Design of Rail Vehicles with Passive and Active Suspensions Using Multidisciplinary Optimization, Multibody Dynamics, and Genetic Algorithms

by

Yuping He

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Mechanical Engineering

Waterloo, Ontario, Canada, 2003

I hereby declare that I am the sole author of this thesis.
I authorize the University of Waterloo to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the University of Waterloo to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Acknowledgements

I am greatly indebted to my supervisors, Dr. John McPhee and Dr. Gordon Andrews, for granting me an opportunity to pursue a doctorate at the University of Waterloo and for their thoughtful advice, continuous support, and patient supervision.

My gratitude goes also to my fellow lab mates: Pengfei Shi, Ivan Bruulsema, Tim Lahey, Chad Schmitke, Scott Redmond, and Yi Liu, who made the time I spent in the lab enjoyable.

I would like to thank Dr. Ronald J. Anderson of Queen's University for providing the A'GEM multibody dynamics software package and for his advice on rail vehicle dynamics. Financial support of this research by the Natural Sciences and Engineering Research Council and Bombardier Inc. is gratefully acknowledged.

I owe my gratitude to my parents for their love, support, and encouragement throughout the course of my studies.

Finally, my thanks go to my wife and son for their understanding of the long hours that I put in the lab every day during my Ph.D. program.

Abstract

A methodology for the design optimization of rail vehicles with passive and active suspensions is presented. The methodology has the following features: (1) multibody dynamics is used for modelling and simulating complex realistic vehicle systems; (2) multidisciplinary optimization (MDO) methods are introduced to make coupled vehicle models and additional control systems a synergistic whole; (3) with genetic algorithms (GAs) and other effective search algorithms, the mechanical and control design variables can be optimized simultaneously; (4) with the scalarization technique, a vector optimization problem is converted into a scalar optimization problem. The proposed methodology is applied to several design optimization problems. First, a rail vehicle is optimized with respect to lateral stability. Second, the rail vehicle is designed so that ride quality is the sole design criterion. Third, the design variables are searched in the feasible design space so as to make the rail vehicle have optimal curving performance. Then, the rail vehicle is optimally designed for obtaining trade-off solutions among conflicting requirements from lateral stability, ride quality, and curving performance. Finally, the methodology is used to optimize the combined mechanical and control systems for vehicles with active suspensions. Of the results obtained, several of them contribute to the fields of rail vehicle dynamics and design, mechatronic systems, and numerical optimization.

For automatically identifying the "critical speed" (above which a rail vehicle's response becomes unstable), a new approach combining sequential quadratic programming (SQP) with the Dynamic Mode Tracking (DMT) technique is proposed and developed. The new approach is compared with that using SQP alone. It is found that without DMT, several more SQP runs are often needed to find the critical speed because the relationship between mode damping and speed deviate from their actual shapes. In the process of optimizing the lateral stability of a rail vehicle model, the existence of sharply-discontinuous "cliffs" in the plots of critical speed versus suspension stiffness is identified and originally interpreted. In recognition of the cliff phenomenon, the definition of critical speed is generalized to make it a more practical measure of lateral stability. In the design optimization of a rail vehicle with respect to the lateral stability, vertical ride quality, and curving performance, the resulting Edgeworth-Pareto (EP) optimal sets clearly demonstrated the trade-off relation between lateral stability and curving performance. Moreover, the resulting EP-optimal

sets visualize a well-known fact that a relatively weak coupling exists between the vertical and lateral motions of a rail vehicle.

To identify effective algorithms for rail vehicle suspension design, the GA, SQP, and Simplex algorithms are compared in the processes of optimizing lateral stability and ride quality. Results show that the reliability of the SQP and Simplex for finding the global optimum decreases with an increase in number of design variables. However, despite non-smooth objective function surfaces with many local optimal points, the GA can reliably find global optima. By means of GAs, important design variables can be identified and the relative significance of design variable sets, e.g. inertial, geometric, and suspension (stiffness and damping) parameter sets, can be decided. When ride quality analysis is performed in the frequency domain based on a linear vehicle model, if SQP is used with a multibody dynamics program, e.g. A'GEM, the numerical differentiation technique for computing gradients can be used efficiently as a link between the multibody program and SQP.

As an application of the proposed methodology, an integrated design approach to mechatronic vehicle systems is used to resolve the conflicting requirements for ride comfort, suspension working spaces, and dynamic wheel loads in the optimization of quarter-vehicle models and half-vehicle models with active suspensions. Both deterministic and random track excitations and both rigid and flexible car body cases are considered. The approach is implemented in a GA-A'GEM-MATLAB simulation environment in such a way that the linear mechanical vehicle models are generated in the multibody program of A'GEM, the controllers and filters are modelled in MATLAB, and the coupled mechanical and control subsystems are optimized simultaneously using the GA. The numerical simulation results are reported and discussed.

Contents

List of Tables

List of Figures

List of Acronyms

A-i-O	All-in-One
DMT	Dynamic Mode Tracking
DOF	Degree Of Freedom
EA	Evolutionary Algorithm
EP	Edgeworth-Parato
GA	Genetic Algorithm
I.B.	In Board
IDF	Individual Discipline Feasible
KT	Kuhn-Tucker
LQG	Linear Quadratic Gaussian
MDO	Multidisciplinary Optimization
NAG	Numerical Algorithms Group
NVH	Noise, Vibration and Harshness
O.B.	Out Board
PSD	Power Spectral Density
QP	Quadratic Programming
R.M.S.	Root Mean Square
SA	Simulated Annealing
SQP	Sequential Quadratic Programming

Chapter 1

Introduction

1.1 Motivation

1.1.1 Multiple Model and Design Criteria Optimization Problem

Because of conflicting requirements for lateral stability, ride comfort, curving performance, track loading, and the economics of manufacturing, designing a rail vehicle is a challenging task to untangle the web of the interactions among these contradictions and to resolve them. Of all contradictory design goals of rail vehicles, two of them, i.e. lateral stability and curving performance, may be the most fundamental and important, which have been bothering rail vehicle designers and researchers since the beginning of the history of rail vehicle dynamics till today [41, 125, 141, 142]. These two contradictory design goals can be defined as follows respectively : (1) rail vehicles should be designed to travel comfortably at high speeds on straight or "tangent" track; (2) rail vehicles should be designed to traverse curved track without excessive noise or wear that arises from misalignment of the wheelsets with the track [8].

To improve compatibility between dynamic stability and curving performance, designers and researchers have used equivalent interaxle shear and bending stiffness relationships to determine optimal suspension characteristics [13, 75, 127]; other researchers have turned to local search numerical optimization algorithms combined with Routh-Hurwitz criteria [29, 32]. Neither approach is well-suited to the multidisciplinary optimization of the com-

plex nonlinear dynamic models now available from multibody dynamics programs such as ADAMS [113] or A'GEM [136], especially when more then a few design parameters and multiple local optima are being considered.

Although multibody dynamics programs such as ADAMS or A'GEM are effective tools for modelling and simulating rail vehicles, from a design point of view, the drawback of these programs is that they only provide analyses of systems whose design variables have been specified. Design optimization, parametric studies, and sensitivity analyses are difficult to perform [87]. Instead, design engineers must decide by trial and error how to change design variable values and repeat the analysis until performance measures satisfy design specifications [97]. This "manual" process is tedious and time-consuming for complex systems with nonlinear performance measures [10].

To find effective trade-off solutions for complicated conflicting design criteria, an effective method is to use various mathematical models, each of which concentrates on a specific aspect of interest, and synthesize the design results based on different objective-oriented vehicle models. Therefore, if the vertical ride quality, lateral stability, and curving performance of a rail vehicle should be optimized simultaneously, it is natural that the corresponding ride quality model, lateral stability model, and curving performance model should be included in the optimization problem.

Based on the above considerations, if the lateral stability, vertical ride quality, and curving performance of a rail vehicle should be optimized simultaneously, if the processes of the selection of appropriate design variables should be automated, and if effective modelling and simulation tools of multibody dynamics software are introduced, one is confronted with the following challenges: which optimization method or methods and which optimization algorithm or algorithms are suitable for the multiple model and design criteria optimization problem? how can we make the various numerical algorithms, multibody dynamics programs, multiple models, and other relevant systems or analysis disciplines a synergistic whole so that the rail vehicle can be optimized effectively?

1.1.2 Lateral Stability Optimization Problem

With the consideration of dimensionality (i.e. the number of design criteria) of the above optimization problem and the complexity and expense of the underlying analysis, it seems

reasonable for us to decompose the above multiple model and design criteria optimization problem into smaller and more manageable subsystems. Accordingly, we are confronted with, at least, three independent problems: lateral stability optimization, vertical ride quality optimization, and curving performance optimization.

Consider first the lateral stability optimization problem. Two approaches have been applied to the problem. First, conventional local search algorithms combined with the Routh-Hurwitz criteria were once used to optimize rail vehicles for maximizing the critical speed (above which a rail vehicle's response becomes unstable) [29, 32]. Recently, Baumal and McPhee [11] replaced the conventional local search algorithms and the Routh-Hurwitz criteria with a genetic algorithm (GA) and a nonlinear programming routine, i.e. sequential quadratic programming (SQP), respectively, and used the resulting approach for optimizing a 7 degree of freedom (DOF) rail vehicle model. With the two approaches to the lateral stability optimization, one faces the following problems. Is it necessary to replace the more computation efficient local search algorithms with the less efficient global search algorithm? Is the SQP valid for identifying the critical speed of rail vehicles when the required system parameters are offered? Moreover, could multibody dynamics programs such as A'GEM be integrated with the existing algorithms or analysis disciplines for optimizing the lateral stability of rail vehicles effectively?

1.1.3 Vertical Ride Quality Optimization Problem

In ground vehicle design, in order to determine optimal suspension characteristics and ride properties in particular, a number of optimization algorithms, both local search and global search algorithms, have been used and investigated. However, to date, very few results on identifying effective optimization algorithms for vehicle ride quality analysis have been published. No literature concentrates its analysis on the following issues: the necessity of global search algorithms for vertical ride quality analysis; preferable options of certain kind of search algorithms (e.g. gradient-guided) to others (e.g. direct search) for different vehicle dynamic models, e.g. symbolic analysis model or numerical model; and the suitability of numerical differentiation for gradient evaluation for linearized vehicle models with the ride quality analysis performed in the frequency domain.

1.1.4 Curving Performance Optimization Problem

For the curving performance optimization problem, conventional practice has been a trial and error method by which designers change design variable values and re-perform the analysis until a set of performance measures becomes acceptable [74, 97]. Therefore, processes that automate the selection of appropriate design variable values can be used to develop invaluable tools for rail vehicle designers.

It seems that no literature reported the application of numerical approaches to the curving performance optimization. There are several factors that hinder the application of numerical optimization to the curving performance problem. First, the dynamic curving behavior of rail vehicles has not been as extensively studied as that of the lateral stability behavior [13]. This phenomenon of insufficient research in curving performance can be interpreted by the fact that large deviations in numerical results of curving behavior simulations arose between software packages, e.g. VAMPIRE, GENSYS, SIMPACK, ADAMS/Rail-MEDYNA, and NUCARS, participated in the Manchester benchmark exercise in 1997 [77]. Notice that the above software packages are commonly used by designers and researchers for predicting dynamic behavior of rail vehicles. The benchmark results showed that for a specified vehicle and track case, which is used for predicting the vehicle behavior in a specified constant radius and superelevation part of track, between the above five packages, the highest deviation in steady-state Lateral/Vertical (L/V) force ratio at the outer wheel on the first wheelset reaches as high as 6.7%, and the highest deviation of the peak value of the L/V ratio reaches as high as 10.3% [77]. Second, it is difficult to formulate a simple performance index to reflect the complicated dynamic behavior of rail vehicles on curves, especially when the wheelset excursion has exceeded the flange clearance [13]. Third, if conventional gradient-guided search algorithms are applied to curving performance optimization problem, the sensitivity analysis could be difficult to perform, especially when nonlinear dynamic curving models are used.

With the above considerations, even if a reliable multibody dynamics program is available for modelling a rail vehicle and simulating the corresponding curving behavior, one is still confronted with the problems of finding a suitable optimization algorithm or algorithms and selecting an effective objective function that reflects the realistic curving performance measures.

1.1.5 Combined Mechanical and Control System Design Problem

Conventionally, a common practice has been a sequential approach, so that a passive mechanical vehicle system is designed first and the active elements of the vehicle suspensions are added subsequently. The resultant design may behave less optimally overall due to not considering simultaneously the mechanical parameters and control parameters as design variables. Since there are inherent couplings among the passive vehicle system, active suspension controller, and sensors, it is expected that an integrated control and mechanical design process may coordinate the different or conflicting requirements from the mechanical system and control system so as to achieve an optimal behavior of the overall vehicle.

It is shown [126] that in the design of a glider, a simultaneous optimization of aerodynamics, structures, and control is more efficient than the corresponding sequential approach. The inadequacy of the sequential approach to the simultaneous optimization motivates the application of multidisciplinary optimization (MDO) to the simultaneous design of a structure and a control system to achieve active flutter suppression.

With the above example, it seems natural to expect that multidisciplinary optimization may offer a promising approach to the simultaneous design of active elements and passive components in the design optimization of vehicles with active suspensions.

1.2 Research Objectives

The primary goal of the research is to develop a methodology for the design optimization of rail vehicles with passive and active suspensions. The framework of the methodology consists of the following main components: application of multidisciplinary optimization to the simultaneous manipulation of design variables for several systems or analysis disciplines between which there are strong interactions; introduction of multibody dynamics for effective modelling and simulation of rail vehicles under different operating conditions; use of genetic algorithms for reliable global optimal solutions; development of an approach for automatically identifying the critical speed of rail vehicles; and selection of strategies for establishing effective objective functions for facilitating the complex optimization problem.

The following steps are taken to develop the proposed methodology:

1. Optimizing a rail vehicle with respect to lateral stability;
2. Optimizing rail vehicles with respect to vertical ride quality;
3. Optimizing a rail vehicle with respect to curving performance;
4. Optimizing a rail vehicle with respect to lateral stability, curving performance and vertical ride quality, simultaneously;
5. Simultaneously optimizing mechanical system and control system of mechatronic ground vehicles with respect to ride comfort, dynamic wheel loads, and suspension working spaces.

In the process of developing the proposed methodology, several results are obtained. Of these results, listed below are those that contribute to the fields of rail vehicle dynamics and design, mechatronic systems, and numerical optimization:

- The existence of sharply-discontinuous "cliffs" in the plots of critical speed versus suspension stiffness is identified and originally interpreted using modal analysis techniques. This sharp discontinuity in the critical speed occurs when the mode (eigenvector) determining the critical speed switches. In recognition of the cliff phenomenon, the definition of critical speed is generalized to make it a more practical measure of lateral stability.

- An effective approach combining SQP and a dynamic mode tracking (DMT) technique [7] is developed for identifying the critical speed automatically. Since the governing equations for rail vehicles are similar to those in rotor dynamics, wind turbine dynamics, aeronautics, and road vehicle dynamics, the approach can also be applied to these problems for identifying the corresponding stability criterion automatically.

- When SQP is combined with a numerical multibody dynamics software for optimizing vehicle suspensions using linear vehicle models for ride quality analysis in the frequency domain, it is found that numerical differentiation techniques can be used for evaluating gradient reliably.

- It is discovered that with the co-existence of passive and active components in vehicle suspensions and the design variables determined by using the optimization method proposed in the research, the corresponding actuator forces can actively resist track disturbances much longer than the actuator forces in active suspensions that have no passive elements and for which the design variables are determined by using the linear quadratic Gaussian (LQG) method.

- Compared with inertial and suspension parameters sets, geometric parameters have the most significant effect on lateral stability, curving performance, and vertical ride quality.

- A clear picture of the trade-off relationship between the lateral stability and curving performance of a rail vehicle is offered. Moreover, the trade-off relationship between the vertical ride quality and lateral stability, and between the vertical ride quality and curving performance of the vehicle confirm a well-known fact that a relatively weak coupling exists between the vertical and lateral motions of a rail vehicle.

- Compared with a vehicle with passive suspensions and the vehicle with active suspensions based on the LQG, the optimized vehicle system based on the approach proposed has the best overall performance in all three aspects including ride comfort, suspension working spaces, and wheel dynamic loads. However, traditional active suspensions often achieve better performance that is a compromise among these three aspects and they can rarely improve the vehicle performance in all the three aspects simultaneously.

1.3 Thesis Organization

The thesis is composed of nine chapters. In Chapter 2, a brief literature review is offered on the issues: vehicle system modelling, optimization methods and algorithms for vehicle suspension design, and design optimization of rail vehicles with passive and active suspensions.

Chapter 3 describes both the All-in-One (A-i-O) and Individual Discipline Feasible (IDF) approaches for MDO problems. Then, the essential concepts of multicriteria opti-

mization are outlined. Finally, three typical optimization algorithms, i.e. GA, SQP, and Simplex, which are extensively used and investigated in the research, are briefly reviewed.

In Chapter 4, several steps are taken to demonstrate an optimization approach integrating multibody dynamics, a GA or Simplex, SQP, and DMT for optimizing the lateral stability of a rail vehicle. First, a hand-derived solution to a 17 DOF linear rail vehicle model is compared to the results from an A'GEM simulation. Second, the algorithm of SQP is compared with a combined algorithm including the SQP and DMT when the two algorithms are used for identifying the critical speed of the rail vehicle in two different cases. The critical speed is optimized using the GA and Simplex algorithms, and the GA is compared against the Simplex algorithm. In the process, the existence of sharply-discontinuous "cliffs" in the plots of critical speed versus suspension stiffnesses is identified. In recognition of the cliff phenomenon, the definition of critical speed is generalized to make it a more practical measure of lateral stability. Finally, the integrated approach is used to identify the relative significance of different design parameter sets, i.e. geometric, inertial property, and suspension parameters, on the critical speed of the rail vehicle.

In Chapter 5, to identify effective optimization algorithms for vehicle suspension design, the GA, SQP, and Simplex are compared by means of the vertical ride quality analysis for a 2 DOF quarter-vehicle model and a 20 DOF rail vehicle model. To further illustrate the effectiveness of using the GA for optimizing vehicle suspensions, results are reported for analyzing the relative significance of different design variable sets on vertical ride quality using the 20 DOF rail vehicle model and optimizing a rail vehicle with active elements using a 36 DOF model with car body flexibility. In addition, numerical differentiation is validated as an effective link between the multibody dynamics package of A'GEM and SQP for evaluating gradients.

Chapter 6 presents the feasibility and efficacy of applying numerical optimization approaches to a rail vehicle design with curving performance considered. The numerical results of the optimization based on a 21 DOF dynamic curving model are offered and discussed. Once again, the proposed optimization approach is applied to investigate the relative significance of different design parameters or different parameter sets on curving performance.

Chapter 7 demonstrates the effectiveness of a hybrid MDO approach combining the A-

i-O and IDF methods for handling the conflicting requirements from the lateral stability, curving performance, and vertical ride quality in the design of a rail vehicle with passive and active suspensions. This hybrid MDO method integrates a GA, SQP, DMT, and multibody dynamics modelling and simulation programs from A'GEM so that the three complex dynamic rail vehicle models, i.e. the 17 DOF lateral stability model discussed in Chapter 4, the 36 DOF vertical ride quality model employed in Chapter 5, and the 21 DOF nonlinear dynamic curving model described in Chapter 6, are accommodated in a synergistic system. Within the system, the GA coordinates the above mentioned conflicting requirements in system level and the suspension, geometric, inertial, and control parameters for the rail vehicle described by the three models are optimized simultaneously. Out of a family of Edgeworth-Pareto optimal solutions obtained from the optimization, the designer can make his final decision on an acceptable design.

Chapter 8 illustrates the effectiveness of the A-i-O method for resolving the conflicting requirements for ride comfort, suspension work spaces, and dynamic wheel loads in the optimization of quarter-vehicle models and half-vehicle models with active suspensions. Both deterministic and random track excitations and both rigid and flexible carbody (for the half-vehicle models) cases are considered. The optimization problem is implemented in a GA-A'GEM-MATLAB simulation environment in such a way that the linear mechanical vehicle models are generated in A'GEM, the controllers and Kalman filters are modelled in MATLAB, and the coupled mechanical and control subsystems are optimized simultaneously using the GA. Numerical results are offered.

Finally, Chapter 9 is dedicated to conclusions drawn from the research and recommendations for future research.

Chapter 2

Literature Review

2.1 Introduction

An integrated design optimization approach for engineering systems generally involves two parts [17, 19, 42]: modelling and optimization. The first part, modelling, can be further divided into three subparts: modelling, parameterization, and criterion definition. To optimize an engineering system such as a rail vehicle, one should first transform the system to a mathematical model. To establish the mathematical model, the finite element method and the multibody dynamics method are commonly used. In applications to vehicle system design optimization, the finite element method is often used for structural optimization, e.g. automotive crashworthiness. The method of multibody dynamics is a well-accepted and widely used method for analyzing the dynamic behavior of vehicle systems in the process of suspension design [114, 115]. With a suitable mathematical model, one is then confronted with the definition of design criteria and design parameters of the model. Since engineering requirements and designers' wishes are sometimes hard to be formulated as mathematical functions, the design criteria are often difficult to be defined. In general, the design of a dynamic system with respect to several specifications leads to a multicriteria optimization problem [17]. Parameters of the model are often classified either as design variables, whose values can be chosen within given bounds, or as system constants whose values are fixed during optimization. Constraints have to be considered for the modelling, design criteria and parameterization chosen. Optimization constraints can be formulated

in the design variable space or in the design criteria space, depending on what is considered to be more convenient. Since the generation of reliable mathematical models is vital to design optimization problems [42, 95], the state-of-the-art related to modelling techniques and principles for model selection, especially for rail vehicle suspension design, will be further addressed in the chapter.

After the above preparation, one faces the tasks of formulating the optimization problem and selecting appropriate optimization algorithms. Notice that in this thesis, the term "optimization method" refers to the method of optimization problem formulation. Traditionally, applications to dynamic system design are often restricted to single design criterion optimization problems. These optimization problems can be classified either as a constrained optimization problem or as an unconstrained problem. Generally a single design criterion optimization problem can be posed in a standard form [56]. With a suitable search algorithm, the solution to the problem can be found. In the case of a multicriteria optimization problem, based on concepts such as scalarization or hierarchization, it is possible to transform the multicriteria optimization problem to a single (or a sequence of) scalar, nonlinear programming problem [42]. For a complicated optimization problem in which strong interaction between systems or disciplines arises, multidisciplinary optimization proved to be effective for coordinating the design considerations at the system and subsystem levels [123]. Obviously, it is vital to select an appropriate optimization method especially for complex multicriteria design optimization problems. Moreover, there are numerous algorithms available and even the choice of an efficient optimization algorithm is a nontrivial problem [20]. Therefore, a survey of the applications of typical optimization methods and algorithms to the design of vehicle suspensions is offered in this chapter.

In the past 2-3 decades, design optimization approaches, an important tool for the synthesis of complex mechanical systems [42], have been widely applied to the design of vehicle suspensions, especially for road vehicles. The state-of-the-art techniques (e.g. multibody dynamics programs) and search algorithms (e.g. genetic algorithms) have been introduced into the design process. In recent years, design philosophies for a mechatronic system have been developed to a stage where the actuators, the controllers, the sensors, as well as the mechanical structure design of the mechatronic system are considered simultaneously [123]. These design philosophies may be helpful to the design of vehicle suspensions. Hence, a

review of the mainstream developments in the study on the design optimization of rail vehicles with passive and active suspensions is presented in this chapter.

2.2 Vehicle System Modelling

2.2.1 Multibody Dynamics Modelling Techniques

Conventionally, in the case of creating dynamic models for rail vehicles, after the routine for including the wheel/rail forces in the equations of motion has been developed and computers were available to solve large numbers of simultaneous equations, the task was to derive the multi-degree of freedom equations and transform them into computer codes manually [4]. To establish complex vehicle models, this manual process has proven to be a very tedious, difficult, time-consuming, and error-prone task [8, 86]. Moreover, it was illustrated that these hand-derived models, especially when embedded in computer programs, can hardly be changed [117]. When a model was found not to contain an essential feature of a new task, e.g. introducing active elements into a suspension, it was necessary to create a new model. In the early 1970's, the foundations of ADAMS and DADS, today's most used multibody dynamics codes, were laid by the work of Orlandea [108]. In the early 1980's, Anderson applied multibody dynamics approach to dynamic modelling of rail vehicles [4]. With the introduction of multibody dynamics approach to various fields including rail vehicle dynamics in the 1980's and 1990's, this approach has been employed by researchers with significant favor for the purpose of improving their mathematical models of both road and rail vehicles. By means of a multibody formulation, the equations of motion for a complex system of rigid and flexible bodies connected by kinematic and dynamic components may be generated automatically. Many successful commercial programs, e.g. ADAMS, SIMPACK, and VAMPIRE, were applied to the dynamic analysis of several rail vehicle benchmark problems and gained varying levels of success [77]. Since these computer packages have been validated against the results from many experiments and benchmarks, a precise simulation of the dynamic response of a vehicle can be obtained as long as an accurate set of vehicle characteristics are offered to these programs.

With these multibody dynamics programs, analysts can develop very complex vehicle models with consideration of nonlinear suspension components, wheel/rail forces and geometry for assessing various aspects of rail vehicle dynamic behavior. For example, the multibody dynamics software A'GEM developed by Anderson [6] contains a set of modules. To list a few, these modules provide the following functions: generating equations of motions for linear dynamic systems with constraints (A'GEM module); simulating nonlinear curving behavior of arbitrary rail vehicles (RACES module); performing linear lateral stability analysis for rail vehicles (STABLE module); conducting dynamic mode tracking for rail vehicle models (MTRACK module); and calculating the ride quality for arbitrary rail vehicles (RLRIDE module).

However, from a design point of view, the drawback of multibody dynamic programs is that they only provide analyses of vehicle systems whose design variables have been specified [10]. Design optimization, parametric studies and sensitivity analyses are difficult, if not impossible to perform [87]. Integrating multibody dynamics programs with numerical optimization methods and algorithms will automate and facilitate the design optimization of rail vehicles.

2.2.2 Vehicle System Models

The complexity of the mathematical models chosen to represent a vehicle is closely related to the specific problem which has to be resolved with the model. In many cases, simple linear models or models with a few DOF are adequate for obtaining the desired information. Examples are the "quarter-vehicle model" for studying the vertical (bounce) mode or the "half-vehicle model" for studying the bounce and pitch modes in many suspension design problems [44]. In other cases, rail vehicle researchers resorted to using only single bogie (see Figure 4.1 and § 4.2 for the definition of bogie) or half-body models and accepted the inaccuracy in exchange for gains in efficiency [8]. Langlois et al [91] found that most controllers were based on quarter-vehicle models because of the extra mathematical complexity resulting from the use of half- or full-vehicle models. Applied to a full-vehicle model, such controllers may behave less than optimally because no provision for roll or pitch motions can be made in the quarter-vehicle. Because of this reason, in recent years, researchers have developed their controllers based on full-vehicle models [26, 34, 45, 50, 90].

The increasing demands on the accuracy of mechanical models of vehicles result in more complex multibody systems. The history of research in rail vehicle curving performance reflects these demands. To evaluate curving performance of rail vehicles, in the 1960's, Newland and Boocock [22, 105] developed a linear steady-state curving model. To calculate the steady-state solution in a constant radius curve and with a single wheel/rail contact point, in the 1970's, Elkins and Gostling [46] used the so-called quasi-static model which includes nonlinearities of the wheel/rail geometry and forces. To study the complete curving behavior from the tangent track, through the transition spiral, and into the constant radius curve, in the 1980's, Fortin [53] developed dynamic curving models that take nonlinear wheel/rail forces and geometry, suspension nonlinearities, and two wheel/rail point contact into account. Today the multibody dynamics program A'GEM can be easily used to automatically develop 21 DOF and 37 DOF dynamic curving models to assess the curving behavior for a conventional rail vehicle and a forced-steering rail vehicle, respectively [136].

An all-inclusive dynamic model of a rail vehicle would be large and complex [40]. Since size and complexity are factors that tend to reduce physical insight and increase the development and usage costs of computer programs, less general models are usually preferable. In fact the best model is not the most complex one, but the most appropriate one. Vehicle dynamic applications also show that analysis of different aspects of a vehicle system has to be based on different models [18]. According to the operating conditions, type of terrain, and wheel/rail profiles, different objective-oriented models are developed and used. These specialized models may be classified as follows:

1. Vertical/Lateral Dynamic Models. These models are designed to study the dynamic response of a vehicle to track irregularities.

2. Curving Performance Models. These models are used to calculate the dynamic or quasi-static forces of a vehicle negotiating curves.

3. Lateral Stability Models. These models are applied to the prediction of the critical speed, wheel/rail forces, and suspension and car body forces and displacements.

Hedrick et al [74] used a 12 DOF vertical dynamic model, a 15 DOF lateral dynamic model, a 15 DOF lateral stability model, and a 15 DOF steady-state curving performance

model to evaluate the performance of a conventional rail vehicle. All these objective-oriented models were developed manually. Supported by the multibody dynamics package A'GEM, MacNaughton [97] employed a 17 DOF lateral stability model, a 20 DOF vertical ride quality model, and a 21 DOF nonlinear dynamic curving performance model to conduct a design modification for an urban transit rail vehicle. The above applications illustrate that specialized vehicle models, instead of an all-inclusive dynamic model are commonly used in designing rail vehicle or evaluating rail vehicle performance.

2.2.3 Track Models

In most cases, the tracks are modelled as rigid, smooth, and straight for a lateral stability analysis. In other cases, for example when the curving behavior of rail vehicles is evaluated and studied, the curving of rail vehicles requires the track geometry such as curves and grades. Parameters to describe these profiles are analytic functions specifying the curvature, cross level, gauge etc [86].

From the ride quality analysis point of view, track unevenness represents the main excitation function to both the passenger and the vehicle structure. Thus, track descriptions have to include irregularities, consisting of either deterministic functions or stochastic functions. Many track profiles have been measured, some being recorded on magnetic tape (particularly test tracks) while most have been Fourier analyzed and characterized by their frequency domain properties [83, 119]. It is important to try to predict vehicle dynamic response using a realistic track irregularity model; unfortunately, a completely realistic track model is difficult to develop. The track models are usually expressed in terms of power spectral density (PSD) function of track profile elevation or "slope". All of the track roughness models have been developed under substantial assumptions, which typically idealize the track profile as a stochastic process and exclude all the singular events. Generally, track irregularities result in displacement inputs to vehicle models. When vehicles have multiple axles, these track irregularities disturb the vehicles in a correlated fashion: the same input occurs at the following axle with a time delay as a function of vehicle speed and wheelbase. These correlations cause significant deviations from cases in which both deterministic and random inputs are assumed to be uncorrelated [34]. In the cases where the unsprung masses of vehicles are modelled with damping, not only the irregularities at

a displacement level but also at a velocity level must be considered.

The contact forces between wheel and rail play a dominant role for the computer simulation of rail vehicle running behavior such as lateral stability and curving [86]. The contact problem can be divided into a purely geometrical problem, i.e. the rolling of a wheel on a rail, and the kinetic problem of the modelling of the contact forces, i.e. the creep force laws.

To represent wheel/rail contact geometry, it is assumed that wheels and rails are rigid bodies and the velocities in the common points of contact can be determined from purely geometrical relationships taking into account the specified contact profiles. As a result, nonlinear constraint functions arise which algebraically describe the dependence within the vehicle coordinates such as lateral displacements, roll and yaw angles etc. Law, Cooperrider, Heller and other have done extensive work in this area and have published reports and software which permits the detailed wheel/rail contact geometry representation that is required for the accurate calculation of wheel/rail forces [53].

To calculate the creep forces arising at wheel/rail contact points, wheels and rails are considered to be elastic halfspaces and the contact patch is assumed to be an ellipse. The modelling and computation of the creep forces due to these assumptions have been developed by Kalker [78] resulting in a number of software algorithms and simplications. Kalker's accurate and practical models of creep forces are used almost exclusively in modern simulations of rail vehicle dynamics [8].

2.3 Optimization Methods and Algorithms Used in Vehicle Suspension Designs

2.3.1 Optimization Methods

In general, the design optimization of aircraft and road vehicles as well as rail vehicles, is multidisciplinary. For example, the simultaneous design of a structure and a control system for the purpose of active flutter suppression for an aircraft is a typical application of multidisciplinary optimization [123]. In the case concerned, there are interactions among the wing structure, the control system, and aerodynamics. These interactions make the

structure, the control system, and aerodynamics a synergistic whole that is greater than the sum of the three. Taking advantage of that synergy is the mark of a good design [85]. In fact, Multidisciplinary Optimization is presently of increasing interest in engineering. MDO received recognition in the aeronautical sciences, first for the structural optimization and later for the aerodynamic design [123]. Currently, we can find the application of MDO to automotive vehicle design for safety and NVH (noise, vibration and harshness) reduction [85, 124, 145, 146]. MDO is also used for ground vehicle suspension design [66, 67].

The current literature shows that most MDO problems of vehicle systems for safety, NVH, and ride comfort are multidimensional (more than a few design variables) and multicriteria (more than a few design criteria) [66, 67, 85, 124]. In these applications, the relevant design criteria are optimized simultaneously. In these cases, it is unlikely that the same values of design variables will result in the best optimal values for all the design criteria. Hence, some trade–off between the design criteria is needed to ensure a satisfactory design. In fact, the solution of a MDO problem is a compromise between disciplines. Moreover, by means of using multicriteria optimization techniques, we can integrate different disciplines into the united computational process.

Most MDO approaches are based on the response surface technique and on the original approximation concept [123]. Due to the framework, for a given application of MDO to the design of vehicle suspension systems, the selection of design space search algorithms is vital. Many algorithms for design space search have been developed, and they differ according to the type of criteria (linear/nonlinear), constraints (unconstrained/bounds/equality/inequality), their approximation type and their demand for information (criteria/ gradients/Hessians) [42].

MDO Methods

One of the primary challenges in MDO is organizational complexity [123]. A successful vehicle system design requires harmonization of a number of criteria and constraints. Such a design problem can be modeled as a constrained optimization in the design variable space. However, for such optimization, due to its dimensionality, complexity, and expense for analysis, a decomposition approach is recommended so as to enable concurrent execution of smaller and more manageable tasks [85]. To preserve the couplings that naturally

occur among the subsystems of the whole problem, such optimization by various types of decomposition must include a degree of coordination at the system and subsystem levels. MDO offers effective methods for performing the above optimization so as to resolve the trade-off relations among the various design criteria at the system and subsystem levels.

Several MDO methods exist, including All-in-One (A-i-O) method [85], Individual Discipline Feasible (IDF) method [33], Collaborative Optimization(CO) method [23], Bi-Level Integrated System Synthesis (BLISS) method [84], and Concurrent SubSpace Optimization (CSSO) method [111], to name a few. Among most of these MDO methods, the shared character is that the system concerned is decomposed into subsystems so that the corresponding subtasks are performed independently in their own modules; then in a system level, the coordination of the different design considerations gives rise to a two-level optimization. One of the most important advantages of this decomposition is the concurrent execution of the subtasks, which is well suited for parallel computations.

Another primary challenge in MDO is computational expense. The computational challenge may simply reflect increased dimensionality, with analysis and design variables that accumulate from all disciplines. Effective methods for improving calculation efficiency are, for example, to use parallel computations and approximation concepts [123]. A discussion on this topic is beyond the scope of this thesis.

Multicriteria Optimization Methods

Conventionally, when optimization methods are used in vehicle system design, the applications are often restricted to single criterion optimization [17]. Designs derived from the optimization method are not fully satisfying for practical applications, because they consider only a single aspect out of a couple of conflicting system requirements and specifications. In most cases, such 'optimal' designs are even inferior to those derived by common sense. Thus, several conflicting design specifications and goals have to be taken into account in practical applications. The multicriteria optimization methods seem to offer a promising way to handle the design problems with conflicting specifications and requirements and find optimal designs [18].

Using multicriteria optimization methods, we can minimize several criteria simultaneously, and define optimal compromise solutions when conflicting criteria co-exist in the

design problems. Generally, such optimal designs are a family of designs that are not comparable to each other. With the consideration of design intuition and additional information, a final solution can then be determined.

A well-known technique used for multicriteria optimization is to reduce a multiple design criteria or vector optimization problem to a single design criterion or scalar optimization problem [18, 42]. This topic will be further discussed in the next chapter.

2.3.2 Optimization Algorithms

Optimization algorithms can be classified into three main types [42, 49, 112]: "hill-climbing", enumerative, and stochastic.

"Hill-climbing" algorithms may further divide into two subclasses: direct search (e.g. Simplex) and indirect search (e.g. SQP). Direct search algorithms are typically based on function comparison techniques. Most such procedures are heuristic in nature and derivative evaluations are not needed. Generally these tend to be slow. The indirect search algorithms solve a non-linear set of equations resulting from setting the derivative of the objective function equal to zero and finding the local optima. They are well-known to exploit all local information in an efficient way, provided that certain conditions are fulfilled and, in particular, that the function to be minimized is well-conditioned in the neighborhood of the unique optimum. These algorithms require a lot of local information to be known, e.g. the gradient and the Hessian matrix. If the basic requirements are not satisfied, the reliability of the methods is greatly jeopardized. Most often, for indirect search algorithms, the restrictive requirements of continuity and differentiability should be satisfied.

Enumerative algorithms are straightforward search algorithms. They evaluate objective fitness in the feasible search space of design variables one point by one point. The lack of efficiency of these algorithms prohibits them from a wide range of applications.

Recently, stochastic algorithms (e.g. GAs, Evolutionary Algorithms (EAs) and Simulated Annealing (SA) Algorithms) have achieved popularity mainly for these properties: performing global optimizations, requiring no or very low accuracy gradient information, using probability rules to guide their searches, and being suitable for solving complex real-world problems. Stochastic algorithms achieve the above merits at the expense of requiring

more function evaluations compared with conventional optimization algorithms.

In ground vehicle design, a number of optimization algorithms have been used to determine optimal suspension characteristics. A modified Simplex was used to optimize the characteristics of the elasto-damping elements for a simple 4 DOF model of a passenger car [37]. The constrained optimization algorithm of M. J. Box (Complex) has been applied to maximize the critical speed for a rail vehicle (based on a 15 DOF model) subject to constraints on lateral ride quality and suspension working spaces [32]. A modified version of the Hook-Jeeves method was employed for seeking the optimization of the characteristics of the elasto-damping elements of automobiles from the aspect not only of ride properties, but also of handling [38]. A SQP has been chosen for the purpose of optimizing a tractor/semi-trailer suspension [14]. In recent years, GAs [12, 36, 49, 66], SA [42], and EAs [93] lengthen the list of optimization algorithms for ground vehicle suspension design.

Among the above mentioned algorithms, Simplex, Complex, and Hook-Jeeves are direct search algorithms and Complex is essentially based on Simplex. These algorithms, together with SQP, belong to the family of local search algorithms unlike the global search algorithms such as GAs, EAs, and SA. It is well-known that the local search algorithms converge to the nearest optimum, since they depend upon the starting values of the design variables. Finding the global optimum is a great problem of these optimization algorithms, and it is solved by using a larger number of combinations of the initial values of the design variables [38]. Furthermore, when the local search algorithms, e.g., SQP, are used for solving multicriteria optimization problems, they usually find only a single point of the whole set of trade-off solutions of the multicriteria design problem. However, for the global search algorithms such as SA, in addition to their high reliability for finding the global optimal point in the design variable space, they can be used to present, a much clearer picture of the trade–off solution set and a better feeling for sensitive or insensitive design variables [42].

