Spatial and Temporal Considerations in Vehicle Path Tracking With an Emphasis on Spatial Robustness

Donald L. Cripps
Utah State University

SPATIAL AND TEMPORAL CONSIDERATIONS IN VEHICLE

PATH TRACKING WITH AN EMPHASIS ON

SPATIAL ROBUSTNESS

by

Donald L. Cripps

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

Approved:

_____
Dr. Kevin Moore
Major Professor

_____
Dr. Matthew Berkemeier
Committee Member

_____
Dr. R. Rees Fullmer
Committee Member

_____
Dr. Robert Gunderson
Committee Member

_____
Dr. Todd Moon
Committee Member

_____
Dr. David Peak
Committee Member

_____
Dr. Thomas Kent
Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2003

Abstract

Spatial and Temporal Considerations in Vehicle Path

Tracking with an Emphasis on Spatial Robustness

by

Donald L. Cripps, Doctor of Philosophy

Utah State University, 2003

Major Professor: Dr. Kevin Moore
Department: Electrical and Computer Engineering

This dissertation researches the task and path management of an autonomous vehicle with Ackerman-type steering.

The task management problem was approached as a path training operation in which a human operator drives the desired path through an environment. A training trajectory is converted into a series of path segments that are driveable by the autonomous vehicle by first fitting a general path to the dataset. Next, transition segments are added to the general path to match the vehicle velocity and steering angle rate limit.

The path management problem has been approached by first deriving a kinematic model of the vehicle. The time domain model is expressed in the frequency domain and then converted into a spatial frequency domain. Next, a stability criterion is derived and used in the synthesis of a spatially-robust path controller.

(176 pages)

To my wife, without whom no dreams are possible.
To my son, who helps me see the world through new eyes.
To my mom and dad, who taught me to love learning.

Acknowledgments

I would like to acknowledge the help, direction, and kindness of both of my graduate advisors, Dr. Moore and Dr. Gunderson. Also, I would like to thank the rest of my committee for their inspiration and help.

I would also like to thank Mel Torrie and my other coworkers at Autonomous Solutions. Their potential drives me to improve. Lastly, I would like to thank my first mentor, Dr. Robert Redlich, who taught me that simplicity is worth the struggle.

Donald L. Cripps

Contents

List of Figures

Chapter 1
Introduction

The field of autonomous ground vehicles has expanded rapidly during the last decade. A description of the progress in the field is illustrated by two position papers. The first paper was written in 1993 [1] and describes the transition from tele-operation to autonomous operation with the second paper written in 2001 [2] describing the current state of autonomous vehicle research. The expansion has largely been made possible due to the availability of accurate position measurement systems such as the Global Positioning System (GPS). Vehicle position can be determined to sub-meter accuracy making roving, autonomous ground vehicles an achievable reality. The possible uses for autonomous vehicles have only begun to be explored. Candidates for automation include tedious, repetitive, and dangerous tasks that are now performed by human operators while driving a vehicle. For example, the task of spraying pesticides on a crop is tedious, but requires an alert operator to avoid crop damage and to effectively use the chemical sprays. Additionally, certain pesticides are harmful to humans so the operator must be appropriately protected and the protection systems can be expensive and uncomfortable. However, before this and other applications of autonomous vehicles can be realized technical advances must be made in several key areas. In particular, the designers of autonomous vehicles must consider areas such as vehicle safety, task management, path management, vehicle function, and system reliability. The widespread use and acceptance of autonomous vehicles depends on excellent performance in all of these critical areas.

Of the five areas described above, task management and path management are particularly required for the vehicle to perform useful work. Autonomous vehicles

follow a path while performing a useful task and the management functions are those performed to ensure that the task is completed correctly. The task, in the case of an autonomous vehicle, is defined as the completion of useful work done while following an appropriate path with a vehicle. Therefore, the task consists of a sequence of path segments plus a sequence of whatever work the vehicle performs while driving. Task management consists of the supervisory functions that plan the execution of the desired task, prescribe the path, perform the desired work along the way, and determine what remains to be done in order to complete the assigned task. Path management consists of the supervisory functions performed to ensure that the vehicle is following the desired path. Hierarchically, task management oversees path management, given that the task manager determines the level of task completion and provides path updates to the path manager.

The task management function varies widely with the task to be performed and is not unique except to a family of similar problems. Path management is more generic to the common problem of guiding the autonomous vehicle through an environment on an assigned path. Given the breadth of possible task management schemes, the focus of this research is narrowed to a single type of task management and to the more general problem of path management.

1.1 Task and Path Management as a Control System

Given a task to be accomplished by an autonomous vehicle, the combination of task and path management enables the vehicle to complete the assigned task. As an example, consider an autonomous tractor given the task of harvesting a forage crop. A multi-segmented path is defined that allows the tractor to travel back and forth on the field in such a manner as to completely mow the crop. The segments of the path could consist of 'straight' and 'curved' portions. The tractor is started on the first path segment and observations and corrections are made to ensure that the

tractor stays on that path segment. As one path segment is completed, the tractor is commanded to a new segment and the tractor is again held to that segment. Finally, when the field is mowed, the tractor stops.

The language of the example in the previous paragraph leads naturally to a feedback structure to model the process of task and path management. Given the assigned task of cutting the crop in the field, an action is planned to accomplish the task. The vehicle is checked periodically while performing the assigned task to ensure that the correct path is being followed, and to determine if a new path segment should be commanded based on the current level of task completion. Eventually, the task is recognized as completed and the process stops. Checking the system to determine correctness and the level of task completeness implies a feedback structure. A block diagram of the task and path management system for an autonomous vehicle is given in Figure 1.1. The input to the system is the assigned task and the output is a measure of the level of task completion defined as the state of the task.

The block diagram of Figure 1.1 represents vehicle path management as a control system with two feedback loops. The outer feedback loop provides a measurement of the level of task state, which when compared to the complete task generates the path segment that the vehicle must follow in order to complete the desired task. The inner feedback loop provides a measurement of the physical deviation from the current commanded path.

The inner and outer feedback loops of Figure 1.1 represent two distinct levels of control of the vehicle. The outer loop holds the intelligence of the system while the inner loop provides the low-level machine control. The outer loop serves the function of translating an abstract task into concrete path commands that an autonomous vehicle can follow, and determining the current state of the task.

Since the addition of feedback to a system adds complexity, expense, and the potential for unstable behavior, it is reasonable to ask if feedback is required for both
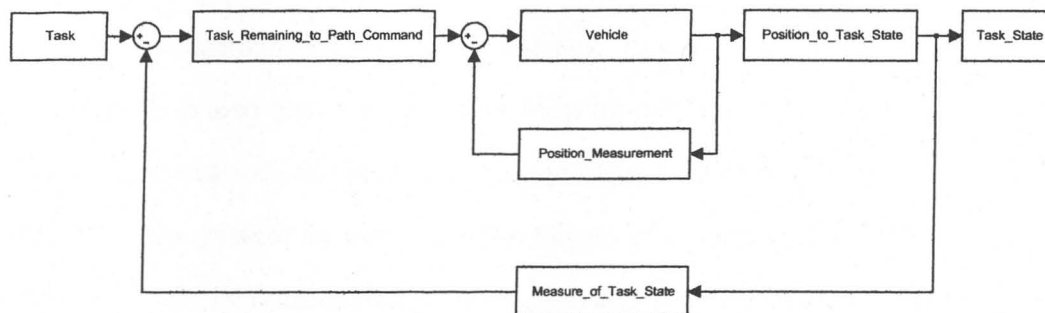
Fig. 1.1: Task and path management feedback structure.

the inner and outer loops of Figure 1.1. If the task is well defined and accurately translated into path segment commands then the vehicle could simply be commanded to complete the task without the need for observing the current task state. Similarly, if the vehicle could exactly follow a commanded path segment then there would be no need for corrective action to force the vehicle back to the path. The need for feedback becomes obvious when the vehicle operating environment is taken into consideration.

The inner loop of the autonomous vehicle system controls the vehicle motion, keeping the vehicle close to the commanded path. Vehicles normally operate in environments that provide significant disturbances to the vehicle motion. Wheels can strike objects, slide on slick surfaces, or be misaligned, all of which disturb the motion of the vehicle, causing the vehicle to deviate from the commanded path. Those who drive automobiles normally keep their hands on the steering wheel for precisely these reasons. The human driver of an automobile is constantly measuring the effects of environmental disturbances and correcting for these effects. In an autonomous vehicle, the inner loop feedback is added primarily to handle the unpredictable and persistent disturbances that force the vehicle off the commanded path.

The outer loop is also subjected to disturbances. Events that interrupt task completion must be dealt with in such a way as to ensure that the vehicle will properly complete the task despite the disturbance. Task disturbances can come in

many forms, including delays caused by obstacles in the commanded path and delays caused when consumables must be replenished. In addition, almost any disturbance of the inner loop also disturbs the outer loop by creating delays in task completion. If the autonomous vehicle were to operate without feedback, no corrections would be added to the system to overcome the effects of environmental disturbances and the vehicle would be commanded to new path segments before the vehicle had completed the previous segment. The outer loop feedback is added primarily to handle disturbances that delay the completion of the commanded task.

Feedback is often added to a system to provide robustness, or some measure of immunity, to external disturbance and to internal system variations. In this research, robustness is divided into two separate types: stability robustness and performance robustness. Stability robustness implies that a system will remain stable, i.e. not exhibit oscillatory behavior about, or exponential departure from, the commanded path. A system performs robustly when the response of a system to command or disturbance is similar despite system variation. While the two forms of robustness are considered separately, they are related in that performance can be considered as a stricter stability requirement. Stability and performance robustness of both the inner and outer loops of Figure 1.1 is considered in detail in this research.

1.1.1 Vehicle Training as a Method of Task Management

There are many reasonable approaches to the translation of the abstract task into path commands as seen in [3], [4], and [5]. Algorithmic approaches and the methods of artificial intelligence have proven effective at this translation. Also, for repetitive tasks, the vehicle can be 'trained' to the desired task by driving the vehicle around the required path. This approach removes the need for machine intelligence from the system. The task control system becomes less abstract when it is simply mimicking the path taken by a human operator. All that remains for the outer loop

design is to make the system insensitive to external disturbances and this can be accomplished through careful design.

In the training mode, a human operator guides the vehicle through the required path for a given task. Information about the path is stored, processed into a recognizable sequence of path commands, and the commands are provided to the vehicle as it traverses the path to complete the desired task. The training process is similar to the so-called pendant training of industrial robots where an operator controls the robot through a pendant, or manual controller, and moves the robot through the required task. The robot then mimics the trained motion to complete the desired task.

There are advantages to the training method over the algorithmic or artificial intelligence approaches mentioned previously. The first advantage is simplicity as it is generally not hard to drive the vehicle over the required path. Secondly, the path restrictions and tolerances are inherently obtained as the human operator drives the path, i.e. the operator performs the desired task in an acceptable manner and the autonomous vehicle simply mimics the operator from that time.

Vehicle path commands are derived from the vehicle position data taken while training. The path commands are those applied to the inner loop of Figure 1.1. The training algorithm translates the spatial vehicle position data into commands that the vehicle can follow. For vehicles with Ackerman steering, the type of steering commonly found in automobiles, the commands will consist of straight lines, segments of circles with constant radii, and curved segments that occur when changing from straight segments to circular segments. The path that is specified by the training algorithm must be one that the vehicle can follow. The path must not contain circular segments that have radius less than the minimum turning radius of the vehicle. Transitions between segments must be continuous and smooth, e.g. if a curved segment is connected to a straight segment, the straight segment must be

tangent to the curved segment. Path segments cannot be spatially shifted from the path that the human operator followed while training the vehicle. Lengths of path segments cannot be changed when processing the vehicle position data. The training algorithm must also account for vehicle size and the dimensions of any trailers that the vehicle is towing.

The task manager also must measure the current state of the task and command specific path segments to the path manager. The criterion that determines the current path segment is based on the observation of the current task state in comparison to the complete task. The remaining parts of the task dictate the current path segment. The path segments are not generally calculated continously, rather the segments are calculated after the vehicle training drive and are then selected when appropriate to complete another part of the task. The path segment selection criterion must allow for a smooth transition to the next path segment. Also, the criterion must be robust, meaning that a segment transition cannot be misplaced in space or missed altogether due to external disturbances.

1.1.2 Path Management

The inner loop of Figure 1.1 is less abstract than the outer loop. The vehicle responds to steering commands that when combined with the vehicle velocity move the vehicle along a path. The inner loop translates the vehicle deviation from the commanded path into steering commands that physically keep the vehicle on the commanded path. The form of the path management control loop is more familiar to those versed in control systems. The input to the inner loop is a physical path and the output is a physical vehicle position. As in all feedback systems, an error must be defined that measures the difference between where the system is and where the system is commanded to be. The error in the autonomous vehicle system is referred

to as the tracking error. The inner loop of Figure 1.1 is drawn in more detail in Figure 1.2.

The path manager consists of two primary parts with the first part being the vehicle controller that translates the tracking error into corrective action that tends to reduce the tracking error. The corrective action takes the form of steering commands in the case of an Ackerman-steered autonomous vehicle. The controller must be stable over the range of vehicle operating conditions, have a reasonable response to command changes, and attenuate the effects of external disturbances. An additional block identified as 'Preprocessing' is shown in Figure 1.2. The function of this block is to shape path segment changes in order improve the vehicle path response to these changes. While the block is shown as a series transfer function, the path segment preprocessing can also take the form of command feedforward.

Two primary methods of path management are used in autonomous vehicle applications. These are referred to in [6] as path following and trajectory tracking with the difference being that trajectory tracking includes time explicitly in the tracking while path following does not. Simply viewed trajectory tracking requires that the vehicle control system attempt to follow or track a point as it moves though time and space. This method works well for systems that are not likely to encounter significant disturbances or delays while tracking the point and requires that both the vehicle position and velocity be jointly controlled. On the other hand, path following is concerned only with the spatial relationship of the vehicle to the path and the vehicle can be slowed or stopped and restarted along the path with no effect on the control system and the vehicle position and velocity can be independently controlled. This research focuses on the path following method of control and seeks to expand this method to allow the design of path following controllers with responses defined in a spatial domain rather than in explicitly in time.
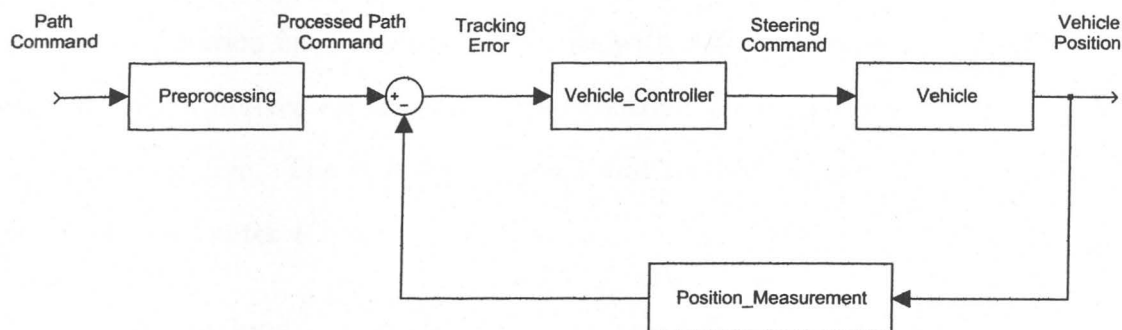
Fig. 1.2: Path management block diagram.

Path following is also the method of choice for solving problems that are inherently spatial with only secondary dependance on time. The field mowing example presented earlier in this chapter is such a problem. The mower must completely cover the field in order to effectively mow the field. While there is an optimal speed for mowing, actual ground speed will vary due to crop conditions, terrain, etc. and a controller that provides complete terrain coverage while being insensitive to normal velocity variation is more useful.

Understanding the inner loop of Figure 1.2 requires that a model of the vehicle be derived to analyze system performance, predict vehicle stability, and synthesize a controller to meet performance requirements. The model of the vehicle must accurately represent the real vehicle response to input and disturbance to provide a reasonable basis for design. The model must include any nonlinearities that exist in the actual vehicle as well as modelling vehicle dynamics. The derivation of the vehicle model and the derivation of the vehicle transfer function in a spatial domain is described in Chapter 3.

The synthesis of the controller and the command preprocessor is based on a combination of stability and performance requirements. The controller is commonly a linear system, but effective controller design for an autonomous vehicle also requires that a nonlinear stability analysis and linear controller synthesis be closely

coupled. A stability analysis is provided in this research that provides a stability and robustness criterion for autonomous vehicles with Ackerman steering. The stability analysis is first derived for the time based control method and then translated into the spatial domain. The stability criterion and method of controller synthesis are described in Chapter 4.

1.2   Contributions

The specific contributions of this research are:

(1) Furthering the development of the vehicle training method of task management.

(2) Development of a unique training data set reduction technique.

(3) Addition of spiral transition segments to customize paths for vehicle and task by producing driveable paths.

(4) Use of spirals segments to achieve a smooth transition from segment to segment.

(5) The reduction of the vehicle model to a single input/single output linear system.

(6) Replacing the independent variable of time with a spatial independent variable in the linear vehicle model.

(7) The development of a nonlinear stability criterion that predicts the stability margins of a specified controller.

(8) The development of a spatially robust controller.

Items (1)-(4) are presented in Chapter 2, items (5) and (6) in Chapter 3, and items (7) and (8) in Chapter 4.

1.3 Summary

The block diagram of Figure 1.1 presents the problem of controlling an autonomous vehicle in its most abstract form. The following chapters will provide a basis for analyzing and synthesizing an autonomous vehicle system thus reducing the abstraction of Figure 1.1.

The task management function will be considered in Chapter 2 of the dissertation. Chapter 2 includes the development of a method of translating from training data to vehicle path commands and producing a path that is generalizable to different vehicle velocities and conditions. The vehicle model will be derived in Chapter 3 prior to the discussion of the path management function in Chapter 4. The path management function is examined around the issues of control, stability, and performance of the inner loop of Figure 1.1. The results of simulation and vehicle experiments will be presented in Chapter 5. Finally, conclusions and directions for future research are presented in Chapter 6.

Chapter 2
Task Management

In the previous chapter, the problem of controlling an autonomous vehicle was described as a task management function coupled with a vehicle control function, referred to as the path manager, which keeps the vehicle on the desired path. This chapter provides a more detailed description of one method of task management: vehicle training via manually driving the path the vehicle must follow in order to complete the prescribed task. This chapter provides a mathematical description of a functional method of translating data obtained while driving a desired path into a general path consisting of tangentially connected straight lines and circular arcs. Then a derivation is presented of the equations for transition segments between circular arcs and thus provides a method for tailoring a general path to a specific vehicle performance. The chapter concludes with step-by-step procedures for processing the general path from the training data set and adding transition segments to the general path.

A subset of the overall task management function for an autonomous vehicle consists of generating of a series of commands that describe where the vehicle must drive to complete a desired task. Additionally, the task manager must determine the level of path completion and then schedule the vehicle transition from one path segment to another. A common path command form for vehicles with Ackerman steering consists of a series of tangentially connected straight lines and circular arcs, given that such vehicles naturally drive on circular arcs of different radii, including arcs of infinite radii, which are straight lines. However, for both high speed path

following and/or high accuracy requirements in path following, the regions of transition from one circular arc to another must be taken into account. This chapter provides a description of the transition regions between arcs and the requirements for generating and commanding the transition segments.

This chapter first presents a formal definition of vehicle trajectories and paths, followed by a description of the training data set, a description of driveable paths segments, and a method of processing the training data set in a more general data set. A method of adding spiral segments to a trained path is presented along with an argument for using transition segments to trigger switching from segment to segment, and the chapter concludes with a discussion of the derived methods.

2.1 Formal Definitions of Vehicle Trajectories and Vehicle Paths

To reduce the abstraction inherent in the task manager, the terms trajectory and path are defined in a formal, mathematical sense. Trajectories are the plane curves that the vehicle follows as it is driven and are normally parameterized in time. Paths are the plane curves that the vehicle is expected to follow and may or may not be parameterized in time depending on the control strategy. The formal definitions that follow allow the development of algorithms and criteria for designing and analyzing the task management system.

A trajectory is defined as a piecewise continuous plane curve C with a parametrization given as a vector function $\mathbf{P}$ in the $x - y$ plane representing the actual curve that a point on the vehicle traces. Such a parametrization is visualized as the curve traced by the tip of a vector when the vector is dependent on some parameter k. A trajectory C is denoted as:

$$C : \mathbf{P}(k) = x(k)\hat{\mathbf{x}} + y(k)\hat{\mathbf{y}}. \tag{2.1}$$

In (2.1) k is the curve parameter, $x(k)$ is the x component of the parametrization, $\mathbf{P}$, $y(k)$ is the y component of the parametrization, and $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ denote unit

vectors defined in the usual way. Using the above definition, trajectories can be parameterized using space, time, or a variety of other parameters unique to points on the curve. Two simple examples follow to help illustrate space and time parameterizations.

Example 1:

If the trajectory of the vehicle is defined as a simple curve in space with no dependence on time then a vector curve describing a trajectory along the line $y = 1.2x + 1$ is given by:

$$C : \mathbf{P}(x) = x\widehat{\mathbf{x}} + (1.2x + 1)\widehat{\mathbf{y}}. \tag{2.2}$$

Geometrically, the tip of the vector, $\mathbf{P}$, traces the line as x varies as shown in Figure 2.1.
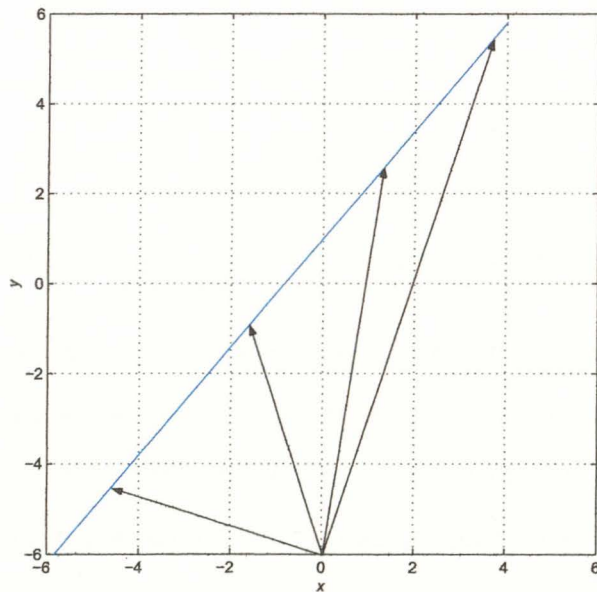
Fig. 2.1: Graphical representation of a space parameterized line.

Example 2:

A time parameterized vector curve has x and y components that are explicit functions of time. A simple example of this would be the time parametrization of a circle. The tip of the **P**-vector traces a circle of radius one in the plane as t varies from zero to one according to:

$$C : \mathbf{P}(x) = \cos(2\pi t)\widehat{\mathbf{x}} + \sin(2\pi t)\widehat{\mathbf{y}}. \tag{2.3}$$

The time parameterized circle of (2.3) is shown graphically in Figure 2.2.
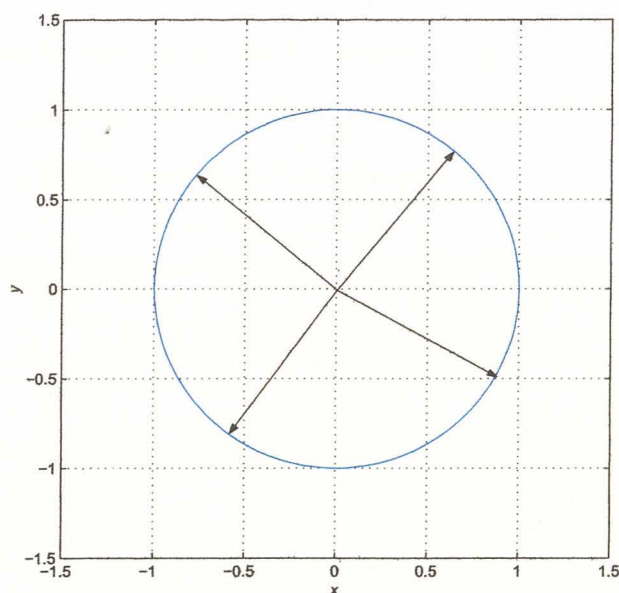
Fig. 2.2: Graphical representation of a time parameterized circle.

As noted, the parametrization of plane curves is not limited to either a time or space parametrization. A space curve can be parameterized using any property of the curve that can be uniquely associated with a point on the curve. The flexibility in choosing curve parameterizations can simplify analysis and enhance the understanding of a problem, as will be illustrated below.

Paths are also plane curves that are defined by a vector parametrization, with an allowable path being defined as one that the vehicle can exactly follow. The

vehicle steering method, the vehicle steering geometry, the steering dynamics, and the vehicle velocity limit the set of allowed vehicle paths. Paths can be represented as tangentially connected segments. Allowable paths for vehicles with Ackerman steering are limited to combinations of straight-line segments, circular arcs, and spirals. An allowable path for a vehicle with Ackerman steering cannot include any non-analytic or non-tangential segments.

2.2   Overview of Research in Path Planning

The research in this area is broad. Researchers have investigated many ways to synthesize paths for robotic vehicles to follow. The methods used to synthesize a planned path or process the data from a trained path tend to be similar. Some commonly used synthesis/processing methods include:

1. Paths consisting of splines (cubic and B-splines are common) [7].

2. Tangentially connected circular arcs and straight lines.

3. Paths consisting of straight line segments and clothoid spiral segments.

4. Paths consisting of clothoid spirals, circular arcs, and straight lines.

The last three items in the above list will be considered in more detail. The most common method of path planning consists of forming a path consisting of tangentially connected straight lines and circular arcs. Sophisticated versions plan optimized paths through cluttered environments and include dynamic replanning should unknown obstacles appear. Replanning also occurs if the path becomes sub-optimal for the task at hand. An overview of planning algorithms can be found in [8].

Shin and Singh [9] present a method of path planning that synthesizes curves in planned paths by using of clothoid spiral segments. Clothoid spiral segments occur

when the turning radius of the vehicle in changed in a continuous linear manner implying that the change of steering angle is not instantaneous. They suggest that the resulting paths are superior to those created from tangentially connected straight lines and circular arcs in that the spirally connected paths is continuous and therefore driveable. Curves in the planned paths are completely composed of spiral segments with no circular arcs. As described below such paths require that the vehicle steering system be continuously moving whenever the vehicle is turning. Similarly, Ishimoto, Tsubouchi, Sarata, and Yuta [10] use a two-spiral segment to provide a forward and reverse path for a wheeled loader.

Scheuer and Fraichard in [11] present a method planning continuous curvature paths using clothoid spiral segments, straight lines, and circular arcs. The paths created are more driveable than paths consisting of tangentially connected circular arcs and straight lines. The paper uses synthesized curvature to produce the path. The problem is addressed geometrically.

Nagy and Kelly in [12] contend that spiral segments are not the best segments for a vehicle with constrained steering. They feel that a curvature with continuous derivative is more representative of the path driven by a vehicle with a rate limited steering system. The curve that they suggest using is a cubic spiral.

The method of path training presented below provides the advantages of continuous curvature paths without requiring the steering system to be in motion for the entire duration of the curve. Additionally, changes in vehicle velocity and steering systems rate can be added to the path prior to a specific mission within limits. Also, if the assertion of Nagy et al. proves to be true, the clothoid spiral segments can be replaced with cubic spirals without difficulty.

Adams in [13] uses a potential field method to derive the path through the environment in order to reach a desired point in the environment. Significantly, he states:

Many mobile robot path planning algorithms produce changing intermediate goal coordinates for a mobile robot to pursue, and provide motoring speed/torque signals based upon local sensor information and the position of the global target. This is often done with little or no regard for the low level vehicle dynamics, which, in practice, must be taken into account for efficient path planning. (p. 230)

One of the advantages of the path training method is that vehicle kinematics are automatically taken in account when the training path is driven. The method of path training described below also allows changes in vehicle velocity and curvature rate limit to be added to the path after training and processing. The limit on steering angle is usually taken in account in path planning algorithms for Ackerman steered vehicles (minimum turn radius), but the effect of the rate limit on curvature or the vehicle velocity are not commonly considered.

2.3 Path Training

The vehicle path-training mode is similar in part to the so-called pendant training of a stationary industrial robot. In pendant training, a skilled operator moves the robot through the motion required to perform a desired task. After training, the robot repeats the trained motion in order to perform the task. The robot is supplied with the exact time-dependent commands that created the original motion. The vehicle training method could precisely mimic pendant training, logging the exact steering commands, and supplying these commands to vehicle as it is driving. An exact imitation of pendant training would work in theory for guiding a vehicle through an environment, but it requires a restrictive set of conditions that include: placing the vehicle exactly in the same starting position and orientation, driving with the exact velocity profile, experiencing no external forces not seen during training,

and receiving an exactly identical performance of the steering system. Meeting all of these conditions would allow 'playing back' the recorded steering commands and, at least in an undisturbed environment, the vehicle would follow the trained path through the environment with the exact time response. However, the operating environment of the vehicle is not predictable and the vehicle could easily be forced off of the desired path by external disturbances or the steering system response could change due to environmental factors such as temperature. In short, simply supplying steering commands to the vehicle makes no use of vehicle position and orientation feedback given that no corrective action is taken in the presence of external disturbance or internal variation. The 'play back' method has the additional restriction of producing a path with a fixed, or at best scalable, velocity profile.

An alternative method of path training converts the training trajectory into a path consisting of segments that a wide range of vehicles can drive. Ideally, the training path would consist of a set of spatial segments that are not dependent on time and independent of specific vehicle limitations. Given a perfect vehicle with instantaneous response to steering commands, removing the time dependence from the vehicle training path would allow any path segment to be driven at any velocity rather than a velocity pattern, or a scaled velocity pattern, fixed during training. Making the trained path independent of vehicle limitations implies that the trained path would not have to modified for any vehicle. Producing a path that is both velocity and vehicle independent is not possible given the possible range of operating velocities and variations in vehicle steering systems. However, a general path that is modifiable for specific driving conditions can be produced and is described in the following sections.

The path training algorithm that is developed in this chapter follows the steps outlined below:

(1) The training data set is acquired by driving the vehicle along the desired vehicle trajectory.

(2) The training data set is filtered to reduce the effects of measurement noise.

(3) The filtered data set is processed to obtain an estimate of the trajectory curvature and tangent vector at each point in the data set.

(4) Critical points in the calculated curvature are identified.

(5) A base path is derived using equivalent circular arc segments calculated to produce equivalent heading and position at data set critical points.

The development of the path training algorithm follows with a general description of the path training data set, details of the subcomponents that will exist in the data set, data set filtering, equations of circular arcs between segments, location of critical points in the data set, and error measurement.

2.3.1 Description of the Path Training Data Set

The path training data set considered in this research consists of a series of $x - y$ data points taken while the vehicle is being driven by a human operator over the desired vehicle path. The vehicle position data set defines a segmented vehicle trajectory and this trajectory can be thought of as a series of time-parameterized spatial curves, C_i, with the complete trajectory being represented as the union of the individual segments, $C = \bigcup_i C_i$. Time parametrization is used at this point in the derivation, given that the vehicle training data set is regularly sampled over time.

The data set is a combination of the actual vehicle position plus some measurement error or measurement 'noise.' The effect of this noise must be considered in the processing of the data set to obtain the spatial curves that define the vehicle path.

2.3.2 Description of Allowable Path Segments

As mentioned above, an allowable path for a vehicle with Ackerman steering will consist of tangentially connected straight-line segments, circular arcs, and spirals. Both straight-line segments and circular arcs occur when the vehicle is driven with a fixed steering angle. The spiral segment is obtained when the steering angle is changing over time. Unlike straight-line segments and circular arcs, the spirals are dependent on vehicle velocity and steering system rate limits. A special case of a spiral occurs when the rate of change of curvature is constant. Such spirals referred to as clothoid spirals. The clothoid spiral is well known to highway designers and such spirals are included in all high speed roads, with high speed being defined as greater than fifty kilometers per hour. The spirals normally act as connecting curves between segments of different radii and/or direction. It should be noted that the shape and length of the spiral segments is dependent on both the vehicle velocity and rate of change of curvature. It is worthwhile to create base, or general, trained paths that are independent of either the velocity or steering system limitations of the training vehicle given that such paths are driveable by vehicles other than the training vehicle.

Given the desirability of velocity independent paths and the impossibility of defining such paths, a compromise is sought. The compromise that is presented requires that a general base path be selected and that this base path be modified to meet velocity and vehicle turning radius requirements. The base path that has been selected in this dissertation consists of a series of circular arcs and straight lines that are connected to enough points of the original training data set to ensure path accuracy. The velocity and steering system limitations are inserted into the base path after a specific vehicle is selected to operate on the path. The resulting combination of base path and vehicle/velocity dependent transition segments produces paths that are driveable by a broad variety of vehicles over a reasonable range of vehicle speeds.

2.3.3 Dataset Processing

The dataset is processed to remove the time dependence and produce a series of space curves. The chosen method uses the curvature of the training trajectory to define critical points in the data set and fit circular arcs to these critical points. Curvature is formally defined as the change in direction per unit of arc length. Alternatively, curvature can be thought as the reciprocal of the radius of the instantaneous tangent circle of best fit at any given point on the curve. In the case of a vehicle, curvature is understood as the reciprocal of the radius of the tangent circle that the vehicle would follow if the instantaneous steering angle were held.

Curvature is the primary tool used in converting the position data set into commands used by the vehicle control system. Obtaining accurate curvature information is therefore critical to the translation of training data set into recognizable path commands. Knowing the curvature at each point on a trajectory is equivalent to knowing the instantaneous steering angle required to follow the trajectory. It is the correlation of curvature to vehicle position on the training path that allows the derivation of driveable path segments. The existence of a driveable path allows for accurate feedback control of the vehicle as it follows the trained path.

Since correlating curvature to vehicle position is necessary in building driveable path segments, the curvature could be obtained from a steering angle sensor and the position information could be obtained from some position measurement system such as a Global Positioning System (GPS) device. The resulting combination of curvature and position data set can then be processed into driveable path segments that are no longer dependent on time. However, as will be shown, measuring the steering angle is unnecessary as the needed curvature information can be obtained from position information only using the formula for the curvature of a time-parameterized curve [14] given by:

$$\kappa(t) = \frac{x'(t)y''(t) - y'(t)x''(t)}{\left([x'(t)]^2 + [y'(t)]^2\right)^{3/2}}. \tag{2.4}$$

The curvature formula presented in (2.4) is slightly modified from the formula in [14] in that the absolute value in the numerator is removed. The formula of (2.4) provides direction information with a counterclockwise turn having a positive sign and a clockwise turn having a negative sign.

If the time functions $x(t)$ and $y(t)$ in (2.1) were known in closed form then the process of converting the position data set into vehicle steering commands would require the simple application of (2.4). However, the position data set is generally sampled at some regular time interval and numerical approximations must be made for the derivative terms. The appearance of first and second derivatives in (2.4) makes the problem difficult to solve numerically due to presence of time-varying measurement error in the position data samples. The measurement error is small, typically limited to less than a tenth of a meter for modern GPS systems, but the error changes rapidly and randomly and the process of differentiation amplifies the rapidly changing error, or measurement 'noise'.

A sample of vehicle position data taken from a tractor driving a series of straight lines, curves, and circles is illustrated in Figure 2.3. The curvature for this vehicle trajectory should consist of relatively smooth sections of very large radius indicating curvature near zero and relatively smooth smaller radius circles indicating larger absolute curvature. While the noise level is small, virtually invisible at the scale shown, the curvature calculated directly from the raw position data set is difficult to use as illustrated in Figure 2.4.

In order for the calculated curvature to be useful, the training data set must be filtered to attenuate the measurement noise. The chosen filter must both attenuate the noise and maintain the critical spatial relationships of the training data set. Filtering that shifts the training data set in space is unacceptable. The design of an effective filtering system for the training data set is described in the next section.
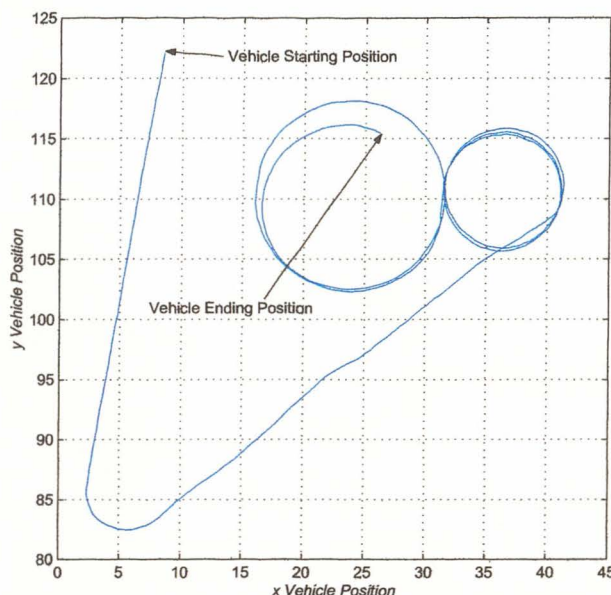
Fig. 2.3: Raw vehicle trajectory data.

