we can then argue the convergence of these states.

Stability during Contact. Once we are satisfied with isolated stability, we consider the case of contacting an arbitrary environment. We assume this environment is also passive, i.e. without any active elements, so that the global system remains

passive. In particular, the total energy function remains upper and lower bounded with a nonpositive derivative. So the stability arguments continue to hold.

Similarly, the power dissipation function can only increase and always maintains the same dependence on the various states. So the same arguments for asymptotic stability can be made.

Furthermore, it has been shown that passivity is not only sufficient, but even necessary for stability when connecting a linear system to *any* passive environment (Colgate and Hogan, 1988).

Human Input. Finally we must consider the human input. Naturally the operator must produce energy to generate this input and interact with the system. However, we assume this input energy is bounded and the above arguments again remain intact.

It has also been suggested that humans appear passive at high frequencies without conscious thought, providing an explanation of why we seem to interact in a stable fashion with passive systems.

Conclusions. So the use of a passive control strategy indeed constitutes a sufficient condition for the stability of a system coupled to any passive environment with a bounded operator input energy. Appropriate power dissipation can furthermore assure asymptotic stability.

2.2 Time Delayed LTI Systems

The following sections discuss the stability and behavior of closed loop linear time-invariant (LTI) systems, which include one or possibly more delays. In particular, we examine root loci with respect to variable delay time T as well as variable open loop gain K. We also briefly cover the Nyquist criterion and the use of rational approximations in place of the delay.

The major difference between delayed and undelayed LTI systems is the number of their closed loop poles. For undelayed systems, this number remains constant and equal to the number of open loop poles. Delayed systems, however, have an infinite number of closed loop poles, stemming from the non-algebraic nature of the characteristic equation. So while the underlying ideas in the following discussions are quite similar to the traditional undelayed cases, the analysis steps and their outcome are quite different. In particular, the root loci take a very different appearance.

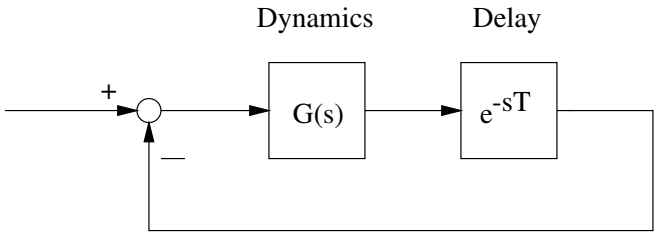


Figure 2-3: Block diagram for a single delay system. We do not distinguish between feedforward and feedback elements, so a single transfer function with unit feedback is sufficient.

Probably due to their increased complexity, delayed systems have received much less attention than their undelayed counterparts and are often mentioned only in passing. Several sources, for example (Marshall et al., 1992), (Malek-Zavarei and Jamshidi, 1987), or (Górecki et al., 1989), do attack the problems directly with varied approaches. Nonetheless a globally accepted and matured set of methods still seems to be lacking. In the following, we attempt to clarify the effects of varied delay or gain on the system behavior, and in so doing develop a basic understanding of delayed systems.

The concepts of zeros and poles, root loci, and frequency responses are, of course, limited to linear time-invariant systems. In contrast, wave variables and passivity are not effected by this limitation. However, LTI system form the foundation of many control applications and provide significant insight into behavior and changes of behavior due to varied parameters and external influences. We thus do believe there is significant benefit to this discussion.

2.2.1 Single Delay Systems

For our initial discussions in Sections 2.2.2 and 2.2.3, we will assume a single feedback loop with a single delay, as shown in Figure 2-3. Should multiple delays be present in the loop, they are combined into the single delay T. All other dynamics are lumped in the rational transfer function $G(s)$.

Multiple internal loops without delays may be solved explicitly, replaced by a simple transfer function, and included in the above assumption. However, multiple loops with independent delays are excluded as they can no longer be represented by a rational transfer function. They are addressed later in Section 2.2.4. Nevertheless single delay systems cover a large domain including basic teleoperation, where the delay is isolated in the communications between local and remote sites.

Also, we neglect the difference between feedforward and feedback elements. This distinction does not effect the characteristic equation and hence the underlying system behavior. Instead, it can be included in later stages when determining the explicit transfer functions between various inputs and outputs.

The characteristic equation of a single delay system is given by

$$G(s)\ e^{-sT} = -1 \tag{2.18}$$

where $G(s)$ is the undelayed open loop transfer function and T is the positive delay time. Note wave variables may use positive feedback instead of the standard negative feedback. This can be included into this form, by substituting the negative transfer function.

The only difference to the standard characteristic equation $G(s) = -1$ is the exponential delay operator. Indeed, it is this exponential function that creates the infinite number of solutions and closed loop poles, thereby causing the increased complexity of delayed systems.

The magnitude and phase criteria can be expressed as

$$|G(s)|\ e^{-\sigma T} = 1 \tag{2.19a}$$

$$\angle G(s) - \omega T = \pi + 2\pi k \tag{2.19b}$$

where k is an arbitrary integer and $s = \sigma + j\omega$. Note that the delay time T appears in both criteria, which are hence more tightly coupled that in the undelayed case.

Finally, we can express the transfer function $G(s)$ as the ratio of a numerator polynomial $N(s)$ and denominator polynomial $D(s)$.

$$G(s) = \frac{N(s)}{D(s)} = K_\infty\ \frac{(s - z_1)\cdots(s - z_L)}{(s - p_1)(s - p_2)\cdots(s - p_M)} \tag{2.20}$$

where z_i are the open loop zeros, p_i are the open loop poles, and K_∞ is the high frequency gain. We define the relative degree $r = M - L$ as well as the type n, which denotes the number of free integrators (poles at the origin). Should the system have zeros at the origin instead, we specify a negative n.

2.2.2 Variable Delay Root Locus

We first examine the root locus with respect to variable delay. This locates and tracks the roots of the characteristic equation (2.18), and hence the closed loop poles, while

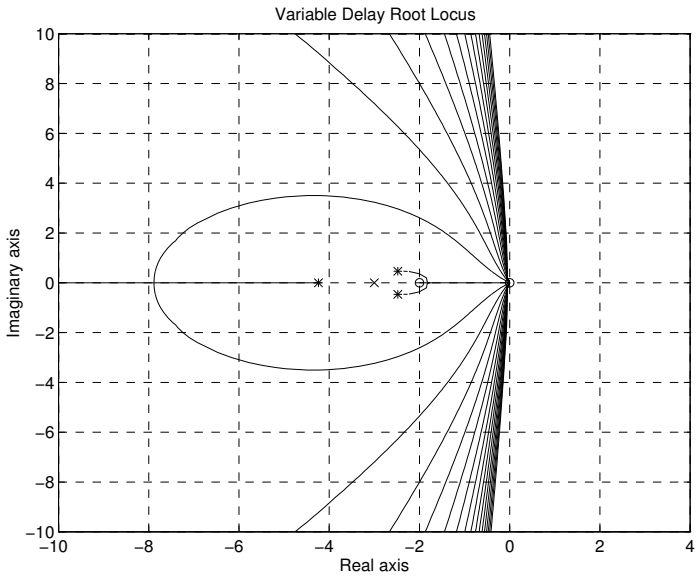


Figure 2-4: Example of a typical variable delay root locus. Notice how additional roots appear from infinite values and all roots tend towards the origin or the LHP zero.

varying the delay time T. The open loop gain K is held constant.

Figure 2-4 shows a typical example of such a root locus, using the system

$$G(s) = \frac{s(s+2)}{5(s+3)^3}$$

The open loop poles and zeros are marked with 'x' and 'o' respectively. The locus itself begins at the undelayed closed loop poles, which are marked with '$*$'. As noted before, the delay introduces an infinite number of closed loop poles at infinite values ($|s| = \infty$). As the delay increases they continually enter the region of interest and eventually tend to the origin ($s = 0$) for large delays. Even the original undelayed poles will tend inward to the origin if they don't approach a zero. Unfortunately, as more poles enter the region of interest, they do so at steeper angles with lower damping ratios. So typical systems with large delays are not only slow to react, but also show strong oscillatory tendencies.

We proceed by examining the initial behavior for small delays and the new distal poles, including their approach paths and order. This also predicts the initial stability. Next we focus on the final behavior for large delays, including the approach paths to the origin. Thereafter we discuss real axis poles, check imaginary axis crossings and the limits of stability they imply, and look at stability windows. We finally complete

the developments with some brief summarizing and concluding remarks.

2.2.2.1 Initial (Small Delay) Behavior and Distal Poles

For small delay values, the system remains dominated by the undelayed closed loop poles. The additional poles all appear either in the left half plane (LHP) or right half plane (RHP), but never mixed. In particular, this depends only on the system degree and possibly the high frequency gain. Therefore, the initial stability can be checked easily.

We proceed by examining the cases of zero delay and infinitesimal delay, then tracking the approach paths of the additional (distal) poles.

Zero Delay. For $T = 0$ the characteristic equation (2.18) collapses to the traditional case

$$G(s) = -1$$

with a constant feedback gain. Using the numerator $N(s)$ and denominator $D(s)$ polynomials we have

$$D(s) + N(s) = 0$$

Traditional tools may be used to solve this algebraic equation and determine the locations of the undelayed closed loop poles. In particular, there is only a finite number of poles corresponding to the higher degree of either polynomial.

Infinitesimal Delay. As the delay time T changes from zero to infinitesimally small, the delay operator shows up in the characteristic equation (2.18) in the form of the exponential e^{-sT}. Simultaneously, the infinite new poles appear as solutions.

But notice that for all finite values of s, the exponential has only an infinitesimal effect (remember T is infinitesimal). Hence the undelayed closed loop poles remain the only finite solutions of the characteristic equation. Instead, the new poles must appear at infinite values of s. The remaining questions are at what angle do they appear, how do they approach from infinity, and in what order.

Distal Poles/Initial Stability. To answer these questions, we study the system at infinite or very large values of s. In this region the transfer function $G(s)$ will be dominated by its highest powers and may be expressed as

$$\lim_{|s|\to\infty} G(s) = \frac{K_\infty}{s^r} \tag{2.21}$$

where K_∞ is the high frequency gain, and r is the relative degree of the system. Substituting this expression into the magnitude criterion (2.19a) leads to the requirement

$$\frac{|K_\infty|}{|s|^r}\, e^{-\sigma T} = 1$$

or

$$\frac{|K_\infty|}{|s|^r} = e^{\sigma T} \tag{2.22}$$

For strictly proper systems with a positive relative degree, the left hand side of (2.22) will approximate zero. So the requirement can only be fulfilled in the left half plane (LHP) where $\sigma < 0$, allowing the exponential to also take small values. Thus all new poles must appear in the LHP and, assuming the undelayed closed poles are also in the LHP, strictly proper systems are initially stable.

For systems with negative relative degree, Equation (2.22) takes on very large values. This forces $\sigma > 0$, so that all new poles must appear in the right half plane (RHP). Hence all such systems will be initially unstable. Interestingly, initial instability implies instability for all possible delays. This because the new poles, of which there are an infinite number, may only cross the imaginary axis at a finite rate so that some poles always remain in the RHP.

The borderline case of zero relative degree depends on the magnitude of the high frequency gain K_∞. Should this value fall below unity, it forces a negative σ and the poles appear in the LHP. Should the value exceed unity, σ is positive and the poles appear in the RHP. And finally, if the value equals unity, the poles appear exactly on the imaginary axis.

Table 2.1 summarizes these results. Basically, the behavior depends on the system gain $|G(j\omega)|$ at high frequencies $\omega \to \infty$. The higher the relative degree and the lower the high frequency gain, the smaller $|G(\infty)|$ and the further in the LHP the new poles appear. The table also gives the angles at which the poles appear and from which they approach, which we discuss next. Note the sign of the gain has no effect, so that the new poles appear in the same area for both negative and positive feedback.

Distal Approach Paths. Substituting the distal transfer function from (2.21) in both the magnitude and phase criteria (2.19), applying the logarithm to the magnitude, and rearranging we find

$$\sigma T = RT\cos(\varphi) = \ln|K_\infty| - r\ln R \tag{2.23a}$$

$$\omega T = RT\sin(\varphi) = \pi + 2\pi k + \gamma - r\varphi \tag{2.23b}$$

System Parameters		Gain $\|G(j\omega)\|$	Initial Pole Locations		Initial Stability
$r > 0$		$\|G(\infty)\| < 1$	LHP	$\varphi \to 180^o$	stable
$r = 0$	$\|K_\infty\| < 1$			$\varphi = 90^o \ldots 270^o$	
	$\|K_\infty\| = 1$	$\|G(\infty)\| = 1$	-	$\varphi = \pm 90^o$	
	$\|K_\infty\| > 1$	$\|G(\infty)\| > 1$	RHP	$\varphi = -90^o \ldots 90^o$	unstable
$r < 0$				$\varphi \to 0^o$	

Table 2.1: The initial pole locations and the initial stability for small delays depend primarily on the system degree r, secondarily on the high frequency gain K_∞. Note this assumes the undelayed closed loop poles are in the LHP.

where

$$s = \sigma + j\omega = Re^{j\varphi} \qquad \text{and} \qquad K_\infty = |K_\infty| e^{j\gamma} \qquad \gamma = 0, \pi$$

The angle γ is included to allow alternate signs of the gain K_∞ for both negative and positive feedback. Also the integer k may be considered an index for the new poles.

First we note that the magnitude portion (2.23a) verifies the previous results. For distal locations, $\ln R$ will dominate $\ln |K_\infty|$ and the sign of σ is opposite to that of r. Or when $r = 0$, the sign of σ is dependent on the relative magnitude of K_∞.

But also we find

$$\tan(\varphi) = \frac{\pi + 2\pi k + \gamma - r\varphi}{\ln |K_\infty| - r \ln R} \tag{2.24}$$

and

$$RT = \sqrt{(\ln |K_\infty| - r \ln R)^2 + (\pi + 2\pi k + \gamma - r\varphi)^2} \tag{2.25}$$

For nonzero relative degree r, $\ln R$ will eventually dominate (2.24) and the angle is governed by

$$\lim_{R \to \infty} \tan(\varphi) = \lim_{R \to \infty} \frac{c}{\ln R} = 0$$

so that all new poles initially approach from a horizontal direction. However, given the logarithmic dependence, this is only true for very large radii indeed. For most radii, the angle will vary within the appropriate half plane. Similarly for $r = 0$ the angle is actually independent of the radius and will take multiple values based on k. Figure 2-5 shows a typical set of approach paths for the first 26 additional poles of a fourth order system with unit gain.

Notice the more oscillatory poles with larger ω approach at a more vertical angle and thus with less damping. From (2.24) we see that larger values of k correspond to larger angles at any given radius. Fortunately, as we discuss next, these poles appear

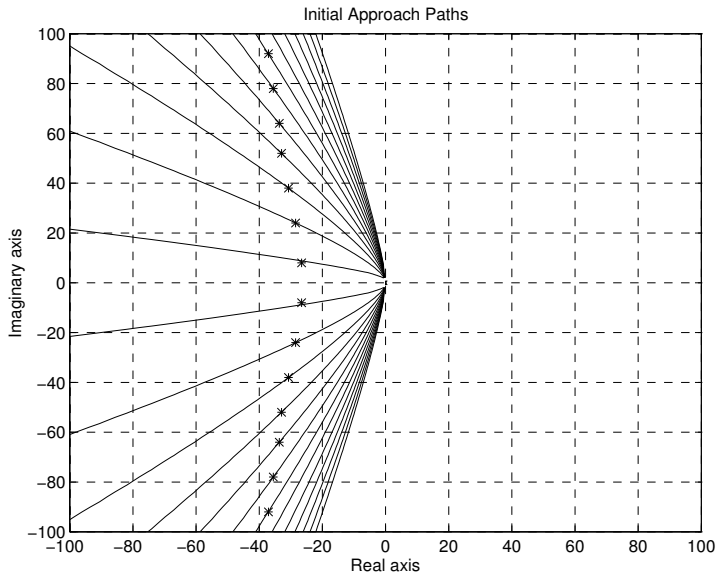


Figure 2-5: The typical approach paths of the first 26 additional poles of a fourth order system are shown. Marked are the poles for a given delay, which verifies that more oscillatory poles approach later.

later with a larger delay T. However, this highlights the finding that larger delays produce more oscillatory systems.

Distal Approach Order. The final question concerns in what order the poles approach from infinity. Or equivalently, for a given delay where along their approach path are these poles located and which ones have entered the region of interest and are thus relevant. For example, Figure 2-5 marks the poles for a delay of 0.5.

To this extent consider (2.23)–(2.25) for a constant delay. First from (2.25) we see that more distal poles at a larger radius R correspond to a higher index k

$$2\pi k = -\pi - \gamma + r\varphi \pm \sqrt{(RT)^2 - (\ln|K_\infty| - r\ln R)^2}$$

as R dominates $\ln R$. Indeed k grows roughly proportional to R. The alternate signs corresponds to poles above and below the real axis.

Then from (2.23b) we find that the more distal poles have higher imaginary components.

$$\omega = \frac{\pi + 2\pi k + \gamma - r\varphi}{T} \tag{2.26}$$

Similarly the real components are given by (2.23a) as

$$\sigma = \frac{\ln |K_\infty| - r \ln R}{T} \tag{2.27}$$

and also increases with R. The later dependence, however, is much weaker and for large radii we find

$$\sigma = -\frac{r}{T} \ln \omega \tag{2.28}$$

So, at any given delay, any distal poles lie on a curve which can be approximated by (2.28). The more distal poles have both higher real and imaginary components and can thus be safely ignored outside the region of interest. Or conversely, these poles would require a larger delay time to reach the region of interest and become relevant.

Of course, as the delay increases, this entire curve of distal poles moves closer to the origin and the imaginary axis, as described in the previous section.

Also, from (2.26) we see that at any given delay the additional poles are spaced vertically by approximately

$$\Delta\omega = \frac{2\pi}{T} \tag{2.29}$$

This spacing becomes smaller as the delay increases and so the poles appear faster.

Systems with zero relative degree form a minor exception, in that the logarithmic curve of distal poles collapses to a straight line. In these cases, the poles lie close to the vertical asymptote

$$\sigma = \frac{\ln |K_\infty|}{T} \tag{2.30}$$

But again this curve moves closer to the origin and the vertical spacing shrinks as the delay increases.

Finally, these approach paths are not limited to small delays. Indeed new poles continually approach from infinity for all values of the delay T. But, as mentioned before, these later poles approaching at larger delays T appear at steeper angles closer to the imaginary axis and are less damped.

2.2.2.2 Final (Large Delay) Behavior

For large delays, the closed loop poles will tend to the origin, any LHP open loop zeros, and any RHP open loop poles. The approach angle to the origin depends on the number of free integrators.

In the following we verify these final pole locations for infinite delays and explicitly determine the approach angle at the origin.

Final Pole Locations. Substituting the rational expression (2.20) for transfer function $G(s)$ into the characteristic equation (2.18), we have

$$\frac{N(s)}{D(s)}\, e^{-sT} = -1 \tag{2.31}$$

which we examine for infinite delay times $T \to \infty$.

First consider any possible solutions in the LHP. To this end we manipulate (2.31) into

$$D(s)e^{sT} + N(s) = 0$$

As $\sigma < 0$ the exponential e^{sT} takes an infinitesimal value, so that we are left with the requirement

$$N(s) = 0$$

Thus the open loop zeros are the only possible solutions in the LHP.

Similarly to obtain solutions in the RHP we manipulate (2.31) into

$$D(s) + N(s)e^{-sT} = 0$$

which for $\sigma > 0$ collapses into

$$D(s) = 0$$

so that the open loop poles are the only possible solutions in the RHP.

Solutions on the imaginary axis, with the exception of the origin, are not possible as the exponential term does not reach any limit. So all poles will converge to LHP open loop zeros, RHP open loop poles, or the origin. At the origin the approach angle is of particular interest, as it determines both instability and the dominant damping ratio.

Final Approach Paths. The approach paths to the origin are actually closely related to the initial approach paths previously discussed. In this case we study the system very near the origin and can express the transfer function $G(s)$ by a simple power

$$\lim_{|s|\to 0} G(s) = \frac{K_0}{s^n} \tag{2.32}$$

where K_0 is the low frequency gain and n denotes the type and thus the number of free integrators in the open loop system. Note the undelayed closed loop system could also have poles at the origin. These will simply remain at the origin for all delays and do not effect the asymptotic behavior or change the value of n. Also n may be negative, in which case it denotes the number of open loop zeros at the origin.

As before, we substitute this expression into both the magnitude and phase criteria (2.19), apply the logarithm to the magnitude criterion, and find

$$\sigma T = RT\cos(\varphi) = \ln|K_0| - n\ln R \tag{2.33a}$$

$$\omega T = RT\sin(\varphi) = \pi + 2\pi k + \gamma - n\varphi \tag{2.33b}$$

where γ now represents the phase and thus the sign of K_0.

From (2.33a) we can determine which half plane the poles approach the origin from. For nonzero n the logarithm $\ln R \to -\infty$ will dominate the right hand side as $R \to 0$. Hence the sign of σ equals the sign of n. For the cases where $n = 0$, the sign of σ depends on the magnitude of the gain K_0. Small gains imply a LHP approach, large gains a RHP approach, and a unit gain means the poles approach perfectly vertical.

The results are summarized in Table 2.2. Not surprisingly, this behavior is dependent on the system gain $|G(j\omega)|$ at low frequencies $\omega \to 0$. The lower the type and the lower the low frequency gain, the smaller $|G(0)|$ and the further from the LHP the poles approach the origin.

In the case of vertical approach paths, with $n = 0$, $|K_0| = 1$, and thus $|G(0)| = 1$, we can still determine in which half plane the paths lie. To do so, we examine the sign of $\frac{\mathrm{d}}{\mathrm{d}\omega}|G(j\omega)|$ at $\omega = 0$. Should this be negative, i.e. the gain $|G(j\omega)|$ be decreasing below unity, all poles near the origin will lie in the LHP. Conversely, a positive sign and increasing gain forces the poles into the RHP.

Unfortunately, these results do *not* predict stability for large delays. Regardless of the number of poles which reach the origin, there are still many other poles moving from infinity into the region of interest. Any of these may or may not move through the RHP and can easily cause instability. Naturally, should the final approach come from the RHP, the system will be unstable and remain so for all large delays.

The exact approach angle is further determined by

$$\tan(\varphi) = \frac{\pi + 2\pi k + \gamma - n\varphi}{\ln|K_0| - n\ln R} \tag{2.34}$$

For nonzero n, all poles must approach the origin from a horizontal direction as

System Parameters		Gain $\|G(j\omega)\|$	Approach Angle	
$n < 0$		$\|G(0)\| < 1$	LHP	$\varphi \to 180^o$
$n = 0$	$\|K_0\| < 1$			$\varphi = 90^o \ldots 270^o$
	$\|K_0\| = 1$	$\|G(0)\| = 1$	-	$\varphi = \pm 90^o$
	$\|K_0\| > 1$	$\|G(0)\| > 1$	RHP	$\varphi = -90^o \ldots 90^o$
$n > 0$				$\varphi \to 0^o$

Table 2.2: Final approach angles of the poles to the origin. This depends primarily on the number of free integrators n, secondarily on the low frequency gain K_0. Note a LHP approach angle is not sufficient to guarantee system stability as there are many other poles to be considered.

$\tan(\varphi) \to 0$ for small radii. However, this is again a logarithmic dependence, so that the angles quickly change away from the origin. And for $n = 0$ the angles cover the half plane independently of the radius. Figure 2-6 shows a typical set of approach paths with the first 26 poles.

Finally, we can determine the order of the approach. From (2.33a) we see that

$$T = \frac{\ln |K_0| - n \ln R}{\sigma}$$

Thus on a fixed radius, the poles closest to the horizontal axis with the largest σ have the smallest delay and approach first. The poles with steeper approaches come in later.

2.2.2.3 Real Axis Branches

Finding the real axis branches of a variable delay root locus is similar to the usual tests, but involves additional checks regarding the undelayed closed loop poles.

Like the standard root locus, the phase criterion can be used to determine any possible regions on the real axis. In particular, substituting $s = \sigma$ in (2.19b) removes the dependence on T and provides

$$\angle G(\sigma) = \pi + 2\pi k \tag{2.35}$$

Now the phase angle $\angle G(\sigma)$ depends only on the real open loop (OL) zeros and poles. All others appear in complex conjugate pairs and their contributions cancel. Assume, for the moment, a negative feedback system with a positive high frequency gain K_∞.

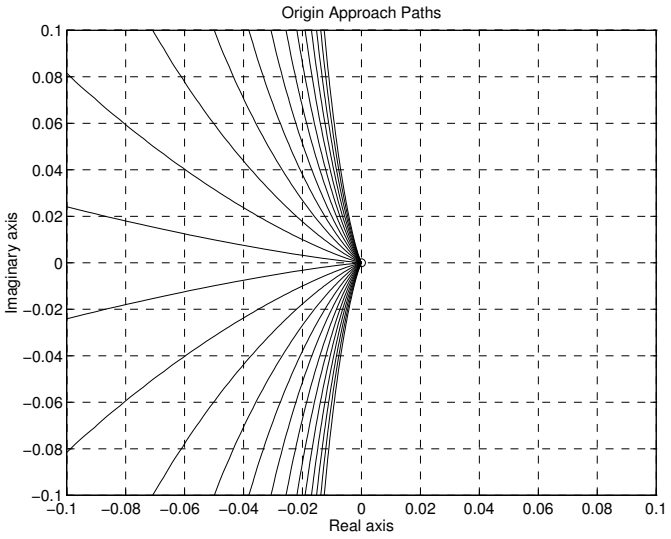


Figure 2-6: Shown is a typical set of approach paths to the origin for large delay times.

Then, as in the standard case, the real axis may only be part of the root locus if it lies to the left of an odd number of OL zeros and poles.

However, this test is only necessary and not sufficient. This because the locus begins at the undelayed closed loop (CL) poles and ends only at LHP OL zeros, RHP OL poles, or the origin. The LHP OL poles and RHP OL zeros are never touched. And so some stretches of the real axis may satisfy the above test, but must be excluded from the root locus.

This may be formalized by adding a second test based on the magnitude criterion (2.19a). Only positive delays are acceptable, so that we require

$$T = \frac{\ln |G(\sigma)|}{\sigma} \geq 0 \tag{2.36}$$

Together, the two tests are sufficient. However, the additional requirement does not lend itself to an easy graphical interpretation.

So in practice, to find the real axis branches we proceed as follows: Separately in the LHP and RHP, and within the possible sets defined by the standard test, we connect the undelayed CL poles to the available final locations. This should automatically avoid the untouched features. We may also connect to the origin or $\pm\infty$ if needed. But bear in mind, that at most one side of the real axis immediately

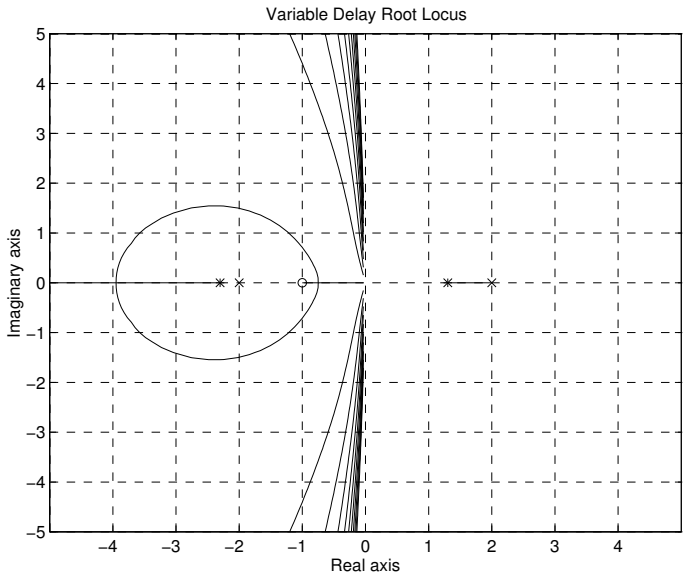


Figure 2-7: As in the standard case, the real axis branches of a variable delay root locus are located to the left of an odd number of OL zeros and poles. But some sections are excluded as the locus begins at the undelayed CL poles.

next to the origin may be part of the root locus. This because the poles only approach the origin from one half plane.

Figure 2-7 shows an example of a variable delay root locus for

$$G(s) = \frac{(s+1)}{(s+2)(s-2)}$$

Notice how the sections $(-2.3 \ldots -2)$ and $(0 \ldots 1.3)$ are excluded and the LHP OL pole at $s = -2$ is isolated.

Finally, for positive feedback systems, or equivalently systems with a negative high frequency gain K_∞, the standard test is inverted. The real axis branches must lie to the left of an even number of open loop zeros and poles. The secondary procedure, however, remains the same.

2.2.2.4 Imaginary Axis Crossings and Stability Tests

One of the most significant parts of any analysis is the determination of whether and when instability occurs. This necessitates knowing whether and when any poles cross the imaginary axis.

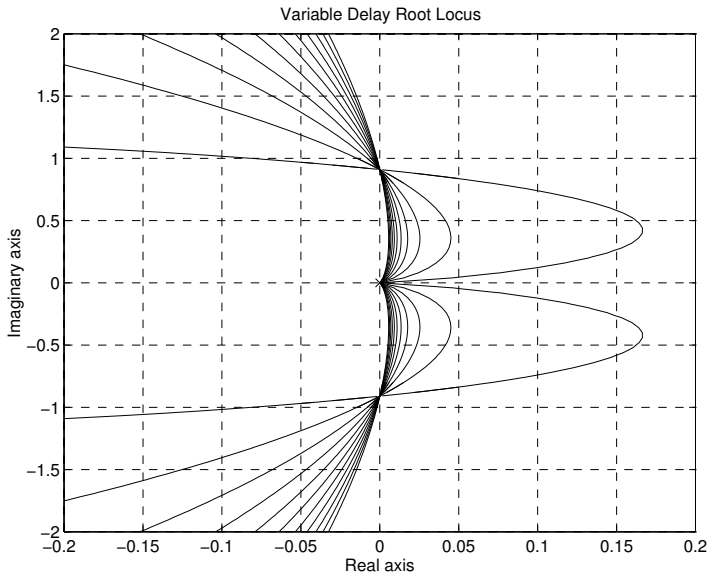


Figure 2-8: Enlarged view of the imaginary axis crossover locations. All branches pass through the same point on the axis but at different delays. They also cross each other, so that the later branches are steeper on both sides.

Figure 2-8 shows an example of imaginary axis crossings based on

$$G(s) = \frac{2}{s(s+2)}$$

Note that all branches cross at the same frequencies, but one after another at different delay times. Furthermore all branches crossing at the same location will cross in the same direction. I.e. the behavior on the imaginary axis is not dependent on the delay, but only dependent on the frequency.

To justify these answers and find the explicit locations, we seek solutions of the characteristic equation (2.18) that lie on the imaginary axis $s = j\omega$. As all solutions appear in complex conjugate pairs or on the real axis we can furthermore assume the frequency ω to be positive or zero. Thus we need values of ω and T that solve

$$G(j\omega)e^{-j\omega T} = -1 \tag{2.37}$$

Separating this equation into magnitude and phase criteria, we find

$$|G(j\omega)| \;=\; 1 \tag{2.38a}$$

$$\angle G(j\omega) \;=\; \omega T + \pi + 2\pi k \tag{2.38b}$$

where k is an arbitrary integer. Having restricted the solutions to the imaginary axis, the magnitude of the exponential is unity and the two equations are decoupled. In particular, we can use the magnitude criterion (2.38a) to determine all possible values of ω and then apply the phase criterion (2.38b) to compute the corresponding delay values.

Magnitude Criterion/Crossover Frequencies. The magnitude criterion is both necessary and sufficient to find all locations at which the root locus touches or crosses the imaginary axis. It is necessary because it is a subset of the original characteristic equation. It is sufficient because the phase criterion can amend *every* solution of ω with some delay value such that the characteristic equation is indeed solved.

To avoid the explicit absolute value, we square the criterion (2.38a)

$$|G(j\omega)|^2 = 1$$

and use the complex conjugate

$$G(j\omega)\bar{G}(j\omega) = 1$$

As the transfer function is real, we may simply use the complex conjugate argument, which in this case is also the negative argument. The magnitude criterion can thus be simplified to

$$G(j\omega)G(-j\omega) = 1 \tag{2.39}$$

Substituting the explicit transfer function (2.20) with numerator $N(s)$ and denominator $D(s)$ polynomial, we can convert the criterion into the polynomial equation

$$W(\omega^2) = D(j\omega)D(-j\omega) - N(j\omega)N(-j\omega) = 0 \tag{2.40}$$

Notice that $W(\omega^2)$ is an even function w.r.t. ω and thus represents an n^{th} order polynomial in ω^2, where n is the original system degree.

Naturally only the positive real roots of $W(\omega^2)$ are valid solutions of the original criterion, as ω is assumed real.

Phase Criterion/Crossover Delay Times. Having determined all locations of poles on the imaginary axis, we need to compute the delays T which accompany and validate these solutions. For each location we substitute the known frequency ω into

the phase criterion (2.38b) and find

$$T = \frac{\angle G(j\omega) - \pi - 2\pi k}{\omega} \tag{2.41}$$

As k may take any integer value, there are indeed an infinite number of delays for each location. I.e. an infinite number of branches of the root locus cross or touch the imaginary axis at every location.

At each location the delay times are evenly spaced and occur at a rate of $\frac{2\pi}{\omega}$. So the further the crossover location is from the origin, the faster the poles will follow one another. However, this rate is finite and therefore it is impossible for all poles to have completely crossed from one half plane to the other. For example, if the additional poles initially appeared in the RHP, there will always be some poles in the RHP.

Also note that these results are consistent with classical tests. For systems which are initially stable, determining the first imaginary axis crossing corresponds directly to the phase margin and the delay which exhausts this margin.

Crossover Direction/Behavior. Finally, we need to know whether the poles are crossing from left to right, right to left, or simply just touching the imaginary axis. This is again independent of the delay value and completely determined by the frequency ω and the transfer function $G(j\omega)$, so that all branches will behave the same.

Starting with the characteristic equation (2.18) and assuming infinitesimal movements along the root locus corresponding to infinitesimal changes in the delay, we find the direction of movement given by

$$\frac{\mathrm{d}s}{\mathrm{d}T} = \frac{s}{\frac{G'(s)}{G(s)} - T} = s\left[\frac{G'(s)}{G(s)} - T\right]^{-1} \tag{2.42}$$

For our purposes we only need the real component and only the sign thereof.

$$S = \operatorname{sgn}\operatorname{Re}\frac{\mathrm{d}s}{\mathrm{d}T} = \operatorname{sgn}\operatorname{Re}\left(\frac{\mathrm{d}s}{\mathrm{d}T}\right)^{-1} = \operatorname{sgn}\operatorname{Re}\frac{1}{s}\left[\frac{G'(s)}{G(s)} - T\right]$$

Also we may restrict ourselves to the imaginary axis, so that we can examine

$$S = \operatorname{sgn}\operatorname{Re}\frac{1}{j\omega}\left[\frac{G'(j\omega)}{G(j\omega)} - T\right] = \operatorname{sgn}\operatorname{Re}\frac{1}{j\omega}\frac{G'(j\omega)}{G(j\omega)} - \frac{T}{j\omega}$$

The delay only appears in the imaginary component and hence drops out. We expand the remaining fraction by $\bar{G}(j\omega) = G(-j\omega)$ to find

$$S = \operatorname{sgn} \operatorname{Re} \frac{1}{j\omega} \frac{G'(j\omega)G(-j\omega)}{|G(j\omega)|^2} = \operatorname{sgn} \operatorname{Re} (-j)G'(j\omega)G(-j\omega)$$

as both ω and $|G(j\omega)|$ are positive. Finally we double the real component and eliminate the imaginary component of the expression by adding its complex conjugate

$$S = \operatorname{sgn} (-j)G'(j\omega)G(-j\omega) + jG'(-j\omega)G(j\omega)$$

and express this in terms of the derivative.

$$S = \operatorname{sgn} \frac{\mathrm{d}}{\mathrm{d}\omega} - G(j\omega)G(-j\omega) = \operatorname{sgn} \frac{\mathrm{d}}{\mathrm{d}\omega} - |G(j\omega)|^2 = -\operatorname{sgn} \frac{\mathrm{d}|G(j\omega)|}{\mathrm{d}\omega}$$

Therefore the poles will move from LHP to RHP if and only if the magnitude of the transfer function $|G(j\omega)|$ is decreasing at the crossover frequency. Equivalently, an increasing magnitude implies that the poles move from the RHP into the LHP. Remember from the magnitude criterion that the magnitude exactly at the crossover frequency is unity.

Cases where the magnitude only touches unity, but does not cross over, correspond to poles that touch the axis without crossing. Similar arguments may be found in (Marshall et al., 1992).

2.2.2.5 Stability Windows

If the branches of a root locus cross the imaginary axis from the LHP to the RHP and back, each pole is inside the RHP only for a finite range of delays. Under certain circumstances, these instability ranges can lead to distinct regions of stability, also known as stability windows.

Based on an example in (Marshall et al., 1992), Figure 2-9 contains the root locus for the system

$$G(s) = \frac{\sqrt{2}}{s^2 + s + 2}$$

The locus verifies two separate imaginary axis crossover frequencies of $\omega_1 = 1$ and $\omega_2 = \sqrt{2}$. The corresponding delay values are

$$T_1 = -\tfrac{1}{4}\pi - \pi - 2\pi k = \tfrac{3}{4}\pi,\ \tfrac{11}{4}\pi,\ \ldots$$

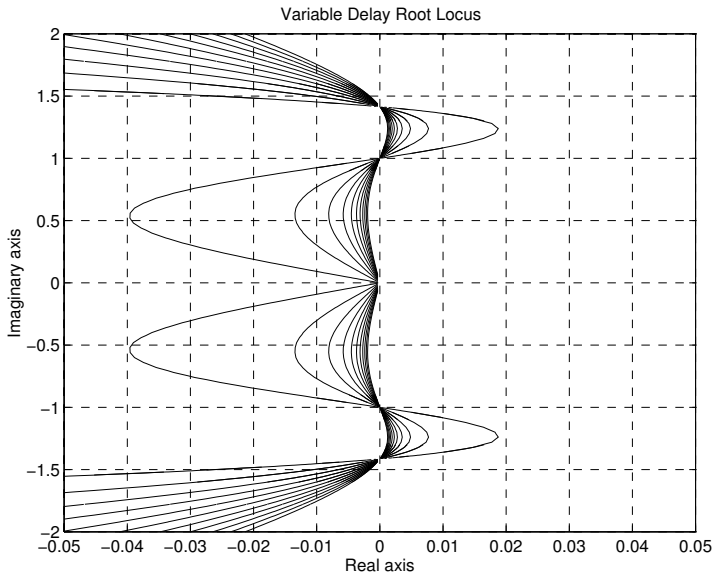


Figure 2-9: If the branches of a root locus cross over the imaginary axis and back, as shown here, the system may exhibit distinct regions of stability or stability windows.

and

$$T_2 = \frac{1}{\sqrt{2}}(-\tfrac{1}{2}\pi - \pi - 2\pi k) = \tfrac{1}{4}\sqrt{2}\pi,\ \tfrac{5}{4}\sqrt{2}\pi,\ \ldots$$

Note that ω_1 is stabilizing, i.e. the poles are moving from RHP to LHP, while ω_2 is destabilizing.

We therefore find the system to be stable for the following ranges of delay

$$\begin{aligned} 0 &< T < \tfrac{1}{4}\sqrt{2}\pi \approx 0.35\pi \\ 0.75\pi &< T < \tfrac{5}{4}\sqrt{2}\pi \approx 1.76\pi \\ 2.75\pi &< T < \tfrac{9}{4}\sqrt{2}\pi \approx 3.18\pi \end{aligned}$$

Thereafter the higher crossover rate corresponding to the larger ω_2 implies that there will always be a pole in the RHP and thus prevents any further stability windows. Of course, note that even within the stable ranges the poles remain very close to the imaginary axis and so the system will be highly underdamped.

2.2.2.6 Summary and Conclusions

The previous sections have shown how this root locus begins at the undelayed closed loop poles with additional branches continually entering the region of interest from infinity. As the delay increases, the poles move to LHP OL zeros and RHP OL poles, while the remainder approach the origin. We conclude our discussion with some final remarks.

- First, we note that the system stability generally depends on the magnitude of the frequency response $|G(j\omega)|$. This is true for initial stability, imaginary axis crossings, and final approach angles to the origin. In all cases, the system is stable when/where the magnitude is below unity and smaller magnitudes imply a larger margin of stability.

  Meanwhile the phase angle $\angle G(j\omega)$ determines the delay time at which instability occurs and consequently the largest acceptable delay. For strictly proper and minimum phase systems, this is consistent with the classic result indicating the maximum tolerable delay as

  $$T_{\max} = \frac{PM}{\omega_c}$$

  where PM is the phase margin and ω_c the crossover frequency.

  Clearly a system is always stable if the magnitude remains below unity. This is a particular incarnation of the small gain theorem and we make use of it during the development of wave variables in Chapter 4.

- When the delay increases beyond other time constants in the system, the closed loop dynamics are, not surprisingly, dominated by its effect. Indeed the bandwidth of the most significant poles are approximately inversely proportional to the delay. Other poles follow at a rate close to $\frac{2\pi}{T}$. Unfortunately, the dominant poles are typically also quite underdamped, so that highly delayed systems are both slow and oscillatory.

  To this end, added filters with time constants near the delay time appear beneficial by providing more damping. They reduce the magnitude $|G(j\omega)|$ and pull the poles further into the LHP. While they also increase the phase lag $\angle G(j\omega)$, its value is already governed by the delay making these changes less relevant.

- In general we notice that the goals of stability and performance are at odds: From high speed performance vs. initial stability at low delay through steady state errors vs. final approach angles at large delay. This is particularly evident

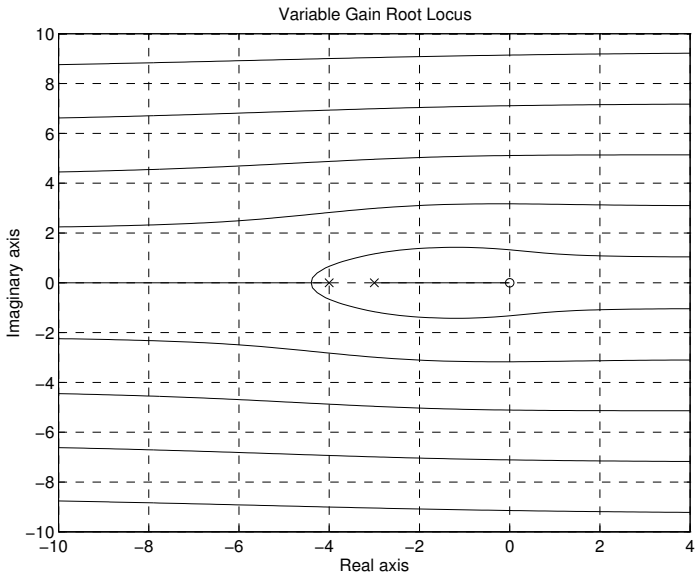


Figure 2-10: Example of a typical variable gain root locus. Notice how the additional poles appear from LHP and all poles move into the RHP toward horizontal asymptotes.

when looking at the magnitude $|G(j\omega)|$, which should be small for stability but large for performance. Though such a trade-off is natural, it does reiterate the importance of maintaining stability while seeking good performance in delayed systems.