2.4 Design Optimization of Rail Vehicles with Passive and Active Suspensions

2.4.1 Optimization of the Lateral Stability

It has been demonstrated that a system subject to non-conservative forces may become unstable under certain conditions. For rail vehicles with conical steel wheels running on straight steel rails, the non-conservative forces arise due to creepage at the contact point between the wheel and rail. In rail vehicle dynamics, based on linear stability analysis, one of the important stability criteria is a forward speed known as "critical speed" above which a rail vehicle becomes unstable [139, 140]. This unstable behavior, also known as "hunting", exhibits increasing frequency and decreasing damping with increasing speed [7]. The linear stability analysis has become state-of-the-art [82] and it provides very useful design information that has not previously been available [8]. At present, the linear stability analysis is applied in industry for newly designed vehicles and the method is integrated in computer programs for multibody systems such as A'GEM, ADAMS, MEDYNA, NUCARS, SIMPACK, VAMPIRE, and VOCO [82].

Development of a rail vehicle that can operate in the 160 to 480 [km/hr] speed regime must avoid the serious problem of hunting [32]. An effective way to do this is to use numerical optimization to determine a set of suspension parameters that maximize the critical speed [11, 29, 32, 66, 69, 72].

Interestingly, the physical basis of wheel/rail and tyre/road rolling contact mechanics are to a great extent the same [82]. This similarity is reflected in the existence of asymmetric matrices in the governing equations of both rail vehicles and road vehicles. Corresponding to the hunting phenomena for rail vehicle wheelsets, there exist the shimmy phenomena for road vehicle steering systems. Moreover, similar asymmetric matrices are found in rotor dynamics, in wind turbine dynamics, and in aeronautics. Therefore, the numerical optimization approaches to maximizing the critical speed for rail vehicles are expected to be applicable to these fields.

Cooperider, Hedrick and Cox [29, 32] have optimized the critical speed of a 3 DOF rail vehicle model using an unconstrained optimization method called the Hooke-Jeeves

algorithm. They also used a constrained optimization method called the M. J. Box or Complex algorithm to optimize the critical speed of a 15 DOF model. In both cases, they employed a two-loop algorithm. In the interior loop, the critical speed is obtained from a stability analysis of the equations of motion for a given set of suspension and wheel parameters. The Hessenberg algorithm is used to find the system characteristic equation, to which the Routh-Hurwitz criteria is applied to determine stability for a given speed. This process is repeated for different speeds until the critical speed is found. In the outer loop, the Hooke-Jeeves or Complex algorithm is used to find the design variables that maximize the critical speed.

In fact, as early as in 1928, Carter [82, 140] used the Routh stability criteria to investigate the stability problem of various configurations of locomotives, identifying the critical speeds that separate regions of stable and unstable motions. In 1957, Matsudaira also used the Routh-Hurwitz criteria to determine the critical speed of a two-axle vehicle. Obviously the Routh-Hurwitz criteria is important in the linear stability analysis. However, the value of the Routh-Hurwitz criteria is diminished if the characteristic equation can be solved using a root-finding routine on the computer [88]. Numerical experiments showed that for a given set of suspension and wheel parameters and a known value of vehicle forward speed, the eigenvalues of the rail vehicle system matrix in state space form can be calculated with ease using Householder's transformation method and the QR algorithm [11, 66, 69]. Thus, for the problem of optimization of the lateral stability of rail vehicles, when the system design variables and vehicle speed are given, the Hessenberg algorithm and the Routh-Hurwitz criteria used by Cooperider, Hedrick and Cox [29, 32] may not be an efficient method for identifying the critical speed automatically.

Both the above mentioned Hooke-Jeeves and Complex algorithms can be classified as local search algorithms because they use only local information to find a better solution. When they converge to a stationary point, there is no guarantee that this is in fact the global optimum.

Baumal and McPhee have reported that a combination of a genetic algorithm (GA) and sequential quadratic programming (SQP) is well-suited to the design optimization of a simple rail vehicle model [11]. Essentially, the GA and SQP are the counterparts of the M. J. Box/Hooke-Jeeves algorithm and the Routh-Hurwitz criteria in the optimization

methods described in [32, 29]. Eleven design variables, including suspension stiffness and damping, geometric parameters, and inertial property parameters, are optimized [11]. It is shown that this combined approach can find the global optimum with a high reliability. The simple 7 DOF model offers a good qualitative understanding of rail vehicle performance trends, but it is not sufficient for quantitative design use.

The combined approach has been extended by introducing multibody dynamics programs and is validated using a 17 DOF rail vehicle model [68, 69].

2.4.2 Optimization of Vertical Ride Quality

Although a number of optimization algorithms have been used to determine optimal suspension characteristics, with particular emphasis on improving ride quality, just a few researchers tackled the task of identifying effective optimization algorithms for vehicle ride quality analysis. Li et al. [93] compared the simulation results based on evolutionary algorithms (EAs) with those based on GAs using a simple 2 DOF quarter-vehicle model for ride quality analysis. Baumal et al. [12] applied GAs to the design optimization of a vehicle suspension system, emphasizing on optimizing ride quality using a 5 DOF half-vehicle model; they also compared the results based on GAs with those based on a gradient-guided algorithm. Eberhard et al. [42] addressed some advantages of stochastic optimization algorithms such as SAs over gradient-guided algorithms such as SQP in multicriteria optimization of a 11 DOF vehicle model for ride quality improvement.

Bestle [15] proposed an approach combining multibody dynamics and a nonlinear programming optimization algorithm for the design optimization of automotive systems. He clearly addressed that the missing link between multibody dynamics codes (for analyzing the dynamic behavior) and optimization codes is the sensitivity analysis of objective and constraint functions with respect to parameter perturbance. In principal, the sensitivity analysis can be performed using three different methods: numerical differentiation, the direct differentiation method, and the adjoint variable method. Bestle investigated advantages and drawbacks of these methods and obtained a helpful result. The result illustrated that the adjoint variable method is more reliable and efficient than the often used numerical differentiation for gradient evaluation when the vehicle ride quality analysis is performed in the time domain and the vehicle model is nonlinear.

2.4.3 Optimization of Curving Performance

Rail vehicles should possess satisfactory dynamic behavior in curves so as to ensure safety, minimize wear and noise, and provide adequate comfort for the passengers. To understand the behavior of rail vehicles in curves, the early studies by Newland [105] and Boocock [22] developed linear steady-state curving analysis. To predict curving performance more realistically, nonlinear wheel/rail forces and geometry were introduced into steady-state curving analysis [31, 46]. To study the complete curving behavior from a tangent track, through a transition spiral, and into a constant radius curve, and to include all suspension nonlinearities, a dynamic curving model was used [53]. Currently, multibody dynamics modelling software, e.g. A'GEM and VAMPIRE [51], are available for evaluating curving performance of rail vehicles. To provide validation of software packages on simulation of rail vehicle dynamics including curving behavior, a benchmark exercise was carried out in Manchester in 1997 [77].

Although current computational power and multibody dynamics modelling software on rail vehicle dynamics makes it possible to perform the extensive synthesis of complex rail vehicle dynamic systems, few researchers have used numerical methods to optimize rail vehicle systems with respect to curving performance.

Hedrick et al. [74] evaluated and optimized a conventional rail vehicle on mainline intercity curves using a steady state curving model with 15 DOF. This curving model was mainly based on the Newland/Boocock approach to linear, creep guidance curving [22, 105] where the performance was measured in terms of the curve radius and speed for which the wheelsets did not "flange" (the wheelset excursions did not exceed a specified flange clearance) or "slip" (the resultant creep force on each wheel did not exceed a friction force limit). This steady-state curving model was cast in linear algebraic equations in matrix format. With given design variables and required system parameters, these equations were solved by the Gaussian elimination method. If a design variable or a set of design variables are allowed to vary within a specified range, these linear algebraic equations are solved repeatedly to obtain the relevant curving performance curves or surfaces. By comparing the resultant curves or surfaces with the relevant wheel slip and flange contact boundary curves or surfaces, one may evaluate a specified design or select an appropriate design variable or variables as an optimal design.

MacNaughton [97] performed a design modification for an urban transit rail vehicle using a nonlinear dynamic model with 21 DOF for curving performance analysis. In the model, the nonlinear creep forces and the nonlinear wheel/rail geometry with two points of contact were taken into account. In predicting the curving performance of the vehicle, the RACES program of A'GEM was used. The ratios of lateral to vertical (L/V) contact forces and the angle of attack (the angle between the axle of a wheelset and the corresponding radius of curve) were selected as design criteria for curving performance design. With the assistance of necessary dynamic analysis and the consideration of the design specifications on lateral stability, ride quality, and vertical dynamics, the RACES routine was run repeatedly until the selected design variables made the resultant curving performance satisfy the design criteria.

Both the above design approaches used by Hedrick and MacNaughton share common characteristics: the number of design variables is small and the designer must decide by trial and error how to change variable values and re-perform the analysis until a set of performance measures becomes acceptable.

2.4.4 Optimization of Lateral Stability, Ride Quality, and Curving Performance

To investigate lateral stability/curving performance compatibility for rail vehicles, Wickens [141] proposed a generalized set of stiffnesses, i.e. interaxle shear and bending stiffnesses, that encompass both conventional and radial bogies. These elastic characteristics of bogies have been used to select appropriate design variable values for improving the compatibility of lateral stability/curving performance [13, 75, 122, 127]. However, this elastic characterization is valid only under the assumptions: (1) the bogie is freely pivoted to a car body [122]; (2) the bogie model for dynamic stability and curve negotiation is linearized [141]; (3) the dynamic effects due to bogie frame inertia and primary suspension damping are neglected [13]. Cox, Hedrick, and Cooperrider [29, 32] have turned to local search numerical optimization algorithms combined with Routh-Hurwitz criteria for optimizing rail vehicles with respect to lateral stability and ride quality in lateral direction using a 3 DOF and a 15 DOF rail vehicle model, respectively. Neither approach is well-suited to

the design optimization of the complex nonlinear models now available from multibody dynamics programs such as ADAMS or A'GEM, especially when more than a few design parameters and the existence of many local optimal points are being considered.

Bestle and Eberhard proposed the "multidisciplinary multimodel design concept" for the design optimization of road vehicles with respect to ride comfort and ride safety [18]. Different aspects of design criteria were based on different models. This multimodel design problem was implemented using multicriteria design methods.

To optimize a rail vehicle with respect to lateral stability, vertical ride quality, and curving performance, He and McPhee have proposed a combined approach using genetic algorithms and multibody dynamics [66]. Lateral stability, curving performance, and ride quality were assessed using realistic multibody models from A'GEM, and combined in a multicriteria objective function. By combining a GA with A'GEM, the suspension stiffness and damping, geometric, inertial, and control parameters for the rail vehicle with active suspensions were optimized simultaneously. This research has been extended and will be presented in Chapter 7 of this thesis.

2.4.5 Combined Mechanical and Control System Design

Vehicle suspensions with active components potentially have significant advantages over passive suspensions [58, 71, 101, 102, 120, 144]. Thus, in the past 2-3 decades, various control strategies, e.g. Linear Optimal Control [143], "Skyhook Damper" [79], Preview Control [91], H_∞ Control [103], and Fuzzy-Logic Control [147], have been applied to the design of vehicle active suspensions. In these control strategies, active elements of vehicle suspensions are designed independently of the passive components [15, 19]. Applied to a vehicle model, such controllers may behave less optimally overall due to not considering the mechanical parameters, e.g. inertial and geometric parameters, as design variables in the control design process. Bestle et al. [15, 19] recommended that an integrated modeling and control design be performed to improve the performance of suspension systems and that parameters of both active and passive elements be adjusted adequately to attain optimal behavior of the overall vehicle.

In a robot design, Pil et al. [109] proposed a recursive experimental optimization method for simultaneously optimizing both the mechanical structure and controller of the

mechatronic system. In the design optimization, the mechanical structure might be modified recursively and quickly with structure reinforcement and rapid prototyping techniques. The feedback control gains were adjusted with the mechanical structure modification and this procedure was iterated until the optimized design criteria were reached.

The above methods applied to vehicle active suspension design, airplane design, and robot design share a common character: the corresponding mechanical system and control system are combined as a synergistic whole and the design optimization is conducted with the coordination between the mechanical system and control system. It is reasonable to expect that a combined mechanical and control design process may help to achieve an optimal behavior of an overall vehicle.

In recent years, several researchers have tackled the task by taking passive parameters and active parameters as design variables simultaneously when designing ground vehicles with active suspensions [12, 15, 66, 116].

However, all these researchers have introduced the "skyhook" control strategy [66] into their design optimizations. Since the 1970's, the linear quadratic Gaussian (LQG) optimal control algorithm [129] has been widely used by researchers for controller design in ground vehicles with active suspensions. This optimal control algorithm was also modified for optimizing a 4 DOF passive vehicle suspension system [94]. It was shown that LQG provides a compact analytical solution with relatively low design and computational time and the stability of the system designed is guaranteed. Moreover, the result of an optimization process is a controller that considers and feeds back all system states with constant gains while any classical structure may not be ensured to be optimal [129]. Although the passive spring stiffness and damping coefficients have been optimized with the control parameters using LQG algorithm [129], it seems that the vehicle inertial and geometric parameters and control parameters have not been considered as design variables simultaneously in the optimization process using LQG. Moreover, one difficulty in using LQG is how to determine the weighting factors of the performance index. Traditionally, a 'trial and error' method is used for choosing the weighting factors or the combinations of factors [129]. To some extent, the optimal design of controllers depends on the experience of designers. For complex design systems and for multicriteria design optimization problems in particular, the choice of the weighting factors is a nontrivial problem [101].

When the LQG algorithm is used for vehicle active suspension design, it is often assumed that perfect measurement of all state variables is available. In practice, not all the state variables are available but only a limited number of the states. Even in the latter case, the corresponding measurements are noisy and thus the performance of the control systems would suffer. In the cases where only limited states are assumed measurable, Kalman filters can be used to estimate required states.

Numerical results have showed that an integrated approach using genetic algorithms, multibody dynamics, LQG control strategy, and Kalman estimator is an effective approach to the design optimization of combined mechanical and control system for vehicles with active suspensions [67].

2.5 Summary

This chapter reviews the state-of-the-art related to modelling techniques, mathematical models, optimization methods, and optimization algorithms applied to the design optimization of rail vehicles. A survey of the applications of numerical optimization approaches to the design of rail vehicles with passive and active suspensions is presented.

Multibody dynamics programs are effective tools for modelling vehicle dynamic systems. The combination of multibody dynamics with appropriate optimization methods and algorithms may automate and facilitate the design process of rail vehicles. The increasing demands on the accuracy of mechanical models of vehicles result in complex multibody systems. It would be an impossible task to construct a single mathematical model that could be universally addressed to all aspects of vehicle dynamic behavior. However, the complicated dynamic behavior can be studied by using various mathematical models, each of which concentrates on a specific area of interest. For a design optimization problem that involves coupled analysis disciplines, multidisciplinary optimization is a suitable option. Genetic algorithms are effective search algorithms for multiple design variable optimization problems that have many local optima.

It is evident that there is a lack of systematic study on the feasibility and efficacy of optimization algorithms and formulation methods used in the design of rail vehicles with passive and active suspensions. Combined mechanical and control design process may help

to achieve an optimal behavior of an overall vehicle.

Chapter 3

Design Optimization Methods and Optimization Algorithms

3.1 Introduction

Roughly speaking, two cases were studied in the research and they will be discussed in the thesis: 1) a rail vehicle is optimized with respect to lateral stability, curving performance, and ride quality in vertical direction [66]; 2) to improve a ground vehicle ride quality with suspension working spaces and wheel dynamic loads being taken into consideration, the vehicle mechanical system, controller, and estimator are optimized simultaneously [67]. In both cases, MDO methods will be applied to the corresponding design optimization problems.

First of all, we are confronted with the problem of choosing the appropriate MDO methods mentioned in Chapter 2 for each case. In the first case, the lateral stability model, the curving performance model, and the vertical ride quality model represent three individual analysis disciplines which could be analysed individually during the optimization. In the process, the optimizer at system level can be used to drive the three individual disciplines to multidisciplinary feasibility and optimality by controlling the coupling design variables. The Individual Discipline Feasible Method is suitable for solving problems in which the coupling design variables are manipulated by the optimizer at system level. Thus, the IDF method is the appropriate option for the first case.

In the second case, for mechatronic vehicles with active suspensions, the vehicle mechanical system, active suspension controller, and Kalman filter co-exist in a synergistic system and each of them represents an analysis discipline. In the optimization, it is expected that a vector of design variables be offered to the coupled system of three analysis disciplines, each discipline feasibility should be achieved, and at the same time the input to one should correspond to the output of the other via interdisciplinary mappings. Via a fixed-point iteration with that value of design variable vector, the Multidisciplinary Feasibility (MDF) is expected to obtain, and the relevant output variables of the coupled system should be fed to the optimizer at the system level for evaluating the objective function and constraint functions. The inherent features of the All–in–One method match the above requirements so well that the A–i–O method is adopted for the second case.

For both cases, to optimize the conflicting design criteria simultaneously, the scalarization strategy commonly used in multicriteria optimization [17] is utilized in the corresponding MDO problem. At the system level, the weighted criteria method is used to construct the required objective function. It is expected that the introduction of the scalarization strategy or the weighted criteria method should facilitate the implementations of the IDF and A–i–O methods for the first case and the second case, respectively.

In the two cases, for these complicated design optimization problems with multidimensional, multicriteria, and multidisciplinary features, the selection of proper design space search algorithms or optimization algorithms is another important issue. Indirect search algorithms or gradient-guided search algorithms, e.g. SQP, often have the advantage of rapid convergence towards a local minimum. However, the convergence rates strongly depend on properties like differentiability or convexity. Moreover, even if these conditions are satisfied, the efficiency of these algorithms is degraded by the computational cost of gradients. Hence, the direct search algorithms or pattern search algorithms are also frequently applied to MDO problems [39, 132]. GAs offer another alternative to gradient–guided search algorithms [62]. Due to the long computation time, GAs cannot completely replace conventional gradient–guided search algorithms. Therefore, to identify effective optimization algorithms for vehicle suspension design, in the thesis, three typical algorithms, SQP, Simplex, and GA, are compared by means of ground vehicle ride quality analysis and rail vehicle lateral stability analysis.

In this chapter, both A–i–O and IDF approaches for MDO problems are introduced. Then, the essential concepts about multicriteria optimization are outlined. Finally, three typical optimization algorithms, SQP, Simplex, and GAs, are briefly reviewed.

3.2 Design Optimization Methods

3.2.1 Multidisciplinary Optimization Methods

All-in-One Method

The All-in-One method (or Multidisciplinary Feasibility (MDF) [33] method) is commonly used for approaching the solution of MDO problems. When this method is used, the optimization problem can be formulated as the following general format:

$$\begin{cases} \textit{minimize} & \mathbf{F}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}_d)) \\ \textit{with respect to} & \mathbf{X}_d \\ \textit{subject to} & \begin{cases} \mathbf{g}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}_d)) \leq \mathbf{0} \\ \mathbf{C}_l \leq \mathbf{X}_d \leq \mathbf{C}_u \end{cases} \end{cases} \tag{3.1}$$

where

$$\begin{cases} \mathbf{U}(\mathbf{X}_d) = \mathbf{A}(\mathbf{X}_d, \mathbf{G}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}_d))) \\ \mathbf{Y} = \mathbf{G}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}_d)) \end{cases} \tag{3.2}$$

and $\mathbf{C}_u$ and $\mathbf{C}_l$ are the upper and lower bounds on the design variable vector $\mathbf{X}_d$, $\mathbf{U}(\mathbf{X}_d)$ is the system output variable vector, $\mathbf{A}(\mathbf{X}_d, \mathbf{G}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}_d)))$ is the analysis mapping from the inputs $\mathbf{X}_d$ and $\mathbf{Y}$ of an analysis discipline to the outputs $\mathbf{U}$, $\mathbf{G}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}_d))$ is the mapping to the inputs required for an analysis discipline from the output of another analysis discipline, and $\mathbf{F}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}_d))$ and $\mathbf{g}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}_d))$ are the objective function vector and constrained function vector, respectively.

Figure 3.1 shows an example using the A-i-O method for an aeroelastic optimization problem [33]. The system consists of an optimizer that controls specified objective $\mathbf{F}$ (drag, structural weight, etc.) and constraints $\mathbf{g}$, two analysis disciplines including aerodynamics (discipline 1) with analysis solver $\mathbf{A}_1$ and structures (discipline 2) with analysis solver

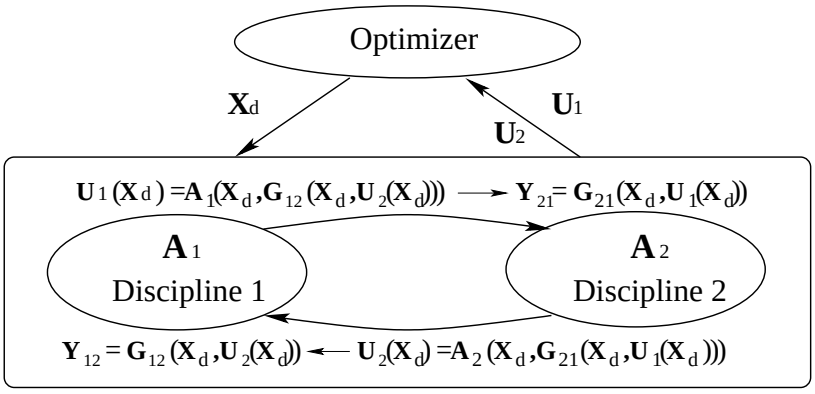

Figure 3.1: All-in-One (A-i-O) method

$\mathbf{A}_2$. For a certain iteration, the fixed design variable vector $\mathbf{X}_d$ (aerodynamic and structural design variables) is provided by the optimizer to the coupled analysis disciplines, then a complete aerodynamic analysis and structural analysis is performed with that value of $\mathbf{X}_d$ to obtain system output variable vectors $\mathbf{U}_1(\mathbf{X}_d)$ (the pressures on the wing surface) and $\mathbf{U}_2(\mathbf{X}_d)$ (the deflections of the wing). The output variable vectors $\mathbf{U}_1(\mathbf{X}_d)$ and $\mathbf{U}_2(\mathbf{X}_d)$ are used for evaluating the objective $\mathbf{F}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}_d), \mathbf{U}_2(\mathbf{X}_d))$ and constraints $\mathbf{g}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}_d), \mathbf{U}_2(\mathbf{X}_d))$. In Figure 3.1, the interdisciplinary mapping $\mathbf{G}_{21}$ means that given the pressures on the aerodynamic grid, the loads on the structures grid ($\mathbf{Y}_{21}$) can be calculated. Similarly, the interdisciplinary mapping $\mathbf{G}_{12}$ indicates that provided the deflections of the wing, the wing shape ($\mathbf{Y}_{12}$) can be obtained. We say that we have single discipline feasibility for aerodynamics when the solver $\mathbf{A}_1$ has been executed successfully and solved for the pressures, given an input shape. Similarly, we have single discipline feasibility for structures when the structures code $\mathbf{A}_2$ has successfully solved the structural analysis equations to produce deflections, given some input forces. Here, "feasibility" for a single discipline means that the equations the discipline code is intended to solve are satisfied.

For this two analysis discipline case, the optimization problem can be rewritten as:

$$\begin{cases} \textit{minimize} & \mathbf{F}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}_d), \mathbf{U}_2(\mathbf{X}_d)) \\ \textit{with respect to} & \mathbf{X}_d \\ \textit{subject to} & \begin{cases} \mathbf{g}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}_d), \mathbf{U}_2(\mathbf{X}_d)) \leq \mathbf{0} \\ \mathbf{C}_l \leq \mathbf{X}_d \leq \mathbf{C}_u \end{cases} \end{cases} \tag{3.3}$$

where

$$\begin{cases} \mathbf{U}_1(\mathbf{X}_d) = \mathbf{A}_1(\mathbf{X}_d, \mathbf{G}_{12}(\mathbf{X}_d, \mathbf{U}_2(\mathbf{X}_d))) \\ \mathbf{U}_2(\mathbf{X}_d) = \mathbf{A}_2(\mathbf{X}_d, \mathbf{G}_{21}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}_d))) \\ \mathbf{Y}_{12} = \mathbf{G}_{12}(\mathbf{X}_d, \mathbf{U}_2(\mathbf{X}_d)) \\ \mathbf{Y}_{21} = \mathbf{G}_{21}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}_d)) \end{cases} \tag{3.4}$$

Notice that if a gradient-guided method is to be used to solve the above problem, then a complete multidisciplinary analysis (MDA) is necessary not just at every iteration, but at every point where the derivatives are to be evaluated. Thus, it is very expensive to attain multidisciplinary compatibility in realistic applications.

Individual Discipline Feasible Method

One way to avoid a complete MDA every time an objective function, constraint, or sensitivity evaluation is needed is to use the IDF method. The essence of IDF is that this approach maintains individual discipline feasibility, while allowing the optimizer to drive the individual disciplines to multidisciplinary feasibility and optimality by controlling the interdisciplinary coupling variables. In the case of IDF approach, some specific analysis variables representing communication, or coupling, between analysis disciplines via interdisciplinary mappings are "promoted" to become optimization variables. These optimization variables are indistinguishable from design variables from the point of view of a single analysis discipline solver. The general IDF formulation can be as follows:

$$\begin{cases} \textit{minimize} & \mathbf{F}(\mathbf{X}_d, \mathbf{U}(\mathbf{X})) \\ \textit{with respect to} & \mathbf{X} = (\mathbf{X}_d, \mathbf{X}_\mathbf{Y}) \\ \textit{subject to} & \begin{cases} \mathbf{g}(\mathbf{X}_d, \mathbf{U}(\mathbf{X})) \leq \mathbf{0} \\ \mathbf{C}_{aux} \stackrel{\triangle}{=} \mathbf{X}_\mathbf{Y} - \mathbf{G}(\mathbf{X}_d, \mathbf{U}(\mathbf{X})) = 0 \\ \mathbf{C}_l \leq \mathbf{X} \leq \mathbf{C}_u \end{cases} \end{cases} \tag{3.5}$$

where

$$\mathbf{U}(\mathbf{X}) = \mathbf{A}(\mathbf{X}) \tag{3.6}$$

and $\mathbf{C}_u$ and $\mathbf{C}_l$ are the upper and lower bounds on the design variable vector $\mathbf{X}$ which consists of the original design variable vector $\mathbf{X}_d$ and "promoted" design variable vector $\mathbf{X}_\mathbf{Y}$. Since the vector $\mathbf{Y}$ is "promoted" as design variable vector, here $\mathbf{X}_\mathbf{Y}$ is introduced to replace the input variable vector $\mathbf{Y}$ for an analysis discipline. $\mathbf{F}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}))$ and $\mathbf{g}(\mathbf{X}_d, \mathbf{U}(\mathbf{X}))$ are objective and constraints, respectively. $\mathbf{U}(\mathbf{X})$ is the system output variable vector and $\mathbf{A}(\mathbf{X})$ is the analysis mapping from the inputs $\mathbf{X}_d$ and $\mathbf{X}_\mathbf{Y}$. $\mathbf{G}$ represents interdisciplinary mapping and the condition $\mathbf{C}_{aux} \stackrel{\triangle}{=} \mathbf{X}_\mathbf{Y} - \mathbf{G}(\mathbf{X}_d, \mathbf{U}(\mathbf{X})) = 0$ converts the interdisciplinary mappings into auxiliary optimization constraints. Notice that in the thesis the symbol '$\stackrel{\triangle}{=}$' means 'defined as'.

It should be noted that an evaluation of $\mathbf{U}(\mathbf{X}) = \mathbf{A}(\mathbf{X})$ involves executing all the single discipline analysis codes simultaneously with available multidisciplinary data $\mathbf{X}$. Therefore, these very expensive computations can be done independently and concurrently and communication costs are likely to be negligible. It is evident that the IDF method is well-suited for applications with the use of parallel computer system.

Figure 3.2 shows an application of the IDF method to a system consisting of an optimizer that controls objective $\mathbf{F}$ and constraints $\mathbf{g}$, $\mathbf{C}_{12}$ and $\mathbf{C}_{21}$, discipline 1 with analysis solver $\mathbf{A}_1$, and discipline 2 with analysis solver $\mathbf{A}_2$. For a certain iteration, the fixed design variable vectors $\mathbf{X}_d$, $\mathbf{X}_{Y_{12}}$, and $\mathbf{X}_{Y_{21}}$ are provided by the optimizer to the analysis disciplines $\mathbf{A}_1$ and $\mathbf{A}_2$. With the offered design variable vectors, each analysis is performed to obtain system output vectors $\mathbf{U}_1(\mathbf{X})$ and $\mathbf{U}_2(\mathbf{X})$ and interdisciplinary mapping vectors $\mathbf{G}_{12}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}))$ and $\mathbf{G}_{21}(\mathbf{X}_d, \mathbf{U}_2(\mathbf{X}))$, respectively. The objective $\mathbf{F}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}), \mathbf{U}_2(\mathbf{X}))$ and constraints $\mathbf{g}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}), \mathbf{U}_2(\mathbf{X}))$, $\mathbf{C}_{12}$, and $\mathbf{C}_{21}$ can be evaluated, given the system output vectors $\mathbf{U}_1(\mathbf{X})$ and $\mathbf{U}_2(\mathbf{X})$ and interdisciplinary mapping vectors $\mathbf{G}_{12}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}))$

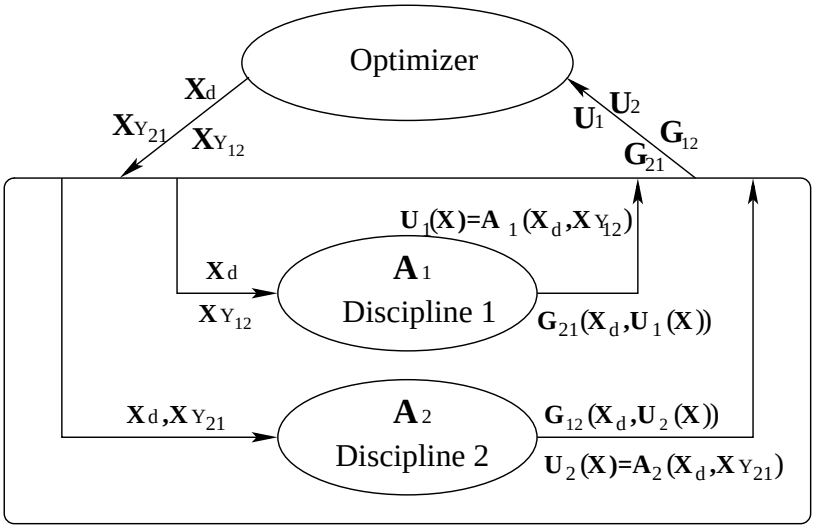

Figure 3.2: Individual discipline feasible (IDF) method

and $\mathbf{G}_{21}(\mathbf{X}_d, \mathbf{U}_2(\mathbf{X}))$. For this case, the optimization problem is formulated as:

$$\begin{cases} \textit{minimize} & \mathbf{F}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}), \mathbf{U}_2(\mathbf{X})) \\ \textit{with respect to} & \mathbf{X} = (\mathbf{X}_d, \mathbf{X}_{Y_{12}}, \mathbf{X}_{Y_{21}}) \\ \textit{subject to} & \begin{cases} \mathbf{g}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X}), \mathbf{U}_2(\mathbf{X})) \leq \mathbf{0} \\ \mathbf{C}_{12} \stackrel{\triangle}{=} \mathbf{X}_{Y_{12}} - \mathbf{G}_{12}(\mathbf{X}_d, \mathbf{U}_2(\mathbf{X})) = 0 \\ \mathbf{C}_{21} \stackrel{\triangle}{=} \mathbf{X}_{Y_{21}} - \mathbf{G}_{21}(\mathbf{X}_d, \mathbf{U}_1(\mathbf{X})) = 0 \\ \mathbf{C}_l \leq \mathbf{X}_d \leq \mathbf{C}_u \end{cases} \end{cases} \tag{3.7}$$

where

$$\begin{cases} \mathbf{U}_1(\mathbf{X}) = \mathbf{A}_1(\mathbf{X}_d, \mathbf{X}_{Y_{12}}) \\ \mathbf{U}_2(\mathbf{X}) = \mathbf{A}_2(\mathbf{X}_d, \mathbf{X}_{Y_{21}}) \end{cases} \tag{3.8}$$

For the above A-i-O and IDF methods, with moderate or no modification, they all have the advantage of using existing single discipline analysis codes. Compared with the A-i-O method, the IDF method avoids the expensive procedure for achieving full multidisciplinary feasibility at each optimization iteration. Moreover, when using the IDF method, one may easily replace one analysis code with another, or add new disciplines, and one can easily

implement parallel and distributed computation. On the other hand, the IDF method requires the explicit imposition in the optimization of the nonlinear constraints resulting from the interdisciplinary maps. If gradient-guided optimization algorithms are used, the calculation of additional sensitivities corresponding to the coupling variables between disciplines may be very expensive. Provided the coupling variables and constraints are small, the overall IDF optimization will be significantly more efficient than A-i-O optimization.

3.2.2 Multicriteria Optimization Concepts

To implement multicriteria optimization methods, the relevant criteria should be sorted based on their purpose. The introduction of some criteria may lead to excluding infeasible designs. This kind of criteria can be treated as equality constraints or inequality constraints. Criteria, whose values are relevant to the behavior or performance indices of the dynamic systems concerned, are often selected as objective functions. Finally, several key objective functions are left and a vector or multicriteria optimization problem can be described as:

$$\begin{cases} \textit{minimize} & \mathbf{F}(\mathbf{X}_d) \\ \textit{with respect to} & \mathbf{X}_d \\ \textit{subject to} & \begin{cases} \mathbf{g}(\mathbf{X}_d) = \mathbf{0} \\ \mathbf{h}(\mathbf{X}_d) \leq \mathbf{0} \end{cases} \end{cases} \tag{3.9}$$

where $\mathbf{X}_d$, $\mathbf{F}$, $\mathbf{g}$, and $\mathbf{h}$ are design variable vector, objective function vector, equality constraint vector, and inequality constraint vector, respectively. Frequently, a design variable vector with which all criteria reach their minimal values simultaneously is not feasible.

Although a unique optimal solution can not be defined generally, non-optimal designs can be eliminated. For example, for every design variable vector $\mathbf{X}_d$ that satisfies the constraints shown in (3.9), if $\mathbf{F}(\mathbf{X}_d) > \mathbf{F}(\bar{\mathbf{X}}_d)$ and $\bar{\mathbf{X}}_d$ is a feasible design variable vector, the design variable vector $\mathbf{X}_d$ is not optimal. Design variable vectors $\tilde{\mathbf{X}}_d$ that satisfy the constraints described in (3.9) are called Edgeworth–Pareto–optimal (EP–optimal), if there is no feasible design variable vector $\mathbf{X}_d$ where $F_i(\mathbf{X}_d) \leq F_i(\tilde{\mathbf{X}}_d),\ \forall i \vee \mathbf{F}(\mathbf{X}_d) \neq \mathbf{F}(\tilde{\mathbf{X}}_d)$, [18]. Usually, EP-optimal solutions are not unique and design points with different images (e.g. curves or surfaces) are not comparable, all of them have to be considered as optimal. The

designer has to choose a special EP–optimal solution based on additional information on the design problem.

A whole picture of EP–optimal solutions of multicriteria optimization problems requires many objective function evaluations. For dynamic system design, objective function evaluations involve a time–consuming numerical integration of differential equations of motion. In high-dimensional problems, the EP–optimal solution cannot be visualized any more. Even computing a representative part of the EP–optimal set is already too time–consuming for dynamic problems. Therefore, not all multicritia optimization strategies are appropriate for dynamic system design.

The commonly used strategy for dynamic system design is to reduce the vector optimization problem to a scalar one that may be solved by existing optimization algorithms [42]. This strategy has proven to be very efficient. This reduction is based on the principles of scalarization, hierarchization or a combination of them.

In the following subsections, the principles of scalarization and hierarchization are briefly introduced.

Scalarization

As shown in Figure 3.3, during the process of scalarization, the objective functions are formulated as a scalar utility function $u(\mathbf{F}(\mathbf{X}_d))$. During the optimization, instead of the vector of objective functions, the scalar utility function is minimized. In the design space, the utility function should have the property of monotonicity, i.e., for two different scalars F^a and F^b, if $F^a < F^b$, then $u(F^a) < u(F^b)$.

To implement the scalarization, generally, the utility function can be formulated either by the weighted criterion method or by the distance method. For the weighted criterion method, the utility function can be expressed as

$$\begin{cases} u(\mathbf{F}) = \sum_{i=1}^{n} \rho_i F_i \\ \rho_i > 0 \end{cases} \tag{3.10}$$

where $\rho_i,\ i = 1, 2, ..., n$, are weighting factors. In the case of the distance method, the

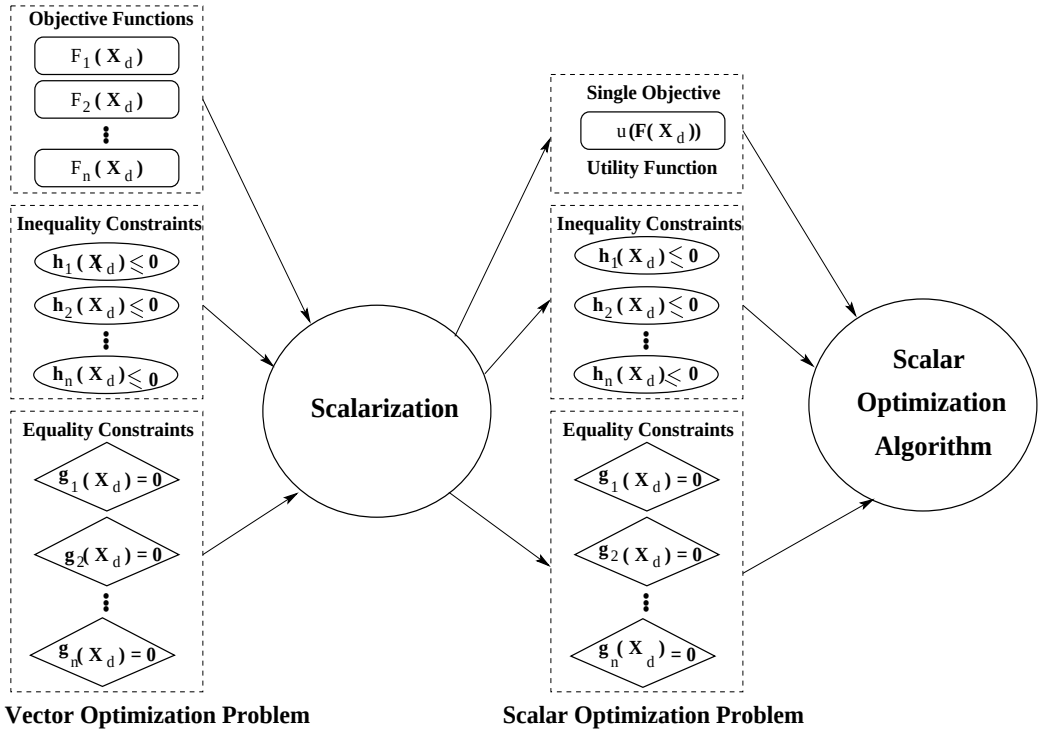

Figure 3.3: Scalarization by introducing an utility function

utility function is written as

$$\begin{cases} u(\mathbf{F}) = (\sum_{i=1}^{n} |F_i - \bar{F}_i|^{\varrho})^{1/\varrho} \\ 1 \leq \varrho < \infty \end{cases} \tag{3.11}$$

where $\bar{F}_i,\ i = 1, 2, ..., n$, are ideal or utopian design goal vectors.