2.3.4 Training Data Set Filter Design

The $x-y$ position data set is divided into two time sequences with the individual time sequences shown in Figures 2.5 and 2.6. The resulting time sequences can be filtered to adequately attenuate the measurement noise. As is indicated in Figures 2.5 and 2.6, the x and y time series consist of pieces of sinusoids (corresponding to circular arcs) and linear segments (corresponding to straight lines). The filter must be designed so as to pass the sinusoidal segments and not adversely affect the linear segments. In other words, both the frequency response and the slew rate of the chosen filter must be considered in order to avoid attenuating desired information while adequately attenuating the measurement noise.

The frequency of the sinusoidal segments is given by:

$$f = \frac{v}{2\pi r}, \tag{2.5}$$

where v is the vehicle velocity and r is the turning radius of the vehicle.
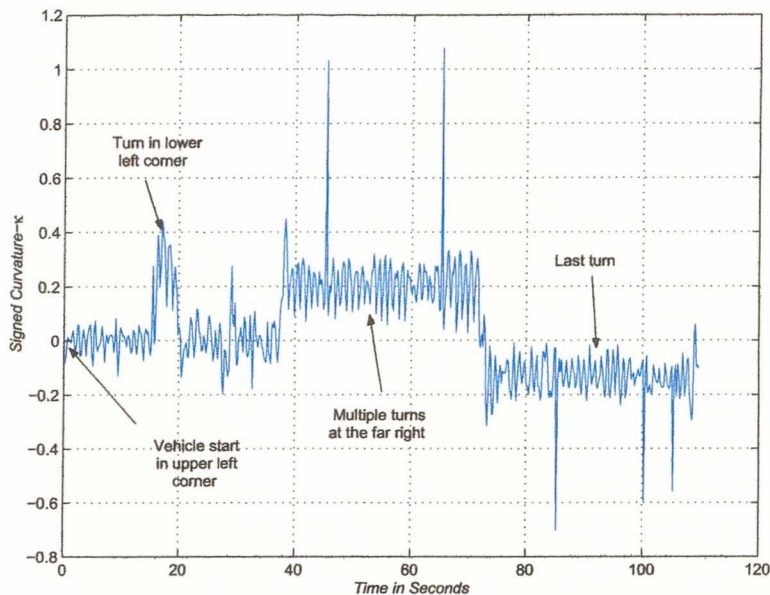
Fig. 2.4: Curvature of the unfiltered trajectory data.

The maximum slope of a sinusoidal waveform occurs when the waveform crosses its horizontal axis of symmetry and this slope is given by the value of the derivative of the waveform at these points. The slope is given by the amplitude of the waveform multiplied by the radian frequency of the waveform, $2\pi f$, with the amplitude being given by r. The resulting maximum slope for either the x or y waveforms is then given by v, the vehicle velocity. The maximum slope of the linear segments are obtained when the vehicle is travelling orthogonally to either the x or y axes and the slope is again simply v.

The result is that the minimum pass frequency of the filter can be specified by the maximum frequency present in the time sequence and that frequency is given by:

$$f_{\max} = \frac{v_{\max}}{2\pi r_{\min}}. \tag{2.6}$$

This is because the filter should accurately reproduce the waveshapes of sinusoids in the filter passband and if the x and y time segments include sinusoidal sections in
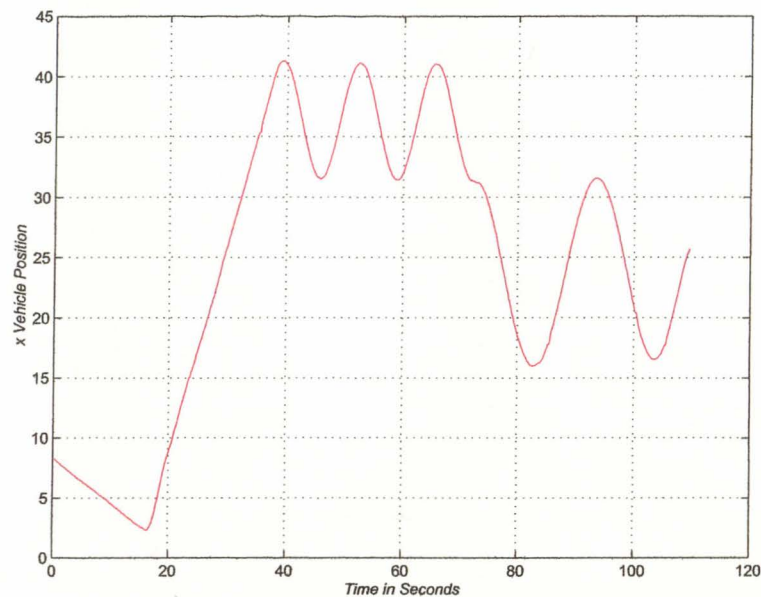
Fig. 2.5: x-Component of the trajectory data versus time.

the passband with maximum slopes v these sections will also be accurately repro-
duced. Therefore, any linear region with slope v will be accurately reproduced by
the filter. Note that (2.6) is dependent on the maximum velocity and the minimum
turning radius of the vehicle, in this case, the maximum velocity of the vehicle is the
maximum velocity that the vehicle reaches during training.

The frequency content of the regularly spaced measurement noise is more diffi-
cult to quantify. The noise is broadband and the spectrum changes with time due to
the random nature of the noise. Although measurement noise exists at all frequen-
cies, the only noise of concern is that noise falling in the frequency range of legitimate
signal data. If the process of differentiating the data set twice significantly lowers
the signal to noise ratio in the passband of the low pass filter then the calculated
curvature will be noise contaminated. The signal to noise ratio goes to zero when
velocity goes to zero since the measurements consist only of noise. Obtaining an es-
timate of a minimum acceptable velocity is desirable to maintain a reasonable signal
to noise ratio in the curvature calculation.
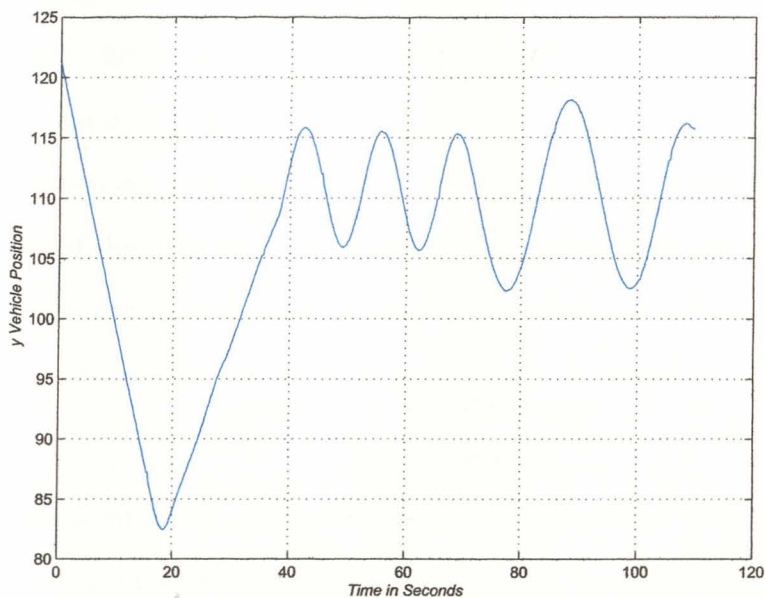
Fig. 2.6: y-Component of the trajectory data versus time.

Noise in the passband is amplified by the process of differentiation. The measurement plus noise signal for a vehicle travelling along the x-axis with velocity, v_x, is given as:

$$x_m(t) = v_x t + N_x(t) \qquad (2.7)$$

$$y_m(t) = N_y(t), \qquad (2.8)$$

where $N_x(t)$ and $N_y(t)$ are normally distributed white noise signals that have been filtered.

The power spectral densities (PSD) of the noise signals are σ_x^2 and σ_y^2 respectively where σ_x and σ_y are the standard deviations of the noise signals. The process of differentiation multiplies the power spectral densities by ω^2 for each differentiation. The resulting noise signals are therefore no longer have a uniform frequency spectrum.

A conservative estimate of the differentiated noise is made by assuming that the noise is uniformly distributed over the filter passband with PSD of $\sigma^2\omega_{max}^2$ for the first derivative and $\sigma^2\omega_{max}^4$ for the second derivative.

The root mean squared amplitude of the differentiated noise signals is simply the square root of the PSD for those signals. Therefore, a conservative estimate of the curvature as a function of forward velocity is calculated as:

$$\kappa_{max} \cong \frac{\sigma_y\omega_{max}^2 v_x}{\left(v_x(v_x - 2\sigma_x\omega_{max})\right)^{3/2}}. \tag{2.9}$$

Given a maximum RMS amplitude variation of curvature that is acceptable and the standard deviation of the noise signals, (2.9) can be solved for v_x, which is the minimum allowable velocity. The estimated velocity is normally much higher than it needs to be and simulation of the noise contaminated signals can be used to obtain a less conservative estimate of the minimum acceptable velocity.

The filter is chosen based on a calculation of the highest frequency component possibly present in the waveform, which is dependent on the minimum turning radius of the vehicle and the maximum vehicle velocity. One might suppose that the Nyquist sampling rate plays some role in the selection of the filter cutoff, but in this dissertation the sample rate of the data set is considered to be substantially higher than the Nyquist rate. Note that sampling at the Nyquist rate is the equivalent of taking only two samples per revolution when driving the vehicle in a circle. Such a slow rate does not make path reconstruction feasible. It will be shown below that the choice of filter type is not important beyond the choice of cutoff frequency, which must be above the highest possible frequency in the waveform. Empirical testing shows that choosing the low-pass filter cutoff frequency at one and one half times the maximum waveform frequency yields good results in both amplitude reproduction and noise attenuation.

2.3.5 Eliminating Phase Shift in the Calculated Curvature

Unfortunately, the typical low-pass filter introduces both magnitude attenuation and phase lag. The magnitude attenuation can be reasonably controlled by selection of the cutoff frequency, but the phase lag is more difficult to control as the phase lag becomes significant at frequencies much below the cutoff frequency of the filter. The filtered data set shows the effects of the phase lag when overlaid with the original, unfiltered data set, as shown in Figure 2.7. This figure shows the effects of the magnitude attenuation and phase lag caused by a low-pass filter that was adequate to reduce the measurement noise in order to provide usable curvature information.

The difficulty introduced due to the phase shift can be eliminated given that the entire position data set exists before filtering commences, i.e. the position data set is not processed into path segment commands until the training mission is completed. This fact permits the use of a zero-phase digital filter or a non-causal low pass filter. The zero-phase filter functions by filtering the data set twice, once forward in time and then backward in time, using the same filter for both passes. Any phase lag introduced by the first filtering pass is cancelled by the second pass. A block diagram illustrating zero phase filtering is shown in Figure 2.8. The resulting filtered data set can then be numerically processed to produce the needed curvature information using Equation 2.4. The curvature calculated using the zero-phase filtered position data set is shown in Figure 2.9. Note that the two passes through the digital filter substantially attenuates the magnitude of the data, which again introduces spatial error when path reconstruction is attempted strictly from the curvature calculation.

Any filtering of the x and y time sequences results in a dilemma. Normal low-pass filtering introduces phase shift that results in spatial movement of the vehicle trajectory, but eliminating the phase shift by using a zero-phase filter introduces significant magnitude attenuation. Critical information is now spread between the unfiltered vehicle position data set and the curvature calculated from the filtered and
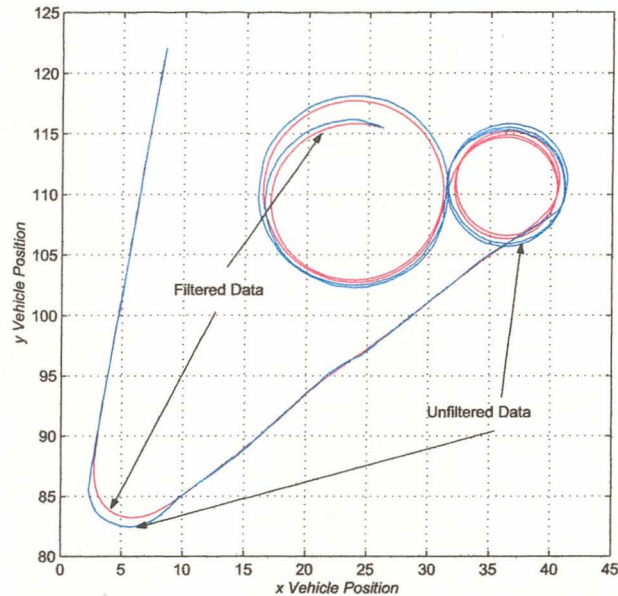
Fig. 2.7: Overplot of vehicle trajectory data-unfiltered and low-pass filtered.

differentiated position data set. The curvature holds accurate time, and therefore event, information because the filter used on the position data does not introduce any significant phase shift. Magnitudes in the unfiltered position data are as accurate as the measurement method itself and assumed to be adequate for locating and controlling the vehicle. After the critical event times are identified, the x and y points associated with the event times are taken from the unfiltered position data. The algorithm that locates the critical event times is developed in the following section.

2.4   Reduction of the Training Data Set to a General Path

The goal of reducing the training data set is to create a series of tangentially connected straight lines and circular arcs from the data set that consists of straight lines, circular arcs, and spiral segments. The reduced training set curvature therefore has no sloped regions and transitions between different values of curvature are discontinuous at the transition. The training data set is essentially 'squared-off' with
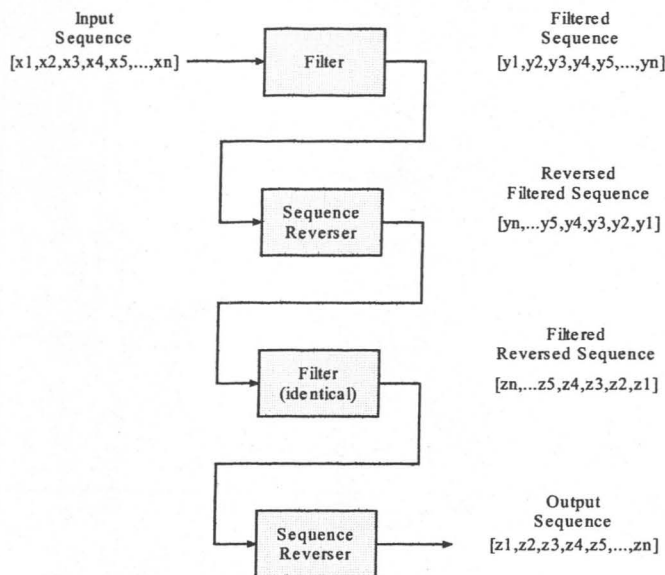
Fig. 2.8: Block diagram representation of zero phase filtering.

relatively slow transitions being replaced by transitions taking place instantaneously. Zero crossings in the training data set curvature indicate the points where the vehicle is again travelling straight forward.

One could select the zero crossings as critical data points and then replace the spirals and arcs between the points with arcs only. At first glance it might seem reasonable to replace the segments between zero crossings with a single circular arc that meets the starting and ending orientation. This is equivalent to ensuring that the integral of the training data set curvature times velocity and the fitted curvature times velocity are equal. However, this is generally not possible given the effect of transition segments on both position and orientation. It is a simple matter to replace the complex curve between zero crossings with one of constant value that has equal area under the curve, but only the vehicle orientation at the endpoint will
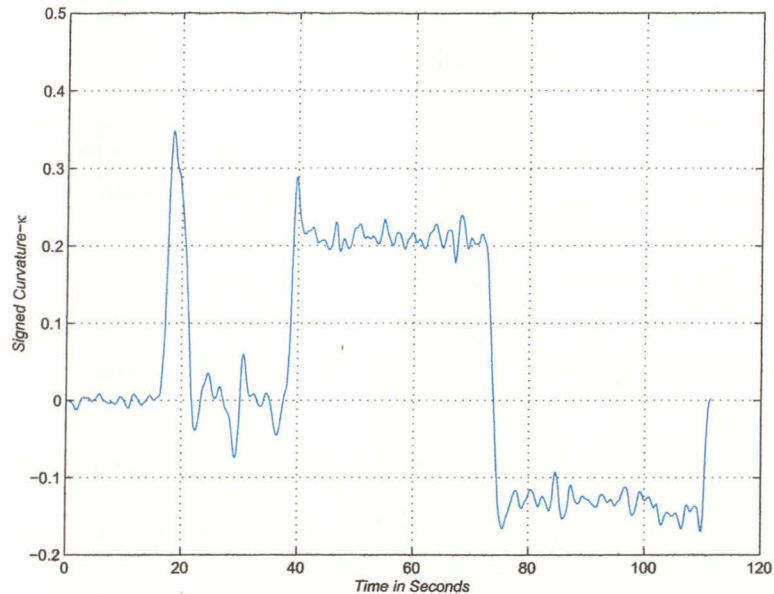
Fig. 2.9: Curvature calculated from filtered trajectory data.

be matched with the position error being significantly large and accumulating from segment to segment.

The solution to this problem is begun by defining the relationship between the time rate of change of heading of the vehicle and its position and orientation. Although the kinematic vehicle model will be explained in detail in Chapter 3, a definition of this relationship is shown here. The product of velocity and curvature is the time rate of change of vehicle heading. The integral of the product of curvature and vehicle velocity is the vehicle heading. The integral of the product of the cosine of vehicle heading and vehicle velocity produces the vehicle x-position and the integral of the product of the sine of vehicle heading and vehicle velocity produces the vehicle y-position. A block diagram representation of the system is shown in Figure 2.10.

In the block diagram of Figure 2.10, matching the integral of the product of velocity and curvature produces identical headings only. The second integration defines the vehicle position and must be considered in the solution of the fitting
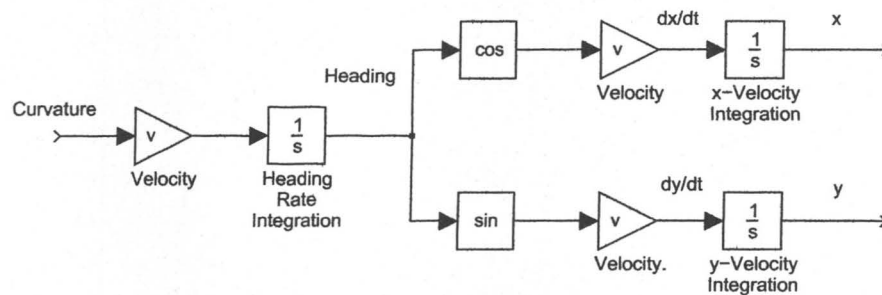
Fig. 2.10: Transfer function relationship of curvature to vehicle position.

problem. Mathematically the problem of fitting the data set to both orientation and position amounts to solving the following system of nonlinear, boundary-value differential equations:

$$
\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} v \cos \phi \\ v \sin \phi \\ v \kappa \end{bmatrix}.
\tag{2.10}
$$

The boundary conditions are the starting and ending points of the data segment and are represented by (x_0, y_0, ϕ_0) and (x_2, y_2, ϕ_2).

2.4.1 Fitting Tangential Circular Arcs Between Points on the Training Trajectory

The problem of matching both starting and ending conditions requires that a sufficient condition be established describing when a given circular arc can be tangentially connected to a vector in space with known position and orientation. A general proof of the case when the circular arc is a full circle is presented. The proof is illustrated in Figure 2.11.

Theorem 1. *Given a circle C of radius R, an arbitrary point C_0 in the interior, and a vector, $\mathbf{v}$ of arbitrary direction starting at C_0. There exists two circular arcs, A_a and A_b, of radius r_a and $r_b < R$ such that the arcs are tangent to a point on the perimeter of C and tangent to the vector $\mathbf{v}$*
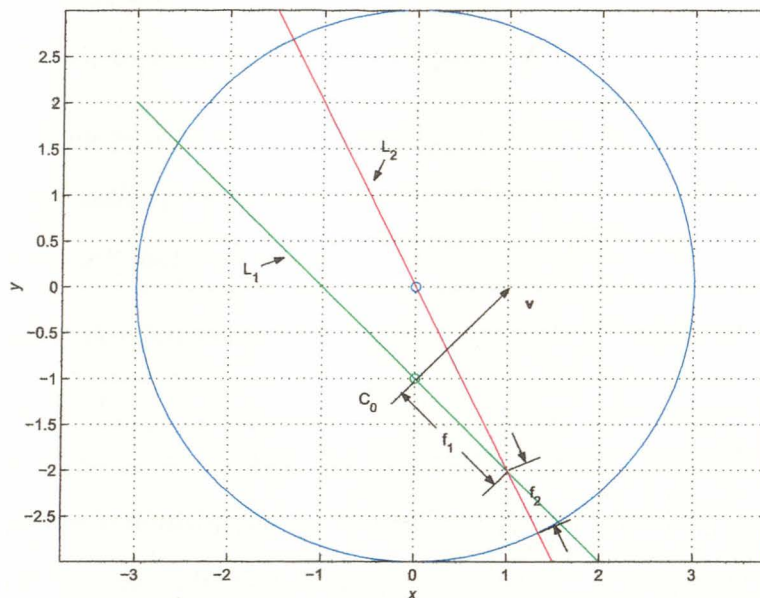
Fig. 2.11: Geometrical illustration of tangential arc fit.

Proof. Construct a line, L_1, that is perpendicular to the vector **v**. Any circle that passes through C_0 will have its center on L_c. Construct a second line, L_2 of arbitrary orientation that passes through the center of the circle C. Now define two functions f_1 and f_2 as follows:

- f_1 is the distance from the intersection of L_1 and L_2 to the point C_0.

- f_2 is the distance from the intersection of L_1 and L_2 to the point at which L_2 intersects the circumference of C.

If a circular arc A passes through C_0, is tangent to **v**, and is tangent to C then f_1 is equal to f_2. Look at the points were f_1 and f_2 are zero:

- f_1 is zero when the line L_2 passes through the point C_0.

- f_2 is zero when the line L_2 passes through one of the points on the circumference of C intersected by L_1.

Let $g = f_1 - f_2$ then when $f_1 = 0$, $g > 0$, and when $f_2 = 0$, $g < 0$ and using a theorem attributed to Bolzano [15] that states:

Theorem 2. *Suppose a numerical function f is continuous on a closed interval $[a, b]$ and takes values with opposite signs at the endpoints of $[a, b]$. Then there exists an interior point c such that $f(c) = 0$.*

Note that g is continuous given that it consists of additive combinations of continuous functions.

If $g = 0$ then $f_1 = f_2$ and the tangential circular arc exists. Since either intersection of the circumference of C can be chosen, two such tangential circular arcs exist. Call the two solutions r_a and r_b.

$\square$

While the above proof establishes that it is possible to create an arc that connects to any point on the interior of the given circle C, it does so only by allowing the point of tangency on C to be unfixed. It is this insight that allows the completion of the fitting algorithm. If at the zero crossings, straight line segments of variable length can be included as part of the data fit, these segments act as heading 'place holders' in that they allow shifts in position, but no change in orientation. Also, curves of larger radius can be extended to allow fitting of the data to a segment of smaller radius.

2.4.2 Solution of the Boundary Value Problem

To fit the training data set to a more general path consisting of a series of circular arcs and straight lines, the matrix nonlinear differential equation of (2.10) must be solved to match the boundary conditions imposed by the training data. Let the time rate of change of heading (velocity times the curvature) be defined as:

$$v\kappa(t) = \begin{cases} v\kappa_0 & \text{if } t \in [t_0, t_1] \\ v\kappa_1 & \text{if } t \in [t_1, t_2] \end{cases}, \tag{2.11}$$

where v, κ_0, and κ_1 are constants. The shape of the trajectory created by such a curvature input consists of a pair of tangentially connected circular arcs.

Equation (2.10) must be integrated to solve for $x(t_2)$ and $y(t_2)$. Integrating (2.10) yields the equations for $x(t_2)$ and $y(t_2)$ below:

$$x(t_2) = v \int_{t_1}^{t_2} \cos\left(\int_{t_1}^{t} v\kappa_1 dt + \phi_1 \right) dt$$

$$+ v \int_{t_0}^{t_1} \cos\left(\int_{t_0}^{t} v\kappa_0 dt + \phi_0 \right) dt + x_0 \qquad (2.12)$$

$$y(t_2) = v \int_{t_1}^{t_2} \sin\left(\int_{t_1}^{t} v\kappa_1 dt + \phi_1 \right) dt$$

$$+ v \int_{t_0}^{t_1} \sin\left(\int_{t_0}^{t} v\kappa_0 dt + \phi_0 \right) dt + y_0. \qquad (2.13)$$

Expanding (2.12) and (2.13) results in the equations for $x(t_2)$ and $y(t_2)$ as shown:

$$x(t_2) = \frac{\sin(v\kappa_1 t_2 - v\kappa_1 t_1 + \phi_1) - \sin(\phi_1)}{\kappa_1}$$

$$+ \frac{\sin(v\kappa_0 t_1 - v\kappa_0 t_0 + \phi_0) - \sin(\phi_0)}{\kappa_0} + x_0 \qquad (2.14)$$

$$y(t_2) = \frac{-\cos(v\kappa_1 t_2 - v\kappa_1 t_1 + \phi_1) + \cos(\phi_1)}{\kappa_1}$$

$$+ \frac{-\cos(v\kappa_0 t_1 - v\kappa_0 t_0 + \phi_0) + \cos(\phi_0)}{\kappa_0} + y_0. \qquad (2.15)$$

Defining the angles $\phi_1 = v\kappa_0 t_1 - v\kappa_0 t_0 + \phi_0$ and $\phi_2 = v\kappa_1 t_2 - v\kappa_1 t_1 + \phi_1$, (2.14) and (2.15) can be written in a time independent form as shown:

$$x(\phi_2) = \frac{\sin(\phi_2) - \sin(\phi_1)}{\kappa_1} + \frac{\sin(\phi_1) - \sin(\phi_0)}{\kappa_0} + x_0 \qquad (2.16)$$

$$y(\phi_2) = \frac{-\cos(\phi_2) + \cos(\phi_1)}{\kappa_1} + \frac{-\cos(\phi_1) + \cos(\phi_0)}{\kappa_0} + y_0. \qquad (2.17)$$

Equations (2.16) and (2.17) can be solved for any two unknowns. In the algorithm presented here, the unknowns will normally be κ_1 and ϕ_1 and the equations constitute a simultaneous pair of nonlinear equations. The equations can be solved using any appropriate nonlinear method. For instance, the multi-variable Newton-Raphson method works well.

An important special case occurs when $\kappa_0 = 0$ or when the large radius circular segment is a straight line. Returning to (2.14) and (2.15), a limit can be taken as $\kappa_0 \to 0$. The limits are:

$$\lim_{\kappa_0 \to 0} x(\phi_2) = \frac{\sin(\phi_2) - \sin(\phi_0)}{\kappa_1} + \frac{0}{0} + x_0 \tag{2.18}$$

$$\lim_{\kappa_0 \to 0} y(\phi_2) = \frac{-(\cos(\phi_2) - \cos(\phi_0))}{\kappa_1} + \frac{0}{0} + y_0. \tag{2.19}$$

In (2.18) and (2.19), both the numerator and denominator of the second terms are zero. Applying L'Hospital's rule to the equations by differentiating both the numerators and denominators of the second terms in (2.14) and (2.14) with respect to κ_0 results in:

$$\lim_{\kappa_0 \to 0} \frac{d\left(\sin(v\kappa_0 t_1 - v\kappa_0 t_0 + \phi_0) - \sin(\phi_0)\right)/d\kappa_0}{d\kappa_0/d\kappa_0} = \cos(\phi_0)v(t_1 - t_0) \tag{2.20}$$

$$\lim_{\kappa_0 \to 0} \frac{d\left(-\cos(v\kappa_0 t_1 - v\kappa_0 t_0 + \phi_0) + \cos(\phi_0)\right)/d\kappa_0}{d\kappa_0/d\kappa_0} = \sin(\phi_0)v(t_1 - t_0). \tag{2.21}$$

Note that the terms $v(t_1 - t_0)$ simply represent the length of the straight segment extended from the point (x_0, y_0) and that these terms can be replaced by an arbitrary parameter p. The equations for a straight line to curve transition thus reduce to:

$$x(\phi_2) = \frac{\sin(\phi_2) - \sin(\phi_0)}{\kappa_1} + \cos(\phi_0)p + x_0 \tag{2.22}$$

$$y(\phi_2) = \frac{-\cos(\phi_2) + \cos(\phi_0)}{\kappa_1} + \sin(\phi_0)p + y_0. \tag{2.23}$$

The unknowns in (2.22) and (2.23) are normally κ_1 and the straight line length parameter p, and in this case, the equations reduce to a pair of simultaneous linear equations.

2.4.3 Fitting the Training Trajectory Data Using Only Curvature Zero Crossings as Critical Points

One possible method of fitting the data set involves selecting the critical points at the zero crossings of the calculated curvature. The method has the advantage of simplicity in that the zero crossings are simple to locate and naturally provide a significant reduction in total points in the resulting trained path. The training path is generated between each successive pair of zero crossings through the following steps:

(1) The curvature zero crossings in the data set are identified using the method described above.

(2) The total heading change is calculated between between the two curvature zero crossings.

(3) The point on the data set that is at half the total heading change between zero crossings is located on the data set.

(4) A straight segment, circular arc, and another straight segment are calculated to fit the end and middle conditions of the original data.

Each resulting segment consists of a straight line segment at each curvature zero and a circular arc passing through the center of the segment. The circular arc accounts for all the heading change in the segment and provides some $x - y$ translation. The straight line segments provide the necessary remaining $x - y$ translation required for matching the end coordinates.

An example is used to clarify the method The complex trajectory of Figure 2.3 is to be fit using this method. First the curvature zeroes are calculated. The resulting zeroes are plotted on the curvature plot shown in Figure 2.12.
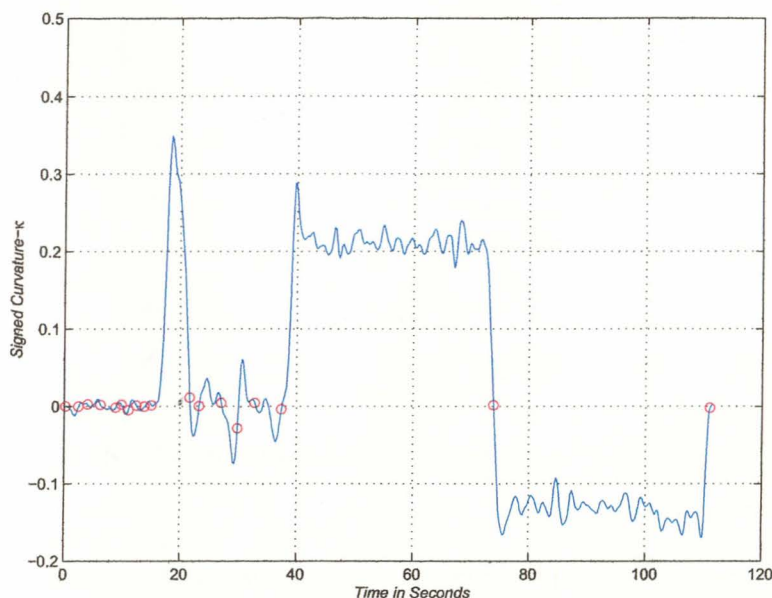
Fig. 2.12: Curvature zero crossings.

After calculating heading at each zero crossing, locating the segment center points, the data set is fitted with the circular arcs and straight segments The fitted data set is shown in red and the original vehicle position data set is shown in blue. The plot points out both the strengths and weaknesses of the method. The method accurately fits the data set if the curvature varies little between zero crossings, which is the case in the data shown until the last multiple turn. The last turn begins with a larger radius with the radius being reduced near the end of the trajectory. The method tends to favor the smaller radius in the fitted circular arc. This problem can be solved by the addition of one additional point to the list of critical points.

Expanding the plot of Figure 2.3 illustrates how individual segments are fit between data set zero crossings. The first area of interest is the single curve at the
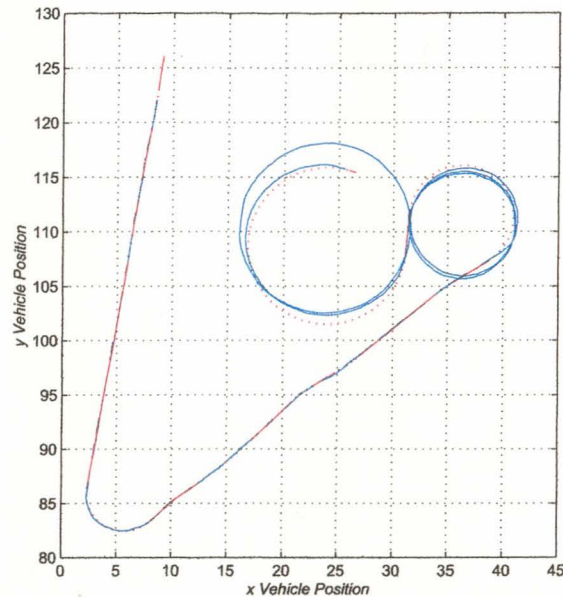
Fig. 2.13: Vehicle trajectory data fitted at zero crossings.

lower left of the trajectory as shown in Figure 2.14. The circular arc is shown as the dotted red line, the straight segments are shown in solid red, and the original data set is shown as a solid blue curve.

The second area of interest is selected as a nearly straight segment on the trajectory. The fitted data set is shown in Figure 2.15.

One of the primary advantages of this method is that it provides significant reduction in the number of segments required to represent the segment as described in the next paragraph. The original data set is reduced from five hundred fifty segments to thirty six segments. Also, no effort has been made to optimize the number of points required to fit the data. It is probable that a data set with many nearly straight segments could be reduced even more.

The number of segments that must be stored is reduced by a factor of approximately fifteen for this data set. A series of ten randomly generated paths was also fit
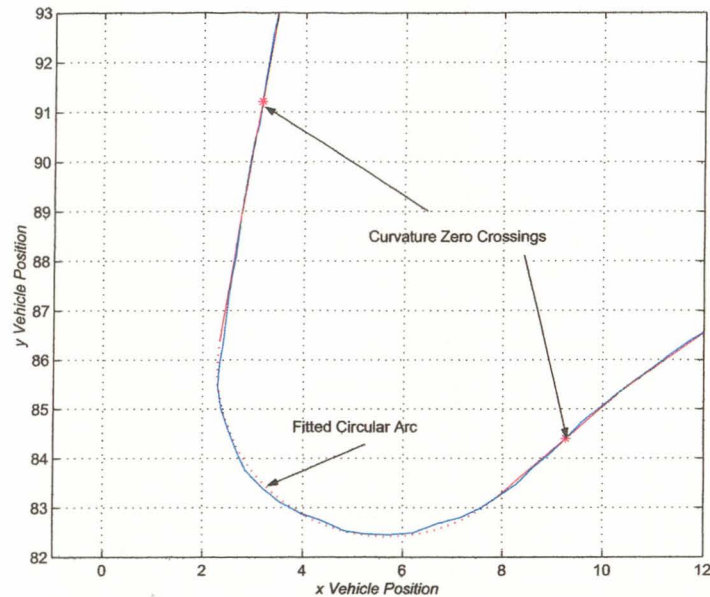
Fig. 2.14: Closeup of the data fit to a single curve.

with the average number of segments being reduced by a factor of greater than eighteen with a standard deviation of approximately seven. A sample random trajectory with the fitted data overlaid is shown in Figure 2.16.

The resulting path derived from the vehicle trajectory is general in the sense that it consists only of circular arcs and straight line segments. Recall that the resulting path is not strictly driveable at any non-zero velocity. The next segment describes the addition of the spiral segments to make the path driveable for a specific velocities and steering parameters.