2.2.3 Variable Gain Root Locus

The closed loop poles of a delayed LTI system also move when the open loop gain K is changed. These motions are plotted in the variable gain root locus, which holds the delay time T constant and varies the gain K from zero to infinity.

This second type of root locus more closely represents the classical undelayed version, but again we see significant differences in the individual diagrams. For example, a typical system is given by the entire open loop transfer function

$$G_{OL}(s) = K \, \frac{s}{(s+3)(s+4)} \, e^{-sT}$$

The corresponding variable gain root locus is shown in Figure 2-10 with $T = 3$.

Like its classical counterpart, this root locus begins at open loop (OL) poles and

ends at OL zeros. But once again the time delay creates an infinite number of additional closed loop (CL) poles. Here, under the variable gain, they enter from negative infinity in the left half plane (LHP). Together with other extra poles, they move towards positive infinity in the right half plane (RHP). The asymptotes are all horizontal and spaced evenly. In general, an overall horizontal behavior is demonstrated. Only very close the OL dynamics does this root locus appear similar to its classical counterpart.

To simplify the following developments, we extract an explicit positive multiplicative gain K from the rational transfer function $G(s)$. The characteristic equation (2.18) is updated to

$$K\ G(s)\ e^{-sT} = -1 \tag{2.43}$$

Again positive feedback may be included in this framework using a negative $G(s)$.

The magnitude and phase criteria are also updated as

$$K\ |G(s)|\ e^{-\sigma T} = 1 \tag{2.44a}$$

$$\angle G(s) - \omega T = \pi + 2\pi k \tag{2.44b}$$

where $s = \sigma + j\omega$ and k is an arbitrary integer. The delay T is constant, so that the two equations are effectively uncoupled. The phase criterion can be used to obtain possible solutions, while the magnitude criterion always determines the appropriate gain.

$$K = \frac{1}{|G(s)|}\ e^{\sigma T} \tag{2.45}$$

We concentrate our efforts on the initial and final CL pole locations and especially on the asymptotes. The real axis branches, imaginary axis crossovers, and departures angles are only discussed briefly as they closely follow classical intuition.

2.2.3.1 Initial and Final Pole Locations

To determine both the initial and final pole locations, when the gain K is zero or infinity respectively, we substitute the rational expression (2.20) for the transfer function $G(s)$ and rearrange the characteristic equation (2.43) to find

$$K = -\ G^{-1}(s)\ e^{sT} = -\ \frac{D(s)}{N(s)}\ e^{sT} \tag{2.46}$$

Initial (Small Gain) Pole Locations. Using $K = 0$, this is solved in two cases. First, when s is finite, we have

$$D(s) = 0$$

and the open loop poles provide the solutions. But second, when the magnitude of s is infinite, the equation may also be satisfied if $\mathrm{Re}(s) < 0$. The exponential term will dominate any polynomial expression and force the equation to zero.

Consequently, the root locus begins at all open loop poles and at negative (LHP) infinite values of s. This is true regardless of the sign of the transfer function $G(s)$, so that new poles always appear from the LHP for both negative and positive feedback systems.

Final (High Gain) Pole Locations. When $K \to \infty$, we again find two types of solutions for (2.46). Finite s imply

$$N(s) = 0$$

which is solved by all open loop zeros. Meanwhile infinite s require a positive real part to prevent the exponential from reaching zero and to guarantee infinite values.

Therefore, the root locus ends at all open loop zeros and at positive (RHP) infinite values of s.

2.2.3.2 Asymptotes

To study the asymptotes in more detail, we focus on infinite values of s, using $s = \sigma + j\omega = Re^{j\varphi}$ and letting $R \to \infty$. Under these circumstances, the transfer function $G(s)$ is dominated by its relative degree

$$\lim_{|s|\to\infty} G(s) = \frac{K_\infty}{s^r} \qquad \text{with} \qquad K_\infty = |K_\infty| e^{j\gamma} \qquad \gamma = 0, \pi \tag{2.47}$$

where, as always, r is the relative degree of the system and K_∞ is the high frequency gain. Note this gain is constant and appears in addition to the variable gain K. It may be either positive or negative to allow all possible systems.

Horizonal Orientation. Substituting this limit into the characteristic equation we find

$$K \, \frac{K_\infty}{s^r} \, e^{-sT} = -1 \tag{2.48}$$

In turn, small variations thereof lead us to the derivative

$$\frac{\mathrm{d}s}{\mathrm{d}K} = \frac{1}{K(\frac{r}{s} + T)} \tag{2.49}$$

which describes the instantaneous direction of motion for a closed loop pole. Taking the limit

$$\lim_{|s|\to\infty} \frac{\mathrm{d}s}{\mathrm{d}K} = \frac{1}{KT}$$

produces a real value and shows that, for large distances from the origin, all closed loop poles must move in a horizontal direction.

Vertical Spacing. The phase criterion (2.44b) at these outer limits provides the condition

$$\gamma - r\varphi - \omega T = \pi + 2\pi k$$

and therefore

$$\omega = \frac{\gamma - \pi - 2\pi k - r\varphi}{T} \tag{2.50}$$

As the poles eventually move outward along horizontal lines, their phase φ tends to values of 0 in the RHP or π in the LHP.

So both LHP and RHP asymptotes occur at evenly spaced intervals of

$$\Delta\omega = \frac{2\pi}{T} \tag{2.51}$$

and both may include or exclude the real axis, depending on the sign of the high frequency gain K_∞ as well as the relative degree r. The two resulting patterns are

$$\omega = 0\,,\ \pm\frac{2\pi}{T}\,,\ \pm\frac{4\pi}{T}\,,\ \ldots \qquad \text{or} \qquad \omega = \pm\frac{\pi}{T}\,,\ \pm\frac{3\pi}{T}\,,\ \pm\frac{5\pi}{T}\,,\ \ldots$$

Also notice that the asymptotes move closer together for larger delays. This reiterates results found in the variable delay root locus.

Approach Order. To compute the gains at which the poles approach the various asymptotes, we use (2.45) and substitute (2.47) to obtain

$$K = \frac{|s|^r}{|K_\infty|}\, e^{\sigma T}$$

We find that at the same horizontal position, i.e. at the same real value σ, poles at larger frequency ω require a higher gain. In other words, more distant poles move from left to right later, with larger gains. And so, at any given gain, we may safely ignore the distant poles outside the region of interest.

LHP to RHP Frequency Shift. Finally, we notice that the additional poles first appear at the LHP asymptotes and eventually move to the RHP asymptotes. However, some regular poles, which do not converge to an open loop zero, also move towards RHP asymptotes. Indeed the relative degree r gives their number.

To make room for the increased number of poles moving into the RHP, the additional poles shift outward during their left to right motion. The amount of this shift is given by

$$\mid \omega_{LHP} - \omega_{RHP} \mid = \frac{r\pi}{T} \tag{2.52}$$

2.2.3.3 Real Axis Branches

The real axis branches of the variable gain root locus are determined by the same test as in a classic root locus.

Substituting $s = \sigma$ in the phase criteria (2.44b) determines all the regions on the real axis.

$$\angle G(\sigma) = \pi + 2\pi k \tag{2.53}$$

But on the real axis the transfer function $G(\sigma)$ produces only real values. So the phase angle $\angle G(\sigma)$ may be either 0 or π. This, in turn, depends only on the real axis open loop (OL) zeros and poles, as well as on the sign of the high frequency gain K_∞. Any contributions from complex conjugate pairs cancel.

The results can be summarized as follows. First consider normal, negative feedback systems with a positive high frequency gain. Then the real axis is part of the root locus if and only if it lies to the left of an *odd* number of OL zeros and poles.

The opposite is true when the high frequency gain takes a negative value. For these systems, the real axis is part of the root locus if it lies to the left of an *even* number of OL zeros and poles.

2.2.3.4 Imaginary Axis Crossings and Departure Angles

The imaginary axis crossings determine the regions of stability and are obviously an important part of any analysis. Departure angles also give additional information

about the shape of a root locus.

Again, as with the standard root locus, these features are also determined by the phase criterion (2.44b). Given their location, the magnitude criterion can then specify the appropriate gain via (2.45).

The only major distinction to the classic case is the infinite number of imaginary axis crossovers. Eventually, at higher frequencies, they are spaced evenly at

$$\Delta\omega = \frac{2\pi}{T}$$

matching the spacing between the asymptotes. But remember the distal poles move from left to right later, i.e. with higher gains, so that we can safely ignore any additional poles not interacting with the regular OL system.

2.2.4 Multiple Delay Systems

Multiple time delays, which occur in different feedback loops within a system, cannot be combined into a single element. They thus violate the assumptions stated in Section 2.2.1 for the above developments. Instead, they lead to characteristic equations with more than one exponential function and are significantly harder to analyze analytically.

In such systems the internal loops already create an infinite number of open loop poles. The corresponding open loop transfer functions are not rational and may contain multiple exponential components. While numerical procedures are still able to handle these problems and plot root loci, the resulting diagrams may show unusual phenomena not predicted or explained by previous tests.

For example, consider a Smith predictor for a first order delayed plant with proportional feedback (compare Appendix A). This contains multiple loops with two independent delays, one being the actual delay, the other implemented by the prediction. The complete compensator can be summarized in the transfer function

$$G_c(s) = \frac{K\hat{a}}{s + (K+1)\hat{a} - K\hat{a}\, e^{-s\widehat{T}}} \; \frac{s+\hat{a}}{\hat{a}}$$

for the plant

$$G_p(s) = \frac{a}{s+a}\; e^{-sT}$$

where K is the feedback gain and $\hat{a}, \widehat{T}$ are estimates of the actual plant pole location a and its delay T. Even assuming the pole is matched accurately ($\hat{a} = a$), the open

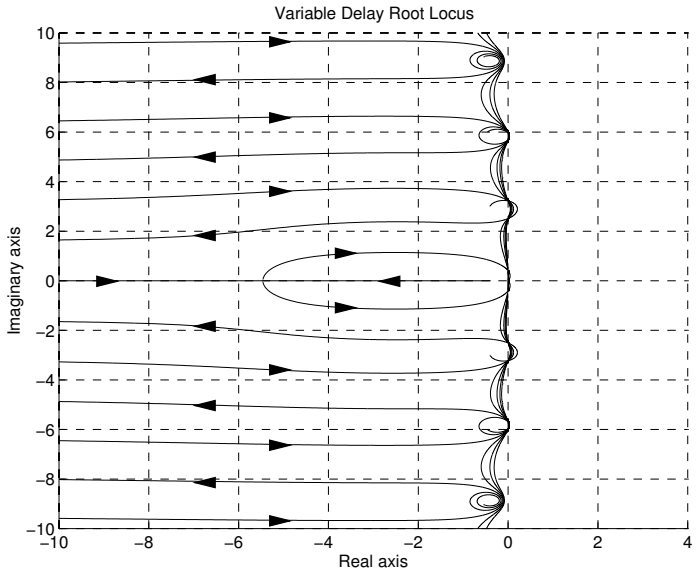


Figure 2-11: Example of a variable delay root locus for a system with multiple delays. Notice how the poles initially circle, move to negative infinity, then return and finally converge to the origin.

loop transfer function

$$G_{OL}(s) = \frac{Ka}{s + (K+1)a - Ka\, e^{-s\widehat{T}}}\; e^{-sT} \tag{2.54}$$

contains an exponential term in the denominator.

Figure 2-11 shows the variable delay root locus of this systems, using $K = 4$, $a = 1$, $\widehat{T} = 2$ and varying the actual delay T from zero to infinity. For small delays the closed loop poles remain fairly stationary and circle before moving to negative infinity. When the actual and predicted delay match ($T = \widehat{T}$), the effects of both delays cancel and the only remaining pole is given by the proportional feedback at $s = -(K+1)a = -5$. Then for larger delays the poles reappear and eventually proceed to the origin. An enlarged region is displayed in Figure 2-12 to highlight the circular motion.

As the example illustrates much of the typical time delayed behavior remains, i.e. new poles appear at infinity and eventually move to the origin. But interactions between multiple delays can cause new effects and invalidate previous results. The following updates initial stability checks and examines overall stability tests in this more general context.

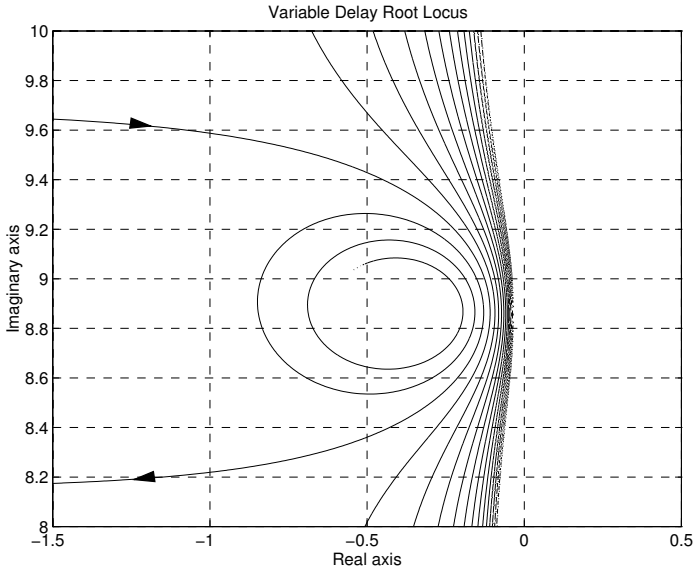


Figure 2-12: Shown is an enlarged section of the variable delay root locus presented in Figure 2-11. Clearly visible is the circular motion of a pole before it exits and reenters the region of interest. This behavior may appear when multiple delays interact.

2.2.4.1 Initial Stability

As before, the existence of any delay operator adds an infinite number of solutions to the characteristic equation and hence an infinite number of closed loop poles. All these additional poles first appear at infinite values when one or more of the delays is changed from zero to infinitesimally small. In the following we test whether or not they appear in the right half plane (RHP). Should they do so, the system is initially unstable.

Furthermore, as an infinite number of new poles appear and they continually spread out from their initial location, any system which is initially unstable will remain so for all delays. So initial stability is necessary for general stability. Note it is not sufficient as it does not consider the other closed loop poles at finite locations.

Characteristic Equation. All systems with multiple delays can be described by a characteristic equation of the form

$$\sum_{i=1}^{N} G_i(s) e^{-sT_i} = -1 \tag{2.55}$$

where N is the number of distinct positive delays T_i and all $G_i(s)$ are rational transfer functions. The relative degree of each transfer function will be denoted by r_i and its high frequency gain by K_i.

In the following, we update one or more T_j from zero to infinitesimal values. This small change does not significantly effect the poles already established at finite values of s, which we ignore for the purpose of this test. Instead we concentrate on the poles with infinite magnitude, where the product sT_j becomes relevant. To this end, we use $s = Re^{j\varphi}$ and increase $|s| = R \to \infty$.

Avoiding RHP Solutions. To guarantee that the characteristic equation (2.55) has no solutions in the RHP, we limit the magnitude of the combined transfer functions.

$$\left| \sum_{i=1}^{N} G_i(s) e^{-sT_i} \right| < 1$$

This is also satisfied if

$$\sum_{i=1}^{N} |G_i(s)| \left| e^{-sT_i} \right| < 1$$

Within the RHP the magnitude of the exponential terms is upper bounded by unity, so that we require

$$\sum_{i=1}^{N} |G_i(s)| < 1$$

We now increase the magnitude of s to infinity to find the new poles. As $|s|$ becomes large, the transfer functions $G_i(s)$ are dominated by their relative degrees and their magnitudes can be obtained on the imaginary axis.

$$\lim_{|s|\to\infty} G_i(s) = \frac{K_i}{s^{r_i}} \qquad \Rightarrow \qquad \lim_{|s|\to\infty} |G_i(s)| = \lim_{\omega\to\infty} |G_i(j\omega)|$$

And so we find that if

$$\sum_{i=1}^{N} \lim_{\omega\to\infty} |G_i(j\omega)| < 1 \tag{2.56}$$

then the new poles cannot appear in the RHP. In turn, this requires that the relative degrees be positive or at least zero with a small high frequency gain. Also, for $N = 1$ this condition collapses to that of single delay systems.

2.2.4.2 General Stability Tests

Determining the imaginary axis crossovers to predict under which conditions a system turns unstable is substantially more complex. For multiple delays, the characteristic equation (2.55) generally does not decouple into the magnitude and phase criteria as for single delays. Though several criteria have been presented in the literature, for example see (Pontryagin, 1955), they do not lend themselves for effective tests. Only in selected cases have analytic solutions been obtained.

One of these exceptions are systems with commensurate delays, which are multiples of some common factor T. They occur when multiple feedback loops pass through a single delay element and are are studied in (Marshall et al., 1992) and (Thowsen, 1981). After sufficient manipulation, the magnitude criterion is converted into a form independent of the delay and provides algebraic rather than transcendental constraints.

2.2.5 Nyquist Criterion

The Nyquist criterion is one of the few classical methods and tests that can be extended to delayed systems without alteration, even with multiple delays. This because it does not rely on a rational form of the open loop transfer function $G_{OL}(s)$. Instead, it is based on the theory of complex variables, in particular on Cauchy's theorem and the logarithmic residue theorem, which apply to all analytic functions.

Nevertheless, the addition of the exponential delay operator e^{-sT} introduces some minor differences. First, the Nyquist plots generally show spiral patterns which may make it harder to determine the number of encirclements. For example, Figure 2-13 displays an unstable Smith predictor with the open loop transfer function

$$G_{OL}(s) = \frac{50\, e^{-sT}}{s + 40 - 50\, e^{-sT}} \qquad , \qquad T = 1$$

Closer examination identifies 9 encirclements of the (-1) point in counterclockwise direction.

Second, when multiple delay operators interact, the number of open loop unstable poles may not be known. Such cases require multiple steps starting with simple internal loops.

We proceed with a very brief review of the criterion. Complete discussions can be found in many textbooks, for example (Ogata, 1990).

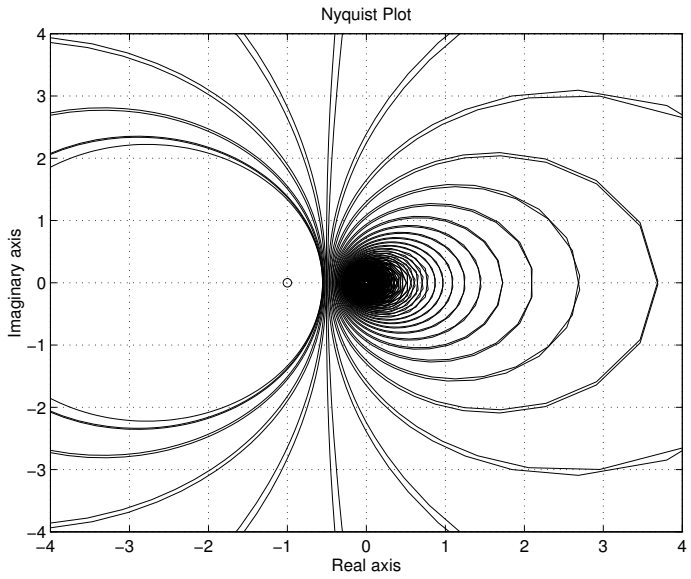


Figure 2-13: The Nyquist plots of irrational transfer functions including the exponential delay operator e^{-sT} typically show strong circular patterns, as evident in this slightly exotic example.

2.2.5.1 The Nyquist Stability Criterion

The Nyquist criterion maps the Nyquist path from the s plane into the $G(s)$ plane. Within the later, it counts the clockwise encirclements about the (-1) point to determines the number of unstable closed loop (CL) poles. It requires knowledge only of the number of unstable open loop (OL) poles and the frequency response $G(j\omega)$.

The Nyquist path defines the region of interest and is designed to enclose the entire right half plane (RHP). Beginning at the origin, it moves up the positive imaginary axis, turns right onto a semi circle of infinite radius, and returns up the negative imaginary axis to the origin. To satisfy the necessary theorems, the path may not pass directly over any OL zeros or poles. Infinitesimal semi circles are inserted to avoid such cases.

The transfer function $G_{OL}(s)$ is nonzero and analytic along this path and thus constitutes conformal map. Its value is traced on the $G(s)$ plane while following the Nyquist path in a clockwise fashion. Any sharp angles (i.e. any turns) in the path are necessarily preserved. Such a mapping process is sketched in Figure 2-14 for

$$G_{OL}(s) = \frac{e^{-s}}{s}$$

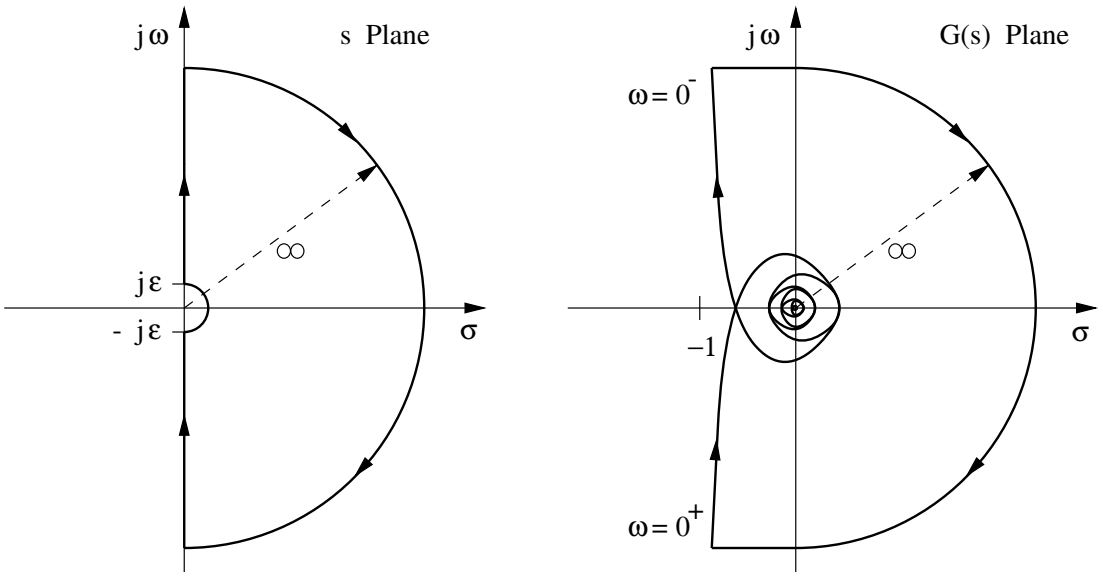


Figure 2-14: The Nyquist criterion interprets the open loop transfer function $G_{OL}(s)$ as a conformal map to project (a) the Nyquist path in the s plane onto (b) the $G(s)$ plane.

The criterion then counts the number of clockwise encirclements N about the (-1) point during this trace. This number equals the difference between the number of zeros and poles of $1 + G_{OL}(s)$ located within the original contour, i.e. within the RHP. But notice that the zeros of $1 + G_{OL}(s)$ are also solutions of the characteristic equation and therefore CL poles. Meanwhile the poles of $1 + G_{OL}(s)$ are actually the OL poles. So the Nyquist criterion may be summarized as

$$N = Z - P \tag{2.57}$$

where N = number of clockwise encirclements of the (-1) point
Z = number of zeros of $(1 + G_{OL}(s))$ = number of CL poles in RHP
P = number of poles of $(1 + G_{OL}(s))$ = number of OL poles in RHP

Both Z and P account for multiplicity. Of course, to determine the actual number of unstable CL poles $Z = N + P$, we need to know the number of unstable OL poles P.

Finally, using other closed paths in the s plane, we can update the criterion to find the CL poles in other areas. For example, we may use this to determine the number of underdamped poles.

2.2.6 Approximation of the Delay Operator

Introducing the delay operator in its exact exponential form into a LTI system leads to non-rational transfer functions and an infinite number of poles. This increased complexity significantly hinders most classical analysis tools and renders many inapplicable. While we have updated some of these tools in the previous sections, they remain somewhat limited and are often plagued by transcendental rather than algebraic equations.

However, it has also become clear that only a very few of these infinite poles are truly relevant. So it is not surprising that many people suggest approximating the delay with appropriate rational compensators which capture the dominant behavior. For example, a crude approximation could contain a single zero, a single pole, or a pole/zero pair

$$e^{-sT} \approx 1 - sT \qquad e^{-sT} \approx \frac{1}{1+sT} \qquad e^{-sT} \approx \frac{1-\frac{1}{2}sT}{1+\frac{1}{2}sT}$$

Substituting such rational compensators for the delay eliminates the transcendental functions and returns the problem to the classical framework.

We proceed to develop and examine the appropriate rational replacements for the delay. To that end, we first briefly introduce Padé approximants. They provide the mathematical foundation for using rational functions as approximations. We then focus our attention on the exponential function and establish the compensators. Lastly, we check their characteristics and limitations. These include bounds in both the frequency range and delays that can be handled adequately. Indeed we find this approach to fail if the delay time is larger than other time constants in the system.

2.2.6.1 Padé Approximants

A Padé approximant represents a function $f(z)$ by the ratio of two polynomials $P_L(z)$ and $Q_M(z)$ of finite degree

$$f(z) \approx \frac{P_L(z)}{Q_M(z)} = \frac{\sum_{k=0}^{L} a_k z^k}{\sum_{k=0}^{M} b_k z^k} \tag{2.58}$$

To eliminate the common factor in this ratio, the normalization constraint

$$Q_M(0) = b_0 = 1$$

is generally enforced. This leaves $(L + M + 1)$ free coefficients which are constructed to match an equal number of coefficients in the Taylor series expansion of $f(z)$. Hence the approximation is accurate up to order $(L + M)$.

$$f(z) - \frac{P_L(z)}{Q_M(z)} = O(z^{L+M+1})$$

Both the numerator and denominator degree L and M may be varied independently, so that a pure polynomial approximation based on the truncated Taylor series appears as the special case of $M = 0$. With the addition of denominator coefficients, the Padé approximant can typically achieve a better accuracy than this simple series, especially at large distances from the origin. A detailed treatment can, for example, be found in (Baker, Jr. and Graves-Morris, 1981) or (Baker, Jr., 1975).

2.2.6.2 Padé Approximants to the Exponential Function

The exponential function e^z has received particular attention and is one of the few cases where explicit formulas for the Padé coefficients exist. From (Baker, Jr. and Graves-Morris, 1981) we find

$$a_k = \frac{(L + M - k)!}{(L + M)!} \frac{L!}{(L - k)!} \frac{1}{k!} \tag{2.59}$$

and

$$b_k = \frac{(L + M - k)!}{(L + M)!} \frac{M!}{(M - k)!} \frac{(-1)^k}{k!} \tag{2.60}$$

In the following we will use a negative argument typical of time delay operators. Also, we shall limit the polynomials to be of equal degree N (i.e. $L = M = N$), so that

$$e^{-z} \approx E_N(-z) = \frac{P_N(-z)}{Q_N(-z)} = \frac{\sum_{k=0}^{N} \frac{(2N - k)!}{(2N)!} \frac{N!}{(N - k)!} \frac{1}{k!} (-z)^k}{\sum_{k=0}^{N} \frac{(2N - k)!}{(2N)!} \frac{N!}{(N - k)!} \frac{1}{k!} z^k} \tag{2.61}$$

This has several advantages, including

- The numerator and denominator coefficients are identical except for the different sign pattern. Consequently the approximant satisfies and shares the property

$$E_N(-z) \equiv \frac{1}{E_N(z)}$$

with the exponential function.

- The approximation is all-pass, meaning the magnitude on the imaginary axis is always unity.

$$|E_N(\pm j\omega)| = 1$$

This, of course, matches the exponential exactly and is particularly important when determining the points of instability, which occur as the poles cross the imaginary axis.

- All poles introduced by the approximation appear in the stable LHP. An equal number of zeros are placed in the RHP. Together they form a symmetry about the imaginary axis, as well as the real axis. (Compare below and to Figure 2-15.)

For example, $N = 1, 2, 3$ lead to the following second, fourth, and sixth-order approximations of e^{-sT} respectively.

$$\begin{aligned} e^{-sT} \approx E_1(-sT) &= \frac{1 - \frac{1}{2}(sT)}{1 + \frac{1}{2}(sT)} \\ E_2(-sT) &= \frac{1 - \frac{1}{2}(sT) + \frac{1}{12}(sT)^2}{1 + \frac{1}{2}(sT) + \frac{1}{12}(sT)^2} \\ E_3(-sT) &= \frac{1 - \frac{1}{2}(sT) + \frac{1}{10}(sT)^2 - \frac{1}{120}(sT)^3}{1 + \frac{1}{2}(sT) + \frac{1}{10}(sT)^2 + \frac{1}{120}(sT)^3} \end{aligned}$$

Note the Padé approximant is not the ratio of two Taylor series. Indeed such a ratio would provide only N^{th} order accuracy rather than the $2N^{\text{th}}$ order achieved above.

2.2.6.3 Pole/Zero Distribution and Selection of the Approximation Order

Using the Padé approximants as rational compensators in place of the delay operator introduces additional poles and zeros not actually present in the real system. For any analysis to be accurate, these new dynamics should not interfere with the rest of the

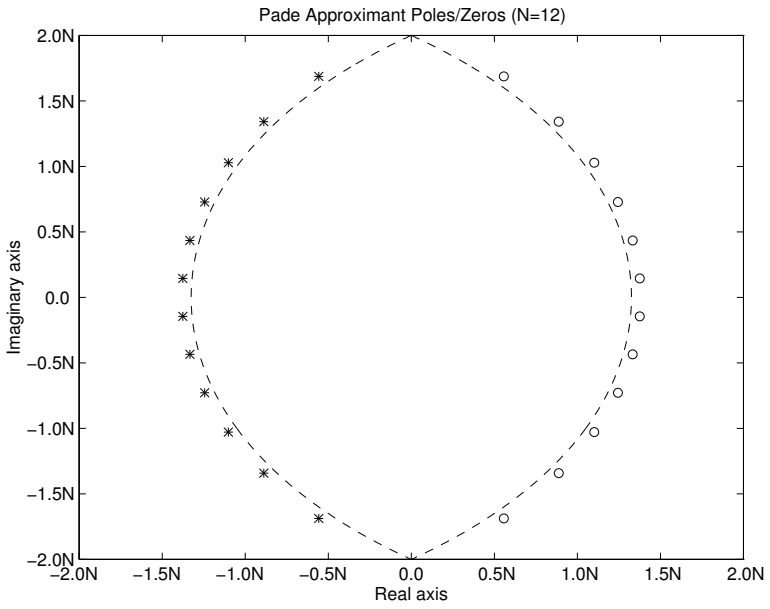


Figure 2-15: Shown are the poles and zeros of $E_N(-z)$ for $N = 12$ approximating e^{-z} to the $2N^{\text{th}}$ order, as well as the limit curves for $N \to \infty$

system and must lie substantially outside the region of interest. This is equivalent to requiring a reasonable accuracy of the approximation within that region.

The distribution of the poles and zeros is studied in detail in (Saff and Varga, 1977). Several results are noteworthy here. First, as mentioned above, *all* poles appear in the LHP, while the zeros are located symmetrically thereto in the RHP. Second, both poles and zeros are located within the annulus

$$2\mu N < |z| < 2N + \tfrac{4}{3}$$

where $\mu = 0.278$ is the unique positive root of $\mu e^{1+\mu} = 1$. In particular, the minimum radius is proportional to the order of the approximation N. Note the constant μ is quite conservative, as it is valid even for unequal numerator and denominator degrees.

Finally, the poles of the normalized approximant $E_N(-2Nz)$ approach the curve defined by

$$\left| \frac{ze^{\sqrt{1+z^2}}}{1+\sqrt{1+z^2}} \right| = 1 \qquad \text{for} \quad |z| \leq 1,\ \text{Re}(z) \leq 0$$

for increasing N. The zeros approach the symmetric curve in the RHP. Figure 2-15 shows the location of the poles and zeros for $N = 12$ as well as these limit curves.

Based on these results, we can determine an acceptable region of approximation. Conversely, given the system bandwidth or some region of interest, we can select the necessary approximation order N. As a rule of thumb, we require

$$N > 3\,\omega_{\max}\,T \tag{2.62}$$

where T is the delay and $\omega_{\max}$ is the largest frequency occurring in the system or the size of the region.

Clearly, the order must increase with larger delays or faster systems. But this implicitly limits the cases in which an approximation of the delay may be useful. As the order becomes too large, numerical problems arise and the results become very sensitive to small perturbations and errors. For example, when the delay is comparable to other time constants, we find

$$T \approx \frac{2\pi}{\omega_{\max}} \qquad \Rightarrow \qquad N \approx 6\pi \approx 20$$

At this order the ratio of smallest to largest polynomial coefficient is already $3{\cdot}10^{-30}$. For $N = 30$ it quickly reaches $3{\cdot}10^{-50}$, and continues to increases roughly exponentially.

So while the use of approximations can be an effective analysis tool, it begins to fail when the delay time grows larger than other time constants in the system. Under such circumstances we must revert to other methods, such as those presented in the earlier sections.

Chapter 3

Basic Force Reflecting Teleoperation

By definition teleoperation encompasses many different applications with many different systems (Sheridan, 1989, 1992). These may include various feedback channels, including audio, visual, tactile, and others, and also could involve different feedforward paths, such as voice commands, mouse input, etc. Our attention, however, is focused on input and force feedback via an actuated joystick. Thus we will consider a teleoperator as a chain of elements leading from the human operator to the remote environment, as shown in Figure 3-1.

At the local site the human operator interacts with the computer controlled master manipulator. Similarly, at the remote site the computer controlled slave robot interacts with the environment. The two computers are connected in both directions via some communications device, so that a closed loop can be generated from the human operator to the remote environment and back. Finally, any and all time delays are isolated within the communications.

Note the local site including the human operator and the master manipulator is always depicted at the left end of the chain. The slave manipulator and the environment form the opposite end of chain at the right side. Also, the main direction is chosen from left to right, so that by convention positive power flows from the human operator through the system into the environment.

In the following, we more carefully describe the basic system and examine the classic methods for dealing with time delays. This is accomplished both with traditional tools and using passivity. We also demonstrate when and how the instability occurs. In general, we find that while teleoperators may be stable at large delays, these methods do not provide sufficient guarantees or do not include any procedures

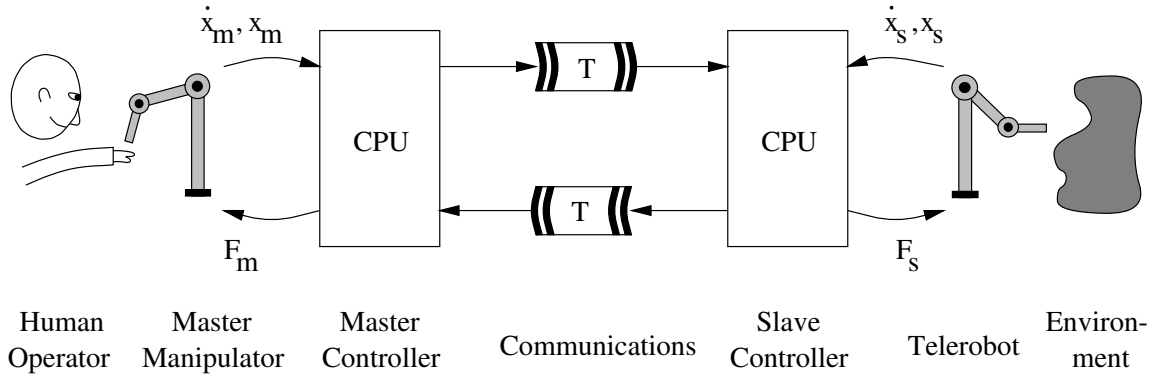


Figure 3-1: We limit our attention to teleoperator systems, in which all interaction occurs via an actuated joystick (the master manipulator) at the local site and a slave robot at the remote location.

to understand or change the system characteristics.

3.1 Basic Undelayed Teleoperator

To determine the basic system layout, we remember that passivity is not only sufficient but even necessary for stability, if we allow all possible passive environments (see Section 2.1.6). So without any knowledge of the environment, and without wanting to place any restrictions thereon, it seems only natural to require passivity of our basic system. Indeed this is the driving argument behind much of our work.

3.1.1 Passive P.D. Control

Both master and slave manipulators are inertial elements and thus already satisfy the passivity constraint. Like any other manipulator, their dynamics can be expressed as

$$\begin{aligned} H_m(\mathbf{q}_m)\ddot{\mathbf{q}}_m + C_m(\mathbf{q}_m, \dot{\mathbf{q}}_m)\dot{\mathbf{q}}_m + G_m(\mathbf{q}_m) &= -\boldsymbol{\tau}_m \\ H_s(\mathbf{q}_s)\ddot{\mathbf{q}}_s + C_s(\mathbf{q}_s, \dot{\mathbf{q}}_s)\dot{\mathbf{q}}_s + G_s(\mathbf{q}_s) &= \boldsymbol{\tau}_s \end{aligned}$$

where $\mathbf{q}$ is the vector of joint positions, and H, $C\dot{\mathbf{q}}$, G, $\boldsymbol{\tau}$ are inertia matrix, coriolis torques, gravitational torques, and motor torques respectively (Asada and Slotine, 1986). We negate $-\boldsymbol{\tau}_m$ to stay consistent with the notation and sign convention used in later results and based on Section 2.1.3.

For the controller, we cancel the gravitational torques (if necessary) and provide

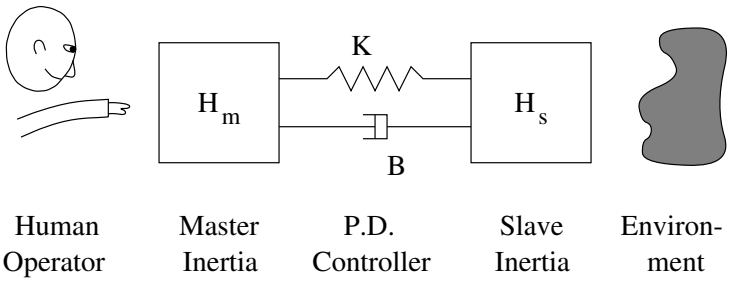


Figure 3-2: The basic undelayed teleoperator mimics the behavior of two inertia connected via a spring and damper arrangement.

simple P.D. feedback based on the tracking error between master and slave joints.

$$\begin{aligned} \boldsymbol{\tau}_m &= -G_m(\mathbf{q}_m) + \boldsymbol{\tau}_{\mathrm{PD}} \\ \boldsymbol{\tau}_s &= \quad G_s(\mathbf{q}_s) \; + \boldsymbol{\tau}_{\mathrm{PD}} \\ \boldsymbol{\tau}_{\mathrm{PD}} &= B(\dot{\mathbf{q}}_m - \dot{\mathbf{q}}_s) + K(\mathbf{q}_m - \mathbf{q}_s) \end{aligned}$$

The feedback gains B and K are constant symmetric positive definite matrices. This makes the computers and actuators appear as a spring and damper arrangement, so that the entire system is passive. In particular, the global stored energy and dissipation function are given by

$$\begin{aligned} E_{store} &= \tfrac{1}{2}\dot{\mathbf{q}}_m^T H_m \dot{\mathbf{q}}_m + \tfrac{1}{2}\dot{\mathbf{q}}_s^T H_s \dot{\mathbf{q}}_s + \tfrac{1}{2}(\mathbf{q}_m - \mathbf{q}_s)^T K(\mathbf{q}_m - \mathbf{q}_s) \\ P_{diss} &= (\dot{\mathbf{q}}_m - \dot{\mathbf{q}}_s)^T B(\dot{\mathbf{q}}_m - \dot{\mathbf{q}}_s) \end{aligned}$$

Figure 3-2 sketches the mechanical equivalent of the controller for a single degree of freedom system.

3.1.2 Kinematic Transformations

The above described controller implicitly assumes that both master and slave robots have the same kinematics. Only then does it make sense to have master and slave joints track each other. If the kinematics differ, tracking should occur between the endpoint of both manipulators.

In the passivity framework we insert a kinematic transformation on both sides of

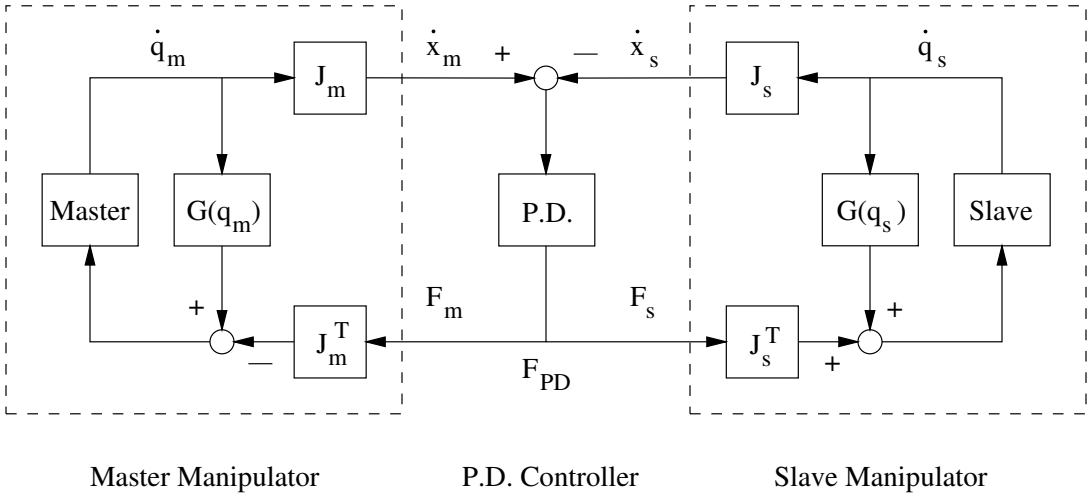


Figure 3-3: Both master and slave robots are gravity compensated, transformed into a cartesian domain, then connected via a P.D. controller mimicking a spring and damper.

the P.D. controller. These convert the joint motion and forces into endpoint values.

$$\begin{aligned} \dot{\mathbf{x}}_m &= J_m(\mathbf{q}_m)\dot{\mathbf{q}}_m & \boldsymbol{\tau}_m &= J_m^T(\mathbf{q}_m)\mathbf{F}_m \\ \dot{\mathbf{x}}_s &= J_s(\mathbf{q}_s)\dot{\mathbf{q}}_s & \boldsymbol{\tau}_s &= J_s^T(\mathbf{q}_s)\mathbf{F}_s \end{aligned}$$

They are also passive and even lossless, as the power in both joint and cartesian domains is equal.

$$\dot{\mathbf{q}}^T\boldsymbol{\tau} = \dot{\mathbf{q}}^T J^T\mathbf{F} = \dot{\mathbf{x}}^T\mathbf{F}$$

Then we replace the joint P.D. controller with a cartesian version

$$\begin{aligned} \mathbf{F}_m &= \mathbf{F}_{\mathrm{PD}} \\ \mathbf{F}_s &= \mathbf{F}_{\mathrm{PD}} \\ \mathbf{F}_{\mathrm{PD}} &= B(\dot{\mathbf{x}}_m - \dot{\mathbf{x}}_s) + K(\mathbf{x}_m - \mathbf{x}_s) \end{aligned}$$

and again we have a passive system, but one which relates the endpoint motions of the master and slave.