The weighted criterion method is widely applied, but the weighting coefficients (ρ_i) are difficult to choose and the optimization result depends on this choice in a highly nonlinear fashion. To facilitate the implementation of the weighted criterion method, it is recommended that each element of the objective function vector $\mathbf{F}$ be normalized to have a value of one for the initial design variables [42]. This recommendation is based on the requirement that all elements of the objective function vector should be optimized simultaneously. Obviously, the selection of the initial design variables and the corresponding values for each element of the objective function vector are vital for the normalization.

Given the utopian goals $\bar{F}_i$, the distance method is a preferable option to the weighted criterion method, since the tedious job of choosing the weighting coefficients for the latter can be avoided. Usually, the goals have some physical meaning. However, if the goals $\bar{F}_i$ are not utopian solutions, EP–optimality can not be reached.

Hierarchization

In the case of hierarchization, each objective function $F_i(\mathbf{X}_d)$ has to be assigned a level of importance. The objective function with the highest priority is optimized first without taking into consideration the objective functions with lower priority. Based on the information obtained during the process of optimizing the objective function with higher priority, constraints on the objective function can be formulated. Then the next important objection function can be optimized with respect to the additional constraints. In essence, the principle of hierarchization is to transform a vector optimization problem to a sequence of scalar optimization problems.

It was shown [42] that by means of conventional nonlinear programming algorithms such as sequential quadratic programming, the solution of the reduced optimization problem will yield only a single point of the whole set of EP–optimal solutions of the original multicriteria design problem. Thus, it is hard to get a clear picture of the EP–optimal set. However, stochastic optimization algorithms, like a simulated annealing algorithm, presents a much clearer picture of the EP–optimal set for multicriteria optimization problems [42].

3.3 Optimization Algorithms

3.3.1 Genetic Algorithms (GAs)

It is well-known that the evolution of living beings is a process based on operating on chromosomes which are organic devices for encoding the structure of living beings. Natural selection plays a role as a link between a chromosome and the performance of their decoded structures. The operations of selection offer more chance for chromosomes that encode successful structures to reproduce than those that do not; the processes of mutation result in the chromosome of offsprings to be different from those of their parents; the operations of

crossover may produce quite different chromosome in offsprings by combining the material from the chromosome of their two parents.

GAs are developed to mimic some of the above processes observed in natural evolution. With an appropriate encoding mechanism, GAs manipulate strings of binary digits, i.e. $1s$ and $0s$, which correspond to chromosomes of living beings. By means of coding, there exists a map between a design variable vector $\mathbf{X}_d$ and the corresponding binary string or chromosome:

$$\text{individual design variable } \mathbf{X}_d \quad \overset{\textit{coding}}{\underset{\textit{decoding}}{\Longleftrightarrow}} \quad \text{binary string } \check{b} = [1, 1, 0, ..., 0, 1]$$

The following function defines the coding–decoding mechanism.

$$\mathbf{X}_d \overset{\triangle}{=} \tilde{g}(\check{b}) \tag{3.12}$$

The binary string $\check{b}$ contains all of the necessary information that the individual $\mathbf{X}_d$ implies, e.g., $\mathbf{X}_d$ may be a set of design variables representing inertial and geometric variables for a vehicle dynamic system. For an admissible individual $\check{b}$, the fitness $\tilde{F}$ can be defined as the value of the objective function F.

$$\tilde{F}(\check{b}) \overset{\triangle}{=} F(\mathbf{X}_d) \tag{3.13}$$

The objective function value of the individual point $\mathbf{X}_d$ in the design variable space, in general, will be maximized or minimized by GAs. The fitness value will be used to determine the probability for the individual to be acted on by genetic operators. Usually, a population of individual design variable sets evolves from generation to generation through the application of genetic operators. The total number of strings included in a population is kept unchanged throughout generations. Simple GAs use three operators: selection, crossover, and mutation.

Selection is a process in which individual strings are copied based on their fitness values. Highly fit strings (good designs) have a higher number of offspring in the succeeding generation. Crossover is a method of combing successful individuals by exchanging equivalent lengths of their chromosome. The two strings from the reproduced population are

mated randomly, and a crossover site is selected at random. Mutation is a technique that introduces new information into the new population at the bit level. A set of bits are selected randomly within the entire population.

After performing selection, crossover and mutation, GAs generate a new population with potentially more individuals of higher fitness value. With enough repetitions of the cycle, the population will converge on the chromosome/design with the highest fitness.

GAs are well suited for unconstrained optimization problems [36]. However for constrained optimization problems, we can use the penalty method that degrades the fitness ranking in relation to the degree of constraint violation. With the method, a standard constrained optimization problem with following form

$$minimize \qquad F(\mathbf{X}_d) \tag{3.14}$$

$$subject\ to \quad \begin{cases} g_i(\mathbf{X}_d) \leq 0, & \text{i=1,2,...,m} \\ h_i(\mathbf{X}_d) = 0, & \text{i=1,2,...,p} \end{cases} \tag{3.15}$$

can be transformed into an unconstrained optimization problem by associating a cost or penalty with each constraint violation. Thus, for the constrained optimization problem described in (3.14) and (3.15), we can write the fitness function as follows

$$\tilde{F}(\check{b}) = F(\mathbf{X}_d) + \sum_{i=1}^{p} \gamma_i |h_i(\mathbf{X}_d)|^{\tau_i} + \sum_{j=1}^{m} \gamma_{p+j} (g_j(\mathbf{X}_d) + |g_j(\mathbf{x}_d)|)^{\tau_{p+j}} \tag{3.16}$$

where $\tilde{F}(\check{b})$ is the fitness value associated to the binary string $\check{b}$ which corresponds to the design variable vector $\mathbf{X}_d$, the constants γ_i and $\tau_i (i = 1, 2, ..., p+m)$ determine the severity of the penalties for the $p + m$ constraints.

GAs offer significant advantages over traditional local search methods because of the following characteristics [57]: a) GAs work on a population of design variables in parallel and not on a unique point, so that GAs have a higher reliability to find the global optima; b) GAs solve the problem of finding good chromosomes (designs) by manipulating the material in the chromosome without any knowledge of the problem they are solving. The only information they require is an evaluation of each chromosome/design — they do not need the gradients of the objective function and constraints; c) they are simple yet powerful in their search for improvement and they are not limited by restrictive requirements about

the search space, such as continuity or existence of derivatives; d) GAs guide their searches using probability rules; this enhances their global explorative properties.

In the research, the GA is implemented using the MechaGen program [10]. The MechaGen program is based on Goldberg's GA [57] and was written in C using pseudo-random number generators linked from the NAG (Numerical Algorithms Group) Fortran library. However, to avoid premature termination of the algorithm, instead of using a weighted roulette wheel based on the fitness sum of the population for the reproduction stage, one based on the ranking of the population according to fitness is used [138]. In addition, to improve the efficiency of the GA, the binary strings and fitness values for each unique design of the current generation are stored in a linear search look-up table. If a design string in the next generation matches one in the table, then the fitness does not have to be re-calculated. This saves significant computing time, especially for expensive fitness evaluations.

3.3.2 SQP Algorithm

The SQP algorithm is a nonlinear programming algorithm known as the projected Lagrangian method. It is used for the purpose of minimizing a smooth nonlinear function subject to a set of constraints with upper and lower bounds. This general constrained minimization problem can be transformed into the standard form offered in (3.14) and (3.15). The objective function and the constraint functions are assumed to be at least twice-continuously differentiable

For a local optimum $\mathbf{X}_d^*$, the following first-order Kuhn–Tucker (KT) condition is a necessary qualification [21]:

$$\bigtriangledown F(\mathbf{X}_d^*) + \sum_{i \in active} u_i \bigtriangledown g_i(\mathbf{X}_d^*) + \sum_{i=1}^{p} v_i \bigtriangledown h_i(\mathbf{X}_d^*) = 0 \tag{3.17}$$

where $\bigtriangledown F(\mathbf{X}_d^*)$, $\bigtriangledown g_i(\mathbf{X}_d^*)$ and $\bigtriangledown h_i(\mathbf{X}_d^*)$ are the gradients of the objective function, active inequality constraints and equality constraints, respectively. The corresponding Lagrange multipliers are $u_i \geq 0,\ i \in active$, and $v_i,\ i = 1, 2, ..., p$. If the optimization problem is a convex programming problem in which $F(\mathbf{X}_d)$ is convex, the equality constraints are linear, and the inequality constraints are concave, the first order KT condition is both

the necessary and sufficient condition that guarantees the point $\mathbf{X}_d^*$ as the global optimal point.

Each iteration step of SQP generally consists of three partial steps [56]:

Step 1: Find the Lagrange multipliers and the direction to move from the current iteration by solving the quadratic programming (QP) subproblem.

Generally a sequence of iterates $\mathbf{X}_d^k$ that converge to $\mathbf{X}_d^*$, i.e. a first-order KT point of (3.17), is generated. At a typical iteration, the new iterate $\mathbf{X}_d^{k+1}$ can be expressed as

$$\mathbf{X}_d^{k+1} = \mathbf{X}_d^k + \alpha \mathbf{s} \tag{3.18}$$

where $\mathbf{s}$ is the search direction, and α is the step length. The search direction in (3.18) is the solution of a quadratic programming subproblem of the form

$$minmize \qquad \bigtriangledown F(\mathbf{X}_d^k)^T \mathbf{s} + \tfrac{1}{2}\mathbf{s}^T H(\mathbf{X}_d^k)\mathbf{s} \tag{3.19}$$

$$subject\ to \quad \begin{cases} h_i(\mathbf{X}_d^k) + \bigtriangledown h_i(\mathbf{X}_d^k)^T \mathbf{s} = 0, & \text{i=1,2,...,p} \\ g_i(\mathbf{X}_d^k) + \bigtriangledown g_i(\mathbf{X}_d^k)^T \mathbf{s} \leq 0, & \text{i=1,2,...,m} \end{cases} \tag{3.20}$$

where $H(\mathbf{X}_d^k)$ is an approximation to the Hessian matrix of the Lagrangian which is obtained by solving KT conditions for the original problem at $\mathbf{X}_d^k$.

Step 2: Determine the step-length α in (3.18) to be taken from the point $\mathbf{X}_d^k$ in the direction obtained in step 1 such that there is "sufficient decrease" in a merit function. A merit function measures the value or worth of the current design point $\mathbf{X}_d^k$ by taking into account the objective function $F(\mathbf{X}_d)$, KT conditions (3.17) and the Lagrange multipliers.

Step 3: Update the approximate Hessian matrix of the Lagrangian using Newton's approach or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) approach.

Using Newton's method for nonlinear equations for solving KT conditions with the original problem considered, we have the Hessian of the form

$$H(\mathbf{X}_d^k) = \bigtriangledown^2 F(\mathbf{X}_d^k) + \sum_{i=1}^{m} u_i \bigtriangledown^2 g_i(\mathbf{X}_d^k) + \sum_{i=1}^{p} v_i \bigtriangledown^2 h_i(\mathbf{X}_d^k) \tag{3.21}$$

The SQP algorithm has very strong, theoretical, local convergence properties. The sequence of iterates converges quadratically to a local minimum that satisfies the KT and sufficient optimality conditions when starting from a point sufficiently close to, or in the neighborhood of, that minimum.

As mentioned in the previous chapter, the SQP algorithm is often applied to the design optimization of ground vehicle suspensions. Of course, convergence rates strongly depend on properties such as convexity or differentiability. Since these properties can rarely be proven for complicated vehicle dynamic systems, there is, in general, no guarantee for convergence. When the algorithm is used, some experience with specific vehicle models is required for the purpose of determining a desired behavior. Due to the fact that the SQP algorithm only takes local information into consideration, the algorithm often gets trapped at local optimal points. For complicated vehicle dynamic systems, often many local minima exist; thus, there is no guarantee for finding global minima with SQP.

In the research, the SQP algorithm is implemented using the E04UCF routine from the NAG (Numerical Algorithms Group) library. This routine is well-documented and offers several user-friendly provisions. It estimates the Hessian matrix of the objective function using a preceding estimation of the matrix and the gradient of the objective function. The user may offer the first-order derivatives of the constraint functions and objective function, or they can be approximated by E04UCF using finite differences. If the objective function and constraint functions can be expressed explicitly in terms of design variables, then maximum reliability is achieved by providing as many partial derivatives as possible.

3.3.3 Simplex Algorithm

The Simplex algorithm minimizes a general function $F(\mathbf{X}_d)$ of n independent variables $\mathbf{X}_d = \{X_{d1}, X_{d2}, ..., X_{dn}\}^T$. This algorithm utilizes a regular geometric figure, called a simplex, consisting of $n+1$ vertices. At each stage of the algorithm, a simplex of $n+1$ points is retained, together with the function values at these points. At each iteration, a new simplex will be generated by producing a new point to replace the "worst" point. The vertex or the worst point of the simplex is reflected in the centroid of the remaining vertices and the function value at this new point is compared with the remaining function values. Depending on the outcome of this test, the new point is accepted or rejected, a further expansion move may be made, or a contraction may be carried out. When no further progress can be made the sides of the simplex are reduced in length and the method is repeated, in such a way that it adapts itself to the function landscape and finally surrounds the local optimum.

If f_i, for $i = 1, 2, ..., n+1$, are the individual function values at the vertices of a simplex and f_m is the mean of these values, then the algorithm will terminate when

$$\sqrt{\frac{1}{n+1}\sum_{i=1}^{n+1}(f_i - f_m)^2} < \varepsilon \tag{3.22}$$

where ε is the desired error tolerance. The algorithm may also terminate when a specified number of iterations has been exceeded.

The Simplex algorithm based on function comparison is suitable for problems in which $F(\mathbf{X}_d)$ is discontinuous [56]. Furthermore, it is robust and therefore very useful for functions that are subject to inaccuracies.

Once the algorithm is used, the function $F(\mathbf{X}d)$ is assumed to be unimodal, i.e. there is an unique optimum in the feasible design variable space, otherwise the identification of the global optimum is not guaranteed. To improve the reliability for finding the global optimum, it is customary to consider several sets of starting points and restart the entire procedure such that the searches are carried out until the simplex repeatedly collapses onto the same solution.

The Simplex algorithm, like the GAs described previously, is an unconstrained optimization algorithm. However, by means of introducing penalty factors in the objective function, the algorithm can be applied to constrained optimization problems.

During the research, the E04CCF routine from NAG library is the implementation of the Simplex algorithm. Before using the routine, the optimization problem should be scaled so that the values of the design variables are of order unity.

3.4 Summary

The IDF method is chosen for the design optimization of a rail vehicle with respect to lateral stability, curving performance, and ride quality and the A–i–O method is selected for improving vehicle ride quality with a vehicle mechanical system, active suspension controller, and corresponding Kalman filter being optimized simultaneously. Both IDF and A–i–O methods are briefly introduced. The multicriteria optimization procedure and essential concepts such as EP–optimal solutions and the scalarization strategy are offered.

Three typical optimization algorithms, SQP, Simplex, and GAs are described and their characteristics are highlighted.

Chapter 4

Optimization of the Lateral Stability

4.1 Introduction

The objective of this chapter is to demonstrate how to combine the GA and SQP algorithms as well as the dynamic mode tracking (DMT) technique [7] with advanced multibody dynamic simulation programs such as A'GEM. The effectiveness of this combined approach is investigated and validated using a 17 DOF rail vehicle model. Although other objective criteria such as ride quality or curving performance could be introduced into the design optimization problem, only the critical speed is considered as the design objective in the chapter. The multidisciplinary and multicriteria optimization problem will be discussed in detail in Chapter 7.

Several steps are taken to validate and show the effectiveness of this integrated approach. First, a hand-derived solution to the 17 degree of freedom linear rail vehicle model is compared to the results from an A'GEM simulation. Second, the calculation of the critical speed is investigated by comparing a combined algorithm including the SQP and DMT with the SQP algorithm alone when they are used for identifying the critical speed of the rail vehicle for two specified examples. In the process, the existence of sharply-discontinuous "cliffs" in the plots of critical speed versus suspension stiffnesses are identified. These cliffs, which are due to switching of the least-damped mode in the system, greatly hinder the application of gradient-based optimization algorithms. The critical speed is optimized using a genetic algorithm, and compared against the Simplex algorithm. In recognition of the

cliff phenomenon, the definition of critical speed is generalized to make it a more practical measure of lateral stability. Finally, the integrated approach is used to identify the relative significance of different design parameter sets, i.e. geometric (7 variables), inertial property (9 variables), and suspension (11 variables) parameters, on the critical speed of the rail vehicle.

4.2 Vehicle System Model

In this study, the dynamic equations for a 17 DOF rail vehicle model are generated and linearized, both by hand and by the A'GEM multibody simulation program. Using a set of nominal design variables, the solutions from the hand-derived model and the A'GEM model are compared.

4.2.1 Model Description and Hand Derived Equations

The 17 DOF model is shown in Figure 4.1, with the leading bogie, car body, and trailing bogie denoted as bodies 2, 4, and 6, respectively. The leading bogie, with the leading and trailing wheelsets denoted as 1 and 3, and trailing bogie, with the leading and trailing wheelsets denoted as 5 and 7, are connected to the car body by secondary suspensions. Both the leading and trailing bogies, in turn, are connected with their own leading and trailing wheelsets by primary suspensions. Each suspension component consists of a parallel spring and damper, with stiffness and damping coefficients in the three coordinate directions.

The vehicle is assumed to be traveling with a constant velocity V along a flat tangential section of railway track. The x axis of a Cartesian reference frame is aligned parallel to the track, the z axis is vertically downward, and the y coordinate is used to measure the lateral displacements of the car body, bogies, and wheelsets from the centerline of the track. The nominal geometry parameters, suspension stiffness and damping coefficients, and inertial property parameters are listed in Table A.2 in Appendix A.

For the bogies and car body, the motions considered are lateral displacements y_i, yawing ψ_i (about axis z), and rolling ϕ_i (about axis x), where $i = 2, 4, 6$. For the wheelsets, the motions considered are lateral displacement y_i and yawing ψ_i, where $i = 1, 3, 5, 7$. The resulting vehicle model has 17 DOF. In deriving the equations of motion, each body's lateral

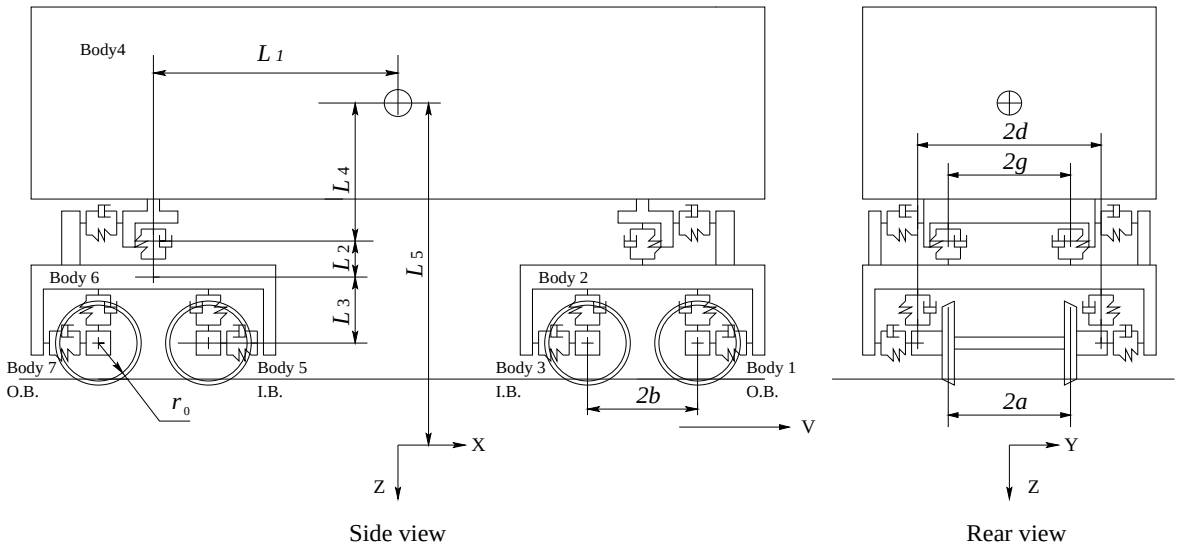

Figure 4.1: Configuration of a 17 DOF rail vehicle model

and yawing displacement from the equilibrium are assumed to be small. For simplicity, in the hand-derived model, the gravitational terms (the contributions of wheelset loads to the lateral forces and yaw moments due to the yaw motions) and spin creep are neglected, and the creep forces are linearized. The longitudinal and lateral creep coefficients f_{11} and f_{22} take the values offered in Table A.1 in Appendix A. Note that in the A'GEM vehicle model, the gravitational terms are included but both the lateral/spin creep coefficient f_{23} and spin creep coefficient f_{33} are set to zero for comparison with the hand-derived model. Otherwise, f_{23} and f_{33} take the values provided in Table A.1 in Appendix A.

The four creep coefficients f_{11}, f_{22}, f_{23}, and f_{33} are obtained using A'GEM's CREEP routine, which is based on Kalker's linear creep theory [78]. These four creep coefficients are calculated using the values shown in Table A.2 in Appendix A for the normal load on a wheel (W), the wheel rolling radius (r_0), the wheel transverse radius (r_w), the rail transverse radius (r_r), and Poisson's ratio (σ) and Young's modulus (E) for both wheel and rail materials.

From Newton's laws of dynamics, the linearized equations of motion can be cast in

matrix form as follows:

$$\mathbf{M}\ddot{\mathbf{r}} + \mathbf{C}\dot{\mathbf{r}} + \mathbf{K}\mathbf{r} = \mathbf{0} \tag{4.1}$$

where $\mathbf{r} = \{y_1, \psi_1, y_2, \psi_2, \phi_2, y_3, \psi_3, y_4, \psi_4, \phi_4, y_5, \psi_5, y_6, \psi_6, \phi_6, y_7, \psi_7\}^T$, and $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are mass, damping, and stiffness matrices and these matrices are offered in Appendix A, respectively.

Due to the unsymmetrical nature of the stiffness matrix $\mathbf{K}$, equation (4.1) can not be transformed into an uncoupled set of differential equations by means of conventional methods of modal analysis, i.e. using only real modes. A solution can still be effected though, by using the complex modal transformation described in [100]. To do this, equation (4.1) is recast into the following state-space form:

$$\dot{\mathbf{q}} = \mathbf{A}\mathbf{q} \tag{4.2}$$

where the relation of the generalized coordinates with the denoted bodies is listed in Table 4.1 and the assembled generalized coordinates $\mathbf{q}$ and coefficient matrix $\mathbf{A}$ are listed as follows:

$$\mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4, \mathbf{q}_5, \mathbf{q}_6, \mathbf{q}_7, \dot{\mathbf{q}}_1, \dot{\mathbf{q}}_2, \dot{\mathbf{q}}_3, \dot{\mathbf{q}}_4, \dot{\mathbf{q}}_5, \dot{\mathbf{q}}_6, \dot{\mathbf{q}}_7\}^T \tag{4.3}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{17\times 17} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \tag{4.4}$$

where $\mathbf{I}_{17\times 17}$ is an identity matrix. After performing an eigenvalue analysis on matrix $\mathbf{A}$, the solution to equations (4.2) is given by:

$$\mathbf{q}(t) = \mathbf{R}e^{[\mu]t}\mathbf{L}^T\mathbf{q}(0) \tag{4.5}$$

in which $\mathbf{q}(0)$ contains the initial conditions specified for the physical variables, $e^{[\mu]t}$ is a diagonal matrix with each entry having the form $e^{\mu_i \cdot t}$, $i = 1, 2, ..., 34$, μ_i is the corresponding eigenvalue of the matrix $\mathbf{A}$, and $\mathbf{R}$ and $\mathbf{L}$ contain the "right" and "left" eigenvectors of $\mathbf{A}$ [100].

With solution (4.5), we can analyze the relationship between the critical speed and the suspension parameters. For any μ_i, assuming its real part is $Re(\mu_i)$, if $Re(\mu_i) < 0$, the corresponding motion mode is stable.

Table 4.1: Relationship between the generalized coordinates and the denoted bodies

Body	*Type*	Generalized Coordinates
1	*wheelset*	$\mathbf{q}_1 = \{y_1, \psi_1\}$
2	*bogie*	$\mathbf{q}_2 = \{y_2, \psi_2, \phi_2\}$
3	*wheelset*	$\mathbf{q}_3 = \{y_3, \psi_3\}$
4	*car*	$\mathbf{q}_4 = \{y_4, \psi_4, \phi_4\}$
5	*wheelset*	$\mathbf{q}_5 = \{y_5, \psi_5\}$
6	*bogie*	$\mathbf{q}_6 = \{y_6, \psi_6, \phi_6\}$
7	*wheelset*	$\mathbf{q}_7 = \{y_7, \psi_7\}$

4.2.2 Validation of A'GEM Using Hand-Derived Model

For the purpose of investigating the integrated approach using GAs, SQP, DMT, and multibody dynamics, the A'GEM multibody dynamics program is used for modelling and simulating the rail vehicle response.

Using influence coefficients that relate the displacement of force-producing components (e.g. springs and dampers) to the DOF defined for the wheelsets, bogies, and car body, A'GEM automatically generates equations of motion that are linear in terms of geometrical effects [6]. For the purpose of lateral stability analysis, A'GEM approximates the wheel/rail forces, i.e. the lateral creep, longitudinal creep, and spin creep, and the gravitational stiffness as linear functions of the generalized coordinates and velocities. The program can calculate the eigenvalues and perform modal analysis by means of tracking dynamic modes of rail vehicles [7].

Using the A'GEM program, we can obtain the eigenvalues for the 17 DOF rail vehicle model for the nominal design parameters. Table 4.2 lists the eigenvalues obtained from A'GEM model and those from the hand-derived model.

To compare the eigenvalues obtained from A'GEM with those from the hand-derived model, the spin creep coefficients in the A'GEM model are set to zero. The forward speed of the vehicle corresponding to the eigenvalues shown in Table 4.2 is $11.0[m/s]$. As shown in the table, except for the 33^{rd} and 34^{th} eigenvalues, the results are in close agreement.

Table 4.2: Eigenvalues for the 17 DOF rail vehicle model

	A'GEM Model				*Hand Derived Model*			
Root	Real	Imag	F/Hz	Damp%	Real	Imag	F/Hz	Damp%
1	-2182.6	0.00	0.00	100.00	-2182.6	0.00	0.00	100.00
2	-2182.6	0.00	0.00	100.00	-2182.6	0.00	0.00	100.00
3	-2182.7	0.00	0.00	100.00	-2182.7	0.00	0.00	100.00
4	-2182.7	0.00	0.00	100.00	-2182.7	0.00	0.00	100.00
5	-1162.0	0.00	0.00	100.00	-1162.0	0.00	0.00	100.00
6	-1162.0	0.00	0.00	100.00	-1162.0	0.00	0.00	100.00
7	-1162.0	0.00	0.00	100.00	-1162.0	0.00	0.00	100.00
8	-1162.0	0.00	0.00	100.00	-1162.0	0.00	0.00	100.00
9,10	-16.761	±193.31	30.766	8.64	-16.761	±193.31	30.766	8.64
11,12	-16.761	±193.31	30.766	8.64	-16.761	±193.31	30.766	8.64
13,14	-18.369	±68.576	10.914	25.87	-18.039	±69.030	10.9865	25.28
15,16	-18.119	±68.712	10.936	25.50	-18.119	±68.713	10.936	25.50
17,18	-36.232	±46.049	7.329	61.83	-36.249	±45.713	7.276	62.13
19,20	-36.250	±45.713	7.276	62.13	-36.912	±43.978	6.999	64.29
21,22	-21.265	0.00	0.00	100.00	-21.281	0.00	0.00	100.00
23,24	-10.973	0.00	0.00	100.00	-10.956	0.00	0.00	100.00
25,26	-3.813	±5.343	0.850	58.09	-3.360	±5.261	0.837	53.83
27,28	-4.006	±4.574	0.728	65.89	-4.006	±4.574	0.728	65.88
29,30	-1.315	±5.109	0.813	24.92	-1.304	±5.103	0.812	24.76
31,32	-1.410	±5.174	0.823	26.301	-1.385	±5.195	0.827	25.76
33,34	-0.588	±3.684	0.586	15.77	-0.706	±3.903	0.621	17.81

The small difference in the results may be due to the gravitational stiffnesses that are included by the A'GEM program.

4.2.3 Dynamic Mode Tracking (DMT) Technique

In order to analyze the results of a linear stability analysis of a rail vehicle, in the A'GEM program the DMT technique is used to track the natural modes as the velocity changes and produce plots of frequency and damping of the modes versus vehicle speed[7, 136]. For convenience, this technique is outlined here.

It is well-known that for a real symmetric matrix $\mathbf{B}$, the eigenvalue problem can be expressed as

$$\mathbf{B}\mathbf{x} = \lambda\mathbf{x} \tag{4.6}$$

where λ is an unknown eigenvalue and $\mathbf{x}$ is the corresponding eigenvector. For this eigenvalue problem, any two eigenvectors, e.g. $\mathbf{x}_i$ and $\mathbf{x}_j$, have the following orthogonality relationships:

$$\begin{cases} \mathbf{x}_i^T\mathbf{x}_j = 0 & i \neq j \\ \mathbf{x}_i^T\mathbf{x}_j \neq 0 & i = j \end{cases} \tag{4.7}$$

The problem concerned is, at a given speed, to identify a mode that is known at a previous speed. Thus, based on the orthogonality relationships, for a symmetric system, when two speeds are close, we could determine whether two modes are the same by evaluating and inspecting the size of the dot product obtained by the way as shown in equation (4.7).

Unfortunately, the matrix $\mathbf{A}$ defined in equation (4.4) is unsymmetric due to the non-conservative forces between the wheels and rails. Hence, the above orthogonality relationships does not hold for matrix $\mathbf{A}$.

However, in the case where matrix $\mathbf{A}$ is real but not symmetric, a pair of related eigenvalue problems can arise as follows [100]:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{4.8}$$

$$\mathbf{A}^T\mathbf{x} = \beta\mathbf{x} \tag{4.9}$$

and (4.9) can be rewritten as

$$\mathbf{y}^T\mathbf{A} = \beta\mathbf{y}^T \tag{4.10}$$

where the vectors $\mathbf{x}$ and $\mathbf{y}$ are the so-called right and left eigenvectors of $\mathbf{A}$, respectively. The equations $\mid \mathbf{A} - \lambda\mathbf{I} \mid$ and $\mid \mathbf{A}^T - \beta\mathbf{I} \mid$ should have the same solutions for λ and β because the determinant of a matrix and the determinant of its transpose are equal.

Hence the eigenvalues of $\mathbf{A}$ and $\mathbf{A}^T$ are identical. For the eigenvectors $\mathbf{x}$ and $\mathbf{y}$, they are bi-orthogonal: if λ_i, $\mathbf{x}_i$, $\mathbf{y}_i$ and λ_j, $\mathbf{x}_j$, $\mathbf{y}_j$ are solutions that satisfy (4.8) and (4.9), then

$$\begin{cases} \mathbf{y}_i^T \mathbf{x}_j = 0 & i \neq j \\ \mathbf{y}_i^T \mathbf{x}_j \neq 0 & i = j \end{cases} \tag{4.11}$$

The bi-orthogonal relationships of eigenvectors of the unsymmetric matrix $\mathbf{A}$ are useful for tracking the modes from speed to speed. A procedure for tracking the modes has been offered by Anderson [7]

4.3 Optimization Problem and Implementation

4.3.1 Objective Function, Constraints and Design Variables

Based on equation (4.5), we can conclude that if the real part of each eigenvalue μ_i, $i = 1, 2, ..., 34$, is negative or zero, the response of the system in the time domain is stable. Otherwise, the response will be unstable. If one eigenvalue has zero real part and all others have negative real parts, the vehicle is traveling at its critical speed.

By analyzing matrix $\mathbf{A}$ (4.4) and equation (4.5), we may find the relationship between the critical speed V_c and the system design variables. These variables may include the suspension stiffness and damping, the inertia properties of the wheelsets, bogies, and carbody, the various geometric variables listed in Table A.2, in Appendix A, the creep coefficients f_{11}, f_{22}, f_{23}, and f_{33}, conicity λ, and forward vehicle speed V. Considering that many previous researchers have investigated the effects of creep coefficients and conicity on the critical speed [30, 54, 63, 92, 107, 140], only the suspension stiffness and damping coefficients, the inertial property parameters, and the geometric parameters are chosen here as design variables. The values of all other parameters are fixed. Thus, the objective function and constraints may be expressed as

$$\begin{cases} \textit{maximize} & V_c(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, V) \\ \textit{subject to} & Re(\mu_i)(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, V) \leq 0, \quad \text{i=1,2,...,34} \end{cases} \tag{4.12}$$

where $\bar{\mathbf{S}}$, $\bar{\mathbf{I}}$ and $\bar{\mathbf{G}}$ represent the suspension stiffness and damping coefficient vector, the inertia property parameter vector, and the geometric parameter vector, respectively.

4.3.2 Implementation of the Optimization Problem

A two-loop method is used for the optimization problem (4.12). The two-loop method, as shown in Figure 4.2, is implemented using the MechaGen program [10] (i.e. a GA) as the outer loop algorithm, and the E04UCF routine alone (i.e. an SQP) from the NAG (Numerical Algorithms Group) library or the E04UCF routine combined with MTRACK (i.e. DMT) from A'GEM as the interior loop algorithm.

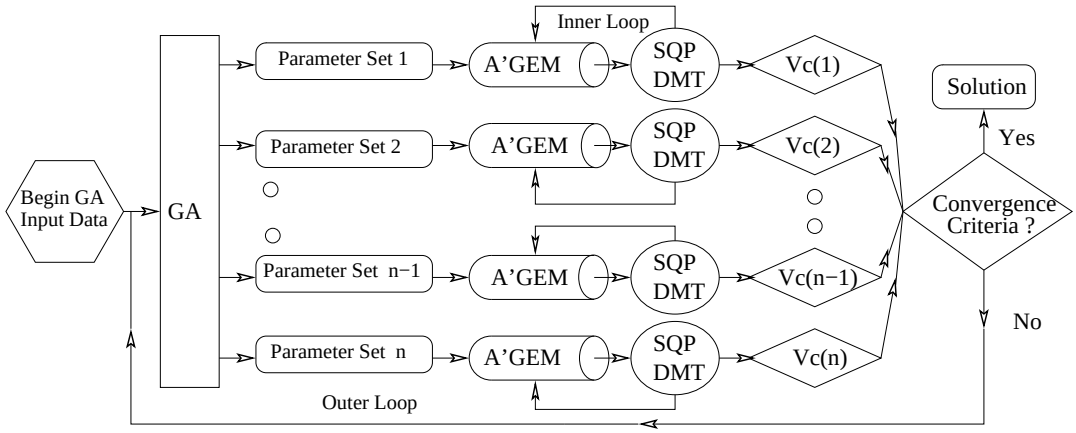

Figure 4.2: The implementation of the GA combined with A'GEM, SQP, and DMT

As shown in Figure 4.2, for a given set of design variables $\bar{\mathbf{S}}$, $\bar{\mathbf{I}}$, and $\bar{\mathbf{G}}$, the corresponding critical speed is determined in the interior loop by the SQP or by the combination of the SQP and DMT and returned to the GA. Using the returned value as the required cost value, the GA in the outer loop produces the next generation of design variable sets using reproduction, crossover and mutation. Then each specified set of design variables is forwarded to A'GEM for deriving the system matrix $\mathbf{A}$ and performing an eigenvalue analysis. Then the SQP or the combination of the SQP and DMT is called one by one. This procedure is repeated until the maximized critical speed is found for a set of optimal design variables.

For the purpose of comparing with the GA used in the integrated optimization algorithm, the Simplex routine E04CCF from NAG is used to replace the MechaGen routine as the outer loop algorithm.

4.3.3 Critical Speed Identification and Algorithm Implementation

In the above two-loop method shown in Figure 4.2, provided that in the interior loop, all the system parameters are known except the vehicle speed, for the SQP algorithm, the objective function and constraint functions contain only one independent design variable, i.e. the forward speed V. Since only oscillatory modes with frequency and damping ratio below specified values are concerned and only these modes are tracked by DMT routine (based on MTRACK from A'GEM), the corresponding eigenvalues, say $i = 1, 2, ..., m$ and $m \leq 34$, are taken into account instead of all 34 eigenvalues. With these considerations, when the DMT is introduced, the SQP solves the optimization problem:

$$\begin{cases} minimize & F(V) = -V_c(V) = -V \\ subject\ to & Re(\mu_i)(V) \leq 0, \quad \text{i=1,2,...,m} \end{cases} \tag{4.13}$$

Notice that if the SQP algorithm alone is used for identifying the critical speed, in equation (4.13), m still takes the number of 34.

For this case, the first-order KT condition at a local optimum V^* and the gradient of the objective function with respect to the design variable V are of the forms:

$$\bigtriangledown F(V^*) + \sum_{i \in active} u_i \bigtriangledown Re(\mu_i)(V^*) = 0 \tag{4.14}$$

$$\bigtriangledown F(V) = -1 \tag{4.15}$$

For the active inequality constraints, the Lagrange multipliers are $u_i \geq 0$.

The numerical algorithm implementation (using the SQP and DMT) for the critical speed identification is shown in Figure 4.3. The numerical algorithms used in the implementation are written in Fortran. The vehicle system parameters are offered by an outer loop algorithm, e.g. a genetic algorithm. With the given system parameters, the A'GEM software program is utilized to generate the system mass, stiffness, and damping matrices, i.e. $\mathbf{M}$, $\mathbf{K}_0$, and $\mathbf{C}_0$, respectively. Then the non-conservative forces or the creep forces between the wheels and rails are added to the stiffness and damping matrices to form the resulting stiffness matrix $\mathbf{K}$ and damping matrix $\mathbf{C}$ accordingly. To improve the efficiency

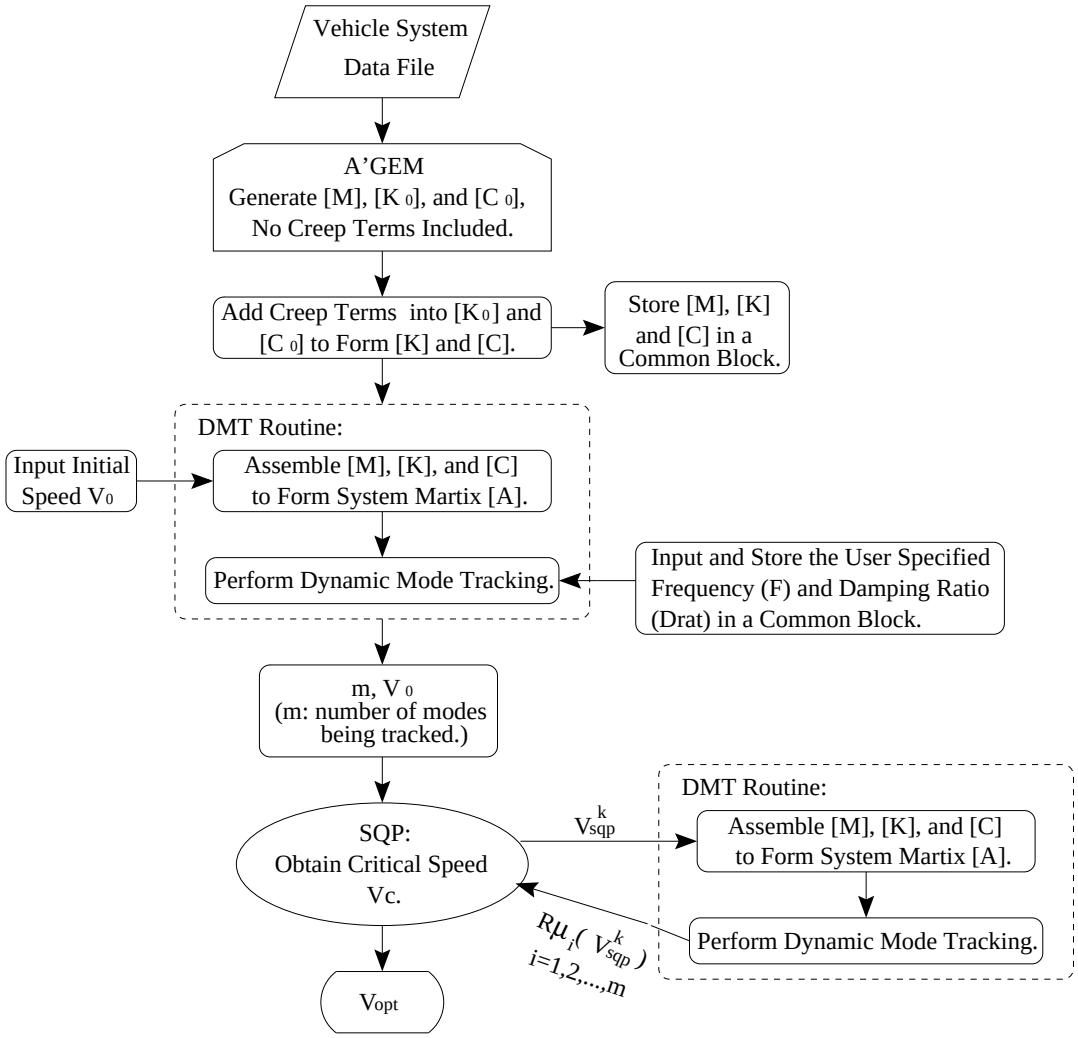

Figure 4.3: The implementation of the SQP combined with DMT

of the calculation, the matrices **M**, **K**, and **C** are stored in a common block for later use. Once the initial vehicle forward speed V_0 is given, the routine DMT assembles matrices **M**, **K**, **C** to form the system matrix **A** defined in (4.4). With the given initial speed V_0 and the specified frequency $Freq$ and damping ratio $Drat$, the DMT routine can determine the number of dynamic modes, i.e. m, of which the DMT routine will track. When the initial speed V_0 and the number of nonlinear constraints, i.e. the number of dynamic modes m,

are offered, the SQP routine with the objective function and constraint functions defined in (4.13) will be used to identify the required critical speed. The SQP routine calls the DMT program repeatedly until it reaches the resulting speed V_{opt} which is assumed to be the critical speed V_c.