2.5 The Need for Spiral Segments

It is obvious that the steering system of a vehicle cannot instantaneously transition from one steering angle to another. During the finite transition time, the vehicle is describing a trajectory with a time-varying curvature. This section describes in detail the connecting spiral segments, the mathematical description of these segments, and finally the addition of these segments to the general path.
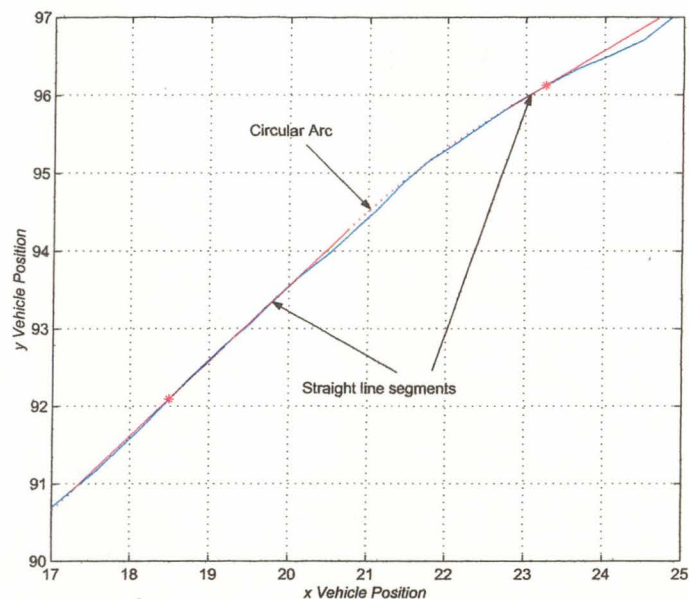
Fig. 2.15: Closeup of the data fit to a nearly straight segment.

The method of the previous section provided a trained path that consisted of tangentially connected circular arcs and ignored any transition region between these arcs. The trained path is reasonable for vehicles travelling at low relative speeds in comparison to the response of the vehicle steering system. However, as the vehicle speed increases or equivalently as the path following accuracy requirement becomes stricter, the transition region between path segments caused by limitations of the vehicle steering system and/or vehicle velocity must be accounted for in deriving paths for high accuracy tracking. Given that the problem of producing driveable paths has been reduced to practice by highway designers, it is reasonable that the methods used for highway design can be adapted to the problem of producing driveable path commands from vehicle training data set. The methods used by highway designers to produce driveable roads are examined in this next section followed by a derivation of the mathematics of transition segments between circular arcs.
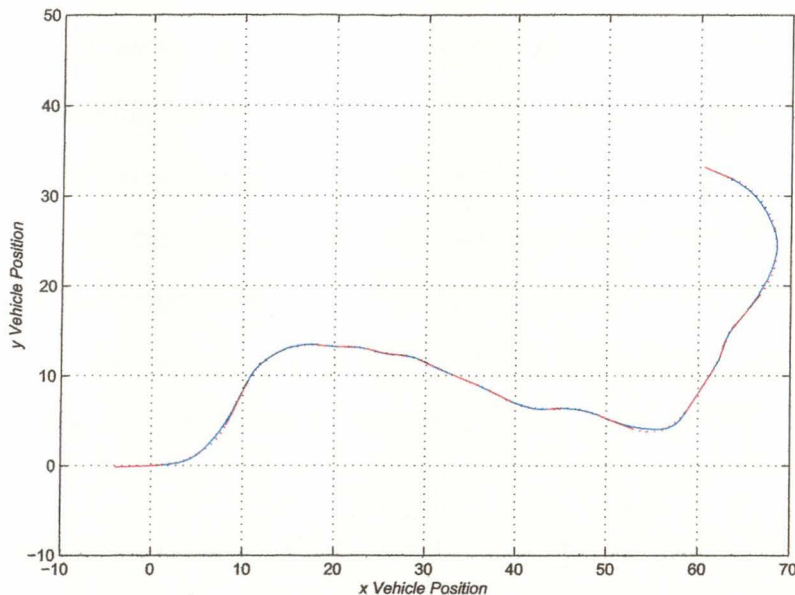
Fig. 2.16: Randomly generated vehicle trajectory with data fit.

2.5.1 Highway Design Methods

The methods used by highway designers to construct driveable roads have application in the area of path planning and training for autonomous vehicles in that the geometry of the two problems and the desired end results are the same. The application of highway design methods to the problem of path planning is obvious; the vehicle must follow a prescribed path through an environment. The application to path training is less obvious unless this problem is viewed as near real-time highway design problems. For example, in the path training problem, the desired path is specified when a human operator drives along the desired path. The path is then converted into segments that are driveable by the autonomous vehicle. Thus the path through the environment is specified by the act of driving through the environment and the 'road' through the environment is designed by translating the driven path into driveable path segments.

The highway designer is constrained by the kinematics of the vehicles that must drive on the road, by the driver's ability to steer the vehicle, and by the response of the vehicle steering system. The most common vehicles seen on roads are those with Ackerman steering and these vehicles can only drive circular arcs, straight lines, and transition segments. The restrictive paths travelled by vehicles with Ackerman steering makes the designation of paths and path transitions important for accurate operation of the vehicles. In the case of a vehicle driving on a highway, accurate operation implies safe operation in that the vehicle stays on the road and in the correct lane.

When designing a highway, the designer should minimize the amount of steering effort that the driver must exert in order to follow the road as well as considering the driver's ability to steer the vehicle, particularly the rate of change of steering angle that the driver can produce. Minimizing steering effort implies that the transitions between straight and curved segments of the road must be smooth so that the driver can meet the required segment radius change in the available space. Highway designers meet the requirement for smooth transitions in two common ways: tangentially connected circular arcs and straight segments or circular arcs and straight segments connected by transition segments. The designers account for the driver's steering response by limiting path segment radius change based on the vehicle velocity. The two common highway layout methods are described in the next two paragraphs.

Highway designers normally lay out roads using tangentially connected circular arcs and straight segments if the speed limit on the road is less than forty miles per hour. The designers are aware that the transition will likely require that the driver use some of the lane width to transition to the next segment, i.e. the driver will probably not stay on the exact lane centerline while making this transition. The centerline deviation required to transition to the next tangential segment is deemed acceptable at lower speeds and the road is designed with no special transition

segments. The amount of centerline deviation depends on a combination of vehicle speed, the x-y point where the driver begins the transition to the new radius, and the time rate of change of steering angle that the driver produces. If the driver begins the transition to the new radius at the correct distance for the vehicle velocity, it is possible at low speeds for the vehicle to transition to the new segment with little deviation from the lane centerline. The low speed transition is shown for a velocity of 0.5 meters per second distances in Figure 2.17.
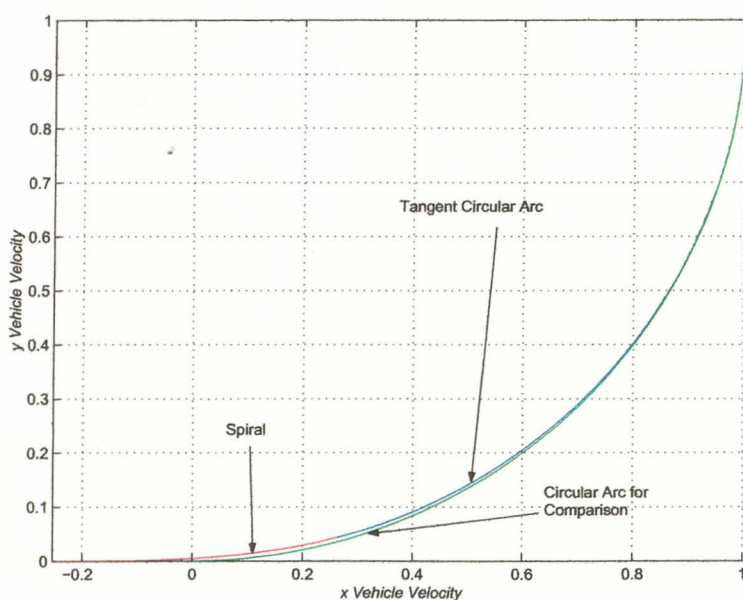
Fig. 2.17: Low-speed spiral transition.

If the vehicle speed is over forty miles per hour, highway designers normally connect circular arcs and straight segments with a transition segment [16]. The transition segment provides space for the driver to transition from one steering angle to the next. While any number of possible transition segments exist, the most commonly used segment is referred to in highway design texts as a logarithmic spiral. In more exact mathematical terms, the spiral used by highway designers is a clothoid spiral. The clothoid spiral has the property that when parameterized by

arc length, the curvature of the spiral is a linear function of arc length. Note that the arc length of a curve being driven at velocity v is given by vt and for constant velocity the curvature is also a linear function of time. The clothoid spiral for a velocity of 1.33 meters per second is shown in Figure 2.18. As mentioned above, the use of the clothoid spiral has been suggested for planning paths in [9] where the authors recommend fitting points on the planned path with multiple clothoid segments. The disadvantage of creating paths that consist of only clothoid segments is that the steering system must be constantly moving when the vehicle driving a curve. Additionally, the steering must be controlled at varying rates and this complicates the system design as well as making the stability analysis more difficult.
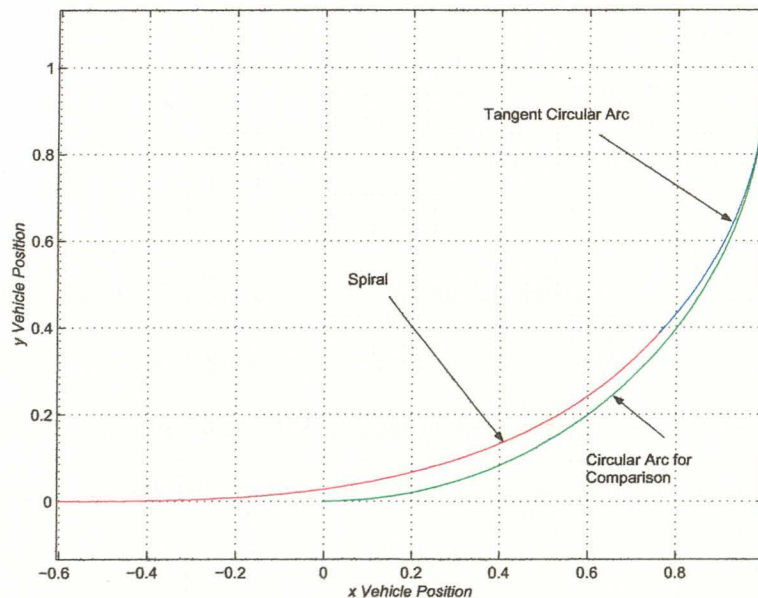
Fig. 2.18: High-speed spiral transition.

The clothoid spiral approximates well the path that a human driver naturally follows when making the transition between straight and circular arc segments and vice versa. Examination of the curvature of Figure 2.19, an expanded view of a section of Figure 2.9, shows that the segment connecting two relatively flat regions

is nearly a straight line. The linear change of curvature with time results in a sloped straight line section in curvature.
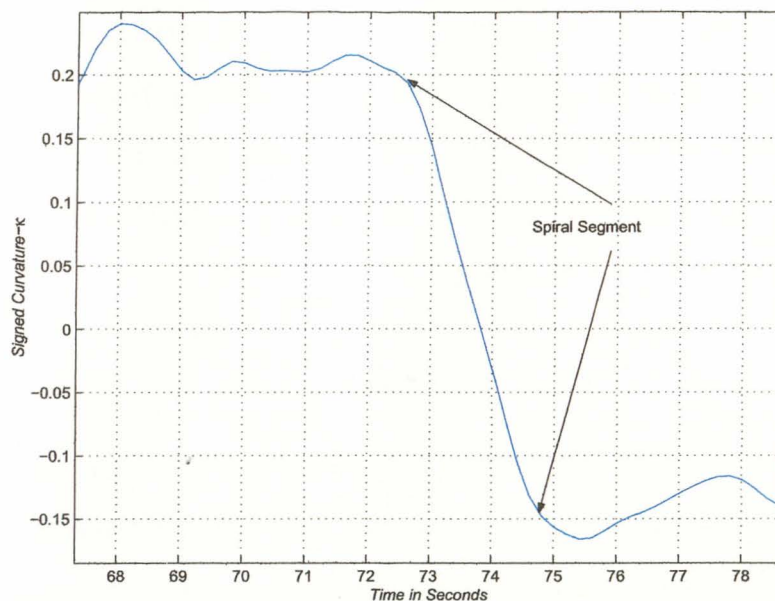
Fig. 2.19: Spiral section in calculated curvature.

2.5.2 Mathematical Definition of Clothoid Spirals

As mentioned above, the clothoid spiral occurs when the vehicle steering is manipulated such that the time rate of change of curvature is constant. The system of Figure 2.10 can be driven with a linearly varying input to produce a clothoid spiral. The equations that describe the spiral are:

$$x = \sqrt{\frac{v\pi}{C_0}} \int_0^{\sqrt{\frac{vC_0}{\pi}}t} \cos(\pi x^2/2)dx \qquad (2.24)$$

$$y = \sqrt{\frac{v\pi}{C_0}} \int_0^{\sqrt{\frac{vC_0}{\pi}}t} \sin(\pi x^2/2)dy, \qquad (2.25)$$

where C_0 is the time rate of change of curvature and v is the vehicle velocity.

The spiral drawn by plotting x vs. y is a clothoid spiral. The integrals of (2.24) and (2.25) are called Fresnel integrals and have no solution in elementary functions. The absence of a solution in terms of elementary functions is not troubling in that most of what are referred to as elementary functions, e.g. the sine function, are numerically estimated at all but a handful of points. A simple approximation for the clothoid spiral can be obtained by expanding both x and y in terms of t using a Taylor's series. The first few terms of the resulting series for x and y are:

$$x \cong vt - \frac{v^3 C_0^2 t^5}{40} \tag{2.26}$$

$$y \cong \frac{v^2 C_0 t^3}{6}. \tag{2.27}$$

The approximations of (2.26) and (2.27) adequately represent the clothoid spiral for Ackerman steered vehicles with reasonable minimum turning radii and steering angle rate limit. A plot of x vs. y is shown in Figure 2.20. The approximate values of equations of (2.26) and (2.27) are plotted with MATLAB's more accurate approximation for comparison. For the plot, $C_0 = v = 1$ and the wheelbase, $L = 2$. The resulting steering angle at $t = 1$ is then 1.1 radians or 63 deg, indicating a steering rate of change of $\frac{63 \, \text{deg}}{\text{second}}$. The approximations can be made more accurate by adding additional terms to the series.

The clothoid curve possesses two useful properties that simplify calculations. The first property is the tangent angle ϕ in terms of time t, which is naturally the vehicle heading:

$$\phi = \kappa \frac{t}{2}. \tag{2.28}$$

The second useful property defines linearly varying curvature in terms of time as:

$$\kappa(t) = C_0 t. \tag{2.29}$$

While the time parametrization of the spiral provides insight into the form of the spiral, it is useful to look for an alternative parametrization that represents the
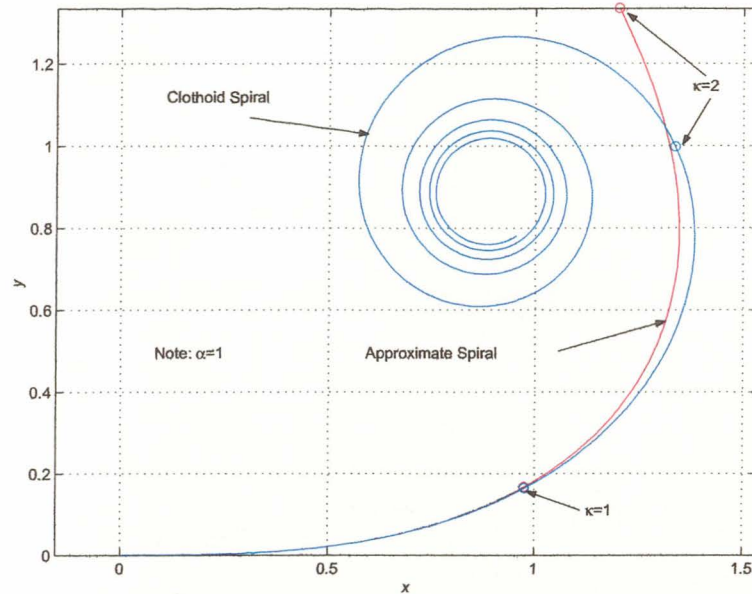
Fig. 2.20: The clothoid spiral and two-term Taylor's series approximation.

spiral as spatial object and defines the spiral in terms of critical parameters. Such a parametrization, based on curvature, is derived in the next section.

2.5.3 Parameterization of Spiral Segments in Curvature

The shape of a connecting spiral between straight segments and/or circular arcs is dependent on both the vehicle velocity and the time rate of change of curvature of the vehicle, and as stated previously, velocity independent paths are not possible. In other words, it is not possible to define a path with spiral segments specified for a velocity and time rate of change of curvature and then change the velocity of the vehicle without adjusting the shape of the spiral segment or the time rate of change of curvature. The dependence on both velocity, time rate of change of curvature, and the curvature is made clear by re parameterizing the spiral segment using curvature, κ, as the vector function parameter. Recalling (2.29) and solving for time results in:

$$t = \frac{\kappa}{C_0} \qquad (2.30)$$

Substituting (2.30) into (2.26) and (2.27) yields the curvature parameterized x and y components of the spiral shown below:

$$: \quad \begin{aligned} x(\kappa) &= \tfrac{v}{C_0}\kappa - \tfrac{v^3}{C_0^3}\kappa^5 \Big/ 40 \\ y(\kappa) &= \tfrac{v^2}{C_0^2}\kappa^3 \Big/ 6. \end{aligned} \qquad (2.31)$$

Noting that a common ratio exists in all the terms of (2.31) results in a further simplification. Let:

$$\alpha = \frac{v}{C_0}; \qquad (2.32)$$

then (2.31) can be written as:

$$x(\kappa) = \alpha\kappa - \frac{\alpha^3\kappa^5}{40} \quad \text{and} \quad y(\kappa) = \frac{\alpha^2\kappa^3}{6}. \qquad (2.33)$$

The form of (2.33) is convenient in that a spiral segment can be completely specified by its beginning and ending curvatures plus the coefficient α. The parameter α uniquely defines the shape of any spiral segment. Qualitatively, if α is large the spiral is longer for a given change in curvature. α is large when velocity is large or when the time rate of change of curvature is small and viceversa.

The tangent angle, or equivalently, the vehicle heading can also be expressed simply in terms of κ and α as:

$$\phi = \frac{\alpha\kappa^2}{2} \qquad (2.34)$$

As stated in a previous section, the spiral segments present themselves in the curvature as nearly straight line sections connecting two different constant curvatures. The geometry of the spiral is completely defined by the parameter α with its orientation defined by the starting and ending orientation of the circular arcs it connects.

2.6 Adding Spiral Segments to the General Path

The previous section described spiral segments, why they are needed for accurate path specification, and the basic mathematical relationships defining the spiral. In this section, the equations defining spiral connections between tangentially-connected segments of different curvature are derived and solved for the general case. The latter part of the section focuses on the special case of connecting straight line segments to circular arcs with a spiral segments of varying shapes. Lastly, the generalized path derived for the example used throughout this chapter is fitted with several spirals of differing parameter α with the results displayed graphically.

2.6.1 The Defining Equation of the Circular Arc/Spiral/Circular Arc Combination

The system of nonlinear differential equations specified in (2.10) were solved as a step in translating the training data set into the general path. Now it is necessary to solve this system of equations to add spiral segments to the general path. The input to this system of equations is the time rate of change of heading or the product of velocity, v, and curvature κ. In this case the input is defined as follows:

$$v\kappa(t) = \begin{cases} v\kappa_0 & \text{if} \quad t \in [t_0, t_1] \\ v(C_0(t - t_0) + \kappa_0) & \text{if} \quad t \in [t_1, t_2] \\ v\kappa_1 & \text{if} \quad t \in [t_2, t_3] \end{cases} \tag{2.35}$$

Equation (2.35) indicates that the vehicle is initially commanded to drive a circle of curvature κ_0, followed by a segment of linearly changing curvature with rate of change of curvature C_0, and concluding with a circle of curvature κ_1.

The equations describing the combination of a leading circular arc, connecting spiral segment, and a trailing circular arc are:

$$x(t_3) = v \int_{t_2}^{t_3} \cos\left(\int_{t_2}^{t} v\kappa_2 dt + \phi_2\right) dt$$

$$+ v \int_{t_1}^{t_2} \cos\left(\int_{t_1}^{t} v\left(C_0\left(t - t_0\right) + \kappa_0\right) dt + \phi_1\right) dt$$

$$+ v \int_{t_0}^{t_1} \cos\left(\int_{t_0}^{t} v\kappa_0 dt + \phi_0\right) dt$$

$$+ x_0 \tag{2.36}$$

$$y(t_3) = v \int_{t_2}^{t_3} \sin\left(\int_{t_2}^{t} v\kappa_2 dt + \phi_2\right) dt$$

$$+ v \int_{t_1}^{t_2} \sin\left(\int_{t_1}^{t} v\left(C_0\left(t - t_0\right) + \kappa_0\right) dt + \phi_1\right) dt$$

$$+ v \int_{t_0}^{t_1} \sin\left(\int_{t_0}^{t} v\kappa_0 dt + \phi_0\right) dt$$

$$+ y_0. \tag{2.37}$$

The above equations include initial conditions and describe an arbitrary combination of arc/spiral/arc in any position or orientation in the $x - y$ plane.

Note the similarities between (2.36) and (2.37) and those of (2.12) and (2.13). The integrals of (2.12) and (2.13) define two tangentially connected circular arcs connecting the points (x_0, y_0) and (x_2, y_2) with starting curvature κ_0 and ending curvature κ_1. The connecting point is (x_1, y_1). Equations (2.36) and (2.37) define a starting circular arc with curvature κ_0 and ending at the point (x_0, y_0). Starting at (x_0, y_0) is the connecting spiral segment that begins with curvature κ_0 and ends at (x_1, y_1) with curvature κ_1. Finally, the last segment is a circular arc beginning at (x_1, y_1) with curvature κ_1 ending at (x_2, y_2).

As previously indicated, the solution of the spiral portions of (2.36) and (2.37) is a combination of Fresnel cosine and sine integrals. It is convenient to look at the spiral segments separately and these segments are shown below:

$$x_{spiral} = v \int_{t_1}^{t_2} \cos \left(\int_{t_1}^{t} v \left(C_0 \left(t - t_1 \right) + \kappa_0 + \phi_1 \right) dt \right) dt \qquad (2.38)$$

$$y_{spiral} = v \int_{t_1}^{t_2} \sin \left(\int_{t_1}^{t} v \left(C_0 \left(t - t_1 \right) + \kappa_0 + \phi_1 \right) dt \right) dt. \qquad (2.39)$$

Given that the choice of t_1 is arbitrary, it can be set to zero without loss of generality. Expanding (2.38) and (2.39) results in:

$$
\begin{aligned}
x_{spiral} = \sqrt{\alpha \pi} \Bigg[&\cos(\frac{\alpha \kappa_0^2}{2} - \phi_1) C \left(\sqrt{\alpha \pi} \frac{(\kappa_1 + \kappa_0)}{\pi} \right) \\
&+ \sin(\frac{\alpha \kappa_0^2}{2} - \phi_1) S \left(\sqrt{\alpha \pi} \frac{(\kappa_1 + \kappa_0)}{\pi} \right) \\
&- \cos(\frac{\alpha \kappa_0^2}{2} - \phi_1) C \left(\sqrt{\alpha \pi} \frac{\kappa_0}{\pi} \right) \\
&- \sin(\frac{\alpha \kappa_0^2}{2} - \phi_1) S \left(\sqrt{\alpha \pi} \frac{\kappa_0}{\pi} \right) \Bigg]
\end{aligned}
\qquad (2.40)
$$

$$
\begin{aligned}
y_{spiral} = \sqrt{\alpha \pi} \Bigg[&\cos(\frac{\alpha \kappa_0^2}{2} - \phi_1) S \left(\sqrt{\alpha \pi} \frac{(\kappa_1 + \kappa_0)}{\pi} \right) \\
&- \sin(\frac{\alpha \kappa_0^2}{2} - \phi_1) C \left(\sqrt{\alpha \pi} \frac{(\kappa_1 + \kappa_0)}{\pi} \right) \\
&- \cos(\frac{\alpha \kappa_0^2}{2} - \phi_1) S \left(\sqrt{\alpha \pi} \frac{\kappa_0}{\pi} \right) \\
&+ \sin(\frac{\alpha \kappa_0^2}{2} - \phi_1) C \left(\sqrt{\alpha \pi} \frac{\kappa_0}{\pi} \right) \Bigg].
\end{aligned}
\qquad (2.41)
$$

C() and S() are the Fresnel cosine and sine integrals, respectively. Note that the Fresnel integrals are evaluated at discrete points only. The resulting equations are combined with the equations for the two circular arcs in order to obtain a solution. The spiral components are functions of the curvature at the beginning and end of the spiral segment. The circular arc segment integrals of (2.36) and (2.37) are solved and

rewritten in the same manner as those used in (2.12) and (2.13) The circular arcs are written in terms of angles at the endpoints of the arcs. The resulting equations with all components included are:

$$x(t_3) = \frac{\sin(\phi_3) - \sin(\phi_2)}{\kappa_1} + x_{spiral} + \frac{\sin(\phi_1) - \sin(\phi_0)}{\kappa_0} + x_0 \qquad (2.42)$$

$$y(t_3) = \frac{-\cos(\phi_3) + \cos(\phi_2)}{\kappa_1} + y_{spiral} + \frac{-\cos(\phi_1) + \cos(\phi_0)}{\kappa_0} + y_0. \qquad (2.43)$$

Recall that the relative tangent angle at the end of the spiral segment is related to the curvature at the endpoint by (2.34). The angle ϕ_2 becomes:

$$\phi_2 = \frac{\alpha \kappa_1^2}{2} + \phi_1, \qquad (2.44)$$

and (2.42) and (2.43) can be written as:

$$x(t_3) = \frac{\sin(\phi_3) - \sin(\frac{\alpha \kappa_1^2}{2} + \phi_1)}{\kappa_1} + x_{spiral}$$
$$+ \frac{\sin(\phi_1) - \sin(\phi_0)}{\kappa_0} + x_0 \qquad (2.45)$$

$$y(t_3) = \frac{-\cos(\phi_3) + \cos(\frac{\alpha \kappa_1^2}{2} + \phi_1)}{\kappa_1} + y_{spiral}$$
$$+ \frac{-\cos(\phi_1) + \cos(\phi_0)}{\kappa_0} + y_0. \qquad (2.46)$$

Equations (2.45) and (2.46) constitute a pair of simultaneous nonlinear equations. Normally, the equations are solved for κ_1 and the segment extension parameter. This parameter is an angle when the curvature is non-zero for one of the arcs and a linear extension term when a straight line segment is one of the segments.

As was the case in (2.16) and (2.17), an important special case is when one of the segments connected to the spiral is a straight line. Assume that the straight line segment is the starting segment, implying that the angles ϕ_0 and ϕ_1 will be equal and that κ_0 will be equal to zero.

Using the results of the derivation of (2.18) through (2.21) the reduced equations are:

$$x(t_3) = \frac{\sin(\phi_3) - \sin\left(\frac{\alpha\kappa_0^2}{2} + \phi_0\right)}{\kappa_1} + x_{spiral} + \cos(\phi_0)p + x_0 \qquad (2.47)$$

$$y(t_3) = \frac{-\cos(\phi_3) + \cos\left(\frac{\alpha\kappa_0^2}{2} + \phi_0\right)}{\kappa_1} + y_{spiral} + \sin(\phi_0)p + y_0. \qquad (2.48)$$

Equations (2.47) and (2.47) can be solved for the parameter p and the ending spiral curvature κ_1.

The data set first plotted in Figure 2.3 was converted to a general path in Figure 2.13. Now it is possible to add spiral segments to the path by solving (2.47) and (2.48). The fit of the curve used in Figure 2.14) used to illustrate the conversion of the training data set to the general path is fitted with spiral segments. The fit is illustrated for two different α and the resulting vehicle/mission specific paths are shown in Figures 2.21 and 2.22.
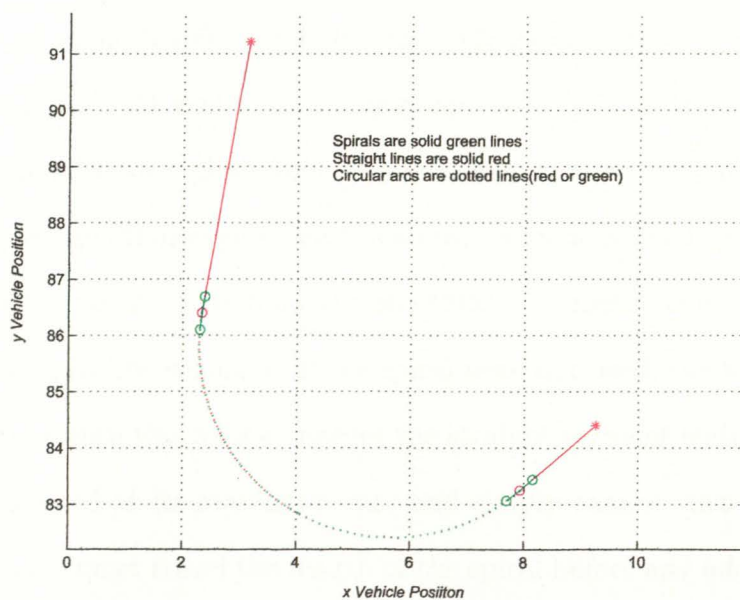
Fig. 2.21: Addition of small α spiral segments to a base path.
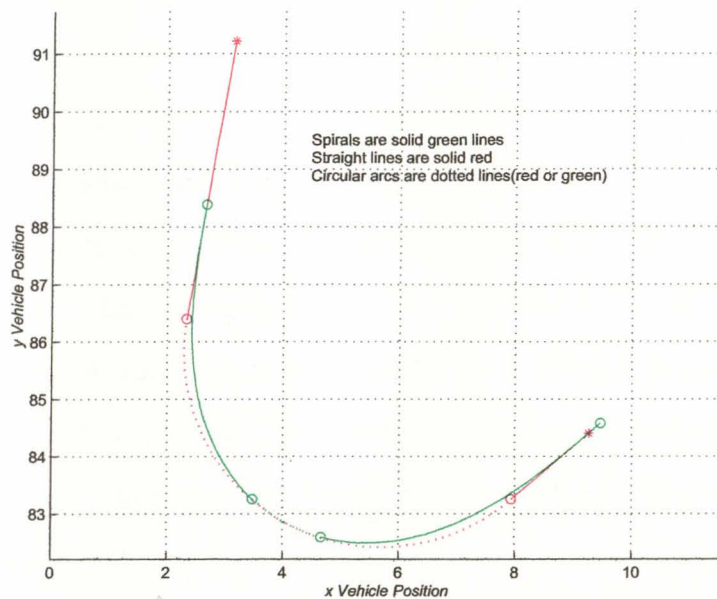
Fig. 2.22: Addition of large α spiral segments to a base path.

The spiral fit of Figure 2.21 has parameter $\alpha = 2$ and the resulting spiral segments are short, as expected. The circular arc following/proceeding the spiral segments has a radius (3.405m) that is nearly identical to that of the general path circular arc (3.411m). At first glance it may appear that adding the spiral segment for such a small value of alpha might be unnecessary. However, the spiral segment supplies a needed function even when the spiral segment is small. Observe that the spiral segment begins at approximately one third of a meter before the end of the general path straight line segment. If the spiral were not used, the the vehicle would not begin turning until the vehicle reaches the straight segment endpoint. Normally, the vehicle is steered at its maximum rate and simply cannot turn the vehicle any faster. The vehicle must travel the length of the spiral before any additional steering correction can be provided to bring the vehicle back to the path. By the time the spiral end is reached, the vehicle will be significantly off of the desired path with an

incorrect heading. Significant control effort must be expended to bring the vehicle back to the path.

The spiral fit of Figure 2.22 has parameter $\alpha = 15$ with the spiral segment being significantly longer. Despite the length of the spiral, the closeness of fit to the general path remains good. The circular arc between the spiral segments is short and the radius (2.765m) is somewhat smaller than the radius of the circular arc of the general path (3.411m). There is no doubt that a spiral is needed for this value of α given that the fitted spiral begins approximately two meters before the end of the general path straight line segment. If one were to wait until the end of the straight segment, significant control effort is needed to return the vehicle to the desired path.

The spiral/arc/spiral combination of the plot of Figure 2.22 includes a short arc between the spiral segments. The length of the included arc will go to zero for some value of α with that value being:

$$\alpha_{\max} = \frac{2(\phi_3 - \phi_1)}{\kappa_1^2}. \tag{2.49}$$

Paths of radius 3.504 meters cannot be made driveable for speeds and/or curvature rates of change with ratio above α_{max}.

Note that the above constraints on driveability of the path assume that the vehicle can travel the path at the maximum possible velocity without sliding, skidding, or rolling.

2.6.2 Discussion of Segment Switching Criteria

In Chapter 1 the need for a criteria for switching from segment to segment was introduced. Given the steering system response, some anticipation of the segment end must exist in whatever criterion that is selected. The simplest method changes the segment at some point before the segment end. The point can be selected based on vehicle velocity and the amount that the steering must change. The switching

distances have been empirically derived for experimental vehicles. The resulting relationships are not easily translated into simple equational relationships and the performance of the vehicle normally degrades with increasing speed. However, the addition of spiral segments eliminates the need for such an empirical approach.

The spiral segment naturally provides a switching point for a smooth transition between segments. Additionally, the spiral segment is driveable and leads the vehicle to a driveable segment. The steering rate limit that is imposed by the steering system can naturally provide the spiral or the steering rate limit can be artificially provided by the control system. Spirals can be added to the path, whether trained or planned, as a post-processing step.

The addition of spiral segments to the general path has advantages at both low and high speeds. The segments provide a semblance of vehicle and velocity independence and a simple, driveable scheme for changing from segment to segment.

2.7 Systematic Summary of the Path Training Methods

The general path is derived using the zero-crossing data fit presented in this chapter are summarized in a stepwise fashion below:

1. Obtain the vehicle position data set by driving the vehicle on the desired path while measuring position.

2. Zero-phase, low-pass filter the $x-y$ time sequences to reduce the high frequency measurement noise.

3. Calculate the curvature of the path at each point.

4. Locate curvature zero crossings, the crossings provide the critical zero crossings.

5. Using equations derived in Section 2.4, calculate the straight line extensions and the circular arc that fit to the segment center.

6. Fit between each zero crossing index until the end of the data set.

The addition of spiral segments to the general path is accomplished via the following steps:

1. Given a straight line/circular arc/straight line find the angular center point.

2. Solve the equations derived in Section 2.6 for the given end and center points and for the vehicle/mission parameter α.

2.8 Summary

This chapter has developed a task management strategy that uses the expertise of a human operator to train the autonomous vehicle. The training method is compared to the pendant training of a stationary robot in that in both cases the human operator moves the machine through the desired task. Vehicle training differs from pendant training in that the data taken from vehicle training consists of the desired vehicle path while the data taken during pendant training consists of the commands that move the stationary robot through the desired motion. The command data set used in pendant training is used in an open loop fashion to move the stationary robot through a desired task without the benefit of feedback. The vehicle training data set is processed to provide a series of path commands that are applied to the closed loop vehicle path management system. The path management system is designed to be relatively insensitive to external disturbances. It has been shown that the task management function for a trained autonomous vehicle can be reduced to the determination of a sequence of driveable path segments from training data set and the addition of unique spiral segments to the path for different vehicles and/or velocities. The training method offers the significant advantage of simplicity over methods that require precise measurement of the environment that the vehicle traverses.

The training data set was converted to produce paths consisting of circular arcs and straight segments. This is the so-called general path. Next, a method was developed to include spiral transition segments that tangentially connect the circular and straight segments and the paths produced by this method are driveable in a mathematical sense.

The next chapter focuses on the development of the vehicle model, which is a necessary precursor to the development of the complete path management function. The vehicle model is derived, analyzed, and simplified in order to facilitate the design of the vehicle controller and enhance the understanding of the stability issues that arise.

Chapter 3
Vehicle Model

Chapter 2 described the task management function for the autonomous vehicle using the training method. The task management function consists of the outer loop of Figure 1.1. This chapter deals with the vehicle model, which is a part of the inner loop of Figure 1.1. The model that is derived in this chapter is for vehicles with so-called Ackerman steering that have a single fixed axle and a steering axle. While the model is restricted to this form of motion it can be easily extended to other Ackerman-equivalent vehicle such as vehicles with two steering axles.

The path management control loop described in Figure 1.2 contains a block that is labelled simply 'vehicle.' The mathematical description of the vehicle derived in this chapter establishes the relationship between the steering angle and the vehicle position as well as a description of the steering actuation system. The kinematic model is frequently used due its simplicity and accuracy. The development of the model allows the analytical design of a controller for the autonomous vehicle and provides the framework for an analysis of stability, both of which are subjects of Chapter 4. The model is presented in two forms, as a nonlinear kinematic model of the vehicle and as a linearized model of the system that contains the critical linear dynamics.