Figure 3-3 sketches the layout of this system. The gravity compensation is still best expressed in the joint domain and remains outside the transformations. For the remainder, we usually consider the system from the cartesian point of view and will assume gravity compensation and kinematic transformations are included as neces-

sary.

3.1.3 Force Measurements

This basic P.D. controller assumes only position and velocity measurements and does not require the use of force sensors. Yet even if they are available, we actually recommend against their inclusion in the overall teleoperator design.

The reasons are basically twofold:

(a) Such an approach is not passive. For example, measuring a force signal at the slave robot, then transmitting this information and applying it directly to the master joystick circumvents the normal path through other passive elements. In essence, the signal is no longer colocated with the corresponding velocity at the slave. And so it may inject unknown amounts of energy into the master in addition to the energy created at the slave.

Naturally the problem is less severe when the master and slave are closely connected with few and stiff dynamics in between. In this case master and slave velocities are nearly identical. However, especially as we include delays, this is no longer true and the system may show unexpected and unstable behavior.

(b) The measured forces cannot be recreated accurately. Indeed the difference between any measured force and the applied force, which we use in the feedback, is governed by acceleration terms. And these are often dominated by high frequency impact signals.

While such signals contain important information, most motors and manipulators cannot recreate them within the needed frequency range. Instead other displays, such as tactile or audio, may be better suited. Meanwhile within the bandwidth of most motors, the applied force should provide a reasonable representation of the contact.

Part of the above argument assumes that the robots can be modeled as pure inertia. Only then do the applied and contact forces differ by acceleration terms. Unfortunately, many robots include large gear reductions, significant friction, backlash, and many other undesirable components. If these substantially distort the force propagation from actuators to contact points, force sensors could and should be used in local force-feedback loops in order to ensure backdrivability.

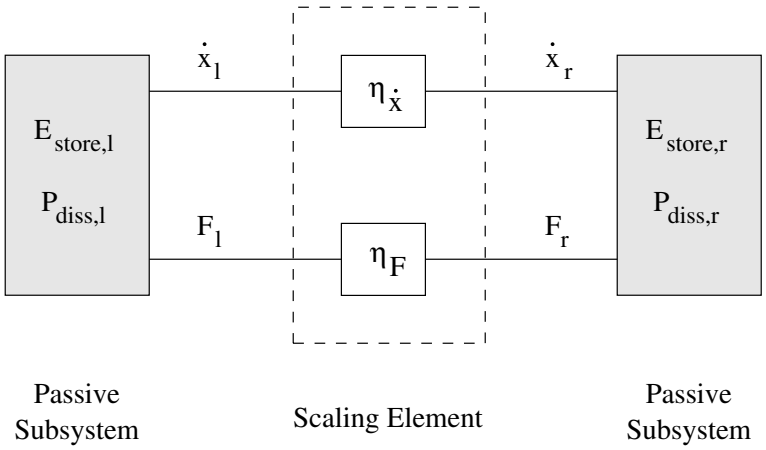


Figure 3-4: A two-port scaling element between passive subsystems may apply a different scale to both velocity and force signals, even if this scales the power flow.

3.1.4 Scaling

Many applications for teleoperation require some form of scaling. This may occur in the spatial dimensions to enlarge or reduce the workspace, or in the force signals to make the operator stronger or more sensitive, or both. In most cases this also implies scaling the power transmitted from the human to the environment.

By definition power scaling is not passive in the normal sense. Nonetheless, the chain like layout of the teleoperator system allows us to include arbitrary scaling within the passivity framework. Consider a two-port element scaling both motion and force, as shown in Figure 3-4. It is described by

$$\dot{\mathbf{x}}_r = \eta_{\dot{\mathbf{x}}}\, \dot{\mathbf{x}}_l \qquad \mathbf{F}_r = \eta_{\mathbf{F}}\, \mathbf{F}_l \tag{3.1}$$

With P_l denoting the power entering the left port,

$$P_l = \dot{\mathbf{x}}_l^T \mathbf{F}_l$$

we find the power P_r exiting the right port

$$P_r = \dot{\mathbf{x}}_r^T \mathbf{F}_r = \eta_{\dot{\mathbf{x}}}\, \eta_{\mathbf{F}}\, \dot{\mathbf{x}}_l^T \mathbf{F}_l = \eta\, \dot{\mathbf{x}}_l^T \mathbf{F}_l = \eta\, P_l$$

to be a scaled version thereof, with $\eta = \eta_{\dot{\mathbf{x}}}\, \eta_{\mathbf{F}}$. That is, power flowing left to right is increased by η, but also power returning from right to left is decreased accordingly. So in general, all power and energy functions on the left side must be scaled when

they are compared to their right side counterparts.

To verify global passivity of the system in Figure 3-4, let us examine the global stored energy function E_{store} and dissipation function P_{diss}. Following the above argument, they should be computed as

$$E_{store} = E_{store,r} + \eta E_{store,l} \qquad P_{diss} = P_{diss,r} + \eta P_{diss,l} \tag{3.2}$$

Both left and right element are individually passive, hence obey

$$\frac{\mathrm{d}}{\mathrm{d}t} E_{store,l} + P_{diss,l} = -P_l \qquad \frac{\mathrm{d}}{\mathrm{d}t} E_{store,r} + P_{diss,r} = P_r$$

where power exits the left element and enters the right element. Substituting this, we find

$$\frac{\mathrm{d}}{\mathrm{d}t} E_{store} + P_{diss} = \frac{\mathrm{d}}{\mathrm{d}t}(E_{store,r} + \eta E_{store,l}) + (P_{diss,r} + \eta P_{diss,l}) = P_r - \eta P_l = 0$$

as required, given no external inputs.

And so arbitrary, power generating scaling can be included between two passive systems, because any power increase in one direction is accompanied by the appropriate power decrease in the other. But beware — this is only acceptable because the two sides interact via a single, well controlled path. No other connections are made. Fortunately, the nature of teleoperation ensures that the sides are indeed well separated.

Also beware that the scaling must be the same for all axes. Should this not be the case, power may shift between axes and circumvent the matched up/down scaling when flowing back and forth between subsystems. This may quickly lead to unbounded power increases and hence instability.

Finally these arguments do not describe how to deal with the distortions associated with large scaling. Objects may appear very different as compared to normal interactions. For example, scaling to small workspace dimensions will accentuate surface tension forces and viscous effects. Only a detailed understanding and modeling of the particular environment can predict and address these changes. Nevertheless, passivity and hence stability is always guaranteed.

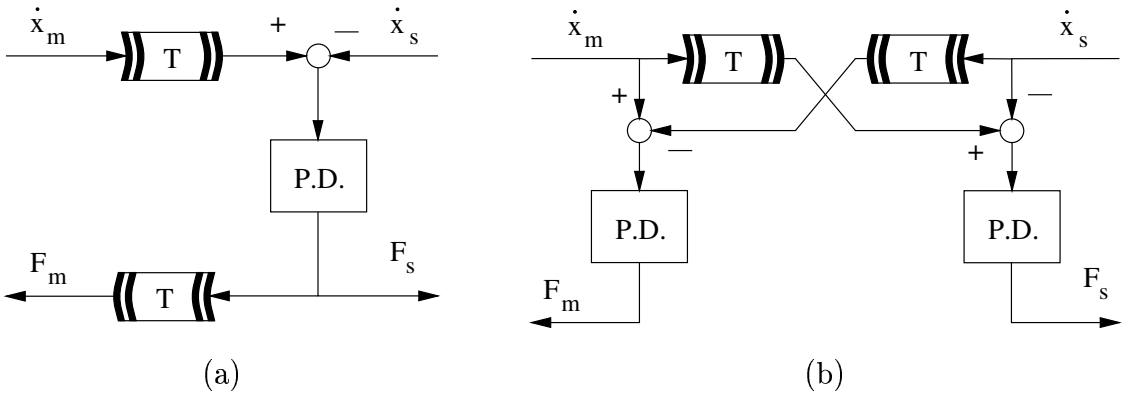


Figure 3-5: Including a communications delay may lead to either a (a) force/position control layout, or (b) a symmetric position/position control structure.

3.2 Including the Communications Delay

We now turn our attention to cases with time delay due to the communications between the local and remote locations. This primarily effects the P.D. controller in the center of the system.

Clearly we can no longer compute and apply the P.D. forces to both master and slave robots simultaneously. This leaves us with two alternatives:

(a) Compute the P.D. force at one location (usually the slave), then transmit and directly apply the force at the other site.

(b) Compute the P.D. force at both locations and transmit only position and velocity information between the sites.

Figure 3-5 sketches these choices. Essentially, case (a) places the slave under position control and the master under force control. In contrast, case (b) places both systems under position control, where the set points are determined by the opposite site. The first method appears to be more popular, perhaps because it more closely resembles the notion of force feedback. In the second method, such a feedback is only indirect.

While both methods revert to the same basic P.D. controller for zero delay, their appearance changes as the delay grows. Consider the underlying equations. For case (a) we have

$$\mathbf{F}_s(t) \;=\; B\Big(\dot{\mathbf{x}}_m(t-T) - \dot{\mathbf{x}}_s(t)\Big) + K\Big(\mathbf{x}_m(t-T) - \mathbf{x}_s(t)\Big)$$

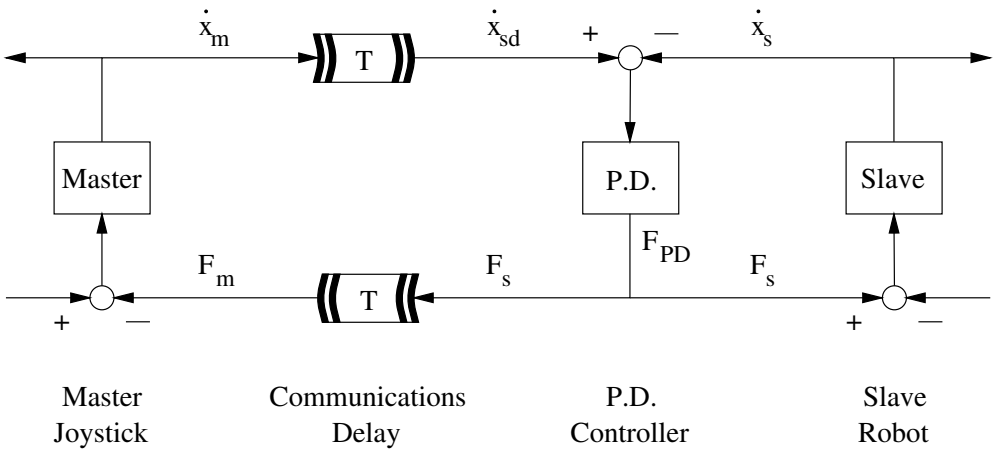


Figure 3-6: The basic delayed teleoperator computes the P.D. forces at the slave site. It transmits master position and velocity, as well as slave force.

$$\begin{aligned}\mathbf{F}_m(t) &= \mathbf{F}_s(t-T) \\ &= B\Big(\dot{\mathbf{x}}_m(t-2T) - \dot{\mathbf{x}}_s(t-T)\Big) + K\Big(\mathbf{x}_m(t-2T) - \mathbf{x}_s(t-T)\Big)\end{aligned}$$

where T is the delay. For case (b) the slave feedback is the same, but the master force is computed locally and contains less delay. The gains B and K must match those of the slave.

$$\begin{aligned}\mathbf{F}_s(t) &= B\Big(\dot{\mathbf{x}}_m(t-T) - \dot{\mathbf{x}}_s(t)\Big) + K\Big(\mathbf{x}_m(t-T) - \mathbf{x}_s(t)\Big) \\ \mathbf{F}_m(t) &= B\Big(\dot{\mathbf{x}}_m(t) - \dot{\mathbf{x}}_s(t-T)\Big) + K\Big(\mathbf{x}_m(t) - \mathbf{x}_s(t-T)\Big)\end{aligned}$$

The main difference is the immediate feedback from the master motion to the master force in the second case. It will fight any input motions immediately and more strongly.

This can be interpreted in the following manner. Strategy (a) assumes the slave motion will match that of the master perfectly and hence provides no correction force to the master until it realizes that some errors exist. Strategy (b), however, assumes that the slave will not move at all. It completely resists the master input, until it sees the slave comply. One could therefore imagine these two options as the extreme cases of possible predictions.

In the following we limit ourselves to the method (a) using a force/position control layout. The entire system is depicted in Figure 3-6. We do so not only because it is more popular, but also because the generally lower levels of resistance to master inputs

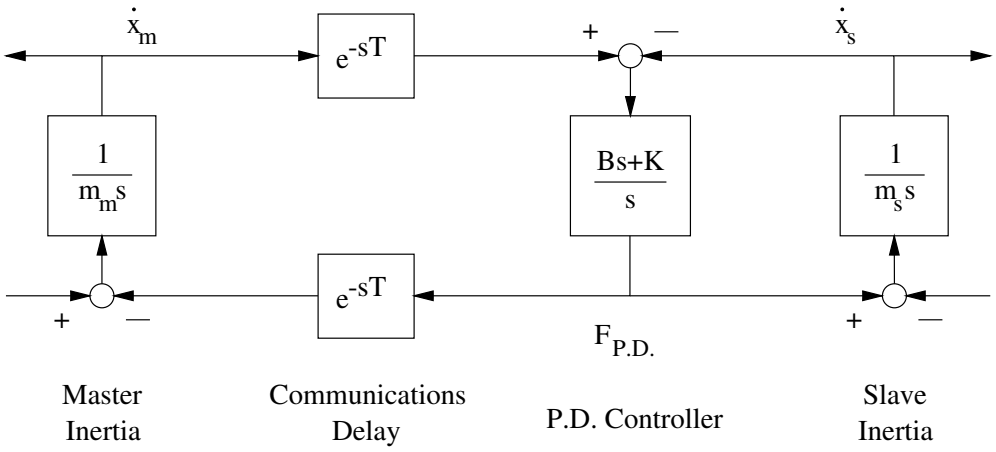


Figure 3-7: Block diagram of a basic single degree of freedom teleoperator. Both master and slave are modeled by simple inertia, while the P.D. force is computed at the slave site and fed back to the master.

make it more comfortable to operate. Nevertheless, a closer examination actually shows that the stability limits are quite similar for both cases.

3.3 Traditional Stability Limits

We proceed to examine the basic delayed teleoperator, using position control at the slave site and placing the master under force control. Using LTI tools, in particular variable delay root locus, we attempt to determine the limits of stability as the delay increases and find requirements on the system parameters. Unfortunately, the environment may vary significantly so that we need to examine numerous models and cannot determine truly global requirements.

In order to use these tools, we must treat each degree of freedom independently. In essence, we linearize the system by assuming a constant diagonal inertia matrix for both master and slave. And so both the master and slave robots collapse to a simple mass. We can also ignore scaling, as the same effect is achieved by changing either inertia value. The appropriate block diagram with the individual transfer functions is shown in Figure 3-7.

The system dynamics may also be written explicitly as

$$\begin{aligned} m_s \ddot{x}_s(t) &= F_{\mathrm{PD}}(t) \\ m_m \ddot{x}_m(t) &= -F_{\mathrm{PD}}(t-T) \end{aligned}$$

where the feedback force is determined by the P.D. controller.

$$F_{\rm PD}(t) = B\Big(\dot{x}_m(t-T) - \dot{x}_s(t)\Big) + K\Big(x_m(t-T) - x_s(t)\Big)$$

The total round trip delay time is $2T$.

3.3.1 No contact

We first examine the system stability assuming no interaction forces. The transfer functions for the master and slave sites can be expressed as

$$\begin{aligned} G_s(s) &= \frac{m_s s(Bs+K)}{m_s s^2 + Bs + K} \\ G_m(s) &= -\frac{1}{m_m s} \end{aligned}$$

For simplicity we assume identical inertia, so the complete OL transfer function is given by

$$G_{OL}(s) = -\frac{Bs+K}{ms^2+Bs+K}\, e^{-2sT} \tag{3.3}$$

Note this is a positive feedback system, as no other negation takes place in the loop. Equivalently, we could use a positive transfer function with negative feedback.

First, as the relative degree is positive, small delays will have negligible effect. However, for increasing delays, the closed loop poles will move into the RHP and the system will become unstable. Figure 3-8 shows the corresponding variable delay root locus.

In particular, let us examine the point of instability as the poles cross the imaginary axis. From (2.38a) we can find the frequency ω independently of the delay as

$$|G(j\omega)| = 1$$

If we square the equation we have

$$\begin{aligned} B^2\omega^2 + K^2 &= (K - m\omega^2)^2 + B^2\omega^2 \\ \Rightarrow \qquad 0 &= m^2\omega^4 - 2mK\omega^2 \end{aligned}$$

which is solved for

$$\omega = \sqrt{\frac{2K}{m}} \tag{3.4}$$

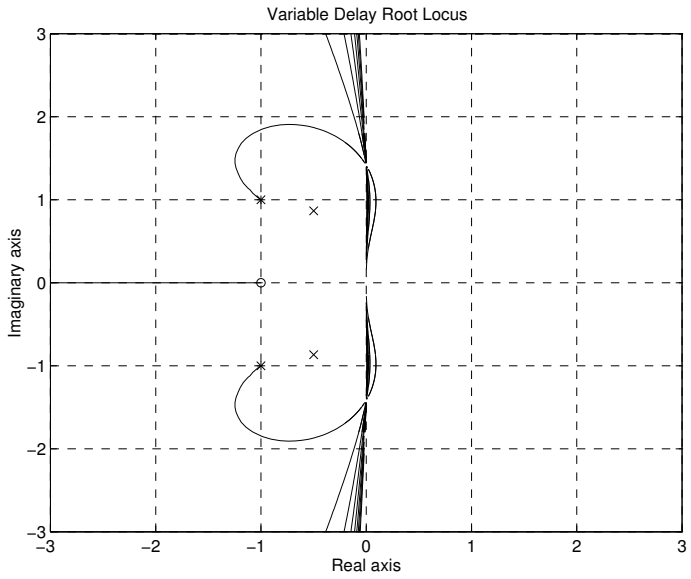


Figure 3-8: Variable delay root locus of the basic teleoperator system. Above the critical delay, the poles move into the RHP and drive the system unstable.

Then from (2.41) the delay itself is computed as

$$2T = \frac{\angle G(j\omega) - \pi - 2\pi k}{\omega} = \sqrt{\frac{2m}{K}} \arctan\left(B\sqrt{\frac{2}{mK}}\right) \tag{3.5}$$

So we have a relation between both P.D. gains, the inertia, and the largest acceptable delay, that we can use to tune the system.

3.3.2 Additional Damping

Intuitively adding damping should also help stabilize the system and thus increase the possible delays. So far we have only considered the damping inherent in the velocity feedforward of the P.D. controller. Now we add an explicit damping term to both master and slave inertia.

The new equations of motion with the added damping are

$$\begin{aligned} m\ddot{x}_s(t) &= -D\dot{x}_s(t) + F_{\mathrm{PD}}(t) \\ m\ddot{x}_m(t) &= -D\dot{x}_m(t) - F_{\mathrm{PD}}(t-T) \end{aligned}$$

where the force computed by the P.D. controller remains unchanged.

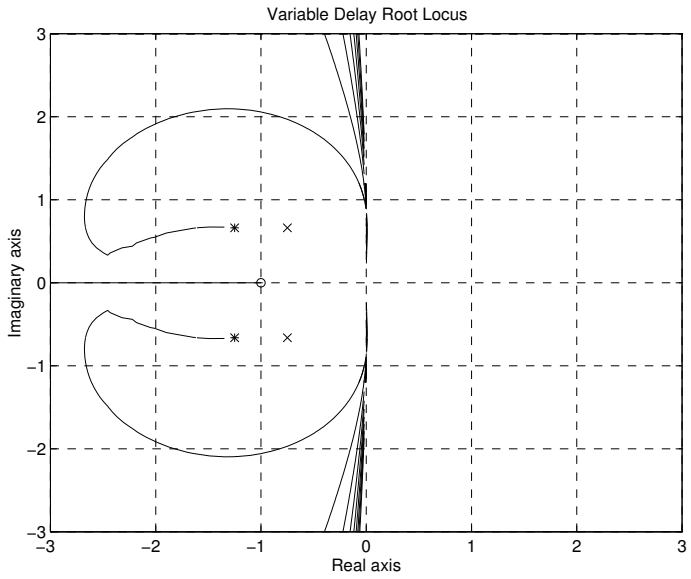


Figure 3-9: Variable delay root locus of the basic teleoperator system with added damping. The damping moves the open loop poles further into the LHP and reduces the region of instability.

The resulting OL transfer function is then

$$G_{OL}(s) = -\frac{Bs + K}{ms^2 + (B + D)s + K}\, e^{-2sT} \tag{3.6}$$

Notice the increased damping ratio in the open loop poles.

In Figure 3-9 we see the corresponding variable delay root locus. In this case, we find the point of instability is given determined by

$$\omega^2 = \frac{2mK - 2BD - D^2}{m^2} \tag{3.7}$$

which collapses to the previous case if $D = 0$.

First of all notice that the frequency ω is reduced by the added damping. This implies an increase in the delay T which is roughly inversely proportional thereto. Then, as the added damping is increased even further, the solution for ω^2 becomes negative so that it disappears completely.

So we find that for sufficient damping the system will be stable regardless of the

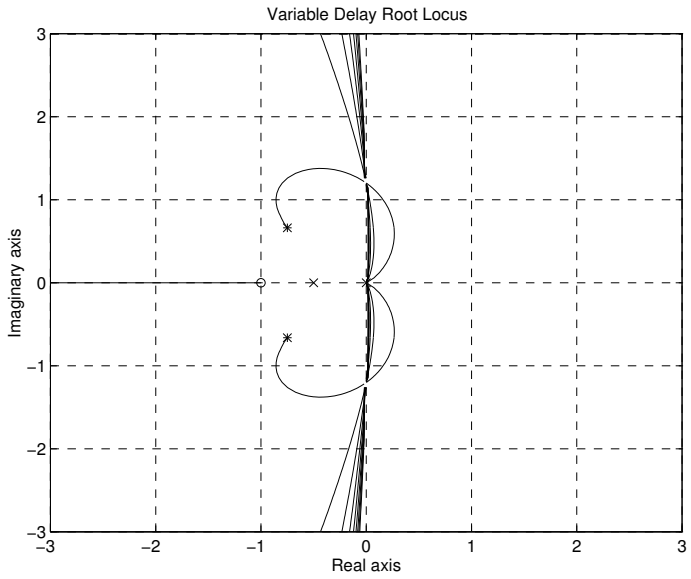


Figure 3-10: Variable delay root locus including added damping but with the slave in rigid contact. The changed model negates the previous results and we see an increased region of instability.

time delay. In particular the necessary minimum amount of added damping is

$$D = -\frac{B}{2} + \sqrt{\frac{B^2}{4} + 2mK} \tag{3.8}$$

This is a very promising result. We no longer need to now the delay precisely and it may even vary between runs. However, keep in mind that we have made assumptions on the system model to obtain these results. As we see next, should the model be inaccurate the result no longer holds.

3.3.3 Slave Contact

Consider the previous controller, but assume the slave has made contact to some rigid environment. That is, we now set $\dot{x}_s = 0$ and observe the result.

In this case, the open loop transfer function turns to

$$G_{OL}(s) = -\frac{Bs + K}{ms^2 + Ds} e^{-2sT} \tag{3.9}$$

Figure 3-10 gives the root locus and again shows a region of instability. Analysis shows that the point of instability occurs at a frequency of

$$\omega^2 = \frac{(D^2 - B^2) + \sqrt{(D^2 - B^2)^2 + 4m^2K^2}}{2m^2}$$

which always has a solution.

So the previous result of delay-independent stability is negated and reverted back to a case where the delay must be limited by some upper bound. While this is not catastrophic, it does imply that we need to consider all possible models and generate a design that works well under all conditions.

Of course, in the context of teleoperation, the set of possible models is quite large. Indeed the very point of teleoperation is the unknown environment and exploration. This is where passivity and wave variables provide significant benefit. They can provide the desired results, i.e. stability independent of delay, *without* requiring detailed models. The system thus becomes simultaneously robust to delay and unknown environments. In essence, wave variables automatically place the added damping in the areas where it is needed and with an amount necessary to guarantee the stability.

3.4 Examples of Delay-Induced Instability

In the previous sections we found that simply adding delay to the basic teleoperator can cause instability. The point at which the instability occurs depends heavily on the feedback gains, the delay time, and especially on the environmental interactions. This section presents the results of some early tests demonstrating two such situations.

In both tests, the operator is tightly holding the master joystick. This to prevent large motions and possible damage they may cause and to add further damping. The slave robot is in free space without any contact. The feedback gains were tuned in the undelayed case to reasonable values and include significant damping, indeed the system appears overdamped. Test 1 introduces a delay of $T = 0.1\,\mathrm{s}$, while test 2 uses a large delay of $T = 1.0\,\mathrm{s}$. And, of course, the tests are aborted when the behavior turns violent.

The graphs in Figures 3-11 and 3-12 show position and force signals for the master (solid line) and slave (dashed line) for all three degrees of freedom. The violent instability is clearly visible. While the operator can generally keep the displacements quite limited, he is not able to stop the strong and increasing oscillations.

Comparing both cases we notice that the large delay effectively modulates the

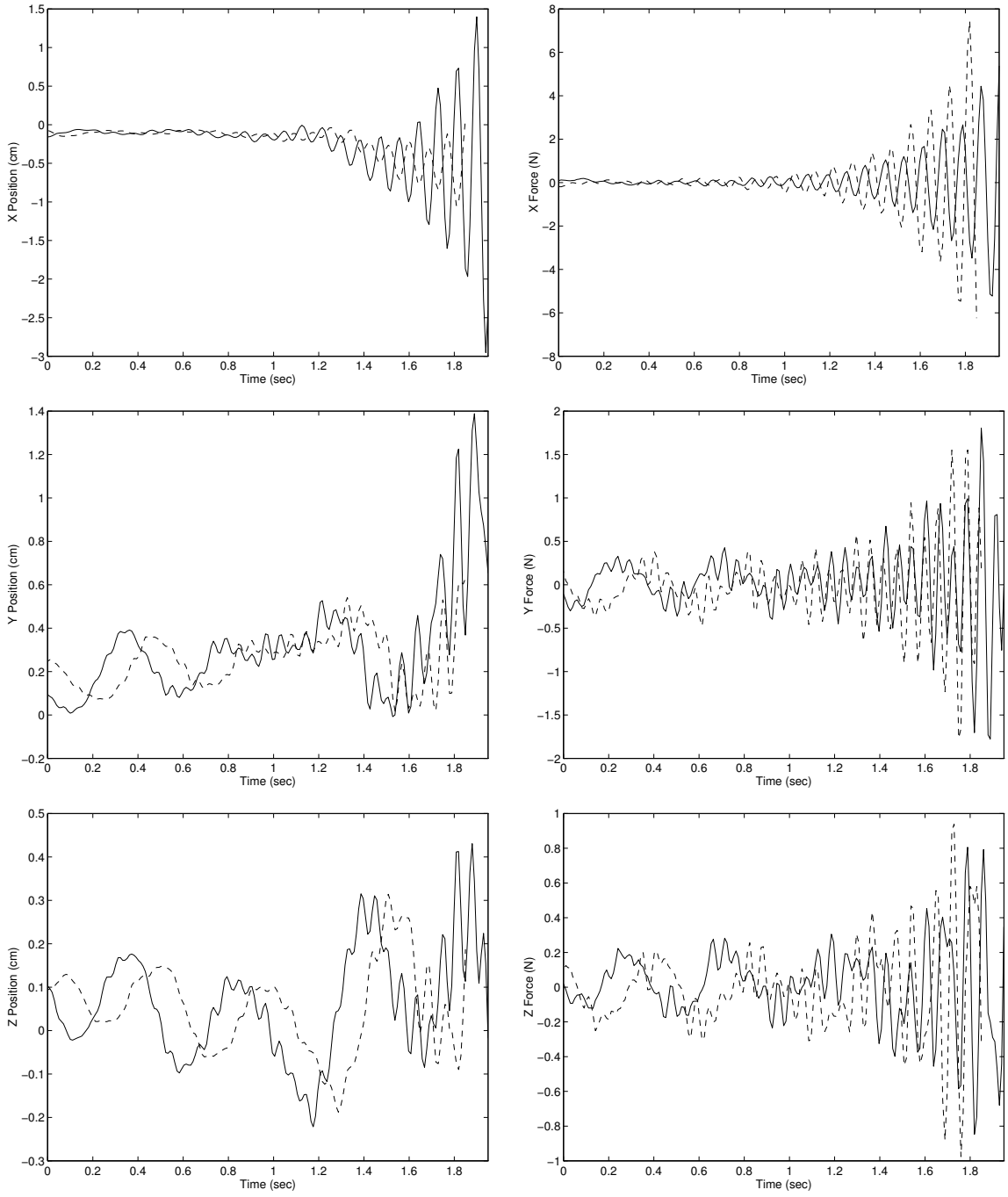


Figure 3-11: Basic delayed teleoperator with a delay of $T = 0.1$. The data clearly shows violent instability, even with the operator preventing large motions.

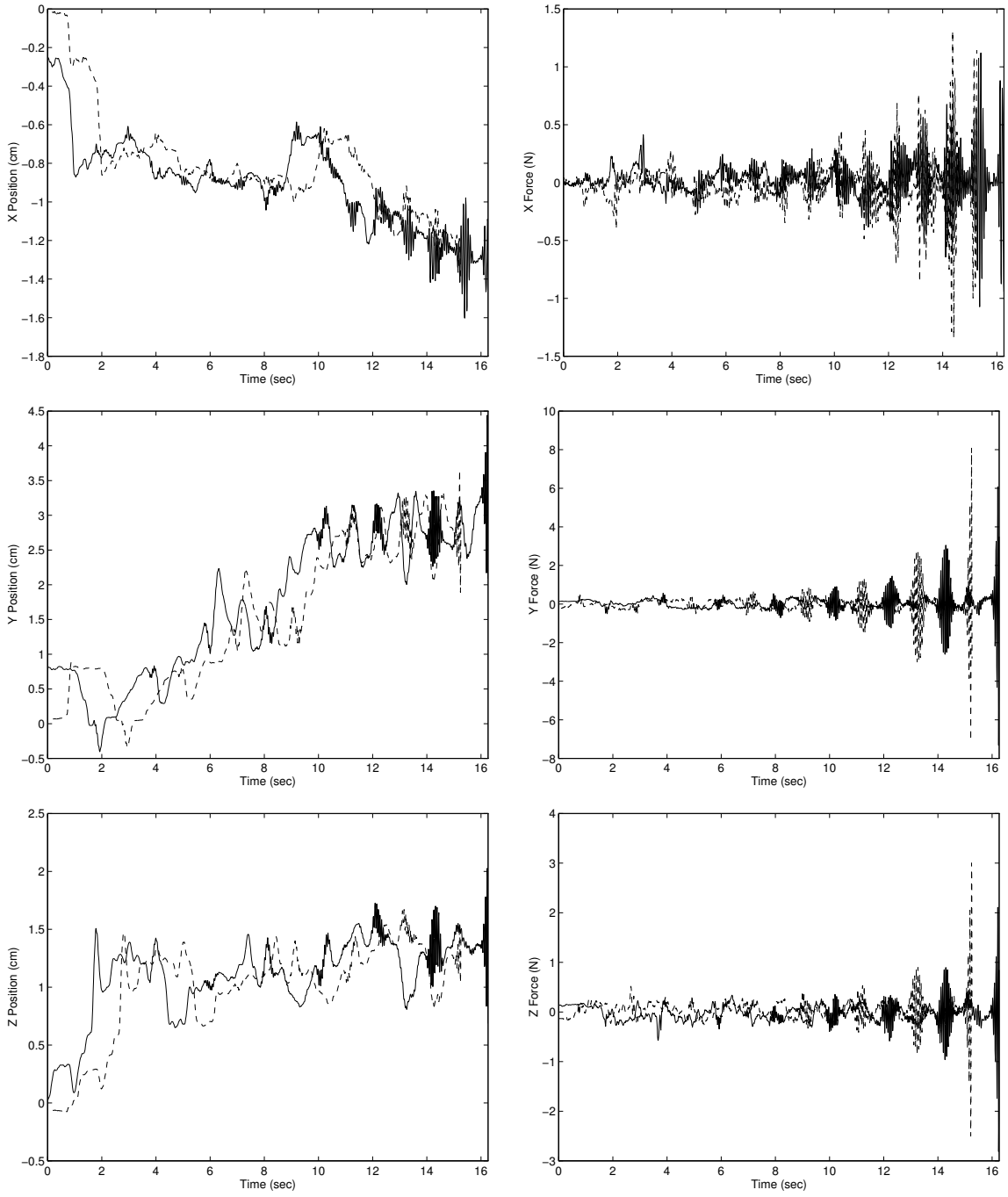


Figure 3-12: Basic delayed teleoperator with a delay of $T = 1.0$. The instability is just as inevitable, though the oscillations are modulated by the large delay.

behavior and we can literally watch the instability grow over a long period of time. From the root loci 3-8 through 3-10 we know that the closed loop poles move close to the imaginary axis and especially the origin as the delay increases. So the larger delay creates a slower instability, but subsequent poles still force oscillations at higher speeds.

3.5 Passivity Analysis and Modification

We now proceed to analyze the basic delayed teleoperation with force feedback shown in Figure 3-6 within the passivity framework. This work is based on the ideas of (Anderson and Spong, 1989a, 1989b).

3.5.1 Standard Communications

Of the four major components, master and slave inertia, P.D. controller, and communications, three are passive naturally or by construction. Only the communications element has not yet been tested. We do so here and find that it is indeed not passive, but creates energy, and causes the instabilities shown in the previous section.

The standard communications are simply governed by

$$\dot{\mathbf{x}}_{sd}(t) = \dot{\mathbf{x}}_m(t-T) \tag{3.10a}$$

$$\mathbf{F}_m(t) = \mathbf{F}_s(t-T) \tag{3.10b}$$

where $\dot{\mathbf{x}}_{sd}(t)$ is the current slave desired velocity. This is linear and time-invariant, so that we can use LTI passivity tools. First we construct the combined input and output vectors

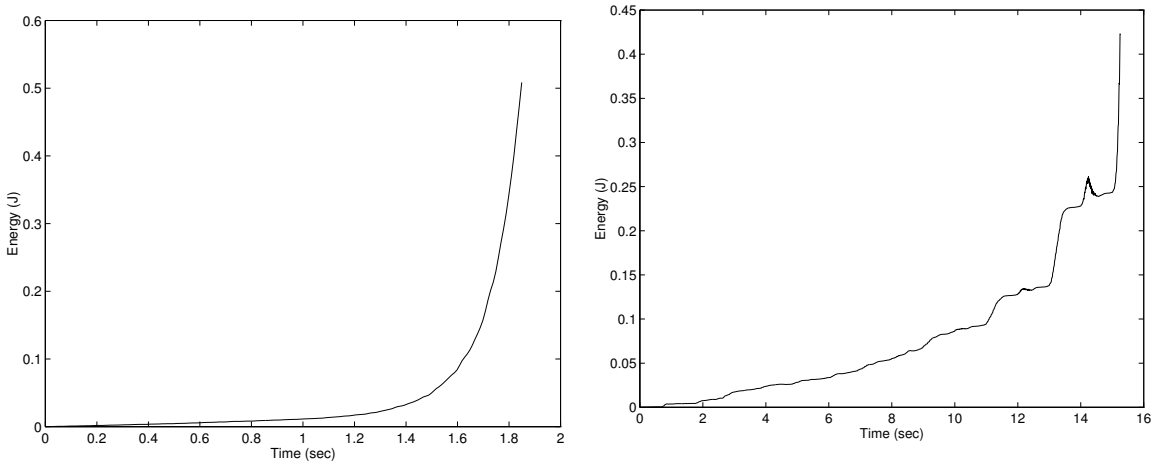
$$\text{input} = \begin{bmatrix} \dot{\mathbf{x}}_m \\ -\mathbf{F}_s \end{bmatrix} \qquad \text{output} = \begin{bmatrix} \mathbf{F}_m \\ \dot{\mathbf{x}}_{sd} \end{bmatrix}$$

where the inputs on the right side are inverted so that their inner product computes the total power input

$$P_{in} = \dot{\mathbf{x}}_m^T \mathbf{F}_m - \dot{\mathbf{x}}_{sd}^T \mathbf{F}_s$$

(see Section 2.1.3). Next we determine the corresponding transfer matrix as

$$\begin{bmatrix} \mathbf{F}_m \\ \dot{\mathbf{x}}_{sd} \end{bmatrix} = G(s) \begin{bmatrix} \dot{\mathbf{x}}_m \\ -\mathbf{F}_s \end{bmatrix} = \begin{bmatrix} 0 & -e^{-sT} \\ e^{-sT} & 0 \end{bmatrix} \begin{bmatrix} \dot{\mathbf{x}}_m \\ -\mathbf{F}_s \end{bmatrix} \tag{3.11}$$



(a) (b)

Figure 3-13: Shown is the energy produced by the communications element of the basic delayed teleoperator for both tests presented in the previous section. The delay times are (a) $T = 0.1\,\mathrm{s}$ and (b) $T = 1.0\,\mathrm{s}$. Instability is caused by this growing amount.

which leads to the Hermitian matrix

$$G(s) + G^*(s) = G(s) + G^T(\bar{s}) = \begin{bmatrix} 0 & e^{-\bar{s}T} - e^{-sT} \\ e^{-sT} - e^{-\bar{s}T} & 0 \end{bmatrix} \tag{3.12}$$

For passivity we need to test positive definiteness of (3.12) in the right half plane. We do this by examining the eigenvalues. First we find

$$e^{-sT} - e^{-\bar{s}T} = e^{-\sigma T}\left(e^{-j\omega T} - e^{+j\omega T}\right) = e^{-\sigma T}\left(-2i\sin(\omega T)\right)$$

Meanwhile the eigenvalues are determined from

$$\lambda^2 = -\left(e^{-sT} - e^{-\bar{s}T}\right)^2 = 4e^{-2\sigma T}\sin^2(\omega T)$$

to be

$$\lambda = \pm 2e^{-\sigma T}\sin(\omega T) \tag{3.13}$$

Clearly they take both positive and negative values, so that the $G(s) + G^*(s)$ is not positive definite or even positive semidefinite and the element is not passive.

This is also verified experimentally in Figure 3-13, which plots the energy output

of the communications element during both previous tests.

$$E_{out} = -\int_0^t P_{in}\,\mathrm{d}\tau = \int_0^t \dot{\mathbf{x}}_{sd}^T\mathbf{F}_s - \dot{\mathbf{x}}_m^T\mathbf{F}_m\,\mathrm{d}\tau$$

The continual grow indicates the energy production inconsistent with the passivity requirements. The delay time dictates the speed of the production and hence the speed of the instability.

3.5.2 Transmission Line Based Communications

Though the standard communications in the basic delayed teleoperator are not passive, there are many natural physical systems which are both bidirectional and passive. Classic examples are electrical transmission lines and tensioned strings. The following sets up a new communications law imitating these phenomena.

The transfer matrix of a transmission line is given by hyperbolic functions as

$$G(s) = \begin{bmatrix} z_0\tanh(sT) & -\operatorname{sech}(sT) \\ \operatorname{sech}(sT) & \frac{1}{z_0}\tanh(sT) \end{bmatrix} = \begin{bmatrix} z_0\frac{\sinh(sT)}{\cosh(sT)} & -\frac{1}{\cosh(sT)} \\ \frac{1}{\cosh(sT)} & \frac{1}{z_0}\frac{\sinh(sT)}{\cosh(sT)} \end{bmatrix} \tag{3.14}$$

where z_0 is the line impedance. Appropriate algebraic manipulations show that $G(s)+G^*(s)$ is positive semidefinite in the right half plane, verifying the passivity of the transmission line.

But perhaps a more effective path is to examine the corresponding scattering matrix. Again after some manipulation we find

$$S(s) = \begin{bmatrix} 0 & -e^{-sT} \\ e^{-sT} & 0 \end{bmatrix} \tag{3.15}$$

The passivity conditions quickly satisfied as the norm is constant at unity.

$$S^*(s)S(s) = \begin{bmatrix} e^{-2\sigma} & 0 \\ 0 & e^{-2\sigma} \end{bmatrix} \qquad \Rightarrow \qquad ||S(j\omega)|| = 1$$

We now replace the standard communications element with this transfer matrix, in essence simulating the behavior of a natural transmission line. This leads to the governing equations

$$\dot{\mathbf{x}}_{sd}(t) \;=\; \dot{\mathbf{x}}_m(t-T) \;-\; \frac{1}{n^2}\Big(\mathbf{F}_s(t) - \mathbf{F}_m(t-T)\Big) \tag{3.16a}$$

$$\mathbf{F}_m(t) \;=\; \mathbf{F}_s(t-T) \;+\; n^2\left(\dot{\mathbf{x}}_m(t) - \dot{\mathbf{x}}_{sd}(t-T)\right) \tag{3.16b}$$

where n is a scaling parameter. Using these modified communications, all elements are passive and we can guarantee system stability regardless of the delay time. This law first appeared in (Anderson and Spong, 1989b, Equations (7.9) and (7.10)).

Of course, such modifications may have significant impact on the behavior of the system, especially for large delays, much of which is hidden in (3.16). For example, Figure 3-14 shows the results of a step input. The operator moves the joystick, then holds it in place as best he can. Notice the slave position and master force oscillate wildly. Only the slave forces remain near zero without any external contact. While admittedly not well tuned, this test does illustrate the existence of additional dynamics caused by signals reflecting back and forth between master and slave at a delay of $T = 1.0\,\mathrm{s}$. This in turn excites oscillations at higher frequencies. After developing appropriate tools in the form of wave variables, we will continue and build on this work to derive a more detailed understanding of these effects.