In the DMT routine, the system matrix **A** is first balanced and then reduced to an upper Hessenberg form using Householder method. The eigenvalues and eigenvectors of the Hessenberg matrix are calculated using the QR algorithm. The eigenvectors of the Hessenberg matrix are back-transformed to give the eigenvectors of the original matrix **A**.

Notice that the algorithm implementation using the SQP alone for identifying the critical speed is almost the same as described above except that the DMT routine is not included. The difference between the results of the two will be offered in the following section.

4.4 Results and Discussion

4.4.1 Validation of SQP and DMT for Identifying the Critical Speed

In this section, to investigate the efficacy of the algorithms for identifying the critical speed, the numerical results based on the SQP alone are compared with those based on the algorithm combining the SQP and DMT under two different circumstances.

In the first case, the vehicle system parameters taking the nominal values listed in Table A.2 in Appendix A, we can find the 7 least damped motion modes plotted in the form of mode damping versus the forward speed of the vehicle as shown in Figure 4.4. These modes are the Leading/Trailing Bogie Lateral (In Phase) motion, Leading/Trailing Bogie Lateral (Out Phase) motion, Leading O.B. (out board)/Trailing I.B. (in board) Wheelset Lateral (In Phase) motion, Leading O.B./Trailing I.B. Wheelset Lateral (Out Phase) motion, Carbody Lateral motion, Leading Bogie Yaw motion, and Trailing Bogie Yaw motion. Plotted in Figure 4.4, the zero damping line intersects the curves corresponding to the motion modes 3 and 4 at points A and B, respectively. The points A and B are very close. When the vehicle speed is higher than the speed of V_A (corresponding to point A), the damping

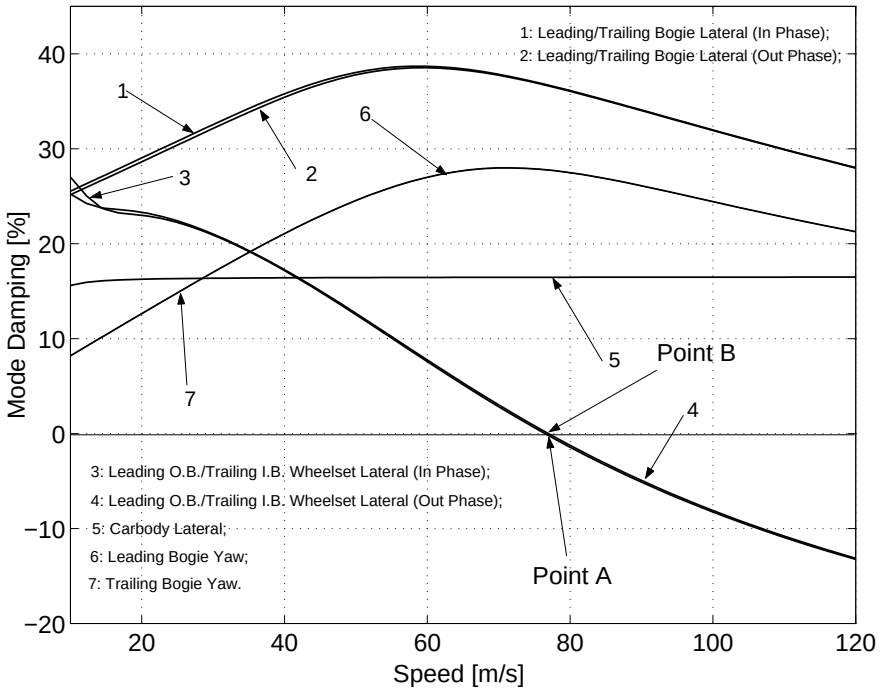

Figure 4.4: Mode damping ratios versus speed (single unstable range case, $k_{2y} = 1.97 \cdot 10^5$ $[N/m]$)

is negative and the vehicle is not stable. Over the speed span offered in Figure 4.4, there is a single unstable range, i.e. $V > V_A$. We call this case the single unstable range case.

In the second case, the lateral spring (in the secondary suspension) stiffness coefficient k_{2y} takes the value of $2.3E+6[N/m]$, with all other parameters taking their nominal values. Again, Figure 4.5 offers the 7 least damped motion modes, i.e. motion modes denoted by 1 to 7. As shown in Figure 4.5, the zero damping line intersects the curves corresponding to the relevant motion modes at points A, B, C and D. The points C and D are very close to one another. Over the speed span offered in Figure 4.5, there are two unstable ranges, i.e. $V_A < V < V_B$ and $V > V_C$. We call this case the double unstable range case.

In both cases, the specified frequency $Freq$ and damping ratio $Drat$ take the values of $60Hz$ and 45%, respectively. Both Figure 4.4 and Figure 4.5 are obtained using the DMT technique.

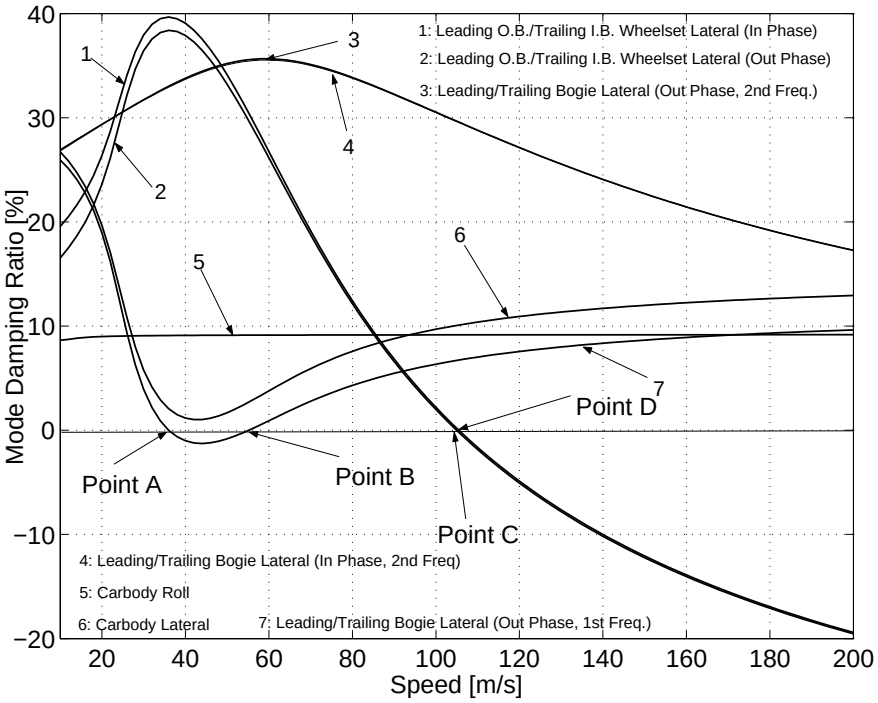

Figure 4.5: Mode damping ratios versus speed (double unstable range case, $k_{2y} = 2.3 \cdot 10^6\ [N/m]$)

Identification of the Critical Speed for Single Unstable Range Case

Based on (4.13), (4.14), and (4.15), one can find that both points A and B in Figure 4.4 are KT points and at point A the critical speed V_c of $76.469[m/s]$ is determined.

Offered in Table 4.3 are the calculation results for determining the critical speed using the algorithm combining the SQP and DMT when different values are selected for the initial speeds. The results based on the SQP alone are also illustrated in Table 4.3. For both of the cases, the Householder's transformation and the QR algorithm are used to solve the eigenvalue problem.

Table 4.3 shows that with the initial speed (V_0) offered, the algorithm combining the SQP and DMT can find the critical speed without exception. However, the SQP alone sometimes gets trapped at certain points which are not even the KT points. For example, with the initial speed selected as $16.5[m/s]$, the resulting optimal speed V_{opt} is $50.273[m/s]$

Table 4.3: Numerical results for single unstable range case

		SQP+DMT			*SQP*		
Run No.	V_0 $[m/s]$	V_{opt} $[m/s]$	*KT Point?*	*Critical Speed?*	V_{opt} $[m/s]$	*KT Point?*	*Critical Speed?*
1	10.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
2	14.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
3	16.5	76.469	*yes*	*yes*	50.273	*no*	*no*
4	17.0	76.469	*yes*	*yes*	29.810	*no*	*no*
5	18.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
6	22.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
7	26.0	76.469	*yes*	*yes*	51.265	*no*	*no*
8	29.0	76.469	*yes*	*yes*	77.509	*no*	*no*
9	31.0	76.469	*yes*	*yes*	32.251	*no*	*no*
10	31.5	76.469	*yes*	*yes*	76.469	*yes*	*yes*
11	32.0	76.469	*yes*	*yes*	32.821	*no*	*no*
12	32.5	76.469	*yes*	*yes*	32.821	*no*	*no*
13	34.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
14	48.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
15	56.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
16	68.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
17	75.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
18	80.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
19	96.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*
20	106.0	76.469	*yes*	*yes*	76.469	*yes*	*yes*

instead of the critical speed of $76.469[m/s]$.

To investigate why the SQP algorithm sometimes gets trapped at certain points which are not even the KT points, we arbitrarily chose and plotted only 4 motion modes in terms of mode damping versus the forward speed as shown in Figure 4.6. It should be noted that the 4 motion modes shown in Figure 4.6 are obtained using the Householder's transformation and the QR algorithm. These motion modes correspond approximately to the modes 1 to 4 from Figure 4.4.

Compared with the corresponding motion modes shown in Figure 4.4, except for mode

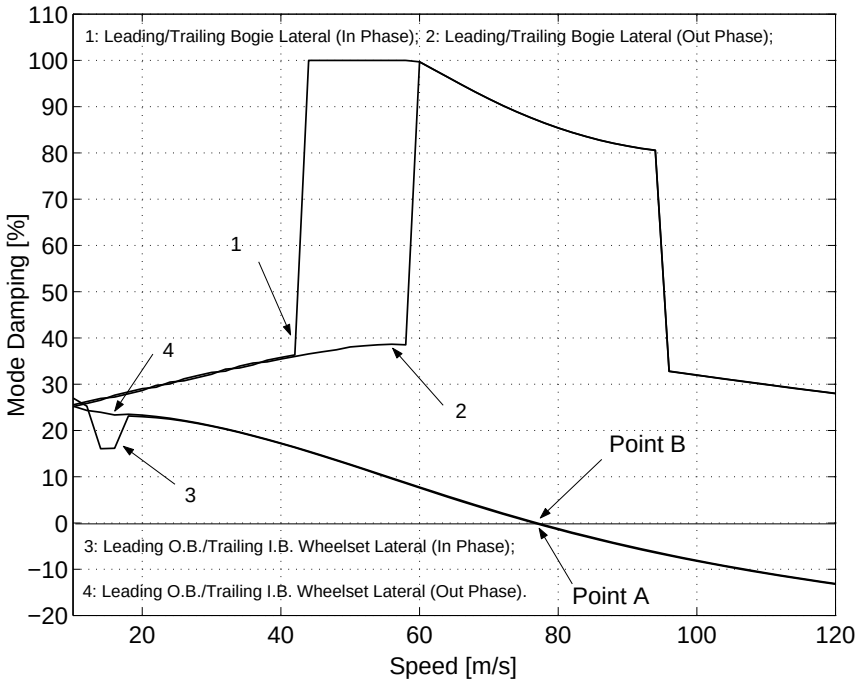

Figure 4.6: Mode damping ratios versus speed

4, the other motion modes shown in Figure 4.6 distort their actual motion modes and these distorted modes become discontinuous over the corresponding speed ranges. These distortions result from the motion mode rank exchanges because the Householder's transformation and QR algorithm ranks the eigenvalues in its own way instead of tracking these eigenvalues once the speed increases.

As mentioned previously, for the SQP algorithm, the objective function and constraint functions should be at least twice-continuously differentiable. Therefore, if a dominant motion mode in terms of mode damping versus the forward speed distorts, the corresponding constraint function becomes discontinuous and the matrix of constraints in the working set is ill-conditioned. During the numerical calculation, once the problem occurs, the E04UCF routine often offers the error information:"Current point cannot be improved upon". During the final line search a sufficient decrease in the merit function could not be attained and the calculation gets trapped at a point that does not satisfy the first-order KT conditions.

Close observations of the results offered in Table 4.3 disclose that even though the SQP alone sometimes gets trapped at certain points which are not corresponding to the critical speed because of the distortions of the relevant motion modes, compared with the initial speed of V_0, the calculated resulting speed V_{opt} is always closer to the critical speed. Moreover, if the initial speed is selected higher than $34.0[m/s]$, the calculation always converges to the critical speed. Numerical experiments show that for the non-convergence case, by adding an appropriate small speed increase to the resulting speed V_{opt}, multiple SQP runs can still find the required critical speed. All these phenomena are also found in the double unstable range case. These phenomena will be discussed in more detail in the following subsection.

Identification of the Critical Speed for Double Unstable Range Case

According to (4.13), (4.14), and (4.15), we find that points A, C, and D shown in Figure 4.5 are KT points and at point A the critical speed V_c of $35.969[m/s]$ is determined. The speed corresponding to point C takes the value of $104.957[m/s]$. The optimization problem corresponding to Figure 4.5 is a nonconvex optimization problem and the points A, C, and D are local optimal points. The objective is to identify the KT point A. Listed in Table 4.4 are the results of numerical experiments for determining the critical speed using the combined algorithm (SQP+DMT) and the SQP algorithm alone, for different values of the initial speed.

Examining the numerical results for the SQP alone, one sees that the calculated speed V_{opt} may or may not converge to the KT point A when the initial speed V_0 is less than the critical speed $V_c(35.9690[m/s])$. As mentioned in the single unstable range case, even if the SQP does not converge to V_c, the calculated speed V_{opt} is closer to the critical speed than V_0. In this situation, the calculated speed V_{opt} does not satisfy the first-order KT condition, and no improved point for the objective function could be found during the final line search. As an example, when the initial speed V_0 is $16.6[m/s]$, the calculated speed V_{opt} is $17.6899[m/s]$; the corresponding values of the real part of each eigenvalue (the constraint functions) are listed in Table 4.5. With the chosen nonlinear search tolerance (i.e. the constraint function value tolerance) of $5.43E-6$, no constraint function value could be considered as 0.0. Among the 34 constraint function values, the value which is closest

Table 4.4: Numerical results for double unstable range case

		SQP+DMT			*SQP*		
Run No.	V_0 $[m/s]$	V_{opt} $[m/s]$	*KT Point?*	*Critical Speed?*	V_{opt} $[m/s]$	*KT Point?*	*Critical Speed?*
1	1.0	104.975	*yes*	*no*	2.036	*no*	*no*
2	2.0	104.957	*yes*	*no*	2.143	*no*	*no*
3	3.0	104.957	*yes*	*no*	3.075	*no*	*no*
4	4.0	104.957	*yes*	*no*	35.969	*yes*	*yes*
5	6.0	104.957	*yes*	*no*	35.969	*yes*	*yes*
6	7.0	104.957	*yes*	*no*	32.733	*no*	*no*
7	9.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
8	10.0	104.957	*yes*	*no*	35.969	*yes*	*yes*
9	11.0	35.969	*yes*	*yes*	32.055	*no*	*no*
10	13.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
11	14.0	35.969	*yes*	*yes*	31.928	*no*	*no*
12	15.0	35.969	*yes*	*yes*	30.206	*no*	*no*
13	16.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
14	16.6	35.969	*yes*	*yes*	17.690	*no*	*no*
15	17.0	35.969	*yes*	*yes*	33.738	*no*	*no*
16	18.0	35.969	*yes*	*yes*	35.256	*no*	*no*
17	20.0	35.969	*yes*	*yes*	31.897	*no*	*no*
18	22.0	35.969	*yes*	*yes*	33.546	*no*	*no*
19	24.0	35.969	*yes*	*yes*	35.942	*no*	*no*
20	25.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
21	26.0	35.969	*yes*	*yes*	33.488	*no*	*no*
22	27.0	35.969	*yes*	*yes*	33.523	*no*	*no*
23	29.0	35.969	*yes*	*yes*	35.942	*no*	*no*
24	29.5	35.969	*yes*	*yes*	34.693	*no*	*no*
25	29.8	35.969	*yes*	*yes*	35.969	*yes*	*yes*
26	30.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
27	31.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
28	32.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
29	33.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
30	34.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
31	34.5	35.969	*yes*	*yes*	35.969	*yes*	*yes*

		SQP+DMT			*SQP*		
Run No.	V_0 $[m/s]$	V_{opt} $[m/s]$	*KT Point?*	*Critical Speed?*	V_{opt} $[m/s]$	*KT Point?*	*Critical Speed?*
32	36.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
33	37.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
34	38.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
35	39.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
36	39.6	35.969	*yes*	*yes*	35.969	*yes*	*yes*
37	39.8	35.969	*yes*	*yes*	33.970	*no*	*no*
38	40.0	35.969	*yes*	*yes*	33.360	*no*	*no*
39	41.0	35.969	*yes*	*yes*	35.969	*yes*	*yes*
40	42.0	35.969	*yes*	*yes*	35.068	*no*	*no*
41	44.0	35.969	*yes*	*yes*	33.970	*no*	*no*
42	47.0	104.957	*yes*	*no*	47.152	*no*	*no*
43	49.0	104.957	*yes*	*no*	54.358	*no*	*no*
44	53.8	104.957	*yes*	*no*	54.359	*no*	*no*
45	53.9	104.957	*yes*	*no*	104.957	*yes*	*no*
46	54.0	104.957	*yes*	*no*	104.957	*yes*	*no*
47	57.0	104.957	*yes*	*no*	104.957	*yes*	*no*
48	80.0	104.957	*yes*	*no*	104.957	*yes*	*no*

to 0.0 is -0.3702. Therefore, at the speed of $17.6899[m/s]$, no active constraint exists. The second term on the left-hand side of equation (4.14) disappears and this equation can not hold, i.e. the resulting speed $V_{opt}(17.6899[m/s])$ does not satisfy the first-order KT condition. Along with the eigenvalues at the speed of $17.6899[m/s]$, the eigenvalues at the corresponding initial speed of $16.6[m/s]$ are also offered in Table 4.5. Results for initial speeds of $16.0[m/s]$, $49.0[m/s]$, and $80.0[m/s]$, and their corresponding calculated speeds, are also provided for comparison purposes.

If the initial speed is selected as $16.0[m/s]$, the corresponding resulting speed V_{opt} is $35.9690[m/s]$. As we can see from the Table 4.5, the upper bound of the constraint function values at the speed of $35.9690[m/s]$ is $-1.78E-12$, which is within the nonlinear search tolerance mentioned above. Thus, the constraint corresponding to the upper bound value is an active constraint. Then, equation (4.14) holds and the resulting speed $35.9690[m/s]$

Table 4.5: Real parts of eigenvalues at different speeds (using the SQP alone)

	V_0	V_{opt}	V_0	V_{opt}	V_0	V_{opt}	V_0	V_{opt}
	$[m/s]$	$[m/s]$	$[m/s]$	$[m/s]$	$[m/s]$	$[m/s]$	$[m/s]$	$[m/s]$
Root	16.6	17.6899	16.0	35.9690	49.0	54.3582	80.0	104.9566
1	-768.33	-1332.9	-1481.8	-582.93	-367.39	-259.59	-67.607	-62.490
2	-1425.5	-1333.4	-1481.8	-590.36	-367.39	-259.59	-67.607	-62.490
3	-768.33	-720.51	-797.41	-582.93	-304.98	-259.59	-67.607	-62.490
4	-1426.0	-1332.9	-1482.2	-590.36	-304.98	-259.59	-67.607	-62.490
5	-1425.5	-720.51	-797.41	-349.97	-304.98	-272.84	-159.94	-120.71
6	-1426.0	-1333.4	-1482.2	-349.97	-304.98	-272.84	-159.94	-120.71
7	-768.33	-720.51	-797.41	-349.97	-264.99	-272.84	-159.94	-120.71
8	-768.33	-720.51	-797.41	-349.97	-264.99	-272.84	-159.94	-120.71
9,10	-21.710	-22.684	-21.176	-39.496	-51.510	-56.059	-148.33	-96.59
11,12	-21.710	-22.684	-21.176	-39.496	-51.510	-56.059	-148.34	-96.60
13,14	-22.942	-23.330	-22.942	-27.683	-109.110	-115.67	-87.111	-30.439
15,16	-22.932	-23.188	-22.792	-27.645	-30.665	-31.620	-87.096	-30.432
17,18	-33.688	-33.670	-33.698	-75.340	-30.686	-31.654	-33.017	-79.473
19,20	-33.327	-33.315	-33.334	-33.213	-55.982	-67.249	-32.987	-79.436
21,22	-32.432	-34.651	-31.218	-33.461	-33.351	-33.472	-34.255	-34.045
23,24	-16.408	-17.502	-15.813	-37.412	-33.550	-33.667	-34.000	-34.335
25,26	-4.0320	-4.1744	-3.9600	-8.2144	-8.7474	-8.3301	-4.2690	-0.0799
27,28	-3.5105	-3.6570	-3.4366	-7.9581	-8.5814	-8.1987	-4.2030	-8.15E-9
29,30	-1.8211	-1.9145	-1.8354	-0.3014	-0.2524	-0.4179	-1.3174	-1.7516
31,32	-1.8684	-1.8644	-1.7892	-1.78E-12	0.14571	0.02398	-0.7428	-1.1584
33,34	-0.3691	-0.3702	-0.3685	-0.3761	-0.3773	-0.3775	-0.3784	-0.3787

satisfies the first-order KT condition.

As shown in Table 4.5, when the initial speed is chosen as $80.0[m/s]$, the calculated speed V_{opt} is $104.9566[m/s]$. At this speed, the upper bound constraint function values is $-8.15E-9$. Therefore, the speed of $104.9566[m/s]$ also satisfies the first-order KT condition. In fact, this speed corresponds to the KT point C from Figure 4.5. When the initial speed is selected within the speed range from 44.5 to $53.8[m/s]$, a solution from the SQP algorithm may be outside the feasible range. For instance, if the initial speed is

selected as $49.0[m/s]$, one obtains a calculated speed V_{opt} of $54.3582[m/s]$ using the SQP. From Table 4.5, the upper bound of the constraint function values corresponding to the calculated speed is 0.02398. Thus, the solution is not feasible.

Figure 4.7 shows the relationship between the major iteration number and the resulting speed V_{opt} when the initial speed is chosen as 16.0, 16.6, 49.0, and $80.0[m/s]$, respectively.

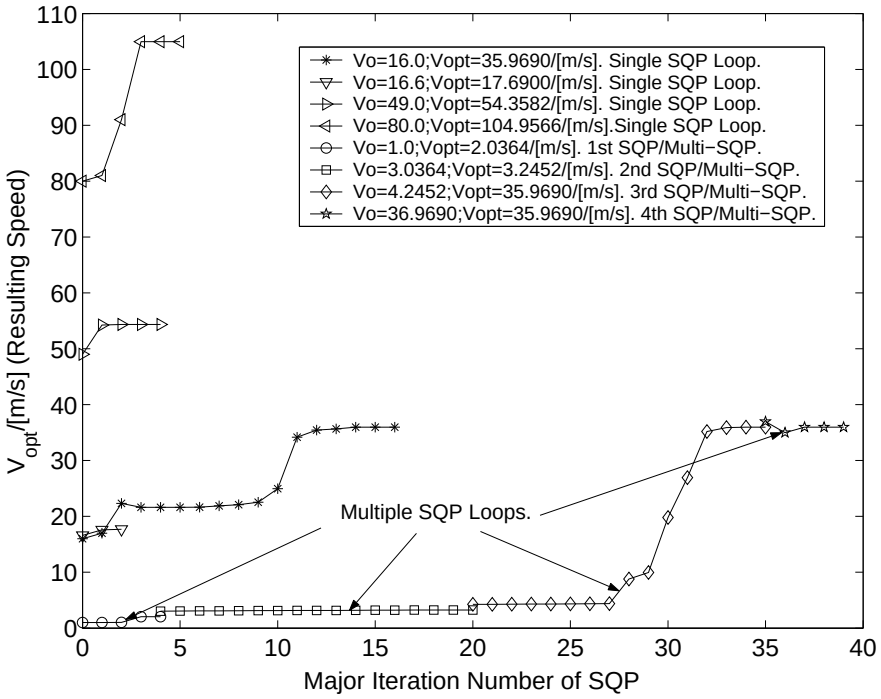

Figure 4.7: V_{opt} versus major iteration number of SQP

When the initial speed is selected in the neighborhood of the critical speed of $35.9690[m/s]$, the calculated speed can be guaranteed to converge to the critical speed. As shown in Table 4.4, if the initial speed selected is within the range from 29.8 to $39.6[m/s]$, the calculated speed always converges to $35.9690[m/s]$. This phenomenon can be explained by the fact that the SQP has very strong, theoretical, local convergence properties and within this speed range there are no distortions of the relevant motion modes.

Therefore, to reliably identify the critical speed using the SQP alone, the initial speed V_0 should be selected in the neighborhood of the critical speed V_c. Unfortunately, for the general case, we do not know beforehand where the critical speed is located. However, based on the above analyses, we know that if the initial speed is chosen less than the critical speed, there are two possibilities for the calculated speed V_{opt}. First, the calculated speed may converge to the critical speed. Second, the resulting speed is still a feasible solution, it should be closer to the critical speed than the corresponding initial speed, and the resulting speed should be less than the critical speed. Thus, if we choose a small initial speed, using the SQP we can obtain an improved resulting speed V_{opt}.

By adding a small speed increase V_{step} to the resulting speed V_{opt}, we can get an increased speed $V_{increase}$. If any of the constraints is violated at the increased speed, it should be higher than the critical speed but still within the neighborhood of the critical speed. If the increased speed is treated as a new initial speed and one more SQP run is carried out, the new calculated speed should converge to the critical speed.

On the other hand, if the constraints are satisfied, the increased speed should be a feasible solution and less than the critical speed. The value of the increased speed is assigned to the initial speed, and the previous procedure is repeated until we reach the critical speed. The flowchart of this multiple SQP loop algorithm is shown in Figure 4.8.

As shown in Table 4.4, an initial speed of $1.0[m/s]$ results in a calculated speed of $2.0364[m/s]$ using only one SQP run. However, using multiple SQP loops for the same initial speed, one can obtain the correct critical speed. The relationship between the major iteration number of the multiple SQP loop algorithm and the resulting speed V_{opt} when both the initial speed V_0 and the speed increase V_{step} are chosen as $1.0[m/s]$ is shown in Figure 4.7. To find the critical speed for the case just discussed, four SQP calls are needed. Numerical experiments show that this multiple SQP inner-loop algorithm has a high reliability for identifying the critical speed of the 17-DOF rail vehicle.

To clarify how this multiple SQP inner-loop algorithm works, the following example is further investigated. As illustrated in Figure 4.9, if the initial speed and the speed step are chosen as $15.0[m/s]$ and $1.0[m/s]$ respectively, we can find the critical speed using two SQP runs. For the first SQP run, the resulting optimal speed is $30.206[m/s]$. For the second SQP run, the initial speed is $31.206[m/s]$.

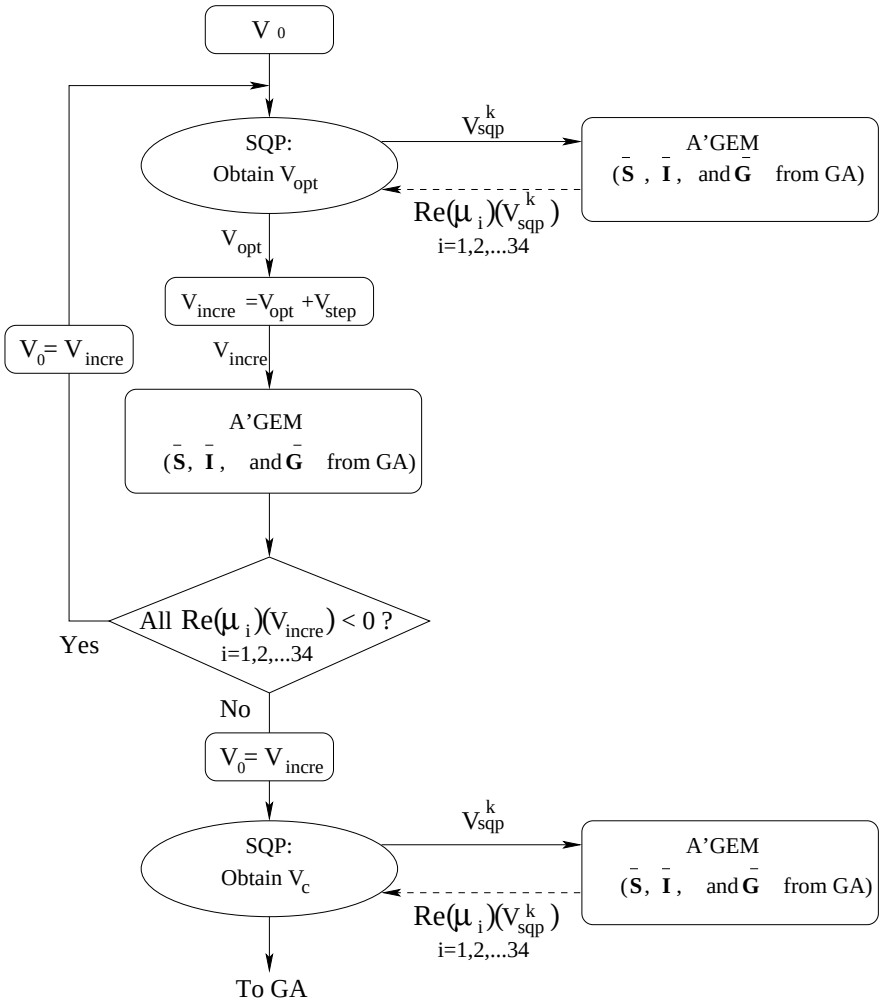

Figure 4.8: Flowchart of interior loop algorithm for identifying V_c

To investigate the SQP's behavior shown in Figure 4.9, the 7 least damped motion modes in terms of mode damping versus speed are offered in Figures 4.10 and 4.11 corresponding to the initial speeds of $15.0[m/s]$ and $31.206[m/s]$, respectively. It should be noted that the 7 motion modes shown in Figure 4.10 may not be the same as those shown in Figure 4.11.

With the initial speed selected as $15.0[m/s]$, the SQP algorithm searches the working space and reaches point Q as shown in Figure 4.10. Around point Q, the dominant motion

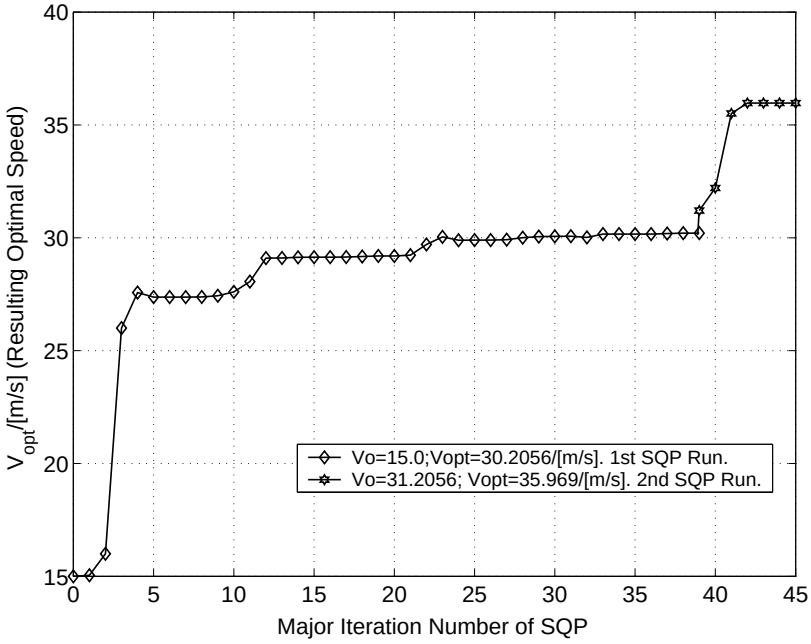

Figure 4.9: V_{opt} versus major iteration number of SQP ($V_0 = 15.0\ [m/s]$, $V_{incre} = 1.0\ [m/s]$)

mode is mode 7. Because this mode is not tracked at point Q and the corresponding motion mode in terms of mode damping versus speed distorts, the dominant constraint function becomes discontinuous. So for the first SQP run as shown in Figure 4.9, during the final line search, point Q does not satisfy the first-order KT conditions offered by equation (4.14), and no improved point for the objective function described as (4.13) could be found. Thus at the end of the first SQP run, the calculation gets trapped at point Q corresponding to the resulting speed of $30.206[m/s]$.

With the resulting optimal speed of the first SQP run plus the speed step as the initial speed for the second SQP run, the SQP algorithm searches the optimal point in the search space. As shown in Figure 4.11, the dominant motion modes should be mode 7, mode 5, and mode 6 respectively over the corresponding speed ranges before the critical speed is determined. During the final line search, point M is found and this point satisfies the first-order KT conditions offered by equation (4.14), and the calculation terminates at point M which corresponds to the critical speed of $35.969[m/s]$.

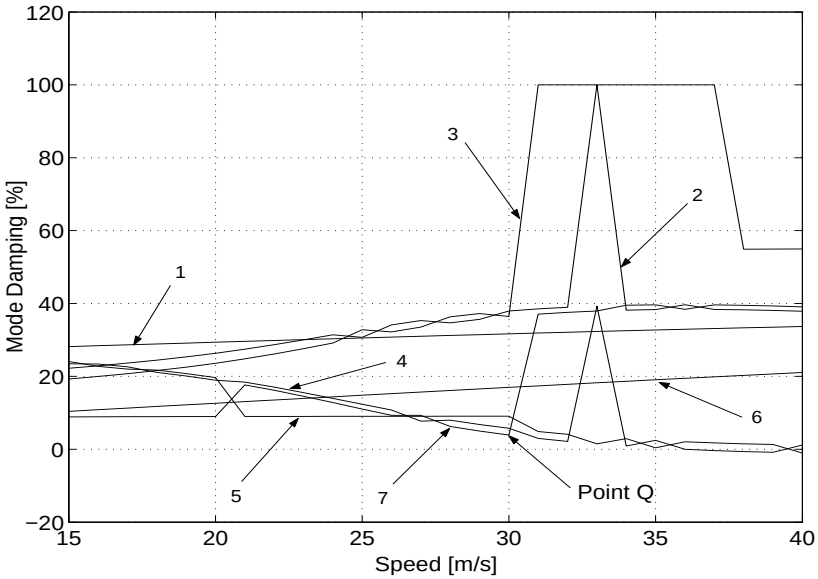

Figure 4.10: Mode damping ratios versus speed ($V_0 = 15.0\ [m/s]$)

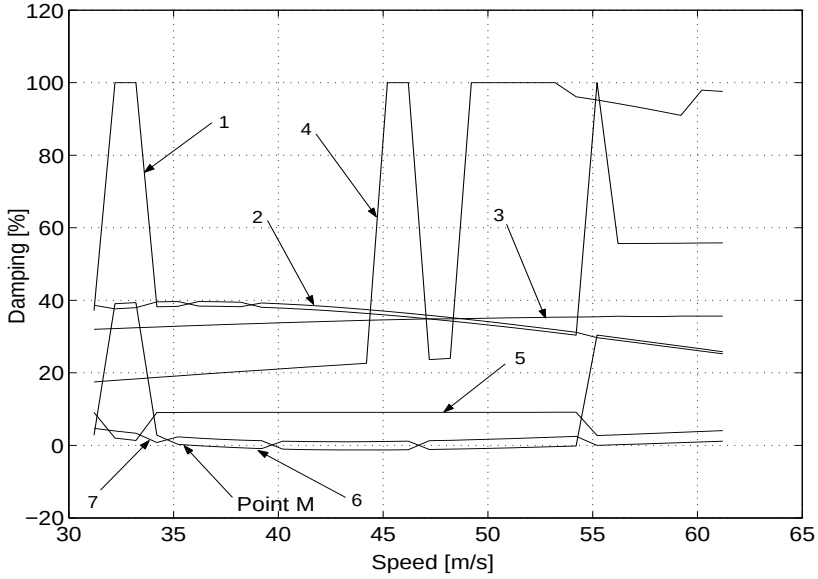

Figure 4.11: Mode damping ratios versus speed ($V_0 = 31.206\ [m/s]$)

Provided in Table 4.4 are also the numerical results for identifying the critical speed using the algorithm combining the SQP with DMT when the initial speeds are selected as different values.

Close observations of the numerical results shown in Table 4.4 reveal that with the initial speeds selected, the resulting optimal speeds V_{opt} always converge to the KT point A or the KT point C (see Figure 4.5) when the algorithm combining SQP with DMT is employed. If the initial speeds are selected within the range from 11.0 to 44.0$[m/s]$, the resulting optimal speeds converge to the critical speed V_c without exception. Once the initial speeds are selected larger than 44.0$[m/s]$ or less than 11$[m/s]$, the corresponding resulting optimal speeds frequently converge to the KT point C.

All these phenomena result from the local convergence properties of the SQP algorithm. To improve the reliability for identifying the critical speed using the combined algorithm including the SQP and DMT, there are two options: the calculation should start from a point sufficiently close to, or in the neighborhood of, the KT point A; with multiple initial speeds selected within the speed span concerned, the critical speed is determined by comparing the resulting optimal speeds corresponding to the initial speeds selected.