A kinematic model assumes that the geometry of vehicle motion is constrained in mathematically defined way, with the cause of motion not being of primary importance [17]. While the assumptions of kinematic motion do not always accurately describe vehicle motion, the results obtained from the kinematic model do reasonably describe the majority of vehicle motion. In the kinematic model, the vehicle

motion is modelled as the motion of the point exactly between the fixed axle of the vehicle and consists of translations and rotations about this point. The translation is strictly caused by the driving wheels and the rotation is caused by the angle of the steering wheels, with no additional external forces directly considered.

The kinematic model is first described in an intuitive fashion based on the common experience of driving a vehicle. The intuitive description is formalized by making use of the vehicle geometry. The nonlinear model provides an accurate mathematical description of the response of the vehicle in an undisturbed environment. The effects of external forces are added to the kinematic model and finally the steering actuation and rigid body effects are added to complete the model.

The resulting nonlinear model is then linearized to provide a basis for the analysis of vehicle control system stability and the design of the vehicle controller. Lastly, the independent variable time is replaced with a spatial independent variable in the direction of the desired path travel. This change of independent variable allows the specification and design of the spatial response of the vehicle rather than the time response. The spatial response of the vehicle is defined simply as the space that the vehicle fills in response to changes in steering or due to external disturbances. Designing for the spatial response of the vehicle is reasonable for vehicles that must follow tightly-defined paths as in the case of the trained vehicle described in Chapter 2. Also, spatial response is important for vehicles travelling in cluttered environments and environments with frequent external disturbances.

3.1 Kinematic Model for Ackerman Steering

A human driver controls a vehicle so that it follows a specified path, or so that the vehicle at least stays within a specified bound. The driver controls the vehicle by turning the steering wheel. Intuitively, steering changes the heading, or direction of travel of the vehicle, and a change in heading implies that the vehicle will travel on a

new trajectory. Therefore, a series of steering commands makes it possible to move from one path to another. As an example, consider a vehicle turning to the right after travelling along a straight road. The steering wheel is turned by the driver to the right and the vehicle heading begins to change, when the correct heading and position are reached the driver straightens the steering wheel and the vehicle proceeds down the new path at the new heading.

The kinematic model of a vehicle with Ackerman steering is relatively simple to derive and understand. Given steering angle as an input and vehicle position as an output the relationship can be thought of in the following sequence:

(1) Non-zero steering angle creates a rate of change of vehicle heading.

(2) Rate of change of vehicle heading is related to vehicle heading through integration.

(3) Vehicle heading changes the x and y velocity components.

(4) The $\dot{x}$ and $\dot{y}$ velocities are related to the x and y components of vehicle position through integration.

The steering axle that controls a vehicle with Ackerman steering is designed such that when the wheels are turned the vehicle begins to follow a circular path. Each steerable wheel follows a circular path with the outside path having a larger radius than the inside path as illustrated in Figure 3.1. If each of the steerable wheels were turned to the same angle relative to the vehicle then one or both of the wheels would skid while the vehicle turned. The difference in wheel angles makes the definition of a single steering angle a necessity. The simplest definition of steering angle can be made by assuming that the vehicle has only one movable wheel, identical to a child's tricycle, and the steering angle is then defined as the angle of this single wheel
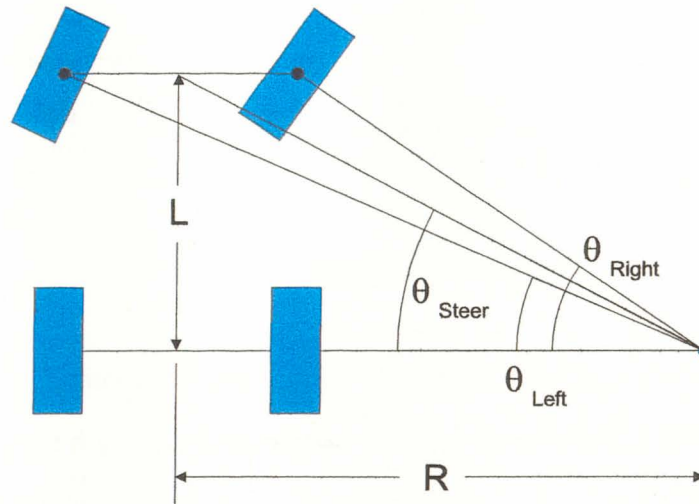
Fig. 3.1: Geometry of Ackerman steering.

relative to the vehicle. This is equivalent to the angle θ_{Steer} shown in Figure 3.1, which we will denote as θ from now on.

The steering angle, θ, is related to the time rate of change of heading, $\dot{\phi}$ as follows:

$$\dot{\phi} = (v/L)\tan(\theta). \tag{3.1}$$

where v is the vehicle velocity and L is the wheelbase.

The vehicle heading, ϕ, relates to the x and y components of velocity as follows:

$$\dot{x} = v\cos(\phi) \tag{3.2}$$

$$\dot{y} = v\sin(\phi). \tag{3.3}$$

The position of the vehicle is found by integrating equations $\dot{x}$ and $\dot{y}$.

The block diagram of the complete system is shown in Figure 3.2.
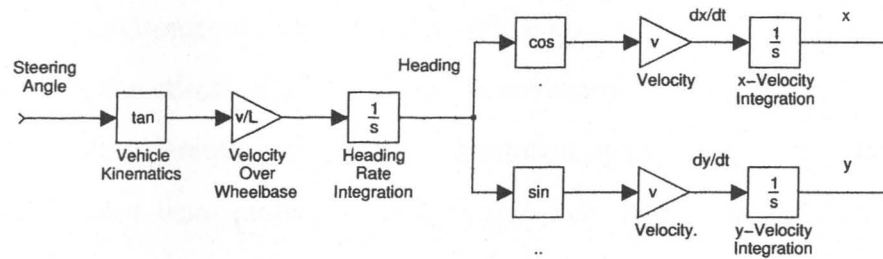
Fig. 3.2: Nonlinear subsystem.

The system in Figure 3.2 is nonlinear with a single input, the steering angle, and two outputs, the x and y. The vehicle position and direction of travel are completely described by the triple (x, y, ϕ).

3.1.1 Modelling External Forces

The model of the vehicle in Figure 3.2 would be complete from steering angle, θ, to the positions x and y if the vehicle operated in an environment free from external disturbance, but external disturbances do exist. The model used in this research is based strictly on vehicle kinematics, i.e. the equations describing the motion of the vehicle do not include the effects of outside forces that cause the vehicle to run outside of its undisturbed geometric constraints. Examples of such forces include the force of gravity on the vehicle when it operates on a steep slope and the forces due to an irregular driving surface that cause the vehicle to be moved off its kinematic course. An alternative to the kinematic model is the kinetic model of the vehicle, which includes the effects of all outside forces on the vehicle.

The planar kinetic model of the vehicle resolves all forces acting on the vehicle into a central force and a couple of moment that act through the center of gravity. External forces such as the effects of stones under the tires or steeply sloped ground cause the vehicle to move in a trajectory that is different from the constrained kinematic motion. These common forces occur constantly as the vehicle is travelling.

The forces are random in nature and their magnitudes and rates of occurrence depend on the driving environment of the vehicle. While a kinetic model of the vehicle could exactly predict the effects of all forces in the environment, this model does not lend itself to ease of understanding or use in a control systems context. However, the kinetic model does lend qualitative insight into vehicle behavior in the real world environment.

The vehicle travelling over rough terrain exhibits displacements and rotations when the wheels collide with objects in the environment. Displacements and rotations are the motions predicted by the kinetic model. As another example, when the vehicle travels on a steep slope it can slide down away from the path predicted by the kinematic model. The sliding is usually a combination of displacement and rotation. Although the environmental forces are not directly modelled in the vehicle model of Figure 3.2, these forces can be included in the kinematic model as external disturbances in heading and position. Linear displacements are modelled as disturbances on the vehicle position while rotations affect the heading. The vehicle model with environmental disturbances included is shown in Figure 3.3.
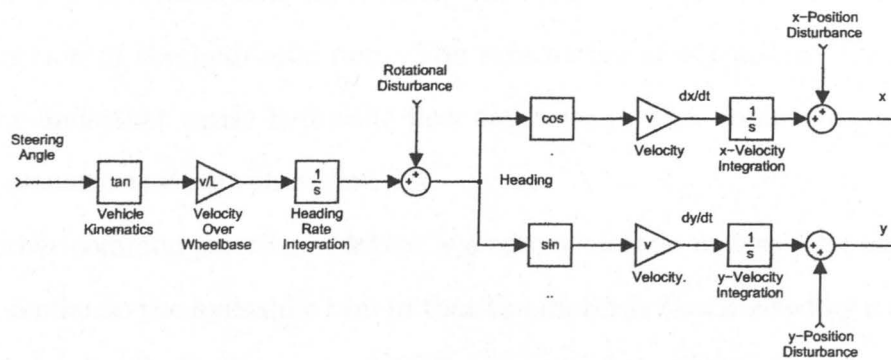
Fig. 3.3: Vehicle kinematic block diagram with external disturbances.

The nature of the disturbances can be understood from their effects on vehicle performance. Disturbances of an impulsive or short duration that cause vehicle rotation are included in the model as causing a heading change with the disturbance

being modelled as a finite duration pulse. Persistent disturbances, such as those associated with the vehicle travelling on a steep slope, are modelled as a step disturbance of the heading together with a displacement disturbance; without corrective action the vehicle slides away from kinematic path while simultaneously rotating. One of the advantages of using a disturbance model instead of the kinetic model is that the effects of disturbances in control systems are well understood. Controllers can be designed so that external disturbances are attenuated.

3.1.2 Steering Actuator Dynamics

An autonomous vehicle includes an actuator for displacing the steering axle and the dynamics of this actuator must be included in order to accurately model the vehicle. Ultimately the maximum speed of the vehicle, at which it can still follow the commanded path, is limited by the rate of response of the steering actuator.

A common steering actuator is a hydraulic ram, which is frequently modelled as an integrator with an appropriate gain term. The steering mechanism model also includes an important nonlinearity that represents the actuator rate limit, or equivalently a hydraulic flow saturation, that acts to limit the speed of extension and retraction of the hydraulic ram. The ram moves at a constant linear rate for flow commands that cause hydraulic flow saturation. The nonlinear model of the steering actuator is shown in Figure 3.4.

Another common steering actuator is a rotary electric motor. The model of the motor is similar to the hydraulic ram in that the motor is rate limited by a saturation identically to the hydraulic ram, with the exception that the saturation is on the input voltage, which is limited either by the motor design or capacity of the motor power source.

Instead of cascading the steering actuator directly with the vehicle model, a feedback loop is placed around the steering actuator with the input to the actuator
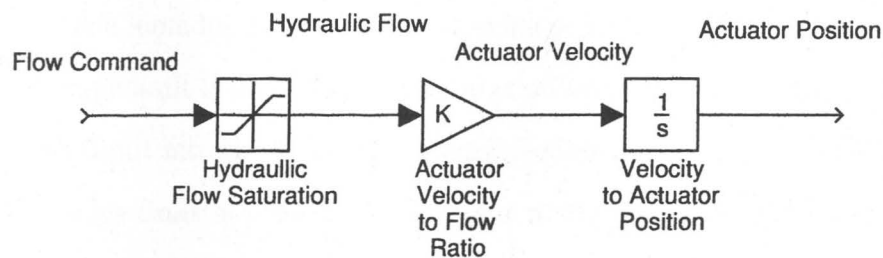
Fig. 3.4: Nonlinear steering actuator model.

in feedback being the desired steering angle and the output being the actual steering angle. The feedback loop around the actuator is sometimes called a minor loop and the addition of this loop has the practical advantage of providing direct control over the vehicle steering angle, thus avoiding problems that could occur when the vehicle is started with an unknown steering angle. Additionally, as will become clear in Chapter 4, the addition of another free integrator to the vehicle system is undesirable for the design of the controller.

Adding the minor loop around the steering actuator also has a mitigating effect on any deadband that the steering system might have. Deadband is understood as a non-zero region or band around zero flow command where, despite non-zero command, no hydraulic flow occurs. The deadband in a hydraulic actuator is analogous to the behavior of a so-called class B power amplifier and the phenomena referred to as crossover distortion. The use of feedback to reduce the effects of a deadband in an amplifier is well understood, as explained in [18]. The effect of feedback on a deadband can be understood as a flow command distortion that adds an offset to the flow command that is adequate so as to overcome the actuator deadband. A similar compensation technique is used for hydraulic actuators and commercially available electrohydraulic servo valve controllers that use this technique nearly completely eliminate actuator deadband.

In addition to the flow saturation and actuator deadband, the steering subsystem of the vehicle includes a limit on the steering angle created by vehicle geometry. The steering angle limit is modelled as a saturation on the commanded steering angle that acts as an input into the steering actuator feedback loop as shown in Figure 3.5. The steering angle limit is placed at this point in the model to effectively stop the model of the actuator from continuing to move when the saturated steering angle is reached thus avoiding the so-called 'wind-up' problem associated with the actuator integrator.
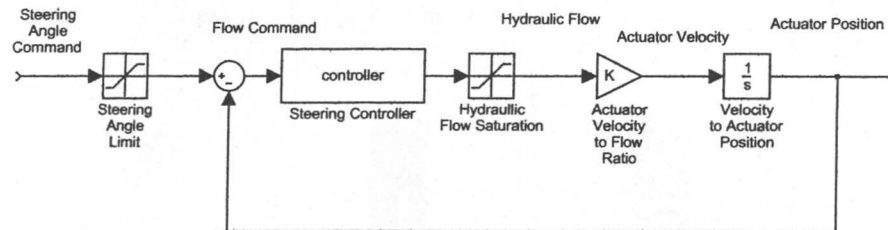
Fig. 3.5: Nonlinear steering actuator model in feedback.

3.1.3 Rigid Body Considerations

The derivation of the kinematic model has treated the vehicle as a single point, the point exactly between the fixed wheels of the vehicle, but the vehicle is a rigid body or collection of points. The model is valid so long as measurements of the vehicle position are made at the modelled point. However, it is not always convenient to measure the vehicle position at this point and the effects of moving the measurement point to another point on the vehicle must be modelled. Additionally, moving the measurement point of the vehicle away from the modelled point has some desirable dynamic effects, e.g. if the measurement point is moved forward of the modelled point the measurement point provides a predictive measure of where the modelled point is going. The predictive measure of the position of the modelled point can be used in the design of the controller.

Any point on the vehicle can be thought of as a vector offset from the point between the fixed wheels of the vehicle. The point is defined absolutely in space by making use of the vehicle heading and the relative position of the measurement point to the point between the fixed wheels of the vehicle, as shown in Figure 3.6.
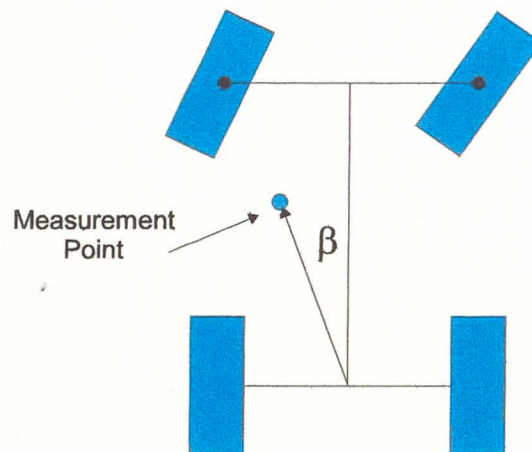
Fig. 3.6: Vehicle with shifted measurement point.

The equations describing the absolute position of the measurement point (x_m, y_m) in space are given as:

$$x_m = x + r_m \cos(\phi + \beta)$$

$$y_m = y + r_m \sin(\phi + \beta).$$

(3.4)

The measurement point is commonly placed along the vehicle centerline, thus making the angle β zero or one hundred and eighty degrees. The addition of the effects of the shifted measurement point completes the nonlinear model of the vehicle.

3.1.4 Completed Nonlinear Model

It is now possible to assemble the complete vehicle model as shown in Figure 3.7. For diagram clarity, the model is shown with the measurement point placed
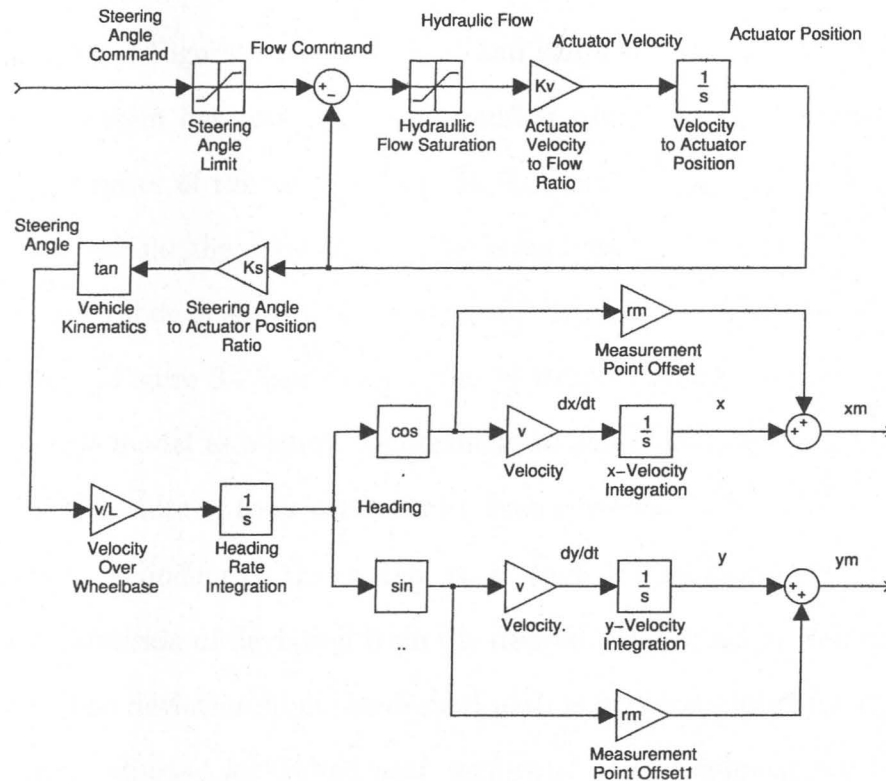
Fig. 3.7: Complete vehicle nonlinear model.

on the vehicle centerline and without external disturbances. The model provides an accurate mathematical description of the vehicle behavior in response to steering commands.

The next step in the modelling process is to linearly approximate the behavior of vehicle. The linear approximation of the vehicle kinematics provides insight into the vehicle behavior, which points the way for the design of the vehicle controller. While the vehicle will be linearized, the effects of the steering angle limit and the hydraulic flow saturation will be considered in detail, given their strong effects on vehicle stability and system performance.

3.2 Linearization of the Vehicle System

The model of Figure 3.7 is linearized and simplified in order to aid controller design. The system has one input, the steering command, and two outputs, the x and y coordinates of the vehicle position. It seems apparent that the linearized vehicle model should therefore be a single-input, multiple-output (SIMO) system and that the methods of state space control will be required. However, examination of the model of Figure 3.7 leads to a series of simplifications and it is possible to view the vehicle model as a single-input/single-output (SISO) system, thus gaining access to the rich suite of tools available for such systems.

The key to reducing of the system to a single-input/single-output system is found in the definition of deviation from the desired path, which is described in the next section. The deviation from the desired path is first completed for straight line paths and then adjusted for curved path segments. Each nonlinear portion is then examined in turn and replaced by a linear approximation. Finally, the independent variable of time is replaced with a spatial independent variable.

3.2.1 Definition of Deviation from Straight Line Path Segments

The choice of coordinate systems for the vehicle is usually limited to either body-centered or world-centered coordinates. A third possibility results in a particularly valuable simplification and this can be thought of as a path-centered coordinate system. In the path-centered system, the vehicle position is measured as the orthogonal distance from the path to the measurement point on the vehicle. This choice of coordinates therefore assumes that the desired path is always along the inertial x-axis, the vehicle position is always on the y-axis, and the coordinate system moves with the vehicle along the desired vehicle path. Deviation from the origin and therefore the desired vehicle path is viewed as the y-coordinate of the vehicle position, which

is the orthogonal error from the origin. This concept is shown in Figure 3.8. The orthogonal error is a convenient measure of deviation from the commanded path. The assumption that the desired path is on the inertial x-axis is always achievable via a coordinate transformation; therefore, no loss of generality is incurred from this simplifying assumption. The definition of orthogonal error from the path coupled with the assumption that the desired path lies on the x-axis, effectively eliminates the x output from the diagram of Figure 3.7 in that all the necessary information about the deviation of the vehicle from the desired path is contained in the y-coordinate of the vehicle position. The system error is the y-position of the vehicle making this system single input/single output.

It is important to note that the absolute magnitude of (3.2) and (3.3) cannot exceed the velocity v. This fact coupled with the definition of deviation from a straight path bounds the maximum rate of growth of error. In the path centered coordinate system, the maximum growth of deviation occurs when the vehicle is travelling in a direction that is orthogonal to the desired path. The result can be generalized to circular paths or paths with smoothly changing curvature given that the rate of growth of deviation from any smooth path is bounded by the velocity.

3.2.2   Definition of Deviation from Circular Paths and Paths with Smoothly Changing Curvature

Defining the deviation from circular segments of any radius is similar to that of a straight segment. Consider a circular path segment at an arbitrary location in the $x - y$ plane with the vehicle at some location not on the path segment as shown in Figure 3.9. Given that the location and orientation of the segment are completely arbitrary, it is possible to use a coordinate transformation to rotate and/or translate the circular segment so that the point on the segment that is closest to the vehicle is at the origin while the vehicle measurement point falls on the y-axis. The vehicle
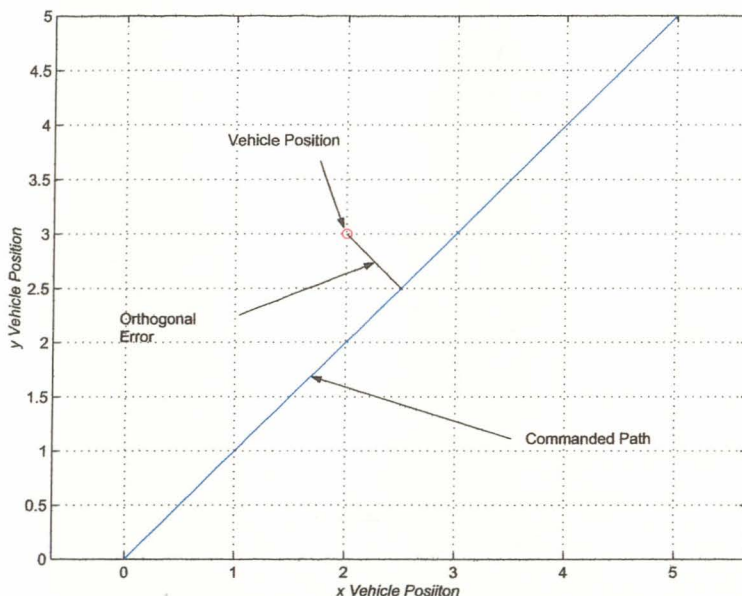
Fig. 3.8: Vehicle deviation from a straight path.

deviation, or orthogonal error, from the path is again defined as the y-coordinate of the vehicle position. This definition of vehicle deviation again eliminates the x-coordinate output from the system, which is important for the linearization that follows. Large deviations that are greater than or equal to the radius and on the concave side of the circular segment, as shown in Figure 3.10, can produce undesired results and care must be taken to deal with such cases. One method of eliminating the ambiguous choice of orthogonal distances is to section paths into shorter segments and measuring the error to the current segment only.

The orthogonal error from an arbitrary plane curve is defined in a similar manner. Let the path be represented as a parameterized curve C with the x and y components defined as functions of a parameter k. The orthogonal distance to the path can be obtained by first differentiating the parametrization of C of (3.5) with respect to the parameter k to get:

$$\mathbf{P}' = x'(k)\widehat{\mathbf{x}} + y'(k)\widehat{\mathbf{y}}. \tag{3.5}$$
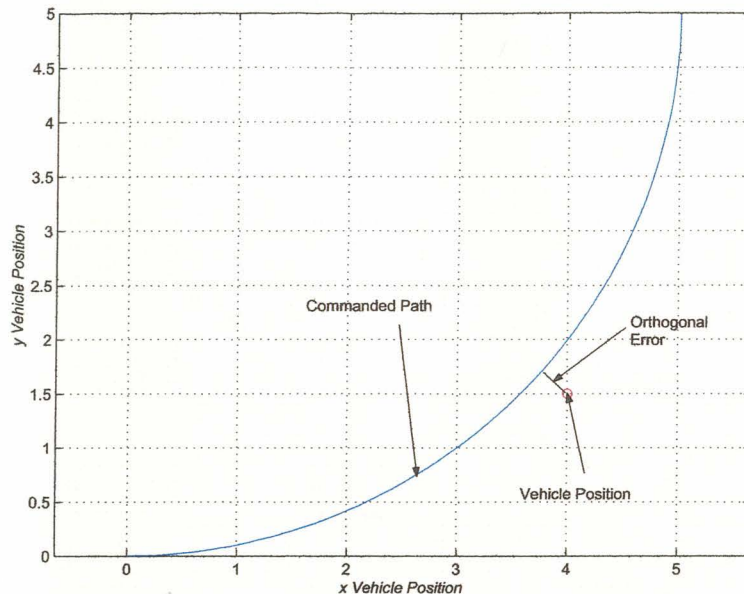
Fig. 3.9: Vehicle deviation from circular path.

The vector, $\mathbf{P}'$, of (3.5) is tangent to the curve C at all points on C as seen in Figure 3.11. The current vehicle position (x_v, y_v) is perpendicular to the curve at some point, $(x(k_0), y(k_0))$, on C. The point $(x(k_0), y(k_0))$ can be found by taking the vector dot product of the tangent vector, $\mathbf{P}'$, and the vector defined by $\mathbf{N} = (\mathbf{x} - \mathbf{x(k_0)}, \mathbf{y} - \mathbf{y(k_0)})$, setting the result equal to zero and solving for k_0. When k_0 is known, the magnitude of the vector $\mathbf{N}$ is the magnitude of the orthogonal error. The sign of the error is found by solving for the z-component of the vector cross product of $\mathbf{P}'$ and the vector $\mathbf{N}$. The vector $\mathbf{N}$ can be rotated such that it lies along the y-axis. The y-component of the vector $\mathbf{N}$ is again the orthogonal error and the output of the system is again reduced to y and the simplification holds for arbitrary curves. The difficulty of solving for the perpendicular point on curve C is of course related to the complexity of the parametrization of C. The solution can become complex for relatively simple parameterizations. As in the case of the circular path segment, the concavity of a curve can create an additional set of problems since the orthogonal
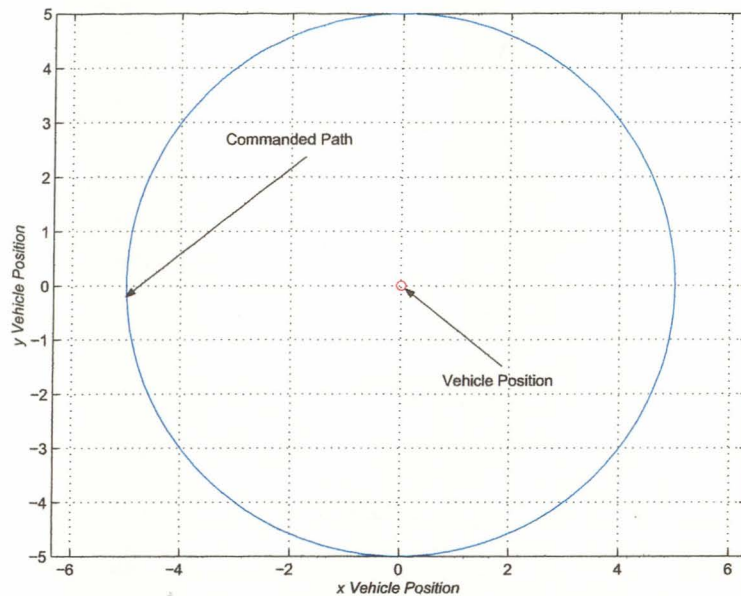
Fig. 3.10: Large deviation from circular path.

error calculation can have multiple solutions. Taking care in the definition of path segments can help solve this problem.

Naturally, the effect of a moving coordinate system must be evaluated for its effect on the kinematic model. The response of the vehicle when it is travelling along a path segment with no orthogonal deviation is instructive. If the vehicle is travelling exactly along a straight line path segment nothing is changing except the x-coordinate, given that the vehicle heading relative to the path is zero. However, for the vehicle exactly following a circular path, the heading, which is related through integration to the vehicle x and y positions, is changing at a constant rate. For the path-centered coordinate system to be valid for circular path segments of finite radii, the vehicle steering must be biased so as to naturally follow the circular segment. The orthogonal error from the path then defines a small corrective steering angle rather than the entire steering angle for the vehicle. In control systems terms, the steering of the vehicle is biased to the drive the radius of the current circular segment.
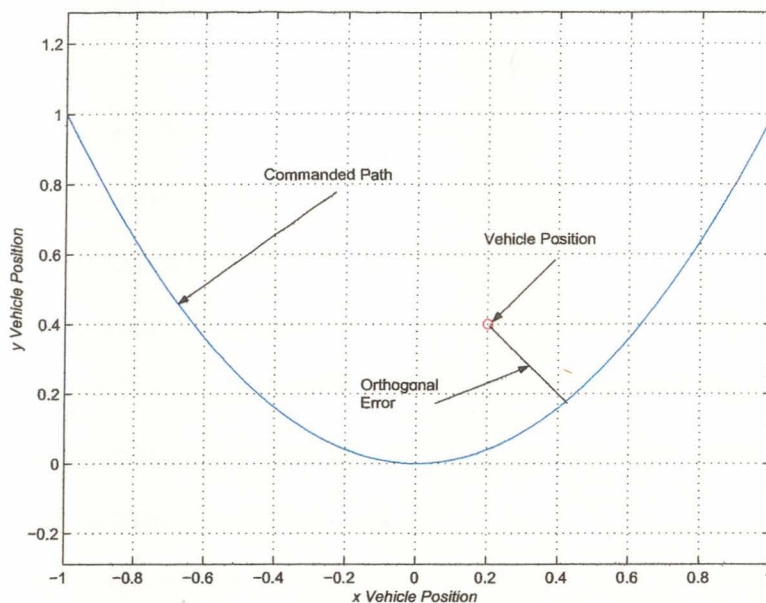
Fig. 3.11: Deviation from an arbitrary curve.

The bias is modelled as a feedforward structure. For circular paths the system is linearized about a non-zero steering angle.

Additionally, the orthogonal deviation can be defined for an arbitrary space curve with continuous curvature. The feedforward bias in the general case is not constant and is required to instantaneously change the vehicle steering angle if the vehicle were to be able to follow the path without corrective steering. The requirement for instantaneous steering response makes precise following of non-circular path segments more difficult.

It has been shown that the system output of Figure 3.7 can be defined for any allowable path segment as the y-coordinate of the vehicle position through a series of rotations and/or translations. The linearized system that is presented below produces a single input/single output system. Before the linearization can be completed, a discussion of the system nonlinearities is required. The method used to calculate the orthogonal error from the desired path is deferred to Chapter 4 in the discussion about the implementation of the control system.

3.2.3 Linear Approximations for Vehicle Nonlinearities

Starting from the steering angle command input at the top left of Figure 3.7 the first nonlinear block that is encountered is the steering command saturation. The linear approximation of this block is a gain of one (simply the gain of the block before the steering command limit is reached).

The next nonlinear block that is encountered following the signal flow from left to right in Figure 3.7 is the hydraulic flow saturation inside of the steering actuator loop. The flow saturation is replaced by a gain of one, this being the gain of the block before flow saturation occurs.

Continuing to follow the signal flow in Figure 3.7 from the output of the steering actuator feedback loop, the next nonlinearity in the path is the tangent function that operates on the steering angle, θ. If the steering angle is small then the tangent function could also be replaced by a gain of one. The range of inputs to the tangent function is limited by the maximum steering command. Typically the tangent function is replaced by a gain of one for the small angle corrections that are required to correct vehicle position errors on straight paths. For any path, the tangent function can be approximated for the steering angle required to produce a trajectory of the path radius. The approximation is simply the slope of the tangent function for a fixed steering angle and is given by:

$$k_{tangent} = \sec^2(\theta_{bias}).\tag{3.6}$$

The equivalent tangent gain can be adjusted based on the current path segment radius, be conservatively set to the maximum slope of the tangent function at the steering angle limit, or set to one if the steering angle limit is relatively small. For the remainder of this deviation the linearized gain of the tangent function will be set to one.

The last nonlinearity of importance is the sine function that operates on the vehicle heading, ϕ. The linear model is valid when the heading angle is small and the sine block can be approximated by a gain of one. The results of the previous section make the x-output zero for any allowable path and the cosine term has no effect on the linearized system. The linearized system is presented in the next section.

3.2.4 Linearized Vehicle Kinematics

Replacing the nonlinear blocks of Figure 3.7 with their linear approximations and then reducing the resulting linear system gives the approximate linear system shown in Figure 3.12.
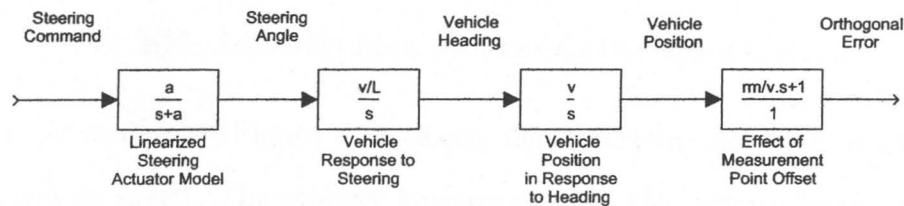
Fig. 3.12: Linearized vehicle kinematic block diagram.

Note that the zero due to the placement of the measurement device is minimum phase only if the measurement device is placed in front of the driving wheels of the vehicle, $r_m > 0$. If the measurement point is placed behind the driving wheels, the zero is non-minimum phase with a negative gain coefficient. It is instructive to look at pole-zero plots of both the minimum and non-minimum phase systems as shown in Figure 3.13. As the measurement point moves away from the driving wheels of the vehicle, the dynamic effect on the vehicle motion becomes more pronounced, or equivalently moves the zero closer to the origin. Note that the plot shown is for a single velocity. The minimum phase system can be stabilized by using a controller and negative feedback, but the non-minimum phase system can only be stabilized by using a controller and positive feedback.
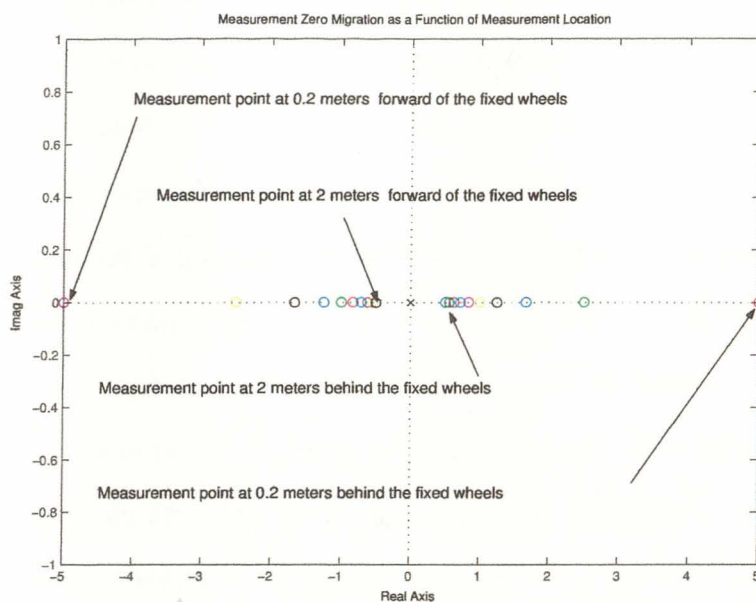
Fig. 3.13: Zero migration with measurement position.

The linear system of Figure 3.12 clearly demonstrates the dependence of the vehicle system on speed. The velocity appears twice in the system shown in Figure 3.12 giving the system a speed squared 'gain' parameter. The strong dependence on vehicle speed implies that the system could be sensitive to even small speed variations. Controllers designed for the autonomous vehicle system must directly account for the variation in speed or be so designed as to make the system insensitive to speed variations.

The section that follows uses the linear system of Figure 3.12 to define a change of independent variable that simplifies the design of a controller for a vehicle with a large range of operating speeds.