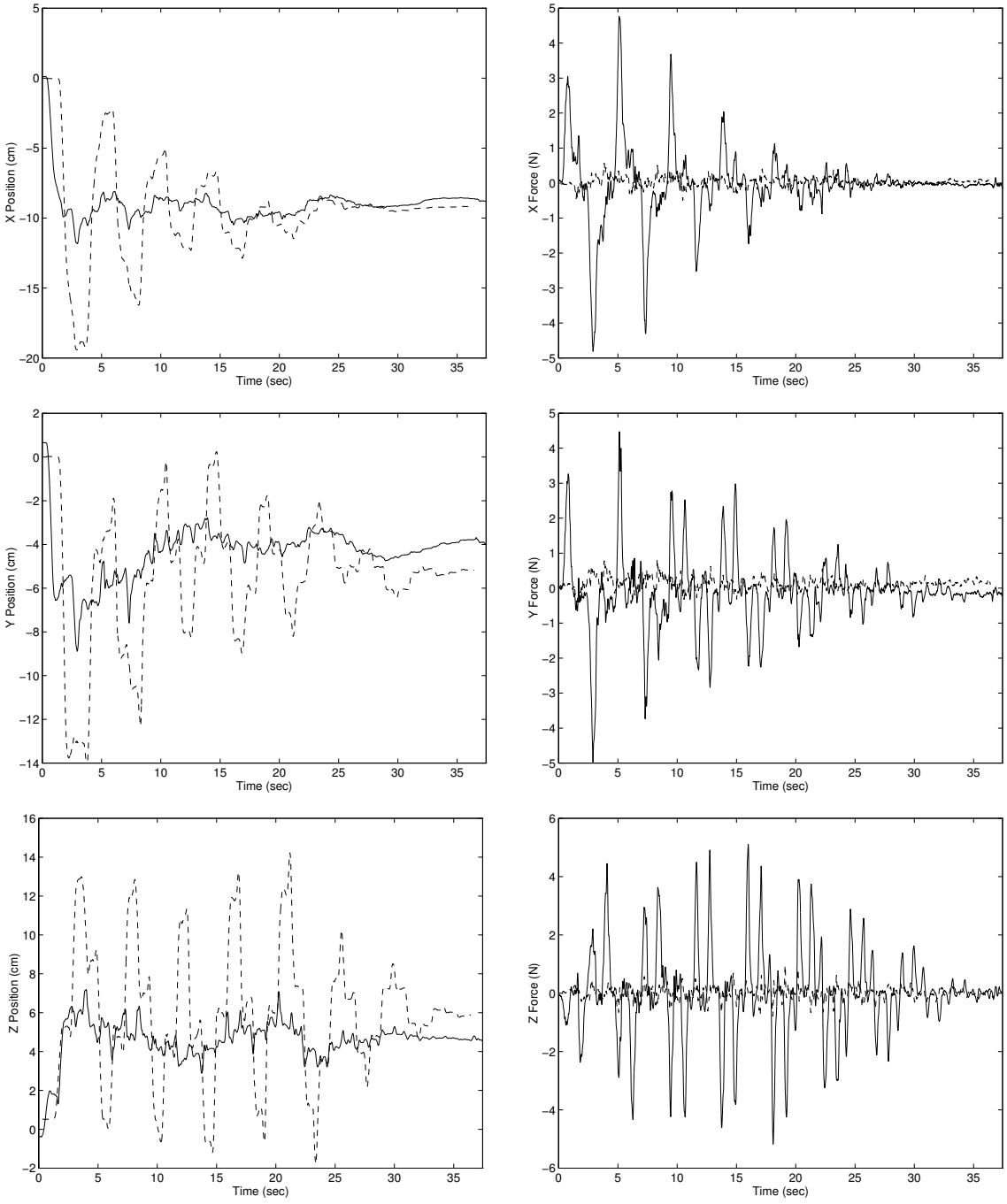


Figure 3-14: The modified communications law implies passivity and stabilizes the delayed teleoperator ($T = 1.0$). However, additional dynamics appear due to reflections between master and slave.

Chapter 4

Wave Variables

Wave variables present a modification or extension to the theory of passivity which creates robustness to arbitrary time delays. Based only on the concepts of power and energy, they are applicable to nonlinear systems and can handle unknown models and large uncertainties. As such they are well suited for interaction with real physical environments and teleoperation.

To achieve their goals, wave variables provide an alternate information encoding scheme to the standard power variables. The required transformations are extremely simple and preserve all information. In combination with delays, this scheme automatically introduces damping elements which limit the best achievable performance to levels attainable within the delay limitations.

The following sections examine wave variables in detail. We first define and interpret the basic waves, including their effect on passivity. Then we proceed by looking at systems from a wave perspective, both open loop elements and closed loop behavior. We conclude with some remarks on where and when the wave representation may be useful.

4.1 Definition

The entire concept of wave variables is based within the framework of passivity and was in particular motivated by scattering operators (Desoer and Vidyasagar, 1975). So not surprisingly, the basic definition of wave variables is also rooted therein. More specifically, we redefine the power flow as

$$P = \dot{\mathbf{x}}^T \mathbf{F} = \tfrac{1}{2}\mathbf{u}^T\mathbf{u} - \tfrac{1}{2}\mathbf{v}^T\mathbf{v} \tag{4.1}$$

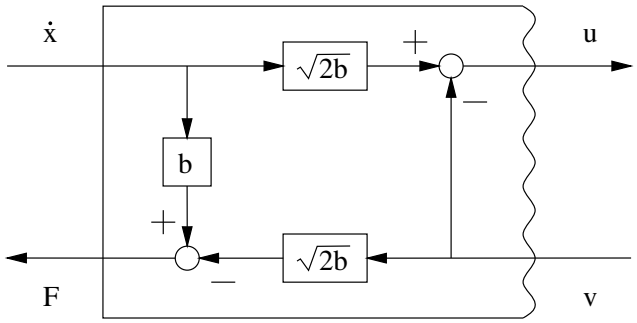


Figure 4-1: The basic wave transformation relates velocity, force, right, and left moving waves, in this case converting $(\dot{\mathbf{x}}, \mathbf{v})$ into $(\mathbf{F}, \mathbf{u})$. The wavy line signifies the wave variable port.

Note the force $\mathbf{F}$ and velocity $\dot{\mathbf{x}}$ variables may be replaced by any other effort and flow pair. However, our application is mechanical in nature, so we continue to use this mechanical notation.

In accordance with the definitions introduced in Section 2.1.3, we find that $\frac{1}{2}\mathbf{u}^T\mathbf{u}$ specifies the power flowing *along* the main direction leading to a positive value for P. In contrast, $\frac{1}{2}\mathbf{v}^T\mathbf{v}$ specifies the power flowing *against* the main direction and introduces a negative element into P. In this sense, we define $\mathbf{u}$ to denote the forward or right moving wave, while $\mathbf{v}$ denotes the backward or left moving wave.

The wave variables $(\mathbf{u}, \mathbf{v})$ can be computed from the standard power variables $(\dot{\mathbf{x}}, \mathbf{F})$ by the following transformation.

$$\mathbf{u} = \frac{b\dot{\mathbf{x}} + \mathbf{F}}{\sqrt{2b}} \qquad \mathbf{v} = \frac{b\dot{\mathbf{x}} - \mathbf{F}}{\sqrt{2b}} \tag{4.2}$$

where b is the characteristic wave impedance and may be a positive constant or a symmetric positive definite matrix. Section 4.4 will describe its meaning in more detail.

This transformation is bijective, i.e. one to one, so that it is always unique and invertible. No information is lost or gained by encoding the variables in this fashion. We can thus also compute the power variables as

$$b\dot{\mathbf{x}} = \sqrt{\frac{b}{2}}\,(\mathbf{u} + \mathbf{v}) \qquad \mathbf{F} = \sqrt{\frac{b}{2}}\,(\mathbf{u} - \mathbf{v}) \tag{4.3}$$

Finally, the wave transformation may be used to determine any combination of power and wave variable. This is, any two of the four variables can be picked as input

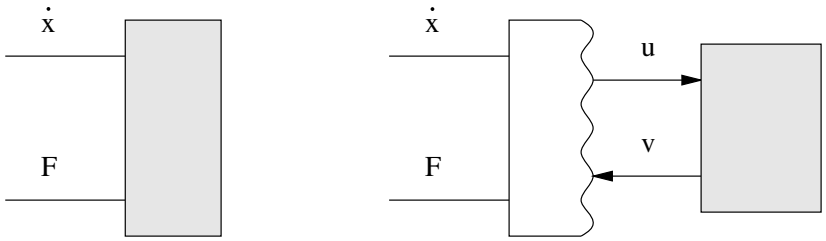


Figure 4-2: Systems in both the power and wave domain can be tested for passivity directly in their respective domain.

or output. For example, consider the velocity $\dot{\mathbf{x}}$ and left moving wave $\mathbf{v}$ as the inputs. The appropriate transformation will then determine the force $\mathbf{F}$ and right moving wave $\mathbf{u}$.

$$\mathbf{F} = b\dot{\mathbf{x}} - \sqrt{2b}\,\mathbf{v} \qquad\qquad \mathbf{u} = -\mathbf{v} + \sqrt{2b}\,\dot{\mathbf{x}} \tag{4.4}$$

Figure 4-1 shows this graphically. It is transformations of this kind that provide the interface between power and wave variables. Notice especially the apparent damping element of magnitude b within the transformation. It retains the power necessary to compose the wave signal. But it also effects the system behavior and is relevant to many discussions in the following chapters (see in particular Sections 4.3.5 and 4.4).

4.2 Condition for Passivity

One of the main reasons for dealing with wave variables is their effect on the condition for passivity. Remembering Section 2.1 we know that a system is passive if it can not produce energy. We now reexpress this condition within the domain of wave variables.

As depicted in Figure 4-2 we assume, for the moment, that power arrives from the left and hence positive power flow is pointing into the element. This also implies that the right moving wave $\mathbf{u}$ is the input, while the left moving wave $\mathbf{v}$ contains the output. In the power domain passivity is then tested by the condition

$$\int_0^t P_{in}\,\mathrm{d}\tau = \int_0^t \dot{\mathbf{x}}^T\mathbf{F}\,\mathrm{d}\tau = E_{store}(t) - E_{store}(0) + \int_0^t P_{diss}\,\mathrm{d}\tau \tag{4.5}$$

where the stored energy $E_{store}(t)$, the initial stored energy $E_{store}(0)$, and the power dissipation $P_{diss}(t)$ must all be nonnegative.

Substituting the redefined power flow (4.1), which formed the basis for the wave

definitions, we find

$$\int_0^t P_{in}\,\mathrm{d}\tau = \int_0^t \tfrac{1}{2}\mathbf{u}^T\mathbf{u} - \tfrac{1}{2}\mathbf{v}^T\mathbf{v}\,\mathrm{d}\tau = E_{store}(t) - E_{store}(0) + \int_0^t P_{diss}\,\mathrm{d}\tau$$

This leads us to the passivity condition in the wave domain

$$\int_0^t \tfrac{1}{2}\mathbf{v}^T\mathbf{v}\,\mathrm{d}\tau = \int_0^t \tfrac{1}{2}\mathbf{u}^T\mathbf{u}\,\mathrm{d}\tau + E_{store}(0) - E_{store}(t) - \int_0^t P_{diss}\,\mathrm{d}\tau \qquad \forall t \geq 0 \qquad (4.6)$$

or in the simplified form

$$\int_0^t \tfrac{1}{2}\mathbf{v}^T\mathbf{v}\,\mathrm{d}\tau \leq \int_0^t \tfrac{1}{2}\mathbf{u}^T\mathbf{u}\,\mathrm{d}\tau + E_{store}(0) \qquad \forall t \geq 0 \qquad (4.7)$$

So not surprisingly we find a system to be passive if the energy in the outgoing wave (in this case $\mathbf{v}$) is limited to the energy provided by the incoming wave (in this case $\mathbf{u}$) or stored initially. Should the interaction occur on the right side, the right moving wave $\mathbf{u}$ must assume the role of output, while the left moving wave $\mathbf{v}$ provides the input. In this case their respective roles in the condition (4.7) are also reversed, so that the output wave remains bounded.

Most importantly, the new condition removes the multiplicative co-dependency of power variables seen in (4.5). In the old form passivity depends on the product of both variables. If the output variable is delayed, the effect on the product and thus passivity is unpredictable. In the new form, however, passivity compares only the integrated magnitudes. If the output variable is delayed, its power is temporarily stored without changing the passivity. For example, consider a simple delay

$$\mathbf{v}(t) = \mathbf{u}(t - T)$$

The passivity condition (4.7) becomes

$$\begin{aligned} & \int_0^t \tfrac{1}{2}\mathbf{v}^T\mathbf{v}\,\mathrm{d}\tau &\leq& \int_0^t \tfrac{1}{2}\mathbf{u}^T\mathbf{u}\,\mathrm{d}\tau + E_{store}(0) \\ \Leftarrow \quad & \int_0^{t-T} \tfrac{1}{2}\mathbf{u}^T\mathbf{u}\,\mathrm{d}\tau &\leq& \int_0^t \tfrac{1}{2}\mathbf{u}^T\mathbf{u}\,\mathrm{d}\tau \\ \Leftarrow \quad & 0 &\leq& \int_{t-T}^t \tfrac{1}{2}\mathbf{u}^T\mathbf{u}\,\mathrm{d}\tau \end{aligned}$$

which is clearly satisfied. Note the initial stored energy $E(0)$ is zero (assuming, of course, that $\mathbf{u} \equiv 0$ for all $t < 0$). From (4.6) we also find that the stored energy is

given as

$$E_{store}(t) = \int_{t-T}^{t} \tfrac{1}{2}\mathbf{u}^T\mathbf{u}\,\mathrm{d}\tau$$

which indeed stores the input power for the duration of the delay, and the power dissipation is zero, making the delay lossless.

So all systems, which were passive in the power variable notation, remain passive after transformation into wave variables. But in addition, time delays are now also passive elements. So systems expressed in wave variables become completely robust to delays of any amount.

Finally, we see that a simple concatenation of passive wave-based elements remains passive. Unfortunately this comes at a price. The standard negative feedback or parallel configurations, which maintain passivity in the power domain, no longer do so in the wave domain.

4.3 Interpretation

This may be the trickiest aspect and possibly the biggest drawback of wave variables. As beneficial as their use can be, which we will detail later, their physical interpretation is not as intuitive as that of power variables. Indeed, while velocity and force clearly correspond to physical quantities and are readily measurable, wave variables are a mathematical construction without direct manifestation. Even their units (being $\sqrt{\text{Watt}}$) are unusual.

Nevertheless, wave variables have a useful meaning in themselves. It has even been suggested by (Massaquoi and Slotine, 1996) that wave variables may be used in biological systems, in particular humans, to control motor functions. We hope this and the remaining sections will provide a better understanding and intuition.

4.3.1 Symmetry

First we notice the symmetry in the wave variable definition (4.2). Both waves are subject to the same interpretation, regardless of the direction they are associated with. In particular, right and left moving waves are only distinguished by a change of sign in the force $\mathbf{F}$.

Consider an interaction between two subsystems or elements, which both push and pull upon each other. Having defined a global reference frame, they will both measure the same velocity and position. However, according to Newton's Law, they

will experience an equal but opposite force. It is this opposite force that is captured by the opposite signs in the wave definitions.

So each wave represents similar information differing only in the point of view.

4.3.2 Lack of Effort/Flow Distinction

On key benefit of this symmetry is the lack of distinction between force and velocity. While this may be confusing and can add to the non-intuitive nature, it actually simplifies several problems.

It eliminates the need to separate force feedback designs from position/velocity feedback designs. Essentially both layouts are combined and performed simultaneously.

Also it removes causality restrictions. In traditional systems, adjoining elements must be of opposite causality, i.e. one element determines forces, while the other computes positions/velocities. Examples are simple mass-spring-mass sequences.

Wave systems do not require such restrictions. Should two elements be joined that perform similar functions, their action is automatically averaged. Much like a series connection of two springs results in a weaker stiffness.

Compare this, for example, with Bond Graphs, where such 'bad' connections result in derivative causality and must be manually eliminated by combining neighboring elements.

4.3.3 Move or Push Commands

This brings us to the core of what wave variables are. They always combine both force and velocity into a single variable, never selecting just one. Therefore, none of the signals can ever be considered simply a position command or a force command.

Instead, every wave should be interpreted as a general-purpose '*move or push*' command. As usual the sign determines the direction and should be labeled 'forward' for positive and 'backward' for negative.

It is now up to the recipient of the command to interpret the action required to fulfill the command. For example, issuing a 'forward move or push' command to a mass will result in a positive motion. Yet the same command will cause a spring to compress and push into its base.

We will see that it is the reflection returning from the recipient that tells us exactly

what happened and how the command was interpreted.

4.3.4 Elastic String Examples

Consider the meaning of the waves in the classic example of a flexible string or rope under tension. In the first case, shown in Figure 4-3, attach one end of the rope to a rigid wall and shake the other. Initially the ingoing wave $\mathbf{u}$ simply traverses without being distorted. This corresponds to a simple time delay. Then it encounters the wall, where it is completely reflected while its sign is reversed, i.e. the outcoming wave $\mathbf{v} = -\mathbf{u}$. The outcoming wave must also traverse for a fixed amount of time before reaching the initial point. Note the boundary condition at the wall demands no motion, so that at the end the rope has returned to its original state.

If we interpret the example more carefully, the original shake issues a 'move or push' command. Now the wall will not allow motion, so the command is executed as a pure push. Of course, the wall will counter with an opposite push. This opposition is encoded and returned in form of the negative 'move or push' return signal.

So, we find that a wall is described by

$$\mathbf{v} = -\mathbf{u}$$

and counters every 'move or push' command by an opposite return command. From (4.2) we find the same result by forcing $\dot{\mathbf{x}} = 0$.

Figure 4-4 shows the other extreme, in which the rope is not attached to anything, but hanging in free space. In this case the wave is also completely reflected, but with the same sign. As there may be no forces at the boundary to free space, the 'move or push' command is interpreted completely as a move, and this is signaled by an equal return signal.

But also notice that in this case the rope does not return to the original position. The command was interpreted as a move and its execution caused the shift in position.

Mathematically, we can verify this behavior by setting $\mathbf{F} = 0$ in (4.2), and indeed find that free space is governed by

$$\mathbf{v} = \mathbf{u}$$

Should we attach other elements to the end of the rope, we will find they all reflect incoming waves in different degrees and with differing signs and distortions. Various elements are analyzed in Section 4.6

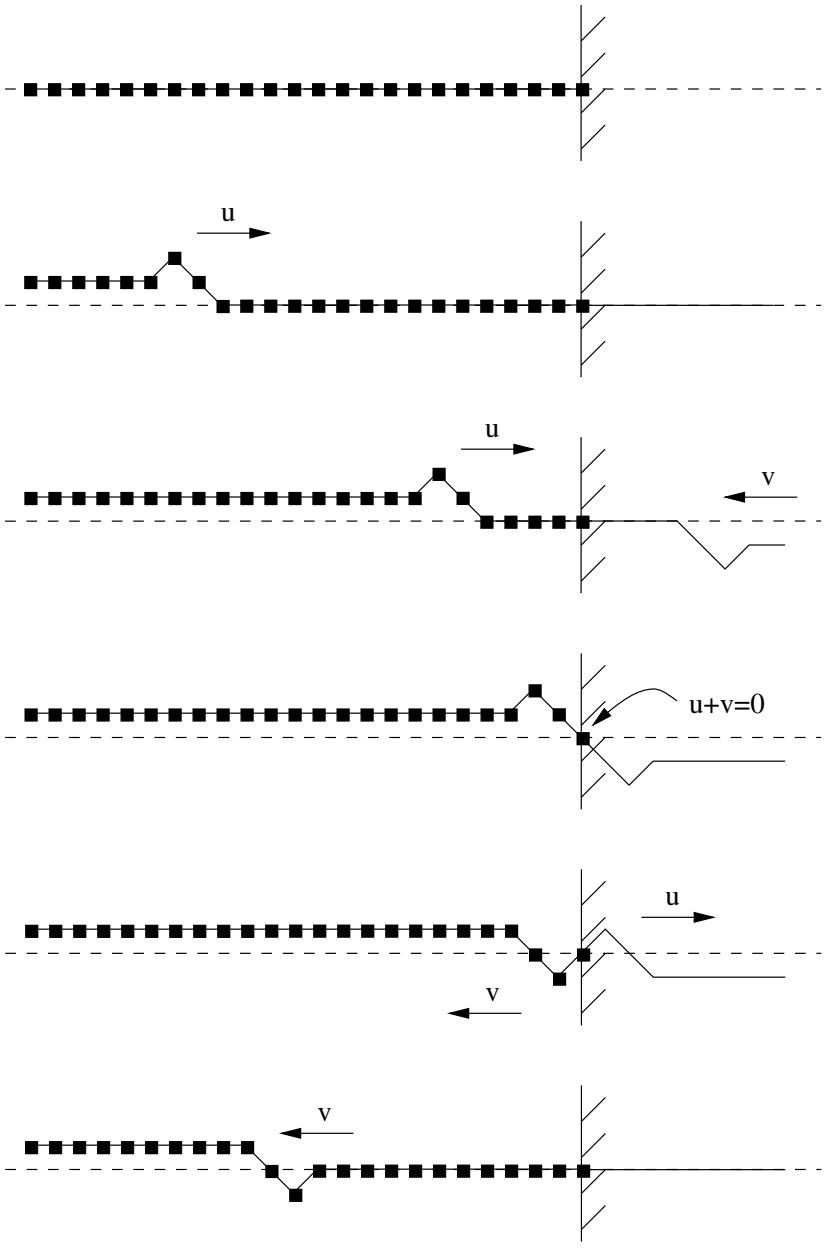


Figure 4-3: A solid wall will reflect any incoming wave with an equal but opposite return signal. This to ensure its boundary condition of no motion.

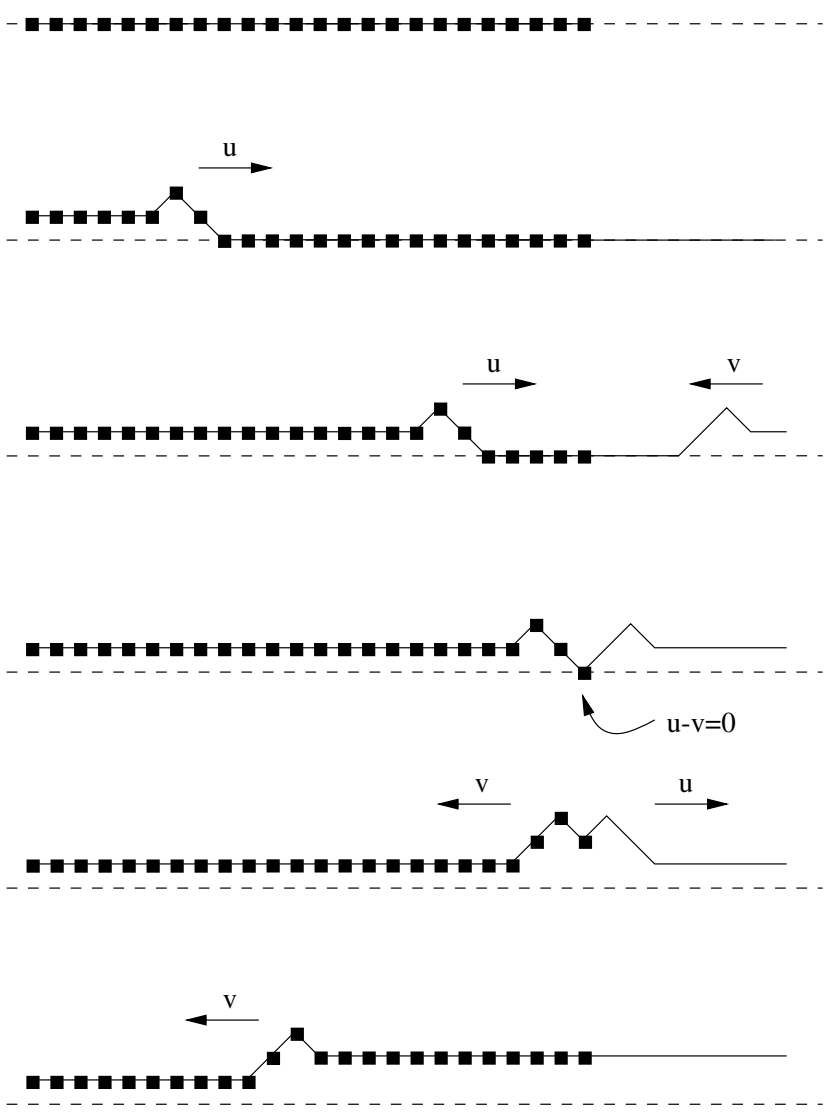


Figure 4-4: An unconnected rope in free space will reflect any incoming wave with an equal return signal. This ensures that no forces are exerted at the boundary.

4.3.5 Power and Energy

From their definition, we see that each wave also contains enough power to execute its own command. This combination of power and information into a single variable is one of the key features and allows the robustness to delays we desire. It is also highlighted by the units in which wave variables are expressed, namely $\sqrt{\text{Watt}}$.

To ensure that enough power is available for a wave command, the wave transformation contains an apparent damping element of magnitude b (compare to Figure 4-1). This element extracts the necessary power from the system to compose the wave. But note that the damping is only apparent. The power is not dissipated but included in the outgoing wave signal.

Furthermore, any returning or incoming wave signal will also provide power. And so the effects of this damping element are often negated and never appear to an operator. Only when no wave signal returns, or it returns after considerable delays, will the damping be noticeable.

As the previous sections have indicated, it is the recipient of a wave that determines how the command is executed. Naturally the recipient thereby also determines how the power is used. Inertial elements will convert the power into kinetic energy and move, while spring-like elements will use it to generate potential energy and thus push.

And so the magnitude of the return signal tells us something about the amount of energy absorbed or lost. At one extreme, both a solid wall and free space can not absorb any energy. They thus return the full magnitude. Inertia and capacitors will retain some energy, though they will eventually return all once they are up to full speed or charge. Only dampers can continually absorb the input energy and always return wave signals of smaller magnitude.

4.3.6 Position Information

In most applications we will be interested in controlling or at least observing the position of a system. This information is still contained in the wave variables, albeit more hidden. From (4.3) we see the velocity can be computed as the sum of outgoing and returning waves. So the position is given as

$$\mathbf{x} = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}(\tau) + \mathbf{v}(\tau)\, \mathrm{d}\tau \tag{4.8}$$

In general, position information and behavior depends on the integral of wave

Power Variables	Wave Variables
Both are based on passivity and contain the same information. Results may be transfered between domains.	
Signal and power are separate.	Signal and power are combined.
Sensitive to time delay as well as co-location.	Robust to arbitrary time delays and co-location.
Inertial vs. capacitive elements, distinction between command and feedback, force and position control.	No distinction between outgoing and returning signals.
Mechanically measurable.	Mathematically constructed via an algebraic transformation — not directly measurable.
Units of Newton and meters/second.	Units of $\sqrt{\text{Watt}}$.

Table 4.1: Comparison of power and wave variables.

signals. Thus when developing and updating controllers, we find that while the wave magnitude should be limited for passivity, the wave integral should not be reduced or altered. This typically leads to smoothing or filter like appearances.

Similarly, to adjust the position of a system, we must issue an input command or wave $\mathbf{u}$ until the integral has reached the desired value. We will discuss exact strategies later, which need to be both passive and take into account physical limitations such as actuator bandwidth.

4.3.7 Power vs. Wave Variables

Finally we summarize the distinctions and commonalities between power variables and wave variables in Table 4.1.

4.4 Wave Impedance

The wave impedance b presents a tuning parameter which can trade off the speed of motion and levels of force. The following briefly discusses this functionality, which will appear repeatedly in later sections.

To allow the wave definitions in Equation (4.2) to combine force $\mathbf{F}$ and velocity $\dot{\mathbf{x}}$, we must include a scaling term that aligns their respective units. This is the primary reason for introducing the wave impedance b. However, it also provides a relative weighting between the two that we can use to adjust the system behavior to our advantage.

Consider a system described in the wave domain and assume a given wave variable $\mathbf{u}$. Now imagine increasing the wave impedance. This will place a larger weight on the velocity $\dot{\mathbf{x}}$ as compared to the force $\mathbf{F}$. The same value of $\mathbf{u}$ therefore leads to smaller values for $\dot{\mathbf{x}}$ and larger values for $\mathbf{F}$. In other words, an increase in impedance will raise force levels while reducing motion. The system will appear more damped.

The opposite is true when the wave impedance is decreased. Now the system associates a larger cost with forces, which are reduced at the price of higher velocities. The system will appear less damped.

Indeed, if we examine Figure 4-1, we see that the wave transformations introduce a direct feedback path from velocity to force, producing an apparent damping element with a magnitude b. The same is verified in (4.4).

However, this is only an *apparent* damping. The energy extracted here is actually the very energy associated with the wave command. It is not dissipated but transmitted to the rest of the system. Now if none of the energy returns, or if it returns only after a significant delay, we will clearly feel the apparent damping. But if the energy is reflected immediately, this apparent damping will be nullified. So the wave impedance becomes more important, as delays increase and the local feedback path in the wave transformation is exposed.

But regardless of this apparent damping, the wave impedance still scales force and velocity information in the wave signals. As we begin to add elements in the wave domain, including the wave transmission for the teleoperation in Section 5.1, this parameter will be built in automatically. Even after all other parameters are fixed, we can still adjust the impedance to emphasize forces or motions.

So we may select between a slow-moving teleoperator system with detailed force information and a fast-moving system without significant force feedback. The impedance thus allows us to explore the various ranges of possible performance bounds.

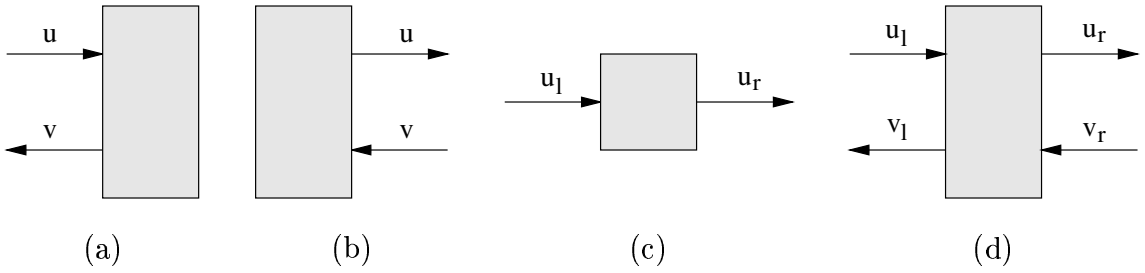


Figure 4-5: Open loop wave systems can map (a) the right moving wave **u** to the left moving wave **v**, or (b-d) other combinations with single or multiple input and output ports.

Of course, part of this tuning process also involves matching the impedance presented by external systems over which we have no other control.

We finally note that this parameter can only provide scaling. It may not include any additional dynamic terms. This limits it to positive constants and constant symmetric positive definite matrices.

4.5 Open Loop Wave Systems

As we discussed above, wave variables do not have a direct manifestation, but instead may be interpreted as a general 'move or push' command. The actual physical action caused by a wave command, whether motion or force, becomes evident when relating the incident and returning wave. So we now take a closer look at this input-output relation of wave systems.

These systems may use any combination of left and right moving waves as input and output signals, for instance see Figure 4-5. However, to simplify the following discussion, we will assume the structure shown under part (a). That is, all interaction occurs on the left side, with **u** providing the input and **v** denoting the output. For other layouts the roles of **u** and **v** should be adjusted accordingly.

4.5.1 General Comments

First, from their very definition, the magnitude of any wave variable describes the power it contains. And so the relative magnitude between **u** and **v** provides a measure of the power losses, in particular when averaged over time. Indeed the passivity condition itself formalizes this.

The power difference between input and output is either be stored or dissipated. So while draining the stored energy, the output magnitude may temporarily exceed the input. But a consistently low output magnitude implies a damped system. And conversely, at typically high output dictates an energy efficient system without much loses.

We also find the relative sign important. It signifies what type of action was taken. More precisely, at each instance in time

$$\text{sgn } \frac{v(t)}{u(t)} = \text{sgn } \Big[v(t)u(t)\Big] = \text{sgn } \Big[(b\dot{x})^2 - F^2\Big] \tag{4.9}$$

In other words the relative sign between input and output is determined by the relative magnitude of $b\dot{x}$ versus F. And so systems dominated by large forces with little motion will return waves of opposite signs. While systems exhibiting fast motions and low forces will return waves of the same sign.

Notice the role of the wave impedance b. As mentioned before, it trades off the relative importance of forces and velocities. Updating its value will change the wave response for a given physical element, or equivalently it will change the physical appearance for a given wave element.

Of course, the input-output map may also be described formally by operators. Indeed the map $\mathbf{u} \to \mathbf{v}$ implements the scattering operator associated with the $b\dot{\mathbf{x}} \to \mathbf{F}$ system (Desoer and Vidyasagar, 1975). This motivated much of the original development and can also be seen by comparing the wave variable passivity condition (4.7) to the scattering operator condition (2.12).

4.5.2 LTI Description

Naturally, wave variables are applicable to linear time-invariant (LTI) systems. In these cases we can proceed in the Laplace domain and define the wave transfer function

$$H(s) = \frac{\mathbf{V}(s)}{\mathbf{U}(s)} \tag{4.10}$$

where again we note the direct relation to scattering matrices. For multi-input multi-output (MIMO) systems, $H(s)$ denotes a transfer matrix.

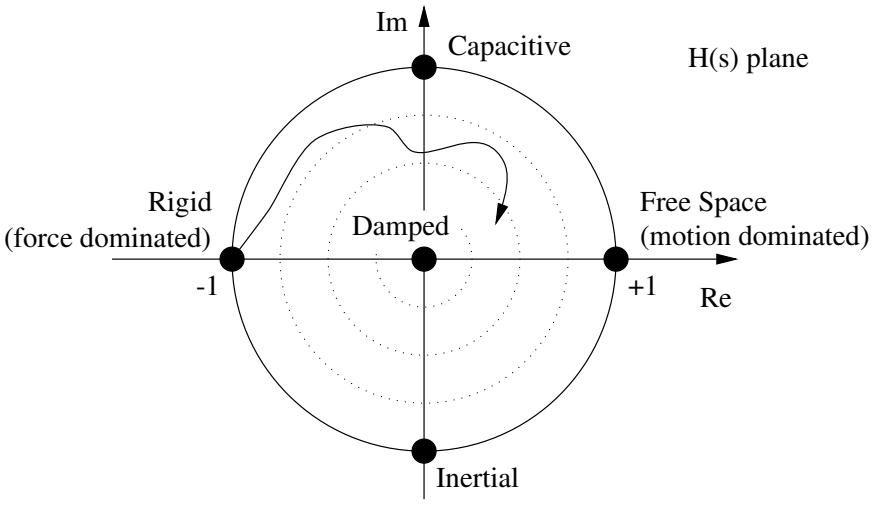


Figure 4-6: The graph of the wave frequency response $H(j\omega)$ remains within the unit circle of the $H(s)$ plane. Various regions are responsible for different behavior.

4.5.2.1 Wave Frequency Response

We pay particular attention to the wave frequency response $H(j\omega)$. It's magnitude represents the power gain at each frequency ω. Not surprisingly, passivity requires this gain be limited below unity for all frequencies to prevent any power increase. Hence the graph of $H(j\omega)$ in the $H(s)$ plane lies entirely within the unit circle.

The graph also presents other information. Various regions of the $H(s)$ plane correspond to different behaviors, as sketched in Figure 4-6. The location on the real axis describes the relative dominance of forces versus motions, positive being motion dominated and negative locations exhibiting higher force values. In the extremes $(+1)$ defines free space with no forces and (-1) rigid interactions with no motion.

On the imaginary axis we find capacitive versus inertial behavior, corresponding to phase lead or lags. And, of course, the distance to the origin signifies the dissipation. Lower magnitudes force more losses, while higher magnitudes provide better efficiency, and unit magnitudes are lossless.

An example is given in Figure 4-7. For low frequencies neither spring nor inertial dynamics appear and the system is dominated by the damper. Then as frequencies increase, first the large inertia dominates, then the spring and finally the light inertia. Correspondingly, the graph moves from the origin through the lower half plane, into the upper part and finally via the lower to the (-1) point.

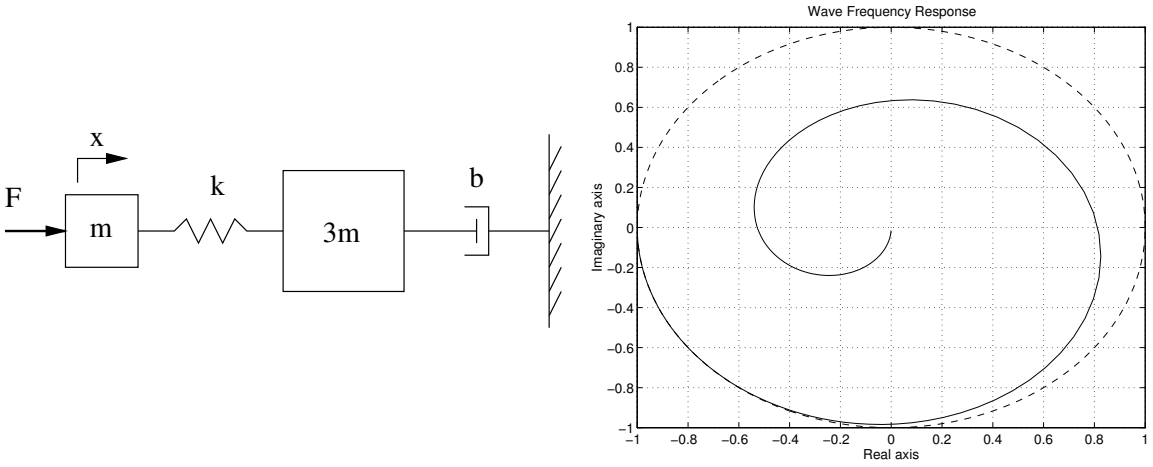


Figure 4-7: Shown is the wave frequency response $H(j\omega)$ of the depicted system. Notice the behavior varies from damped, inertial, spring-like to inertial for increasing frequencies. This corresponds to the various elements which dominate in those ranges.

4.5.2.2 Relation to Power Transfer Function

The wave transfer function (scattering operator) is related to the classic transfer function between the power variables velocity and force

$$G(s) = \frac{\mathbf{F}(s)}{s\mathbf{X}(s)}$$

via the relation

$$H(s) = \frac{b - G(s)}{b + G(s)} \tag{4.11}$$

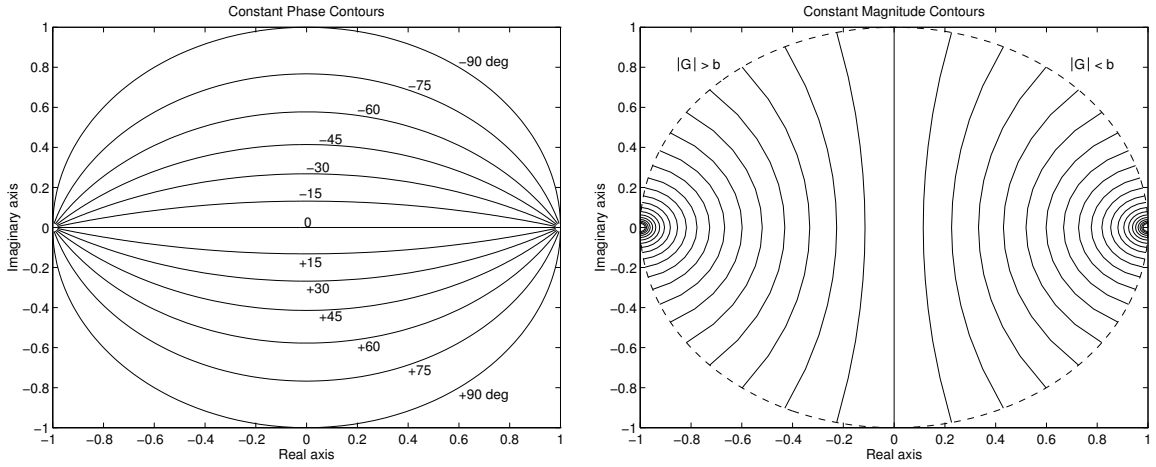
which we can also invert to provide

$$G(s) = b\,\frac{1 - H(s)}{1 + H(s)} \tag{4.12}$$

This establishes a conformal map between the entire right half $G(s)$ plane and the unit circle for $H(s)$. As such it reassures us that the two passivity conditions

$$\mathrm{Re}\,(G(s)) \geq 0 \quad \Leftrightarrow \quad |H(s)| \leq 1 \qquad \text{for all} \quad \mathrm{Re}(s) \geq 0$$

are consistent. But it also provides another interpretation for the graph of $H(j\omega)$. It can be constructed by mapping the Nyquist plot of $G(j\omega)$. Contours of constant phase and magnitude of $G(j\omega)$ are shown in Figure 4-8. They reaffirm the previous



(a) (b)

Figure 4-8: The wave frequency response $H(j\omega)$ is a mapped version of the Nyquist plot of $G(j\omega)$. Shown here are contour lines of (a) constant phase $\angle G(j\omega)$ and (b) constant magnitude $|G(j\omega)|$.

discussion. Large values of $|G(j\omega)|$ provide large forces and move to the left closer to the (-1) point. Phase lag in $G(j\omega)$ implies spring like behavior and appears in the upper half $H(s)$ plane.

In general, the description of a system in wave space provides similar information from a different perspective compared to the classic power variable approach. We encounter more examples in the next section looking at particular physical elements. But perhaps this discussion and perspective become more important in later chapters, when we develop elements based entirely within the wave space, for example wave filters. These may not have a clear mechanical interpretation, making information gained from the wave transfer function more valuable.

4.6 Classic Elements in Wave Space

To further illustrate the meaning and behavior of wave variables and systems, let us examine some basic elements in the wave space. In particular, we consider the spectrum from rigid wall, inertia, damper, spring, to free space.

As these elements are all linear, we present their transfer functions $H(s)$ and step responses. The results are summarized in Table 4.2 and Figure 4-9 respectively.

Individual Element	Governing Equation	Wave Transfer Function
Rigid Wall	$\dot{\mathbf{x}} = 0$	$H(s) = -1$
Inertia	$\mathbf{F} = m\ddot{\mathbf{x}}$	$H(s) = -\dfrac{s-\lambda}{s+\lambda} \qquad \lambda = \dfrac{b}{m}$
Damping	$\mathbf{F} = B\dot{\mathbf{x}}$	$H(s) = \text{const} = \dfrac{b-B}{b+B}$
Spring	$\mathbf{F} = K\mathbf{x}$	$H(s) = \dfrac{s-\lambda}{s+\lambda} \qquad \lambda = \dfrac{K}{b}$
Free Space	$\mathbf{F} = 0$	$H(s) = 1$

Table 4.2: The transfer functions for the basic individual elements are given.