Compared with the results based on the algorithm combining the SQP with DMT, for the case of using the SQP alone, the speed range for the initial speeds to ensure the corresponding resulting optimal speeds converge to the critical speed is much narrower. For the latter case, as mentioned previously, this speed range is from 29.8 to 39.6$[m/s]$ which is only 30% of that for the former case. For the latter case, if the initial speeds are selected less than 29.8$[m/s]$ or higher than 39.6$[m/s]$, the SQP algorithm frequently gets trapped at a certain point which is not even the KT point.

4.4.2 Cliff Phenomenon and Interpretation

Figures 4.12 and 4.13 show the relationship between the the critical speed and k_{1x} (the longitudinal stiffness of the primary suspension) and that between the critical speed and k_{2y} (the lateral stiffness of the secondary suspension), with all other system parameters taking nominal values. The two-dimensional plot 4.12 reveals that there are several local optimal points and, when k_{1x} take its nominal value, the corresponding critical speed (76.4688$[m/s]$) is very close to the global optimal value (76.4779$[m/s]$). In addition to

several local optimal points, one can see in Figure 4.13 the existence of a steep "cliff" at which the critical speed changes dramatically with small changes in k_{2y}. This cliff is also clearly shown in the corresponding three-dimensional plot in Figure 4.14. This phenomenon appeared previously in two-dimensional form in 1984 [107], but its significance was not discussed.

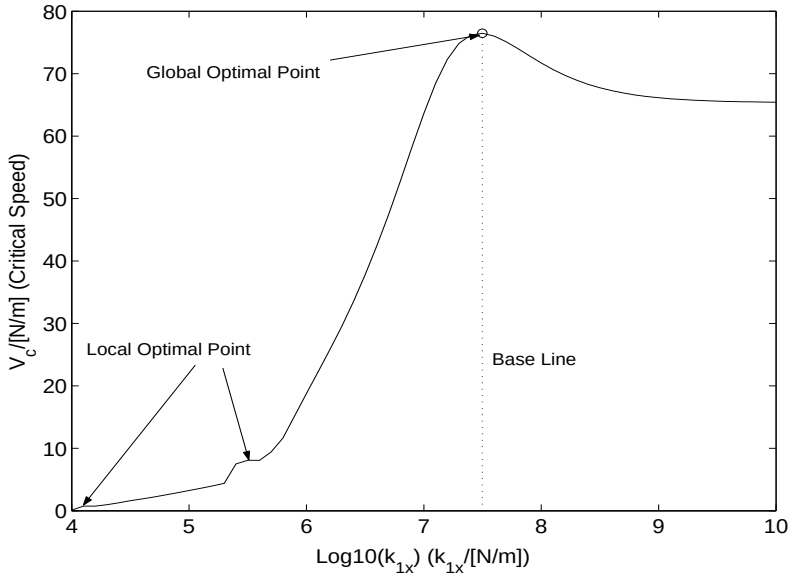

Figure 4.12: Critical speed versus stiffness k_{1x}

The cliffs in Figure 4.14 are due to the fact that the mode (eigenvector) corresponding to the critical speed switches in a discontinuous fashion. This phenomenon can be interpreted clearly using Figure 4.15, in which the parameter k_{2y} takes the value of $1.995E+6[N/m]$ while all the other parameters take their nominal values. In Figure 4.15, the intersection (point A) of the zero damping line and the Leading O.B./Trailing I.B. Wheelset Lateral (Out Phase) motion mode damping curve determines the critical speed; its value is $101.43[m/s]$. However, when k_{2y} increases from $1.995E+6[N/m]$ to $2.300E+6[N/m]$ (see Figure 4.5), the Leading/Trailing Bogie Lateral (Out Phase) motion mode damping curve will now intersect with the zero damping ratio line and the critical speed will plummet to $35.9690[m/s]$ accordingly. This cliff phenomenon might be interpreted as a bifurcation

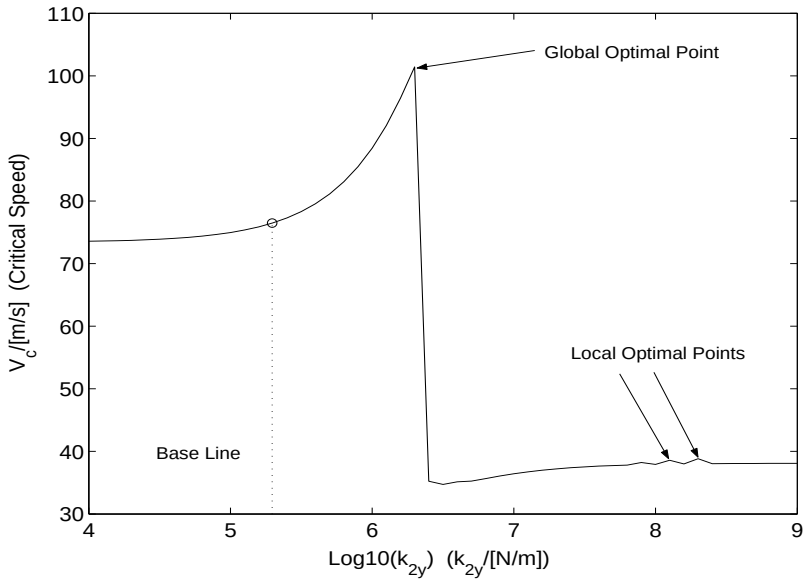

Figure 4.13: Critical speed versus stiffness k_{2y}

in the solution, for which two values of the critical speed correspond to one value of the parameter k_{2y}.

4.4.3 Comparison of Simplex and GA for Critical Speed Optimization

When the Simplex or the GA is used as the outer loop algorithm shown in Figure 4.2, the optimal values of the single design variables k_{1x} and k_{2y} (and the corresponding critical speeds) are listed in Tables 4.6 and 4.7, respectively.

For these two cases involving a single design variable, both the Simplex and the GA can reliably find the global optimum even though the objective functions are discontinuous and have a number of local optimal points (see Figures 4.12 and 4.13). However, the Simplex is more accurate and efficient than the GA.

Using terminology from the literature on GAs [112], the "explorative" property is the ability to explore the whole function space and identify the subdomain in which the global

Table 4.6: Results of optimization ($k_{1x} : 10^4 \sim 10^9/[N/m]$)

Nominal Values		*Simplex Algorithm*		
k_{1x} $[N/m]$	V_c $[m/s]$	Initial k_{1x} $[N/m]$	Optimized k_{1x} $[N/m]$	Maximized V_c $[m/s]$
$3.150000E+7$	76.468754	$3.162377E+4$	$3.039786E+7$	76.477942
		$1.258925E+8$	$3.039834E+7$	76.477942
		$3.981070E+3$	$3.039940E+7$	76.477942
Nominal Values		*Genetic Algorithm*		
k_{1x} $[N/m]$	V_c $[m/s]$	Run No.	Optimized k_{1x} $[N/m]$	Maximized V_c $[m/s]$
$3.150000E+7$	76.468754	1	3.038547E+7	76.477940
$3.150000E+7$	76.468754	2	3.039855E+7	76.477942

Table 4.7: Results of optimization ($k_{2y} : 10^4 \sim 10^8/[N/m]$)

Nominal Values		*Simplex Algorithm*		
k_{2y} $[N/m]$	V_c $[m/s]$	Initial k_{2y} $[N/m]$	Optimized k_{2y} $[N/m]$	Maximized V_c $[m/s]$
$1.970000E+5$	76.468754	$6.309573E+4$	$2.041588E+6$	101.976942
		$1.584893E+7$	$2.041541E+6$	101.976394
		$2.511886E+6$	$2.041588E+6$	101.976464
Nominal Values		*Genetic Algorithm*		
k_{2y} $[N/m]$	V_c $[m/s]$	Run No.	Optimized k_{2y} $[N/m]$	Maximized V_c $[m/s]$
$1.970000E+5$	76.468754	1	$2.035138E+6$	101.901514
$1.970000E+5$	76.468754	2	$2.041567E+6$	101.976692

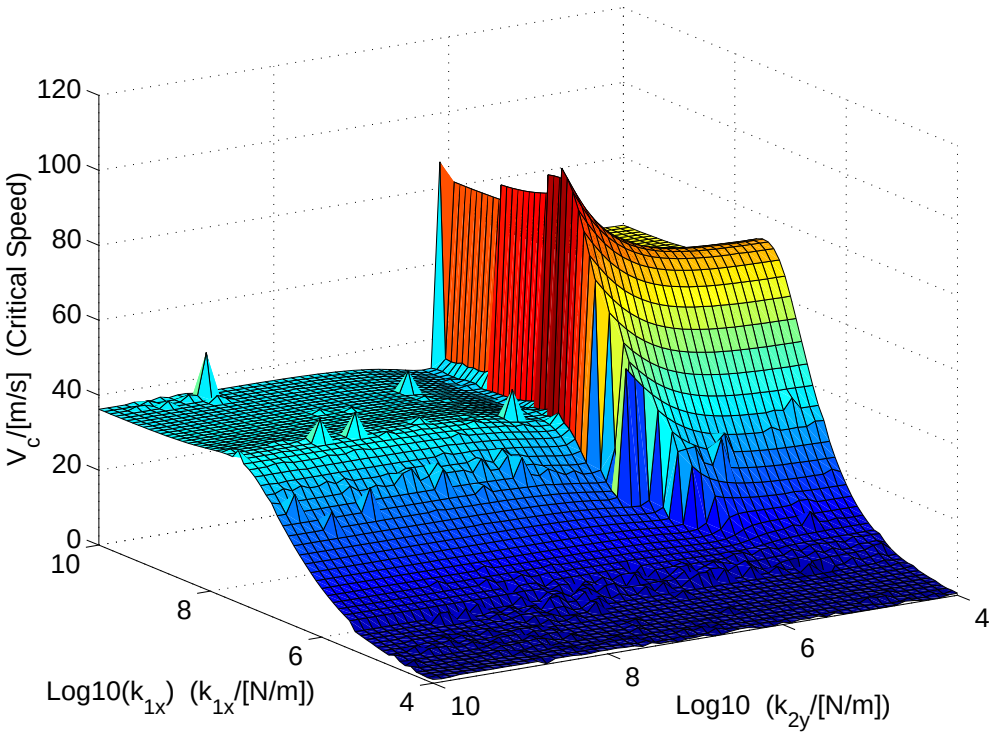

Figure 4.14: Critical speed versus stiffnesses

optimum is located, while the ability to exploit all the local information for refining progressively and efficiently the solution is called the "exploitative" property. From the literature, the Simplex (a local search method) should have better exploitative properties than those of the GA. As shown in Tables 4.6 and 4.7, the results agree with this expectation. However, when the genetic algorithm parameters are chosen properly, the results obtained are very close to those from the Simplex. As an example, when the population size and the mutation ratio are selected as 80 and 0.015 (in Run 2) instead of 40 and 0.01 (in Run 1), respectively, the result of the second run is closer to those from the Simplex than that of the first run, as shown in Table 4.6.

As shown in Table 4.8, when both k_{1x} and k_{2y} are selected as design variables, the Simplex usually converges to the neighborhood of the global optimal point. However, it occasionally gets trapped at local points. In contrast, the GA can find the global optimal

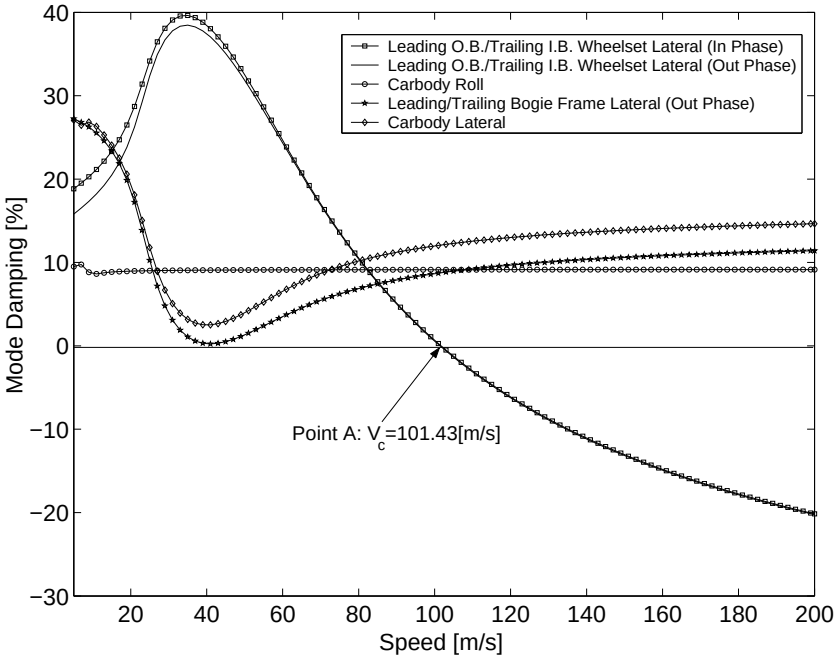

Figure 4.15: Motion mode damping ratios versus speed

point with high reliability. Moreover, as shown in Table 4.9, when k_{1x}, k_{2y} and k_{1y} are all chosen as design variables, the Simplex often gets trapped at local points and only occasionally finds the global optimum. As the number of design variables increases, the number of local optimal points also increases. Therefore, the more the design variables, the more likely that the Simplex will converge to a local optimum instead of the global optimum. Again in contrast, the results show that the number of design variables has no effect on the GA's ability to find the global optimum.

Based on the characteristics of Simplex and GAs mentioned above and analyzed in [112], it is expected that the GA should have better explorative properties than the Simplex. Obviously the results shown in Tables 4.8 and 4.9 are consistent with the expectation.

In order to visualize the optimized results, a plot of motion mode damping ratios versus forward speed for the case of $k_{1x} = 1.53287E + 7[N/m]$, $k_{2y} = 2.94724E + 6[N/m]$ and $k_{1y} = 9.99999E + 8[N/m]$ (see Table 4.9) is offered in Figure 4.16. Based on Figure 4.16,

Table 4.8: Results of optimization ($k_{1x} : 10^4 \sim 10^9; k_{2y} : 10^4 \sim 10^8/[N/m]$)

Initial Values		*Simplex Algorithm*		
k_{1x} $[N/m]$	k_{2y} $[N/m]$	Optimized k_{1x} $[N/m]$	Optimized k_{2y} $[N/m]$	Maximized V_c $[m/s]$
1.00000E+4	1.00000E+4	3.19665E+7	2.03307E+6	101.97620
3.16228E+4	6.30957E+4	2.18999E+7	2.24189E+6	100.31639
3.16228E+4	2.51189E+6	3.21269E+7	2.03017E+6	101.97497
3.16228E+4	1.58489E+7	3.08257E+7	2.05414E+6	101.97001
1.25893E+8	1.58489E+7	3.17328E+7	2.03680E+6	101.97098
1.25893E+8	1.00000E+4	3.14018E+7	2.04337E+6	101.97619
1.00000E+9	1.25893E+7	2.02207E+7	2.27847E+6	99.43702
1.00000E+9	1.00000E+8	3.14152E+7	2.04254E+6	101.96945
5.01187E+7	6.30958E+7	3.10679E+7	2.04913E+6	101.96820
7.94328E+6	1.99526E+6	4.57241E+7	7.12881E+7	38.13778
2.51189E+5	6.30957E+4	3.18438E+7	2.03530E+6	101.97681
1.25893E+4	3.16228E+6	3.16875E+7	2.03815E+6	101.97716
		Genetic Algorithm		
Run No.		Optimized k_{1x} $[N/m]$	Optimized k_{2y} $[N/m]$	Maximized V_c $[m/s]$
1		3.41242E+7	1.98787E+6	101.83254
2		3.13561E+7	2.03927E+6	101.91828
3		3.43900E+7	1.98861E+6	101.88361
4		3.12991E+7	2.02834E+6	101.77798
5		3.04822E+7	2.04572E+6	101.79017
6		3.40126E+7	1.99469E+6	101.89663
7		3.46853E+7	1.98324E+6	101.86394

we can deduce that the optimized point is located on an edge of the cliff that is the same as that shown in Figure 4.14. The reason is that when the curve of the Leading O.B./Trailing

I.B. Wheelset Lateral (Out Phase) motion mode damping ratio intersects with the zero damping line, the curves of the Leading I.B./Trailing O.B. Wheelset Lateral (Out Phase) motion mode damping ratio and the Car Body Lateral motion mode damping ratio are also very close to the zero damping ratio line. Once k_{2y} increases slightly (e.g. 0.68%), the curve of the Leading I.B./Trailing O.B. Wheelset Lateral (Out Phase) motion mode damping ratio intersects with the zero damping line, and the critical speed will drop from $115.56269[m/s]$ to $50.85899[m/s]$ discontinuously. From a design perspective, the critical

Table 4.9: Results of optimization ($k_{1x} : 10^4 \sim 10^9; k_{2y} : 10^4 \sim 10^8; k_{1y} : 10^4 \sim 10^9/[N/m]$)

Initial Values			*Simplex Algorithm (Optimized)*			
k_{1x} $[N/m]$	k_{2y} $[N/m]$	k_{1y} $[N/m]$	k_{1x} $[N/m]$	k_{2y} $[N/m]$	k_{1y} $[N/m]$	V_c $[m/s]$
1.99526E+6	6.30957E+5	5.62341E+6	1.48184E+7	3.02786E+6	9.99996E+8	115.71418
3.16228E+8	1.00000E+5	1.99526E+6	9.99776E+8	1.71973E+6	2.28329E+6	102.85233
1.25893E+8	3.98107E+5	1.58489E+7	1.49328E+7	3.01170E+6	9.99999E+8	115.71824
1.25893E+4	3.16228E+7	3.98108E+8	1.49432E+7	3.01023E+6	9.99907E+8	115.71824
4.46684E+7	1.00000E+6	5.62341E+6	3.79171E+7	1.92954E+6	4.00342E+6	101.64395
8.91251E+4	6.30957E+6	1.25893E+8	1.46263E+7	2.99217E+6	4.85742E+7	113.55293
1.30794E+7	1.54437E+6	1.00000E+6	1.56402E+7	9.99728E+7	9.99999E+8	47.63845
7.74662E+7	4.66745E+5	2.51189E+5	1.48738E+7	3.01649E+6	4.53793E+8	115.58863
7.94328E+6	1.99526E+6	5.62341E+6	1.70687E+7	2.71486E+6	2.76539E+8	114.39930
1.25893E+4	3.16228E+7	1.00000E+5	1.53287E+7	2.94724E+6	9.99999E+8	115.56269
1.00000E+9	1.00000E+8	1.00000E+9	1.00000E+9	1.00000E+8	1.00000E+9	7.56709
3.08257E+7	2.05414E+6	3.96000E+6	1.48139E+7	2.90785E+6	2.42014E+7	111.51704

	Genetic Algorithm (Optimized)			
Run No.	k_{1x} $[N/m]$	k_{2y} $[N/m]$	k_{1y} $[N/m]$	V_c $[m/s]$
1	1.52864E+7	2.93182E+6	6.57839E+8	115.26521
2	1.43278E+7	3.08832E+6	7.26158E+8	115.46558
3	1.59885E+7	2.84442E+6	9.94991E+8	115.16006
4	1.52418E+7	2.95933E+6	2.54195E+8	115.37914
5	1.58708E+7	2.86164E+6	7.69511E+8	115.21562

speed of $115.56269[m/s]$ is not a very useful indicator of lateral stability because a very small disturbance in k_{2y} (e.g. manufacturing tolerances) will cause the critical speed to drop sharply. Because of this instability, we call the critical speed corresponding to the optimized point located on the edges of a cliff a "pseudo-critical speed".

For the purpose of finding a "stable" critical speed in the presence of this cliff phenomenon, it is necessary to redefine the critical speed and adjust the optimization algorithms accordingly.

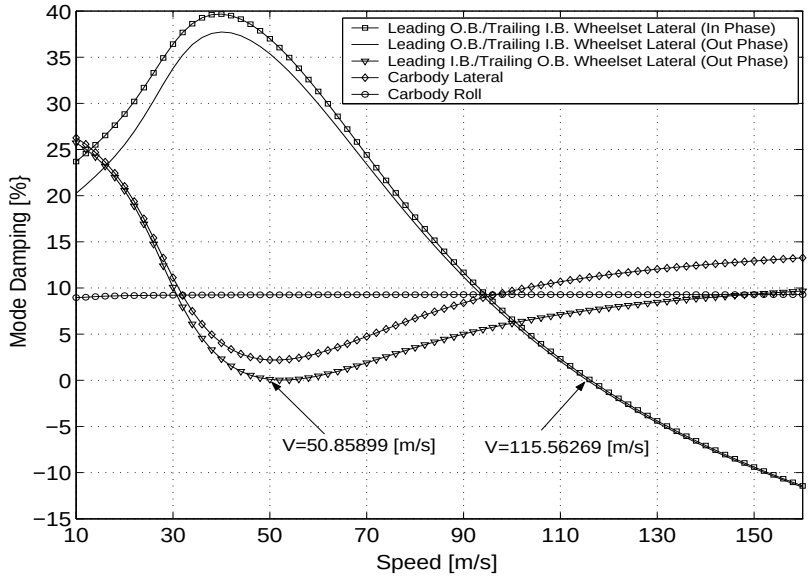

Figure 4.16: Motion mode damping ratios versus speed (conventional definition)

4.4.4 Redefinition of Critical Speed and Adjustment of Algorithms

For safety reasons, the critical speed should be well above the normal operating speed of the vehicle. Moreover, it is further required that all modes have "sufficient damping" over the range of operating speeds [7]. Based on this requirement, the definition of the critical speed can be generalized as follows: when a motion mode reaches the zero damping ratio,

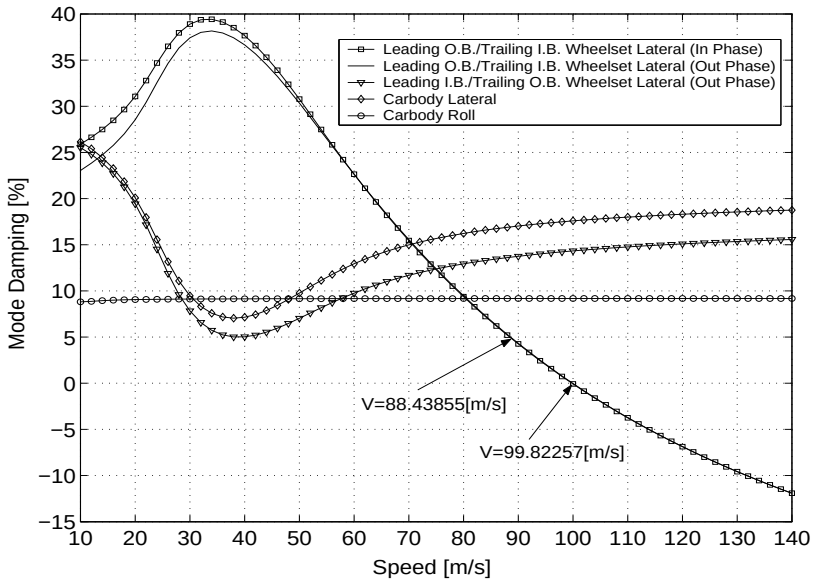

Figure 4.17: Motion mode damping ratios versus speed (generalized definition)

the vehicle is traveling at the critical speed if all other motion modes have a prescribed amount of damping (e.g. 5 % damping ratio). We call the prescribed amount of damping the "cliff damping ratio".

According to this generalized definition, the two-loop algorithm shown in Figure 4.2 should be adjusted. First, the previous two-loop algorithm is used to find the design variables that maximize the pseudo-critical speed with the cliff damping ratio. Then, for the optimized system, the SQP algorithm or the algorithm combining the SQP and DMT is employed to determine the critical speed with zero damping ratio. Thus, within a certain amount of disturbance, the maximized critical speed will not change discontinuously.

Based on the generalized definition, when the cliff phenomenon exists, the optimized design variables and the corresponding critical speed are different from those obtained using the previous two-loop algorithm. The two optimization methods are different only in that the two-loop method optimizes the design variables for maximizing pseudo-critical speed with zero damping ratio, whereas the adjusted method optimizes the design variables for maximizing the pseudo-critical speed with the cliff damping ratio. Therefore, both methods

are the same in nature and the results obtained in the previous sections still hold.

Applying this generalized definition to the case of three design variables, i.e. k_{1x}, k_{2y}, and k_{1y}, Figure 4.17 shows example results from the adjusted optimization algorithm. The critical speed is now $99.82257[m/s]$. For this case, the cliff damping ratio is selected as 5.0%, and the pseudo-critical speed with the cliff damping ratio is $88.43855[m/s]$.

4.4.5 Relative Significance of Different Design Parameter Sets

It is observed that different design parameters impose different effects on the critical speed. As discussed in the previous sections, it is clear that the GA is suitable for multiple design variable optimization problems with multiple local minima. Therefore, the GA could be used to investigate the relative significance of different design parameter sets, i.e. geometric, inertial, and suspension parameters, on the critical speed.

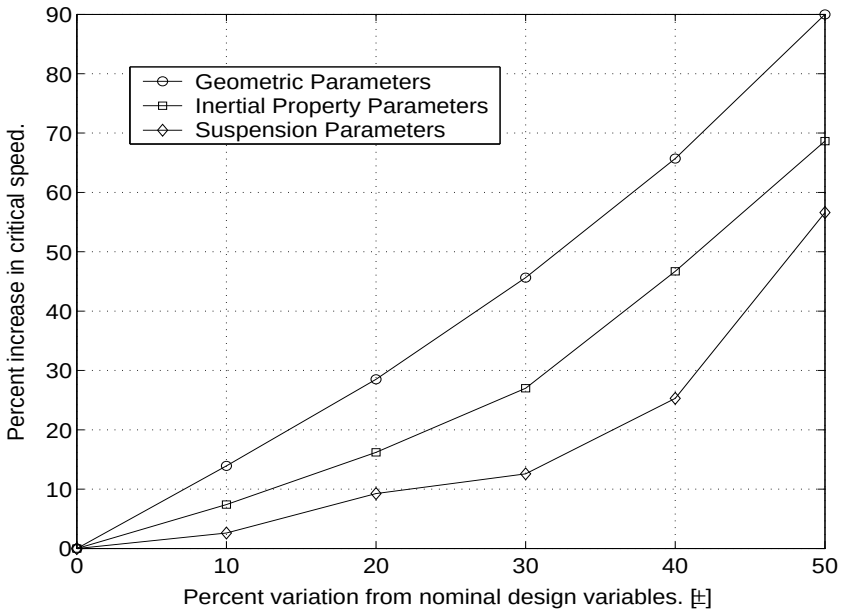

Figure 4.18: Effects of design variable sets on critical speed

Plotted in Figure 4.18 is the percent increase in the critical speed versus percent variations (from their nominal values) of suspension, inertial, and geometric parameters, e.g.

allowing the suspension stiffness and damping values to vary by 30 % leads to a 12.6 % increase in the critical speed. One can see that variations in the geometric variables (e.g. suspension locations) have much greater influence on the lateral stability than variations in the suspension stiffness and damping values. Compared with the geometric and suspension parameter sets, the inertial parameter set has medium effect on the critical speed.

4.5 Summary

An integrated approach using multibody dynamics, genetic algorithms (GAs), the sequential quadric programming (SQP) algorithm, and the Dynamic Mode Tracking (DMT) technique is an effective approach to optimizing the lateral stability of rail vehicles. In the outer loop of this two-loop algorithm, a GA is well-suited to finding the global optima even though the objective function is discontinuous and has multiple local optima, especially when there are more than a few design variables. In the interior loop, the critical speed can be determined using either a combination of the SQP and the DMT algorithms or the SQP alone.

Numerical experiments demonstrate that the algorithm combining the SQP with DMT is suitable for identifying the critical speed of rail vehicles. Over the speed span concerned, if there is only one unstable speed range, this algorithm can always find the critical speed; if there are multiple unstable speed ranges, i.e. there are multiple critical speeds, the algorithm can still always find one of those critical speeds even though this critical speed is not the least critical speed. To improve the reliability of the algorithm for identifying the least critical speed, two options are recommended: the initial speed selected should be close to, or in the neighborhood of, the least critical speed; with the concerned speed span, multiple initial speeds are chosen and the least critical speed is determined by compared the corresponding resulting optimal speeds.

When used for identifying the critical speed, if the initial speed is selected less than the critical speed, the SQP algorithm alone may converge to the critical speed or finds a non-critical speed that corresponds to a search point that does not satisfy the first-order Kuhn-Tucker conditions. Even in the latter case, the resulting calculated speed is still closer to the critical speed than the initial speed. As for the algorithm combining

the SQP with DMT, when the initial speed is selected within the speed range located in the neighborhood of the critical speed, the SQP alone can always find the critical speed. However, compared with the combined algorithm, for the SQP algorithm alone, this speed range is much narrower. To improve the reliability of the SQP alone for identifying the critical speed, a multiple SQP loop method is recommended.

In the case when the SQP algorithm alone finds a non-critical speed that corresponds to a search point that does not satisfy the first-order Kuhn-Tucker conditions, this failure is due to the fact that when the Householder's transformation and the QR algorithm are used for solving the eigenvalue problems, they rank the eigenvalues corresponding to the motion modes of a rail vehicle in their own way. Thus at two different speeds the rank of an eigenvalue corresponding to a certain motion mode may switch with that of another eigenvalue corresponding to another motion mode. This rank exchange causes the distortion of the mode in terms of motion mode damping versus speed or the discontinuity of the constraint functions for the SQP. During the final line search, if the SQP reaches this discontinuous point, no improved point for the objective function can be found and the calculation terminates. Moreover, the multiple SQP loop method recommended for identifying the critical speed is well explained by this motion mode distortion.

The multibody dynamics software A'GEM can be used to combine the outer loop algorithm, i.e. a GA, and the interior loop algorithm, i.e. the SQP alone or the combination of the SQP and the DMT, effectively by automatically generating and solving the equations of motion for realistic vehicle models, given the design variables from the GA; the results from the eigenvalue analysis are in turn offered to the SQP or the combined algorithm using the SQP and the DMT. Using this integrated optimization algorithm, the geometric parameters are found to have greater influence on lateral stability than inertial or suspension parameters.

With only one or two parameters selected as design variables, the Simplex algorithm can often find the global optimum. However, the reliability of the Simplex algorithm decreases with increases in the number of design variables.

The cliff phenomenon is observed and explained. This sharp discontinuity in the critical speed occurs when the mode (eigenvector) determining the critical speed switches. For the purpose of defining a stable critical speed for practical design problems, a cliff damping ratio

is introduced into the definition of critical speed. The integrated optimization approach is easily modified to accommodate this generalized measure of lateral stability.

Since the governing equations similar to those of rail vehicles are found in rotor dynamics, in wind turbine dynamics, in aeronautics, and in road vehicle dynamics [82], the integrated approach using multibody dynamics, GAs, the SQP or the combination of the SQP and DMT can also be applied to these problems in searching the design variables for optimizing the corresponding stability.

Chapter 5

Optimization of Vertical Ride Quality

5.1 Introduction

To identify effective optimization algorithms for vehicle ride quality analysis, in this chapter, three typical algorithms, GAs, SQP, and Simplex, are compared. The design objective is to optimize the vertical ride quality for a 2 DOF quarter-vehicle and a 20 DOF rail vehicle model. To further illustrate the effectiveness of using the GA for optimizing vehicle suspensions, results are reported for analyzing the relative significance of different design variable sets on vertical ride quality using the 20 model and a 36 DOF model. For the 36 DOF rail vehicle model, the car body flexibility is considered and active elements are introduced in the secondary suspensions for improving vertical ride quality.

These linear vehicle models are analyzed in the frequency domain, and for the 20 and 36 models, the equations of motion are generated automatically by the A'GEM multibody dynamics software. Numerical simulation results are offered to demonstrate the effectiveness of numerical differentiation as a link between A'GEM and SQP for sensitivity analysis.

5.2 Methodology for Optimizing Vertical Ride Quality

The process of vehicle ride quality analysis can be schematically represented by Figure 5.1. For a vehicle vibration system, the excitations arise when the vehicle travels at a given speed over a road or track profile which can be treated as a stationary random process. By way of wheels, suspensions, and unsprung masses, the excitations are transmitted to the sprung mass or human-body. Generally the accelerations of the sprung mass or human-body are considered as outputs due to the excitations. Based on the root mean square (R.M.S.) value of the acceleration over a given third-octave frequency band, the ISO 2631 standard [1] is generally recommended to evaluate the vehicle's ride quality. Sometimes suspension working spaces and wheel dynamic loads are also considered as outputs, because they influence the ride quality and ride safety, respectively.

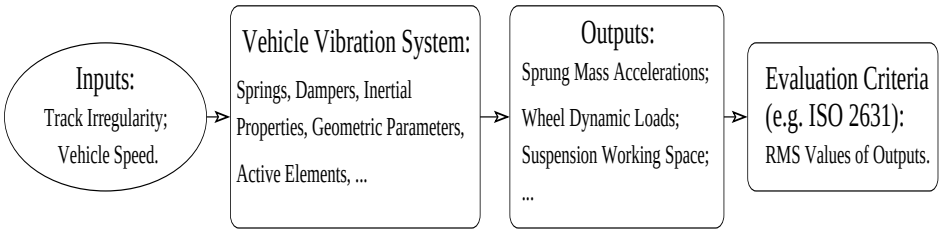

Figure 5.1: Schematic representation of a method for vehicle ride quality analysis

Due to advances in computational power and theoretical methods, the focus of research has switched from pure analysis to extensive synthesis of the vibration system. After modeling and simulating the system's behavior, the design variables can be determined to achieve optimal goals and system specifications. A combined approach using optimization algorithms and multibody dynamics is an effective tool for the synthesis of a complex vehicle suspension system [18, 42, 65, 66]. The combined approach used in our research is shown in Figure 5.2.

As shown in Figure 5.2, in the research, for the purpose of modeling realistic complex multibody vehicle models, the A'GEM multibody dynamics software package is used. With the introduction of the multibody dynamics software, the governing equations of motion

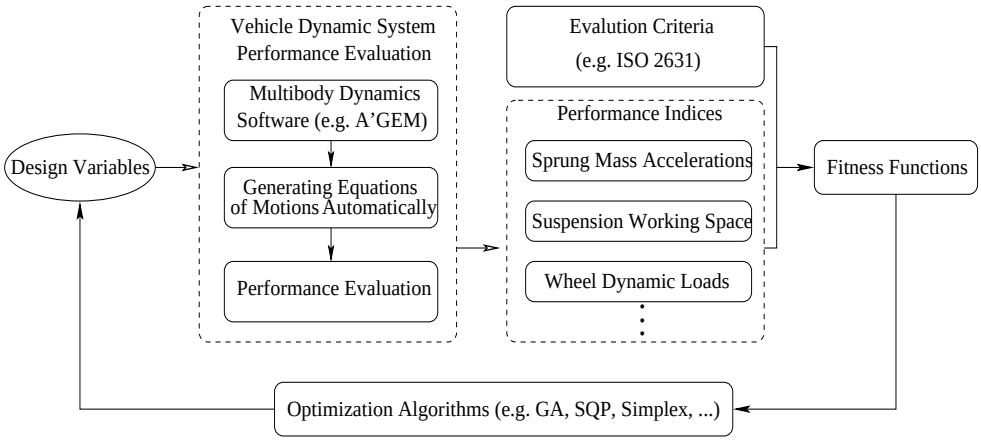

Figure 5.2: Schematic representation of the methodology for vehicle ride quality optimization

of the specified vehicle model can be generated automatically. After performing necessary transformations or integrations, the required performance evaluations can be carried out. By altering parameter values in a search to minimize a fitness function, which reflects performance characteristics, optimal design variables can be found. Thus, the numerical optimization helps automate the design process. Based on the methodology shown in Figure 5.2, different optimization algorithms such as GAs, SQP, and Simplex can be evaluated and compared.

5.3 Linear System Response to Random Excitation

For a stable linear multiple DOF system subjected to n stationary random excitations (displacements), i.e., $w_1(t), w_2(t), ..., w_m(t)$, the response power spectral density (PSD) corresponding to a DOF, i.e., $x_i(t)$, reads as follows [106]:

$$S_{x_i}(\omega) = \sum_{r=1}^{m} \sum_{s=1}^{m} H_{ir}^{*}(\omega) H_{is}(\omega) S_{w_r w_s}(\omega) \tag{5.1}$$

where $H_{is}(\omega)$ stands for the complex frequency response function of output $x_i(t)$ with respect to input $w_s(t)$, $S_{w_r w_s}$ for the cross-spectral density between $w_r(t)$ and $w_s(t)$, and H^* for the complex conjugate of H.

In the research, the stationary random excitations $\{w_1(t), w_2(t), ..., w_m(t)\}$ are track irregularities. The number of excitations matches that of vehicle wheels. The excitations or inputs may be correlated, because between two of them there may be a time delay which is generally a function of vehicle speed and geometry.

The complex frequency response function $H_{is}(\omega)$ could be determined from the following linear differential equations of motion:

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{B}(w_s, \dot{w}_s, \ddot{w}_s) \tag{5.2}$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the inertia, damping, and stiffness matrices respectively, $\mathbf{x}$ is the vector of position coordinates or DOF, and $\mathbf{B}(w_s, \dot{w}_s, \ddot{w}_s)$ is the forcing function vector relating to excitations, $w_s(t)$ (displacement), $\dot{w}_s$ (velocity), and $\ddot{w}_s$ (acceleration).

The Fourier Transform of (5.2) yields

$$(\mathbf{K} - \omega^2\mathbf{M} + i\omega\mathbf{C})\mathbf{X}(\omega) = \bar{\mathbf{B}}(\omega)W_s(\omega) \tag{5.3}$$

where vector $\mathbf{X}(\omega)$ and scalar $W_s(\omega)$ are Fourier Transforms of $\mathbf{x}$ and w_s, respectively. Provided that $(\mathbf{K} - \omega^2\mathbf{M} + i\omega\mathbf{C})$ is nonsingular, equation (5.3) can be rewritten as

$$\mathbf{X}(\omega)/W_s(\omega) = \mathbf{H}_s(\omega) = \bar{\mathbf{B}}(\omega)/(\mathbf{K} - \omega^2\mathbf{M} + i\omega\mathbf{C}) \tag{5.4}$$

where $\mathbf{H}_s(\omega)$ is a vector whose components are the complex frequency response functions relating the input $w_s(t)$ to the outputs $\{x_1(t), x_2(t), ..., x_n(t)\}$:

$$\mathbf{H}_s(\omega)^T = \{H_{1s}(\omega), ..., H_{ns}(\omega)\} \tag{5.5}$$

When the model inputs correspond to ground vehicle track irregularities, in equation (5.2), the forcing function vector relating to input $w_s(t)$ takes the form as $\mathbf{B}(w_s, \dot{w}_s)$ and the corresponding Fourier Transform in equations (5.3) and (5.4) can be expressed as [25, 48].

$$\bar{\mathbf{B}}(\omega) = \mathbf{N}(k_{rs} + i\omega c_{rs}) \tag{5.6}$$

where k_{rs} and c_{rs} are stiffness and damping coefficients for suspension elements relating vehicle wheel to track. The vector $\mathbf{N}$ can be further expressed as

$$\mathbf{N}^T = \{g_{s1}, g_{s2}, ..., g_{sn}\} \tag{5.7}$$

where the coefficients g_{sj} $(\forall j = 1, 2, ..., n)$ are determined by the vehicle model geometry.

The R.M.S. value of x_i is computed by integrating the spectral density, S_{x_i}, over all positive frequencies, i.e.,

$$\sigma_{x_i} = \left[\int_0^\infty S_{x_i}(\omega)\, d\omega\right]^{1/2} \tag{5.8}$$

Alternatively, based on the ISO 2631 standard, the acceleration of the sprung mass or human-body can be formulated in the format of (5.8) over different third-octave bands or the total weighted R.M.S. acceleration values.