3.3 Spatially Dependent Transfer Functions

The Laplace transform of the linearized model of Figure 3.12 is given by:

$$G_v = \frac{av^2(\frac{r_m}{v}s + 1)}{Ls^2(s + a)}. \tag{3.7}$$

The strong dependence of (3.7) on velocity indicates that the system performance changes significantly as the velocity changes. In other words, the identical steering correction commanded at a low speed has a significantly different effect at a high speed. Given the experience of driving an automobile, the difference in reaction to steering command is not unexpected, since large steering angles commanded at low speeds create manageable, and even desirable, vehicle reactions while such large commands at high speeds cause the vehicle to skid or worse. Additionally, the space that the vehicle passes through varies with speed. The question arises as to whether it is possible limit the effect of velocity on the vehicle system and obtain similar responses to steering commands in space rather than time. A solution to this problem is deferred to the controller design in Chapter 4, but the modification of the vehicle model necessary for the controller design is presented in this section.

System responses to inputs and disturbances are traditionally measured in terms of time and magnitude dependent values with terms such as 'rise time' and 'percent overshoot' being applied. It has been observed that autonomous vehicles under closed loop control exhibit the properties normally associated with time but in a spatial sense. The perturbed vehicle that is displaced from the desired path can approach the path, overshoot the desired path, and eventually settle on the path in a manner that is reminiscent of a traditional control system approaching a final value in time. The oscillation of the vehicle about the final value appears in space as well as in time.

The distance along the desired path is related to the time axis while the orthogonal error from the path is related to the output magnitude. The relationship between the distance along the path and the time axis is direct and this distance can be thought of as a spatial time axis. The spatial axis expands and contracts with increasing and decreasing vehicle speed, respectively, and with vehicle angle to the

path. The varying spatial axis causes the response of exactly equivalent dynamic systems to be different spatially for different speeds. To illustrate the effect on velocity on the vehicle, a stabilizing controller is added to a simulated system and run at two different speeds. The dynamic systems are made identical by adjusting the system gain with velocity so that the gain of the transfer function of (3.7) is constant. The areas covered by the vehicle at two different speeds are shown in Figure 3.14 are for identical dynamic systems, the same initial conditions, and different speeds.
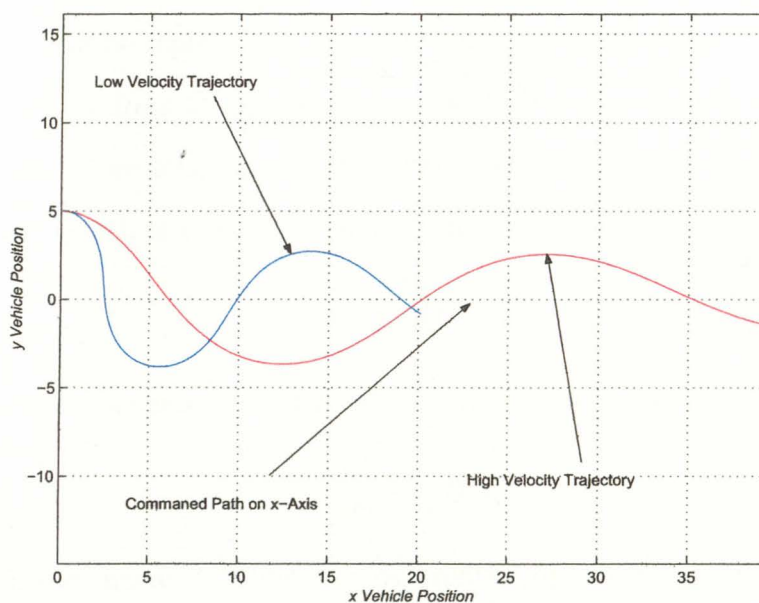
Fig. 3.14: Time identical systems with different spatial responses.

Differences in spatial response at different velocities occur because the system response is not improving with higher velocities. Identical deviations from the path and initial conditions create the same time response but the vehicle takes more space in order to settle on to the desired path. In a classical control sense, the closed loop bandwidth of both the low velocity and high velocity systems is identical while a quicker response at high velocities requires a higher bandwidth. Increased system bandwidth is required for increased velocity in order to ensure the same spatial

response. This implies that a new controller would have to be designed for each vehicle velocity, or alternately a broad bandwidth controller could be designed for the entire range of vehicle velocities. Both alternatives have unattractive components and neither method directly addresses the issue of spatial control.

It is possible to remove the dependence on vehicle velocity from the transfer function, $G_v(s)$, by defining a spatial Laplace differential operator. The development of this operator is based on a vector analysis of the vehicle trajectory. The output of the system shown in Figure 3.2 represents the trajectory of the vehicle in space and this trajectory can be represented as a time parameterized curve C. Two important quantities available from the parameterization of C are arc length and speed.

Arc length is the total length of a curve between two points, a and b, on the curve. The arc length of C is given by $l(C)$ and is defined:

$$l(C) = \int_a^b \|P'(t)\| dt. \tag{3.8}$$

Speed is the time rate of change of arc length and is a scalar value is given by:

$$\frac{dl}{dt} = v(t). \tag{3.9}$$

In the case of linear systems, the differential operator, $\frac{d}{dt}$, is replaced by s as the system is transformed into the Laplace domain. It is possible to manipulate this operator into one where the time dependence is replaced by a spatial dependence.

Using the definition of speed given in (3.9) and the chain rule of differentiation, the differential operator, $\frac{d}{dt}$, is modified to be:

$$\frac{dy}{dt} = \frac{dy}{dl}\frac{dl}{dt} = v\frac{dy}{dl}. \tag{3.10}$$

It is now possible to define a spatial Laplace differential operator in terms of the time Laplace differential operator given by:

$$s_v = \frac{s}{v}. \tag{3.11}$$

The new operator s_v can now be substituted into (3.7) with the resulting equations being in a spatial Laplace domain. $G_v(s)$ in the time Laplace domain translates to $G_v(s_v)$ in the spatial Laplace domain to be:

$$G_v(s_v) = \frac{av^2(r_m s_v + 1)}{v^2 L s_v^2 (v s_v + a)} = \frac{a(r_m s_v + 1)}{L s_v^2 (v s_v + a)}. \tag{3.12}$$

Note that the dependence of G_v on speed squared is eliminated in (3.12), but that the dependence on speed returns in the pole due to the steering actuator minor loop.

The effect of the speed dependence in the steering command to steering angle transfer function is to move the pole of this transfer function in closer to the origin as speed increases. The pole of this transfer function therefore has a more noticeable affect on vehicle system response as speed increases. If speed continues to increase, the spatial response of the vehicle will eventually be dominated by the steering system pole.

In all cases it is desirable that the response time of the steering system be faster than the vehicle response time. The speed effect on the steering actuation system is minimal when the steering actuation system is properly designed for the maximum speed of the vehicle.

While the change of variables made in this section provides a measure of spatial independence, the controller must vary with velocity in order to achieve the spatial independence described in this section. The effect of varying the system controller on vehicle stability must be considered. The analysis the velocity varying controller and its effects on stability is included in Chapter 4.

3.4 Summary

The vehicle model that will be used for the design of a controller has been derived in this chapter. The model chosen was based on the vehicle kinematic equations of motion with additions made to account for the effects of external forces, or

disturbances. The vehicle model was extended to include the model of the steering actuator and steering angle feedback structure, with a specific actuator model being presented for a hydraulic actuator. Finally, the location of the position measurement point on the vehicle was modelled. The resulting nonlinear model was then converted into a single input/single output system for all types of commanded path segments via coordinate transformations. The nonlinear system was then linearized as each of the nonlinear elements is evaluated and replaced by a linear approximation.

The deviation from the desired path was defined as the orthogonal distance from the path for the cases of straight lines, circular segments, and curves with smoothly varying curvature. This definition of deviation reduces the single-input/multiple output vehicle system to a single-input/single output system. Also, the coordinate system of the vehicle becomes centered on and moving with the path. The path-centered coordinate system is shown to be valid for the kinematic model of the vehicle for straight lines and curves.

This chapter concludes with an examination of the spatial response of the vehicle in contrast to the time response of the vehicle. It is shown that by a simple change of variable that the system can be modified to ensure a spatially similar response even though the vehicle is travelling at different velocities.

The next chapter describes the design of a controller for the vehicle. The linearized model derived in this chapter as well as the change of variables made to ensure a consistent spatial response, will be utilized in the next chapter, for controller design. The next chapter also includes a crucial analysis of stability and the development of an analytical method of stability and stability margin prediction.

Chapter 4
Vehicle Stability and Trajectory Control

In the previous chapter, a mathematical model of the vehicle system was derived along with a method of measuring tracking error. The model and the tracking error are used in this chapter in the design of a feedback system and a controller for the vehicle. Continuing with the description of the individual blocks in Figure 1.2, the vehicle model and tracking error blocks of Figure 1.2 were specified in Chapter 3. The only unspecified blocks remaining in Figure 1.2 are the controller block, the measurement block, and the command pre-processing block. The functions of these blocks are explained in this chapter. This chapter presents, in order, a description of the addition of feedback to the system, the derivation of a stability criterion, and the design of the vehicle controller.

The next section of this chapter describes the addition of feedback to the vehicle system. The vehicle position is measured and used to calculate the deviation from the desired path. An understanding of the dynamic characteristics of the measurement device is necessary to the design of the controller. The modelling requirements for the measurement system are described. The addition of feedback to the vehicle system allows the vehicle to accurately drive along commanded path segments and correct for deviations from the desired path, but the addition of feedback also creates the potential for instability. An analysis of system stability follows the description of feedback. The approach taken to controller design is integrally linked to the stability analysis and the controller design procedure therefore follows this analysis.

4.1 Addition of Feedback to the System

This section deals with the addition of the feedback loop shown in Figure 1.2. Feedback is added to the system in order to overcome or mitigate the effects of disturbances, system uncertainties, and unmodelled dynamics, including the effects of system nonlinearities. A vehicle position measurement device is added to the system and the vehicle position is processed to provide steering corrections to return the vehicle to the desired operating path. The dynamic effects of the vehicle position measurement device are examined next.

4.1.1 Vehicle Position Measurement

Some device must be used to measure the position of the vehicle in the (x, y) operating plane. Choices of measurement devices include, but are not limited to, inertial reference units, Global Positioning Systems (GPS), and computer vision systems. Whatever the selected system, understanding the characteristics of the measurement device is critical to the successful design of the vehicle controller. In an ideal world the measurement device would have no dynamic effect on the vehicle system and would respond instantly to changes in vehicle position. However, practical measuring devices can contribute significantly to the system dynamics. An effort must be made to understand the dynamic effect of the measurement device and then reduce these effects when possible. In many cases, the designer is constrained by the measuring device and must deal directly with its dynamics in the design of the feedback system. Four critical properties must be identified for the measurement device.

(1) The position update rate of the measurement sensor, if the sensor provides discrete data.

(2) The approximate, linear, input/output dynamics of the vehicle measurement device, i.e. its small input signal frequency response.

(3) Any pure time delays that the measurement system displays from vehicle position change to vehicle position indication.

(4) Any sources of noise and/or disturbances in the system and their characteristics.

Items (1), (2), and (3) in the above list define how the measurement system responds to change of vehicle position. If the vehicle were moved from one point to another instantaneously, the measurement system sees a step change in position. The measurement system has a response to this change and this response can be used to identify the dynamics of the measurement system. Instantaneous changes in position are not achievable and alternately the dynamics can be estimated from a sensor frequency response. The frequency response can be estimated at a number of discrete frequencies by placing the position measurement system on a large pendulum or rotating table that has an adjustable period and known instantaneous position.

The linear dynamics (2) of the measurement system include the dynamics of the measurement device plus any filtering that the device performs on the data it receives. Both inertial and GPS systems filter input data in an attempt to enhance accuracy by reducing noise and the filters can have a significant dynamic effect on the closed loop behavior of the system.

Pure time delays (3) tend to destabilize a system and with the prevalent use of digital processing for filtering and calculation, the evaluation of time required for data acquisition, conversion, computation, and communication is a must. Even a series of seemingly small delays can accumulate to degrade performance and lessen the stability margins of the closed loop system.

Item (4) in the above list includes any sources of inaccuracy that are inherent to the measurement system. The types of inaccuracy that a vehicle position measurement system exhibits include drifts, offsets, and zero-mean random changes. Drifts

and offsets are modelled as combinations of ramp and step disturbances in vehicle position. Zero-mean random noise is modelled as an additive noise source to the position measurement. The three noise sources can occur in any combination.

Offset and drift in the vehicle position measurement are impossible to correct for using feedback alone. The feedback system depends on the information generated by the measurement device in order to calculate the proper corrective action. If the measurement device is providing inaccurate information, the corrective action derived from the measurement will misguide the vehicle. In general, if the measurement is offset from the true position the vehicle will follow a path that is offset from the actual path by the magnitude of the offset. Drift can be thought of as a time rate of change of offset. Corrective action for offset and drift can only be accomplished through the use of multiple measurement devices.

Random measurement noise, if it truly has a zero-mean, is more tractable than offset and drift inaccuracies. A filter can be added to the system to lessen the effect of the noise or, as is often the case, the random noise level is small enough that it can be ignored, particularly given the fact that the vehicle itself functions as a low pass filter for rapidly and randomly changing position measurements of small magnitude.

A general block diagram of the vehicle position measurement device is shown in Figure 4.1, which includes an unspecified filter, a time, or transport, delay element, and a summing junction for the addition of measurement noise.

4.1.2 Calculation of the Vehicle Orthogonal Error

The vehicle deviation from the desired path was defined in Chapter 3 as the orthogonal distance from the vehicle to the desired path. In order to be useful in a control system, the error term must provide an input into the controller that will create the proper corrective action on the system being controlled. The orthogonal
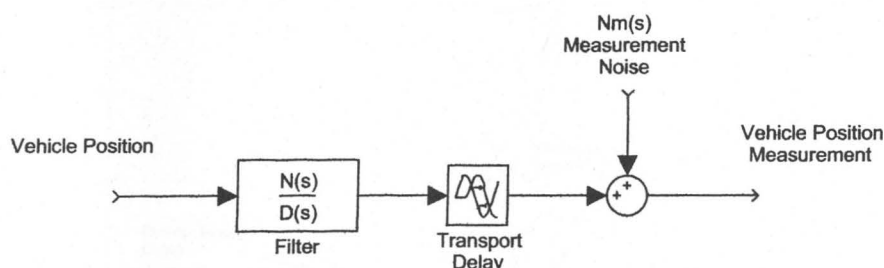
Fig. 4.1: Vehicle position measurement device.

distance by itself does not contain enough information to correct for an error condition. This lack is eliminated by adding the correct sign to the orthogonal distance.

The sign of the orthogonal error is based on the direction of travel on the segment, as shown in Figure 4.2, and upon the side of the segment that the vehicle is travelling. This formula:

$$\text{error} = \gamma \left(\sqrt{(x_v - x_c)^2 + (y_v - y_c)^2} - r_c \right), \tag{4.1}$$

calculates the orthogonal error for circular segments, given a desired direction of travel along the segment, γ, the segment radius, r_c, the segment center, (x_c, y_c), and the vehicle position, (x_v, y_v). For circular segments, γ is defined as:

$$\gamma = \begin{cases} 1 & : & \text{direction CW, inside radius} \\ -1 & : & \text{direction CW, outside radius} \\ 1 & : & \text{direction CCW, outside radius} \\ -1 & : & \text{direction CCW, inside radius} \end{cases} . \tag{4.2}$$

The formulae of (4.1) and (4.2) can also apply to straight line segments given that a straight line segment can be thought of a segment of a circle with a center at infinity. Define the segment as starting at (x_0, y_0) and ending at (x_1, y_1), with the vehicle located off of the segment at (x_v, y_v). The center of the straight segment can lie on any line that is perpendicular to the segment and it is convenient to choose a line perpendicular to the segment at the end point of the segment. In Figure 4.3,
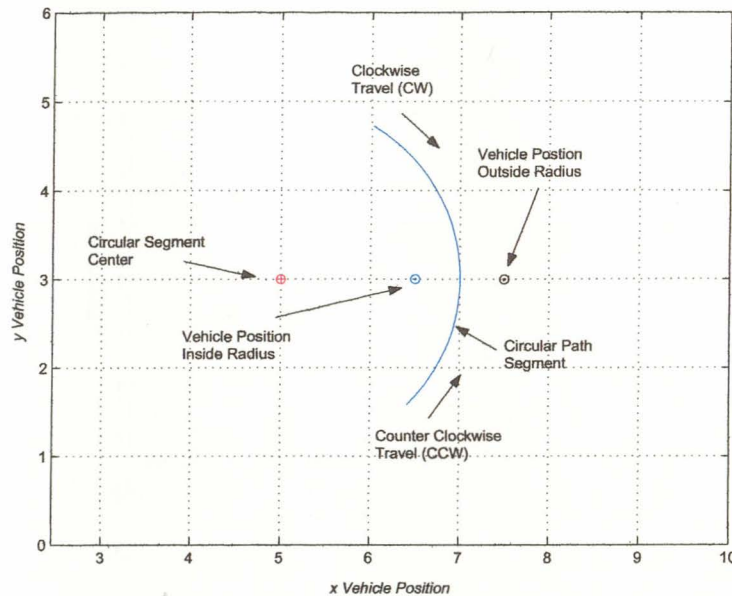
Fig. 4.2: Orthogonal error for a circular segment.

the segment is shown with a line perpendicular to the endpoint with a finite center, (x_c, y_c). Call the angle formed by the sides labelled r and R_n, ψ. The angle formed by the sides labelled N and n is equal to ψ. The tangent of ψ is given by:

$$\tan(\psi) = \frac{d}{r}. \tag{4.3}$$

As r is allowed to approach infinity, the tangent of θ approaches zero given that d is finite. The orthogonal error, or the length of the line labelled n, is given by:

$$n = N \cos(\psi). \tag{4.4}$$

As ψ approaches zero N approaches n, the difference between N and n can be made arbitrarily small by the choice of the angle ψ or, equivalently, the length of r. Therefore, the orthogonal error is approximately given by (4.1), with (x_c, y_c) moved sufficiently far from (x_1, y_1).
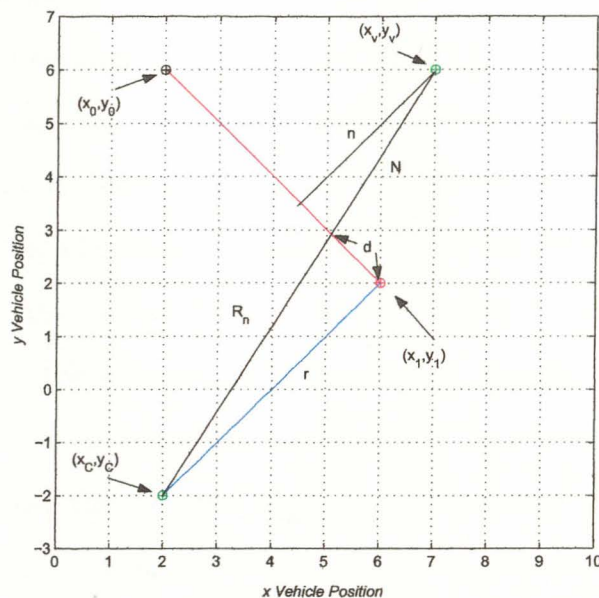
Fig. 4.3: Orthogonal error for a straight line.

The orthogonal error from an allowable, arbitrary curve, C, can also be derived from (4.1) and (4.2), but the calculations require the calculation of the instantaneous curvature of C, which is needed to find the instantaneous center, (x_c, y_c).

4.2 Stability Analysis of the Nonlinear Vehicle System with a Controller

The stability of the vehicle in path tracking is of paramount concern to the designer. An unstable vehicle could exhibit unexpected motions and deviations from the planned vehicle path and being in close proximity to an unstable vehicle is dangerous. Additionally, an unstable vehicle can perform no useful work. Previous research into the problem of trajectory following stability is presented. Also, before proceeding with a mathematical treatment of the stability of the system shown in Figure 3.7, a qualitative analysis of the possible unstable modes of the vehicle is presented to provide valuable insight and guides the mathematical treatment of the control system stability.

As mentioned above, with the addition of feedback to the system the control system design must account for the effects of disturbances, system uncertainties, and unmodelled dynamics, including the effects of system nonlinearities. The designer desires a system that is stable under all allowed operating conditions and the analysis that follows provides a basis for designing a control system that addresses, directly or indirectly, all of the above-mentioned effects. System simulation and testing can provide insight into the behavior of the controlled vehicle, but the chance that a critical dynamic effect or operating condition will be missed is high. Analysis of stability also provides the designer with an understanding of the stable operating range of the vehicle. An understanding of system stability and the factors that influence it is essential to the design of a safe autonomous vehicle.

It will be demonstrated in the analysis that follows that an incomplete consideration can result in a system that has both overt and latent unstable modes. The overt unstable modes are easily seen in vehicle simulation and testing, but the latent modes are less easily found. Latent unstable modes can be initiated by disturbances that are 'large enough' with the vehicle beginning to oscillate around a desired path despite apparently stable operation prior to the beginning of this behavior. The causes of the latent unstable mode are explained as part of the stability analysis.

4.2.1 Existing Approaches to Vehicle Trajectory Tracking and Path Following Stability

Many researchers have recognized the importance of vehicle tracking stability. This section will examine in some detail the problem of stability and the approaches that have been taken in seeking to understand this problem.

An extensive treatment of trajectory tracking, path following, and point stabilization has been compiled by De Luca, Oriolo, and Samson in [19]. The primary concern of De Luca et al. is that the vehicle be capable of finding and tracking a path

(function of spatial variables only) or a trajectory (function of both time and spatial variables) despite the initial position and orientation of the vehicle with respect to the path.

De Luca et al. use a kinematic model that is identical to the one presented in Chapter 3 from steering angle input to $x - y$ outputs. Mention is made of the actuation and other non-ideal effects, but these are not felt necessary to the problem at hand, that of deriving the requirements for stability for on or off path following stability.

It is appropriate at this point to discuss the philosophical differences between the approach to path following taken by De Luca et al. and the approach taken in this dissertation. The problem of controlling the vehicle when it is far off of the path is considered below. However, the path was probably created either by a path training method or by a path planning method. Both methods of producing the path encapsulate a great deal of knowledge about the environment that the path passes through. The training method uses the knowledge of the human driver that drives the training path and the path planning method usually relies on stored knowledge of the environment to avoid hazards, etc. Placing the vehicle off of the path and then letting it generate a sub-path back to the main path bypasses the specific knowledge that might prevent collisions or worse. This research therefore focuses on near path performance and stability with the control system being present to attenuate disturbance, compensate for vehicle variation, and correct for intrinsic path following inaccuracies.

The control laws developed in [19] are generally complex given the difficult nature of the problem they must solve. The conclusion is that the control laws must be time varying. De Luca et al. chose to use a Lyapunov stability analysis on the kinematic model. Choice of this method precludes detailed analysis of so-called 'hard' nonlinearities.

Concerning the nonlinearities presented in Chapter 3, De Luca et al. conclude:

> Concerning the application of the proposed feedback controllers to real mobile robot systems, there are several non-ideal conditions that may affect the actual behavior of the controlled robot, notably: uncertain kinematic parameters of the vehicle (including, e.g., the wheels radius); mechanical limitations such as backlash at the steering wheels and limited range of the steering angle; actuator saturation and dead-zone; noise and biases in the transformation from physical sensor data to the robot state; quantization errors in a digital implementation. Control robustness with respect to these kinds of uncertainties and/or disturbances is an open and challenging subject of research. (p. 78)

This research examines in some detail the effects of the nonlinearities mentioned in the above quotation. In particular, the two critical nonlinearities for system stability are identified as the steering angle rate limit and the steering angle limit or saturation. The effects of these nonlinearities are examined both qualitatively and quantitatively for the near path control problem.

Ollero and Heredia in [20] also approach the problem of autonomous vehicle stability from the kinematic model and use a Lyapunov stability analysis on a kinematic model. The contribution of this paper is that the pure time delay in the measurement device is accounted for in their analysis. The method that is developed below is easily adapted to include the effects of time delays.

Adams in [13] presents a method of tracking to a point in an environment using potential field navigation. He also analyzes stability using the Lyapunov method, but included a describing function analysis to predict and prevent limit cycles at points in the potential field far from the desired point. Specifically, the describing

function analysis looks at the effects of wheel velocity saturation on the potential field path finding algorithm.

Using a scalar error measurement to define deviation from a path has been extended to other vehicle steering schemes in a thesis by Bahl [21]. Additional types of steering systems include differential steering and omni-directional steering. The thesis also presents wheel velocity-based, controller design methodologies.

Bak, Bendtsen, and Ravn in [22] present a hybrid control method that is based on switching control schemes in order to avoid singular points in the control solution. Again, the vehicle model is a kinematic model and the stability analysis of each of the control schemes is analyzed by using the Lyapunov method. Feedback linearization is used as the control scheme and mention is made of the problems that this method has with uncertainties and non-invertible nonlinearities. No direct analysis of stability of the system with the nonlinearities is attempted.

While not dealing with autonomous vehicle control, Ackermann and Bünte in [23] detail the describing function analysis of a steering system that exhibited a limit cycle during testing. The cause of the limit cycle is identified as the steering actuator rate limit. A criterion is derived that allows the design of a system without limit cycle behavior. The analysis presented below will draw similar conclusions regarding the destabilizing effect of the steering angle rate limit.

Numerous other researchers have presented stability analyses that are based on the kinematic model and Lyapunov method.

No direct analyses of the critical steering system nonlinearities were found while researching work done in the area of stability of mobile robots and autonomous ground vehicles.

4.2.2 Vehicle System Unstable Modes

The measure of vehicle deviation from a commanded path was defined in Chapter 3 as the perpendicular distance from the vehicle measurement point to the path

and it was shown that this definition effectively reduced the single input/multiple output system to a single input/single output system. In this chapter, the orthogonal distance was given a sign and is used as the system error. The orthogonal error is input into the controller and the controller produces steering commands based the magnitude and dynamic behavior of the error. Qualitatively, the controller produces large steering angle commands for large orthogonal errors and vice versa.

There are three possible types of instability that the vehicle could exhibit. Each type of unstable behavior could also exhibit more than one mode of operation.

1. An exponential departure of vehicle from the commanded path.

2. A regular oscillatory behavior.

3. A combination of the first two, exponential growth while oscillating about some constant value.

A relatively simple analysis eliminates the possibility of any unstable mode that involves exponential departure from the desired path. If the vehicle were to depart from the path in any manner such that the vehicle deviation from the path is monotonically increasing, the controller would cause the steering angle command to increase. As the steering angle increases, the vehicle follows tighter circular paths until the maximum steering angle is reached. The vehicle then drives in the tightest possible circle that the vehicle can follow and the error is no longer monotonically increasing and in fact the orthogonal error increases and decreases periodically indicating one possible mode of oscillatory instability. By a similar argument, any exponential growth of error coupled with oscillatory behavior is not possible because the vehicle would eventually be far enough away from the desired path that the steering angle could no longer increase. The steering angle limit therefore eliminates the possibility of exponential instability.

The oscillatory mode described in the above paragraph results in the system error periodically increasing and decreasing about a nonzero value. Another possible unstable mode would have the system error increasing and decreasing about zero or 'wandering' about the desired path. No simple elimination of this unstable mode is obvious.

The remaining mode of unstable behavior to be considered is regular oscillation about zero error or a so-called 'limit cycle' in nonlinear systems stability analysis. It will be shown that the vehicle can exhibit two the distinct modes of limit cycle described above and these limit cycles and their causes are described in the following sections.

4.2.3 Large Orthogonal Error and Orbital Behavior

The first mode of limit cycle results when the vehicle is placed at some distance far off the desired path and then the path tracking controller is activated. The orthogonal error is large when the vehicle is far off the path and the controller commands a non-zero steering angle in an attempt to force the vehicle back to the desired path. The further the vehicle is from the path, the larger the steering angle becomes. The steering angle can limit, or saturate, for a large error, leaving the vehicle with the steering angle at an extreme position. When the steering angle is held to an extreme position, the vehicle begins to drive in a tight circle, or orbit as shown in Figure 4.4. The vehicle continues to travel in the tight circle unless the steering angle is reduced, and such a reduction can only be achieved by a closer approach to the path while the vehicle orbits. If the orthogonal error is sufficiently reduced in some portion of the orbit such that the controller commands a steering angle that is less than the extreme value, then the vehicle may approach and attach to the path. The conditions required for the vehicle to reattach to the path include that the overall vehicle system be sufficiently low-pass and that the system is not
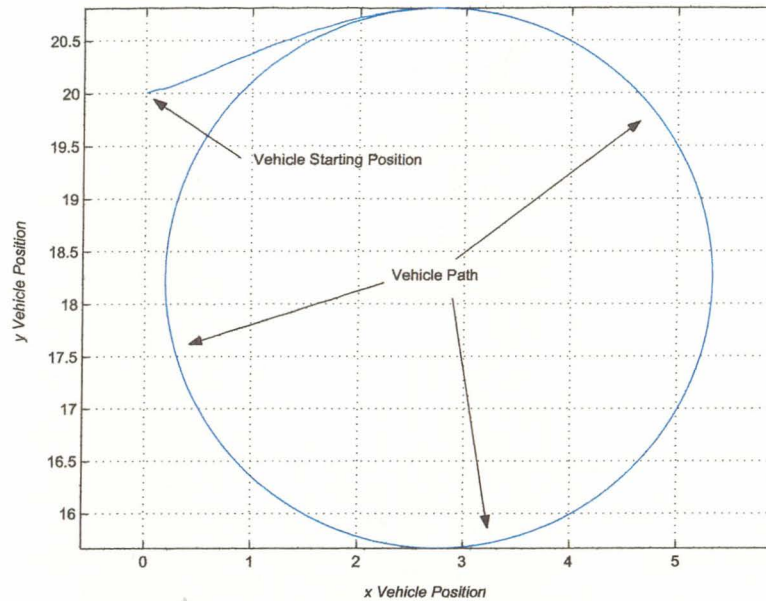
Fig. 4.4: Large error orbit.

conditionally stable. Given the undesirable nature of this motion a detailed analysis is not presented. An approximation of the maximum allowable error is derived in the following paragraph.

The steering angle limit, $M_{steering}$, and the sine/cosine output nonlinearities dominate this unstable mode with the controller playing only a minor role in setting the maximum non-orbiting distance from the path and providing unconditional stability to the overall system. When the vehicle is placed at a point that is close to but less than the maximum possible orthogonal error the path that the vehicle takes in returning to the desired path is nearly orbital and this condition should be avoided. The effect of being close to the maximum error is illustrated in Figure 4.5. Given that the poor behavior of the vehicle near the maximum orthogonal error, it not useful to exactly define the maximum orthogonal error and an approximation to this error suffices. The approximation to the maximum orthogonal error from a
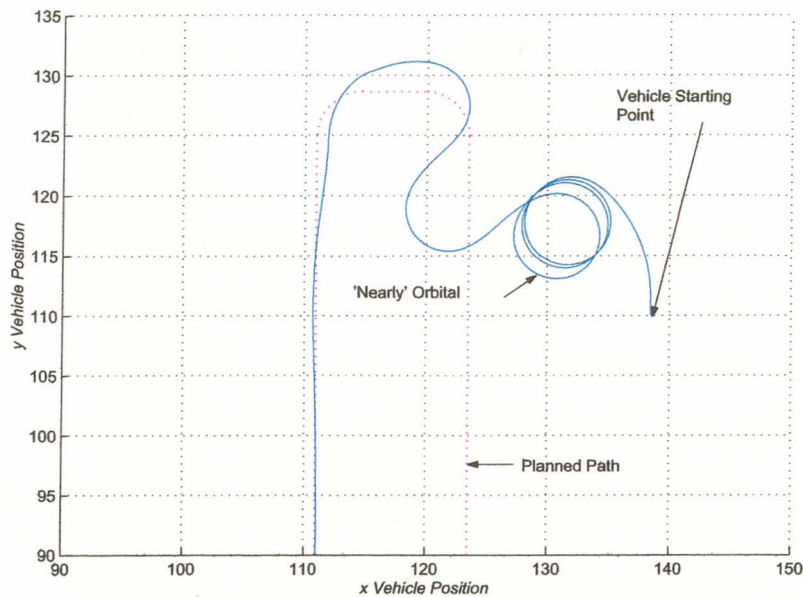
Fig. 4.5: Nearly orbital behavior.

straight line path is made using the controller steady state gain is:

$$\max(\epsilon_{line}) \cong \frac{M_{steering}}{k_{controller}}. \tag{4.5}$$

The maximum orthogonal deviation from a circular path differs from that of a straight line. The undisturbed vehicle follows the circular path by using a non-zero steering angle. An orthogonal error that is stable for a straight line can drive the steering system into an unrecoverable orbit if the error is to the concave side of the circular path segment. The approximation to this limit is given as:

$$\max(\epsilon_{circle}) \cong \frac{M_{steering}}{k_{controller} \sec^2 \theta_s}. \tag{4.6}$$

A method for avoiding the orbital limit cycle while preserving the ability to control the vehicle to an arbitrary path in the plane is presented in the next subsection. The method presented provides control over the vehicle for any magnitude of orthogonal error and is one of many possible methods for overcoming this problem.

4.2.4 Control of the Vehicle When Orthogonal Errors Are Large

Given that the vehicle exhibits an undesirable behavior when the error is greater than the maximum orthogonal error, a possible solution is to replace the orthogonal error controller with a controller that controls the vehicle heading. An appropriate heading leading back to the path can then be chosen and when the orthogonal error is reduced to some preset acceptable value control can be returned to the orthogonal error controller. The design of such a controller involves closing the control loop around vehicle heading. A simple proportional controller is all that is required as shown in Figure 4.6.
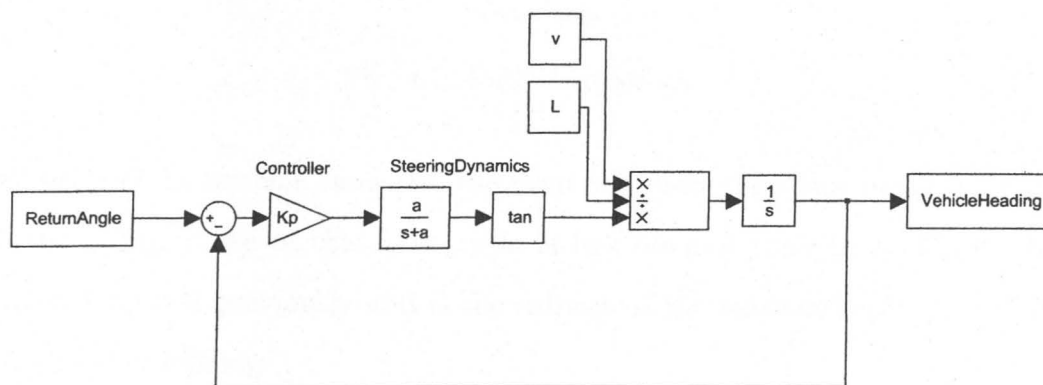
Fig. 4.6: Far path controller block diagram.

4.2.5 Oscillations about the Desired Path

The second mode of oscillatory behavior is a limit cycle that is best described as a 'wandering' behavior in that the vehicle wanders back and forth across the desired path while more or less progressing down the desired path as illustrated in Figure 4.7. This behavior is a strong function of the controller, the rate limit of the steering actuator, the steering angle controller transfer function, and the linear vehicle model. Also important, but less dominant is the steering angle saturation. The sine term has no significant effect in the establishment of this type of limit
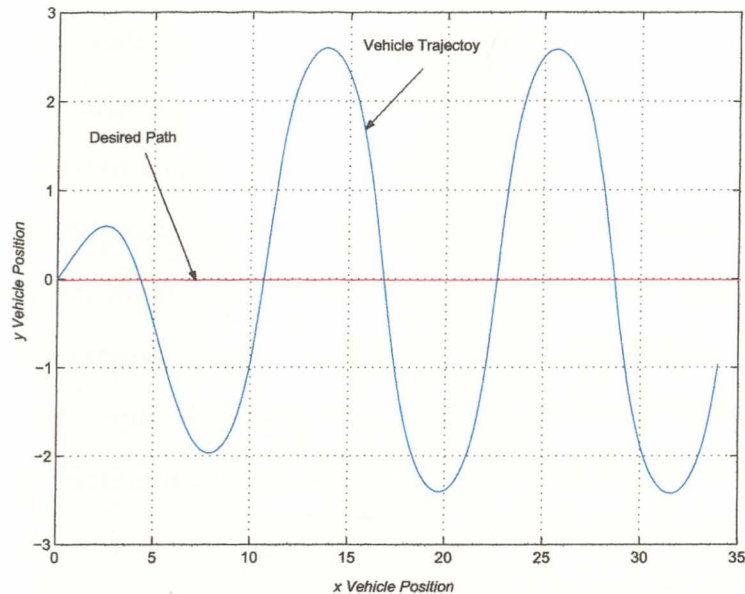
Fig. 4.7: Path oscillations.

cycle, although in extreme cases the sine term can effect the shape of the wandering trajectory. The cause of this limit cycle is less obvious than that of the orbital behavior described previously and is the subject of the mathematical treatment of instability that follows.