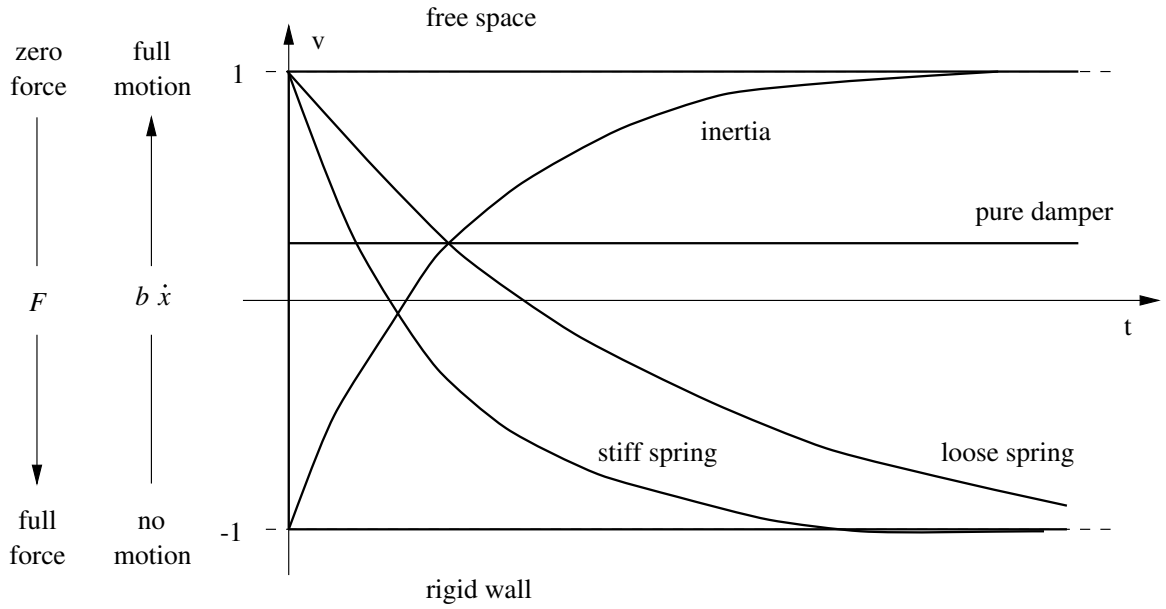


Figure 4-9: Step responses of individual elements in the wave domain. The magnitude is bounded by ± 1, while the sign determines the type of action: motion or force.

4.6.1 Rigid Wall and Free Space

First consider the extreme cases of a rigid wall and free space. As we saw above, they have transfer functions of -1 and $+1$ respectively.

$$H_{\text{wall}}(s) = -1 \qquad H_{\text{free}}(s) = 1 \tag{4.13}$$

This already captures the essence of wave responses. The unit values imply no energy losses, while the signs determine the action. Positive for motion-dominated systems, negative for force-dominated elements.

Also notice that these extremes are handled without any special treatment or causality constraints, simplifying the analysis in the wave domain.

4.6.2 Pure Spring

Next consider the response of a pure spring, which is governed by

$$\mathbf{F} = K\mathbf{x}$$

If we substitute this into the original wave definitions (4.2) and convert to Laplace domain, we see that

$$\mathbf{u} = \frac{b\dot{\mathbf{x}} + K\mathbf{x}}{\sqrt{2b}} \qquad \Rightarrow \qquad \mathbf{U}(s) = \frac{bs + K}{\sqrt{2b}}\,\mathbf{X}(s)$$

$$\mathbf{v} = \frac{b\dot{\mathbf{x}} - K\mathbf{x}}{\sqrt{2b}} \qquad \Rightarrow \qquad \mathbf{V}(s) = \frac{bs - K}{\sqrt{2b}}\,\mathbf{X}(s)$$

And therefore we find the wave transfer function

$$H_{\text{spring}}(s) = \frac{\mathbf{V}(s)}{\mathbf{U}(s)} = \frac{bs - K}{bs + K} = \frac{s - \lambda}{s + \lambda} \tag{4.14}$$

where the $\lambda = \frac{K}{b}$ is the bandwidth.

Correspondingly, the response to a step input $\mathbf{u}(t) = \mathbf{1}(t)$ is

$$\mathbf{v}(t) = -1 + 2e^{-\lambda t} \tag{4.15}$$

Initially, the response jumps to $\mathbf{v}(0) = 1$ and is perfectly equal to the input, much like free space. This because the initial push into the spring sees very little forces but large motion. Then as the spring begins to compress, the motion is limited and

forces are applied. Consequently, the returning wave goes negative and eventually approaches a limit of $\mathbf{v}(\infty) = -1$. So at steady state pushing into a spring is the same as pushing into a wall, neither will move any further.

In essence, a spring changes from motion-dominated to force-dominated as it compresses. So the wave response varies from positive to negative.

Also note the effects of the wave impedance b. In controlling the relative importance of force and velocity, it determines when the wave response changes sign. But also, it changes the bandwidth of the response and can make the spring more or less important in a closed loop system.

4.6.3 Inertia

The inertia forms the opposite element to the spring. Governed by

$$\mathbf{F} = m\ddot{\mathbf{x}}$$

we find the transfer function

$$H_{\text{inertia}}(s) = \frac{\mathbf{V}(s)}{\mathbf{U}(s)} = \frac{bs - ms^2}{bs + ms^2} = -\frac{s - \lambda}{s + \lambda} \tag{4.16}$$

where the the bandwidth is now $\lambda = \frac{b}{m}$, as well as the step response is

$$\mathbf{v}(t) = 1 - 2e^{-\lambda t} \tag{4.17}$$

So the response of an inertia is the exact negative of a spring. Indeed the physical behavior is initially dominated by forces as the mass accelerates, then my motion in steady state. Even the effect of the impedance b is opposite, i.e. it is now proportional to λ rather than inversely proportional.

4.6.4 Damper

Finally in the center of this force $\leftrightarrow$ motion spectrum, we find the damper, which is simply described by

$$\mathbf{F} = B\dot{\mathbf{x}}$$

The transfer function does not contain any dynamics, but is computed as

$$H_{\text{damper}}(s) = \frac{\mathbf{V}(s)}{\mathbf{U}(s)} = \frac{b - B}{b + B} \tag{4.18}$$

so that the step response is also a step of a reduced magnitude.

Here notice that the amount of damping will determine the sign of the response. Heavy damping implies a negative response and makes the element feel more like a wall. Light damping moves it closer to the other extreme of free space. Of course the impedance b again controls this trade off.

This is also the only element that does not reach the ± 1 bound. This is simply due to the dissipative nature, which is unique among this group.

4.7 Applications

We conclude with some remarks concerning the applicability of wave variables. Remember they are a purely algebraic transformation and so we can describe any system in wave space. However, the more interesting question asks for what systems can we find a passive representation to make effective use of these tools.

This is the case for all systems which are inherently passive. Applying the wave transformation to the power input-output variables maintains their passivity. They include most mechanical systems, like robots etc.

If a system is not passive, its wave description and scattering operator show the power gains. Especially in linear systems, it may be possible to cascade a power reduction element to bring these gains below unity. This essentially places sufficient damping in the closed loop system to force global passivity and counteract any energy production.

But let us take a closer look at delayed system, which constitute the primary objective behind this work. A delay is passive when it occurs in wave variables. But it is not passive when it effects power variables. Hence we must assure the proper description such that only wave variables are passed though delay elements.

In the case of teleoperation, all delays are isolated in the data transmission and we can select the data structure. Furthermore we have computer control on both sides of the delay. So the delay effectively occurs in software and we can enforce the correct layout. In practice, we need to implement a wave transformation at both sides of the delay in order to encode the information appropriately. Essentially, we wrap the delay element between two transformations to achieve passivity.

So what happens when the delay is not in software? We still require a transformation on both sides. But these transformations implicitly add a local feedback loop with an apparent damping element equal to the wave impedance (compare Figure 4-1). Without a local computer, we may not be able to generate this local feedback loop. Therefore we cannot implement the transformation and cannot achieve passivity. So systems with hardwired delays may pose problems.

Yet in some cases we can use the physical dynamics to achieve this apparent damping. For example consider the system

$$m\ddot{x} + b\dot{x} = F(t-T)$$

Clearly it is not passive. Also the delay is hardwired, i.e. we cannot control $\dot{x}$ locally without delay. However, the system already has a damping element. We can use this element for our needs.

Instead of including the damping in the system description, we can consider it a lossless and interpret the damping as part of the wave transformation. The energy dissipated therein drives the wave output confirming that no power is generated. So the above example is equivalent to a lossless inertia plus a wave delay.

So in cases with physical damping, we may be able to divert those elements to effectively implement the needed wave transformations and allow the use of the available passivity tools.

Finally we note that, besides forcing passive delays, the wave representation also provides a new perspective on a system and may suggest procedures and methods not typically considered. For example, we discuss filters and predictors in later chapters.

Chapter 5

Wave Based Teleoperation

Previously in Chapter 3 we examined traditional approaches to force-reflecting teleoperation in the presence of time delays and found limited regions of stability. Following work of (Anderson and Spong, 1989b) we traced the instability to the non-passive nature of the communications and came upon a modification to the standard control law that will compensate for this.

We now continue this development utilizing the wave variables just defined. By encoding and transmitting information in this fashion, the system is stabilized regardless of the delay and even without knowledge of the delay. The inherent symmetry of wave variables also increases the flexibility of selecting various configurations for different purposes.

After presenting the new configuration and discussing its implementation, we proceed to examine its characteristics and behavior. This in turn leads to tuning and design methods intended to optimize the achievable performance.

5.1 Wave Communications

As we noticed in Section 4.2, passivity of wave variables is robust to time delays. Indeed wave variables are constructed to achieve this very goal. Also, in the case of teleoperation, we know the time delays are isolated within the communications between the master and slave location.

This suggests performing the communications directly in wave domain. Figure 5-1 shows such an arrangement. As always we denote the master variables at the local site with subscript 'm' and the slave variables at the remote location with 's'.

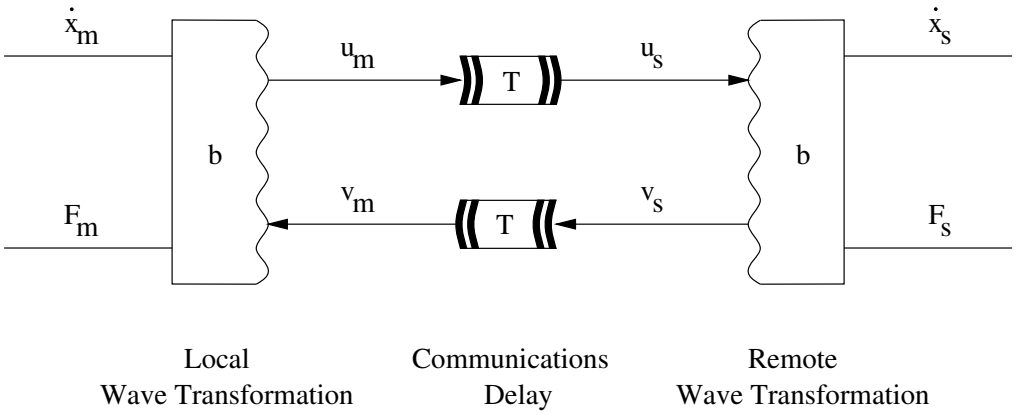


Figure 5-1: The wave based communications transforms both local and remote information into wave variables before transmission to the other side.

5.1.1 Definition

At both sites the information, i.e. the velocity $\dot{\mathbf{x}}$ and force $\mathbf{F}$, is transformed into wave domain via an appropriate transformation. Only the wave variables themselves are then transmitted. In particular, we continue to use left to right as the major direction, so that the right moving wave is transmitted from the master to the slave and vice versa.

The equations governing the transmission are simply

$$\mathbf{u}_s(t) = \mathbf{u}_m(t-T) \tag{5.1a}$$
$$\mathbf{v}_m(t) = \mathbf{v}_s(t-T) \tag{5.1b}$$

while, based on the transformation, the input waves are given as

$$\mathbf{u}_m(t) = \frac{b\dot{\mathbf{x}}_m(t) + \mathbf{F}_m(t)}{\sqrt{2b}} \tag{5.2a}$$

$$\mathbf{v}_s(t) = \frac{b\dot{\mathbf{x}}_s(t) - \mathbf{F}_s(t)}{\sqrt{2b}} \tag{5.2b}$$

Notice this does not specify nor restrict whether force or velocity are considered the input or output at either side. Indeed any combination is possible and we later use this flexibility to design alternate configurations, reminiscent of the selection made in Section 3.2. In essence, the communications acts as both impedance and admittance, depending on the surrounding elements. For now we give both possible

output equations on both sides.

$$\dot{\mathbf{x}}_m(t) = \sqrt{\frac{2}{b}}\,\mathbf{v}_m(t) + \frac{1}{b}\,\mathbf{F}_m(t) \qquad \text{or} \qquad \mathbf{F}_m(t) = b\dot{\mathbf{x}}_m(t) - \sqrt{2b}\,\mathbf{v}_m(t) \tag{5.3a}$$

$$\dot{\mathbf{x}}_s(t) = \sqrt{\frac{2}{b}}\,\mathbf{u}_s(t) - \frac{1}{b}\,\mathbf{F}_s(t) \qquad \text{or} \qquad \mathbf{F}_s(t) = -b\dot{\mathbf{x}}_s(t) + \sqrt{2b}\,\mathbf{u}_s(t) \tag{5.3b}$$

Of course, when the delay time T reduces to zeros, these definitions provide a simple identity between master and slave. That is, the element completely vanishes.

5.1.2 Passivity

Though this element is designed to be passive, we verify this in the following. To do so, we examine the overall power input P_{in} entering the communications.

$$P_{in} = \dot{\mathbf{x}}_m^T\mathbf{F}_m - \dot{\mathbf{x}}_s^T\mathbf{F}_s \tag{5.4}$$

Remember from Section 2.1.3 that we define power entering the left side as positive, as well as power exiting the right side.

Substituting the transformation equations, or simply using the original wave definition (4.3), we can also compute this power input as

$$P_{in} = \tfrac{1}{2}\mathbf{u}_m^T\mathbf{u}_m - \tfrac{1}{2}\mathbf{v}_m^T\mathbf{v}_m - \tfrac{1}{2}\mathbf{u}_s^T\mathbf{u}_s + \tfrac{1}{2}\mathbf{v}_s^T\mathbf{v}_s \tag{5.5}$$

where all variables are measured at the current time t.

We now substitute (5.1), which leaves us with

$$P_{in}(t) = \tfrac{1}{2}\mathbf{u}_m^T(t)\mathbf{u}_m(t) - \tfrac{1}{2}\mathbf{u}_m^T(t-T)\mathbf{u}_m(t-T) + \tfrac{1}{2}\mathbf{v}_s^T(t)\mathbf{v}_s(t) - \tfrac{1}{2}\mathbf{v}_s^T(t-T)\mathbf{v}_s(t-T)$$

and integrate to check passivity. We find that all input power is stored

$$\int_0^t P_{in}\,\mathrm{d}\tau = E_{store}(t) \geq 0$$

where the initial storage $E_{store}(0)$ and power dissipation P_{diss} are both zero, assuming that the initial conditions are also zero. The storage function itself is

$$E_{store}(t) = \int_{t-T}^t \tfrac{1}{2}\mathbf{u}_m^T\mathbf{u}_m + \tfrac{1}{2}\mathbf{v}_s^T\mathbf{v}_s\,\mathrm{d}\tau \geq 0 \tag{5.6}$$

which temporarily stores the wave energy while they are in transit.

So the communications is not only passive, but also lossless. This is true, regardless of which output equation (5.3) we select and independent of the delay time T. Furthermore it does not require knowledge of the delay nor equal delays in forward and reverse paths.

The only restriction we must place on the system is that the delay remain constant. This, however, will also be removed in later sections.

5.1.3 Spring-Like Characteristics

We noted that communications may output either force or velocity. For now, let us examine the case of both sides accepting velocity and dictating force. Thus the communications takes the role of a general impedance or a spring like object.

In particular, we are interested in the overall spring constant K_{comm} that this impedance will exhibit. In steady state we expect to find a relation of the form

$$\mathbf{F}_m = \mathbf{F}_s = K_{comm}\,\Delta\mathbf{x} = K_{comm}\,(\mathbf{x}_m - \mathbf{x}_s) \tag{5.7}$$

First we compute the deflection or position error between the two sides. From (4.8) we have

$$\Delta\mathbf{x}(t) = \mathbf{x}_m(t) - \mathbf{x}_s(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_m(\tau) + \mathbf{v}_m(\tau) - \mathbf{u}_s(\tau) - \mathbf{v}_s(\tau)\,\mathrm{d}\tau$$

After substituting the transmission equations (5.1), we get

$$\Delta\mathbf{x}(t) = \frac{1}{\sqrt{2b}} \int_{t-T}^t \mathbf{u}_m(\tau) - \mathbf{v}_s(\tau)\,\mathrm{d}\tau \tag{5.8}$$

Now in steady state we assume both sides have reached a constant position, making their velocities zero.

$$\dot{\mathbf{x}}_m = \dot{\mathbf{x}}_s = 0$$

Also, both sides should be applying some force to the system. The wave variables are hence computed as

$$\mathbf{u}_m = -\mathbf{v}_m = \frac{1}{\sqrt{2b}}\,\mathbf{F}_m \qquad\qquad \mathbf{u}_s = -\mathbf{v}_s = \frac{1}{\sqrt{2b}}\,\mathbf{F}_s$$

But having reached a steady value, the delay has no effect and the master and slave

waves should be identical at some constant value

$$\mathbf{u}_s = \mathbf{u}_m = \mathbf{u} \qquad \mathbf{v}_m = \mathbf{v}_s = \mathbf{v}$$

As a first conclusion, we see that both master and slave forces must also be equal, as was expected.

$$\mathbf{F}_m = \mathbf{F}_s = \sqrt{2b}\,\mathbf{u} \tag{5.9}$$

Second, now that the waves have become constant, we can further simplify the deflection to

$$\Delta\mathbf{x} = \frac{T}{\sqrt{2b}}\left(\mathbf{u}_m - \mathbf{v}_s\right) = \frac{2T}{\sqrt{2b}}\,\mathbf{u} \tag{5.10}$$

And finally, we can relate the force directly to the deflection as

$$\mathbf{F}_m = \mathbf{F}_s = \frac{b}{T}\,\Delta\mathbf{x} \tag{5.11}$$

So the steady state master and slave force output will be proportional to the deflection between the two, making the communications truly appear like a spring. The spring constant furthermore depends on both the wave impedance and the delay time

$$K_{comm} = \frac{b}{T} \tag{5.12}$$

Of course, this does not imply an immediate stiffness to the input. Naturally such signals only appear after the waves have completed at least one round trip of $2T$. It is only intended to demonstrate the steady state proportionality.

Also notice that for $T = 0$ the spring constant becomes infinite, so that both sides are rigidly connected and the communications, including the transformations, have no effect.

5.1.4 Admittance Characteristics

We may also select velocities and positions as our output from the wave communications. In such a case, the element assumes the role of a general admittance or inertia like object.

A similar derivation as seen above leads us to the conclusion that both the master and slave velocities will take a steady state value proportional to the input momentum,

i.e. the integral of the input forces.

$$\dot{\mathbf{x}}_m = \dot{\mathbf{x}}_s = \frac{1}{M_{comm}} \int_0^t \mathbf{F}_m - \mathbf{F}_s \, \mathrm{d}\tau \tag{5.13}$$

where the mass is given by

$$M_{comm} = bT \tag{5.14}$$

In particular, this input momentum is computed as

$$\int_0^t \mathbf{F}_m - \mathbf{F}_s \, \mathrm{d}\tau = \sqrt{\frac{b}{2}} \int_{t-T}^t \mathbf{u}_m(\tau) + \mathbf{v}_s(\tau) \, \mathrm{d}\tau$$

Plus in steady state we assume no further input forces, so that both right and left moving waves take the same constant value

$$\mathbf{u}_m = \mathbf{u}_s = \mathbf{v}_m = \mathbf{v}_s = \sqrt{\frac{b}{2}} \, \dot{\mathbf{x}}$$

So like before, the steady state behavior is equivalent to a normal inertia, though this does not imply immediate inertial feedback forces. And again the parameter depends on both the wave impedance and the delay time. As the delay reduces to zeros, the element has no inertia and becomes completely transparent.

An additional important point to make is that both master and slave position will track each other perfectly (in steady state). While the input forces may be applied to one side only, and this side may lead the other for a short time, eventually both sides will follow the same path at the same time with no tracking error. This can be quickly verified by computing the actual tracking error from (5.8).

In general, the wave communications will automatically take the role of either impedance or admittance as required by the rest of the system. Indeed this selection occurs during the actual execution, when the environment may provide a stiff contact or no contact at all. And so we the system becomes very flexible in dealing with all types of interactions.

5.2 Simple Wave Teleoperator

We now turn our attention to incorporating these transmissions into the entire teleoperator system. A general layout is shown in Figure 5-2. Both master and slave robots are controlled locally in power variables. This assumes measurements of po-

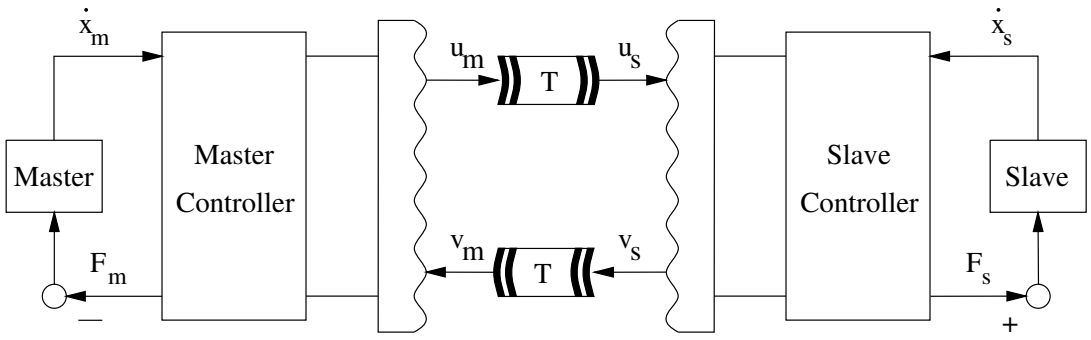


Figure 5-2: In the general layout, a teleoperator may have a separate controller for both master and slave manipulator. But on both sides, the information is transformed into wave variables, which are delayed in the transmission.

sition and velocity, while applying forces to the manipulators. Before transmission, the information is transformed into wave space, so that all delays are incorporated passively.

5.2.1 Basic Teleoperator Layout

Notice that we have not yet specified whether or which of the controllers track a desired position and velocity or regulate to a desired force. We will use this flexibility in later sections to derive alternate layouts. For the basic layout, we now assume that the slave follows a desired motion and is placed under P.D. control. Meanwhile, the master applies the desired force directly to the joystick. As such, this layout is closest to the traditional delayed system with position feedforward and force feedback first discussed in Section 3.2. A block diagram of the transfer functions for a single degree of freedom system is depicted in Figure 5-3

The slave P.D. controller uses constant symmetric positive definite matrices for both the velocity gain B and position gain K to force the slave to track the desired velocity $\dot{\mathbf{x}}_{sd}$. The necessary force is

$$\mathbf{F}_s = -B(\dot{\mathbf{x}}_s - \dot{\mathbf{x}}_{sd}) - K(\mathbf{x}_s - \mathbf{x}_{sd}) \tag{5.15}$$

The desired velocity is decoded from the wave transformation via (5.3b) as

$$\dot{\mathbf{x}}_{sd} = \frac{\sqrt{2b}\,\mathbf{u}_s - \mathbf{F}_s}{b} \tag{5.16}$$

and the return wave is computed by

$$\mathbf{v}_s = \frac{b\dot{\mathbf{x}}_{sd} - \mathbf{F}_s}{\sqrt{2b}} = \mathbf{u}_s - \sqrt{\frac{2}{b}}\,\mathbf{F}_s \tag{5.17}$$

Notice the combination of P.D. controller (5.15) and wave transformation (5.16) creates an algebraic loop. The slave force depends on the desired velocity, which in turn depends on the current force. This loop is quickly solved by writing

$$\dot{\mathbf{x}}_{sd} = \frac{\sqrt{2b}\,\mathbf{u}_s + B\dot{\mathbf{x}}_s + K(\mathbf{x}_s - \mathbf{x}_{sd})}{B + b}$$

Also, the actual desired position must be computed from this velocity.

$$\mathbf{x}_{sd} = \int_0^t \dot{\mathbf{x}}_{sd}(\tau)\,\mathrm{d}\tau \tag{5.18}$$

Any numerical errors may cause the desired position to drift and are addressed below in Section 5.3.

On the master side, the desired force is applied to the joystick without additional changes. The value is determined by the local transformation via (5.3a) as

$$\mathbf{F}_m = b\dot{\mathbf{x}}_m - \sqrt{2b}\,\mathbf{v}_m \tag{5.19}$$

Finally, the right moving wave is computed from

$$\mathbf{u}_m = \frac{b\dot{\mathbf{x}}_m + \mathbf{F}_m}{\sqrt{2b}} = \sqrt{2b}\,\dot{\mathbf{x}}_m - \mathbf{v}_m \tag{5.20}$$

We should also point out that all the elements are passive and so this system is stable, regardless of the delay T. Indeed the value need not even be known, as it is not used anywhere in the controller. And for small delays, the system reverts to a simple P.D. connection between master and slave robots.

5.2.2 Feedback Paths

This simple teleoperator layout contains multiple internal loops, in particular introduced by the wave transformation. Remember from Section 4.1 that each wave transformation contains an apparent damping element. So we find three distinct paths that may carry signals back to the joystick and the operator. They are sketched within

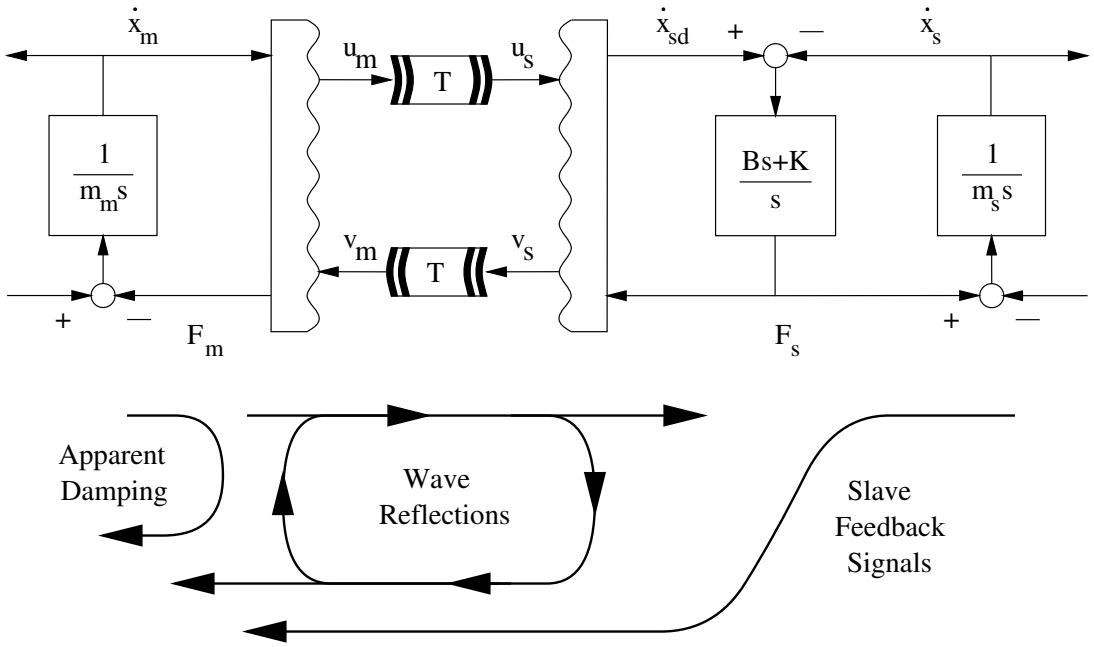


Figure 5-3: The basic wave teleoperator places the slave robot under P.D. control while applying forces directly to the master. The wave transformations both encode information for transmission and decode the incoming waves to compute the desired values. The resulting structure provides three separate paths through which signals may travel to the operator.

Figure 5-3.

First, for every motion of the master manipulator, there is an immediate feedback in the form of damping created in the wave transformation. This energy is not dissipated, but rather used to construct the wave signal and transmitted to the remote location. Nevertheless, this feedback appears to the operator like a simple damper. Notice it is clearly visible in (5.19). Also, it is the only feedback that does not travel through the delay.

The second path contains and is based on wave reflections at the both transformations. When the right moving wave $\mathbf{u}_s$ reaches its destination, part of the signal may return with the left moving wave $\mathbf{v}_s$ back towards the operator. In (5.17) we see this effect. But when the returning wave $\mathbf{v}_m$ arrives at the local transformation, it may again be reflected into the forward path as $\mathbf{u}_m$. Compare to (5.20).

So not only do wave reflections contain little useful information, but the cyclic layout may allow them to last for several cycles before dying out. They easily create unexpected disturbances and distractions and can cause mayor problems. It is this feedback path that we need to eliminate and which is addressed in more detail in Section 5.4.

The third and final path moves signals from the remote robot via the P.D. controller back to the operator. This is the main feedback path and provides the information needed to complete the tasks. Whether motion occurred or contact was made is described by the slave position and velocity and appears in the tracking error. The slave force encodes this information and provides a close representation of the actual contact forces. Its feedback to the operator is the main goal of the system.

Of the three feedback components, the last contains the important data and must be protected. The first creates appropriate damping and is easily tolerable if not even beneficial. But the second, being the wave reflections, may cause significant problems and overshadow the others. It is the aim of the alternate layouts introduced and described below to eliminate or at least strongly reduce this undesired part.

After experimentally demonstrating and verifying the predicted behavior, we address the remaining problems in the following sections. In particular, these are the explicit position feedback and the elimination of wave reflections. This leads to a new teleoperator configuration, which we present and analyze in Section 5.5.

5.2.3 System behavior

The three feedback paths can lead to various behaviors for the teleoperator. First note that for small delays the system is quite transparent. When the delay is below

the human reaction time, the operator will not be able to distinguish the types of feedback. In particular, the returning wave will bring back power quickly, canceling the effects of the apparent damping. Also while the reflections may survive several cycles, these are quite short and the signals die out quickly. So the operator is left with the desired slave feedback.

This transparent behavior has been verified in experiments and is shown in Figure 5-4. The system is not tuned in any special fashion and the gains are similar to an undelayed system. The response shows nice tracking and no unexpected phenomena in both position during a horizontal move (in x, y for $t < 1.5$) and force during a vertical contact (in z for $t > 1.5$). Note the large force signals in x, y during the motion correspond to acceleration and deceleration forces, which the operator applies to the master joystick. And the minor residual position errors in x, y during the contact are caused by friction at the contact point.

For larger delays, however, the operator can distinguish between signals coming from the immediate feedback, i.e. the apparent damping, and the delayed feedback. When wave reflections occur in addition to the desired slave signals, the system can become quite oscillatory.

Indeed experiments show strong reflections that continue for several cycles until the system comes to rest. Compare to Figure 5-5, in which the operator initiates a brief input or disturbance. Throughout the remainder of the test no contact is made with either master or slave robot. Nevertheless, wave signals continue to bounce back and forth and create disturbances at multiples of the delay time $T = 1.0$. In practice such a system is useless without corrective tuning or modification. The slow vertical drift (in z) is the result of small gravity forces.

Consider also the characteristic equation for a single degree of freedom.

$$\frac{b - m_m s}{b + m_m s} \, \frac{b(m_s s^2 + Bs + K) - (Bs + K)m_s s}{b(m_s s^2 + Bs + K) + (Bs + K)m_s s} \, e^{-2sT} = 1$$

where the total round trip delay is $2T$. Its root locus with variable delay is shown in Figure 5-6. Notice how the dominant poles move closer to the imaginary axis and the system becomes more oscillatory.

5.3 Position Feedback

We have noted that the only data being transmitted between the two locations consists of wave signals and implicitly contains velocity and force information. So far we

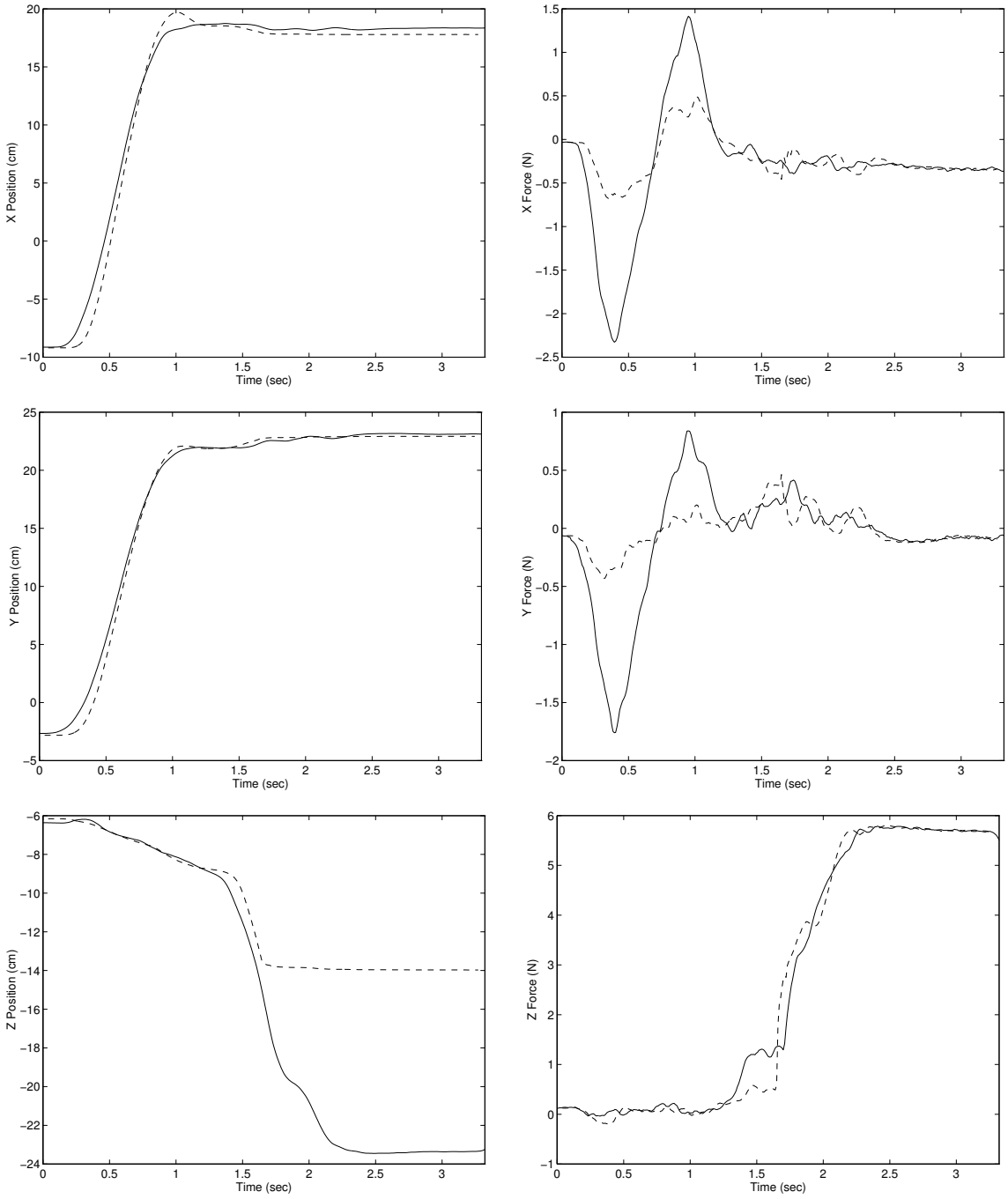


Figure 5-4: The total delay time of 0.1 seconds is comparable to the human reaction time and the system behaves almost transparently. Both position tracking during a horizontal motion and force tracking during a vertical contact is good.

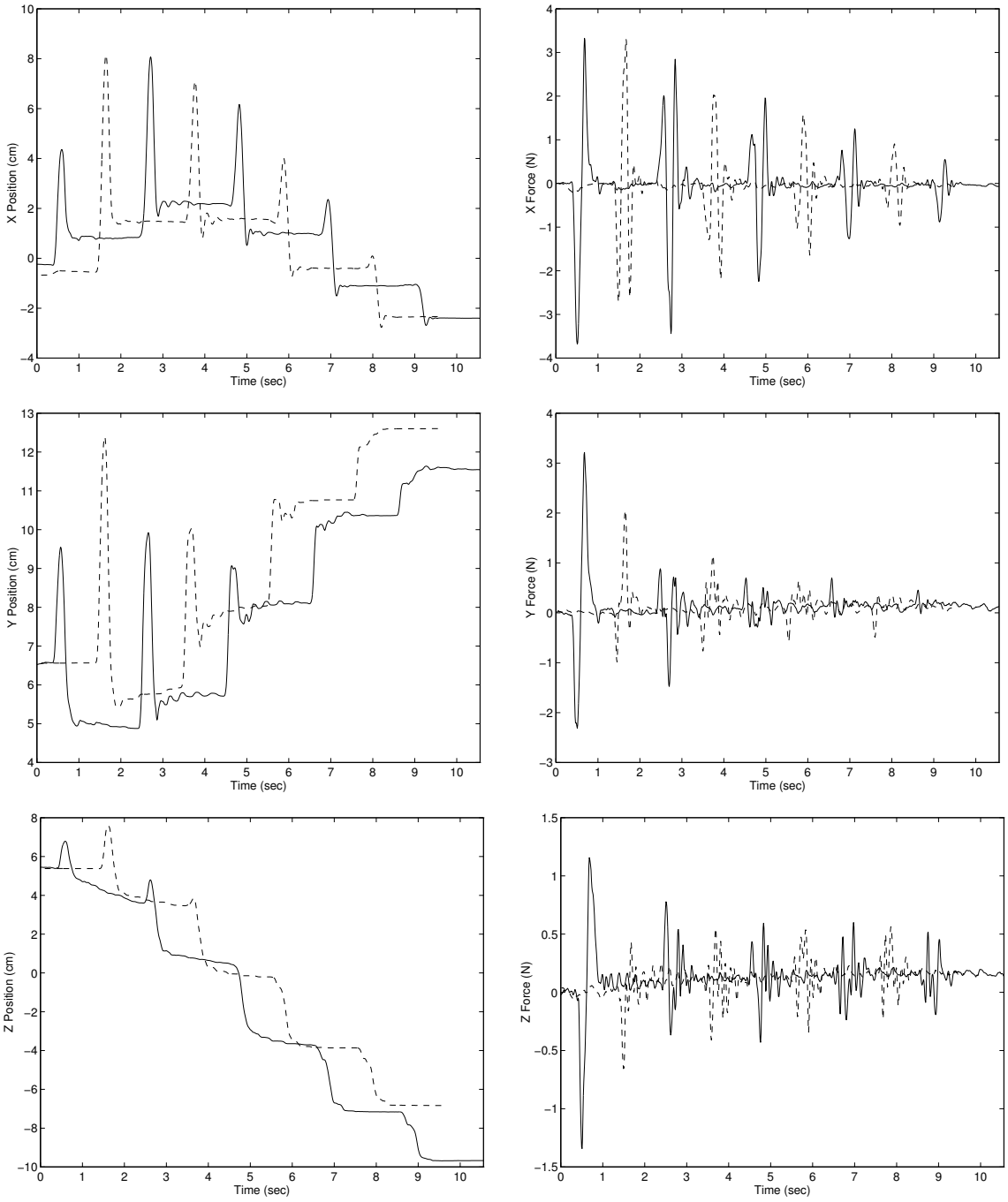


Figure 5-5: This experiment is initiated by a brief input or disturbance, after which both master and slave robots are untouched. The long response shows clear evidence of wave reflections that appear at multiples of the delay time of $T = 1.0$

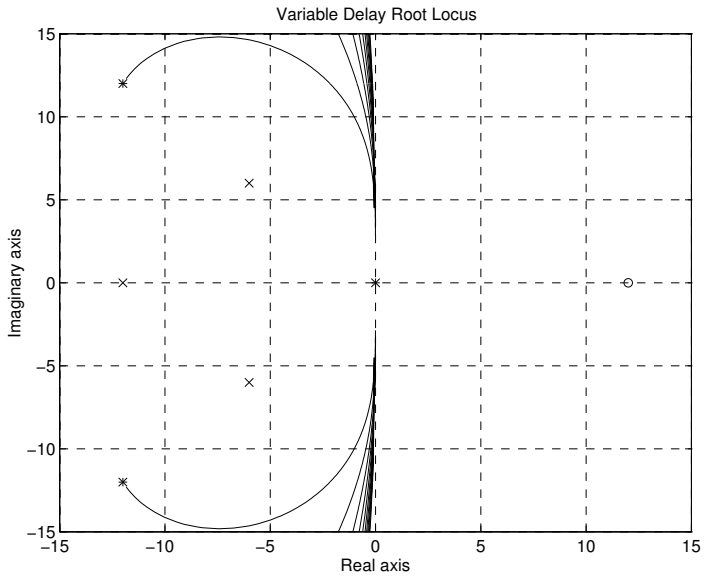


Figure 5-6: The variable delay root locus of the simple wave teleoperator shows the increasingly oscillatory poles.

have not seen any direct position feedback. Indeed the current configuration may be susceptible slow position drift between the two robots, if numerical errors or data loss occur in the transmission.

In the following we review the theoretical tracking which occurs under ideal conditions and its problems, then propose two solutions to provide absolute position feedback in a practical system. The first explicitly includes the integral of wave signals, the second adjusts the wave commands to guarantee tracking.

5.3.1 Theoretical Position Tracking

In theory the wave based teleoperator layout proposed in the previous section will force the master and slave robots to track each other. In steady state the error between both positions should reduce to zero. However, this relies on the numerical integration of desired slave velocity into position (5.18). In practice the master and slave may drift apart.

But first, consider ideal conditions without numerical errors. From (4.8) we see that both the master and the desired slave positions can be computed as

$$\mathbf{x}_m(t) \;=\; \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_m(\tau) + \mathbf{v}_m(\tau)\, \mathrm{d}\tau \tag{5.21a}$$

$$\mathbf{x}_{sd}(t) \;=\; \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_s(\tau) + \mathbf{v}_s(\tau)\, \mathrm{d}\tau \tag{5.21b}$$

Their difference forms the position error across the wave communications

$$\Delta\mathbf{x}_{comm}(t) = \mathbf{x}_m(t) - \mathbf{x}_{sd}(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_m(\tau) + \mathbf{v}_m(\tau) - \mathbf{u}_s(\tau) - \mathbf{v}_s(\tau)\, \mathrm{d}\tau \tag{5.22}$$

But the wave signals are delayed copies of each other, so that we find

$$\Delta\mathbf{x}_{comm}(t) = \frac{1}{\sqrt{2b}} \int_{t-T}^t \mathbf{u}_m(\tau) - \mathbf{v}_s(\tau)\, \mathrm{d}\tau \tag{5.23}$$

In steady state, when the wave signals have decayed to zero without any velocity or force inputs, this position error is zero. Hence the desired slave position exactly equals the master location. And without force inputs, the P.D. controller will guarantee that the actual slave position also converges there.