For a given third-octave band, the upper frequency f_u, the lower frequency f_l, and the center frequency f_c have the following relations:

$$\begin{cases} f_u/f_l = 2^{1/3} \\ f_c = (f_u f_l)^{1/2} \end{cases} \tag{5.9}$$

Based on equation (5.9), we can express f_u and f_l in terms of f_c as $f_u = 1.12 f_c$ and $f_l = 0.89 f_c$, respectively. The width of the third-octave band, thus, is defined as $\triangle f = f_u - f_l$.

With the aid of a one-third-octave-level analyzer, the R.M.S. acceleration value over the range of one third-octave interval could be obtained by integrating the corresponding PSD of the acceleration felt by human body over the third-octave band:

$$\sigma_{\ddot{x}_i} = \left[\int_{0.89 f_{ci}}^{1.12 f_{ci}} S_{\ddot{x}_i}(f)\, df\right]^{1/2} \tag{5.10}$$

where, $i = 1, 2, ..., n$, f_{ci}, $\sigma_{\ddot{x}_i}$, and $S_{\ddot{x}_i}(f)$ are the center frequency, R.M.S. and PSD of the acceleration over the i^{th} third-octave band.

5.4 Vehicle System Models

5.4.1 2 DOF Quarter-Vehicle Model

Figure 5.3 illustrates a quarter-vehicle model considered for optimal design. The system's 2 DOF are represented by the independent generalized coordinates x_1 and x_2, which are measured from the static equilibrium position. The track profile irregularity is represented

by displacement w. M_1, M_2, k_1, k_2, and c_2 are unsprung mass, sprung mass, tire stiffness, suspension stiffness, and damping coefficient, respectively. The corresponding nominal design variables are also offered in Figure 5.3.

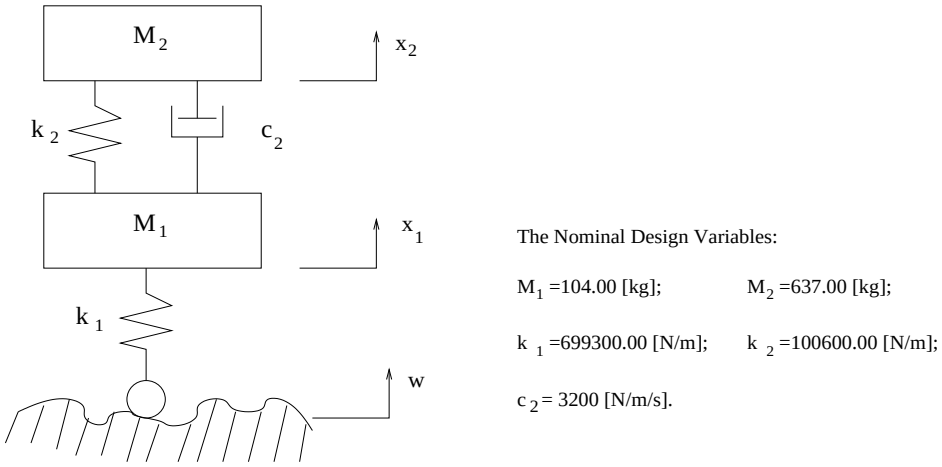

Figure 5.3: 2 DOF quarter-vehicle model

To explicitly express the objective function and constraint functions in terms of design variables, the governing equations of motion of the model are derived by hand and written in the format of equation (5.2). If $\mu = M_2/M_1$, $\omega_2 = (k_2/M_2)^{1/2}$, $\omega_1 = (k_1/M_1)^{1/2}$, $\alpha = c_2/(2M_2)$, $\xi = c_2/[2(M_2k_2)^{1/2}]$, we have the frequency response function of $\ddot{x}_2$ with respect to track excitation w as

$$H_{\ddot{x}_2 w}(\omega) = -\omega^2(i2\alpha\omega_1^2\omega + \omega_1^2\omega_2^2)/A \tag{5.11}$$

where $A = \{\omega^4 - i2\alpha(\mu+1)\omega^3 - [\omega_1^2 + (\mu+1)\omega_2^2]\omega^2 + i2\alpha\omega_1^2\omega + \omega_1^2\omega_2^2\}$. If the relative displacement of M_2 with respect to M_1 is h, and $h = x_2 - x_1$, the frequency response function of h with respect to w is

$$H_{hw}(\omega) = \omega_1^2\omega^2/A \tag{5.12}$$

If the dynamic wheel load with respect to the track is $p = k_1(x_1 - w)$ and the static vehicle load is $G = (M_1 + M_2)g$, where g is the acceleration of gravity, we assume that p/G is the

relative dynamic wheel load with respect to the track. Therefore, the frequency response of p/G with respect to w is

$$H_{p/Gw}(\omega) = \omega_1^2\omega^2[\omega^2/(\mu+1) - \omega_2^2 - i2\alpha\omega]/(gA) \tag{5.13}$$

For the 2 DOF quarter-vehicle model, there is only one random track profile displacement excitation w. The PSD of the excitation [148] may be expressed as

$$S_w(n) = S_w(n_0)(n/n_0)^{-\gamma} \tag{5.14}$$

where n is the spatial frequency, $n = 2\pi/\iota\ [rad/m]$, ι is wavelength $[m]$, n_0 is a reference spatial frequency, generally $n_0 = 1.0[rad/m]$, and γ is a constant frequency index, generally $\gamma = 2.0$. When a vehicle travels at speed of $V[m/s]$ on the track with spatial frequency n, the time angular frequency of track excitation ω is

$$\omega = Vn \tag{5.15}$$

In the time frequency domain, the PSD of the track excitation can be expressed in terms of the PSD in the spatial frequency domain [106] as

$$S_w(\omega) = (1/V)S_w(n) \tag{5.16}$$

If $\gamma = 2$, $R = n_0^2 S_w(n_0)$; substituting equations (5.14) and (5.15) into equation (5.16) yields

$$S_w(\omega) = RV/\omega^2 \tag{5.17}$$

Thus, for the 2 DOF model, for a dynamic response, e.g. x, the response PSD in the format of equation (5.1) reduces to

$$S_{xw}(\omega) = |H_{xw}(\omega)|^2 S_w(\omega) \tag{5.18}$$

Based on equation (5.8), the R.M.S. values of $\ddot{x}_2$, h, and p/G can be expressed as

$$\begin{cases} \sigma_{\ddot{x}} = \{\pi RV[\omega_1^2\xi\omega_2/\mu + (1+\mu)\omega_2^3/(4\mu\xi)]\}^{1/2} \\ \sigma_h = \{\pi RV(1+\mu)/(4\mu\xi\omega_2)\}^{1/2} \\ \sigma_{p/G} = (\pi RVA_1)^{1/2}/g \end{cases} \tag{5.19}$$

where $A_1 = \{\omega_1^4/[4\xi\omega_2\mu(1+\mu)^2] + (1+\mu)\omega_2^3/(4\mu\xi) - \omega_2\omega_1^2/(2(\mu(1+\mu)\xi) + \xi\omega_2\omega_1^2/\mu\}$.

5.4.2 20 and 36 DOF Rail Vehicle Models

The rail vehicle configuration with car body flexibility is shown in Figure 5.4, with the leading bogie, car, and trailing bogie denoted as bodies 2, 4 and 6 respectively. For the 20 DOF model, the car body is treated as a rigid body, while for the 36 DOF model with the car body flexibility, the car body 4 is further divided as 4(1), 4(2), ..., 4(5). The adjacent sections are connected by a group of bending, torsion and shear springs [76]. The leading bogie, with the leading and trailing wheelsets denoted as 1 and 3, and trailing bogie, with the leading and trailing wheelsets denoted as 5 and 7, are connected to the car by secondary suspensions. Both the leading and trailing bogies, in turn, are connected with their own leading and trailing wheelsets by primary suspensions. Each suspension component consists of a parallel spring and damper, with stiffness and damping coefficients in the three coordinate directions. In the vertical direction, the secondary suspension is equipped with 2 active components per bogie.

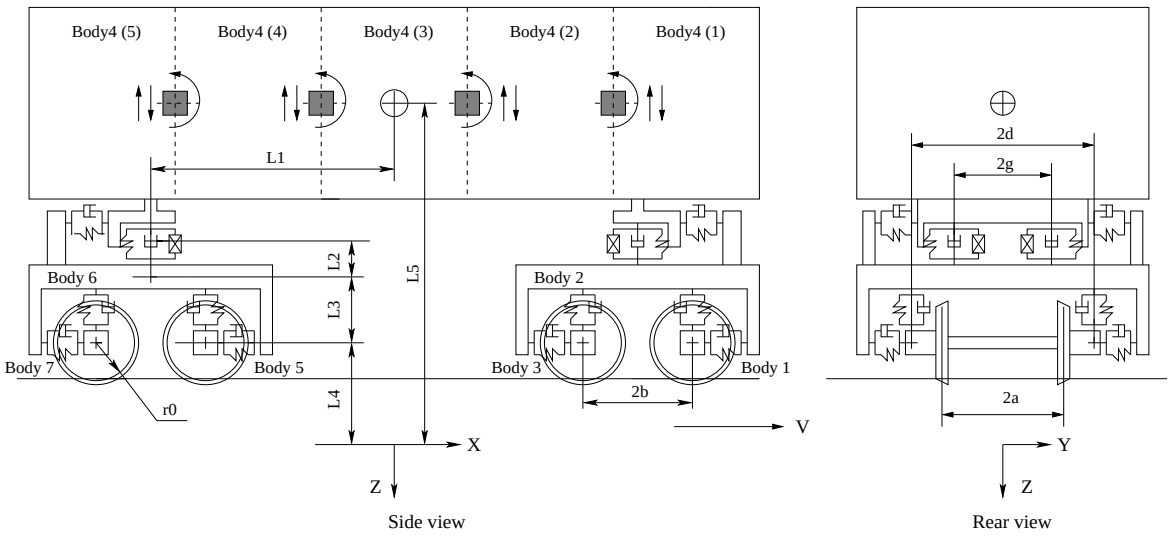

Figure 5.4: Rail vehicle configuration for dynamic models

For the 20 and 36 DOF models, the corresponding nominal design variables, which are adopted from reference [136] and [53], are listed in Tables A.3 and A.2 in Appendix A, respectively.

The degrees of freedom of the 36 DOF model are summarized in Table 5.1. Notice

that for the 20 DOF model, the car body's DOF are lateral displacement (y_4), vertical displacement (z_4), roll motion (ϕ_4), and pitch motion (θ_4). The DOF of the bogies and wheelsets are the same as those of the 36 DOF model.

Table 5.1: Vertical ride quality model components and DOF.

Component	Lateral (y)	Vertical (z)	Roll (ϕ)	Pitch (θ)
Body1		z_1	ϕ_1	
Body2	y_2	z_2	ϕ_2	θ_2
Body3		z_3	ϕ_3	
Body4(1)	$y_{4(1)}$	$z_{4(1)}$	$\phi_{4(1)}$	$\theta_{4(1)}$
Body4(2)	$y_{4(2)}$	$z_{4(2)}$	$\phi_{4(2)}$	$\theta_{4(2)}$
Body4(3)	$y_{4(3)}$	$z_{4(3)}$	$\phi_{4(3)}$	$\theta_{4(3)}$
Body4(4)	$y_{4(4)}$	$z_{4(4)}$	$\phi_{4(4)}$	$\theta_{4(4)}$
Body4(5)	$y_{4(5)}$	$z_{4(5)}$	$\phi_{4(5)}$	$\theta_{4(5)}$
Body5		z_5	ϕ_5	
Body6	y_6	z_6	ϕ_6	θ_6
Body7		z_7	ϕ_7	

For these complex rail vehicle models, the governing equations of motion are generated automatically by A'GEM. The PSD of the random rail profile recommended by the Federal Railroad Administration of the U.S.A. is used as excitation. It is assumed that the wheels follow the rails at all times so that the motions of the wheels are prescribed. According to the procedures described in previous sections, the frequency responses, i.e. passenger point accelerations and secondary suspension working spaces, of the model to random rail profile inputs are also determined by A'GEM. Once the PSD is computed for these dynamic responses, the $ISO/(2631-1985)$ ride quality criterion can be evaluated by integrating the PSD over third-octave bands [1] to obtain the R.M.S. acceleration in the frequency bands.

Note that for the 36 DOF method the flexible car body has been represented by a five-element discretization of a beam undergoing vertical bending, axial torsion, and lateral and vertical shear. In order to determine the values for vertical bending springs, the car body

was first modeled as a free-free Bernoulli-Euler beam and the corresponding fourth-order partial differential equation describing the bending deflection is obtained. Based on the fact that the mode shape function is independent of time, by means of variable separation method and necessary transformations, the partial differential equation can be reduced to two ordinary differential equations with the introduction of the beam eigenfrequencies. Because the lowest frequency of flexible modes of a railway passenger vehicle is usually the body vertical bending modes with a frequency in the range $8 - 12[Hz]$ [43], the first bending mode frequency for the model is chosen as $8.0[Hz]$. With the general solution to the partial differential equation, based on the boundary conditions and the method recommended in [110], the nominal spring constant for each bending spring is computed. The nominal spring constant for each torsion and shear spring is taken from reference [40].

Two active elements per bogie are introduced in the secondary suspension in the vertical direction. The 'skyhook' control strategy [80] is used. The active torque and force acting on the corresponding car body section can be expressed as follows:

$$\begin{cases} T_{x4(i)} = -K_{tr}\dot{\phi}_{4(i)} \\ F_{z4(i)} = -K_{fr}\dot{z}_{4(i)} \end{cases} \tag{5.20}$$

where $i = 1, 5$, and K_{tr} and K_{fr} are the rolling torque rate gain and vertical force rate gain, respectively.

5.5 Optimization Problem and Implementation

5.5.1 Objective Function, Constraints and Design Variables

2 DOF Quarter-Vehicle Model

For simplicity, ω_1, ω_2, and ξ are selected as design variables. Once ω_1, ω_2, and ξ are determined, we can easily obtain the corresponding k_1, k_2, and c_2.

Ride comfort is chosen to be the objective function. If the dynamic wheel load were greater than the static wheel load, the wheel would bounce out of the track and the vehicle would become unstable. Therefore, the relative dynamic wheel load is included as a constraint. If the suspension working space is not restricted, the unsprung mass may

strike the sprung mass and the ride quality will become worse. Thus, this characteristic is also included as another constraint. The constrained minimization problem can be transformed into the following standard form:

$$minimize \qquad \sigma_{\ddot{x}_2}(\omega_1, \omega_2, \xi) \tag{5.21}$$

$$subject\ to \quad \begin{cases} \sigma_{p/G}(\omega_1, \omega_2, \xi) \leq b_1 \\ \sigma_h(\omega_1, \omega_2, \xi) \leq b_2 \\ 0.1 \leq \xi \leq 0.5 \\ b_3 \leq \omega_2 \leq 20.0 \\ b_4 \leq \omega_1 \leq 95.0 \end{cases} \tag{5.22}$$

In (5.22), if $b_1 = 0.3333$, the probability for the wheel to bounce out of the road is 0.3%. Generally we select b_1 as 0.4472; in such a case, the possibility for the wheel to bounce out of the track is 2.51%. If $b_2 = 0.3333 h_d$, where h_d is the limitation of suspension working space (maximum dynamic deflection), the possibility for the unsprung mass to strike the sprung mass is less than 0.3%. In the case study, b_2 is chosen as $0.016667[m]$ (i.e., $h_d = 0.05[m]$). According to the limitation of wheels, b_4 is chosen as $76.66667[rad/s]$. Based on the different requirements, b_3 takes values of 8.0, 8.35, 8.5, 9.0, 9.6, and 10.0 $[rad/s]$.

SQP is very suitable for the general constrained minimization problem expressed in the standard form of equations (5.21) and (5.22). However, if a GA or Simplex is used for the constrained minimization problem, the objective function and constraints expressed in equations (5.21) and (5.22) can not be used directly since GAs and Simplex are suited for unconstrained optimization problems. But we can use the penalty methods that degrade the fitness ranking in relation to the degree of constraint violation. With these methods, a constrained problem in optimization is transformed into an unconstrained optimization problem by associating a cost or penalty with all constraint violations. Therefore, the solution to the problem is to find an appropriate fitness function to be maximized, which depends on the objective function and constraints, as follows:

$$Fit = f - \eta \sigma_{\ddot{x}_2} - \beta max[0, (\sigma_{p/G} - b_1)] - \gamma max[0, (\sigma_h - b_2)] \tag{5.23}$$

where the constants f and η and penalty multipliers β and γ are determined according to the ranges of the design variables ω_1, ω_2, and ξ. Notice that the last 3 constraints in

equations (5.22) are not included in equation (5.23) because these constraints are satisfied by specifying the ranges of allowable values for design variables.

20 and 36 DOF Rail Vehicle Model

The design variables consist of suspension stiffness and damping coefficients ($\bar{S}$), inertial property parameters ($\bar{I}$), geometric parameters ($\bar{G}$), and active control parameters ($\bar{A}$).

The fitness function to be maximized is a combination of R.M.S. acceleration values at different points of the car body, and secondary suspension working spaces:

$$Fit = f - \zeta \left[\int_{\omega_l}^{\omega_u} S_{\ddot{z}_4}(\omega)\, d\omega\right]^{1/2} - \nu max[0, (\left[\int_{\omega_l}^{\omega_u} S_{h_k}(\omega)\, d\omega\right]^{1/2} - h_k)] \qquad (5.24)$$

where f, ζ and ν are constants, h_k $(k = 1, ..., 4)$ are limits on the secondary suspension working spaces, ω_l and ω_u define the frequency interval of interest, and $S_{\ddot{z}_4}(\omega)$ and $S_{h_k}(\omega)$ are the PSDs of the car body vertical acceleration and the working space of the k^{th} secondary suspension, respectively.

5.5.2 Implementation of the Optimization Problem

As shown in Figure 5.5, the optimization problem is implemented by using a GA (MechaGen program [10]), an SQP (E04UCF) algorithm, and a Simplex (E04CCF) algorithm, respectively. For the 20 or 36 DOF rail vehicle optimization problem, the A'GEM software is included.

SQP is based not only on function evaluations but also on gradient information of the objective and constraint functions. For the 2 DOF quarter-vehicle model, since the objective function and constraint functions are expressed explicitly in terms of design variables, reliable analytical gradient information is available during the optimization. However, for the 20 and 36 DOF models, due to the use of A'GEM, the gradients can not be computed analytically. For this case, three techniques are available for offering gradient information: numerical differentiation, the direct differentiation method, and the adjoint variable method. Very often the direct differentiation or adjoint variable method is implemented using symbolic packages to derive the necessary matrices and equations to be included in the derivative code [16]. However, symbolic packages can be very cumbersome to use,

e.g. when differentiated, the various functional dependencies have to be identified and carefully defined [10]. Moreover, for complicated rail vehicle models, it is not practical for symbolic packages to handle programming constructs such as conditional statements, loops and subroutines that are common to numerical multibody vehicle models.

Using finite differences to compute gradients may yield poor results, which may cause slow convergence of the iterative optimization process or a total break down, due to relatively large errors in evaluating objective function and constraint functions by numerical integrations [19]. However, when these functions are algebraic, the error of the gradient of the function is the first order of the variation of the variable. This error is the so-called truncation error. Thus, if the variation is selected small enough, the gradient could be computed within machine precision. Moreover, the computer has a limited accuracy. Subtracting function values that are almost equal and dividing by a small number, i.e. the variation, may lead to large errors in the approximation of the derivative due to numerical truncation. However, these problems may be solved by choosing an appropriate difference interval and by linearly or even nonlinearly scaling the optimization problem.

In our case study, as shown in equation (5.19), for the 2 DOF vehicle model, the objective function and constraint functions are algebraic. For the 20 and 36 DOF rail vehicle models, based on the features of R.M.S. value of dynamic frequency response as shown in Section 5.3, the objective functions and constraint function can be written in the following general form:

$$F(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}}) = P_0(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}}) + \sum_{i=1}^{n} P_i(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}}) \left[\int_{\omega_l}^{\omega_u} S_i(\omega)\, d\omega \right]^{1/2} \tag{5.25}$$

where n is the total number of the vehicle system responses of interest and $P_0(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}})$ and $P_i(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}})$ are algebraic functions of design variable vectors $\bar{\mathbf{S}}$ (suspension stiffness and damping coefficients), $\bar{\mathbf{I}}$ (inertial property parameters), $\bar{\mathbf{G}}$ (geometric parameters), and $\bar{\mathbf{A}}$ (active parameters). For a frequency interval of interest ($[\omega_l, \omega_u]$), given track condition and vehicle speed, $\left[\int_{\omega_l}^{\omega_u} S_i(\omega)\, d\omega\right]^{1/2}$ is constant. Therefore, $F(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}})$ is an algebraic function of design variables.

In our case study, the derivatives of the objective and constraint functions for the SQP are computed using a numerical differentiation method. Since the GA or Simplex is only based on function evaluations, the link between the numerical multibody dynamics

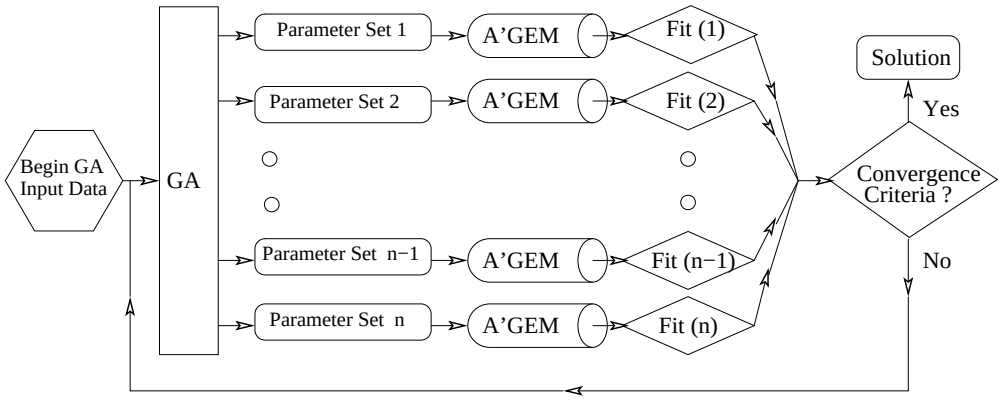

Figure 5.5: The implementation of the GA combined with A'GEM

software A'GEM and the optimization algorithm (GA or Simplex) for computing gradients is superfluous.

Figure 5.5 shows the flowchart for implementing the GA algorithm when the 20 or 36 DOF rail vehicle model is optimized. First, a random population of designs is selected by the GA; the corresponding sets of design variables are sent in parallel to the A'GEM routines that automatically generate the equations of motion. These equations of motion are transformed by means of Fourier Transformations, the PSDs of dynamic frequency responses are integrated in the interval of interest, and then the performance indices are directly obtained. From these performance indices, the fitness function (5.24) is evaluated for each design in the population. At this point, if the convergence criteria are satisfied, the calculation terminates; otherwise these fitness values are returned to the GA. Based on the returned fitness values corresponding to the given sets of design variables, the GA produces the next generation of design variable sets using reproduction, crossover and mutation operators. This procedure repeats until the optimized design variable set is found.

5.6 Results and Discussion

5.6.1 Validation of Numerical Differentiation Method

For the 20 or 36 DOF model optimization problems, how to use a numerical differentiation method for offering gradient information to the SQP is a nontrivial problem. The main point of this problem is to select an appropriate difference interval of a specific design variable. In fact, the finite difference interval is chosen so as to minimize the following computable bound on the error of gradient [56]

$$\begin{cases} Err_g = \mid \Phi \mid h_i/2 + \bar{C}(2/h_i) \\ h_i = \delta(1+ \mid X_i \mid) \end{cases} \tag{5.26}$$

where for a function of $f(\mathbf{X})$, $\mathbf{X} \in R^n$, given a variation of h_i of a variable X_i, Φ is an estimate of $f''(\vartheta)$, $X_i \leq \vartheta \leq X_i + h_i$, $\bar{C}$ is a bound on the condition error, and δ is a difference interval factor. Based on the formula, one would expect the value of δ to have a considerable effect on the number of iterations and the final solution. Practical results did confirm this expectation.

Figure 5.6 shows a simulation result from A'GEM, which offers the R.M.S. maximum weighted acceleration at the rear left corner of the car body of the 20 DOF model versus the half bogie space ($L1$, see Figure 5.4). Obviously the two optimal points are located at the lower and upper bounds. Table 5.2 illustrates the effect of the difference interval on the number of iterations and the final solution of the SQP. A number of numerical experiments showed that when δ lies between $1.0E-3$ and $1.0E-5$, the SQP performs quite satisfactorily.

Practical experience showed that appropriately scaling the optimization problem can also improve the performance of the optimization algorithm. This phenomenon may result from the fact that some design variables may have a very small derivative with respect to the objective function. For this reason, the optimization process may terminate prematurely.

Linear or nonlinear transformations can be used to change the magnitude of a derivative. With the increase of the small derivatives, the SQP may terminate at optimal points. For the case of Figure 5.6 and Table 5.2, when δ takes the value of $1.0E-1$ and the initial value of $L1$ is $5.2[m]$, Figure 5.7 shows that three different linear scaling schemes have totally

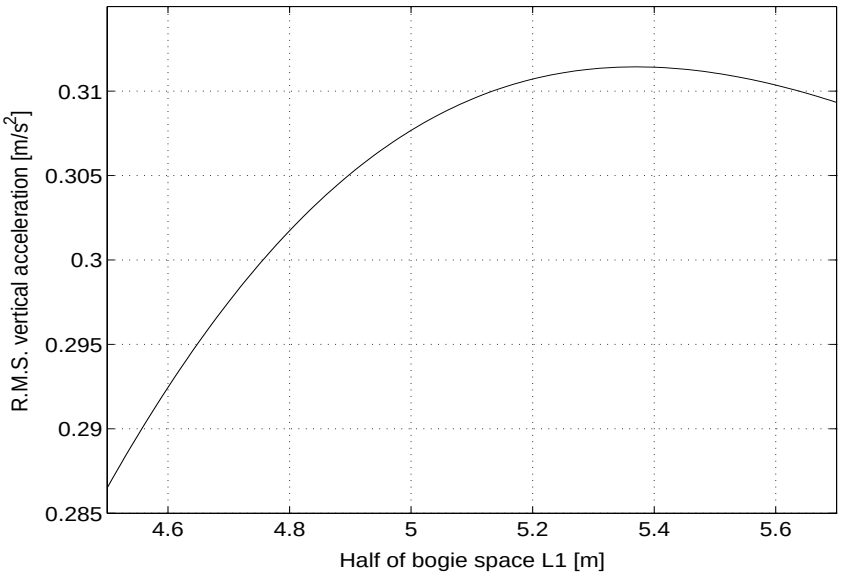

Figure 5.6: R.M.S. vertical acceleration versus half bogie space

Table 5.2: The effect of difference interval on the optimization performance

	$\delta = 1.0E-1$			$\delta = 1.0E-5$		
Ini†$L1[m]$	Opt‡$L1[m]$	Iter#*	acc⋆ $[m/s^2]$	Opt$L1[m]$	Iter#	acc $[m/s^2]$
5.0000	4.5000	2	0.2865	4.5000	1	0.2865
5.1000	4.5000	2	0.2865	4.5000	1	0.2865
5.1500	4.5000	16	0.2865	4.5000	1	0.2865
5.2000	4.5000	31	0.2865	4.5000	1	0.2865
5.2500	4.5000	47	0.2865	4.5000	1	0.2865
5.3000	5.1523	50	0.3102	4.5000	1	0.2865
5.5000	5.7000	2	0.3093	5.7000	1	0.3093
5.6000	5.7000	2	0.3093	5.7000	1	0.3093

† Initial value; ‡ Optimal value; * Iteration number; ⋆ acceleration.

different effects on the SQP's performance. Notice that in Figure 5.7, for the scaling schemes 1, 2, and 3, the scaling factors for the design variables L_1 are chosen as 0.1, 0.5, and 1.0, respectively.

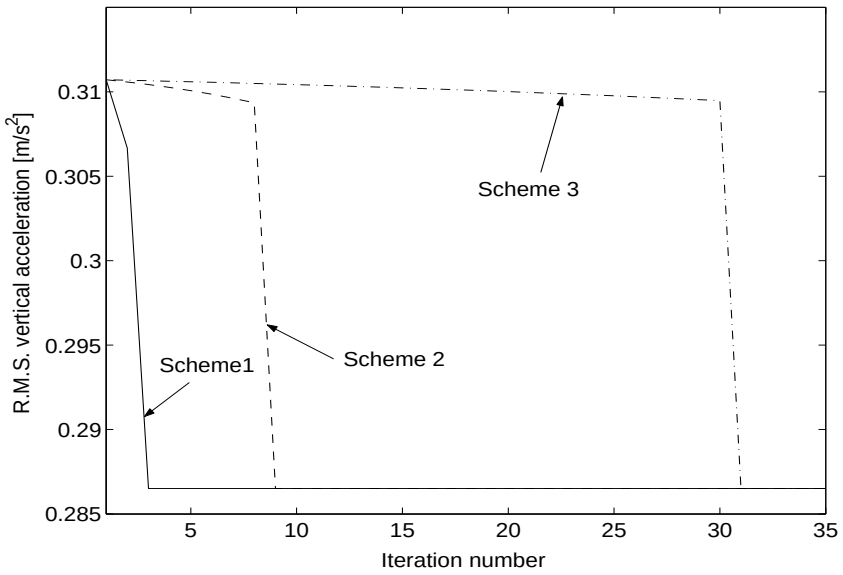

Figure 5.7: The effect of scaling schemes on the optimization performance

5.6.2 Optimization of the 2 DOF Vehicle Model

Table 5.3 displays the results of the numerical optimization for vehicle speed of $40[m/s]$ and a track with the PSD constant $S_w(n_0)$ taking the value of $6.5 \times 10^6[m^3]$. The GA and SQP approaches are used. When using the SQP, an initial set of design variables within the corresponding boundary is guessed and offered. For the case study, the lower bound of ω_1 takes the value of 76.6667 $[rad/s]$ and the lower bound of ω_2 takes the values of 8.0, 9.0, and 10.0 $[rad/s]$ respectively. The maximum number of generations of 2000 was set for the GA. In Table 5.3, the CPU time used by an SGI Indigo 2 XZ workstation is also offered.

Table 5.3 shows that with the appropriate selection of algorithm parameters, e.g. probability of mutation and population size, within the same feasible range of design variables,

Table 5.3: The result of design optimization of the 2 DOF model suspension system

	ξ	ω_2 [rad/s]	ω_1 [rad/s]	$\sigma_{\ddot{x}_2}$ $[m/s^2]$	$\sigma_{p/G}$	σ_h [m]	t [s]
SQP	0.13924	8.0000	76.6667	1.3217	0.4442	0.0146	0.87
GA	0.13905	8.0000	76.6667	1.3216	0.4445	0.0146	48.07
SQP	0.15664	9.0000	76.6667	1.4869	0.3971	0.0130	0.68
GA	0.15625	9.0000	76.6667	1.4869	0.3975	0.0130	47.01
SQP	0.17406	10.0000	76.6667	1.6521	0.3623	0.0117	0.80
GA	0.17405	10.0000	76.6667	1.6521	0.3623	0.0117	48.16

the result obtained from the GA is almost the same as that from the SQP. However, the GA is much more time consuming than the SQP. Table 5.3 also illustrates that the result is the global optimum for the corresponding design variable range. In fact, within a certain design variable range, there is an unique minimum. This can be illustrated by Figure 5.8.

With the continuity property of the objective and constraint functions and the available analytical gradients for these functions, numerical experiments show that the SQP can efficiently and reliably find the optimum in the neighbourhood of the initial point. SQP is also called "exploitative" algorithm with the characteristics that the more intensive the exploitation, the stronger the need of specialized information about the function to be minimized. In contrast, the GA uses only the order relations between the points of individuals of the current population. Its search is performed partially randomly and may include bad intermediate solutions that the SQP method would not have explored.

5.6.3 Optimization of the 20 DOF Vehicle Model

In order to further investigate the performance of SQP and compare the algorithm with the GA and Simplex, a benchmark is carried out by means of optimizing the 20 DOF rail vehicle model. Figure 5.9 shows the relationship between acceleration and half bogie space (all other parameters are fixed and take nominal values) for the 20 DOF model at the same point of the car body as for the case of Figure 5.6. Clearly, the curve shown in Figure 5.6 is a smooth section of that shown in Figure 5.9. By inspection, there are at least two

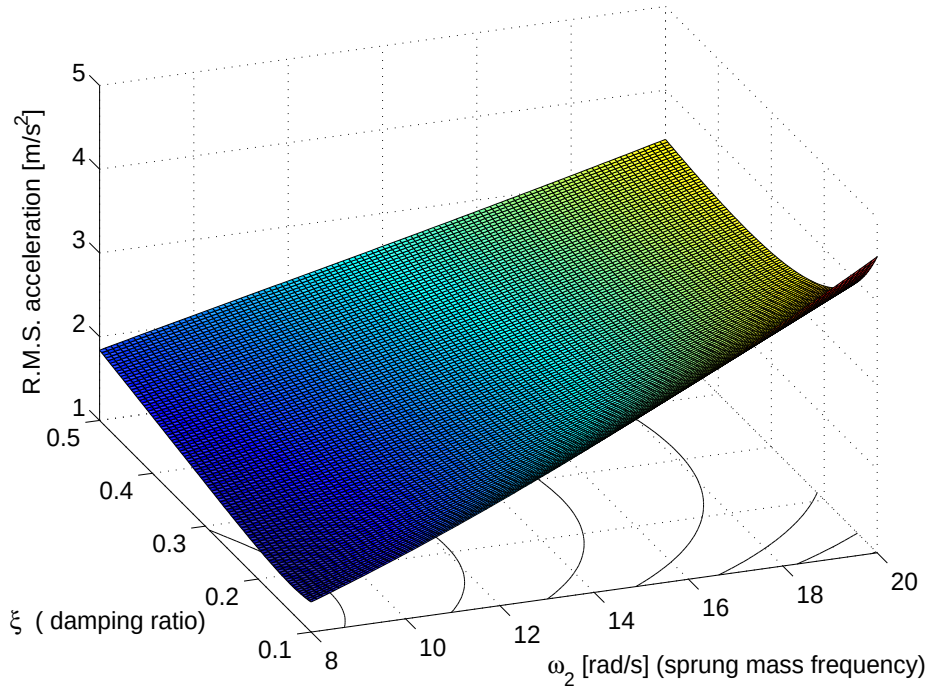

Figure 5.8: $\sigma_{\ddot{x}_2}$ versus ω_2 and ξ with ω_1 fixed at optimal value

local optimal points A and B on the curve. Based on the SQP's requirement that the objective and constraint functions should be "at least twice-continuously differentiable", it is expected that the algorithm will not be reliable for finding the desired solution. Table 5.4 offers the result from the SQP along with those from the GA and Simplex.

From Table 5.4, we notice that the SQP can reliably find the local optimal points A or B with only few function evaluations. Sometimes, when the initial points are selected on the right side of local optimal point A, the search may jump over the point and find the global point B. To some extent, for this one design variable optimization problem, the Simplex behaves like the SQP but with higher reliability for finding the global point B. However, the GA can always find the global point at the expense of more function evaluations.

Figure 5.10 shows the relationship between the mass of car body (again, all other parameters are fixed and take nominal values) and the acceleration at the same point of car body of the 20 DOF model. The optimization results from the SQP, Simplex and GA are offered in Table 5.5. The same features of the performance of the three algorithms as

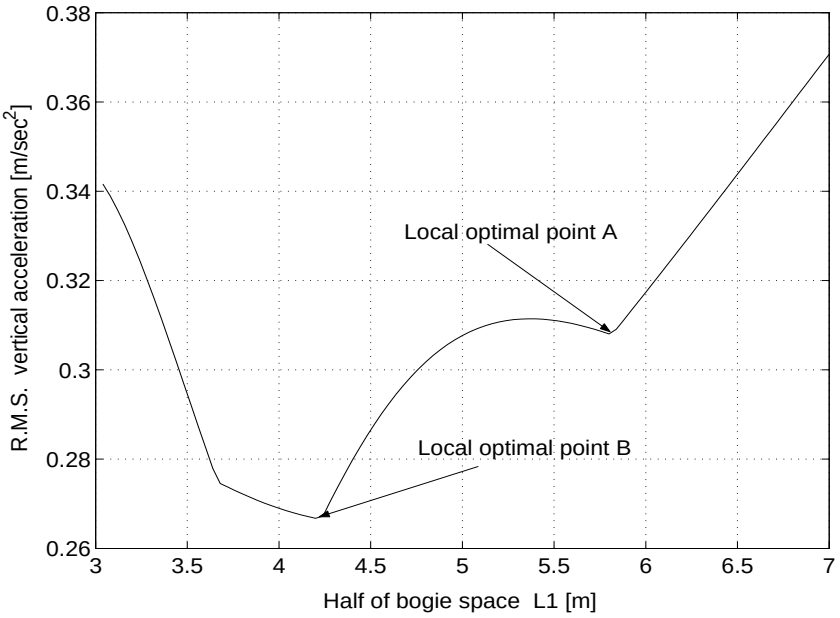

Figure 5.9: R.M.S. vertical acceleration versus half bogie space

mentioned in the previous case can be revealed by analyzing the optimization results.

Corresponding to the one-dimensional simulation results shown in Figures 5.9 and 5.10, the two-dimensional simulation result is displayed in Figure 5.11 in which the car body mass and half bogie space are optimized simultaneously. Compared with the objective function surface shown in Figure 5.8, the objective function surface for the 20 DOF model is much more complicated and there are more local optimal points. The optimization results from the SQP, Simplex and GA are given in Table 5.6.

For the two design variable optimization problem, as shown in Table 5.6, from different initial points, the SQP converges to different points including the global optimal point. Moreover, compared with the previous one design variable optimization problem, the reliability of the algorithm to find the global optimum for the two design variable problem becomes less. This is also the case for the Simplex. Further numerical experiments show that for the multibody rail vehicle model, the more the design variables, the less the reliability for the SQP and Simplex to find the global optimum. However, for the multiple design variable optimization problems, the GA is robust to find the global optimum

Table 5.4: R.M.S. vertical acceleration versus half bogie space

	SQP		*Simplex*	
Int† $L1[m]$	Opt‡ $L1[m]$	R.M.S. acc* $[m/s^2]$	Opt $L1[m]$	R.M.S. acc $[m/s^2]$
3.2000	4.2312	0.2665	4.2312	0.2665
4.0000	4.2312	0.2665	4.2312	0.2665
4.5000	4.2312	0.2665	4.2312	0.2665
5.4000	5.8155	0.3078	4.2312	0.2665
5.5000	5.8155	0.3078	4.2312	0.2665
5.7000	5.8155	0.3078	5.8155	0.3078
5.7500	5.8155	0.3078	5.8155	0.3078
5.8500	5.8155	0.3078	5.8155	0.3078
5.9000	5.8155	0.3078	4.2312	0.2665
6.0000	5.8155	0.3078	4.2312	0.2665
6.5000	4.2312	0.2665	4.2312	0.2665
6.9000	4.2312	0.2665	4.2312	0.2665
GA				
Norm⋆ $L1[m]$	Lower° $L1[m]$	Upper• $L1[m]$	Opt $L1[m]$	R.M.S. acc $[m/s^2]$
5.0000	3.0000	7.0000	4.2317	0.2665

† Initial value; ‡ Optimal value; * Acceleration; ⋆ Nominal value;
° Lower bound; • Upper bound.

although it is time consuming.