4.2.6 Stability Analysis of the Wandering Behavior

The system shown in Figure 3.7 contains a mixture of nonlinear and linear subsystems. The nonlinear blocks include so-called 'hard' nonlinearities, the steering angle limit and the hydraulic flow saturation. Given the current state of nonlinear systems analysis, and the presence of the hard nonlinearities [24], the method of describing functions [25] is the most tractable method available that yields useful information. The method of describing functions excels at the prediction of limit cycles and is a good fit for the prediction and prevention of the wandering behavior. The describing function approach to stability analysis has been validated both mathematically [26], [27] and by control systems practitioners.

Describing function analysis is a linear analysis of a nonlinear system using an assumption of sinusoidal oscillation. Successful use of this analysis method requires that the overall nonlinear system be separable into linear and nonlinear subsystems, that the nonlinearities are symmetric about zero (a requirement that can be relaxed under certain conditions), and that the overall nonlinear system is significantly low-pass i.e. the system response to a sinusoidal input approaches zero as the frequency of the input is increased beyond some finite limit. The vehicle modelled in this research meets the required assumptions, as is demonstrated below.

Describing function analysis begins with the assumption that the input to the nonlinear subsystem has an oscillatory signal of the form $A\sin(\omega t)$. Analysis is then done to determine the conditions under which the assumption of oscillation is true.

The describing function of the nonlinear subsystem is found by applying a sinusoid of known amplitude and frequency to the input of the subsystem, $A\sin(\omega t)$. If the assumptions of the previous paragraph are true, the output of a nonlinear system so driven will be a periodic waveform with a fundamental frequency of ω. The fundamental component of the output waveform is written as, $B\sin(\omega t + \sigma)$, where σ is the phase angle between the input and output waveforms. The input and output waveforms can also be written in phasor form as:

$$A\angle 0^o$$
$$B\angle \sigma^o. \tag{4.7}$$

The ratio of the output to input waveforms is the approximate frequency-domain transfer function given as:

$$N(A,\omega) = \frac{B\angle \sigma^o}{A\angle 0^o} = \frac{B}{A}\angle \sigma^o. \tag{4.8}$$

The formula shown in (4.8) shows the describing function, N, as an explicit function of both the amplitude of the input, A, and the frequency, ω. The explicit dependence on ω implies that the phase and/or amplitude of the output waveform

changes with frequency. This is typical of nonlinear functions that limit the rate of change of the input waveform. N can be real or complex, depending on the phase difference between the input and output. In general, the ratio of B to A is not constant and it is common for B/A to tend toward zero as A tend towards infinity.

Describing function analysis provides an extension to the familiar Nyquist criterion for determining the stable or unstable behavior of the system. The Nyquist criterion is explained in detail in [28] and expressed as:

$$1 + G(s) = 0. \tag{4.9}$$

In (4.9), $G(s)$ represents entire loop transfer function and for a separable system $G(s)$ can be separated into the linear subsystem and the describing function of the nonlinear subsystem as shown:

$$G(s) = G_l(s) N(A, \omega). \tag{4.10}$$

Substituting (4.10) into (4.9) and rearranging yields the revised Nyquist criterion shown in (4.11):

$$G_l(s) = \frac{-1}{N(A, \omega)}. \tag{4.11}$$

While the Nyquist criterion is concerned only about the net encirclements of the minus one point on a complex polar plot, the describing function of the system nonlinear portion extends the minus one point into a locus of points, one point for each amplitude A, or in the case of a frequency dependent nonlinearity, a locus of points for every A for each ω such that the locus of points forms a family of curves. Oscillatory behavior is predicted by examining intersections of the locus or surface representing the nonlinearity and a frequency-domain polar plot of the linear portion of the system. Oscillatory behavior is prevented by designing so as to avoid intersections of the describing function locus or surface.

Describing function analysis predicts oscillation in systems that always begin to oscillate when started and those that require an outside stimulus before the onset of oscillation. The latent tendency towards oscillation of the second type of system is in and of itself justification for analyzing system stability rather than depending on simulation and test to predict unstable behavior. Misdesign of a controller for such a system can produce a seemingly stable system with oscillation occurring only after an unexpected external disturbance. Proper stability analysis and the resulting controller design can obviate the sudden and dangerous onset of vehicle oscillations.

The first requirement for the application of the method of describing functions is that the nonlinear system must be separated into linear and nonlinear subsystems. The nonlinear system illustrated in Figure 3.7 does not appear to be easily separated. The steering angle limit is outside of the steering actuator minor loop while the hydraulic flow saturation is inside of the minor loop. The tangent function is separated from both of the saturations and the output sine and cosine terms are completely isolated from the other nonlinearities. The next section analyzes each of the system nonlinearities in turn and presents a series of simplifying assumptions that make the application of the describing function method possible.

4.2.6.1 System Nonlinearities

The goal in analyzing the nonlinear system of Figure 3.7 for stability is to predict and eliminate the existence of the oscillatory behavior dubbed 'wandering.' The nonlinear model as presented in Figure 3.7 has four nonlinearities that must be considered in addition to the linear dynamics. Following the flow in Figure 3.7 from steering command to y-position output, the first nonlinearity encountered is the steering angle limit. The steering angle limit is modelled as a saturation with the limits being the maximum left and right steering angles. The steering saturation

model is accurate for a vehicle with Ackerman steering in that the steering angle of such vehicles is abruptly limited mechanically.

The next nonlinearity is the hydraulic flow saturation that is inside of the steering angle control system. The saturation is shown in the position that it physically occupies in the hydraulic actuator, but its effect is to limit the maximum rate of travel of the actuator. If the rate limit of the actuator is significantly slower than the maximum rate of change of the small amplitude linear model of the actuator controller, the normal case, then the saturation can be pulled out of the steering angle control loop, converted into a equivalent actuator rate limit, and placed to the right of the steering angle limit as shown in Figure 4.8. The block is placed to the left of the steering angle minor loop in order prevent the 'wind-up' problem that would occur if the actuator were on the right side of the steering actuator minor loop.
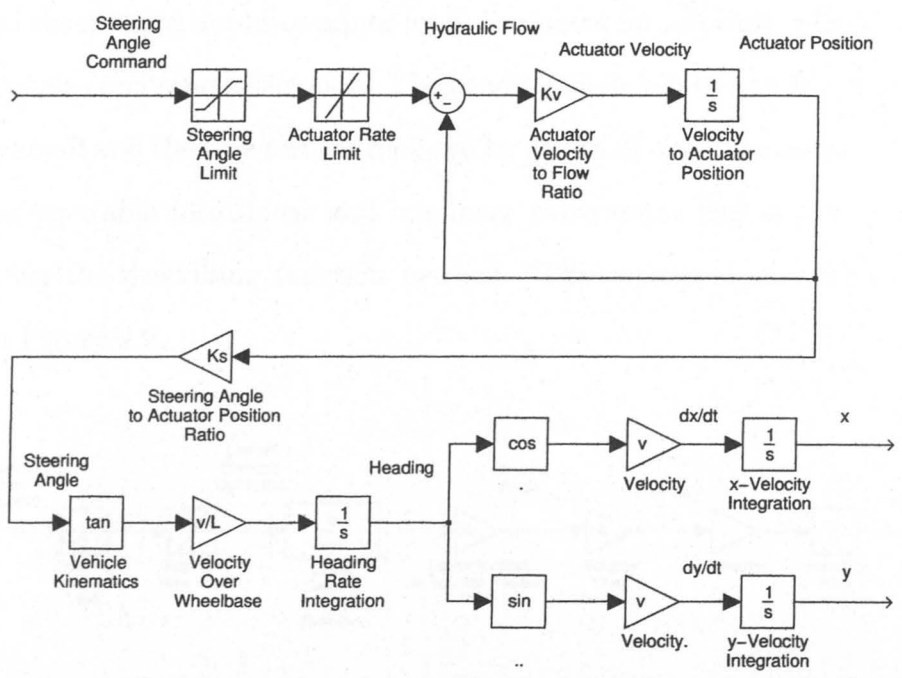
Fig. 4.8: Nonlinear system with rate limit.

The third nonlinearity is the tangent function following the steering angle control loop. The tangent function has a limited input, the steering angle, and therefore will be replaced with an equivalent linear gain. The gain can be conservatively chosen to degrade stability for this analysis.

The last nonlinear block in Figure 3.7 is the sine term preceding the last integrator. If the heading angle from the desired heading is small, then the sine term can be accurately replaced by a gain of one. Given that the goal of the controller is to keep the vehicle travelling down the desired path, the heading angle for a stable controller will always be small for a stable, well-behaved system. Simulation of the system indicates that the describing function method produces valid results even when the heading angle versus the desired path heading is relatively large, in excess of sixty degrees.

The resulting system has the tangent function conservatively replaced with a gain. The hydraulic flow saturation is replaced with an actuator rate limit, moved outside of the steering actuator minor loop, and, most importantly, placed next to the steering angle command saturation. The heading angle relative to the desired path is assumed small and the sine term is replaced by a gain of one. The resulting nonlinear system is separable into linear and nonlinear subsystems and is now amenable to analysis via the describing function method. The separated, nonlinear system is shown in Figure 4.9.
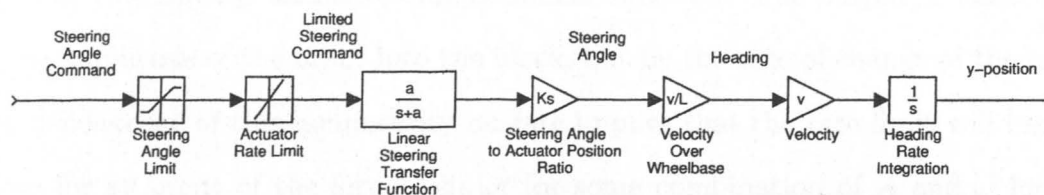
Fig. 4.9: Vehicle system separated into linear and nonlinear subsystems.

4.2.7 The Describing Function of the Saturation-Rate Limit Combination

A describing function of the nonlinear subsystem of the system shown in Figure 4.9 can be thought of as an approximate, frequency-domain transfer function. The approximation consists of magnitude and phase information obtained when the subsystem is stimulated by a sinusoidal input of the form, $A \sin \omega t$. In general, the magnitude and phase information is a function of both the input amplitude, A, and the frequency, ω. The describing function is obtained by examining the Fourier coefficients of the output of the nonlinear subsystem for the fundamental frequency of the input with higher harmonics ignored. The general form of the describing function is given in (4.8).

The nonlinear subsystem in Figure 4.9 consists of a saturation cascaded on the right by a rate limit. Both of the blocks appear as gains of one if the input to the blocks does not exceed the limits of the individual blocks in the nonlinear subsystem i.e. input amplitudes lower than the saturation limit or input rates lower than the rate limit. The describing function of this nonlinear subsystem must take into account the actions of both blocks, with neither block active or the rate limit, the saturation, or both blocks being active for a given input amplitude and frequency. The three active cases are described in the following sections.

4.2.7.1 Rate Limit Active and Saturation Inactive

The rate limit is an interesting nonlinear function. The output is limited not by the amplitude of the input into the block, but by the rate of change of the input. The dependence of this nonlinearity on rate implies that the rate limit will become active for an input of the form $A \sin \omega t$ for some combination of A and ω for any ω. The describing function of the rate limit is therefore dependent not only on the amplitude of the input, but the frequency of the input as well. The dependence

on rate limit implies that the locus of points describing $-1/N_e(A,\omega)$ is a family of curves, one for each ω. The response of a rate limit to several sinusoidal inputs is shown in Figure 4.10. Note that the output of the rate limit does not immediately begin tracing the input again once the rate of the input has dropped below the rate threshold.
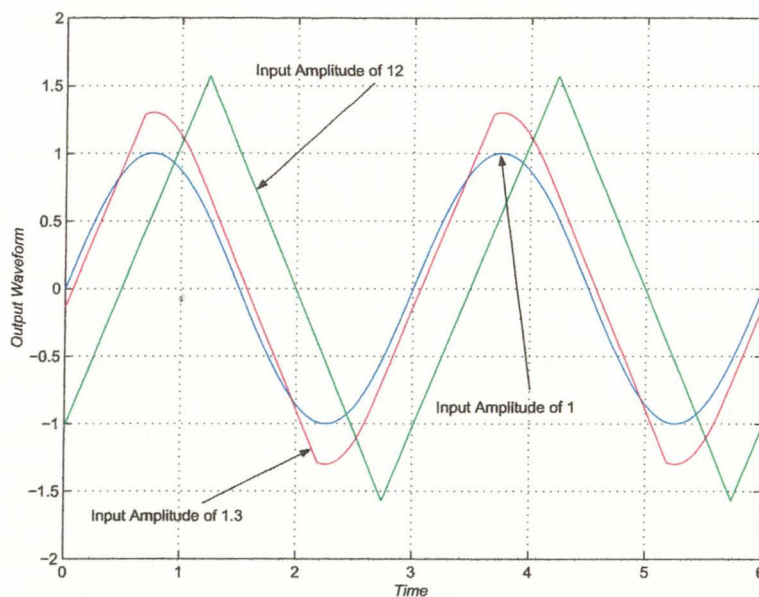
Fig. 4.10: Rate limit response to sinusoidal inputs.

It is clear from Figure 4.10 that the rate limit effects not only the magnitude of the input waveform, but also the phase. Additionally, the phase is dependent on both the amplitude and frequency of the input waveform. The result is that the describing function of the rate limit is best described as a family of loci for each frequency, or as a surface when plotted on a Nyquist plot with an additional vertical axis of radian frequency.

4.2.7.2 Saturation Active and Rate Limit Inactive

The describing function of a saturation function is described in almost all basic controls systems textbooks [24], [25]. The locus of $-1/N(A)$ runs along the negative

real axis on the Nyquist plot from minus one to minus infinity. The formula for $1/N(A)$ for a saturation nonlinearity is given as:

$$\frac{-1}{N(A)} = \frac{-1}{(\frac{1}{\pi}(2\sin^{-1}(M/A)-\sin(2\sin^{-1}(M/A)))+\frac{4M}{\pi A}\cos(\sin^{-1}(M/A)))}, \quad (4.12)$$

where M is the saturation amplitude.

4.2.7.3   Both Saturation and Rate Limit Active

For the case of the saturation cascaded with the rate limit, the locus of $-1/N(A)$ follows the negative real axis until the rate limit becomes active for a given A and ω. The saturation element contributes no phase shift from input to output and the resulting Fourier analysis of the output has real part only.

4.2.8   Results of Numerical Describing Function Analysis

As stated previously, the existence of the rate limit block in Figure 4.9 implies a dependence on the frequency of the input. The describing function of the nonlinear subsystem shown in Figure 4.9 is therefore not a simple locus of points, but consists of a separate curve for each frequency that is applied to the input of the nonlinear subsystem. When viewed on the polar Nyquist plot, the describing function loci appear as a solid region that intersects the Nyquist plot of the linear subsystem at infinitely many points. The only intersections that are important are those intersections that occur when the frequencies of a describing function locus and the Nyquist plot of the linear subsystem are equal as illustrated in Figure 4.11.

While it is possible to obtain information about the system through a careful investigation of the intersections described in the previous paragraph, a simpler and clearer picture is obtained by adding a third axis to the standard Nyquist plot. If both the loci of the describing function of the nonlinear subsystem and the Nyquist plot of the linear subsystem are plotted in three dimensions with the third dimension being ω, the resulting plot shows more clearly the critical intersections as well as
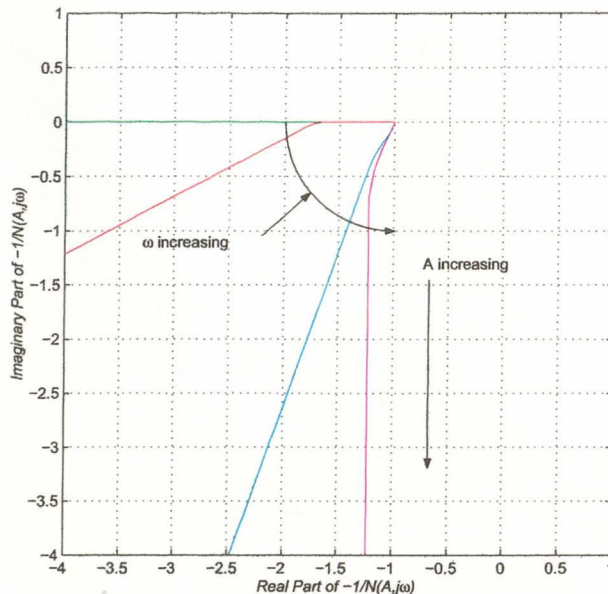
Fig. 4.11: Two-dimensional Nyquist plot.

providing a much simpler interpretation task. A representative three-dimensional plot is shown in Figure 4.12. The steering angle rate, K_{max}, is 30 degrees per second and the steering angle saturation, M, is 38 degrees.

4.2.9   The Extended Nyquist Plot

The approximate frequency response behavior of the nonlinearities described in the previous sections can be plotted as a surface on a Nyquist plot to which a third axis of radian frequency has been added. The surface consists of level curves at each frequency with the curves consisting of the locus of $-1/N_e(A)$ at the level frequency. Intersections of this surface by the locus of the linear portion of the vehicle transfer function can indicate the existence of a limit cycle. It is observed that the plot of Figure 4.12 consists mainly of nearly straight line level curves emanating from the negative real axis. Given that the numerical plot is computationally intensive and that little insight is gained from the numerical calculation, an asymptotic approximation to the stability surface is derived in the next section.
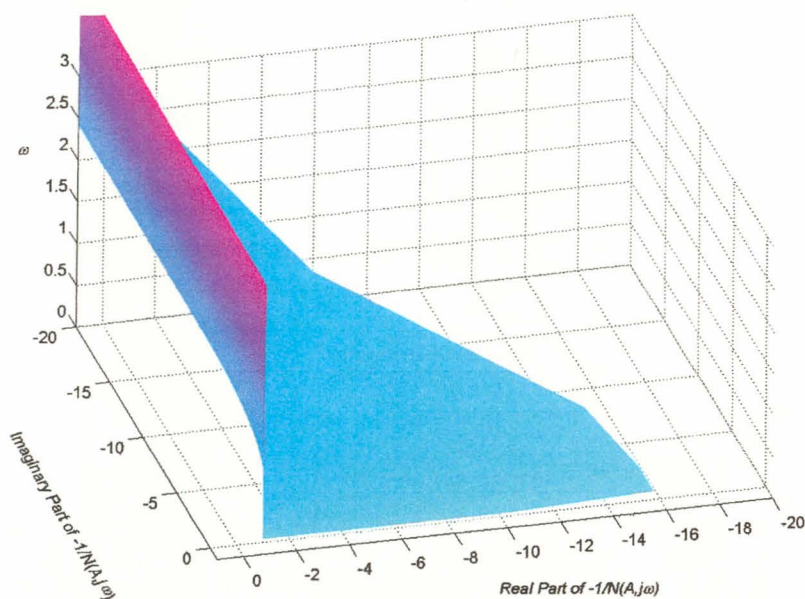
Fig. 4.12: Three-dimensional Nyquist plot.

4.2.10   The Asymptotic Stability Surface

The surface described in the previous section is mathematically complex with no closed form representation. It is of limited usefulness aside from the graphical analysis demonstrated in the previous section. However, the surface can be represented by an asymptotic surface that is a good approximation of the more exact stability surface except near the negative real axis. The asymptotic surface replaces the calculated level curves with straight lines that intersect the real axis at the point that the steering rate limit becomes active. Calculation of the slopes of the asymptotic lines and a conservative estimate of the intersection point on the negative real axis in the real/imaginary plane follows.

Firstly, the angle of the asymptotic line is calculated. When the saturation is active before the rate limit the locus of the nonlinear combination at a single frequency ω_0 travels for some distance along the negative real axis. At some amplitude A the rate limit will also become active, thus adding phase lag to the output of the

saturation/rate limit combination. At this amplitude the locus departs from the real axis and moves into the third quadrant in the real/imaginary plane. As the amplitude is increased further the phase contribution will increase and eventually the output of the nonlinear combination will no longer depend on the amplitude A of the input sinusoid, as shown in Figure 4.13.
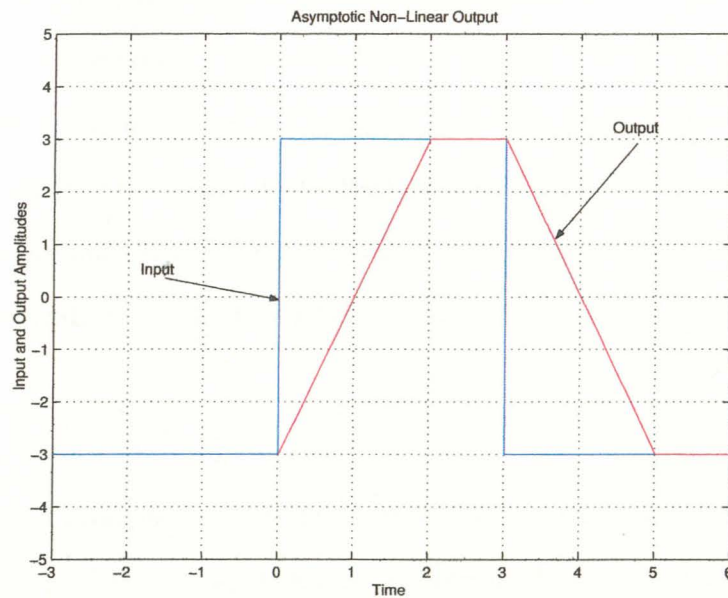
Fig. 4.13: Limiting output of the saturation/rate limit nonlinearity.

Mathematically, the angle of the asymptote can be calculated by letting:

(1) $A \gg M$ where A is the amplitude of the sinusoidal input to the nonlinear combination and M is the saturation level.

(2) $A \gg A_{min}$ where A_{min} is the minimum amplitude at which the rate limit becomes active at this frequency.

The saturation level M is given as well as the rate limit K_{max}. Note that $A_{min}\omega_0$ is the maximum slope of a sinusoidal input of amplitude A_{min}. This maximum slope

occurs at the zero crossings of the input waveform and for any given ω_0 the rate limit will become active when the amplitude A_{min} is defined as:

$$A_{min} = \frac{K_{max}}{\omega_0}. \tag{4.13}$$

As the amplitude of the input approaches infinity the output of the saturation block approaches a square wave of amplitude $\pm M$ and at the limit the input to the rate limiter is a square wave. The output of the rate limit block is now a flattened and shifted triangle wave as shown in Figure 4.13. A Fourier analysis is now performed on the output of the rate limiter to derive the describing function of the combination of the saturation and rate limiters. Recall the definition of the describing function in (4.8). The equation can be rewritten in a complex form as:

$$N(A) = g(A) + jb(A). \tag{4.14}$$

The real and imaginary components of (4.14) are the scaled Fourier coefficients of the output waveform. The analysis is simplified if ωt is replaced by θ. The resulting equations for the real and complex components of (4.14) become:

$$g(A) = \frac{2}{\pi A} \left[\int_0^\pi f(\theta) \sin(\theta) d\theta \right] \tag{4.15}$$

$$b(A) = \frac{2}{\pi A} \left[\int_0^\pi f(\theta) \cos(\theta) d\theta \right]. \tag{4.16}$$

The function $f(\theta)$ is the piecewise continuous output of the combination of saturation and rate limiter. For the limiting case of an infinite amplitude input sine wave, $f(\theta)$ is defined as:

$$f(\theta) = \begin{cases} [A_{min}\theta - M] & : \quad 0 \leq \theta < \theta_1 \\ M & : \quad \theta_1 < \theta \leq \pi. \end{cases} \tag{4.17}$$

At present, the angle θ_1 is unknown, but is related to the frequency of the input waveform and the rate limit of the nonlinearity. Solving for $g(A, \theta)$ and $b(A, \theta)$ yields a more general solution and the definition of θ_1 as follows:

$$g(A) = \frac{2}{\pi A} \left[\int_0^{\theta_1} (A_{min}\theta - M) \sin \theta d\theta + \int_0^{\pi} M \sin \theta d\theta \right] \tag{4.18}$$

$$b(A) = \frac{2}{\pi A} \left[\int_0^{\theta_1} (A_{min}\theta - M) \cos \theta d\theta + \int_0^{\pi} M \cos \theta d\theta \right]. \tag{4.19}$$

$$\tag{4.20}$$

The above equations simplify in the form as shown:

$$g(A) = \frac{2}{\pi A} \left[A_{min} (\sin \theta_1 - \theta_1 \cos \theta_1) + 2M \cos \theta_1 \right] \tag{4.21}$$

$$b(A) = \frac{2}{\pi A} \left[A_{min} \cos \theta_1 - A_{min} \right]. \tag{4.22}$$

The slope of the output waveform between 0 and θ_1 is A_{min} and is also equal to $2M/\theta_1$ and θ_1 is given as:

$$\theta_1 = \frac{2M}{A_{min}}. \tag{4.23}$$

Substituting (4.23) into (4.21) and (4.22) and simplifying yields expressions for $g(A)$ and $b(A)$ as shown below:

$$g(A) = \frac{2}{\pi A} \left[A_{min} \sin \frac{2M}{A_{min}} \right] \tag{4.24}$$

$$b(A) = \frac{2}{\pi A} \left[A_{min} \cos \frac{2M}{A_{min}} - A_{min} \right]. \tag{4.25}$$

The functions $g(A)$ and $b(A)$ represent the real and imaginary parts of the saturated waveform when the input amplitude is tending to infinity. The angle of the asymptote is given by:

$$\delta = -\tan^{-1} \frac{b(A)}{g(A)} = -\tan^{-1} \frac{\cos \theta_1 - 1}{\sin \theta_1} = \frac{\theta_1}{2}. \tag{4.26}$$

The real axis intercept point is taken as the point of departure of the locus from the real axis. The locus departs the real axis as the rate limit becomes active. Defining the intercept point thusly adds a measure of conservatism to the stability surface as the asymptotic surface will always be intercepted before the actual surface. As noted previously, the rate limit becomes active when the slope of the input waveform exceeds K_{max} or, equivalently, when the amplitude of the input waveform is A_{min} for a given ω. The magnitude of the describing function for the saturation nonlinearity only is found in [25] and given in:

$$g_{int}(A_{min}) = \frac{1}{\pi} \left[2\sin^{-1}\frac{M}{A_{min}} - \sin\left(2\sin^{-1}\frac{M}{A_{min}}\right) + \frac{4M}{A_{min}}\cos\left(\sin^{-1}\frac{M}{A_{min}}\right) \right].$$

(4.27)

Using the trigonometric identities $\sin(2\alpha) = 2\sin(\alpha)\cos(\alpha)$ and $\cos\left(\sin^{-1}x\right) = \sqrt{1-x^2}$, (4.27) can be reduced as shown below:

$$g_{int}(A_{min}) = \frac{2}{\pi} \left[\sin^{-1}\frac{M}{A_{min}} + \frac{M}{A_{min}}\sqrt{1 - \frac{M}{A_{min}}^2} \right].$$

(4.28)

The value of the x-intercept is given as:

$$x_{int} = \frac{-1}{g_{int}}.$$

(4.29)

The asymptotic surface is plotted by forming a family of level lines for each ω. An overplot of Figure 4.11 of the asymptotic lines is shown in Figure 4.14. The asymptotic surface is conservative in that it moves the surface closer to the origin and intersections of the linear locus occur for lower system gains.

4.3 Using the Stability Surface for Qualitative Analysis of System Non-linearities

The effects of the two critical nonlinearities can be evaluated based on the effects each nonlinearity has on the shape of the stability surface. The understanding
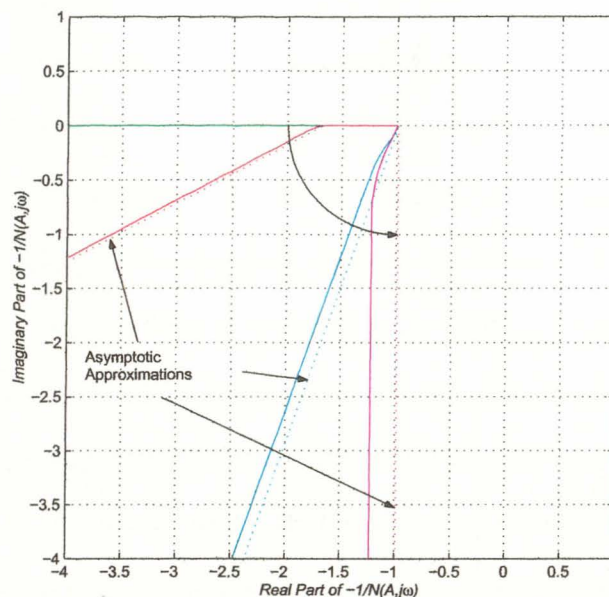
Fig. 4.14: Overplot of the Nyquist plot with the asymptotic approximation.

gained from this analysis is useful in system design in that the steering angle saturation can be judiciously used to enhance stability or improve performance. Also, an understanding of the effect of steering angle rate limit indicates the steering system performance needed to meet system specifications.

Generally, the absence of limit cycles is achieved when the locus of the linear sub-section of the data passes beneath the stability surface without intersecting the surface. Stability margins can be thought of as additional gain and phase needed to create an intersection of the stability surface. Thus stability surfaces that are completely above other surfaces tend to be more stable. This qualitative evaluation of surface position allows the qualitative effects of each nonlinearity to be understood.

4.3.1 Effect of the Steering Angle Limit

The steering angle limit has both positive and negative effects on system behavior. The limit tends to degrade high speed system response and thus performance.

Physically, limiting the steering angle keeps the vehicle from turning more sharply limiting the ability of the vehicle to correct for errors from the path. The limit has the overall effect of lowering the system gain accounting for the observed performance degradation. Normally, the steering limit of a vehicle is designed to be adequate for vehicle maneuverability, but the designer needs to be aware of the effects of this limit both when creating a new vehicle or modifying an existing vehicle.

In terms of vehicle stability, limiting the steering angle has the effect of lifting the stability surface away from the locus of the linear subsystem thus improving the system gain margin. This is consistent with the observation that the limit effectively lowers system gain. Nonconditionally stable systems tend to have larger gain margins. This nonlinearity can be used at higher vehicle speeds to improve stability and to lessen the effects of certain system failures that cause the vehicle to steer to a limit. The effect of reducing the steering angle limit is shown in Figure 4.15.
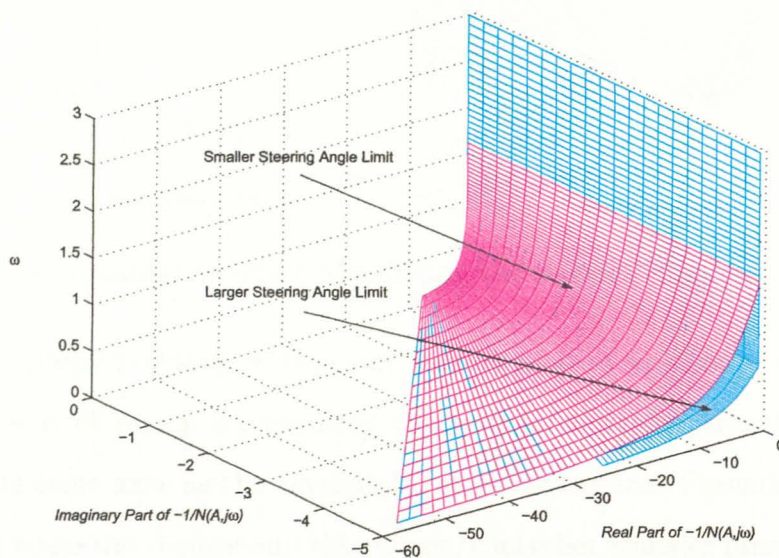
Fig. 4.15: Qualitative effect of reducing the steering angle limit.

4.3.2 Effect of the Steering Angle Rate Limit

The steering angle rate limit also has positive and negative effects on system performance. Unfortunately, the negative effects tend to outweigh the positive for this nonlinearity. Vehicles that have a low rate limit tend to have a 'softer' response to error, but this softer response comes at the cost of performance.

Lowering the rate limit tends to destabilize the system. In classical control systems terms, this is due to the additional phase lag that is added to the system. The effects of changing the steering rate limit on the shape of the stability surface are illustrated in Figure 4.16.
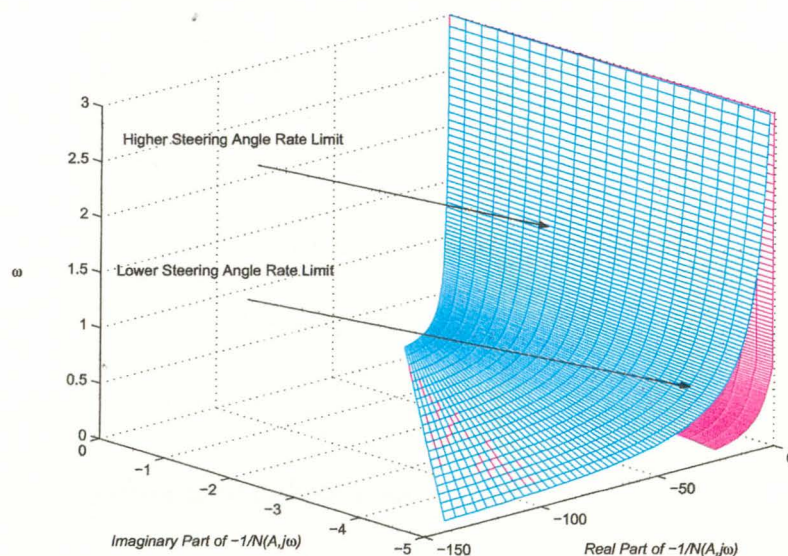
Fig. 4.16: Qualitative effect of reducing the steering angle rate limit.

The asymptotic stability surface can now be plotted as a series of level curves in ω. The locus of points representing the vehicle/controller transfer function is plotted on the same axes as the asymptotic stability surface. Potential limit cycles are indicated when the surface and the vehicle/controller transfer function intersect. Intersections always occur in pairs with the limit cycle occurring at the lower amplitude, or equivalently the higher frequency, intersection point. The proof that the

second intersection occurs is sketched by considering the behavior of the function along the negative real axis. This axis is chosen because the vehicle/controller transfer function traces this axis as the frequency ω tends to zero. The vehicle/controller locus tends to zero as $\frac{1}{\omega^2}$ while the stability surface tends to zero along the negative real axis as $\frac{1}{\omega}$. Therefore, if the vehicle/controller locus is above the surface at any point, it will pass below the surface as ω tends to zero.

4.4 Examples of Limit Cycle Prediction

A series of stability surface plots follows. The effect of increasing vehicle velocity on vehicle path tracking behavior is illustrated. The linear subsystem that is simulated is:

$$G(s) = G_{control} G_{steering} G_{vehicle},\tag{4.30}$$

where:

$$G_{control} = 10\frac{1/0.3s + 1}{s + 1}\tag{4.31}$$

$$G_{steering} = \frac{\pi}{180}\frac{12.5}{s + 12.5}\tag{4.32}$$

$$G_{vehicle} = \frac{v^2}{1.27}\frac{1}{s^2}.\tag{4.33}$$

The critical nonlinear parameters are:

$$M = 40\,\text{deg}\tag{4.34}$$

$$K_{max} = 30\,\text{deg}\,/sec.\tag{4.35}$$

The system of (4.30), (4.34), and (4.35) is based on the model of and measurements taken on a two hundred horse power agricultural tractor. The diagrams that follow are the results of simulation. The simulated velocities are too high to allow the vehicle itself to follow the induced limit cycles.

Example 1

In the first example, the vehicle velocity is set at 1 meter per second. No limit cycle is predicted from the plot of the stability surface and the linear subsystem in Figure 4.17. The simulation of the vehicle at this velocity is shown in Figure 4.18. The vehicle tracks the desired path within reason and there is no indication of oscillation.
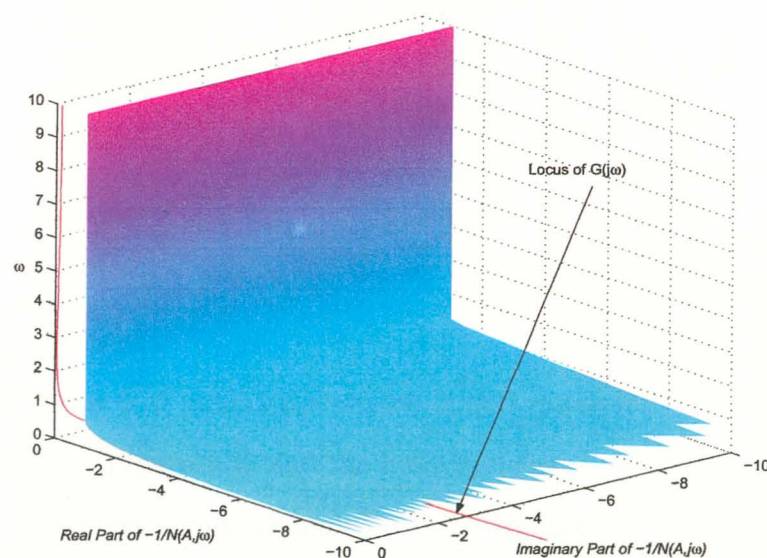
Fig. 4.17: No limit cycle indicated.