Unfortunately, this argument fails or is at least susceptible to errors between steps (5.22) and (5.23). Here we assume that

$$\int_0^{t-T} \mathbf{u}_m(\tau)\, \mathrm{d}\tau \equiv \int_0^t \mathbf{u}_s(\tau)\, \mathrm{d}\tau \qquad \text{and} \qquad \int_0^t \mathbf{v}_m(\tau)\, \mathrm{d}\tau \equiv \int_0^{t-T} \mathbf{v}_s(\tau)\, \mathrm{d}\tau \tag{5.24}$$

Moreover the resulting algorithm must actually compute desired slave position from (5.18) or (5.21b) as

$$\mathbf{x}_{sd}(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_s(\tau) + \mathbf{v}_s(\tau)\, \mathrm{d}\tau = \int_0^t \dot{\mathbf{x}}_{sd}(\tau)\, \mathrm{d}\tau \tag{5.25}$$

In practice this position may drift from its theoretical value and the system may see slow drift between the two robots for several reasons:

- **Discrete Sampling:** The algorithms are usually implemented on digital computers with finite sampling rates. The actual integration is necessarily replaced by a finite summation.
- **Numerical Errors:** The data is also represented with finite precision, so that the continual summation may accumulate the effect of many roundoff errors.
- **Data Loss:** Should the transmission ever lose any data, the integration will always be shifted by the corresponding amount.

But we should also point out that given todays digital computers, these errors are typically very small and problems do not surface for very long periods of operation.

Two solutions have presented themselves to this problem. First we may explicitly transmit the integral of the wave signals. This eliminates the need to assume (5.24) or compute (5.25). Or second, we may observe the resulting drift error and add a corrective term to the wave command. This has the added benefit of solving the initial condition problem.

5.3.2 Transmitting the Wave Integrals

Just as the wave signals encode velocity and force, their integrals encode position and momentum information. Computing and transmitting these values can provide explicit position information and prevent the above mentioned problems.

The integrated wave variables are defined as

$$\mathbf{U}(t) = \int_0^t \mathbf{u}\,\mathrm{d}\tau = \frac{b\mathbf{x}+\mathbf{p}}{\sqrt{2b}} \tag{5.26a}$$

$$\mathbf{V}(t) = \int_0^t \mathbf{v}\,\mathrm{d}\tau = \frac{b\mathbf{x}-\mathbf{p}}{\sqrt{2b}} \tag{5.26b}$$

and encode position $\mathbf{x}$ and momentum $\mathbf{p}$, which is the integral of force

$$\mathbf{p} = \int_0^t \mathbf{F}\,\mathrm{d}\tau \tag{5.27}$$

The position information can be measured directly without problems. However, the momentum data is not as easy to obtain and needs to be integrated numerically. In essence, we have shifted the problem from integrating velocity into position to integrating force into momentum. But while the same errors may still appear, we place little or no importance on the actual momentum value. Remember velocity signals are still contained in the original wave variables and do not depend on this new quantity. Plus we may actually circumvent the long term integration, as noted below.

The communications should transmit both wave and integrated wave signals. And to use this added information, the system should be extended by the following equations. On the slave side integrate the momentum, compute the desired position as well as the returning wave integral.

$$\mathbf{p}_s = \int_0^t \mathbf{F}_s\,\mathrm{d}\tau$$

$$\mathbf{x}_{sd} = \frac{\sqrt{2b}\,\mathbf{U}_s - \mathbf{p}_s}{b}$$

$$\mathbf{V}_s = \frac{b\mathbf{x}_{sd} - \mathbf{p}_s}{\sqrt{2b}}$$

On the master side, just determine the command wave integral from the actual position

$$\mathbf{U}_m = \sqrt{2b}\,\mathbf{x}_m - \mathbf{V}_m$$

A closer examination shows that this extension actually implements the following behavior

$$\mathbf{x}_{sd}(t) = 2\mathbf{x}_m(t-T) - \mathbf{x}_{sd}(t-2T) - \frac{1}{b}\int_{t-2T}^{t} \mathbf{F}_s(\tau)\,\mathrm{d}\tau \tag{5.28}$$

which we can also compute directly. In this alternate form the position information is transmitted directly and we need not use the momentum. But we do need to integrate the force $\mathbf{F}_s$ over a finite period of time.

Using either form the desired slave position is now based on the master position, preventing any drift. Also notice that these developments do not effect passivity. Indeed they only explicitly guarantee what theoretically should be true already. As such they do not interfere with any other elements of the system and are generally ignored during the following sections. If needed, this type of position feedback via the wave integrals can be added to any wave system.

5.3.3 Single Channel Transmissions

At first glance, transmitting the wave integral, as proposed above, would seem to require additional channels or bandwidth in the communications. In fact this is not the case, as we can combine a wave signal with its integral into a single quantity. Then after reception on the other side, the values can be separated again.

This coding/decoding process is depicted in Figure 5-7 for the right moving wave. It is accurate and numerically stable. In particular, add both values before transmission,

$$\bar{\mathbf{U}}_m = \mathbf{U}_m + \frac{1}{\lambda}\mathbf{u}_m$$

send,

$$\bar{\mathbf{U}}_s(t) = \bar{\mathbf{U}}_m(t-T)$$

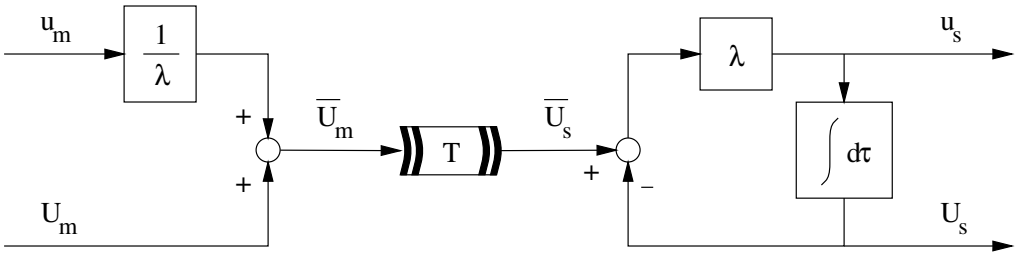


Figure 5-7: Transmitting the sum of both wave and integrated wave signals allows all information to flow through a single channel. The receiving side can decompose the signal with a numerically stable first order filter.

and separate via a stable first order filter

$$\mathbf{u}_s = \lambda(\bar{\mathbf{U}}_s - \mathbf{U}_s) \qquad \mathbf{U}_s = \int_0^t \mathbf{u}_s \, \mathrm{d}\tau$$

where λ is the positive constant bandwidth of the filter.

This makes use of the fact that the two values are closely related via integration or differentiation. The choice of the bandwidth λ is arbitrary and may be used to scale the average magnitudes of $\mathbf{U}$ and $\mathbf{u}$. It should, however, always be significantly less than sampling rate of the digital implementation.

In effect, the combination $\bar{\mathbf{U}}$ encodes all information required to operate the system into a single value: position, velocity and force. And the decoding and wave transformations extract the information as needed.

5.3.4 Adjusting the Wave Command

This second method for providing absolute position feedback observes the global drift error and adds a corrective term to the wave command. This can also help avoid problems of mismatched initial positions. If the two robots are started at unequal locations, the corrective feedback will slowly bring the manipulators together as operations begin. It does so without the sudden step inputs often associated with such large initial errors. Similarly, if the transmission is temporarily interrupted, the system will slowly return to normal without jumping upon reconnection.

Using the notation of Figure 5-8, we define the actual wave command $\mathbf{u}_m$ based on the uncorrected value $\hat{\mathbf{u}}_m$ and the corrective term $\Delta\mathbf{u}_m$ as

$$\mathbf{u}_m = \hat{\mathbf{u}}_m + \Delta\mathbf{u}_m \tag{5.29}$$

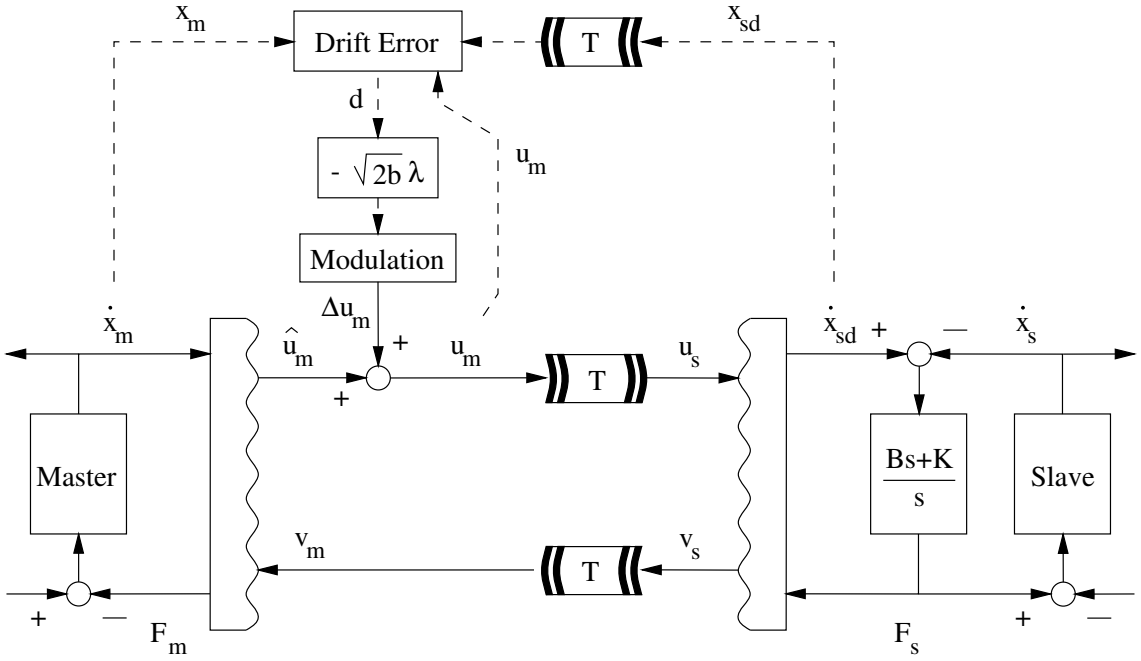


Figure 5-8: By adjusting the wave command, the system can provide absolute position feedback and guarantee tracking even with numerical errors and occasional data loss. In addition, initial position differences are slowly removed without sudden changes.

But before we proceed, let us first examine the position difference we expect to see between the two sides of the communications during normal operation.

$$\Delta\mathbf{x}(t) = \mathbf{x}_m(t) - \mathbf{x}_{sd}(t-T) \tag{5.30}$$

Given the dynamics of the delay, this value need not be zero exactly. Also, we compute this difference on the master side. So we use the delayed slave desired position to account for the transmission. Substituting the wave variables, we find

$$\begin{aligned}
\Delta\mathbf{x}(t) &= \frac{1}{\sqrt{2b}} \int_0^t \hat{\mathbf{u}}_m(\tau) + \mathbf{v}_m(\tau) - \mathbf{u}_s(\tau-T) - \mathbf{v}_s(\tau-T)\,\mathrm{d}\tau \\
&= \frac{1}{\sqrt{2b}} \int_0^t \hat{\mathbf{u}}_m(\tau) - \mathbf{u}_m(\tau-2T)\,\mathrm{d}\tau \\
&= \frac{1}{\sqrt{2b}} \int_{t-2T}^t \mathbf{u}_m(\tau)\,\mathrm{d}\tau \;-\; \frac{1}{\sqrt{2b}} \int_0^t \Delta\mathbf{u}_m(\tau)\,\mathrm{d}\tau
\end{aligned}$$

Under ideal conditions, we do not need to use the corrective term $\Delta\mathbf{u}_m$ and this

position difference is based only on the actual wave commands $\mathbf{u}_m$ during the past $2T$.

Now define the drift error $\mathbf{d}$ as

$$\mathbf{d} = \underbrace{\frac{1}{\sqrt{2b}} \int_{t-2T}^{t} \mathbf{u}_m(\tau)\, \mathrm{d}\tau}_{\text{expected } \Delta\mathbf{x}} - \underbrace{\Big[\mathbf{x}_m(t) - \mathbf{x}_{sd}(t-T)\Big]}_{\text{actual } \Delta\mathbf{x}} \tag{5.31}$$

which compares the actual position difference to its expected value based on ideal conditions. So without numerical errors or other imperfections, this drift error will be exactly zero.

But from the above calculations we see that the drift error $\mathbf{d}$ is also related to the corrective term $\Delta\mathbf{u}_m$ via

$$\mathbf{d} = \frac{1}{\sqrt{2b}} \int_{0}^{t} \Delta\mathbf{u}_m(\tau)\, \mathrm{d}\tau$$

or more importantly

$$\dot{\mathbf{d}} = \frac{1}{\sqrt{2b}}\, \Delta\mathbf{u}_m \tag{5.32}$$

and so we can adjust the wave command to change and reduce any existing drift errors. In particular, we would like to use

$$\Delta\mathbf{u}_m = -\sqrt{2b}\, \lambda\, \mathbf{d} \tag{5.33}$$

where λ is the bandwidth of this new feedback path. Any other value would also work as long as $\Delta\mathbf{u}_m$ and $\mathbf{d}$ are of opposite sign.

Unfortunately, this added feedback path does not contain any power source or any power flow. And so, as not to disturb the passivity criterion, the corrective term may not introduce any power into the system. Hence, the corrected wave command must be bound by the uncorrected (original) version.

$$\tfrac{1}{2}\mathbf{u}_m^T\mathbf{u}_m \leq \tfrac{1}{2}\hat{\mathbf{u}}_m^T\hat{\mathbf{u}}_m$$

In addition we do not want the final system to confuse the operator, so we require that both corrected and uncorrected wave commands are of the same sign (for each degree of freedom). Together this implies that the correction may only reduce the wave magnitude.

We satisfy these requirements by modulating the magnitude of the desired correc-

tion (5.33) in each degree of freedom according to

$$\Delta \mathbf{u}_m = \begin{cases} 0 & \text{if} \quad d\,\hat{u}_m < 0 \\ -\sqrt{2b}\,\lambda\, d & \text{if} \quad d\,\hat{u}_m > 0 \quad \text{and} \quad \sqrt{2b}\,\lambda\,|d| < |\hat{u}_m| \\ -\hat{u}_m & \text{if} \quad d\,\hat{u}_m > 0 \quad \text{and} \quad \sqrt{2b}\,\lambda\,|d| > |\hat{u}_m| \end{cases} \tag{5.34}$$

In the first case the correction is set to zero, because it wants to increase the magnitude of the original wave command in violation of passivity. And the last case is saturated, because the desired correction is too strong and would change direction of the original wave command.

Using this modulation we are assured of passivity and hence stability. Plus the corrective term $\Delta \mathbf{u}_m$ and the drift error $\mathbf{d}$ are never of the same sign, so that the drift error will never increase. Instead, it will decrease once the system detects appropriate wave commands. Note this 'waiting' behavior is a side effect of passivity, where the system cannot force tracking until it receives power input from the operator.

Figure 5-8 displays the procedure. The added feedback path is depicted by dashed lines, because it does not contain wave variables or any power flow. The new bandwidth λ may be quite small as the drift errors should be minimal. It is best chosen to remove initial position differences in an acceptable time frame.

Finally two remarks. First, the uncorrected wave signal $\hat{\mathbf{u}}_m$ is computed as before from (5.20) as

$$\hat{\mathbf{u}}_m = \frac{b\dot{\mathbf{x}}_m + \mathbf{F}_m}{\sqrt{2b}} = \sqrt{2b}\,\dot{\mathbf{x}}_m - \mathbf{v}_m$$

Second, once we begin adding elements in the wave domain, such as wave filters in Section 6.2, we will have to adjust the definition of the drift error (5.31). In particular we will need to update the expected position difference to account for the additional dynamics.

5.4 Wave Reflections and Impedance Matching

Among the various loops in the teleoperator system, we have noticed that wave transformations can reflect incoming wave signals. Such reflections carry little useful information and can easily distract the operator. Plus repeated reflections lead to oscillatory behavior, which may be compared to natural standing wave phenomena. In general wave reflections should be avoided.

The reflections occur when a wave signal hits an element or subsystem with an

impedance unequal to its own wave impedance. So, to reduce and possibly eliminate reflections, we try to match impedances between the various elements. In effect any subsystem should appear just like a pure damper of magnitude b and dissipate all incoming energy associated with a wave command. Without any energy no reflections can return.

This generally requires tuning for some expected operating condition and possibly modifications. In the teleoperator, both the slave P.D. controller and the master provide complex valued impedances that cannot directly match the transmission. While good tuning may lower the resulting reflections, the basic configuration cannot completely eliminate them. We proceed to examine both force and position controllers at both sides more closely and suggest appropriate changes.

5.4.1 Force Control

First, let us examine the master side, which is operating under force control in the basic configuration. The incoming wave signal $\mathbf{v}_m$ is interpreted as a force command $\mathbf{F}_m$ according to (5.19) and applied to the joystick. The position information $\mathbf{x}_m$ is used only to determine the outgoing wave command $\mathbf{u}_m$ and no additional dynamics are introduced.

Unfortunately, such a force controller cannot be perfectly impedance matched. Indeed the inertial effects will always be evident and prevent the master from appearing as a pure damper. But an added damping element may help reduce the reflections and is shown in Figure 5-9. The combined impedance is given as

$$\frac{\mathbf{F}_m(s)}{s\mathbf{X}_m(s)} = m_m s + D$$

The resulting wave transfer function is described by

$$\mathbf{U}_m(s) = \frac{\sqrt{2b}}{m_m s + D + b}\,\mathbf{F}_h(s) - \frac{m_m s + D - b}{m_m s + D + b}\,\mathbf{V}_m(s)$$

where $\mathbf{F}_h$ are human input forces generated by the operator. Clearly no choice of D eliminates the reflection term. However, using

$$D = b$$

can minimize the low frequency components. Nevertheless, the high frequency reflections always remain.

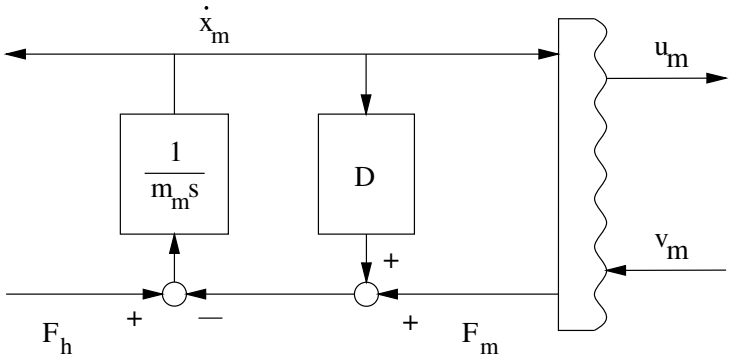


Figure 5-9: It is not possible to completely avoid reflections under force control, but a better impedance match can be achieved by adding a damping element to the master.

Using other damping elements, in particular an inverted element which takes a force input and velocity output, is not beneficial. Such an element would create a relative velocity to dissipate energy, which would force the master and slave positions apart.

So, in our simple wave teleoperator, it will never be possible to prevent reflections at the master side. Fortunately, reflections at one side need not turn into continual oscillations. But we do suggest using an improved layout, as described later.

5.4.2 Position and Velocity Control

In contrast to the above scenario, the slave uses a P.D. controller to regulate its position and velocity and thus adds dynamics in the form of a spring element. With appropriate adjustments, the inertial and capacitive effects can be made to counter balance each other. Hence P.D. controllers can be impedance matched perfectly.

In the basic layout, the slave presents a complex valued impedance of

$$\frac{\mathbf{F}_s(s)}{s\mathbf{X}_{sd}(s)} = \frac{m_s s(Bs+K)}{m_s s^2 + Bs + K}$$

to the incoming wave signals. Selecting the gains as

$$B = b \qquad K = \frac{b^2}{m_s}$$

will provide the best match to the wave impedance b and hence the least reflections. But it cannot eliminate the low frequency mismatch and completely restricts the

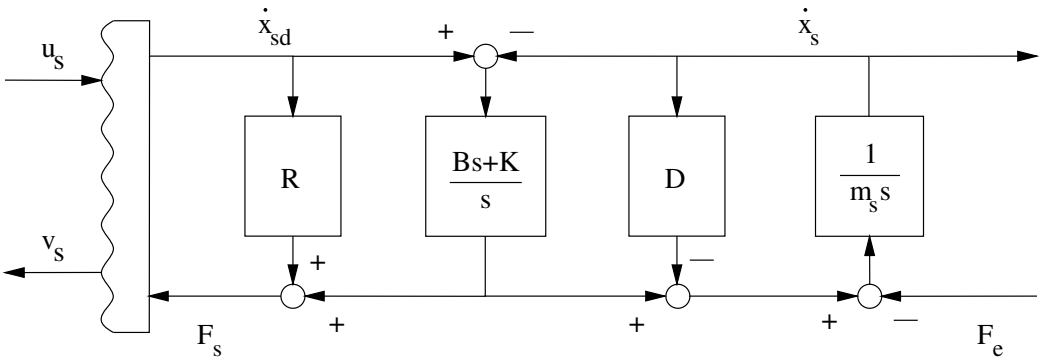


Figure 5-10: In order to avoid reflections and match the impedance at the wave transformation, two additional damping elements are inserted into the P.D. controller.

choice of gains.

Instead, to perfectly match the impedance between the P.D. controller and the wave transmission, we propose using a layout as shown in Figure 5-10. Two damping elements are added to the standard configuration. The resulting slave force, which is seen at the transformation and eventually feed back the master, is given in the Laplace domain by

$$\mathbf{F}_s(s) = \frac{Bs+K}{m_s s^2 + Ds + Bs + K}\,\mathbf{F}_e(s) + \left[\frac{(m_s s + D)(Bs+K)}{m_s s^2 + Ds + Bs + K} + R\right] s\mathbf{X}_{sd}(s)$$

where $\mathbf{F}_e$ is the environment contact force.

Furthermore, if we select the gains as

$$D = \lambda_s m_s \qquad B = \lambda_s m_s \qquad K = {\lambda_s}^2 m_s \qquad R = b - B \tag{5.35}$$

the equation simplifies to

$$\mathbf{F}_s(s) = \frac{\lambda_s}{s+\lambda_s}\,\mathbf{F}_e(s) + b\,s\mathbf{X}_{sd}(s)$$

which we can substitute into the wave transformation to find

$$\mathbf{V}_s(s) = -\frac{1}{\sqrt{2b}}\,\frac{\lambda_s}{s+\lambda_s}\,\mathbf{F}_e(s) \tag{5.36}$$

Notice that the returning wave is now a pure indication of environmental interaction forces. All reflections have been eliminated. The remaining parameters are the slave bandwidth λ_s and the wave impedance b itself. As R must be positive, the impedance

b is lower bounded.

Also, using these gains the actual slave position is related to the incoming wave via

$$s\mathbf{X}_s(s) = \frac{\lambda_s}{s+\lambda_s}\frac{1}{\sqrt{2b}}\,\mathbf{U}_s(s) \;-\; \frac{2bs + B\lambda_s}{2Bb(s+\lambda_s)}\frac{\lambda_s}{s+\lambda_s}\,\mathbf{F}_e(s)$$

which shows that the slave velocity tracks the incoming wave via a first order filter. Meanwhile, the effect of the contact force $\mathbf{F}_e$ is reduced in the low frequency range.

Of course, such perfect impedance matching is only possible in the linear case. For a state dependent inertia matrix of a multi degree of freedom system, the overall impedance will also be state dependent and the matching can only be approximate. However, good tuning still provides a substantial reduction in reflections and performs well in practice.

Furthermore, when the environment contact provides additional inertial or damping effects, the above analysis will be changed. Clearly we have to tune the gains for a desired operating configuration. So while we definitely suggest at least approximate matching, we will also consider other methods for reducing the effects of reflections in later chapters, in particular wave filters.

5.4.3 General Impedance Matching

So far we have considered the master and slave robots to present simple inertial dynamics. Any friction terms may be lumped into the damping elements. While this is usually a sufficient model, some cases may warrant the inclusion of other dynamic effects in the impedance matching process. For example, we could incorporate the expected operator behavior or some environmental dynamics, to further reduce reflections during a particular task or scenario. With care, the resulting controllers may even be substituted as needed during operations. We now briefly discuss this general case.

Extending the previous developments, we expect that appropriate dynamics in the controller can counter act any complex dynamic effects of an existing element. This is indeed true for all passive LTI systems, with one restriction detailed below. Unfortunately, similar general statements cannot be made about all nonlinear systems. Nevertheless, even approximate matching is always suggested in these cases and can greatly reduce reflections.

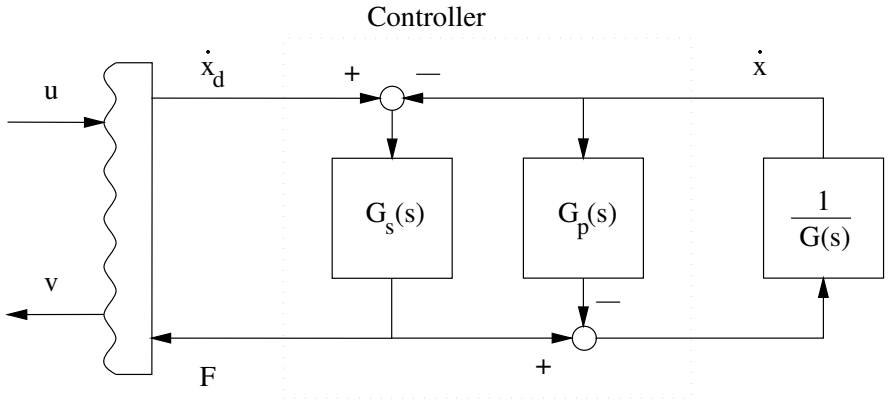


Figure 5-11: A general passive element may be impedance matched by adding appropriate dynamics in the controller. For example, a parallel $G_p(s)$ and series $G_s(s)$ impedance may be used.

Consider a passive element given by the impedance

$$G(s) = \frac{\mathbf{F}(s)}{s\mathbf{X}(s)} \tag{5.37}$$

Together with the controller, the element should appear as a simple damper of magnitude b to the rest of the system. Then the expected reflections are zero and only when the element is subjected to initial conditions, external inputs, or disturbances will a return wave form.

To achieve this goal using only passive controllers, the element must satisfy one condition. It must be position controllable with limited friction, so that it may track the motion of the simulated damper with the energy supplied by the wave command. Thus it must be capable of reaching any position and remaining there at rest without further inputs. This requirement manifests itself as a limited low frequency gain

$$|G(0)| < b \tag{5.38}$$

where b is the desired impedance.

Many different controllers can achieve the objective. In particular, we can combine the existing element with a parallel $G_p(s)$ and series $G_s(s)$ impedance, as shown in Figure 5-11. Should the magnitude of the existing impedance fall below the desired value at any frequency ($|G(j\omega)| < b$), then it is increased via the parallel element $G_p(s)$. The combination of the two should be no less than the desired value b. The series element $G_s(s)$ will reduce this combination and can enforce a constant value

for all frequencies.

Naturally $G_p(s)$ and $G_s(s)$ should be chosen as light and stiff as possible to satisfy these needs. In addition, to assure true position feedback, we require that the series element $G_s(s)$ contain a spring force and hence $G_s(0) = \infty$. This also effects the parallel element as

$$G_p(0) = b - G(0)$$

Fortunately, this is consistent with the original condition (5.38) and does not hinder the matching.

For example, consider a mass-spring system attached via the spring

$$G(s) = \frac{mKs}{ms^2 + K}$$

This is a fairly extreme example with the impedance varying from zero to infinity. We increase the impedance to at least the desired value with the simple parallel element

$$G_p(s) = b$$

Using the series element with mass, damper, and spring components

$$G_s(s) = \frac{b^2}{K}s + b + \frac{b^2}{ms}$$

provides the intended constant value. The final combined impedance is indeed

$$G_{total}(s) = \frac{G_s(s)\ \ [G_p(s) + G(s)]}{G_s(s) + G_p(s) + G(s)} = b$$

5.5 Impedance Matched Teleoperator

The above sections have pointed out that wave reflections may occur and should be avoided if possible. This is only possible when controlling position and velocity, which motivates us to present a new symmetric configuration. Figure 5-12 shows the block diagram of transfer functions, again for a single degree of freedom.

5.5.1 System Description

In this case, the wave communications acts as a true admittance. It observes forces on both sides and updates the desired motions as needed. Both master and slave

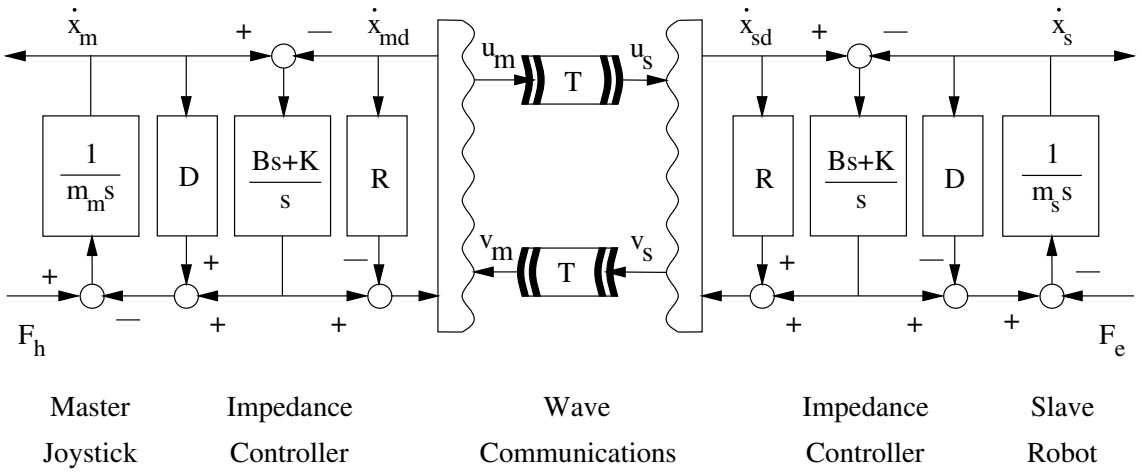


Figure 5-12: The symmetric impedance matched teleoperator uses P.D. control on both sides to eliminate wave reflections.

are controlled via an impedance matched P.D. controller. The gains are selected as described in (5.35), leaving three parameters to tune: λ_m, λ_s, and b. Both bandwidth values should be increased as far as the individual actuators and unmodeled dynamics permit, to provide the best possible stiffness. This then leaves the wave impedance which can be altered during execution to provide different appearances.

Without reflections, the dynamics are also simplified and provide the following single d.o.f. transfer functions.

$$s\mathbf{X}_s(s) = \frac{\lambda_s}{s+\lambda_s}\frac{1}{\sqrt{2b}}\mathbf{U}_s(s) - \frac{2bs+B_s\lambda_s}{2B_sb(s+\lambda_s)}\frac{\lambda_s}{s+\lambda_s}\mathbf{F}_e(s) \tag{5.39}$$

$$s\mathbf{X}_m(s) = \frac{\lambda_m}{s+\lambda_m}\frac{1}{\sqrt{2b}}\mathbf{V}_m(s) + \frac{2bs+B_m\lambda_m}{2B_mb(s+\lambda_m)}\frac{\lambda_m}{s+\lambda_m}\mathbf{F}_h(s) \tag{5.40}$$

$$\mathbf{V}_s(s) = -\frac{\lambda_s}{s+\lambda_s}\frac{1}{\sqrt{2b}}\mathbf{F}_e(s) \tag{5.41}$$

$$\mathbf{U}_m(s) = \frac{\lambda_m}{s+\lambda_m}\frac{1}{\sqrt{2b}}\mathbf{F}_h(s) \tag{5.42}$$

Both wave signals are only present if and when external forces are applied, either by the human operator or the slave environment. In addition they are filtered providing a smooth operation.

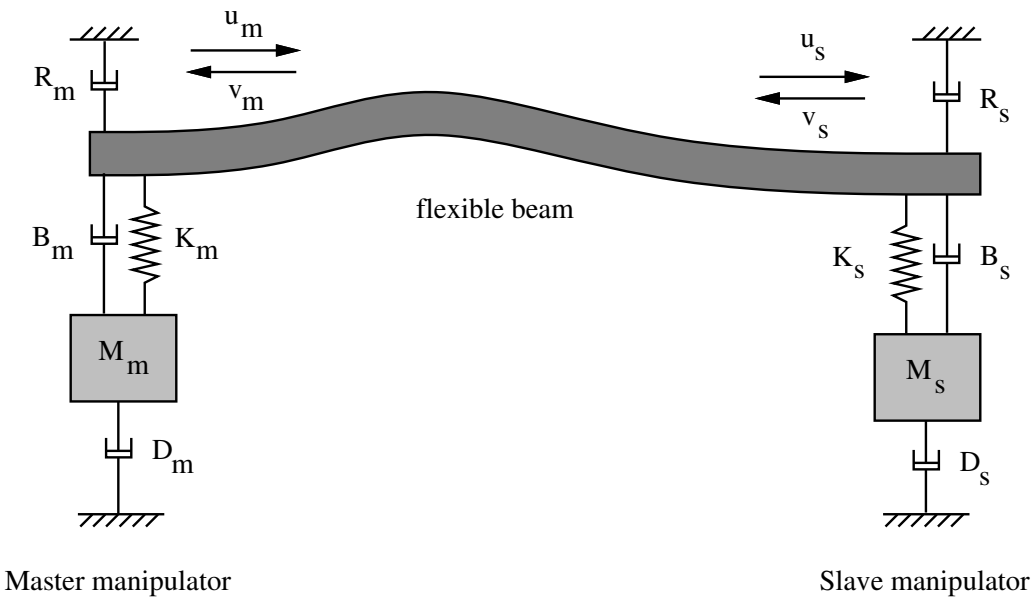


Figure 5-13: The impedance matched wave teleoperator can be visualized by mechanical elements, where the communications is replaced by a flexible beam.

5.5.2 Appearance and Interpretation

In an attempt to visualize the system, Figure 5-13 shows the mechanical equivalent of the entire system. The wave transmission is replaced by a flexible beam, which connects both locations.

The beam is an idealization with simple dynamics. It is best represented by a rotational beam or a tensioned string. More complex beam models, such as Bernoulli-Euler or Timoshenko are beyond our needs. In general the beam length is proportional to the time delay and the beam rigidity and weight characterize the wave impedance.

The properties can also be summarized in the mechanical parameters of total inertia, apparent damping, and steady state stiffness (Slotine and Niemeyer, 1993).

$$M = M_m + bT + M_s \tag{5.43a}$$

$$B = 2b \tag{5.43b}$$

$$K^{-1} = K_m^{-1} + \left(\frac{b}{T}\right)^{-1} + K_s^{-1} \tag{5.43c}$$

They clearly show the effect of the wave impedance b and time delay T. The im-

pedance determines the trade off between high inertia and high stiffness. Meanwhile delay deteriorates both parameters equally, increasing the mass and reducing the spring constant.

Together these parameters define the characteristics of the virtual tool (see Section 1.2). For small delays, the system is dominated by the master and slave inertia and stiffness, so that it should appear quite light and stiff. As such it may be compared to a simple tool such as a screwdriver.

For large delays, the transmission parameters are dominant and the choice of b becomes important. A large impedance provides high inertia and high stiffness. The tool may be represented by a heavy drill. At the other end of the spectrum, a small impedance reduces inertia but also stiffness. The system then appears more like a soft sponge.

5.5.3 Limitations and Behavior

The impedance matching teleoperator was designed based on the assumption that both master and slave robots are described by pure inertias. In addition their values were also assumed. As long as the system moves freely, these assumptions are valid and the wave reflections are indeed eliminated. However, if and when the slave makes contact with some environment or the operator significantly alters the behavior of the master, the wave reflections will reappear. This is natural consequence of impedance matching, which is tuned for some expected operating condition.

Three solutions present themselves. First we can add wave filtering to help reduce the reflections as discussed in Section 6.3. This is independent of the expected configuration, but also reduces the general performance. Second we can attempt to derive alternate controllers for different operator conditions and tasks. And third, we can accept the reflections and adjust the wave impedance to minimize their disruptive effect.

Indeed imagine the slave robot in rigid contact. In this case we find the reflections determined by

$$\mathbf{V}_s(s) = -\frac{K_s}{2bs + K_s}\,\mathbf{U}_s(s) \tag{5.44}$$

which is a low pass filter of bandwidth $\frac{K_s}{2b}$. So increasing the wave impedance will lower the filter bandwidth and prevent the reflections from becoming oscillatory or disorienting the operator. This indeed matches the normal tuning process for the wave impedance well: Low b during quick motions in free space - no reflections will appear. High b for good force clarity during contact operations - the reflections are

small and slow.

Hence the impedance matched teleoperator together with appropriate online tuning of the wave impedance can achieve good performance both in free space and in contact.

5.5.4 Experimental Validation

The developments are also verified by experiments. Consider Figure 5-14 which shows the results of two quick motion commands executed while the system was in free space without slave contact. It is impedance matched at both sides and the data shows no evidence of reflections. The forces correspond to operator inputs pushing against the apparent damping to achieve the large velocities. The delay was set at 1 second.

In contrast, Figure 5-15 shows the impedance matched system while the slave is in contact in the vertical z direction. The operator tries to hold a constant deflection in an attempt to create a constant force. Clearly the impedance matching has broken down and the wave reflections are oscillating between the two sides. And consequently the force feedback also oscillates strongly. The period is equal to the delay time and the amplitude decreases only very slowly.

However, if we adjust the wave impedance to a higher value and repeat the same test, the wave reflections are gone. Figure 5-16 shows the results after b was increased by a factor of six. Also note that the deflection is now substantially smaller. The change in wave impedance simultaneously raises the system stiffness and gives the operator a better spatial picture of the remote environment. The deflections in x and y directions are caused by contact stiction.

And so we verify that the impedance matched teleoperator works well, if the wave impedance is tuned online according to the current task.

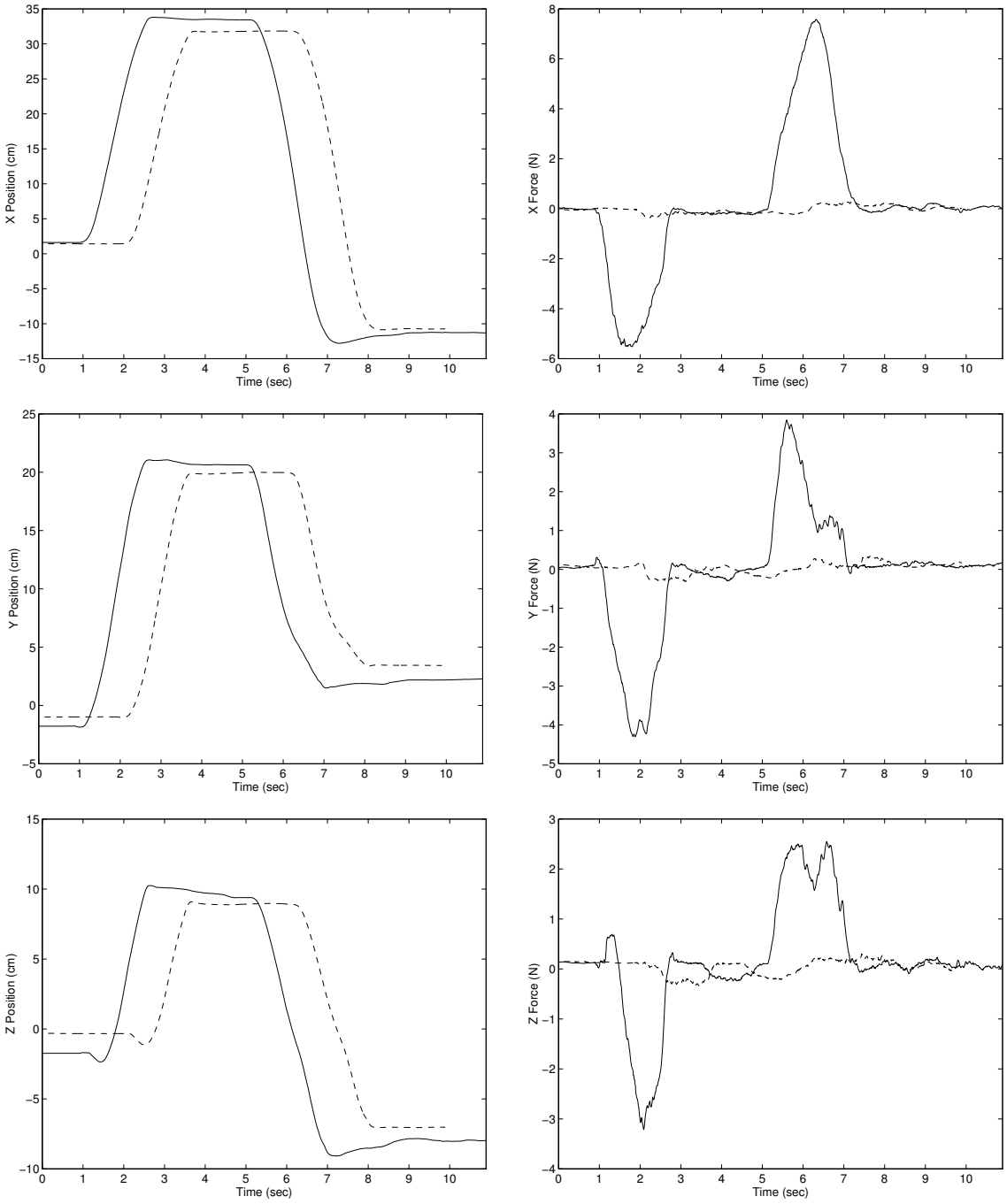


Figure 5-14: Impedance matching can eliminate wave reflections when the telerobot is operated in free space. Here the motion commands are executed after the delay of $T = 1.0$ without further problems.