5.6.4 Using the GA for Optimizing Vehicle Suspensions

In order to further show the effectiveness of using the GA for optimizing vehicle suspensions, the GA is used to analyze the relative significance of different design variable sets using the 20 DOF rail vehicle model and to optimize a rail vehicle suspension with active elements using the 36 DOF model.

Sensitivity Analysis for Design Variable Sets

By optimizing design variables for improving ride quality, the relative significance of different design variable sets, i.e. $\bar{\mathbf{S}}$, $\bar{\mathbf{I}}$ and $\bar{\mathbf{G}}$, is investigated. Plotted in Figure 5.12 is the

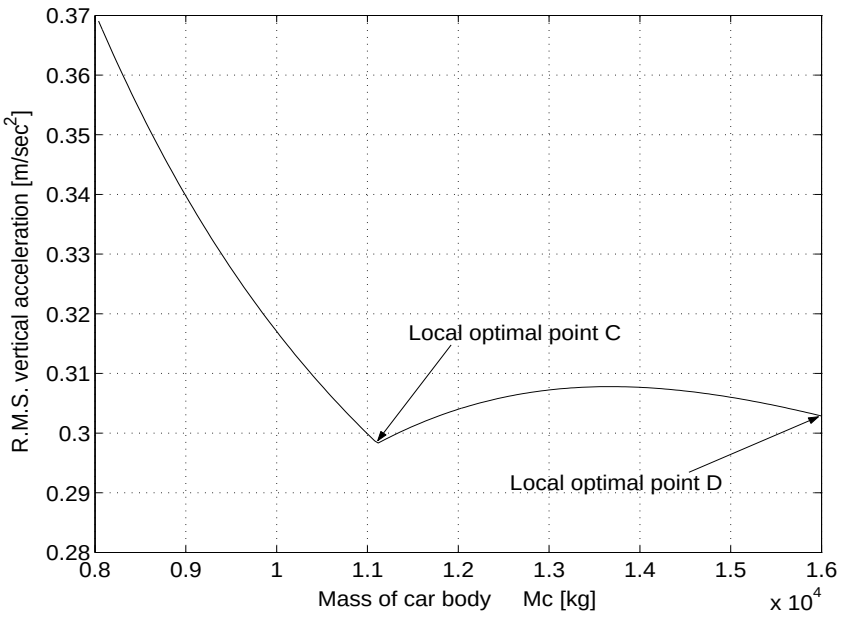

Figure 5.10: R.M.S. vertical acceleration versus mass of car body

Table 5.5: R.M.S. vertical acceleration versus mass of car body

	SQP		*Simplex*	
Int Mc[kg]	Opt Mc[kg]	RMS A $[m/s^2]$	Opt Mc[kg]	RMS A $[m/s^2]$
9000.0	11110.6	0.2983	11111.0	0.2983
12000.0	11110.6	0.2983	11111.0	0.2983
13000.0	11110.6	0.2983	11111.0	0.2983
14000.0	16000.0	0.3029	11111.0	0.2983
15350.0	16000.0	0.3029	11111.0	0.2983
15360.0	16000.0	0.3029	16000.0	0.3029
15500.0	16000.0	0.3029	16000.0	0.3029
GA				
Norm Mc[kg]	Lower Mc[kg]	Upper Mc[kg]	Opt Mc[kg]	RMS A $[m/s^2]$
14000.0	8000.0	16000.0	11112.4	0.2983

percent decrease in the vertical acceleration at the rear left corner of car body versus changes in the suspension, inertial, and geometric parameters, e.g. if the geometric parameters are allowed to vary by 20.0%, the optimized design variables lead to a 23.2% decrease in the acceleration. One can see that among the three design variable sets, the geometric parameters have the most significant effect on the ride quality and the inertial property parameters have the least effect. One may also notice that the effect of suspension parameter set on the acceleration is very close to that of geometric parameter set. Considering the fact that for practical vehicle suspensions, varying suspension stiffness and damping coefficients is more convenient than varying geometric or inertial parameters, the results obtained explain why suspension stiffness and damping coefficient control and optimization are often used to improve vertical ride quality of rail vehicles.

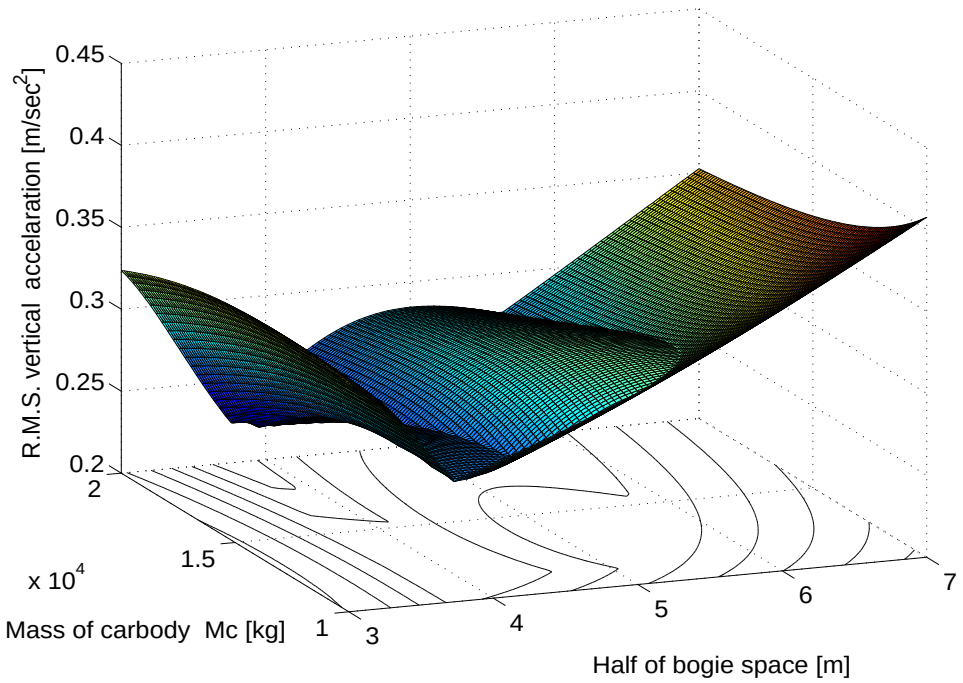

Figure 5.11: R.M.S. vertical acceleration versus half bogie space and mass of carbody

Table 5.6: R.M.S. vertical acceleration versus half of bogie space and mass of car body

Method	*Initial Values*		*Optimal Values*		
	$L1[m]$	Mc [kg]	$L1[m]$	Mc [kg]	RMS acc $[m/s^2]$
SQP	3.2000	8100.0	4.3926	11562.8	0.2679
Simplex	3.2000	8100.0	4.3660	11493.0	0.2679
SQP	4.5000	8500.0	4.3186	12846.5	0.2688
Simplex	4.5000	8500.0	4.0730	16000.0	0.2558
SQP	4.8000	9800.0	4.3931	11564.0	0.2679
Simplex	4.8000	9800.0	4.4170	11629.0	0.2679
SQP	5.0000	15000.0	4.0727	16000.0	0.2558
Simplex	5.0000	15000.0	4.4170	11629.0	0.2679
SQP	5.5000	12000.0	4.1934	11076.5	0.2682
Simplex	5.5000	12000.0	4.0730	16000.0	0.2558
SQP	6.3000	14000.0	5.8181	14079.9	0.3075
Simplex	6.3000	14000.0	4.4170	11629.0	0.2679
SQP	6.9000	15000.0	4.0727	16000.0	0.2558
Simplex	6.9000	15000.0	4.4170	11629.0	0.2679
SQP	7.0000	15000.0	4.0727	16000.0	0.2558
Simplex	7.0000	15000.0	4.0720	16000.0	0.2558
GA	$L1[m]$ Bounds	$Mc/10^4[kg]$ Bounds	*Optimal Values*		
	$3.0 \sim 7.0$	$0.8 \sim 1.6$	4.0714	16000.0	0.2558

Optimization of Vehicle Suspensions

Figure 5.13 gives the PSDs (vehicle speed is 22.22[m/s] and rail profile class is 5) of the accelerations at the rear left corner of the car body based on the 20 DOF and 36 DOF models. For both of these models, the nominal design variables are offered in Table A.2 in Appendix A. It is shown that within the low frequency band (below 4 Hz), the PSDs for the two models are very close. However, within the frequency band from 4 Hz to 10 Hz (which includes the first bending frequency), the ride quality is significantly degraded by the carbody flexibility.

Figure 5.14 illustrates the PSDs of the accelerations at the same point of the car body

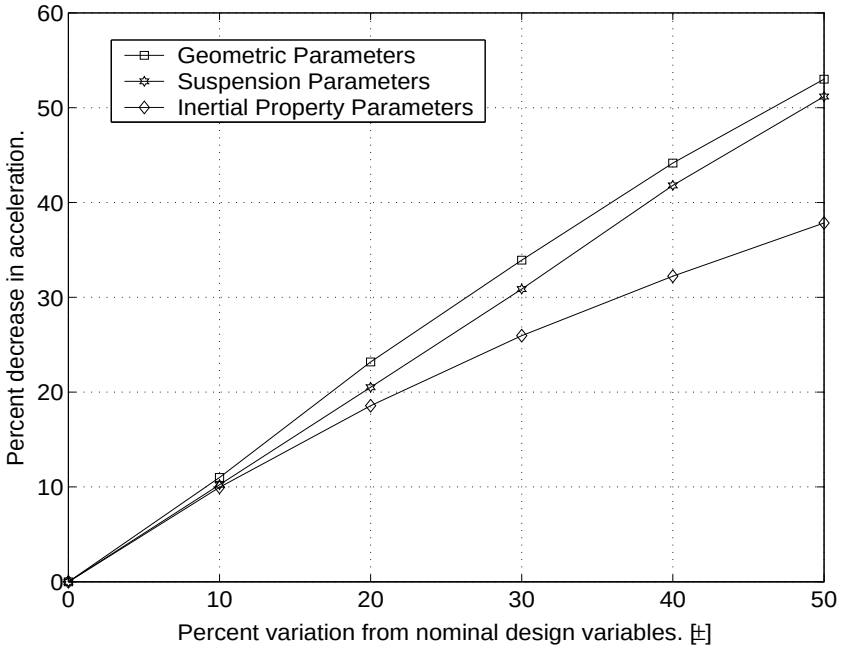

Figure 5.12: Effects of design variable sets on vertical ride quality

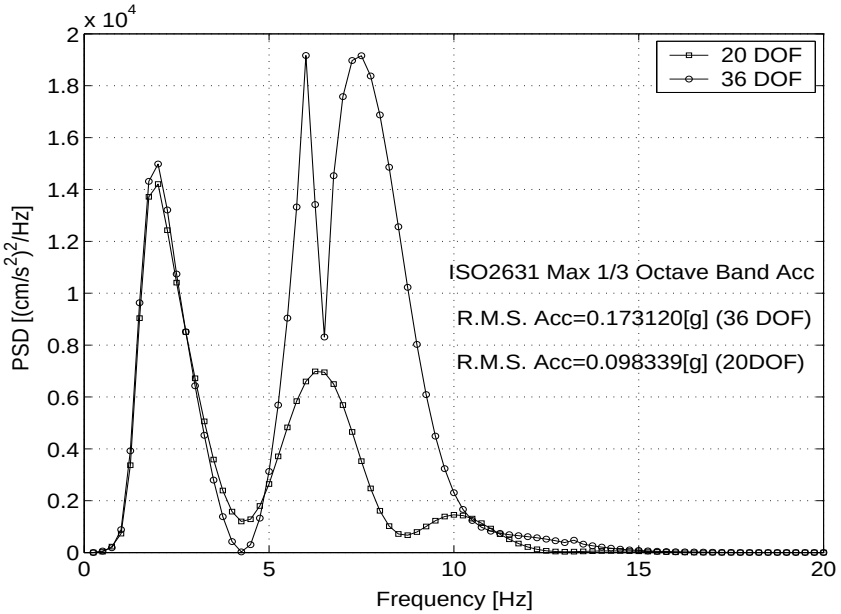

Figure 5.13: Power spectrum of acceleration for 20-DOF and 36-DOF vehicle models

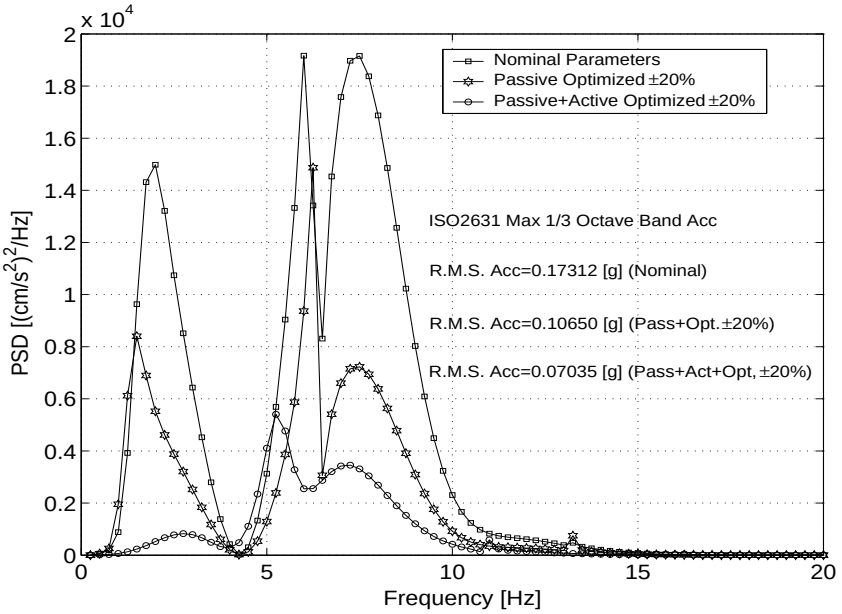

Figure 5.14: Power spectrum of acceleration of the 36 DOF vehicle model for nominal, passive optimized, and passive and active optimized cases

based on the 36 DOF model for three cases: the design variables take nominal values, the passive design variables are optimized (the 29 variables may vary by 20% from the nominal values), and the 31 passive and active design variables are optimized simultaneously. Results show that the ride quality (R.M.S. maximum weighted third-octave band acceleration) in the second and third case is improved by 38.5% and 59.4%, respectively, over the nominal case.

Numerical experiments show that it is difficult for the SQP and Simplex to handle vehicle suspension optimization problems with such a large number of design variables. Moreover, even using a number of combinations of the initial values of the design variables, it is still hard for these local search methods to find the global optimum that the GA can find with high reliability.

5.7 Summary

The GA is an effective optimization algorithm for ground vehicle ride quality analysis. When used with a numerical multibody dynamics package, e.g. A'GEM, the GA can be used to optimize vehicle suspensions including active control parameters by using complex realistic multibody vehicle models, even when many design variables are being considered. For these optimization problems, the algorithm has a high reliability for finding the global optimal point on objective function surfaces with many local optimal points. For this algorithm, sensitivity analysis for gradient information is superfluous, but the associated computational burden is heavy. Parallel processing, for which the GA is ideally suited, could be used to reduce the burden.

In contrast, in order to ensure the validation for SQP, even for local optimum searching, sensitivity analysis of the objective and constraint functions and scaling the optimization problems are necessary and important steps. When SQP works with a numerical multibody dynamics software for optimizing vehicle suspensions using linear vehicle models for ride quality analysis in frequency domain, since the objective function and constraint functions are algebraic functions in terms of design variables, numerical differentiation techniques can be used to offer gradients to the SQP reliably. Choosing an appropriate difference interval for specific design variables and using suitable scaling schemes for the optimization problem are important for the numerical differentiation technique to link the multibody dynamics software and SQP efficiently.

When both SQP and Simplex are used for optimizing vehicle suspensions using complicated multibody vehicle models, the reliability for these algorithms to find the global optimum decreases with the increase of the number of design variables. For a simple vehicle model, where the objective function and constraint functions can be explicitly expressed in terms of design variables and these functions are smooth, SQP outruns Simplex and GA. In both computation efficiency and reliability for global search, Simplex is generally a compromise option between SQP and GA.

Chapter 6

Optimization of Curving Performance

6.1 Introduction

The objective of this chapter is to demonstrate the feasibility and efficacy of applying numerical optimization approaches to rail vehicle suspension design with curving performance considered. The design optimization is to search optimal design variables so that the noise or wear, arising from misalignment of the wheelsets with the track, is reduced to the minimum level when the rail vehicle traverses curved track. The RACES routine for curving simulation from A'GEM is modified and combined with the GA. Due to the fact that a large number of parameters affect the curving behavior [9], the optimization problem is a multiple design variable problem. To accurately predict the curving behavior, time integration and nonlinear dynamic curving models [53] are used.

In the case of using a highly nonlinear dynamic curving model and time integration, the optimization problem may have lots of local optimal points. Furthermore, the gradient information for the objective function and constraint functions with respect to the design variables are difficult to obtain. Therefore, it is justified to use GAs as the optimization algorithm.

A conventional transit rail vehicle model with 21 DOF is optimized. In this chapter, the numerical simulation results of the optimization are offered. The selected objective function

is justified. Finally, the proposed optimization approach is applied to the investigation of the relative significance of different design parameters and different design parameter sets on curving performance.

6.2 Vehicle System Model

Figure 6.1 shows a schematic diagram of a 21 DOF dynamic model used for evaluating the curving performance of a conventional transit rail vehicle. The vehicle configuration is the same as that shown in Figure 4.1. The nominal geometry parameters, suspension stiffness and damping coefficients, and inertial property parameters are listed in Table A.2 in Appendix A. Relevant vehicle and track parameters are listed in Figure 6.1.

The programs RGEM and RACES from A'GEM are adopted to automatically generate the governing equations and to numerically integrate the nonlinear dynamic equations. RACES is used to simulate the vehicle as it travels from a tangent track, through a spiral of constantly decreasing radius, to a constant radius curve [53].

This model takes nonlinear wheel/rail geometry with two points of contact into account. In calculating creep forces, the following factors are considered:

1. The effect of wheel load changes on creep coefficients;
2. Creep force saturation due to combined actions of longitudinal, lateral, and spin creepages;
3. Nonlinear creep force relationships.

Nonlinear suspension elements, i.e. lateral bump stops at each secondary suspension, are included to restrict the relative lateral motion between the car body and the bogie frames.

The dynamic behavior of the model can be described by a set of simultaneous nonlinear ordinary differential equations. These equations can be integrated numerically to obtain the curving behavior. Along the specified track, the following curving performance indices can be obtained:

- Flange contact forces $\mathbf{F}_{flng}$: for the i^{th} wheel, the flange contact force is F_{flngi}, $i = 1, 2, ..., 8$.

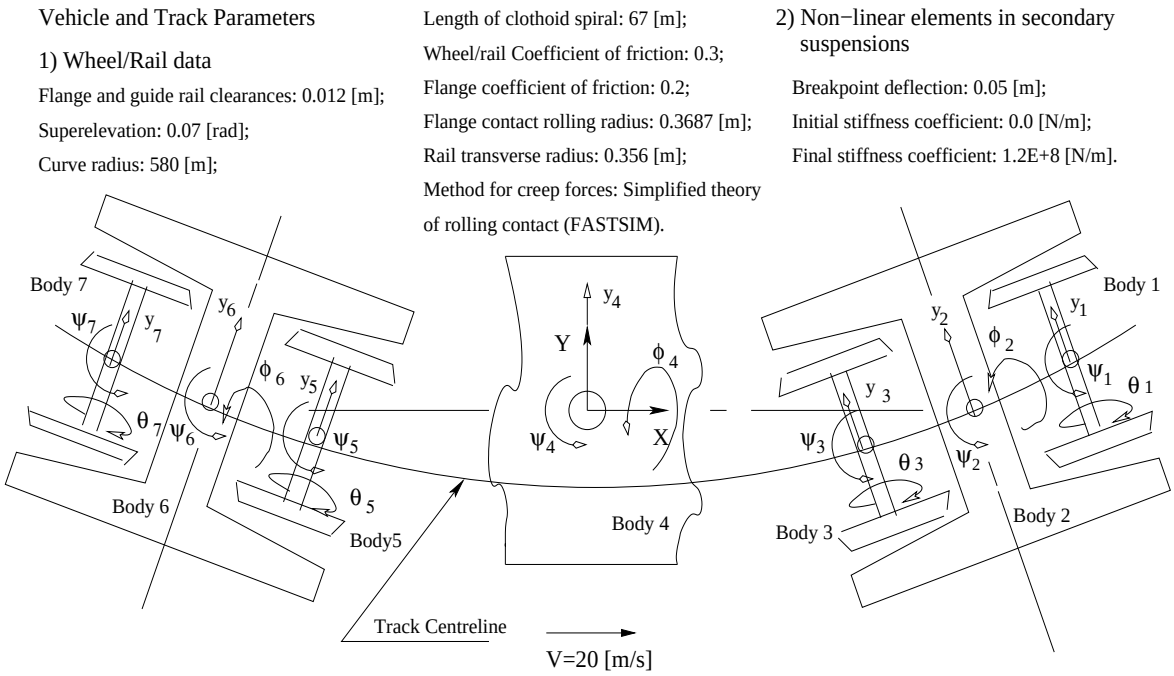

Figure 6.1: Schematic diagram, showing the degrees of freedom of the curving model

- Motions of vehicle components.
- Angles of attack $\mathbf{A}_{an}$: for the i^{th} wheelset, the angle of attack is A_{ani}, $i = 1, 2, 3, 4$. For a wheelset, the A_{ani} is defined as the angle between the axle of the wheelset and the corresponding radius of the curve.
- The ratios of lateral to vertical (L/V) contact forces $\mathbf{L}_v$: for the i^{th} wheel, the ratio is L_{vi}, $i = 1, 2, ..., 8$.
- The work $\mathbf{W}$ done by the wheel/rail force: for the i^{th} wheelset, the work is W_i, $i = 1, 2, 3, 4$.
- The saturation ratios $\mathbf{S}_{ra}$ of creep forces: for the i^{th} wheel, the ratio is S_{rai}, $i = 1, 2, ..., 8$. For a wheel, the S_{rai} is defined as the ratio of the creep force to the limited friction force.

6.3 Optimization Problem and Implementation

6.3.1 Performance Indices and Design Variables

In the optimization of curving performance, the design variables consist of suspension stiffness and damping coefficients ($\bar{\mathbf{S}}$), inertial property parameters ($\bar{\mathbf{I}}$), and geometric parameters ($\bar{\mathbf{G}}$). The total number of design variables is 29. The design variables are listed in Table A.2 in Appendix A. Note that the parameters k_{2x}, c_{2x}, a, and λ listed in Table A.2 are not selected as design variables.

As mentioned previously, since rail vehicles traveling on mainline intercity curves are often thought to negotiate curves by creep guidance, the curving performance may be measured in terms of the curve radius and the speed for which the wheelsets do not flange or slip. However, for urban transit rail vehicles, they must negotiate a large number of short radius curves. When transit vehicles negotiate narrow curves, there will be significant levels of slippage between rail and wheel, often leading to complete saturation of wheel/rail creep forces so that frictional forces are guiding the vehicles [5]. Additionally, although flange free curving is desirable, it is generally unachievable on tight curves. It was shown that flange contact occurs on curves with radii as large as 800.0 $[m]$ [9].

It is difficult to formulate a simple indicator for the complex curving behavior of rail vehicles, especially when the wheelset excursion has exceeded the flange clearance. Since the L/V ratio (L_{vi}, $i = 1, 2, ..., 8$) may be viewed as a wheel climb derailment indicator and the wheelset angle of attack (A_{ani}, $i = 1, 2, 3, 4$) can be used as a measure of wheel/rail wear and noise [66], we choose the L/V ratio and the wheelset angle of attack as the components of a performance index to represent the curving behavior of rail vehicle. Therefore, the objective function is:

$$minimize \quad F(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}) = \xi max(|\ \frac{A_{ani}(\bar{\mathbf{S}},\bar{\mathbf{I}},\bar{\mathbf{G}})}{\tilde{A}_{ani}}\ |) + \eta max(|\ \frac{L_{vk}(\bar{\mathbf{S}},\bar{\mathbf{I}},\bar{\mathbf{G}})}{\tilde{L}_{vk}}\ |) \tag{6.1}$$

where $i = 1, 2, 3, 4$, and $k = 1, 2, ..., 8$, ξ and η are constant weighting factors, and $\tilde{A}_{ani}$ and $\tilde{L}_{vk}$ are nominal values of the angle of attack and the L/V ratio.

6.3.2 Implementation of the Optimization Problem

As shown in Figure 6.2, the optimization is implemented by using the MechaGen program (a GA), and A'GEM software (RGEM and RACES programs for curving performance analysis).

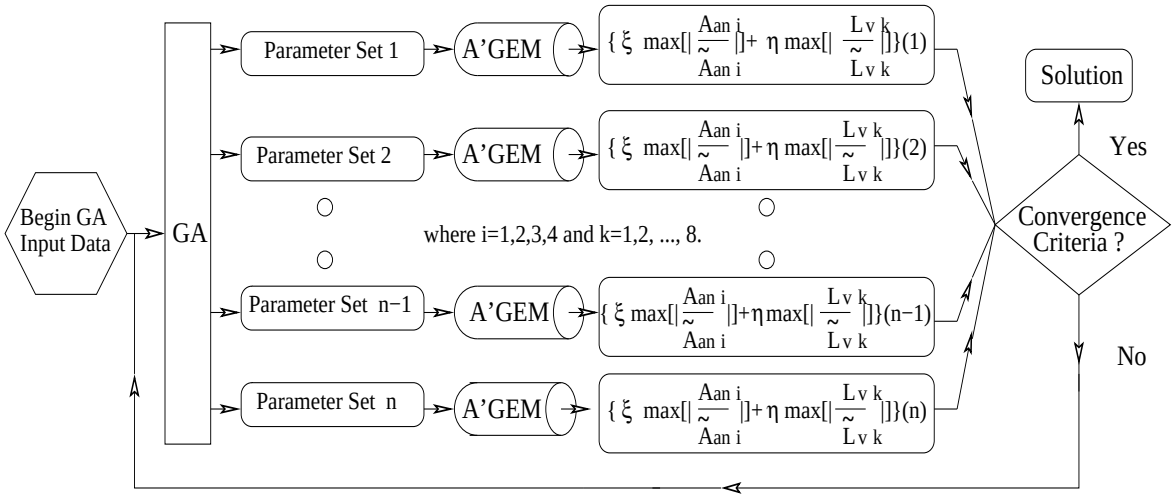

Figure 6.2: Schematic representation of the implementation of the optimization problem

As illustrated in Figure 6.2, each set of design variables of a population generated by the GA is forwarded side by side to the A'GEM programs for calculating the required performance indices of curve negotiation. With a given set of design variables, the corresponding A'GEM programs generate the required nonlinear differential equations of motion automatically. After performing numerical integrations in the time domain, the performance indices can be obtained directly. To ensure reliable measures of curving behavior of the rail vehicle concerned, after 18.0 $[m]$ of the entry point along the constant radius curve until the end point of the curve, the average value of angle of attack per wheelset and that of L/V ratio per wheel are measured. Based on equation (6.1), the corresponding performance index is calculated. Note that the total number of fitness values is the same as that of the individual design parameter sets in the population. At this point, if the convergence criteria are satisfied, the calculation terminates, otherwise these fitness values are returned to the GA. Based on the returned fitness values corresponding to the given

sets of design variables, the GA produces the next generation of design variable sets using reproduction, crossover and mutation. This procedure repeats until the optimized design variable set is found.

6.4 Results and Discussion

6.4.1 Justification for the Selected Objective Function

If both ξ and η in the objective function (6.1) take the value of 1.0, when the geometric parameters are selected as design variables and permitted to vary $\pm 10\%$ from their nominal values, we obtained the selected optimal results shown in Figures 6.3 and 6.5. Offered in Figure 6.3 is the relationship between the angle of attack and position on curve. Figure 6.5 illustrates the dependence of the work done on each wheelset versus the position on curve. For the purpose of comparison, the counterparts for which the design variables take the nominal values are also shown in Figures 6.4 and 6.6. Resulting from the optimization, for the axle that has the maximum angle of attack and for the axle on which the maximum work is done, the average steady-state angle of attack and work decrease by 27.94% and 20.84% (see Table 6.1), respectively. Table 6.1 also offers the results of L/V ratio, flange contact force, and the saturation ratio of creep forces for the axle or wheel which has the maximum value among the four axles or among the eight wheels respectively.

Table 6.1: Results of curving performance for nominal and optimized cases

cases	$A_{an}\ [deg]$	L_v	$W\ [J/m]$	$F_{flng}\ [N]$	S_{ra}
Nominal	0.1308	0.2267	$0.2114E+05$	$0.1478E+05$	0.9665
Optimal	0.0942	0.1795	$0.1776E+05$	$0.1236E+05$	0.9134

Note that the results for wheelset lateral displacements are not listed in Table 6.1 because flange contact occurs for both optimal and nominal cases. Thus, the maximum lateral displacements for both cases are the same, and approximately equal to the flange clearance, as shown in Figures 6.7 and 6.8 for the optimal case and nominal case, respectively.

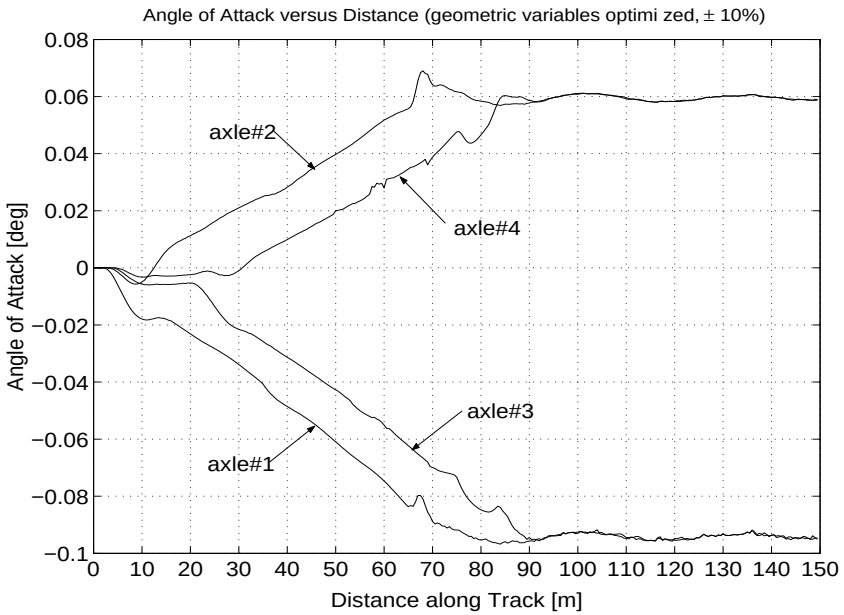

Figure 6.3: Angle of attack versus position on curve for optimal case

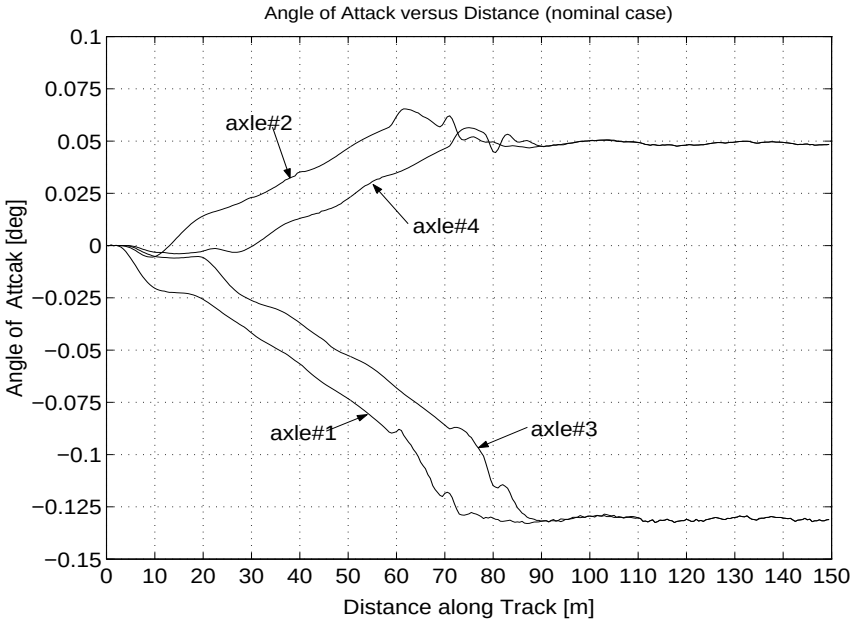

Figure 6.4: Angle of attack versus position on curve for nominal case

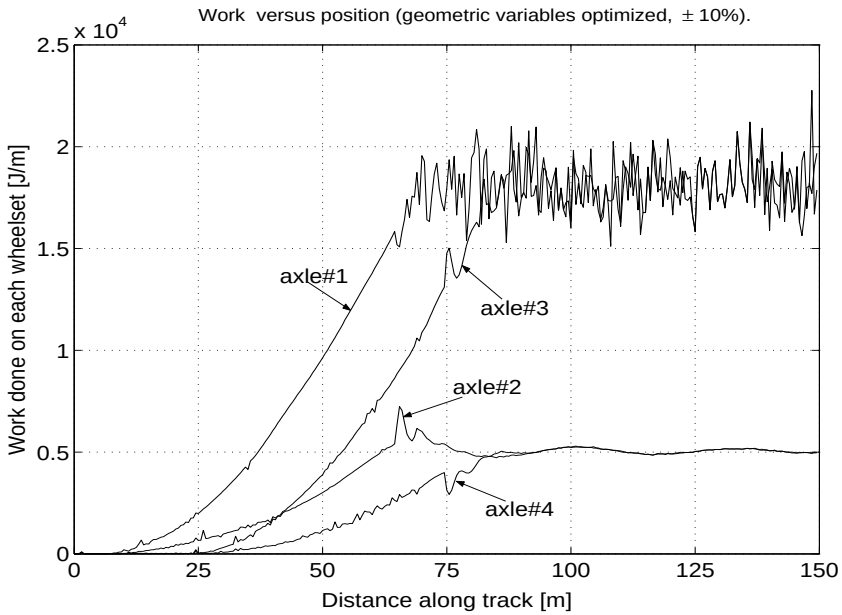

Figure 6.5: Work done on wheelset versus position on curve for optimal case

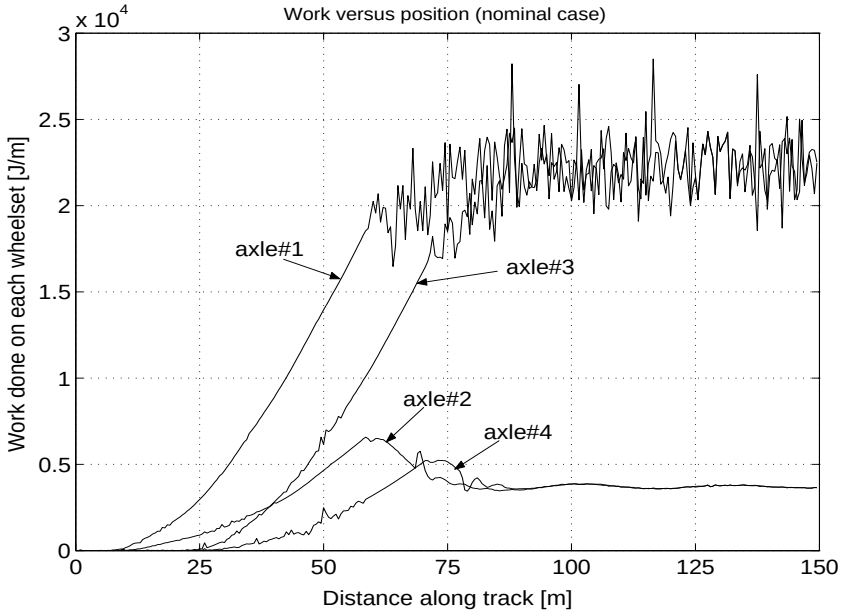

Figure 6.6: Work done on wheelset versus position on curve for nominal case

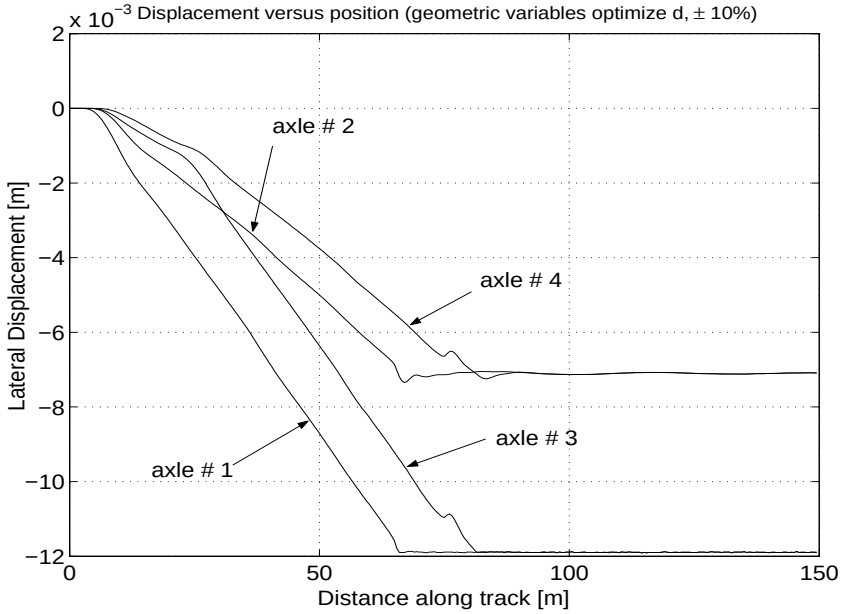

Figure 6.7: Lateral axle displacement versus position on curve for optimal case

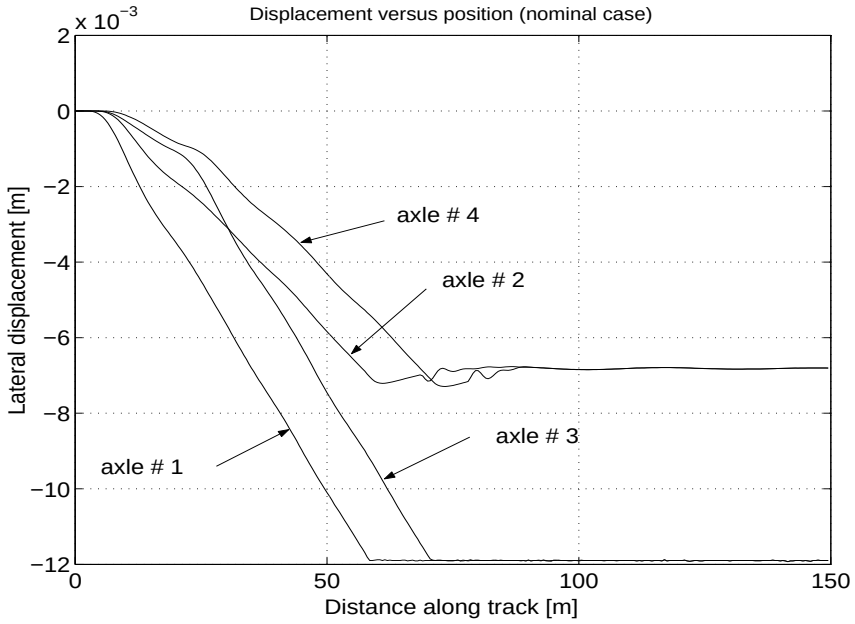

Figure 6.8: Lateral axle displacement versus position on curve for nominal case

To investigate the trends of the optimized design variables, Table 6.2 offers the optimized geometric variables, which are permitted to vary by ±20% from the nominal values, and the corresponding nominal values.