Example 2

In the second example, the vehicle velocity is set at 6.5 meters per second. A limit cycle is predicted by the stability surface plot of Figure 4.19. The simulation of the vehicle shows that a limit cycle does exist with the plot of vehicle position in Figure 4.20 showing large excursions across the path. Of particular interest is the fact that the oscillation does not start immediately after the vehicle starts to move. The limit cycle is latent in this case and must be simulated before it will start. Also worth noting is that near approaches to the path do not stop the limit cycle. Once initiated, the cycle will continue until the vehicle velocity is reduced.
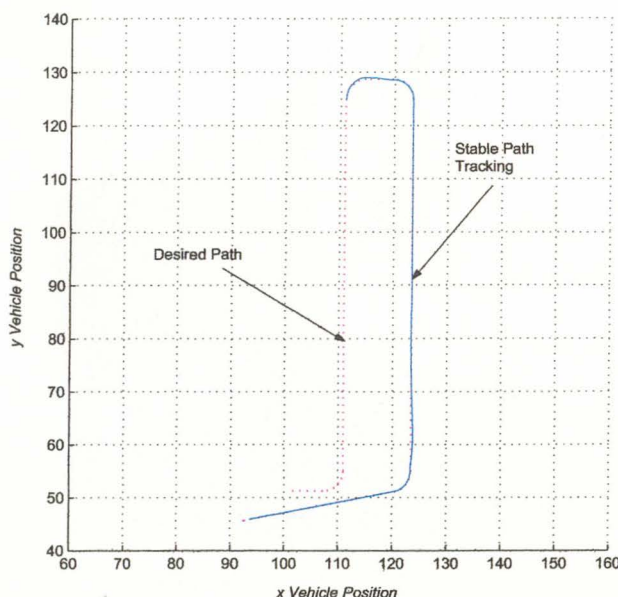
Fig. 4.18: No limit cycle.

Example 3

The velocity is set to 8 meters per second in this example. A limit cycle is again predicted as shown in Figure 4.21 and the limit cycle exists as shown in Figure 4.22. The frequency of the oscillation is high both in the distance it takes to cross the path per period and the time it takes to cross the path. Note that if the frequency of oscillation were the same as in Example 2, the space between path crossings would be greater in Example 3.

Example 4

As stated above, the asymptotic stability surface is conservative in estimating limit cycles. This last example illustrates the case when the method predicts a limit cycle when none exists. The vehicle velocity is set to 5.5 meters per second. The stability surface plot of Figure 4.23 predicts that the vehicle will oscillate, but the results of the simulation shown in Figure 4.24 shows no limit cycle. The simulation indicates that the vehicle is very active, but the oscillatory behavior damps out over time.
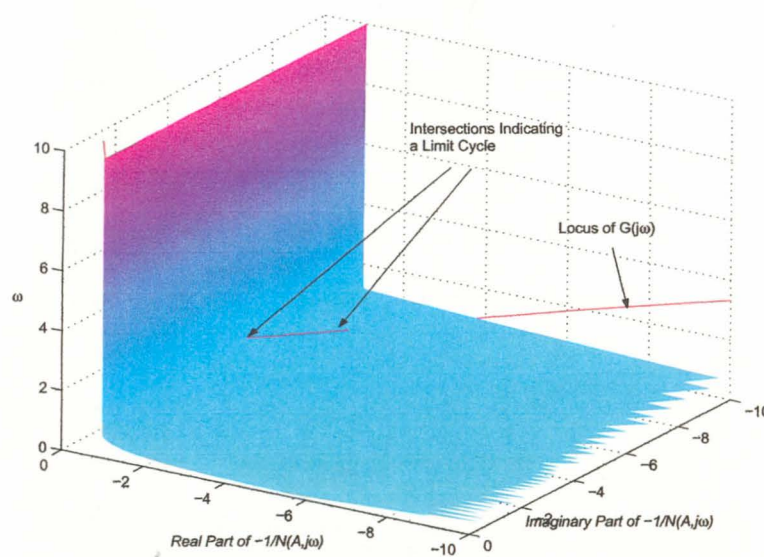
Fig. 4.19: Low-frequency limit cycle indicated.

The examples above illustrate the effectiveness of the method developed in this chapter in predicting limit cycles. The conservative prediction of limit cycle behavior allows a designer to avoid problem areas and aids in understanding the causes of limit cycle behavior in Ackerman steered vehicles.

4.5 Spatial Asymptotic Stability Surface

As presented to this point the stability design is concerned with the temporal behavior of the vehicle system. The analysis can be simply extended if an approximation of spatial independence is desired. The asymptotic level curves are functions of ω and in the frequency domain ω is related to the Laplace variable s by the substitution of $j\omega = s$. Subsequently, the asymptotic stability surface can be adjusted for the velocity by making the substitution:

$$\omega_v = \frac{\omega}{v}. \tag{4.36}$$
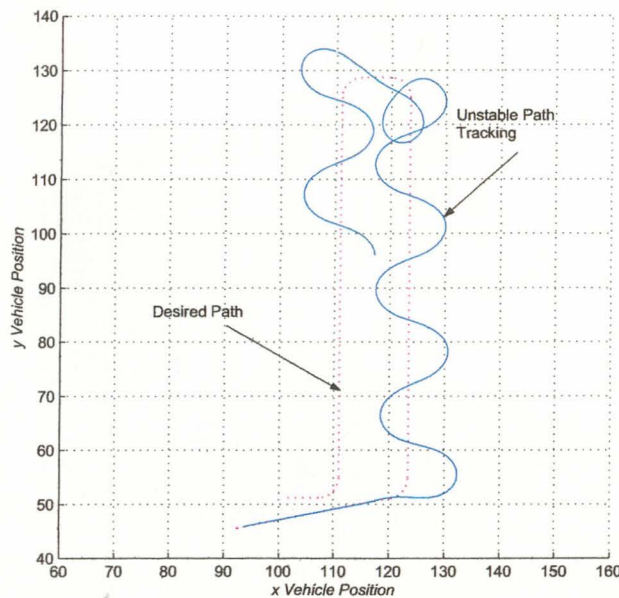
Fig. 4.20: Low-frequency limit cycle.

The stability surface of Fig. (4.12) is dependent on ω at three key points. The angle of the asymptotic level curve, the x-axis intercept point, and the ω-axis itself. If a further substitution is made by creating a new variable ω_v the stability surface can be plotted as a family of surfaces versus ω_v. The family of surfaces can be reduced to a bounding pair of surfaces representing the maximum and minimum operating velocities. As will be shown, the maximum velocity surface is the ultimate bound on stability and the stability surface is again reduced to a single bounding surface in the new variable ω_v with a dependence on velocity for the intersections and slopes of the asymptotic level curves. The vehicle and the controller are also transformed to the new variable ω_v with velocity dependence now found in the steering system pole and the measurement location zero.

The two critical quantities that define the level curves of the asymptotic stability surface are the angle of the level curve given in (4.26) and the x-intercept of the level curve given in (4.29). The values in these equations both depend on the radian
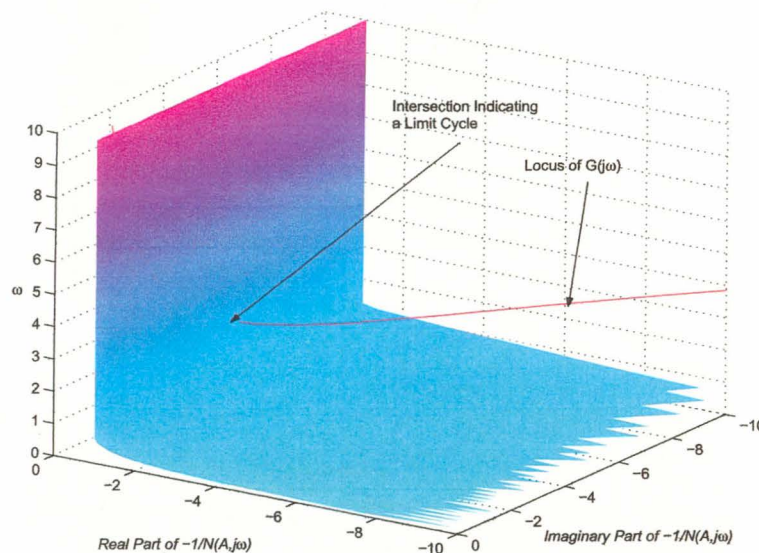
Fig. 4.21: High-frequency limit cycle indicated.

frequency ω. The explicit relationships in ω_v are given as:

$$\delta(\omega_v) = \frac{M}{2A_{min}} = \frac{M\omega}{2K_{max}} = \frac{Mv\omega_v}{2K_{max}} \tag{4.37}$$

$$g(\omega_v) = \frac{1}{\pi}\left[2\sin^{-1}\frac{Mv\omega_v}{K_{max}} + 2\frac{Mv\omega_v}{K_{max}}\cos\left(\sin^{-1}\frac{Mv\omega_v}{K_{max}}\right)\right] \tag{4.38}$$

$$x_{\text{int}} = \frac{-1}{g_{\text{int}}(\omega_v)}. \tag{4.39}$$

In order to show that the high velocity surface defines the stability bound for the system it is sufficient to show that the angles of the asymptotic lines are steeper and asymptotic intercepts are closer to the origin at the maximum operating velocity than at the minimum operating velocity for a fixed ω_v. In other words, if at any given frequency, ω_v, the asymptotic line for the higher velocity is completely to the right of the line for the lower velocity then the high velocity surface will be below the low velocity surface. Crudely, the higher velocity surface can be thought of as the low velocity surface shifted to the right and rotated clockwise.
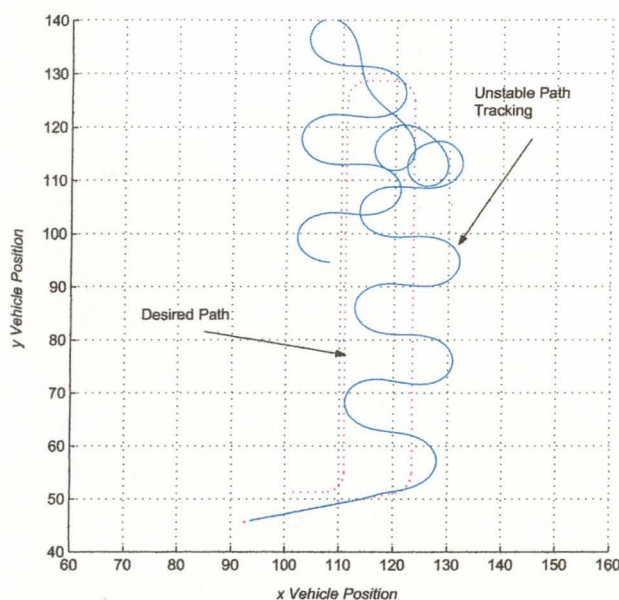
Fig. 4.22: High-frequency limit cycle.

It is easily seen from (4.37) that the angle increases when the velocity increases for a given ω_v. The requirement that the asymptotic intercept point move towards the origin is the equivalent of saying that $g_{\text{int}}(\omega_v)$ increases with increasing velocity.

The simplest way to show that the intercept points are always closer to the origin for the maximum operating velocity surface is to show first that $g(\omega_v)$ is monotonically increasing over the allowed range. This can be easily achieved by examining the derivative of $g(\omega_v)$ with respect to v. The derivative, after simplification, is shown below:

$$\frac{\partial g}{\partial v} = \frac{1}{\pi} \left[4 \frac{M\omega_v}{K_{max}} \left(1 - \left(\frac{M\omega_v}{K_{max}} \right)^2 v^2 \right)^{1/2} \right]. \tag{4.40}$$

The value of the derivative is always positive in the allowed interval indicating that the function g is always increasing over the interval. An examination of the function g at the extremes and the application of the central value theorem proves that the limiting stability surface is the one created using the maximum operating velocity of the vehicle. A representative plot of g is shown in Figure 4.25.
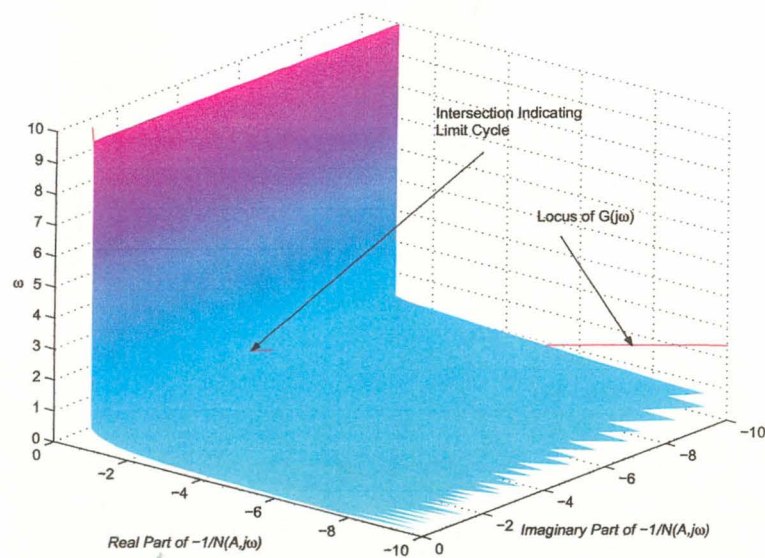
Fig. 4.23: Low-frequency limit cycle indicated.

4.6   Controller Synthesis

Methods of controller synthesis are many and varied ranging from simple pole placement, optimal control, and the ubiquitous proportional/integral/derivative controller (PID) plus other robust control methods intended to overcome the effects of parameter variation and unmodelled dynamics. Robust control techniques divide themselves into three primary classes; those based on classical frequency/complex domain design such as PID and quantitative feedback theory (QFT), additions made to modern control methods such as h-infinity, and exotic methods such as fuzzy logic and neural networks. The controller synthesis method specified in this section will be based in the classical domain in order to take full advantage of the stability surface derived in the previous section. It is worth noting that the design of the controller in the spatial frequency domain is identical to design in the normal frequency domain. The advantage of designing in the spatial domain is that the design needs to be done once for all velocities.
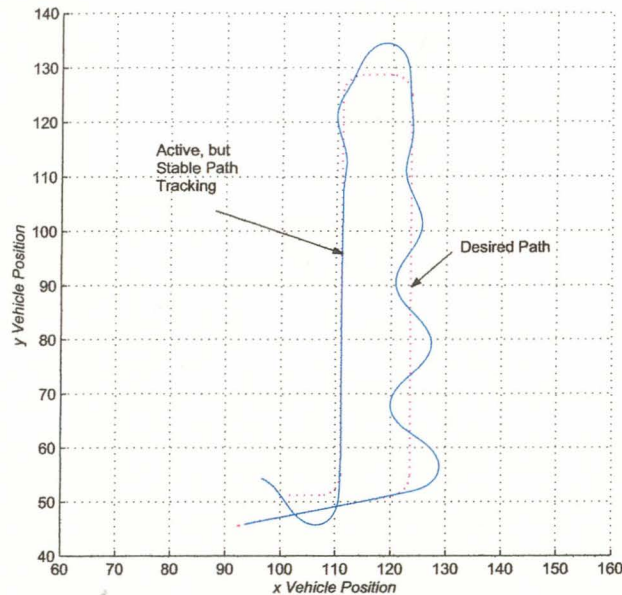
Fig. 4.24: No limit cycle exists in simulation.

The specification of the stability surface has preceded the controller design for two primary reasons; the first being that use of the surface makes it possible to design a controller that is stable over the operating range without the necessity of iterating the design. The second reason is that the shape of the surface itself leads to design of a controller that also meets performance requirements. Up to this point the stability surface has been drawn as an extended Nyquist plot with the extra axis being the radian frequency. However, the Nyquist plot is difficult to design with given that the addition of controller poles and zeros seems to have no quantifiable or predictable effect on the shape of the controller/plant locus. The Nichols chart is more useful as a synthesis tool for a number of reasons including the fact that both phase and magnitude are additive. Arguments encouraging the use of the Nichols chart are described by Horowitz in [29]. For the purposes of this research, the chart will be expanded in exactly the same manner as the Nyquist plot except that the radian frequency axis will be displayed on a logarithmic rather than linear scale.
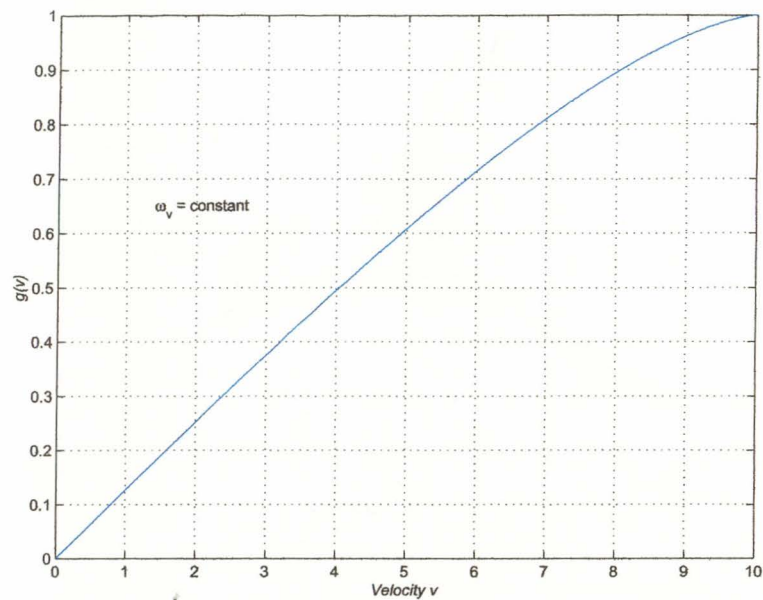
Fig. 4.25: Monotonic nondecreasing $g(v)$.

The asymptotic stability surface of the maximum operating velocity will be plotted on the axes and will be used both as a stability bound and a design guide for the design of a robust, effective controller. An example of a stability surface plotted on a three dimensional Nichols chart is shown in Figure 4.26.

The surface shown in Figure 4.26 consists of the elements of a standard Nichols chart with the addition of the logarithmic scale of radian frequency for the z-axis. The x-axis of the plot is the phase angle in degrees, the y-axis is gain in decibels. The chosen controller design must keep the locus of the linear vehicle subsystem/controller combination below the stability surface in order for the system to be stable. The shape of the surface is similar to the shape of the Nyquist plot. The vertical section near the 0-dB line relates to the vertical section on the Nyquist plot that runs along the $x = -1$ line. The base of the vertical section occurs at the radian frequency where the steering angle rate limit becomes dominant. The locus of the open loop system must be modified in order to avoid intersecting the surface. The
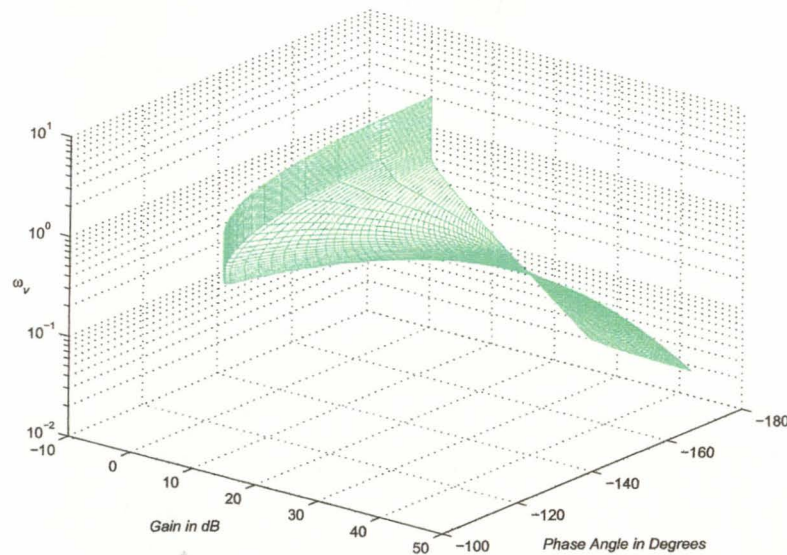
Fig. 4.26: A stability surface plotted on a Nichols chart.

method for modifying the locus and a simple method of achieving a robust controller are presented through an example.

Let the vehicle system be described below by:

$$G_v = \frac{1}{s_v^2(s_v/100 + 1)}\pi/180 \tag{4.41}$$

$$K_{\max} = 100 \tag{4.42}$$

$$M = 30. \tag{4.43}$$

The choice of controller type is open to individual preferences and some success has been obtained with a lead/lag or equivalently a proportional/derivative (PD) controller with a bandlimiting filter. The plot of Figure 4.27 shows the three dimensional stability surface. A simple design procedure is to use the stability surface as a

bound on system bandwidth and then design a controller with a reasonable compromise between bandwidth and performance. Also, a reasonable gain on the controller > 10 speeds the reduction of error to zero and aids in attenuation of transients and offsets.
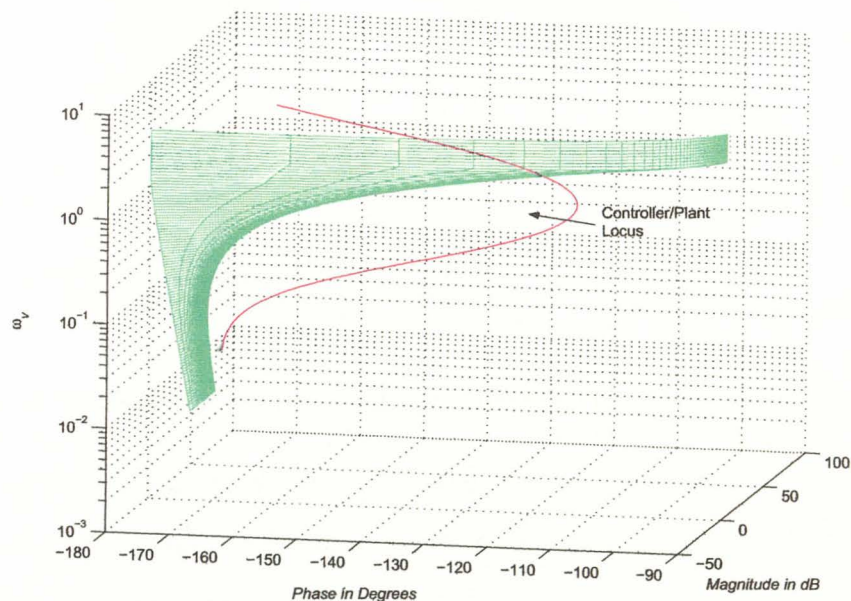
Fig. 4.27: Stability surface and vehicle/controller locus.

The surface and locus are shown in two more views in Figures 4.28 and 4.29 to clarify the shape and position of the data.

The controller used in the represented in Figure 4.27 was:

$$G_c = 10\frac{1/0.2s + 1}{1/5s + 1}. \tag{4.44}$$

The controller was chosen by the following procedure:

1. Observe the frequency where the vertical section of the stability surface begins (approximately 2radians/second in the example).

2. The controller zero is placed one decade below the frequency of (1).
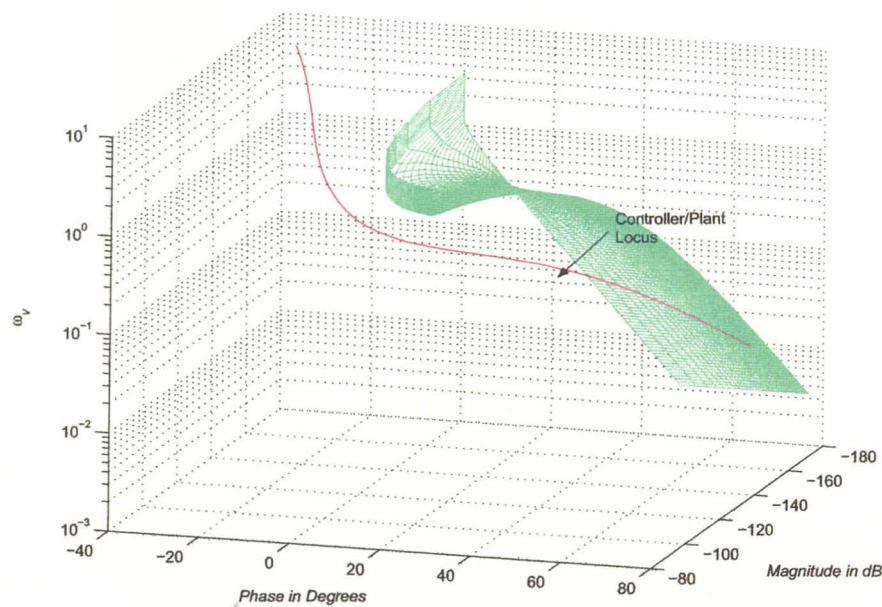
Fig. 4.28: Stability surface and vehicle/controller locus-second view.

3. The controller pole is placed at a higher frequency than the frequency in (1). The precise amount is subject to performance evaluations, but placing the pole a factor of twenty five worked well.

4. The controller is converted to the velocity adjustable form and implemented in simulation.

The results of the vehicle simulation with this controller are deferred to Chapter 5.

4.6.1 Implementation of the Velocity Adapted Controller

The implementation of the controller can be simply achieved by translating the stabilizing controller into the state-space domain while still using the spatial Laplace operator, s_v. The resulting state-space system has form:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{4.45}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}. \tag{4.46}$$
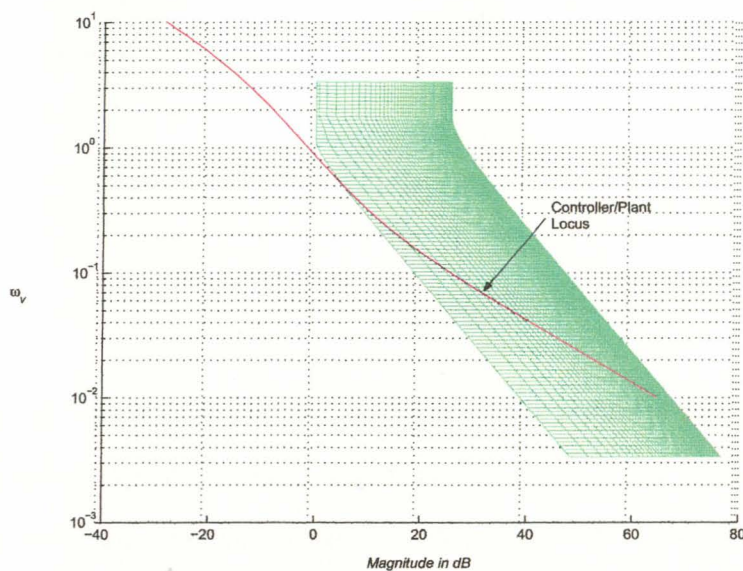
Fig. 4.29: Stability surface and vehicle/controller locus-third view.

The velocity dependence of the system matrices is illustrated by performing a spatial Laplace transform on the equation for $\dot{\mathbf{x}}$ as shown in the following progression:

$$(s_v\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{BU} \tag{4.47}$$

$$(s/v\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{BU} \tag{4.48}$$

$$(\mathbf{sI} - v\mathbf{A})\mathbf{X} = v\mathbf{BU} \tag{4.49}$$

$$\mathbf{X} = (\mathbf{sI} - v\mathbf{A})^{-1}v\mathbf{BU}. \tag{4.50}$$

The velocity adaptation is achieved by multiplying the $\mathbf{A}$ and $\mathbf{B}$ matrices of the controller by the vehicle velocity, v. The adaptive controller is shown in block diagram form in Figure 4.30.

Fig. 4.30: Implementation of the velocity adapted controller.

The velocity adapted controller of Figure 4.30 provides a simple method for controlling a vehicle over a broad range of vehicle velocities. The spatial response of the vehicle is similar over the entire velocity range. The controller does require a velocity sensor in order to perform the adaptation. It should also be noted that the vehicle model that is presented is derived for relative low speeds and higher speed vehicle operation may necessitate modelling of additional dynamics.

4.7 Systematic Design Method for Systems with Saturation/Rate Limit Nonlinearities

The methods presented in this chapter are not limited to the control of autonomous vehicles. Any system that includes a saturation/rat limit combination that is separable from a linear system can be analyzed using the methods presented. The steps that are taken follow:

1. Obtain the system model.

2. Separate the model into linear and nonlinear subsystems.

3. Identify the two critical nonlinear parameters, the saturation level, M, and the rate limit, K_{max}.

4. Using the above formulae, plot the stability surface on either a Nyquist plot or Nichols chart.

5. Plot the locus of the linear subsystem with the stability surface.

6. Identify system requirements such as DC gain, desired phase margin, desired gain margin, etc.

7. Shape the linear locus plot to meet the stability/performance requirements.

8. Simulate the system to ensure proper operation.

The method chosen for loop shaping is left to the individual designer. Both Dorf et al. [28] and Horowitz [29] provide good guidance on controller design techniques applicable to the locus shaping problem.

The procedure is identical for spatially independent controllers, but care must be taken to account for the velocities in all sections of the system.

4.8 Summary

This focus of this chapter was in the addition of feedback to the vehicle system. The addition of feedback to the vehicle system makes accurate control of the vehicle possible, but creates the possibility of unstable behavior. In order to understand this behavior, each element in the system was examined for its effects on the closed loop system.

The analysis of closed loop system began with an examination of the position measuring device, defining several key parameters that must be understood by the control system designer. Critical parameters included: accuracy, dynamic effects, pure time delays, and noise estimates. A sign was added to the orthogonal error calculated in Chapter 3 to make it useful as an error term in feedback.

A stability analysis of the closed loop system followed beginning with an overview of the research to date in the area of path following stability of Ackerman steered autonomous vehicles. The stability analysis of trajectory tracking and path following

systems has largely been approached through the Lyapunov method as applied to the kinematic model of the system and excluding any hard nonlinearities.

A qualitative analysis of the unstable modes of a vehicle with controller indicated that a stable limit cycle was the only unstable mode possible given the steering system limitations. Two distinct limit cycles were identified; an orbital behavior that occurs when the vehicle is too far off the planned path and a wandering behavior that occurs due to gain and phase lag related issues. An analysis of the two limit cycles was then presented. The analysis of the wandering behavior provided a unique look at the describing function of a cascaded saturation/rate limit nonlinearity and its effects on stability. The resulting criterion provides a conservative prediction of the onset of the wandering behavior. Additionally, the stability analysis provides a basis for the design of a robust controller with the traditional measures of robustness of gain and phase margin. The contribution of this research in the area of stability was the analysis of the effects of the steering system hard nonlinearities on near path stability.

The time/frequency based analysis of stability was converted to a spatially based criterion using the methods developed in Chapter 3. The resulting space/spatial frequency based system provides a basis for the design of a system that is spatially robust.

The chapter concluded with a presentation of a systematic design methodology. The methodology was presented for for time and spatial controllers.

The next chapter provides the results of simulation and experiment with the methods presented in Chapters 2 and this chapter. The chapter also discusses limitations of the methods presented and implementation issues that must be considered.

Chapter 5
Results of Simulation and Experiment

The previous chapters of this dissertation have dealt with the theoretical development of a control system for an autonomous vehicle. This theoretical development owes a great deal to the lessons learned while developing and testing many of the methods described in this dissertation on a series of agricultural vehicles. The algorithms in this dissertation were developed, tested, and refined while testing. Testing always exposes the shortfalls of models, the flaws in the hardware used to build physical systems, and the simple misinterpretations of how such a system should function.

This chapter will present descriptions of: the automated vehicle and its components, the practical difficulties encountered while developing the autonomous vehicle, and display the results of simulation and experiment.

5.1 Implementation Details

Several agricultural tractors were automated during the course of this research. The experimental data that is presented in this chapter was taken on a sixty five horsepower tractor converted to autonomous operation through the addition of an onboard computer system and the necessary sensors and actuators. The description of the tractor conversion will be limited to the components pertinent to the path tracking control of the vehicle.

The vehicle system modelled in Chapter 3 was reduced to an effective single input single output system as the result of some simplifying assumptions. The input to the vehicle system is the steering angle and the output of the system is the vehicle

deviation from the desired path. Minimally, the automated steering system must be designed so that it will respond to a steering command and the vehicle position must be measured so that the deviation from the commanded path can be calculated.

5.1.1 Steering System Implementation

The tractor came equipped with a power steering system that positioned the front wheels of the vehicle by moving a hydraulic ram in response to the motion of the steering wheel. The hydraulic system was automated by using an electro-hydraulic servo valve (ESHV) and a commercially available feedback controller, which controlled the movement of the valve to a commanded position. The chosen ESHV and controller combination performed well and included an effective deadband compensator that is similar to the one presented in Chapter 3.

In order to complete the steering system or so-called minor loop of Figure 3.5, the steering angle must be measured. Any of several techniques can be chosen to measure the steering angle. Measuring the linear extension of the hydraulic actuator is a possibility, but this measurement requires a sensor with an exceptionally long linear stroke and sensors with such range are normally expensive. Additionally, the sensor is difficult to protect from the environment without the added expense of placing the sensor inside of the hydraulic actuator. The more direct measurement of the wheel angle was chosen as a better alternative given that common angular sensors are compact, the sensor can be located in a more protected place, and the price of angular sensors is generally more reasonable than those of long linear stroke sensors. The combination of ESHV, controller, actuator, and sensor completes the steering system.

5.1.2 Vehicle Position Measurement System

The vehicle position measurement system chosen was a real time kinematic global positioning (RTK GPS) system device that was mounted one meter forward

of the driving wheels of the tractor. The RTK GPS chosen claims an accuracy of plus and minus two centimeters, which is true when the device has a clear view of the sky and there are an adequate number of satellites in view. The device provides vehicle position updates every 0.2 seconds. The global positioning system (GPS) used to measure vehicle position is at once invaluable to the vehicle system and challenging to use. The GPS unit was placed along the vehicle center line at approximately one meter in front of the rear driving wheels. Without the GPS unit measuring the vehicle location would be difficult and probably inaccurate, but GPS presents a series of problems that must be addressed in order to use the system effectively.

The first problem identified was the update rate of the chosen system. Given that the unit provides position updates every 0.2 seconds, the sample rate becomes a significant factor in the design of the vehicle control system and also effectively limits the vehicle top speed and/or performance characteristics. The position sampling time effectively provides a limit on system bandwidth. Augmented systems that use inertial estimates between samples have also been tested with the performance of the system being improved.

The second problem is that the position data produced by the unit is filtered in attempt to mitigate the effects of environmental 'noise' and the manufacturer of the GPS units used either could not or would not discuss the transfer function of the filter used. The filters that are used are generally low pass and will contribute measurable phase lag in a critical frequency range of the vehicle. In order to use the stability analysis method developed in Chapter 4 to the greatest affect, the GPS filter characteristics need to be understood.

The next problem compounds the update rate problem and the filtering problem. The measurement that is made available to the control algorithm every two hundred milliseconds is a delayed version of the actual vehicle position. This delay exists for all measurements and is caused by computational and communication delays in

the digital processing of GPS data. The amount of delay varies from manufacturer to manufacturer and is never clearly stated in the device specifications. Claims for the measurement delay range from microseconds to tens of milliseconds, but the performance of the GPS units in closed loop systems indicated a much longer delay as indicated by an early onset of limit cycle behavior with velocity, or equivalently, gain increase. The stability analysis of Chapter 4 will properly predict the onset of limit cycles if both the GPS filtering and delay properties are taken into account. Additionally, if the filtering and delay properties are well understood, the controller can be designed to compensate for the additional phase lag added by these devices. In the GPS system used, the total pure time delay was on the order of 0.4 seconds and had a significant influence on vehicle dynamics.

The final problem encountered with the GPS was the loss of reliable position measurement when the vehicle did not have a clear view of the sky or when the number of satellites in the constellation was simply too few for an accurate position fix. In the case of a limited number of in view satellites, the GPS could suddenly jump as much as ten meters from one position measurement to the next. This problem necessitated the addition of monitoring to the GPS system in order to identify false position measurements.

5.1.3 Vehicle Computer System

The onboard tractor computer is a single board, PC compatible computer running a Linux based, real-time operating system. The computer controls the interface to the steering controller and the RTK GPS. The path tracking controller and the segment switching criterion are written in the C computer language. The computer also controls the actuation of clutch, throttle, brake, three-point hitch, and power take-off, but these components are not considered pertinent to this research.