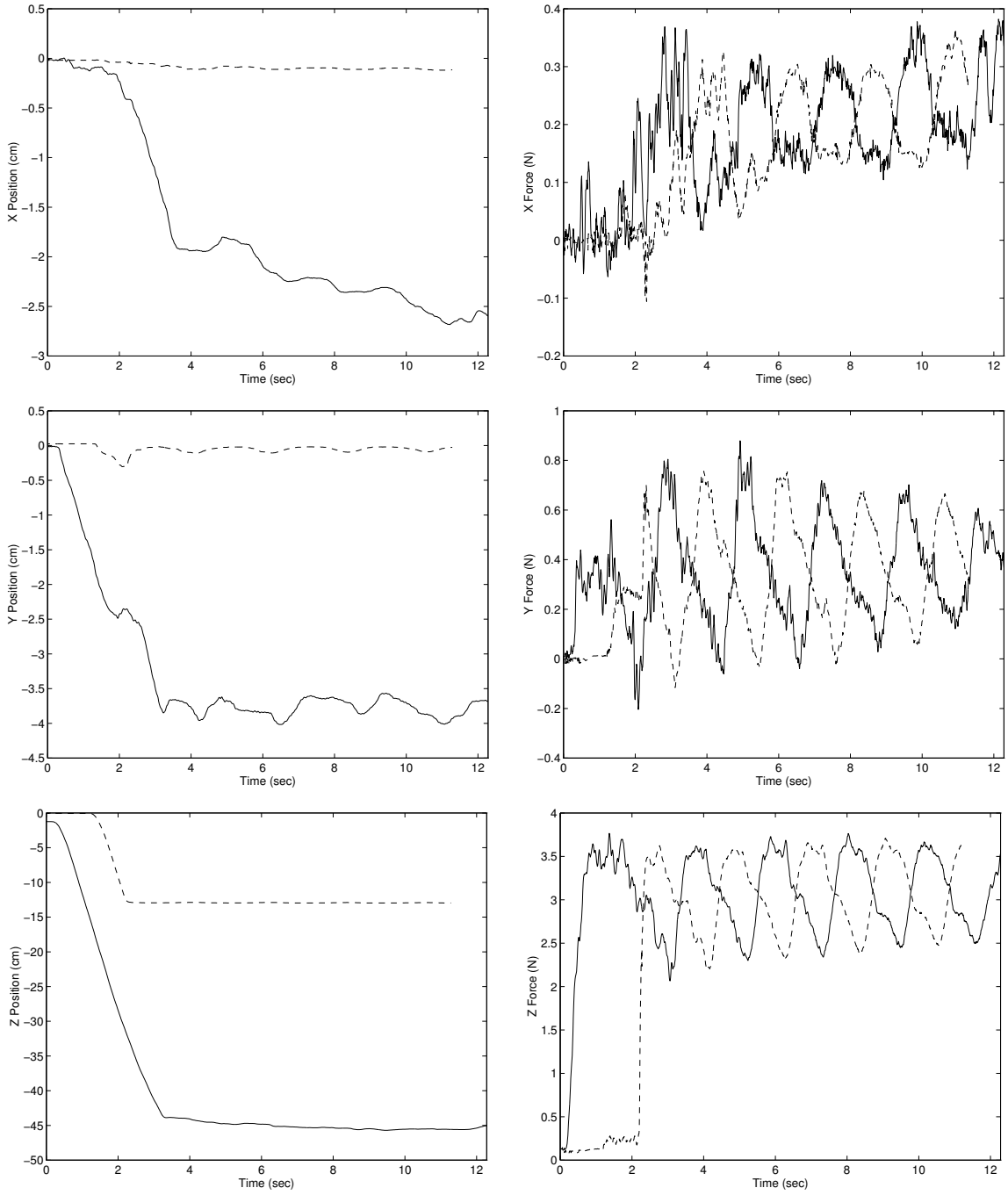


Figure 5-15: If the wave impedance b is low while contact is made, wave reflections may again appear. Here the operator tries to apply a constant force. Instead the force signals oscillate with the delay of $T = 1.0$.

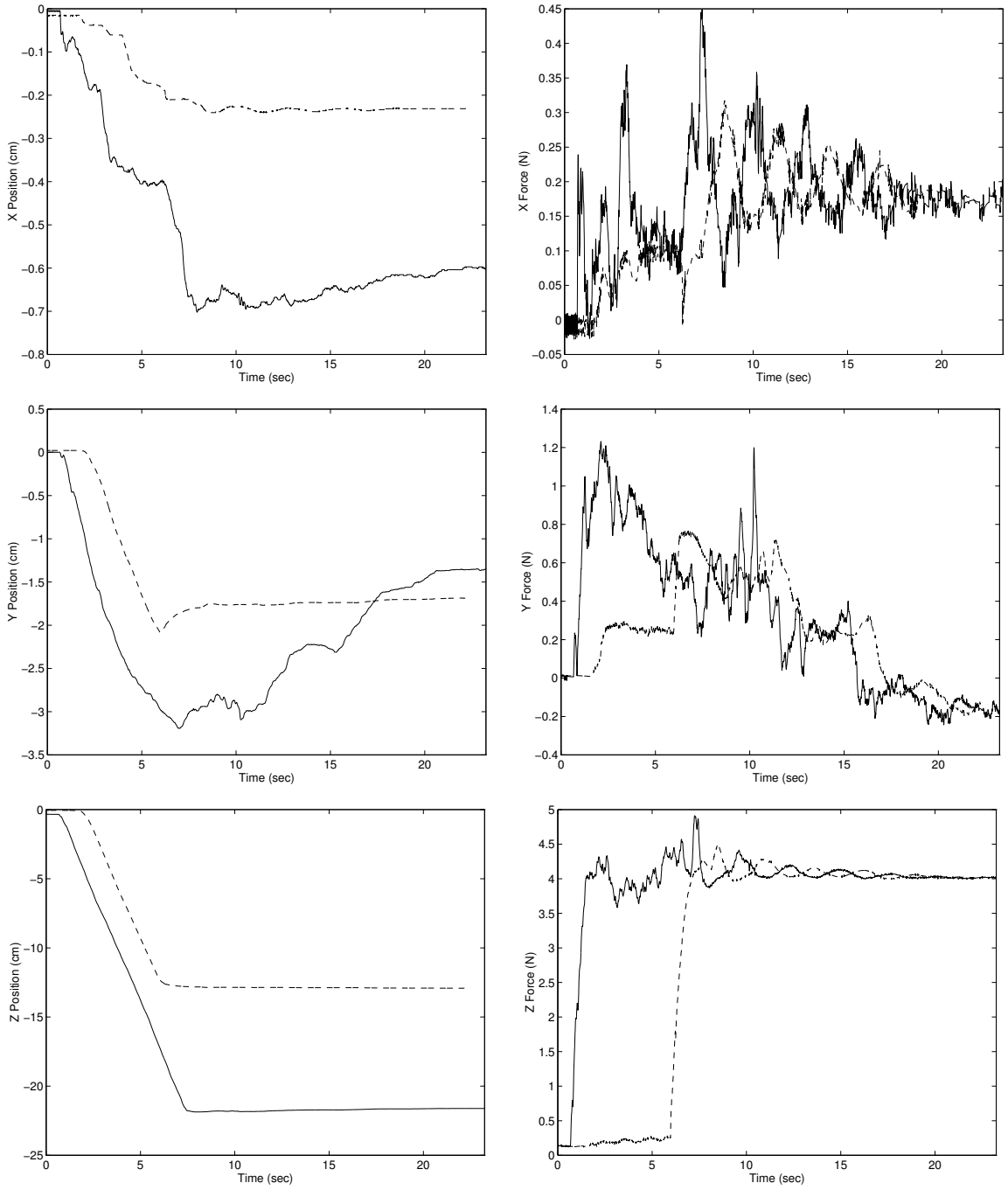


Figure 5-16: Tuning the wave impedance b high during contact again minimizes the wave reflections. Now a constant force may be applied and felt more clearly. All other gains are unchanged and the delay is $T = 1.0$.

Chapter 6

Designing in Wave Space

In the previous chapter we utilized wave variables to replace the communications element in a teleoperator. This wave based element stabilized the system, but also introduced new dynamics especially in the form of wave reflections. Fortunately, we were able to eliminate or at least reduce such reflections using impedance matched controllers.

However, the procedures continued to use controllers designs based on power variables and much of the analysis was performed in power space. We now examine some applications of design directly within the wave domain. This opens new possibilities and suggests different layouts.

We begin by reviewing two-port wave systems, which will be prevalent in the remainder. Then we concentrate on wave filters, which can help reduce the reflections even when impedance matching fails. Thereafter, we look at general wave controller designs, leading us to a purely wave based teleoperator. And finally we discuss the possible applications of wave variables to shared control via wave junctions and to predictors.

6.1 Two-Port Wave Systems

We begin by reviewing the conditions for passivity and position tracking within the context of two-port wave systems, as shown in Figure 6-1. Many of the later elements and systems are actually linear, so that we detail a Laplace domain view. Nevertheless, given the nature of passivity, these systems can still be combined with other nonlinear elements.

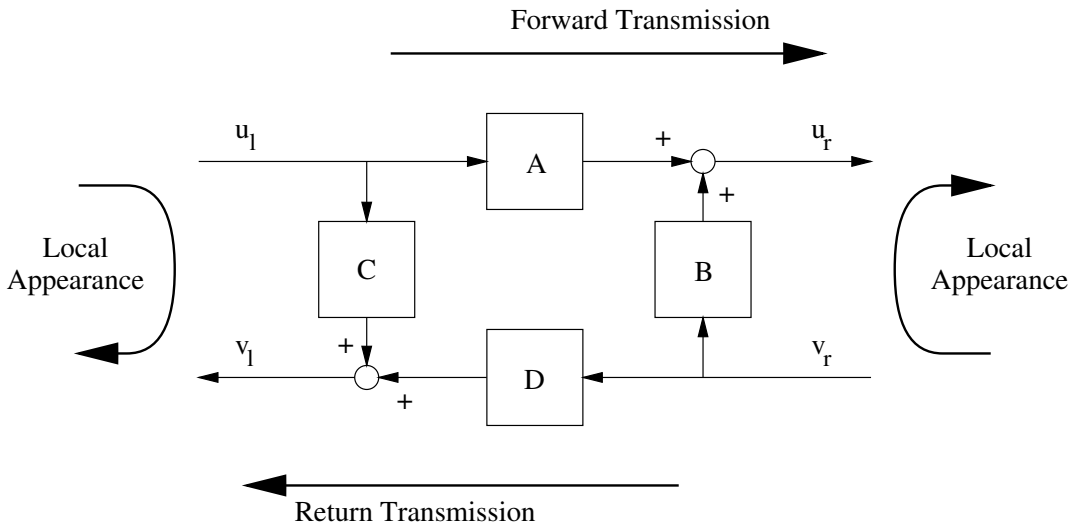


Figure 6-1: A general two-port wave system relates the wave signals at the left and right port. The inputs $\mathbf{u}_l$ and $\mathbf{v}_r$ are mapped to the outputs $\mathbf{u}_r$ and $\mathbf{v}_l$.

6.1.1 Passivity

As previous discussions have shown, passivity can be tested in a variety of ways. In their simplest form, such tests compare the input power directly to the output power. Following the developments of Section 4.2 we find the wave system passive if and only if

$$\int_0^t \tfrac{1}{2}\mathbf{u}_r^T\mathbf{u}_r + \tfrac{1}{2}\mathbf{v}_l^T\mathbf{v}_l \, \mathrm{d}\tau \leq E_{store}(0) + \int_0^t \tfrac{1}{2}\mathbf{u}_l^T\mathbf{u}_l + \tfrac{1}{2}\mathbf{v}_r^T\mathbf{v}_r \, \mathrm{d}\tau \qquad \forall t \geq 0 \tag{6.1}$$

However, this may not always be easy to verify. For linear time-invariant (LTI) systems, we can also confirm passivity via the norm of the wave transfer or scattering matrix $H(s)$.

$$\begin{bmatrix} \mathbf{u}_r \\ \mathbf{v}_l \end{bmatrix} = \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} \begin{bmatrix} \mathbf{u}_l \\ \mathbf{v}_r \end{bmatrix} \quad \Rightarrow \quad H(s) = \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} \tag{6.2}$$

From (2.14) we find a system passive if and only if

$$||H||^2 = \sup_\omega ||H(j\omega)||^2 = \sup_\omega \lambda_{\max}\left(H^*(j\omega)H(j\omega) \right) \leq 1 \tag{6.3}$$

where $\lambda_{\max}(.)$ is the largest eigenvalue and H^* denotes the complex conjugate transpose of H. Substituting the 2 by 2 form of $H(s)$ and after appropriate manipulation,

we find this equivalent to the following three conditions

$$\text{(a)} \quad |A|^2 + |C|^2 \leq 1 \tag{6.4a}$$

$$\text{(b)} \quad |B|^2 + |D|^2 \leq 1 \tag{6.4b}$$

$$\text{(c)} \quad |A|^2 + |B|^2 + |C|^2 + |D|^2 \leq 1 + |AD - BC|^2 \tag{6.4c}$$

which must be satisfied for all $\omega \geq 0$. The third item can also be expressed as

$$|AD - BC|^2 \;\geq\; 1 \;-\; \Big(1 - |A|^2 - |C|^2\Big) \;-\; \Big(1 - |B|^2 - |D|^2\Big)$$

The first two conditions are quite straight forward and simply state that the power entered via one input wave must be conserved as the wave signal is split into two parts. The third condition adds a twist. It is satisfied if the power moves mostly across the system in both directions ($|AD| \gg |BC|$), or if the power is mostly reflected at both sides ($|AD| \ll |BC|$). It is also satisfied if the power is split up evenly but out of phase ($AD \simeq -BC$),or the power is substantially dissipated ($|A|^2 + |B|^2 + |C|^2 + |D|^2 \leq 1$). It only prevents cases where all power is distributed equally and in phase. Remember that adding two waves can double the wave signal and hence quadruple the power. So to remain passive, simple wave additions and subtractions must be done with care and can require constraints as is the case here.

6.1.2 Delayed Wave Transmissions

In the case of delayed wave transmissions, the delay operator may be included explicitly, as shown in Figure 6-2. Also, we assume the delay time T is unknown or variable, so that the system should be passive for any value of T. This obviously effects the passivity conditions (6.4), which become

$$\text{(a)} \quad |A|^2 + |C|^2 \leq 1 \tag{6.5a}$$

$$\text{(b)} \quad |B|^2 + |D|^2 \leq 1 \tag{6.5b}$$

$$\text{(c)} \quad |A|^2 + |B|^2 + |C|^2 + |D|^2 \leq 1 + \Big(|AD| - |BC|\Big)^2 \tag{6.5c}$$

and again must be satisfied for all $\omega \geq 0$. Notice the first two conditions are not altered, because the power input from each wave must still be conserved. But the

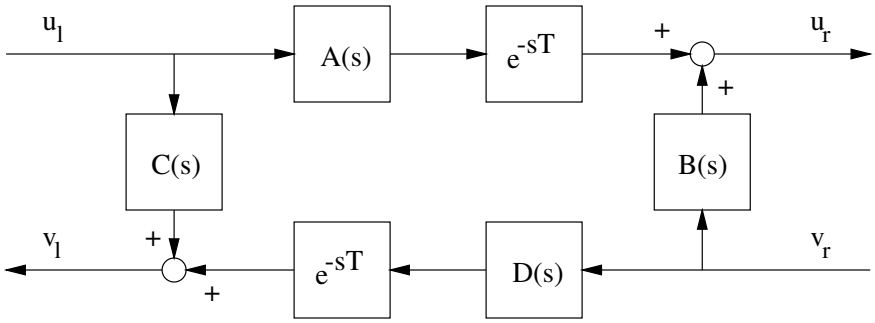


Figure 6-2: The delayed two-port wave system explicitly includes the delay operator in the forward and reverse transmission transfer functions.

third condition is now somewhat more restrictive to account for the variable delay. It may again be expressed in an alternate form

$$\Big(|AB| + |CD|\Big)^2 \;\leq\; \Big(1 - |A|^2 - |C|^2\Big)\,\Big(1 - |B|^2 - |D|^2\Big)$$

For symmetric systems ($D(s) = A(s)$, $C(s) = B(s)$) this simplifies nicely into

$$2|AB| \;\leq\; 1 - |A|^2 - |B|^2$$

and finally into

$$|A| + |B| \leq 1 \tag{6.6}$$

which also satisfies the first two conditions. So a symmetric delayed wave controller is passive regardless of the delay T, if (6.6) is satisfied.

6.1.3 Position Tracking and Convergence

Wave variables encode both velocity and force information. Designing systems in these terms can achieve appropriate behavior and guarantee convergence of velocity and force signals. However, this alone is not sufficient to assure position tracking and convergence as well. For example, if an output velocity typically lags its input or has some other nonzero errors, the output position may be quite different from the desired location.

The solution to this problem comes in two parts. First, place an additional requirement on the wave controller to assure position tracking at least in theory. This implies that the velocity error should be zero on average. Second, include an additional position feedback path to enforce the theoretical result. We proposed two such

methods in Section 5.3 and thus now concentrate on the first part.

Position information is contained in the integral of wave signals. And so position tracking is achieved if these integrals are conserved. Indeed using the wave integrals explicitly is suggested in Section 5.3. But more carefully, we know

$$\mathbf{x}_l(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_l(\tau) + \mathbf{v}_l(\tau)\,\mathrm{d}\tau \qquad \text{and} \qquad \mathbf{x}_r(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_r(\tau) + \mathbf{v}_r(\tau)\,\mathrm{d}\tau$$

so that the position error between the left and right side is given by

$$\Delta\mathbf{x}(t) = \mathbf{x}_l(t) - \mathbf{x}_r(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_l(\tau) + \mathbf{v}_l(\tau) - \mathbf{u}_r(\tau) - \mathbf{v}_r(\tau)\,\mathrm{d}\tau \tag{6.7}$$

To guarantee position tracking, we must assure that the error converges to zero as and when the wave signals die out

$$\lim_{t\to\infty} \mathbf{u}_l(t) = \lim_{t\to\infty} \mathbf{v}_r(t) = 0 \qquad \Rightarrow \qquad \lim_{t\to\infty} \Delta\mathbf{x}(t) = 0 \tag{6.8}$$

While the waves are nonzero, the system will still be moving or exerting forces and some position errors may remain.

For LTI controllers we can substitute the transfer functions to obtain

$$\begin{aligned} \Delta\mathbf{X}(s) &= \frac{1}{\sqrt{2b}\,s}\Big(\mathbf{U}_l(s) + \mathbf{V}_l(s) - \mathbf{U}_r(s) - \mathbf{V}_r(s)\Big) \\ &= \frac{1 - A(s) + C(s)}{\sqrt{2b}\,s}\,\mathbf{U}_l(s) \;-\; \frac{1 - D(s) + B(s)}{\sqrt{2b}\,s}\,\mathbf{V}_r(s) \end{aligned}$$

The final value theorem dictates that $\lim_{s\to 0} s\mathbf{U}_l(s) = \lim_{s\to 0} s\mathbf{V}_r(s) = 0$ and therefore we find the following conditions for position tracking of a wave controller

$$\lim_{s\to 0} \frac{1 - A(s) + C(s)}{s} = \text{finite} \qquad \lim_{s\to 0} \frac{1 - D(s) + B(s)}{s} = \text{finite} \tag{6.9}$$

which also imply

$$A(0) = 1 + C(0) \qquad D(0) = 1 + B(0) \tag{6.10}$$

Perhaps not very intuitive, these conditions will hopefully become more clear as we examine typical controllers in later sections. Generally systems without explicit damping elements pass wave signals directly through in steady state ($A(0) = D(0) = 1$) while avoiding reflections ($B(0) = C(0) = 0$) and thus satisfy this constraint. Damping elements will divide the wave signal, letting some portion pass and reflecting

the rest, but also satisfy (6.10).

6.1.4 Concatenation Form

While the normal description (A, B, C, D) is useful in understanding the wave behavior, it does not allow easy concatenation of multiple controllers or elements. Instead it may be helpful to rewrite the system in a non-causal form relating both left waves to both right waves.

$$\underbrace{\begin{bmatrix} \mathbf{u}_r \\ \mathbf{v}_l \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{u}_l \\ \mathbf{v}_r \end{bmatrix}}_{\text{causal form}} \quad \Leftrightarrow \quad \underbrace{\begin{bmatrix} \mathbf{u}_r \\ \mathbf{v}_r \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} \mathbf{u}_l \\ \mathbf{v}_l \end{bmatrix}}_{\text{non-causal form}} \tag{6.11}$$

Using this non-causal form, any elements may be cascaded simply by multiplying their $(\alpha, \beta, \gamma, \delta)$-matrices.

Conversion between the two forms is governed by the following rules.

$$\begin{aligned} A &= \frac{\alpha\delta - \beta\gamma}{\delta} & B &= \frac{\beta}{\delta} & C &= -\frac{\gamma}{\delta} & D &= \frac{1}{\delta} \\ \alpha &= \frac{AD - BC}{D} & \beta &= \frac{B}{D} & \gamma &= -\frac{C}{D} & \delta &= \frac{1}{D} \end{aligned} \tag{6.12}$$

6.2 Wave Filtering

In many control applications filtering is often necessary to eliminate high frequency noise, but also presents a risk of instability due to the added phase lag. In this respect they are closely related to time delays. So not surprisingly, wave variables provide an effective way to include filtering within wave based control systems.

Filters may be used for several purposes. They include the reduction of noise, the computation of derivations, and shaping of the frequency spectrum to enhance or hide certain characteristics.

6.2.1 Definition and Passivity

In its basic form, a wave filter is simply a first order stable linear filter executed in the wave space. The block diagram in Figure 6-3 shows the right moving wave passing

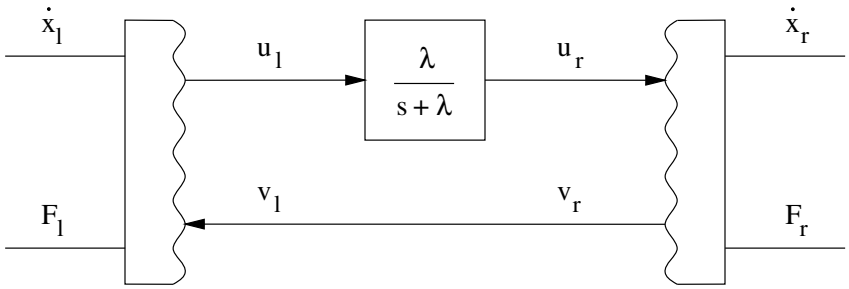


Figure 6-3: Filtering a wave variables is passive and thus inherently stable, regardless of the introduced phase lag.

through the filter. Other layouts may do the same to the left moving wave.

The governing equation is given by the first order differential equation

$$\dot{\mathbf{u}}_r + \lambda \mathbf{u}_r = \lambda \mathbf{u}_l \tag{6.13}$$

where λ is the corresponding bandwidth or cutoff frequency.

To verify passivity, let us examine the power input to this element.

$$P_{in} = \mathbf{u}_l^T \mathbf{u}_l - \mathbf{u}_r^T \mathbf{u}_r$$

Substituting the definition (6.13), we find

$$P_{in} = \left(\frac{\dot{\mathbf{u}}_r + \lambda \mathbf{u}_r}{\lambda} \right)^T \left(\frac{\dot{\mathbf{u}}_r + \lambda \mathbf{u}_r}{\lambda} \right) - \mathbf{u}_r^T \mathbf{u}_r = \frac{1}{\lambda^2} \dot{\mathbf{u}}_r^T \dot{\mathbf{u}}_r + \frac{2}{\lambda} \mathbf{u}_r^T \dot{\mathbf{u}}_r$$

The system satisfies the passivity requirement (2.2) of the form

$$P_{in} = \frac{\mathrm{d}}{\mathrm{d}t} E_{store} + P_{diss}$$

by defining the positive energy storage function

$$E_{store} = \frac{1}{\lambda} \mathbf{u}_r^T \mathbf{u}_r \tag{6.14}$$

and the power dissipation

$$P_{diss} = \frac{1}{\lambda^2} \dot{\mathbf{u}}_r^T \dot{\mathbf{u}}_r \tag{6.15}$$

The total output energy is thus limited by the input energy

$$\int_0^t \mathbf{u}_r^T \mathbf{u}_r \, \mathrm{d}\tau \leq \int_0^t \mathbf{u}_l^T \mathbf{u}_l \, \mathrm{d}\tau$$

assuming zero initial conditions.

The filter continually dissipates energy if the wave signal is rapidly changing and so removes any high frequency power. It also stores some power for a nonzero wave signal, so that it can smoothly reduce the output even if the input drops suddenly.

Of course other, more complex and higher order filters can also be used. In each case the output energy should be limited by the input. For example, this excludes underdamped second order filters with a resonant peak. But it allows nonlinear filtering involving saturation.

6.2.2 Steady State Tracking

As we have discussed on previous occasions, the steady state position error between left and right side of any wave element depends on the integral of the wave signals. From (4.8) we can compute the individual positions as

$$\mathbf{x}_l(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_l(\tau) + \mathbf{v}_l(\tau) \, \mathrm{d}\tau \qquad \mathbf{x}_r(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_r(\tau) + \mathbf{v}_r(\tau) \, \mathrm{d}\tau \quad (6.16)$$

so that the error is given by

$$\Delta\mathbf{x}(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_l(\tau) + \mathbf{v}_l(\tau) - \mathbf{u}_r(\tau) - \mathbf{v}_r(\tau) \, \mathrm{d}\tau \quad (6.17)$$

In the case of the above described filter, the left moving wave is unaffected and drops out of (6.17). And the filter equation (6.13) determines the right moving wave, so that we have

$$\Delta\mathbf{x}(t) = \frac{1}{\sqrt{2b}} \int_0^t \mathbf{u}_l(\tau) - \mathbf{u}_r(\tau) \, \mathrm{d}\tau = \frac{1}{\sqrt{2b}} \int_0^t \frac{1}{\lambda} \dot{\mathbf{u}}_r(\tau) \, \mathrm{d}\tau = \frac{1}{\sqrt{2b}\,\lambda} \mathbf{u}_r(t) \quad (6.18)$$

So we see that the momentary error is proportional to the wave amplitude. The steady state error, when the wave signals have died down and converged to zero, is hence zero as we desire.

But beware that this may not be the case for more complex filters. In normal applications, we therefore must place a second requirement on the allowable set of

filters. Not only show they limit the output energy for passivity, but the output wave integral should converge to the input wave integral.

$$\int_0^t \mathbf{u}_r(\tau)\,\mathrm{d}\tau \to \int_0^t \mathbf{u}_l(\tau)\,\mathrm{d}\tau \tag{6.19}$$

For linear filters this implies a unit low frequency gain.

6.2.3 Spring-Like Characteristics

Like the wave communications in Section 5.1 and other wave systems, the modification of wave signals creates an appearance with both impedance and admittance characteristics. Depending on the external interactions, the element will display a behavior which is appropriate for the situation.

As before we are interested in determining the overall spring constant or stiffness which the element exhibits. For a constant deflection, we expect to find steady state forces of

$$\mathbf{F}_l = \mathbf{F}_r = K_{filter}\,\Delta\mathbf{x} = K_{filter}\,(\mathbf{x}_l - \mathbf{x}_r) \tag{6.20}$$

The deflection is related to the wave variables via

$$\Delta\mathbf{x}(t) = \mathbf{x}_l(t) - \mathbf{x}_r(t) = \frac{1}{\sqrt{2b}}\int_0^t \mathbf{u}_l(\tau) + \mathbf{v}_l(\tau) - \mathbf{u}_r(\tau) - \mathbf{v}_r(\tau)\,\mathrm{d}\tau$$

Substituting the filter definition (6.13) we find

$$\Delta\mathbf{x}(t) = \frac{1}{\sqrt{2b}}\int_0^t \frac{1}{\lambda}\,\dot{\mathbf{u}}_r(\tau)\,\mathrm{d}\tau = \frac{1}{\sqrt{2b}}\frac{1}{\lambda}\,\mathbf{u}_r(t)$$

In steady state, both sides are assumed to be at a constant position. The velocities are zero, and the wave variables reflect the forces as

$$\mathbf{u}_l = -\mathbf{v}_l = \frac{1}{\sqrt{2b}}\,\mathbf{F}_l \qquad\qquad \mathbf{u}_r = -\mathbf{v}_r = \frac{1}{\sqrt{2b}}\,\mathbf{F}_r$$

Meanwhile the effect of the filter has dissipated and the left and right wave variables are equal

$$\mathbf{u}_r = \mathbf{u}_l = \mathbf{u} \qquad\qquad \mathbf{v}_l = \mathbf{v}_r = \mathbf{v}$$

And so we find both forces proportional to the deflection

$$\mathbf{F}_l = \mathbf{F}_r = 2b\lambda\,\Delta\mathbf{x} \tag{6.21}$$

where the spring constant

$$K_{filter} = 2b\lambda \tag{6.22}$$

is dependent on the filter bandwidth and wave impedance. Not surprisingly a faster filter will create a stiffer appearance closer to a perfect transmission.

Unlike the wave communications, the effect is immediately available without any delay. Nevertheless, this element is *not* a spring and there will be some transient forces related to the apparent damping in the wave transformation not typical of springs.

Finally, note that filtering both forward and returning waves will naturally half the stiffness.

6.2.4 Admittance Characteristics

If the element is subject to input forces, it will determine motion and behave much like an inertia. In particular, the steady state speed should satisfy

$$\dot{\mathbf{x}}_l = \dot{\mathbf{x}}_r = \frac{1}{M_{filter}} \int_0^t \mathbf{F}_l - \mathbf{F}_r \, \mathrm{d}\tau \tag{6.23}$$

where the mass is given by

$$M_{filter} = \frac{b}{2\lambda} \tag{6.24}$$

In this case, the input momentum is computed by

$$\int_0^t \mathbf{F}_l - \mathbf{F}_r \, \mathrm{d}\tau = \sqrt{\frac{b}{2}} \int_0^t \mathbf{u}_l(\tau) - \mathbf{v}_l(\tau) - \mathbf{u}_r(\tau) + \mathbf{v}_r(\tau) \, \mathrm{d}\tau$$

which becomes

$$\int_0^t \mathbf{F}_l - \mathbf{F}_r \, \mathrm{d}\tau = \sqrt{\frac{b}{2}} \int_0^t \frac{1}{\lambda} \dot{\mathbf{u}}_r(\tau) \, \mathrm{d}\tau = \sqrt{\frac{b}{2}} \frac{1}{\lambda} \mathbf{u}_r(t)$$

And, if we assume no further input forces, both right and left moving waves take the same constant value

$$\mathbf{u}_l = \mathbf{u}_r = \mathbf{v}_l = \mathbf{v}_r = \sqrt{\frac{b}{2}} \dot{\mathbf{x}}$$

So in steady state the behavior is equivalent to a normal mass. Though we reiterate that the transient behavior will differ and show damping as the filter dissipates

energy according to (6.15). A faster filter will cause less deviations from a perfect transmission and hence present a smaller inertia. And filtering in both wave directions will double the effective mass.

6.2.5 Filtered Differentiation

It is common for robot controllers to require and incorporate knowledge of the desired acceleration, especially if they are constructed for tracking purposes. Such information enables application of the control torques needed to sustain the desired motions without buildup of tracking errors. However, in the context of teleoperation, the desired motion for the slave is measured at the master joystick. It is not known in advance, so that acceleration terms are not readily available.

Filtering the commanded motion provides a smooth signal with known derivatives. But executing such filters in power variables, especially velocity, is not passive and can cause significant stability problems due to the increased phase lag. A possible solution is to filter wave variables instead.

As with any linear first order filter, the above defined wave filter provides an exact expression for the output derivative

$$\dot{\mathbf{u}}_r = \lambda(\mathbf{u}_l - \mathbf{u}_r)$$

As the left moving wave is not changed, we can write this as

$$\dot{\mathbf{u}}_r = \lambda(\mathbf{u}_l + \mathbf{v}_l - \mathbf{u}_r - \mathbf{v}_r)$$

And substituting the wave transformations, this leads to

$$b\ddot{\mathbf{x}}_r + \dot{\mathbf{F}}_r = 2b\lambda(\dot{\mathbf{x}}_l - \dot{\mathbf{x}}_r)$$

This allows us to compute an acceleration term provided we have knowledge of the force derivative. Note this is not the force applied to the slave motors, but instead the value submitted to the master for force feedback. While tracking motion in free space this force will typically be zero and while in contact it is generally determined by an analytic expression which we can differentiate explicitly. And with this information, we can compute the desired acceleration commanded by the operator.

Naturally, the phase lag of the filter implicitly prevents a perfect trajectory tracking with respect to an input device. And so the apparent increase of information we get from computing the acceleration term is paid for by the loss of high frequency data. The resulting tracking system cannot overcome the physical limitations of

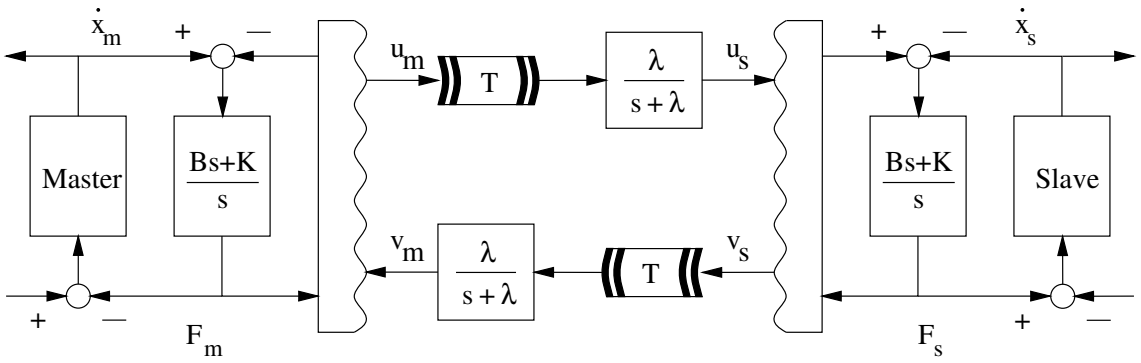


Figure 6-4: Using a filter in the wave space automatically reduces the wave reflections and creates a smoother teleoperator response.

tracking a measured command signal.

6.3 Filtering Teleoperator

Perhaps one of the most important applications of wave filtering is the reduction of wave reflections within the teleoperator context. As we noted extensively in Chapter 5, wave reflections can corrupt or overshadow useful information passing through the system. Most significantly, reflections at both master and slave side can create internal loops with repeated reflections and thereby oscillatory behavior.

While impedance matching can and should be used to minimize the reflections, this process assumes an expected response and thus a nominal operating scenario. Consequently, changes in the environment or operator behavior can obstruct those efforts.

As an alternate or additional method for reducing reflections, we propose a teleoperator configuration as shown in Figure 6-4. Placing a filter element in the wave transmission will eliminate the high frequency signals typical of the reflections. Meanwhile the overall bandwidth of the teleoperator is dominated by the delay time, so that a modest filter delay will have minimal effect.

In general the cutoff frequency of the filter should be selected relative to the delay time. Remember one effect of the reflections is the tendency for oscillations between the two sides, which occur at a frequency of $\frac{1}{2T}$. Placing the filter bandwidth near this value will provide strong damping for these oscillations.

Unfortunately, the filter also effects the desired information contained in the wave

signals, i.e. the feedback and command signals between master and slave. So while the response of the system with a slow filter at $\frac{1}{2T}$ will be smooth, it may also slow down normal operations. In each individual case, tradeoffs between impedance matching, wave filtering, and adjustments of the wave impedance will be necessary. As an approximate rule we use

$$\lambda \geq \frac{1}{T} \tag{6.25}$$

To better describe the effect of filtering on the overall performance, examine the updated system properties first presented in Section 5.5.2. Adding the filter properties from (6.22) and (6.24), we find the system characterized by the steady state parameters

$$M = M_m + bT + \frac{b}{\lambda} + M_s \tag{6.26a}$$

$$B = 2b \tag{6.26b}$$

$$K^{-1} = K_m^{-1} + \left(\frac{b}{T}\right)^{-1} + (b\lambda)^{-1} + K_s^{-1} \tag{6.26c}$$

Clearly the filters increase inertia and reduce stiffness, much like the pure delay. At the same time, the wave impedance can still be used to emphasize either mass or inertia as before.

In summary, the filtering trades off smooth and predictable responses with the speed of operation. And as a final remark, we note that the high frequency data lost by the filtering may be presented to the operator in another form. For example, placing these components on an audio channel will allow the operator to hear important high frequency events such as impact or stick/slip. Tactile displays are also feasible.

6.4 Wave Controllers

Given our general goal of connecting two separate robots or systems, we again focus on two-port controllers. We first examine some basic controllers, which exactly implement the behavior of classic elements, and use them to describe some general characteristics. We then discuss how unmodeled dynamics effect wave controller design. And finally we give rules to convert power domain designs into wave space.

6.4.1 Classic Elements and Basic Examples

Figure 6-5 shows three controllers equivalent to classic elements. In parts (a) and (b) the controllers are first order and implement a spring and mass respectively. Part (c) presents a second order example corresponding to a mass-spring combination.

Naturally all three satisfy the passivity conditions (6.4), as well as the position tracking condition (6.10). They also share the following characteristics:

- The transmission paths consist of low pass filters. This implies that the controller isolates the two sides with respect to high frequency noise. Yet in steady state and at low frequencies, signals are passed true without modification. The cutoff frequency is determined by the spring constant or mass.
- The reflective paths are high pass filters. In steady state or at low frequency they have no effect. But during the transients at high frequency, they give each side a local feedback and create a local appearance.
- The cutoff frequencies are the same, so that the wave signal is effectively split between the transmission and reflection path. This is also a consequence of the passivity conditions.

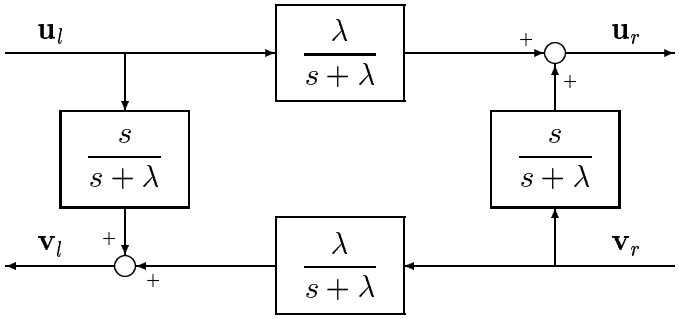
Indeed many wave controllers will share this initial local appearance and eventual transmission form. The transmission path determines the tightness of the coupling between the two sides. And the reflective path shapes the local appearances, whether inertial or spring like. Indeed from Sections 4.5 and 4.6 we know that a positive transfer function is typical of a spring, while a negative transfer function is indicative of a mass. The above examples verify this.

Other wave controllers may place band-limited filters into the transmission to isolate particular frequency ranges. This could be beneficial if one side is vibrating. Or the controller may wish to select only certain frequencies containing the most important information. Nonlinear components are naturally acceptable.

Finally notice that the basic examples do not include any damping. Such dissipation fundamentally reduce the wave signals and the power they provide. With damping, the transmission path will not reach a unity value even in steady state. And the reflective path may provide some continual (low frequency) feedback to establish an appropriately damped appearance.

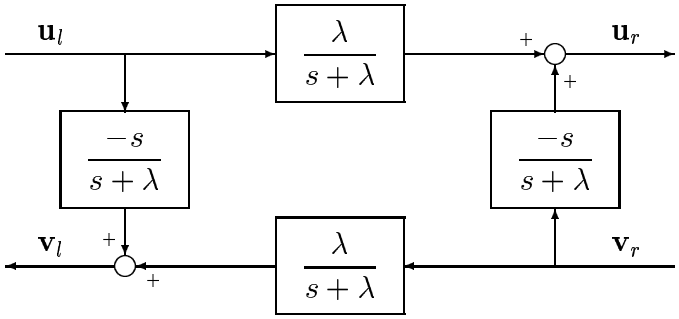
(a) Spring

$$\lambda = \frac{2K}{b}$$



(b) Mass

$$\lambda = \frac{2b}{m}$$



(c) Mass/Spring

$$\lambda = \frac{b}{m} = \frac{K}{b}$$

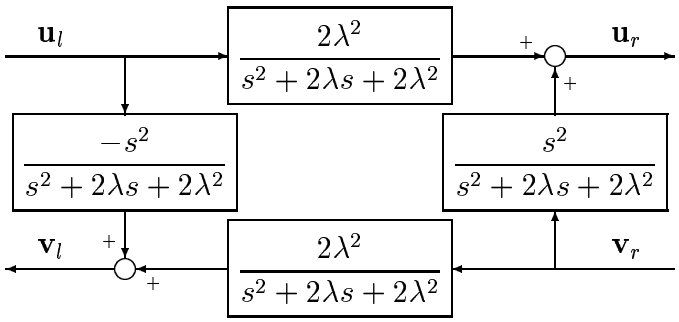


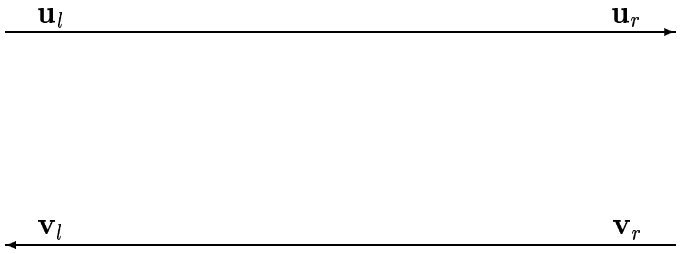
Figure 6-5: These basic wave controllers implement the behavior of classic elements. The sign of the reflection determines the local appearance.

6.4.2 Incorporating Unmodeled Dynamics

Consider a classic robot or telerobot system using passive P.D. control. In theory, such a system is stable regardless of the position or velocity feedback gains. Of course in practice, increasing these gains beyond some bound while cause instability. This because the system starts to interact with the unmodeled dynamics present in every physical system.

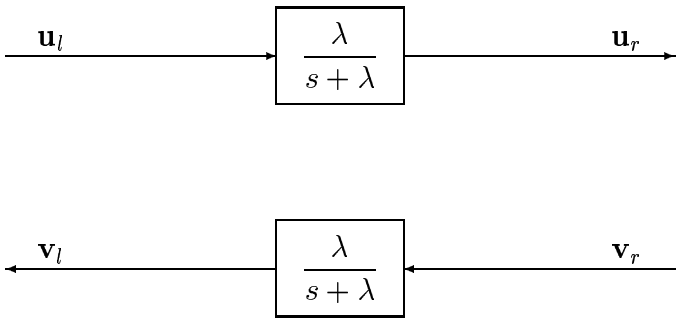
In general it is argued that the signals within the closed loop system should not contain any frequency components that range beyond known limits. These limits are given by: (a) the actuator or power amplifier bandwidth, (b) the sampling rate, (c) structural modes and dynamics (Slotine and Li, 1991). But the frequency components are determined by the location of the closed loop poles, which are in turn determined by the system parameters and gains. And so with knowledge of a robot inertia, we can estimate bounds on its position and velocity gains.

Now let us reconsider these arguments in the wave domain. First imagine a perfect wave controller.



It simply passes the wave commands directly and immediately from the left side to the right side and vice versa. Also no reflective components are needed as transmission is instantaneous. In essence this controller has infinite bandwidth.

Of course this may allow signals to contain very high frequency components in violation of the above mentioned limits. To enforce the limit, we can insert filters with appropriate cutoff frequencies.



The parameter λ can be chosen directly from the known system limitations.

But implementing any wave controller on a physical system also requires a wave transformation and the specification and tuning of the wave impedance b. The transformation contains a local feedback loop to create the apparent damping (see Section 4.1) of magnitude b. Consequently, we may have to limit this value. For example, when controlling a robot or some inertia, we require

$$\frac{b}{m} \leq \lambda_{limit} \quad \Rightarrow \quad b \leq m\,\lambda_{limit} \tag{6.27}$$

In essence the tuning process is reduced to a single gain, as compared to the traditional power variable approach.

Experimental tests verify that this wave controller indeed behaves much like a P.D. controller with gains close to their limits, providing near optimal performance.

6.4.3 Power to Wave Domain Conversions

As we have noted before, every system which is expressed in the power variables $(\dot{\mathbf{x}}, \mathbf{F})$ can also be written in terms of the wave variables $(\mathbf{u}, \mathbf{v})$. It may be helpful to convert existing controllers or elements from one domain to the other. In the following, we present come basic conversions of simple impedance and admittance transfer functions, as well as scaling elements.