The onboard vehicle computer communicates with a base station computer through a wireless link. The base station computer derives and sends the desired path commands and vehicle control commands to the vehicle computer. Control commands consist of operations such as start the vehicle, start the mission, halt the mission, and manual operation of the vehicle. The base station computer also includes a redundant radio link that provides human operators an emergency stop switch.

5.1.4 Controller Implementation

The vehicle controller is converted to a discrete equivalent using the pole-zero matching technique. The pole-zero matching technique was chosen because of the close frequency response match that it provides between the continuous controller and the discrete. The sample time of the controller is tied to the vehicle position sample time of 0.2 seconds/sample.

5.2 Simulation Results for Spatially Similar Systems

The results of the velocity adapted controller developed in Chapter 4 are presented in this section. The section will present time and spatial responses of the vehicles with different starting positions and orientations to the path. The section will also demonstrate the degradation of performance as the linear subsystem is moved closer to the stability surface thus reducing the gain and phase stability margins.

5.2.1 Time and Spatial Response Plots

The following results are based on simulation the system and controller of subsection 4.6.

The time and spatial responses of the system when the vehicle velocities are one and three meters per second are shown in Figures 5.1 and 5.2. The vehicle is

positioned at $(0, 2)$, the desired path is along the x-axis, the vehicle is oriented along the x-axis. As expected, the time responses are very different from one another. In order to follow the same spatial trajectory, the time response of the high velocity system must be faster. The spatial responses show almost no perceptible differences Differences do occur in the spatial responses but they are not generally large if operation occurs within the bounds of the design.

The time and spatial responses for the system when the vehicle is located at $(0, 1)$, the desired path is the x-axis, and the vehicle is oriented at angle $\pi/2$ radians to the x-axis are shown in Figures 5.3 and 5.4. Note that some differences do exist between the spatial responses in the spatial response plot. The cause of this difference is related to the remaining velocity dependent or time invariant values in the system. In this case a combination of rate limit and saturation causes the difference.

The response plots so good spatial agreement. A limitation of the method is that the high velocity response limits the low velocity response. Limitations of this method will be discussed in Chapter 6.

5.2.2 Performance Degradation in Relation to Proximity to the Stability Surface

Classically, a rule of thumb for design of controllers on a Nyquist plot has been that performance degrades as you approach the -1 point on the plot. In Eveleigh [25] a statement is made to the effect that reduction of over shoot is achieved by "general -1 point avoidance." It is therefore reasonable to ask the same question of the stability surface method. The following plots demonstrate the effects of moving the surface closer to the linear subsystem in both magnitude and phase.

The first pair of plots show the performance of the vehicle in response to a initial perturbation. Figure 5.5 shows the effect on time response of adjusting loop gain. Figure 5.6 is the corresponding stability surface plot. The gain of 47 was chosen to be
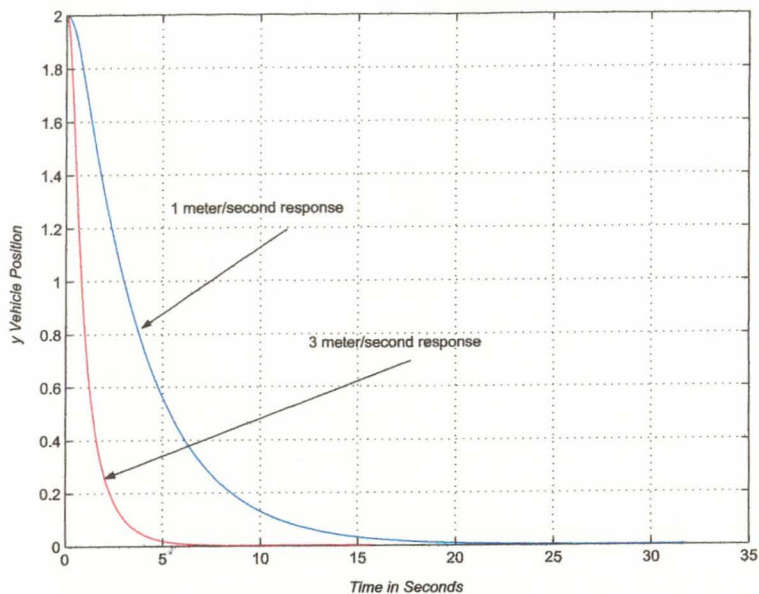
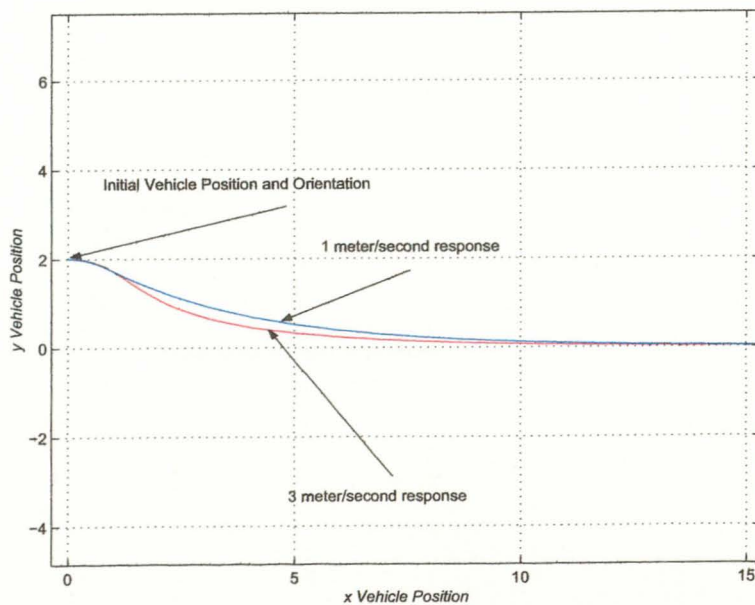Fig. 5.1: Time response plot-orientation 1.

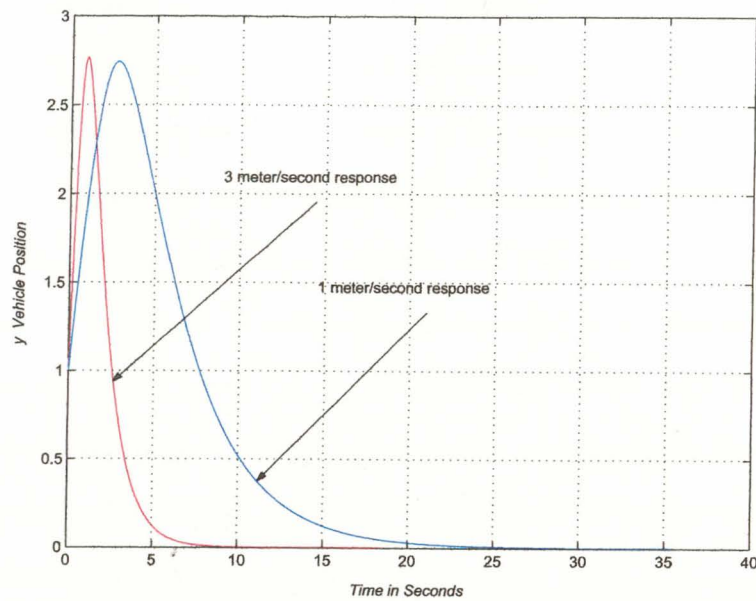Fig. 5.2: Spatial response plot-orientation 1.
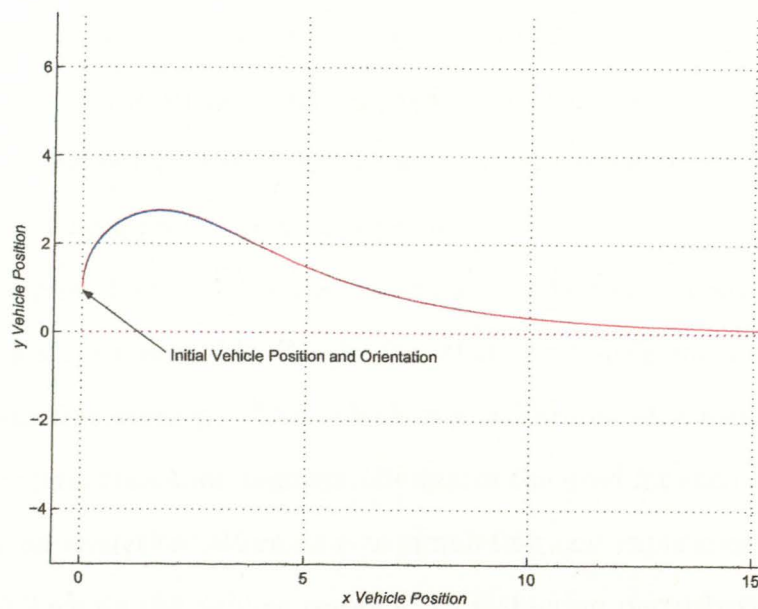
Fig. 5.3: Time response plot-orientation 2.

Fig. 5.4: Spatial response plot-orientation 2.

on the edge of stability. Oscillations are nearly sustained, but the system stabilizes, The locus of the linear subsystem is barely intersecting the stability surface. Again, the surface is drawn the provide a conservative estimate of the onset of a limit cycle.

Adding additional phase to the system is less straight forward. Phase was added by including a time delay or transport lag. The resulting system shows the onset of instability at a time delay of 0.3 seconds and full limit cycle with a time delay of 0.5 seconds. The system is stable for a time delay of 0.1 seconds. The resulting responses are shown in Figure 5.7. A corresponding surface plot is not shown given the lack of direct correlation between time delays and constant phase shifts.

As expected, performance as well as stability can be enhanced by avoiding the stability surface to the extent possible.

5.3   Results of Vehicle Experiments

The results presented in this section were measured on the autonomous vehicle described above. The path switching algorithm was an empirically tuned algorithm that served as the impetus for developing the spiral transition segments presented in Chapter 2. The controller is a lead/lag style controller that is not velocity adapted. The data is presented to show the similarity in vehicle response to transient conditions between simulation and actual experiment.

The first plot, Figure 5.8, shows the vehicle repeatedly making the same turn. The performance is reasonable due the fact that the vehicle was manually tuned to start the transition segment. The vehicle was calibrated at a series of discrete velocities. The spiral transition segment eliminates the need for such hand adjustment and provides an analytical alternative to simulation and experiment.

Figure 5.9 shows the vehicle response to a steering perturbation. The vehicle recovers and continues on the path. The shape of the return response is similar to that of Figure 5.1.
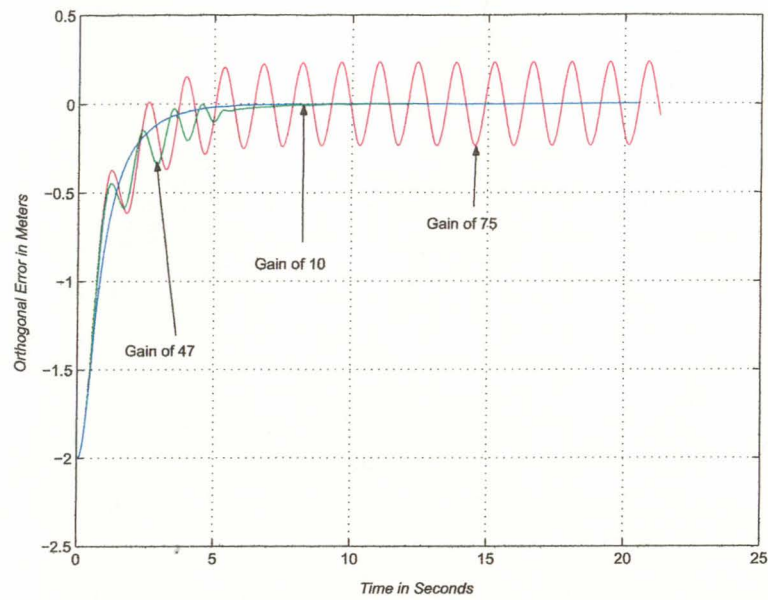
Fig. 5.5: Reduction in gain margin-performance degradation.
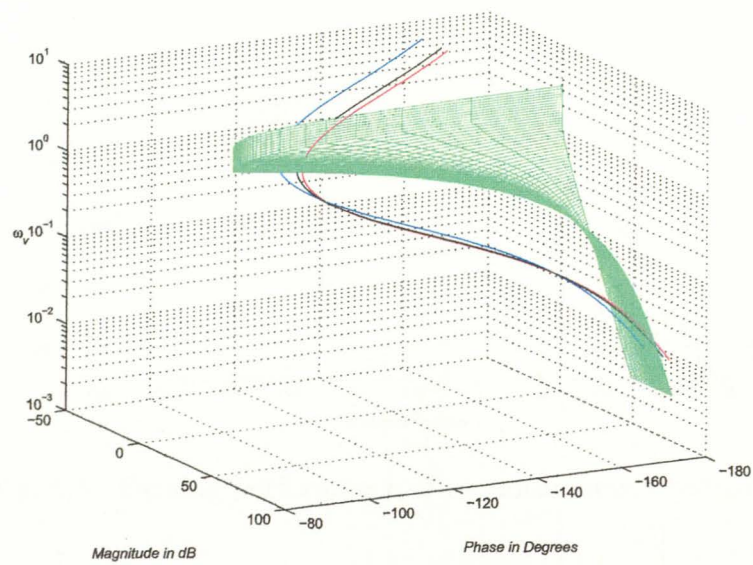
Fig. 5.6: Reduction in gain margin-stability surface plot.
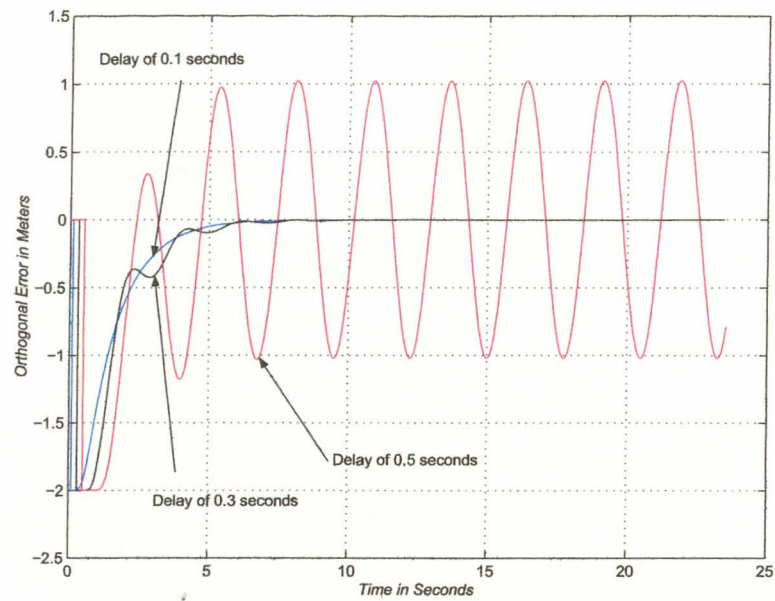
Fig. 5.7: Reduction in phase margin-performance degradation.
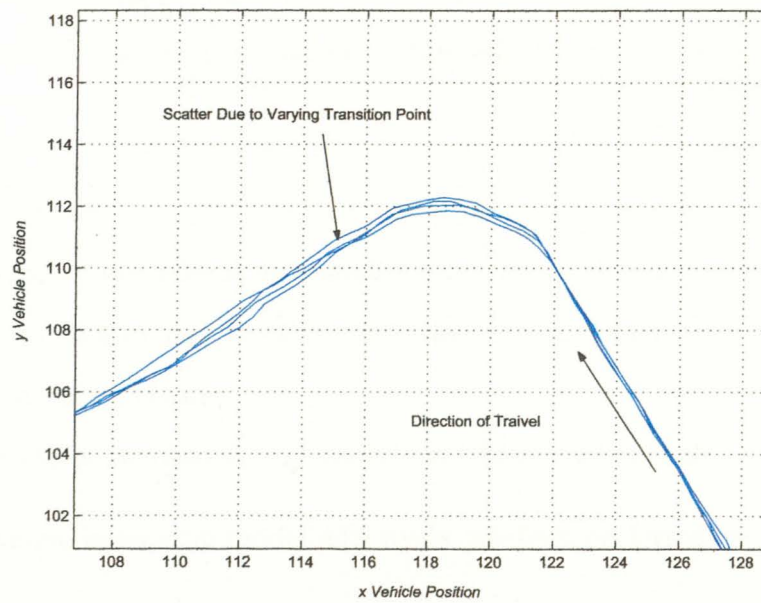
Fig. 5.8: Turning performance of an autonomous vehicle.
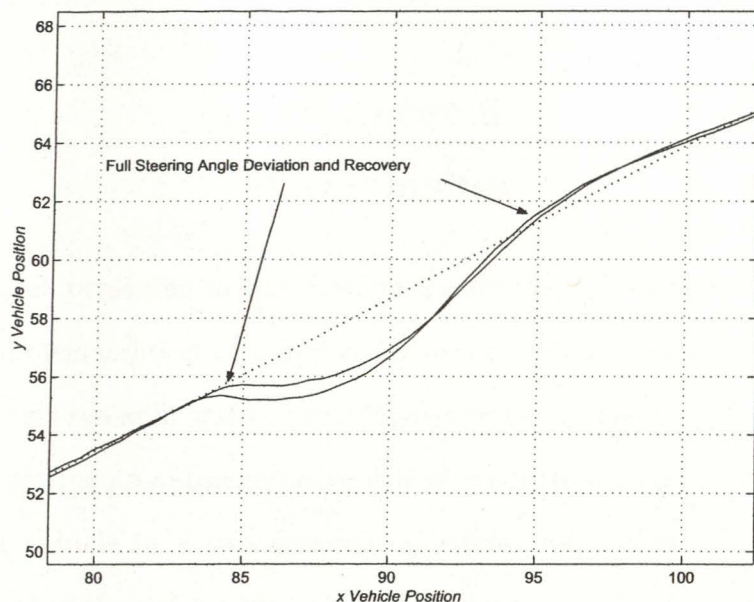
Fig. 5.9: Path following response to a steering perturbation.

In general, the simple kinematic model agrees well with the experimental results. Modelled controllers are transferred to the vehicle with little or no modification required.

5.4 Conclusion

Simulation of the vehicle was used extensively in this dissertation. Experiment agrees with simulation and effective controllers are implemented on working vehicles without the need for change.

The results of simulation and experiment have shown that:

1. The simple kinematic model effectively predicts performance of the vehicle.

2. The use of spiral segments provides an analytic replacement for hand tuning. The implementation of this approach needs more work.

3. The stability surface has proven to be an accurate and useful tool in controller design for stability and prediction of stability margins.

Chapter 6
Conclusion

The research presented in this dissertation developed methods for the analysis, design, and implementation of autonomous vehicle path tracking control systems. The intent of this research was to provide an alternative method of solution for the problem of operating an autonomous vehicle along a preplanned path. The problem of guiding the vehicle in a two-dimensional environment was approached from a spatial view rather than the more common time-based view. Contributions to the field of autonomous ground vehicles were made in four primary areas:

(1) Production of a generalized path from training data.

(2) The spatial modelling of the vehicle with steering system.

(3) The analysis of vehicle stability.

(4) The design of a spatially robust controller.

The specific contributions to the path training technique include:

(1) Reduction of the training path to a generalized path consisting of circular arcs and straight lines.

(2) Significant reduction of the training set data size.

(3) The addition of transition segments, specific to the vehicle or task, to the generalized path.

(4) The use of spirals to provide a point on the path for changing segments that is analytically derived.

The specific contributions to the modelling of autonomous ground vehicles include:

(1) An intuitive view of the kinematic model of an autonomous ground vehicle with Ackerman steering.

(2) The conversion of the vehicle model from a time dependent to spatially dependent base.

The specific contributions to the analysis of stability include:

(1) The separation of the vehicle into nonlinear and linear subsystems.

(2) The describing function analysis of the saturation rate limit combination.

(3) The creation of the frequency dependent 'stability surface' for visualizing system stability.

(4) The asymptotic approximation of the stability surface to reduce computational load and enhance qualitative understanding of the surface.

(5) A qualitative analysis of the effects of critical nonlinearities on stability and performance.

(6) The conversion of the stability surface to the spatial frequency domain.

(7) The design of the vehicle controller using the spatial Laplace transform in concert with the spatial stability surface.

The methods presented in this research are by no means the only possible approaches to the problems encountered in the development of autonomous vehicles, but they present viable alternatives and also provide analytical explanations for observed vehicle behavior. The design methods allow an easy adaptation from vehicle to vehicle with minimal adjustments required.

The design methods also meet the real-world requirements faced while designing autonomous vehicles. The number of sensors required have been reduced to the steering angle sensor, a single position measurement sensor, and an inexpensive heading sensor. The need for accurate heading measurements is eliminated, and the heading sensor is needed only for the far path controller, which does not require great accuracy.

The remainder of the chapter presents in order directions for future research in the areas of task management, path management, and areas related to the general development of autonomous ground vehicles. A brief discussion of the economic viability of autonomus vehicles follows. The chapter concludes with a discussion of the strengths and weaknesses of the methods presented.

6.1 Directions for Future Research in the Areas of Task Management

As stated in Chapter 1, the area of task management is broad, problem specific and not amenable to a single solution. The training method presented in this research is only one of numerous possibilities for providing a path to an autonomous vehicle. This section describes possible extensions to the method presented.

In the area of the vehicle training method of task management, the reduction of vehicle position data to vehicle path commands can be improved. The present selection criteria do not include any attempt at optimization of the resulting data set and further reduction of the data set is certainly possible. Also, the case where the path training is done with a vehicle that is different from the autonomous vehicle

remains to be considered. The differences in vehicle minimum turning radii presents an interesting problem in data reconstruction and reduction.

6.2 Direction for Future Research in the Area of Path Management

The stability criterion that was developed in this research provides a rudimentary performance criterion through the gain and phase margins found by using the criterion. An additional surface can be added to the three-dimensional Nyquist plot representing performance rather than stability bounds. The controller can then be designed so as to avoid intersecting both the describing function of the nonlinear subsystem of Figure 4.9 and the performance surface.

Another research area would be in the automation of controller design based on the stability and performance surfaces described in the previous paragraph. The development of such an algorithm could include the specification of desired controller order or search for a controller of minimum order that still meets the stability and performance specifications.

Spiral segments are most effective as a segment switching tool when the measurement point is at the turning center of the vehicle and the time rate of change of curvature is reasonably related to the rate limit of the steering actuator. Additional study is needed to determine effect of both the measurement point and the steering actuation geometry.

6.3 Directions for Future Research in Related Areas

In Chapter 1 five critical areas were presented as being important to the design an implementation of autonomous vehicles. These areas are: vehicle safety, task management, path management, vehicle function, and system reliability. This research focused on only two of the five areas, but research is needed in all five of the areas listed plus assuredly in other unthought of areas.

Safety is absolutely critical to the success of autonomous vehicles. Autonomous vehicles can be large and fast and operate in environments that do not have completely controlled access. Safe operation in uncontrolled environments presents an interesting and important area of research. The development of reliable obstacle sensors, and more importantly the development of sensors that accurately detect human and animal presence is an area of active research.

Sensing the presence of any obstacle, living or inanimate, is only part of the problem. Can a failed sensor be detected, or does the system run with a failed, critical sensor without the human operator knowing? This question spawns many more questions. How reliable is the detection system? If the vehicle system is designed to detect failures, can the failure detection scheme falsely indicate a failed system when no real failure is present? False indication of system failures requires the performance of expensive maintenance and causes lost work time.

It is clear from the series of questions in the previous paragraph that the design and development of an autonomous vehicle is a complex system and a systems design approach is needed to successfully create working autonomous vehicles that are more than laboratory curiosities. The complex and sometimes conflicting requirements that a autonomous vehicle must meet is reminiscent of the requirements that airplane designers face and those pursuing the design of autonomous vehicle systems would be wise to access the body of systems design knowledge existing in that industry.

The quasi-spatial control methods developed in this research generalize well to higher dimensional systems that follow space curves. Examples include aircraft and robotic manipulators. The primary advantage of the method, that of being relatively insensitive to disturbances that slow progress along the desired paths, translates well into multidimensional paths in disturbed environments. An example is found in the automation of backhoe style manipulators that are required to follow paths through the ground in normal operation.

6.4 Economics of Autonomous Vehicles

Another, and yet unspoken, goal of this research was to design an autonomous vehicle that represents the economic viability of autonomous vehicles in industrial and agricultural environments. The vehicle sensor suite was chosen so as to limit the total cost and increase the probability of the widespread acceptance of the vehicle. The only sensors required are the GPS units (RTK GPS requires a fixed base station) and the steering angle sensor.

The autonomous vehicle described in this research is capable of being commanded to drive complex paths without constant operator intervention. Several manufacturers have designed and marketed tractors with operator assist functions that allow the tractor to follow parallel paths after the driver has trained the vehicle on the first pass. After completing the first pass the human operator turns the tractor at the row end, aligns the tractor with the next row, and turns command over to the augmentation system. The process is repeated with the driver turning and aligning the tractor at each new row end. These systems are considered particularly valuable for large farms with crops requiring long straight rows.

Driving under automatic control with a human operator providing emergency response and obstacle detection certainly simplifies automatic row following, but the benefits of such a system are limited to the production of consistent rows and the reduction of operator fatigue. The operator remains in the vehicle and must be alert at all times as the vehicle drives along the desired path. The cost of the operator and the operator station on the vehicle are added to the nearly identical components required for producing a fully autonomous vehicle. The cost of the human operator and operator station present a target cost for the safety sensing systems that must be added to the vehicle for full autonomy.

6.5 Strengths and Weaknesses of the Methods Developed in This Dissertation

The three primary derivations of this dissertation were the derivation of the general path from training data with spiral segments, the asymptotic stability surface analysis, and the design of a controller in the spatial frequency domain. Each of the derived methods has strengths and weaknesses. The weaknesses and strengths of the general path method are presented first. Next, the weaknesses and strengths of the asymptotic stability surface. The section concludes with a description of the weaknesses and strengths

The general path is derived from training data and at the present time the method presented here does not work well when the vehicle taking the data is not the same as the autonomous vehicle. Research into path adjustments for this case needs to be completed. The method also provides paths that are somewhat general. Obviously, if the vehicle steering rate were slowed too much connecting spirals could not synthesized. The strengths of the method are that a general path exists for most cases of practical vehicle motion and that spiral transition segments can be added to the system to enhance path driveability. The other strength is that the derivation has reduced the equations to a simple series of equation that bypass complex geometry.

The asymptotic stability surface represents an approximation to the actual nonlinear system. Care should be taken in its application and the assumptions of the describing function method should be kept in mind. The surface needs more work to determine more analytic approaches to controller design.

The spatially independent controller has several weaknesses. The first being that the high velocity vehicle spatial response limits the spatial response of the low velocity. Also, the velocity adaption is a complexity that would be better avoided. Finally, on a positive note: the spatial responses are very similar.

6.6   Summary

Research in the area of autonomous vehicles remains an interesting and challenging endeavor and no single approach will solve all the problems. The vehicles are moving towards commercial viability in several arenas: vehicle/vehicle component testing, agricultural, and military applications to name a few. It is impossible to predict what the future will be, but it will likely include autonomous ground vehicles in some form. It is my hope that the research presented in this dissertation has furthered knowledge and will provide a impetus for future research in the areas presented.

References

[1] K. N. Murphy, M. Juberts, S. A. Legowik, H. S. Marilyn Nashman, H. A. Scott, and S. Szabo, "Ground vehicle control at NIST: from teleoperation to autonomy," tech. rep., National Institute of Standards and Technology, 1993.

[2] H. Durrant-Whyte, "A critical review of the state-of-the-art in autonomous land vehicle systems and technology," Unlimited Release SAND2001-3685, Sandia National Laboratories, Albuquerque, New Mexico, Nov. 2001.

[3] T. Fraichard, A. Scheuer, and R. Desvigne, "From Reed's and Shepp's to continuous-curvature paths," in *In Proceedings of the IEEE International Conference on Advanced Robotics*, 1999.

[4] Y. Kanayama and S. I. Yuta, "Vehicle path specification by a sequence of straight lines," *IEEE Journal of Robotics Automation*, vol. 4, pp. 265–276, June 1988.

[5] S. Gray, S. Hansen, and N. Flann, "Dynamic replanning for multiple unmaned ground vehicles using the JAUGS architecture," in *Association for Unmanned Vehicle Systems International*, 2002.

[6] F. D. del Rio, G. Jimenez, J. L. Sevillano, C. Amaya, and A. C. Balcells, "A new method for tracking memorized paths: Application to unicycle robots," in *Proceedings of the 10th Mediterranean Conference on Control and Automation - MED2002*, (Avda. Reina Mercedes s/n. 41012 Sevilla. SPAIN), Robotics and Computer Technology for Rehabilitation Laboratory (Universidad de Sevilla), July 2002.

[7] K. Komoriya and K. Tanie, "Trajectory design and control of a wheel-type mobile robot using B-spline curve," in *International Workshop on Intelligent Robots and Systems*, pp. 398–405, 1989.

[8] Y. K. Hwang and N. Ahuja, "Gross motion planning - a survey," *ACM Computer Surveys*, vol. 24, no. 3, pp. 219–290, 1992.

[9] D. H. Shin and S. Singh, "Path generation for robot vehicles using composite clothoid segments," Tech. Rep. CMU-RI-TR-90-31, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 1990.

[10] H. Ishimoto, T. Tsubouchi, S. Sarata, and S. Yuta, "A practical trajectory following of an articulated steering type vehicle," in *International Conference on Field and Service Robotics (FSR'97)*, pp. 412–419, 1997.

[11] A. Scheuer and T. Fraichard, "Continuous-curvature path planning for car-like vehicles," in *IEEE-RSJ International Conference on Intelligent Robots and Systems*, 1997.

[12] B. Nagy and A. Kelly, "Trajectory generation for car-like robots using cubic curvature polynomials," in *Proceedings of Field and Service Robots (FSR 2001)*, 2001.

[13] M. D. Adams, "Stability and high speed convergence in mobile robotics," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 2, pp. 230–237, 1999.

[14] S. L. Salas and E. Hille, *Calculus: One and Several Variables*. Lexington, MA: Xerox College Publishing, 2 ed., 1974.

[15] G. E. Shilov, *Elementary Real and Complex Analysis*, p. 140. New York: Dover Publications, Inc., 1 ed., 1996.

[16] C. H. Oglesby and L. I. Hewes, *Highway Engineering*. New York: John Wiley and Sons, 1 ed., 1963.

[17] R. C. Hibbeler, *Engineering Mechanics: Statics and Dynamics*. New York: Macmillan, 6 ed., 1992.

[18] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*. New York: Oxford University Press, 4 ed., 1998.

[19] A. D. Luca, G. Oriolo, and C. Samson, "Feedback control of a nonholonomic car-like robot," in *Planning Robot Motion* (J.-P. Laumond, ed.), Berlin, DE: Springer-Verlag, 1998.

[20] A. Ollero and G. Heredia, "Stability analysis of mobile robot path tracking," pp. 461–466, IEEE/RSJ International Conference on Intelligent Robots and Systems, 1995. Pittsburgh, PA.

[21] V. Bahl, "Modeling and control of a class of autonomous wheeled mobile robots," Master's thesis, Utah State University, Logan, UT, 2002.

[22] T. Bak, J. Bendtsen, and A. P. Ravn, "Hybrid control design for a wheeled mobile robot," in *Lecture Notes in Computer Science*, vol. 2623, pp. 50–65, 2003.

[23] J. Ackermann and T. Bünte, "Actuator rate limits in robust car steering control," in *Proceedings of the 36th IEEE Conference on Decision and Control*, 1997.

[24] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ: Prentice Hall, 1 ed., 1991.

[25] V. W. Eveleigh, *Introduction to Control Systems Design*. New York: McGraw-Hill, 1 ed., 1972.

[26] J. Kudrewicz, "Theorems on the existence of periodic vibrations based upon the describing function method," in *Proceedings of the International Federation on Automatic Control*, The International Federation on Automatic Control, 1969. Warsaw, Poland.

[27] A. R. Bergent and R. L. Franks, "Justification of the describing function method," *SIAM Journal of Control*, vol. 9, pp. 568–589, Nov. 1971.

[28] R. C. Dorf and R. H. Bishop, *Modern Control Systems*. Upper Saddle River, NJ: Prentice Hall, 9 ed., 2001.

[29] I. M. Horowitz, *Quantitative Feedback Design Theory*. Boulder, CO: QFT Publications, 1 ed., 1993.

Curriculum Vitae

Donald Lewis Cripps

Professional Experience

January 2001 to Present

Senior Controls Engineer, Autonomous Solutions, Inc., Wellsville, Utah

Responsible for control of autonomous vehicles, design of control/measurement algorithms for vehicle and non-vehicle projects, and project/personell management for various projects. Our vehicles range in size from a ten horse power electric cart to a two hundred horse power agricultural tractor.

July 1995 to December 2000

Research/Design Engineer, Utah State University, Logan, Utah Center for Self-Organizing and Intelligent Systems (CSOIS)

Responsibilities included control system analysis and controller design, analog and digital electronic design, and a vehicle-training algorithm. I have worked on an autonomous wheelchair, an autonomous omni-directional vehicle, an ice sensor used in a weather modification experiment, an autonomous, a two hundred horsepower farm tractor, and an autonomous orchard-spraying tractor.

March 1997 to December 2000

Lecturer, Utah State University, Logan, Utah

I have taught the following courses, including lectures and laboratories, at Utah State University:

I have taught the following courses, including lectures and laboratories, at Utah State University:

- Classical Control Systems

- State-Space Control Systems

- Discrete/Digital Control Systems

- Sensors and Actuators in Control Systems

- Analog and Digital Electronics for Mechanical Engineers (graduate elective)

- Basic Microprocessors (Industrial Technology Department)

January 1990 to July 1995

Senior Specialist Design Engineer, Boeing Commercial Airplane Company, Seattle, Washington Mechanical Systems Division, Flight Control Electronics

Responsible for the spoiler control system on the 757 and 767 airplanes. I oversaw the technical content of the redesign of the Spoiler Control Module (SCM). The original SCM was a triple channel redundant, fail-operate module with a poor maintenance history. The redesigned module became a model of reliability and ease of use at Boeing.

Additional design responsibilities included oversight of the technical content for the redesigned flight control electronics power supply, design of an electronic replacement for a mechanical force feedback computer, and design and construction of electronic analogs of flight surface hydraulic actuators.

I also worked as an inter-group consultant on diverse projects on the 737, 747, 757, 767, and 777 airplanes.

June 1982 to December 1986

Senior Research and Development Engineer, Sunpower Incorporated, Athens, Ohio

Sunpower conducted research and development work on Stirling cycle engines, heat pumps, and cryogenic coolers. I was hired to provide instrumentation design and support. My responsibilities expanded to include analog and digital electronics design, engine analysis and simulation, and control system design.

Educational Experience

October 1995 to Present

Ph.D. Studies, Utah State University, Logan, Utah

Completed course work for degree.

Successfully defended dissertation on December 1, 2003

Dissertation title: Spatial and Temporal Considerations in Vehicle Path Tracking with an Emphasis on Spatial Robustness

August 1988 to August 1989

Graduate studies, University of Colorado, Boulder, Colorado

Studied for two semesters while teaching electronics laboratories

Attended on a teaching fellowship

September 1981 to June 1982

Graduate studies, Ohio University, Athens, Ohio

Completed course work for Masters Degree

Attended on a Stocker fellowship

September 1974 to June 1978

Bachelors degree studies, University of Utah

Graduated Magna Cum Laude

Member Tau Beta Pi

Publications

D. Cripps and B. Abbott, "The use of capacitance to detect icing," The Journal of Weather Modification, vol. 29, no.1, pp. 84-87, April 1997.

M.W. Torrie, R. Koch, V. Bahl,, and D. Cripps, "Ultramaneuverable steering control algorithms for terrain transitions," Proceedings of SPIE, Unmanned Ground Vehicle Technology, vol. 3693, 1999.

D. Cripps, "Spatially robust vehicle path tracking using normal error feedback," Proceedings of the 2001 SPIE AeroSense, April 2001.

M. Torrie, P. Lewis, and D. Cripps, "Commercialization of a JAUGS autonomous development vehicle," Proceedings of 2002 SPIE AeroSense, April 2002.

D. Cripps, "Conversion of vehicle position data into a driveable set of path segments," Proceedings of the 2003 IEEE International Symposium on Intelligent Control, October 2003.

Life Experience

December 1986 to August 1988

Working on family farm

<u>December 1978 to December 1980</u>

Missionary Service, Western Samoa

Volunteer Work

Assistant Varsity Scout Master

Tutor at a Washington State Reformatory over a period of three years

Grade School Lunch Buddy/Mentor over a period of three years