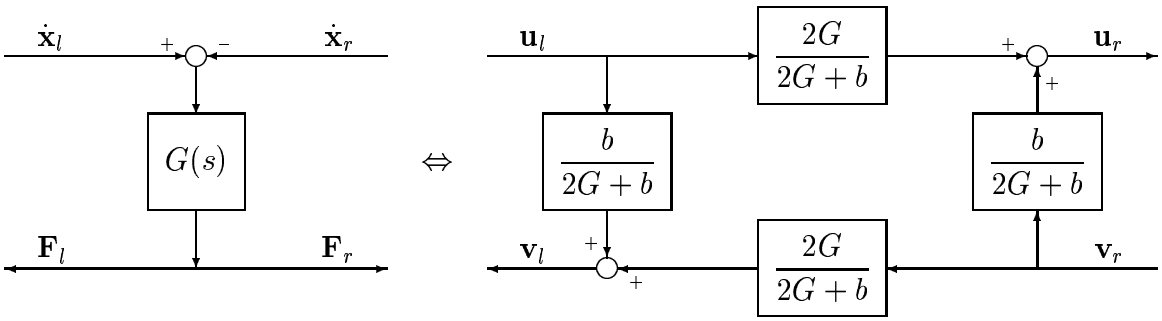
6.4.3.1 Pure Impedance Elements

An impedance element will accept velocity information from both sides and compute the force outputs. In many cases only the relative motion is relevant and both output

forces are equal and determined by the velocity difference.

$$\mathbf{F}_l = \mathbf{F}_r = G(s)\left(\dot{\mathbf{x}}_l - \dot{\mathbf{x}}_r\right)$$

This is equivalent to a 0-junction in bond graphs. The element is converted to the wave domain via



Or in the concatenation form relating both left waves to both right waves, it is given as

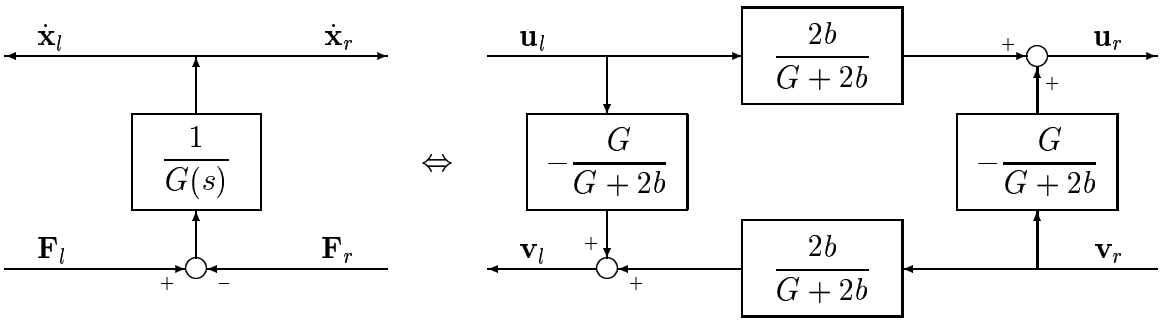
$$\begin{bmatrix} \mathbf{u}_r \\ \mathbf{v}_r \end{bmatrix} = \begin{bmatrix} 1 - \dfrac{b}{2G} & \dfrac{b}{2G} \\ -\dfrac{b}{2G} & 1 + \dfrac{b}{2G} \end{bmatrix} \begin{bmatrix} \mathbf{u}_l \\ \mathbf{v}_l \end{bmatrix} \tag{6.28}$$

6.4.3.2 Pure Admittance Elements

An admittance element determines the velocity output for both sides based on the relative forces. It may be described by

$$\dot{\mathbf{x}}_l = \dot{\mathbf{x}}_r = \frac{1}{G(s)}\left(\mathbf{F}_l - \mathbf{F}_r\right)$$

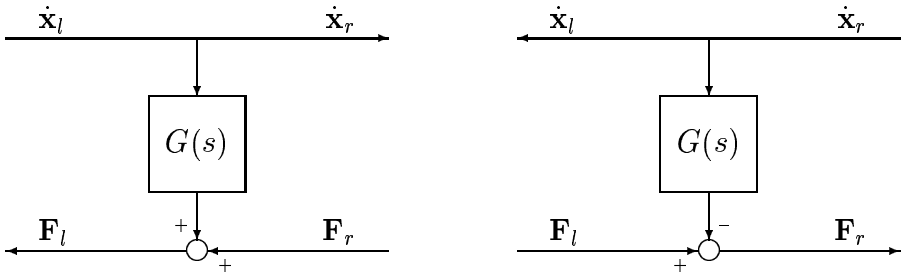
This is equivalent to a 1-junction in bond graphs. The conversion to the wave domain takes the shape



Notice that the roles of $G(s)$ and b are inverted as compared to the impedance above, as well as the sign of the reflections. Both can be attributed to the switched input-output relation. The concatenation form is

$$\begin{bmatrix} \mathbf{u}_r \\ \mathbf{v}_r \end{bmatrix} = \begin{bmatrix} 1 - \frac{G}{2b} & -\frac{G}{2b} \\ \frac{G}{2b} & 1 + \frac{G}{2b} \end{bmatrix} \begin{bmatrix} \mathbf{u}_l \\ \mathbf{v}_l \end{bmatrix} \tag{6.29}$$

Also note that the same wave domain representation is valid for the following two configurations, which are both alternative 1-junction layouts.



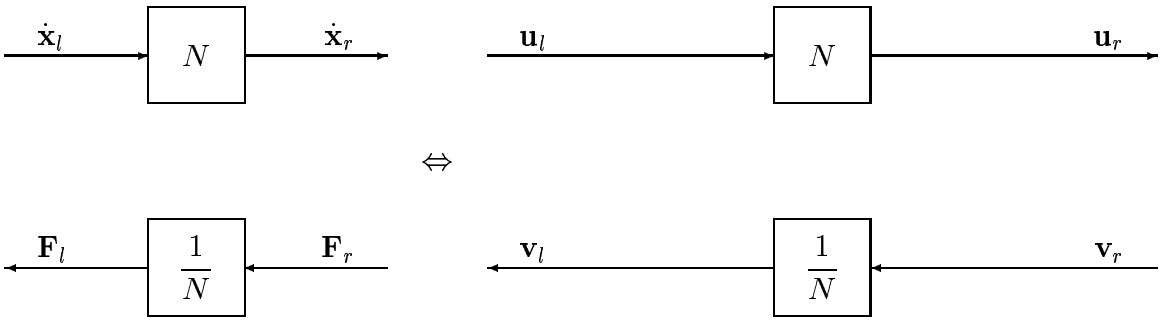
6.4.3.3 Power Scaling

While power scaling is not passive, we have seen before that it can be incorporated into the passivity framework because the two sides are only connected via this single path. The scaling factor N must be constant and equal in all degrees of freedom,

with

$$\dot{\mathbf{x}}_r = N\,\dot{\mathbf{x}}_l \qquad \text{and} \qquad \mathbf{F}_l = \frac{1}{N}\,\mathbf{F}_r$$

We find this simple relation also holds in the wave domain



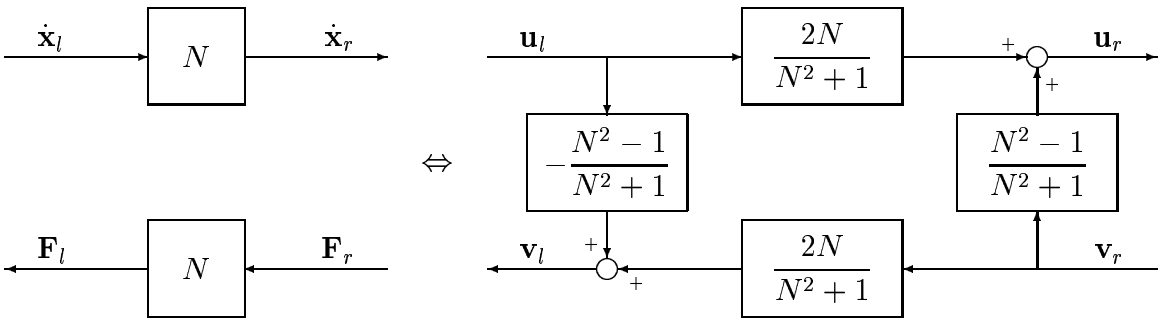
where the forward power flow is amplified while the reverse direction contains the appropriate decrease. In particular no reflections are introduced.

6.4.3.4 Dimensional Scaling

For dimensional scaling the power is conserved and the scaling constant N is equal for both velocity and force.

$$\dot{\mathbf{x}}_r = N\,\dot{\mathbf{x}}_l \qquad \text{and} \qquad \mathbf{F}_l = N\,\mathbf{F}_r$$

This leads to a more complex wave domain behavior including reflections



Indeed the effect of dimensional scaling is the same as that of changing the wave impedance b. After all, the impedance provides a relative scale between velocity and force. So as waves pass through the scaling element, they are partially reflected much as though the wave were to encounter different wave impedances.

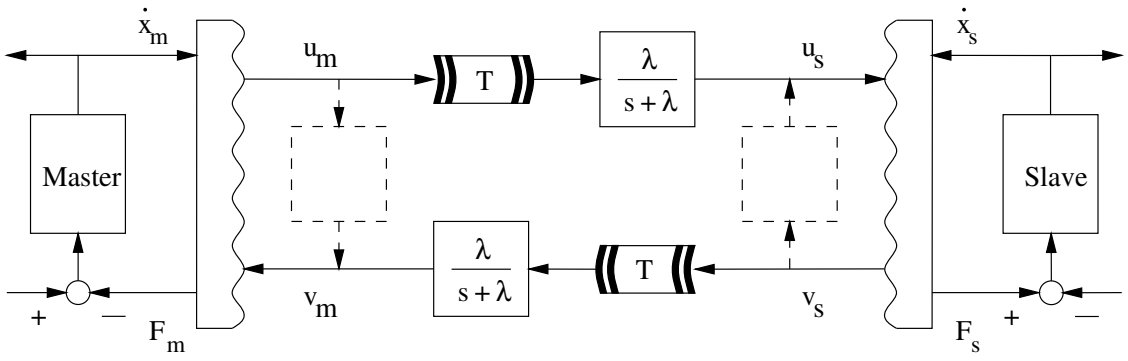


Figure 6-6: Designing the system entirely in wave space removes the need for both master and slave P.D. controllers.

6.5 Pure Wave Teleoperator

As the logical extension of wave controllers, we finally propose a purely wave based teleoperator, as shown in Figure 6-6. It completely replaces all P.D. and other controllers in the teleoperator with systems designed directly and entirely in wave space. In the simplest form, these are just the wave filters introduced in Section 6.3.

The local wave feedback components may be used to shape the initial (high frequency) responses. Without these elements the response would appear as pure damping of magnitude b. If this damping exceeds acceptable amounts, a high pass filter could limit initial forces like motion in free space.

Of course, both the filters in the transmission and the local feedback components may be higher order as well as nonlinear. To assure passivity they must satisfy the condition (6.5) presented earlier.

Having eliminated the P.D. controller, the wave transformations now directly drive the physical hardware. This has the benefit of removing the controller stiffness. Indeed this system is characterized by the steady state parameters

$$M = M_m + bT + \frac{b}{\lambda} + M_s \tag{6.30a}$$

$$B = 2b \tag{6.30b}$$

$$K^{-1} = \left(\frac{b}{T}\right)^{-1} + (b\lambda)^{-1} \tag{6.30c}$$

On the downside, the impedance matching is no longer possible. So this configuration is best suited for systems with unpredictable responses, where the matching does

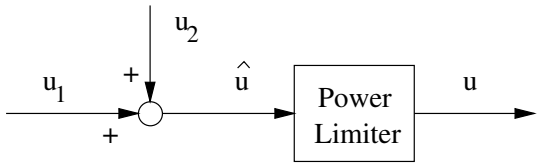


Figure 6-7: Wave junctions combine two wave signals, while making sure the output power does not exceed the input power.

a poor job of eliminating wave reflections. Instead the filtering takes over completely.

But regardless of the teleoperator design, the wave impedance still remains the tunable parameter, which can select between a slow and stiff system versus a light and soft setup.

6.6 Wave Junctions

Wave junctions provide an alternative to 0-junctions and 1-junctions, as defined by bind graphs, for combining command and possibly feedback signals between various sites. For example two operators may assist each other while performing some task.

The biggest concern is again passivity. Each wave signal is self contained, carrying both signal and power. The wave junction must hence combine not only the command signals, but also assure that the associated power is conserved. This may require some modulation of the junction output.

Figure 6-7 sketches the layout of a wave junction. The two wave signals are combined into a desired output value

$$\hat{u} = u_1 + u_2$$

This is filtered or modulated to create the actual output value u and assure that

$$\int_0^t u^2 \, \mathrm{d}\tau \leq \int_0^t u_1^2 + u_2^2 \, \mathrm{d}\tau$$

To implement the modulation process, we define the 'distance' D as the integrated difference between the desired and actual output waves.

$$D = \int_0^t u_1 + u_2 - u \, \mathrm{d}\tau \tag{6.31}$$

Also, we define an energy reservoir E as the amount of input energy yet unused by the output.

$$E = \int_0^t u_1^2 + u_2^2 - u^2 \, \mathrm{d}\tau \geq 0 \tag{6.32}$$

The object of the modulation is to force the distance D to zero while respecting the available energy E. Though numerous procedures are possible, we typically want to reduce D in the shortest time and hence provide the least amount of lag. Selecting

$$u = \frac{E}{D} \tag{6.33}$$

uses the energy in the reservoir at a constant rate to reach the desired output integral as fast as possible.

If the inputs combine to reduce D themselves, excess energy may remain after the convergence is complete and D has reached zero. This energy is removed from the reservoir and thereby dissipated.

6.7 Wave Predictors

Using a priori knowledge, a predictive element may help the operator overcome the delay during normal operation speed up the task. However, a predictor cannot react to unknown or disturbance inputs until after the delay. In fact, it may even worsen the system behavior under such circumstances. And the environment of telerobot is often unknown and the contact forces are very difficult to model accurately. So we do not believe prediction will significantly impact or benefit force reflection. However, it may assist other delayed systems with good models.

Executing predictors in the wave domain can provide stability guarantees even when the model is inaccurate. This because the global power flow is controlled explicitly to assure passivity. However, it also requires that the model itself be passive.

Figure 6-8 shows the general layout. Like other predictors, such as the Smith predictors described in Appendix A, the ideal wave response is given by the prediction output and a delayed prediction error feedback.

$$v_d(t) = v_p(t) - \Big[v_p(t-2T) - v_a(t-T) \Big] \tag{6.34}$$

where $v_p(t–2T)$ and $v_a(t–T)$ are the delayed predicted and actual responses. However, this desired output must be filtered or modulated to limit its power requirements to the currently available energy.

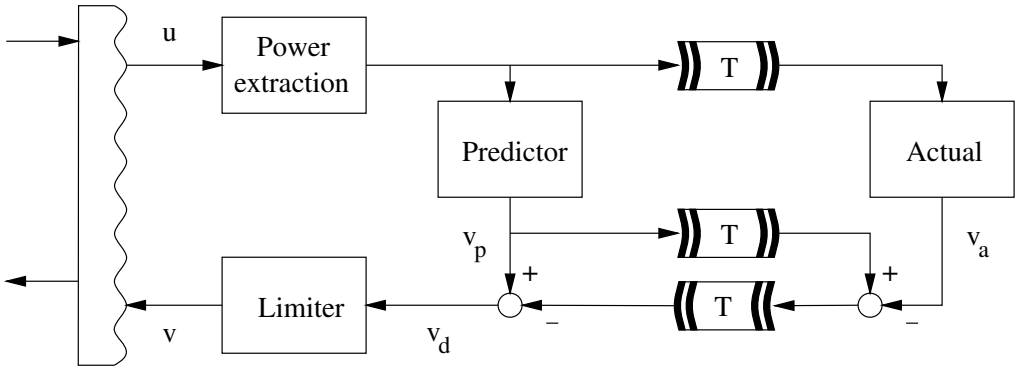


Figure 6-8: Wave predictors feed back the predicted response as well as a delayed error signal like other prediction methods. But they also include an energy reservoir, which is fed from the input signal and used to provide the prediction power needs.

For the purpose of this modulation, we define an energy reservoir

$$E_{reservoir} = \int_0^t v_a^2 \, \mathrm{d}\tau + \int_0^t P_{extracted} \, \mathrm{d}\tau - \int_0^t v^2 \, \mathrm{d}\tau \tag{6.35}$$

which is fed by the actual returned power as well as the power extracted from the input command and drained by the output power to the operator. We also define the distance D as the integrated error between the desired and actual output waves.

$$D = \int_0^t v_d - v \, \mathrm{d}\tau \tag{6.36}$$

The actual modulation then selects the output v to force the distance D to zero while respecting the available energy. With

$$v = \frac{E_{reservoir}}{D} \tag{6.37}$$

the distance is reduced to zero in the least amount of time for the available energy.

With a good model, the actual return power should continually replenish the energy spent, maintaining an appropriate reservoir. But for bad models and during initial startup operations, the energy reservoir may be too low to allow the predicted response to reach the operator. In these cases, additional power is extracted from the input wave command via some form of filtering. Most any type may be used, which removes some frequency components and thereby some power.

In general, a wave predictor should initially feel damped until the reservoir has

been established. Thereafter the process should be self supporting and provide the operator with the direct feedback from the model. In all cases stability is maintained.

Chapter 7

Concluding Remarks and Contributions

In the preceding chapters we have addressed the problem of force reflecting teleoperation in the presence of significant time delays. Force feedback is generally considered to benefit the human operator, but it has often been avoided due to the inherent stability problems caused by the delay. While the basic stability solution has been known for a few years, a practical system with acceptable performance for large delays of several seconds has been missing. Developing the notion of wave variables, we have been able to overcome these problems and establish a system which is robust to arbitrary delays and provides reliable and predictable operation.

In the following, we summarize the contributions of this work, as well as the various options for the working teleoperator system. We also briefly discuss some extensions and other applications to which we hope this work may lead.

7.1 General Contributions

Most fundamentally, we have introduced the notion of wave variables which provide inherent robustness to delays. Based solely on the general theory of passivity, wave variables are particular combinations of velocity and force signals which require very little additional computation and are applicable to a very large set of systems including nonlinear, discrete, and many others. In every case, encoding the information in this manner will ensure robustness to arbitrary delays, known or unknown, and will decrease the global performance as the delay increases.

Using the wave variables, we have proposed various teleoperator configurations

which address the transient dynamics and behavior introduced by the delay and provide the desired performance. We find that for small delays below or near the human reaction time, the system is nearly transparent and allows operations to continue at normal speeds.

For large delays, wave reflections may disrupt the normal feedback signals. They can be eliminated or reduced with appropriate tuning methods, such as impedance matching and/or wave filtering. As a guideline for the tuning process we have suggest the concept of a virtual tool, which summarizes the best behavior allowed by the delay. Indeed the delay adds both inertial and capacitive effects according to

$$M = bT \qquad\qquad K = \frac{b}{T} \tag{7.1}$$

where b is the tunable wave impedance and T is the actual delay.

Notice the tradeoff between high inertia and high stiffness versus low inertia and low stiffness. This selection can provide either quick and easy motions (low inertia) or good and clear force resolution (high stiffness). This selection can be adjusted during operations to optimize for the various tasks.

Also notice the implicit reduction caused by the delay. Large values of T will both increase the inertia and reduce the stiffness, making the system slower and more sloppy. This is the fundamental limitation of delayed feedback and appears automatically in this framework.

In addition, we have proposed some design procedures entirely within the wave domain. The wave signals provide a different point of view, which may lead to a better understanding of the physical behavior of a system and alternatives not readily apparent in the traditional notation. For example, wave junctions may be applicable to shared control between multiple operators, while wave predictors may provide a robust method for including complete a priori information.

The developments were tested using two robots with three degrees of freedom each. The results support our objective of creating a reliable and predictable system which can tolerate and operate with delays of at least one second.

7.2 Teleoperator Options

The developments presented several teleoperator systems, which have individual benefits and possible limitations. In particular, the work made the following basic steps:

(a) Most importantly, wave transformations should be added to any existing system. Without delays they make no changes. Yet they provide robustness against any possible delays and automatically limit performance if the delays are substantial. The transformations are essential to provide the needed stability guarantees.

In many cases with small delays this may already be sufficient to achieve stability and performance. However, larger delays quickly introduce wave reflections that corrupt the useful flow of information.

(b) To reduce reflections we may tune the controller gains. While these simple adjustments cannot eliminate all reflections, any reduction also decreases the level of distraction which the reflections present to the operator. Again this is most likely to work well for small delays.

(c) When the delays increase beyond a half second, any remaining reflections will appear and hit the operator at times much separated from the original inputs. This effect is quite disorienting, so that we suggest using the new impedance matched configuration of Section 5.5. Without reflections, the operator can concentrate on the true interaction signals.

As noted before, the matching process is performed for a particular expected behavior. If the system is not modeled well or comes into contact with additional complex dynamics, reflections will again appear. While this does not interfere with stability it may again limit the practical usefulness.

(d) To overcome the impedance matching limitation, the final configurations uses wave filters to smooth out the reflections. This works well without knowledge of the environment and for varying tasks, but also filters the useful information in the feedback path. As a result, the final system will behave nicely and predictably, but may limit performance more than fundamentally necessary.

To retrieve the lost information, it may be possible to redirect the high frequency components via other feedback channels, for example audio displays. Contact signals would then appear as impact noise providing the operator with the additional cues.

In general, the added controller elements introduced by either impedance matching or wave filtering provide better behavior when the system is subjected to large delays. The impedance matching is preferred when good models are available, whereas the wave filtering requires no model. But both may also limit performance if the system is not understood well or has only small delays.

For example, if we expect only free space motions and rigid contacts, impedance matching may be best suited, accompanied by appropriate real time adjustments of the tunable wave impedance.

7.3 Other Applications and Future Directions

As wave variables are based solely on the theory of passivity, we hope they will find application in other areas which also involve bidirectional information flow and possible time delays.

Networking Increasingly communications are being handled using packet switched networks, such as the internet. The biggest added difficulty is the random delay which individual signals may encounter. Fortunately wave variables can be adjusted to handle these variations, while typical delays of 0.5 seconds fall nicely into their applicable range. Besides providing the communications for teleoperation, one might envision interactions with remote or distributed virtual worlds.

Shared Control Building on wave junctions may allow shared control of a robot between multiple operators. For example a remote expert may watch over and assist a local technician during difficult tasks. Besides the obvious robustness to delays, wave junctions provide an alternative to the traditional 0 and 1 junctions defined in Bond Graphs.

Biological Systems Biological systems, in particular the human nervous system, demonstrate both closed loop control and substantial time delays. As suggested in (Massaquoi and Slotine, 1996), wave variables may play a role in maintaining the system stability.

Appendix A

Smith Predictors

All predictors use a priori information in the form of a system model, as well as the current system state to estimate the future outputs or inputs needed to achieve some goal. Within a closed loop, they represent a feedforward controller, which improves the response to known or predetermined inputs. Unfortunately they have no effect on the feedback path and cannot help reject unknown disturbance inputs, as these are only detectable after the delay. Not surprisingly, predictors perform best when the model is accurate and the system experiences few disturbances, so that it may be controlled mostly via feedforward terms.

Smith predictors represent the classical method for including prediction within the LTI domain. They implement a linear compensator which attempts to cancel the actual dynamics and replace them with a desired set. They were first suggested and developed by (Smith, 1957, 1959).

A.1 Definition

Consider a basic linear time-invariant system, consisting of a non-delayed transfer function $G(s)$ and a pure delay e^{-sT}.

$$G_{OL}(s) = G(s)e^{-sT}$$

Closing a negative feedback loop around this system, as shown in Figure A-1, the closed loop transfer function becomes

$$G_{CL}(s) = \frac{G(s)e^{-sT}}{1 + G(s)e^{-sT}}$$

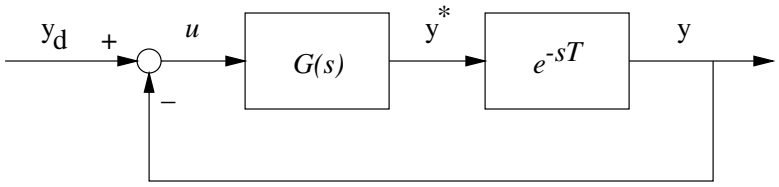


Figure A-1: Basic feedback loop including a pure delay.

In this case the maximum tolerable delay T_{max} before the system goes unstable is given by the phase margin PM divided by the crossover frequency ω_c of $G(j\omega)$.

$$T_{max} = \frac{PM}{\omega_c}$$

To enhance the stability, it would be desirable to measure the undelayed output y^* of $G(s)$ directly before passing through the delay. Using this signal for the feedback loop, instead of the normal output y, would isolate the delay outside the feedback loop and thus eliminate it from the stability issues.

Unfortunately, in many real systems it is not possible to isolate the delay in this fashion. Instead, it may be possible to simulate the real system and predict what the undelayed output should be. This leads to the following layout for Smith predictors, as is depicted in Figure A-2.

The main feedback loop is driven by the predicted value y_p^* of the undelayed output, rather than any real measurements. This local control loop thus has the desirable stability properties. The input to the simulated system is then also applied to the real system. Assuming good prediction, the real system will mimic the simulation and its good behavior. To ensure that these two systems do not diverge, the simulation error measured between the two delayed outputs is also feed back. This secondary feedback loop necessarily includes the delay, as it includes real measurements, but should only provide a weak correction term.

To avoid the multi-loop layout, Smith predictors can also be expressed in the form of a compensator, as seen in Figure A-3, and given by

$$G_c(s) = \frac{1}{1 + G_p(s) - G_p(s)e^{-sT_p}}$$

While this may hide the original intentions, it simplifies the system structure and analysis.

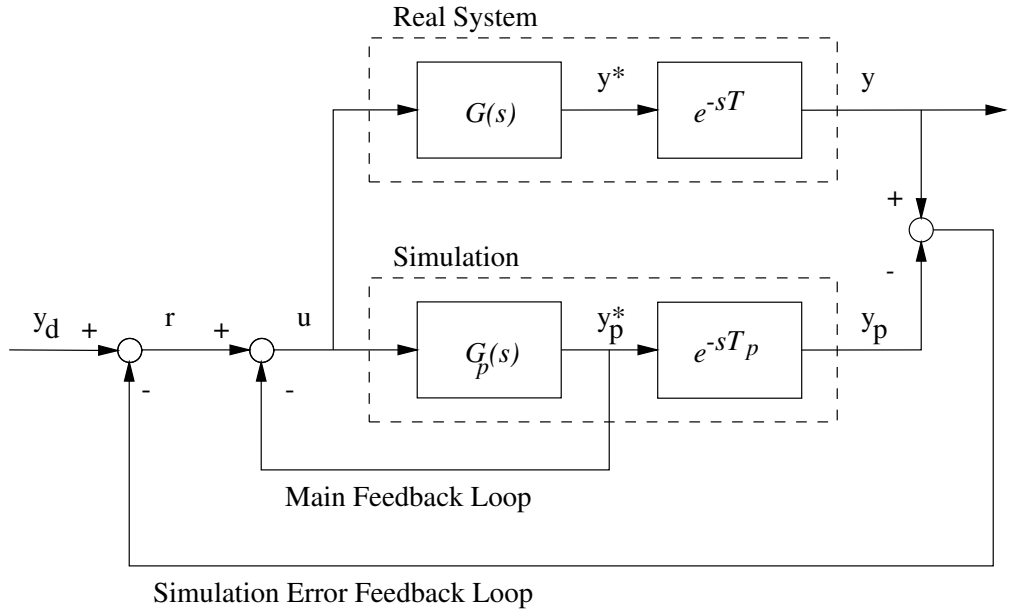


Figure A-2: Smith predictor layout including the real and simulated system, as well as the main and simulation error feedback loops.

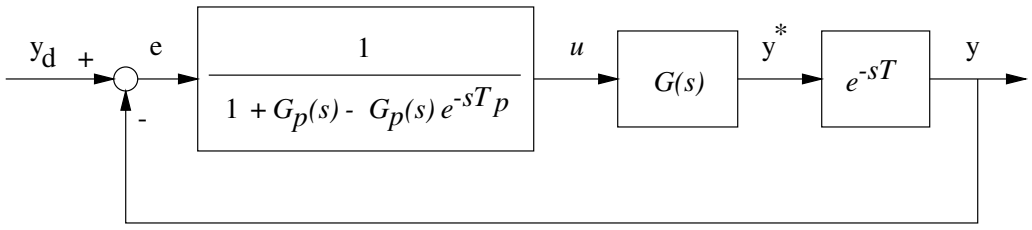


Figure A-3: Smith predictor expressed in the form of a compensator.

In either form, the closed loop transfer function for a Smith predictor is

$$G_{CL}(s) = \frac{G(s)e^{-sT}}{1 + G(s)e^{-sT} + G_p(s) - G_p(s)e^{-sT_p}}$$

where $G_p(s)$ is the predicted transfer function and T_p is the predicted delay.

Notice that the denominator of $G_{CL}(s)$, which provides the characteristic equation, now contains the additional terms due to the prediction. Should the predicted transfer function and delay match the actual, then all delay operators vanish from this expression leaving the desired form

$$G_{CL}(s) = \frac{G(s)e^{-sT}}{1 + G(s)}$$

However, when the prediction is significantly different from the actual system, the additional terms may in fact cause problems.

Furthermore, while command following is significantly improved, any disturbance rejection necessarily remains poor, as external influences can still only be detected after the delay.

A.2 Robustness

To examine the effects of Smith predictors consider the following stable first order system with a delay:

$$G_{OL}(s) = K \frac{a}{s+a} e^{-sT}$$

The Smith predictor for this system, expressed as a compensator, is given by

$$G_c(s) = \frac{1}{1 + G_p(s) - G_p(s)e^{-sT_p}} \qquad \text{with} \qquad G_p(s) = K_p \frac{a_p}{s + a_p}$$

where K_p, a_p, T_p are the estimates of the system parameters. The following tests use $K_p = 5$, $a_p = 10$, and $T_p = (0, 0.1, 0.2, 0.5, 1.0)$ for various compensator configurations as well as varying $K = 5$, $a = 2 \ldots 50$, and $T = 0 \ldots 1.2$ for the actual system. By changing the actual pole location a and the actual delay T we can verify robustness of the system.

Figure A-4 shows the regions of actual parameter values, for which the closed loop system is stable. This is repeated using four different compensators corresponding to different values of T_p. Note that $T_p = 0$ collapses to the case where no compensator

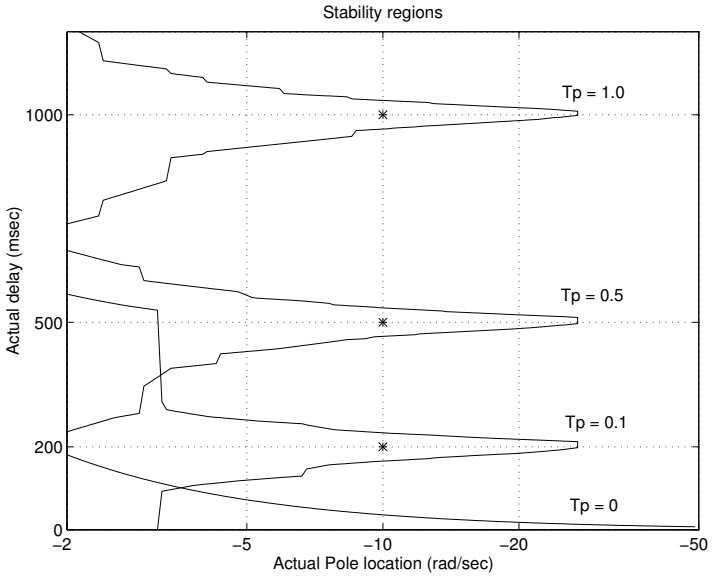


Figure A-4: Stable regions of actual pole location and actual delay for the first order system using different compensator configurations. Note $T_p = 0$ is equivalent to no predictor and the nominal parameters are marked with stars.

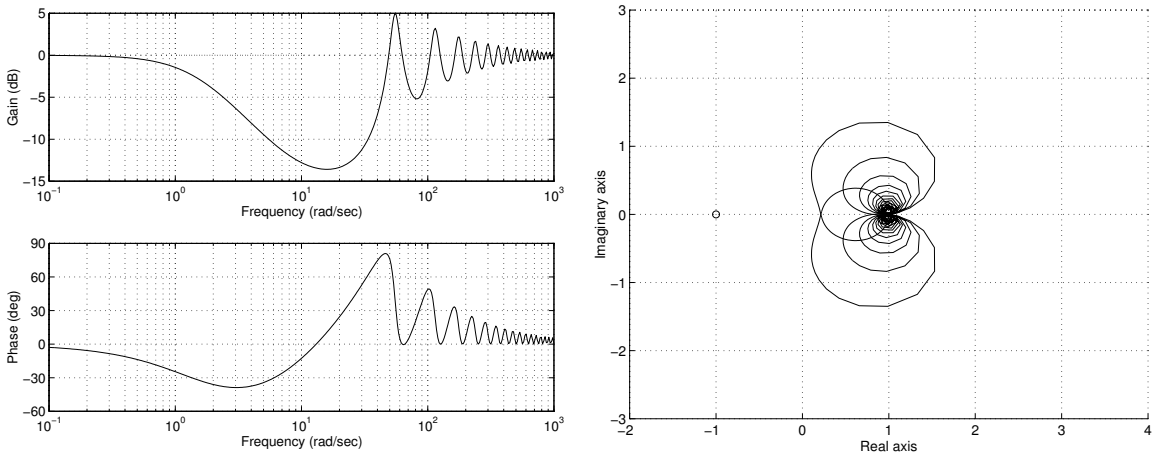
is used. Indeed the Smith predictors perform as advertised, creating a stable region around each operating point. Similar plots are generated by varying the gains rather than the pole location. Notice that the regions expand as the actual pole becomes slower and vanish as the pole becomes faster. However, they do not significantly change shape as a function of the delay.

The compensators themselves are examined in Figure A-5, which shows a Bode and Nyquist plot for the compensator with $T_p = 0.1$. In particular, the exponential (i.e. delay) term in the compensator transfer function creates strong oscillations in the frequency response.

Finally the Nyquist plot for the entire open loop system, including the compensator, is depicted in Figure A-6. It clearly shows a substantial phase and gain margin, even though it includes the actual delay which is larger than the original phase margin would have allowed.

A.3 Limitations

Unfortunately, Smith predictors cannot be applied to unstable systems. While attempting to cancel the open loop dynamics, the predictors generate pole/zero can-



(a) (b)

Figure A-5: Bode and Nyquist plots for the compensator with $T_p = 0.1$.

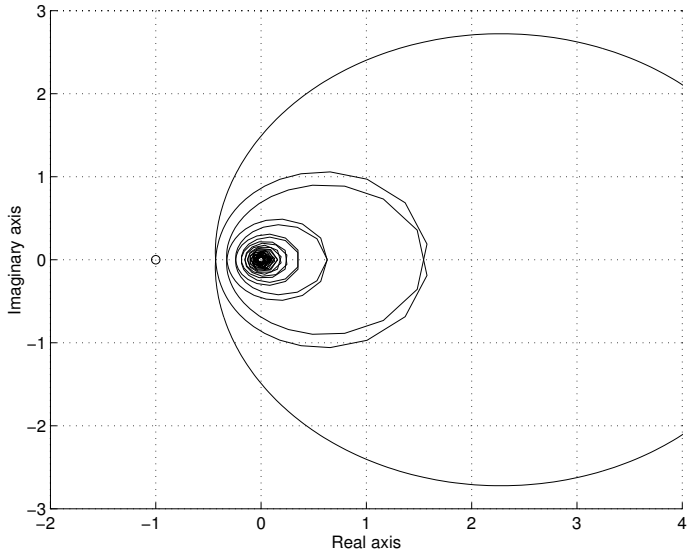


Figure A-6: Nyquist plot for the entire open loop system including the compensator with $T_p = 0.1$.

cellations, which would remain in the right half plane. This limitation can also be understood by realizing that the stabilization process is strongly dependent on the feedback path. Meanwhile predictors try to eliminate this path in favor of feedforward signals. The two goals are thus opposing.

In addition, there are no guarantees for stability when the model is inaccurate. Indeed Smith predictors are not passive. So while they are well suited for many other situations, they do not meet our needs of interacting with unknown passive environments. Instead we suggest the development of wave predictors within the passivity domain.

Bibliography

Anderson, R. J. (1989). *A Network Approach to Force Control in Robotics and Teleoperation*. Ph.D. thesis, University of Illinois at Urbana-Champaign, Department of Electrical and Computer Engineering, Urbana, IL.

Anderson, R. J. and Spong, M. W. (1989a). Asymptotic stability for force reflecting teleoperations with time delay. In *IEEE Conference on Robotics and Automation*, pages 1618–1625, Scottsdale, AZ.

Anderson, R. J. and Spong, M. W. (1989b). Bilateral control of teleoperators with time delay. *IEEE Transactions on Automatic Control*, 34(5):494–501.

Arimoto, S. (1995). Joint-space orthogonalization and passivity for physical interpretations of dexterous robot motions under geometric constraints. *Int. Journal of Robust and Nonlinear Control*, 5(4):269–284.

Asada, H. and Slotine, J.-J. E. (1986). *Robot Analysis and Control*. John Wiley and Sons.

Aström, K. J., Hang, C. C., and Lim, B. C. (1994). A new smith predictor for controlling a process with an integral and long dead-time. *IEEE Transactions on Automatic Control*, 39(2):343–345.

Aström, K. J. and Wittenmark, B. (1984). *Computer-Controlled Systems*. Prentice Hall, Englewood Cliffs, New Jersey.

Baker, Jr., G. A. (1975). *Essentials of Padé Approximations*. Academic Press, New York.

Baker, Jr., G. A. and Graves-Morris, P. (1981). *Padé Approximants, Part I: Basic Theory*, volume 13 of *Encyclopedia of Mathematics and its Applications*. Addision-Wesley, Reading, MA.

Colgate, J. E. (1988). *The Control of Dynamically Interacting Systems*. Ph.D. thesis, M.I.T. Department of Mechanical Engineering, Cambridge, MA.

Colgate, J. E. and Hogan, N. (1988). Robust control of dynamically interacting systems. *Int. Journal of Control*, 48(1):65–88.

Desoer, C. A. and Vidyasagar, M. (1975). *Feedback Systems: Input-Output Properties.* Academic Press.

Górecki, H., Fuksa, S., Grabowski, P., and Korytowski, A. (1989). *Analysis and Synthesis of Time Delay Systems.* PWN—Polish Scientific Publishers, Warsaw, Poland.

Hogan, N. (1982). Mechanical impedance control in assistive devices and manipulators. In Brady, M. et al., editors, *Robot Motion: Planning and Control*, pages 361–371. MIT Press.

Hogan, N. (1985). Impedance control: An approach to manipulation: Parts I–III. *Journal of Dynamic Systems, Measurements, and Control*, 107(1):1–24.

Lawrence, D. A. (1993). Stability and transparency in bilateral teleoperation. *IEEE Transactions on Robotics and Automation*, 9(5):624–637.

Malek-Zavarei, M. and Jamshidi, M. (1987). *Time-Delay Systems: Analysis, Optimization and Applications*, volume 9 of *Systems and Control Series.* North-Holland, Amsterdam.

Marshall, J. E., Górecki, H., Korytowski, A., and Walton, K. (1992). *Time-Delay Systems: Stability and Performance Criteria with Applications.* Mathematics and its Applications. Ellis Horwood.

Martins de Carvalho, J. L. (1993). *Dynamical Systems and Automatic Control.* Prentice Hall, New York.

Massaquoi, S. and Slotine, J.-J. E. (1996). The intermediate cerebellum may function as a wave variable processor. *Neuroscience Letters*, 214:1–5.

Minsky, M. L. (1992). Internal Report.

Mukherjee, S. K. and Dasgupta, S. (1979). The analysis of control systems with transport lag. *IEEE Transactions on Industrial Electronics and Control Instrumentation*, 26(2):116–119.

Niemeyer, G. and Slotine, J.-J. E. (1991a). Shaping the dynamics of time-delayed force-reflecting teleoperation. Technical Report NSL-910501, Nonlinear Systems Lab, MIT.

Niemeyer, G. and Slotine, J.-J. E. (1991b). Stable adaptive teleoperation. *IEEE Journal of Oceanographic Engineering*, 16(1):152–162.

Ogata, K. (1990). *Modern Control Engineering*. Prentice Hall, Englewood Cliffs, New Jersey, second edition.

Palmor, Z. J. and Blau, M. (1994). An auto-tuner for smith dead time compensator. *Int. Journal of Control*, 60(1):117–135.

Pontryagin, L. S. (1955). On the zeros of some elementary transcendental functions. *American Mathematical Society Transactions*, 2(1):95–110.

Pontryagin, L. S. (1958). On the zeros of some elementary transcendental functions — supplement. *American Mathematical Society Transactions*, 2(8):19–20.

Popov, V. M. (1973). *Hyperstability of Control Systems*. Springer Verlag.

Saff, E. B. and Varga, R. S. (1977). On the zeros and poles of Padé approximants to e^z. ii. In Saff, E. B. and Varga, R. S., editors, *Padé and Rational Approximation: Theory and Applications*, pages 195–213. Academic Press, New York.

Salisbury Jr., J. K. (1980). Active stiffness control of a manipulator in cartesian coordinates. In *IEEE Conference on Dynamics and Control*, Albuquerque, NM.

Sheridan, T. B. (1989). Telerobotics. *Automatica*, 25(4):487–507.

Sheridan, T. B. (1992). *Telerobotics, Automation, and Human Supervisory Control*. The MIT Press, Cambridge, MA.

Sheridan, T. B. (1993). Space teleoperation through time delay: Review and prognosis. *IEEE Transactions on Robotics and Automation*, 9(5):592–606.

Slotine, J.-J. E. and Li, W. (1991). *Applied Nonlinear Control*. Prentice Hall, Englewood Cliffs, New Jersey.

Slotine, J.-J. E. and Niemeyer, G. (1993). Telerobotic system. U.S. Patent #5266875.

Smith, O. J. M. (1957). Closed control of loops with dead time. *Chemical Engineering Progress Transactions*, 53(5):217–223.

Smith, O. J. M. (1959). A controller to overcome dead time. *ISA Journal*, 6(2):28–33.

Suh, I. H. and Bien, Z. (1982). A root-locus technique for linear systems with delay. *IEEE Transactions on Automatic Control*, 27(1):205–208.

Thowsen, A. (1981). An analytic stability test for a class of time-delay systems. *IEEE Transactions on Automatic Control*, 26:735–736.

Williamson, S. E. (1969). Accurate root-locus plotting including the effects of pure time-delay. *Proceedings of the Institution of Electrical Engineers*, 116(7):1269–1271.

Yamanaka, K. and Shimemura, E. (1987). Effects of mismatched smith controller on stability in systems with time-delay. *Automatica*, 23(6):787–791.