Table 6.2: Optimized geometric variables (permitted to vary by ±20% from their nominal values) †

	$b\ [m]$	$L_1\ [m]$	$g\ [m]$	$d\ [m]$	$L_3\ [m]$	$L_2\ [m]$	$L_4\ [m]$
Nominal Values	1.0420	8.3200	0.5860	0.8130	0.0000	0.3050	0.8150
Upper Bounds	1.2504	9.8760	0.7032	0.9756	0.0000	0.3660	0.9780
Lower Bounds	0.8336	6.5840	0.4688	0.6504	0.0000	0.2440	0.6520
Optimal Values	0.8336	7.4271	0.4688	0.6707	0.0000	0.2692	0.6558

† See Figure 4.1 for the definition of the symbols in the table.

Based on the results offered in Table 6.2, we notice that the half of wheelbase on bogie (b), vertical distance from car body center of mass to secondary suspension (L_4), and vertical distance from bogie center of mass to secondary suspension (L_2) take lower bound values or approach the lower bound. Moreover, the half of bogie spacing (L_1) takes a lower value. The results obtained are consistent with the lateral dynamics design requirements: for improving curving performance of rail vehicles, we should choose a short wheelbase on the bogie, a short distance between bogie centers, and a low center of vehicle mass [35].

According to the above analysis, we can expect that the objective function in the form of (6.1) can well reflect curving performance including work done on wheelset, flange force, saturation ratios of creep forces, angle of attack, and L/V ratios.

6.4.2 Sensitivity Analysis for Design Variables and Design Variable Sets

In the research, we notice that different design variables or different design variable sets impose different effects on curving performance. By optimizing design variables for improving curving performance, the relative significance of different design variables or different design variable sets are investigated.

Based on the method recommended by Eberhard et al. [42], but using the GA instead of Simulated Annealing Algorithms, we can easily identify important design variables in the design optimization. Figure 6.9 shows the selected results where only geometric parameters are chosen as design variables for improving curving performance. The design variables are permitted to vary by $\pm 50\%$ from their nominal values. According to the dynamic analysis [35], it was found that in order to reduce misalignment of wheelset axles on curves, the half of bogie wheelbase (b) should be as small as possible. Numerical results show that secondary suspension lateral spacing (g) has a minor effect on curving performance. In Figure 6.9, it can be seen that most samples for points b are taken in the vicinity of the optimal variable value (lower bound value), i.e. $b_{Opt} = 0.521[m]$. This further agrees with the above mentioned lateral dynamics design requirements. Variable g, however, illustrates a different behavior. The sampling density around optimized g is only slightly higher than in the other areas, but the overall distribution is somewhat uniform. The results show that the feasible values of g do not change the objective function value very much.

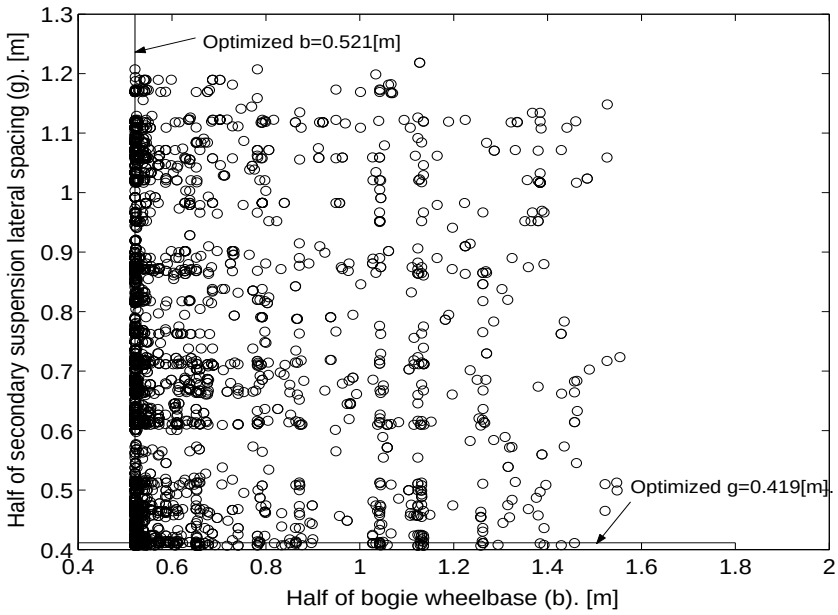

Figure 6.9: Variable distribution to identify important variables (b, g)

Figures 6.10, 6.11, and 6.12 show the selected numerical results where only primary

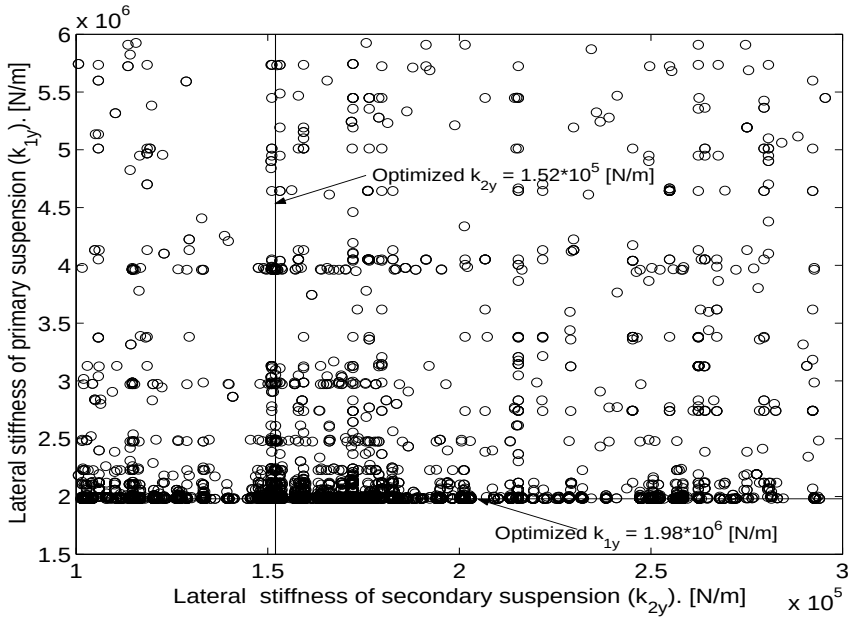

Figure 6.10: Variable distribution to identify important variables (k_{2y}, k_{1y})

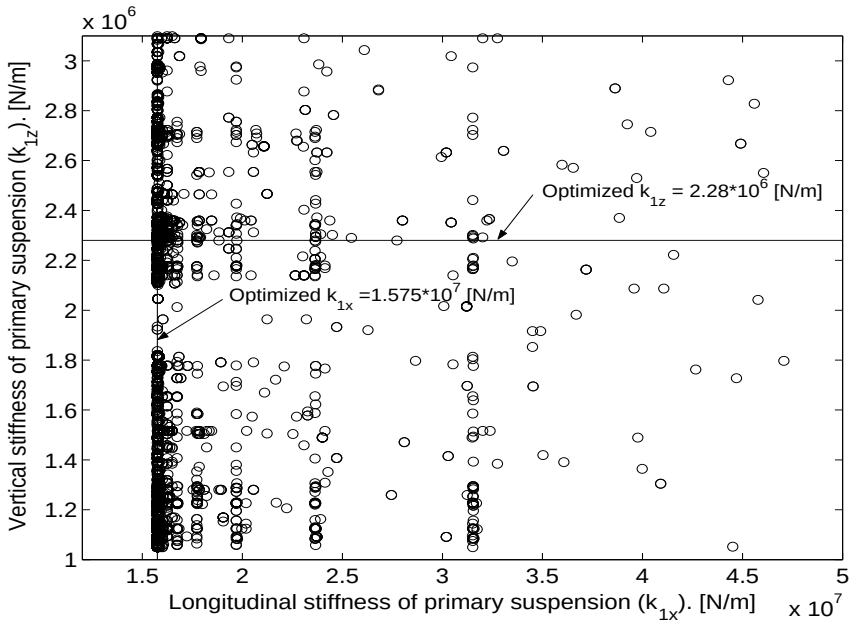

Figure 6.11: Variable distribution to identify important variables (k_{1x}, k_{1z})

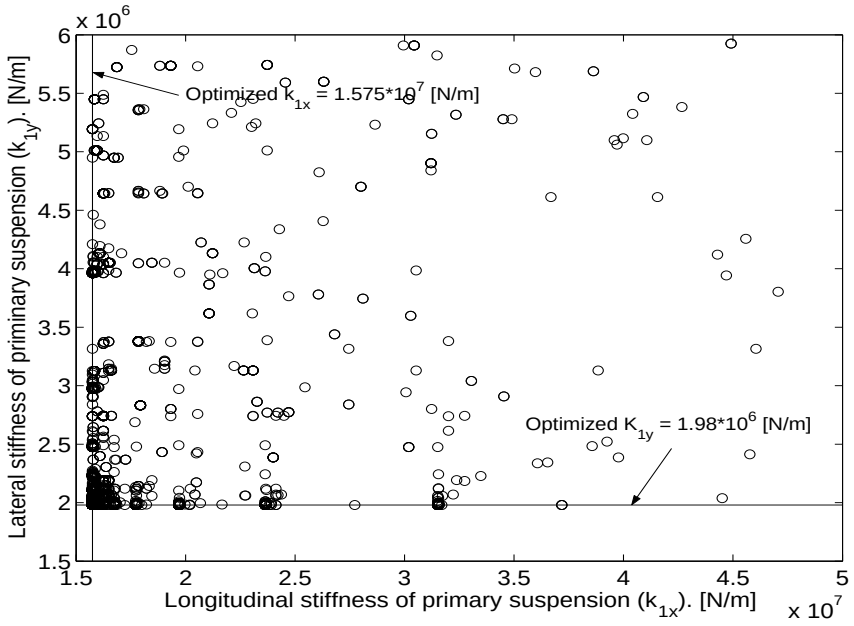

Figure 6.12: Variable distribution to identify important variables (k_{1x}, k_{1y})

and secondary suspension stiffness and damping coefficients are chosen as design variables for improving curving performance. Again the design variables are allowed to change by $\pm 50\%$ from their nominal values. As discussed above, in Figures 6.10, 6.11, and 6.12, most samples for points k_{1y}, k_{1x} are taken in the vicinity of the optimal variable values, i.e. the lower bound values of the feasible domains for k_{1y} and k_{1x} respectively. On the other hand, the sample densities around optimized k_{2y} and k_{1z}, like the case of design variable g analyzed previously, are only a little bit higher than the other areas and the overall distributions are uniform. Again the feasible values of k_{2y} and k_{1z} do not significantly change the objective function value. The results obtained demonstrate that in order to improve curving performance, the longitudinal and lateral stiffness coefficients of primary suspensions of the vehicle should take values as low as possible. This agrees well with the conclusions offered by Wickens [142].

It is clear that by investigating the sampling distribution of each variable, one can judge the global performance and sensitivity over its feasible domain. This global performance and sensitivity over a design variable's feasible domain is, however, difficult to judge by

using its gradient information in a local sense. Thus, with the assistance of GAs, we can visualize the nonlinear functional relation between design criteria and design variables and predict the topology of the criteria space from the space of feasible design variables.

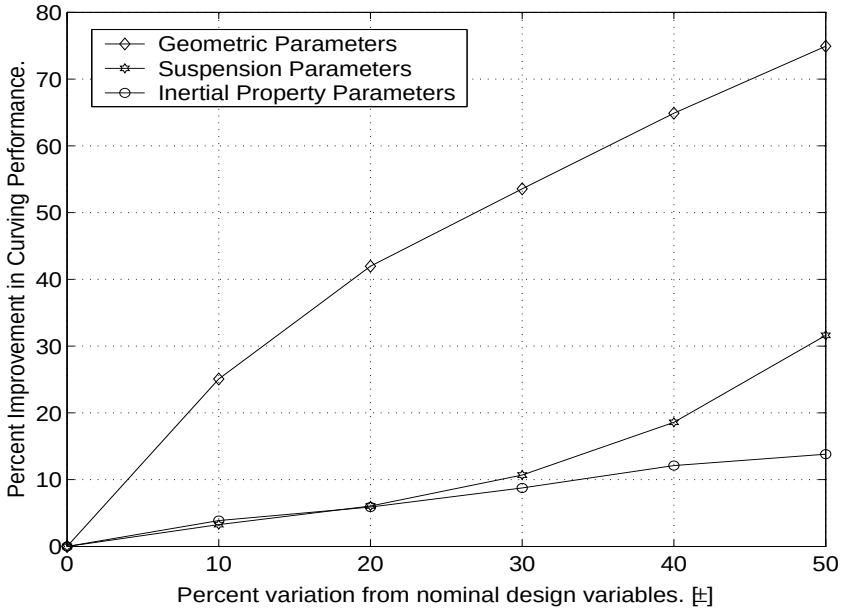

Figure 6.13: Effects of design variable sets on curving performance

Not only different design variables but also different design variable sets impose different effects on curving performance. Figure 6.13 illustrate the effects of different design variable sets, i.e. geometric parameters, suspension parameters (stiffness and damping coefficients), and inertial property parameters, on curving performance. As shown in Figure 6.13, if the geometric parameters, suspension parameters, and inertial property parameters are permitted to vary by $\pm 30\%$ during the optimization, these three optimized design variable sets will lead to 53.56%, 10.69%, and 8.75% decrease of the objective function described in equation (6.1), respectively. One may see that, as in the case of lateral stability, for curving performance, the variations in the geometric variable set (e.g. suspension locations) have a much greater influence than variations in the suspension variable set or in the inertial property variable set. Moreover, the effects of the variations of the latter two variable sets are very close when the percent variation from nominal design variables is less than 30%.

Thus, for the purpose of improving curving performance, varying geometric parameters may be an effective approach.

6.5 Summary

The chapter proposes a numerical optimization approach, that can automate the selection of appropriate design variable values for rail vehicle design with respect to curving performance. In the proposed approach, a global optimization algorithm, a GA, is combined with a multibody dynamics modelling program, A'GEM, so that the generation of governing equations of motion for the complex realistic nonlinear dynamic rail vehicle models and the search for optimal design variables can be carried out automatically. To facilitate the optimization, we choose angle of attack on wheelsets and ratios of lateral to vertical forces on wheels as the only factors for the objective function.

To demonstrate the feasibility and efficacy of the proposed approach, this approach is applied to the optimization of a conventional transit rail vehicle with respect to curving performance using a nonlinear dynamic curving model with 21 DOF. Numerical results show that the GA is an effective algorithm for rail vehicle design for improving curving performance when working with A'GEM. The combination of angle of attack and L/V ratio well reflects curving performance and is suitable as an objective function or a factor of fitness function. By means of the GA, we can also investigate the sampling distribution of each design variable, and judge the design variable's global performance and sensitivity over its whole feasible domain. Compared with inertial and suspension parameters sets, the geometric parameter set has the most significant effect on curving performance.

Chapter 7

Multidisciplinary Optimization of Stability, Ride Quality, and Curving Performance

7.1 Introduction

The objective of this chapter is to synthesize the techniques discussed in Chapters 4, 5, and 6, and to show how multidisciplinary optimization methods combining a genetic algorithm, sequential quadratic programming, dynamic mode tracking, and a multibody dynamics modelling package (A'GEM) can effectively handle the conflicting requirements of rail vehicle design problems. Lateral stability, curving performance, and ride quality are assessed using realistic multibody models from A'GEM, and summarized in a multicriteria objective function. By coordinating the conflicting requirements from lateral stability, curving performance, and vertical ride quality at the system level, the suspension, geometric, inertial, and control parameters for a rail vehicle with passive and active suspensions are optimized simultaneously.

The multidisciplinary optimization (MDO) method used in this chapter combines the 17 DOF lateral stability model used in Chapter 4, the 36 DOF vertical ride quality model used in Chapter 5, and the 21 DOF nonlinear dynamic curving performance model used in Chapter 6 into a synergistic whole. For the lateral stability problem, the dynamic

equations for the 17 DOF model of the rail vehicle are generated and linearized by A'GEM; the corresponding eigenvalue problem is solved. To evaluate curving performance, A'GEM is used to generate and numerically integrate the nonlinear dynamic equations for the 21 DOF nonlinear model of the same rail vehicle. For the problem of vertical ride quality, the frequency response of the 36 DOF model with car body flexibility to stochastic inputs is determined by A'GEM.

From the results of these three discipline analyses, a weighted objective function is constructed for the three-criteria optimization problem. By solving the optimization problem for a set of weighting factors, the GA provides a family of Edgeworth-Pareto (EP-) optimal solutions [42].

7.2 Vehicle System Models

It is important to note that the models for lateral stability, curving performance and vertical ride quality all correspond to the same design configuration; however, for the ride quality model the car body flexibility is considered. The rail vehicle's configuration is shown in Figure 7.1. For the ride quality model with 36 DOF, as described in Chapter 5, the car body is divided into 5 identical rigid bodies. The adjacent car body sections are connected by a group of bending, torsion and shear springs. Each suspension component consists of a parallel spring and damper, with the stiffness and damping coefficients in the three coordinate directions. In the secondary suspension, both vertical and lateral directions are equipped with 2 active components per bogie, respectively. The nominal design variables, which are adopted from Reference [53], are listed in Table A.2 in Appendix A. Note that nominal wheel radius of r_0 $(0.356[m])$ and conicity λ (0.1) as well as the half-distance between contact points a $(0.756[m])$ are fixed in the research.

The lateral stability model with 17 DOF is the same as that described in Chapter 4 except that 2 active elements, as shown in Figure 7.1, are introduced per bogie in the secondary suspension in the lateral direction. A control strategy proposed by Celniker et al. [27] is adopted. The two active elements per bogie operate out-of-phase to produce resultant yaw torques on the leading or trailing bogie:

$$T_{zi} = -K_{fl}\dot{y}_i - K_{tl}\dot{\psi}_i \tag{7.1}$$

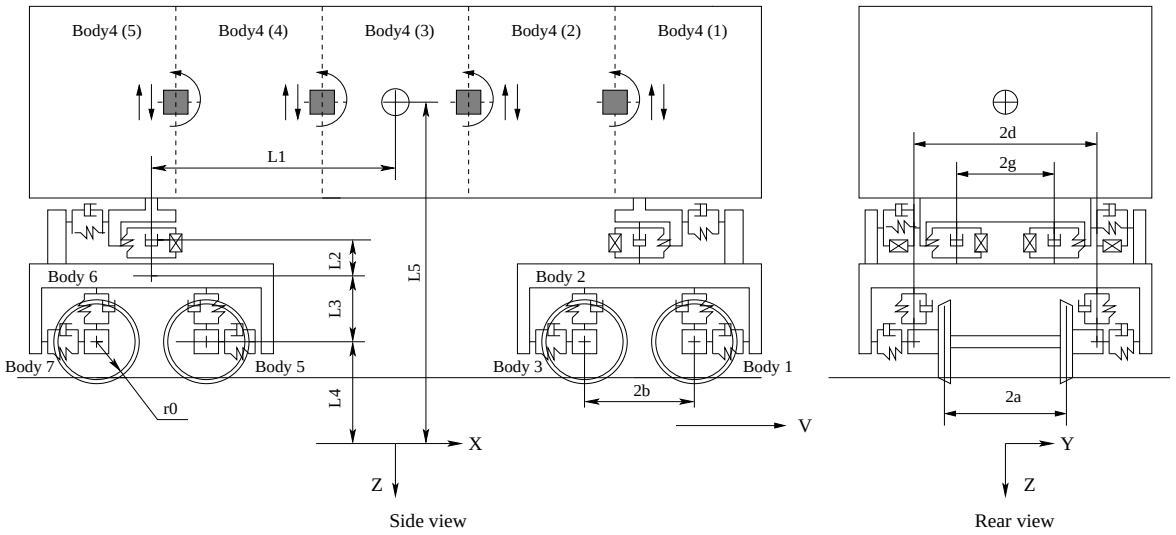

Figure 7.1: Rail vehicle configuration for dynamic models

where y_i and $\dot{\psi}_i$ $(i = 2, 6)$ are defined in Chapter 4, and K_{fl} and K_{tl} are the lateral rate gain and yaw rate gain, respectively.

The 21 DOF curving model and 36 DOF vertical ride quality model with active elements are exactly the same as those described in Chapters 6 and 5, respectively.

7.3 Multidisciplinary Optimization and Implementation

7.3.1 Design Optimization Approach

For the multidisciplinary optimization problem formulation, the optimization method used in Chapter 4, i.e. optimization of the lateral stability, matches the definition of the All-in-One (A-i-O) method described in Chapter 3. Figure 7.2 shows the A-i-O formulation for optimizing the lateral stability of the corresponding 17 DOF vehicle model. With the introduction of active elements, the design variables will be $\mathbf{X}_d = \{\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}}\}$, where $\bar{\mathbf{S}}$, $\bar{\mathbf{I}}$, $\bar{\mathbf{G}}$, and $\bar{\mathbf{A}}$ are suspension parameters, inertial parameters, geometric parameters, and control parameters, respectively. As shown in Figure 7.2, with a set of design variables

$\mathbf{X}_d$, the three coupled analysis disciplines, i.e. Multibody Dynamics (A'GEM), Dynamic Mode Tracking (DMT), and Critical Speed Identification (SQP), will cooperate to find the corresponding critical speed V_c. During the k^{th} iterative search for the critical speed, SQP sends the potential critical speed V_{sqp}^k to A'GEM, assembling the speed V_{sqp}^k with relevant matrices already obtained. A'GEM may obtain the required system matrix $\mathbf{A}^k$ and offer the matrix to DMT, then DMT will perform mode tracking and return to SQP the required real parts of corresponding eigenvalues $Re(\mu_i)(V_{sqp}^k)$, for all $i = 1, 2, ..., n$. This process will continue until the corresponding critical speed V_c and corresponding system matrix A_c are determined. At the end of the process, the resulting V_c, $\mathbf{A}_c$, and $Re(\mu_i)(V_c), i = 1, 2, ..., n$, are returned to the system optimizer, a GA, for further use.

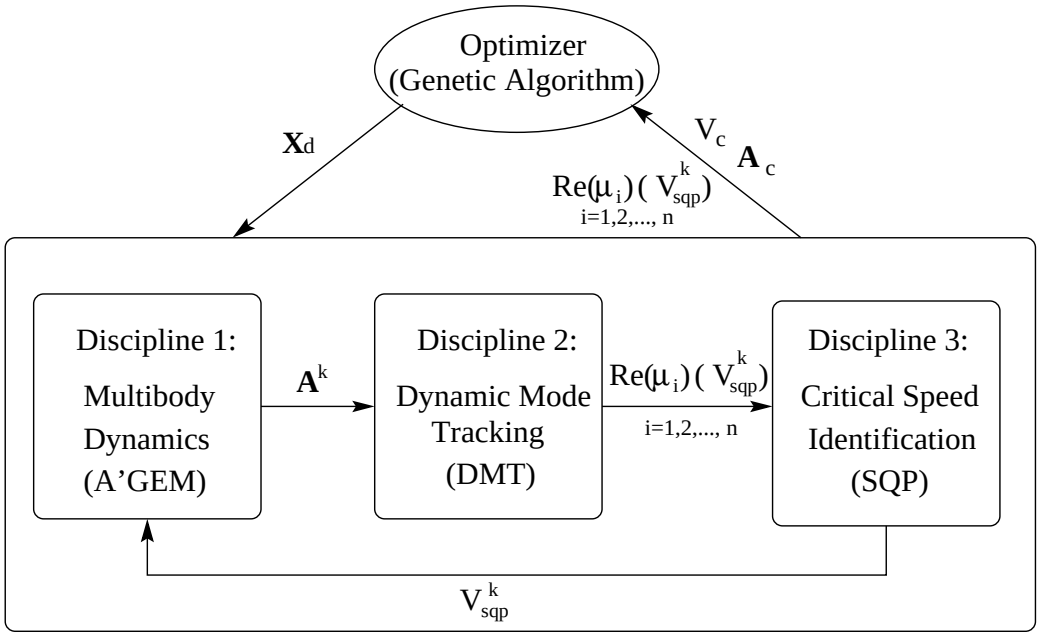

Figure 7.2: All-in-One (A-i-O) formulation for optimizing the lateral stability

Similarly, the optimization of curving performance discussed in Chapter 6 can also be regarded as an application of the A-i-O method, because the program RACES (from A'GEM) consists of linear/nonlinear wheel/rail geometry models, linear/nonlinear wheel/rail creep force models, and all these models together with the genetic algorithm and module R'GEM (from A'GEM) for automatic generation of the equations of motion make the system a syn-

ergistic whole. In this MDO method, once the required design variables are provided to the above coupled models or analysis disciplines, a complete multidisciplinary analysis (MDA) is carried out via a fixed-point iteration with those values of the design variables to obtain the system (MDA) output variables that are later utilized for evaluating the objective function value and the required constraints.

As shown in Figure 7.3, three boxes represent three disciplines, i.e. the vertical ride quality, curving performance, and lateral stability. For the discipline of the lateral stability, the three subdisciplines or subsystems (multibody dynamics, dynamic mode tracking, and critical speed identification) and their coupling relations are also illustrated. With the systems shown in Figure 7.3, the individual discipline feasible (IDF) method is used to synthesize the three disciplines at the system level and the A-i-O method is applied, for example, to the discipline of the lateral stability for making its three subdisciplines a synergistic whole. At the system level, a GA is used as the optimizer. Due to the fact that the formulation method used here is a combination of the IDF and the A-i-O methods, we call it a hybrid MDO method. With the hybrid MDO method and the selected optimization algorithms, the optimal passive and active design variables are searched in the design space so that the rail vehicle's lateral stability, vertical ride quality, and curving performance can be optimized simultaneously.

As shown in Figure 7.3, for the three analysis disciplines of the lateral stability, curving performance, and vertical ride quality, the corresponding analysis solvers are denoted as V_c, C_p, and R_q respectively. A comparison of Figure 7.3 with Figure 3.2 reveals that in the case of Figure 7.3, the specific analysis variables (vector $\mathbf{X}_Y$) representing communication, or coupling, between analysis disciplines vanish. Furthermore, there are no explicit interdisciplinary mappings (vector $\mathbf{G}$) among the three disciplines. However, the three disciplines, are coupled by means of the original design variable set $\mathbf{X}_d$ and their implicit interdisciplinary mappings are coordinated and manipulated by the optimizer at the system or discipline level. If we treat the three disciplines as three black boxes, the optimization method shown in Figure 7.3 is equivalent to the "Multi-Criteria Multi-Model Optimization (MMO)" method proposed by Bestle and Eberhard [18]. Obviously, based on the above analysis, the so called system-level MMO method is a specific case of the IDF method.

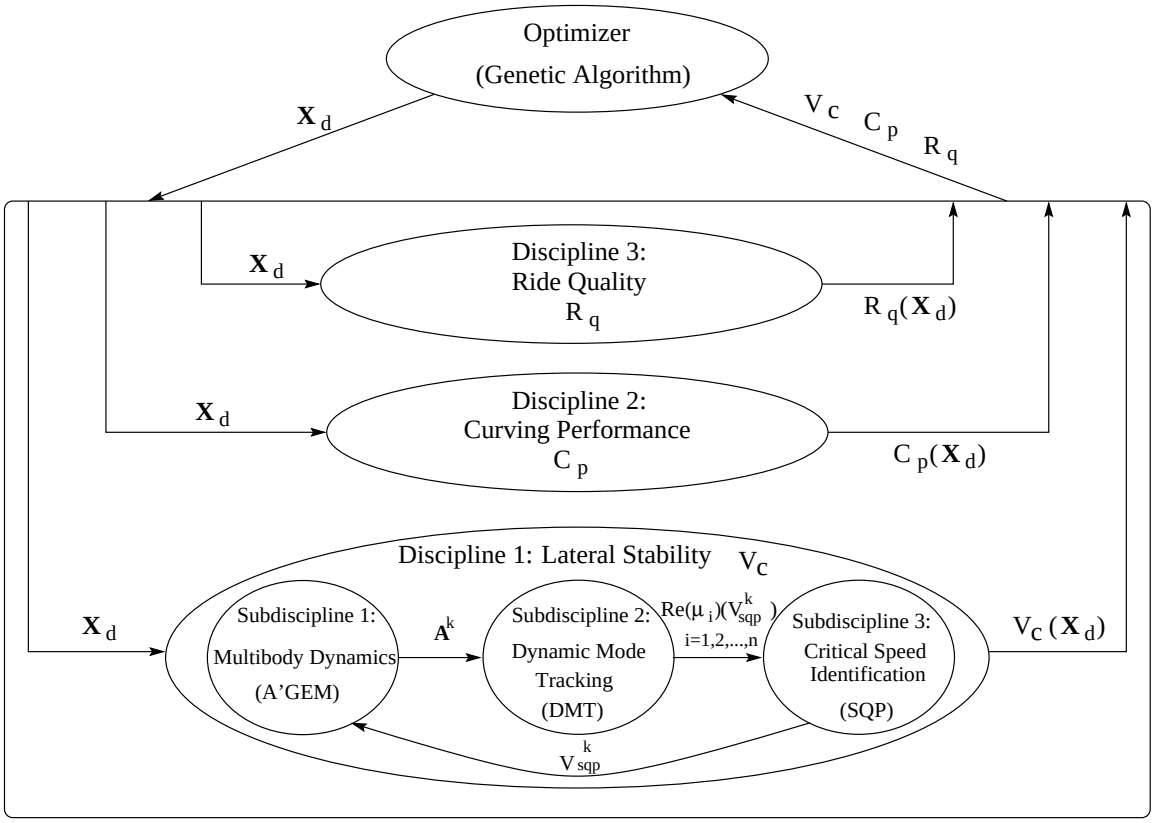

Figure 7.3: Hybrid MDO method combining IDF and A-i-O for optimizing the lateral stability, curving performance, and vertical ride quality simultaneously

7.4 Optimization Problem and Implementation

7.4.1 Objective Function, Constraints and Design Variables

For the combined rail vehicle model including the lateral stability model, dynamic curving model, and vertical ride quality model, the design variables consist of suspension stiffness and damping coefficients ($\bar{\mathbf{S}}$), inertial property parameters ($\bar{\mathbf{I}}$), geometric parameters ($\bar{\mathbf{G}}$), and active control parameters ($\bar{\mathbf{A}}$). The total number of design variables reaches 33 including 29 vehicle system parameters and 4 active control parameters. The vehicle system parameters are listed in Table A.2 in Appendix A. Note that the parameters k_{2x}, c_{2x}, a, and λ listed in Table A.2 are not selected as design variables. The four control parameters

are K_{tr}, K_{fr}, K_{fl}, and K_{tl} (see equation set (5.20) and equation (7.1)).

For the lateral stability discipline, based on equations (4.12) and (4.13), the objective function and constraints may be expressed as

$$\begin{cases} maximize & V_c(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}}, V) \\ subject\ to & Re(\mu_i)(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}}, V) \leq 0, \quad \text{i=1,2,...,m} \end{cases} \tag{7.2}$$

For the ride quality discipline, the function A_R, to be minimized is a combination of R.M.S. acceleration values at different points of the car body and secondary suspension working spaces:

$$A_R = \left[\int_{\omega_l}^{\omega_u} S_{\ddot{z}_{4(i)}}(\omega)\, d\omega\right]^{1/2} + \upsilon max[0, (\left[\int_{\omega_l}^{\omega_u} S_{h_k}(\omega)\, d\omega\right]^{1/2} - h_k)] \tag{7.3}$$

where $i = 1, 2, ..., 5$, υ is a weighting fact, and A_R is a function of design variables $\bar{\mathbf{S}}$, $\bar{\mathbf{I}}$, $\bar{\mathbf{G}}$, and $\bar{\mathbf{A}}$. The other symbols are defined in Section 5.5 of Chapter 5.

For the curve negotiation discipline, the function to be minimized is a combination of angles of attack $\mathbf{A}_{an}$, and L/V ratios $\mathbf{L}_v$. Based on equation (6.1), the function can be described as:

$$minimize \quad \xi max(\mid \frac{A_{ani}(\bar{\mathbf{S}},\bar{\mathbf{I}},\bar{\mathbf{G}},\bar{\mathbf{A}})}{\tilde{A}_{ani}} \mid) + \eta max(\mid \frac{L_{vk}(\bar{\mathbf{S}},\bar{\mathbf{I}},\bar{\mathbf{G}},\bar{\mathbf{A}})}{\tilde{L}_{vk}} \mid) \tag{7.4}$$

where $i = 1, 2, 3, 4$, $k = 1, 2, ..., 8$, $\tilde{A}_{ani}$ and $\tilde{L}_{vk}$ are the angle of attack and L/V ratio when $A_{ani}(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}})$ and $L_{vk}(\bar{\mathbf{S}}, \bar{\mathbf{I}}, \bar{\mathbf{G}}, \bar{\mathbf{A}})$ take nominal values, respectively, and ξ and η are weighting factors.

Obviously, the design optimization of the combined dynamic model with respect to the above three criteria, which are described in equations (7.2), (7.3), and (7.4), leads to a multicriteria optimization problem. Based on the concept of scalarization discussed in Chapter 3, by introducing an appropriate utility function, we can obtain a weighted objective function as follows:

$$minimize \quad \nu_1\{\xi max(\mid \frac{A_{ani}}{\tilde{A}_{ani}} \mid) + \eta max(\mid \frac{L_{vk}}{\tilde{L}_{vk}} \mid)\} + \nu_2(\frac{A_R}{\tilde{A}_R}) + \nu_3(\frac{\tilde{V}_c}{V_c}) \tag{7.5}$$

where ν_1, ν_2, ν_3 are the weighting factors, $i = 1, 2, 3, 4$, $k = 1, 2, ..., 8$, and $\tilde{A}_R$ and $\tilde{V}_c$ are the nominal values of A_R and V_c respectively.

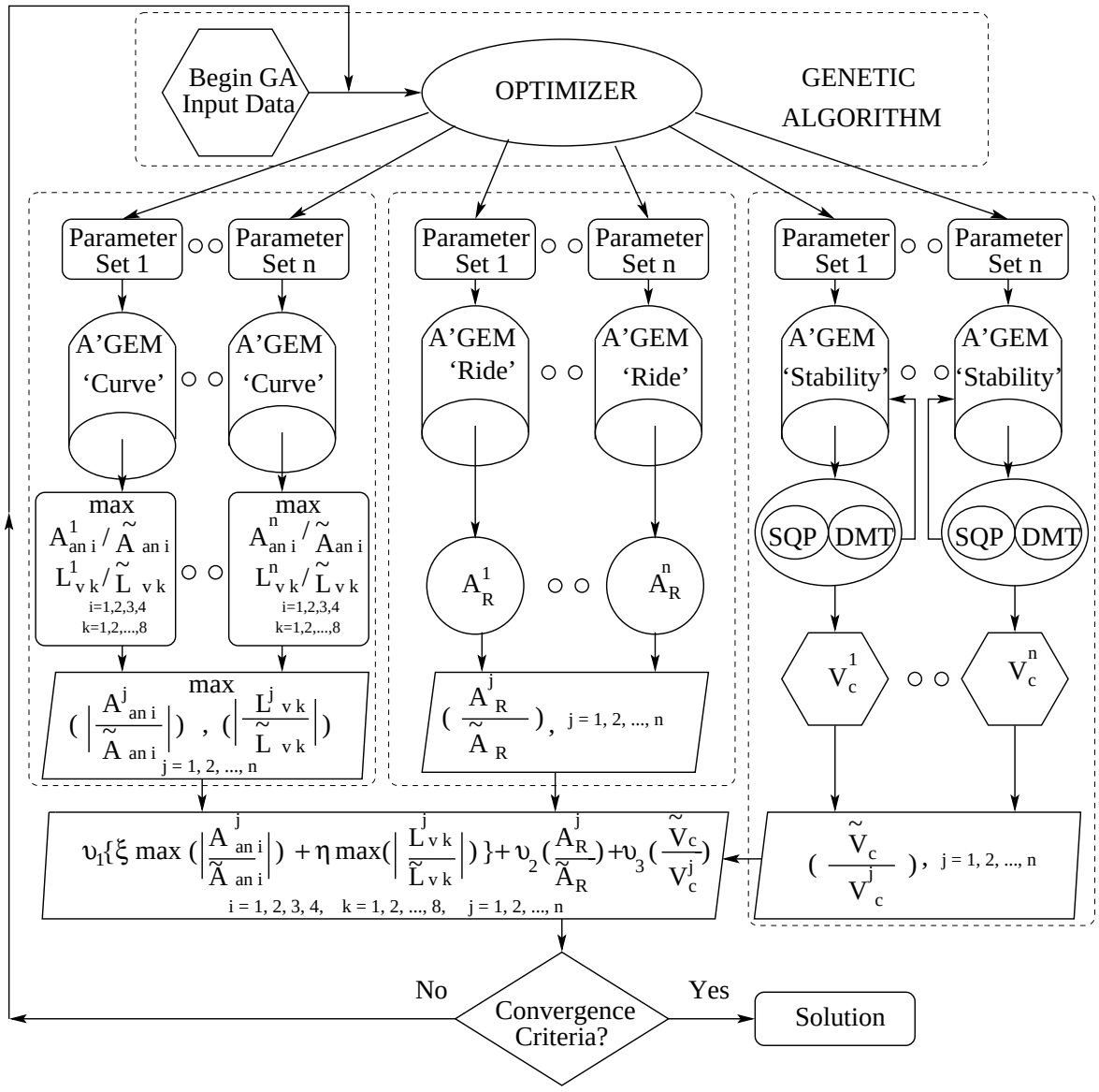

Figure 7.4: Schematic representation of the implementation of the optimization problem

7.4.2 Implementation of the Optimization Problem

As shown in Figure 7.4, the hybrid MDO method combining IDF and A-i-O discussed previously is implemented using:

- MechaGen program (a GA),
- E04UCF routine (a SQP) from the NAG library,
- Dynamic Mode Tracking (DMT) technique,
- A'GEM 'Stability' module (STABLE program for lateral stability analysis),
- A'GEM 'Ride' module (RLRIDE program for vertical ride quality analysis), and
- A'GEM 'Curve' module (RACES routine for curving performance analysis).

As shown in Figure 7.4, each set of design parameters of a population generated by the GA is forwarded to the corresponding A'GEM module for calculating the required performance indices for curve negotiation, vertical ride quality, and lateral stability. With the given set of design variables, the corresponding programs of A'GEM generate the required equations of motion or system matrices automatically. For the cases of curve negotiation and vertical ride quality, after numerical integration in the time domain and necessary transformation in the frequency domain, respectively, the performance indices can be obtained directly. For lateral stability, however, with the system matrix generated in the form of equation (4.4), the SQP and DMT are used to determine the critical speed. Then the corresponding fitness value is obtained by converting the vector optimization problem into a scalar optimization problem using the concept of scalarization by introducing an utility function in the format of (7.5). Note that the scalarization is performed at the system level so that the performance indices of the three disciplines (i.e. the lateral stability, curving performance, and vertical ride quality) are coordinated and manipulated at the system or discipline level by the genetic algorithm. It should be noted that the total number of fitness values is the same as that of the individual design parameter sets in the population. At this point, if the convergence criteria are satisfied, the calculation terminates, otherwise these fitness values are returned to the GA. Based on the returned fitness values corresponding to the given sets of design variables, the GA produces the next generation of design variable sets using reproduction, crossover and mutation. This procedure repeats until the optimized design variable set is found.

Table 7.1: Optimized suspension variables (permitted to vary by ±20% from their nominal values) †

	k_{1x} $[N/m]$	k_{1y} $[N/m]$	k_{1z} $[N/m]$	c_{1x} $[N/m/s]$	c_{1y} $[N/m/s]$
Nominal Values	$3.1500 \cdot 10^7$	$3.9600 \cdot 10^6$	$2.1000 \cdot 10^6$	666.00	5220.00
Upper Bounds	$3.7800 \cdot 10^7$	$4.7520 \cdot 10^6$	$2.5200 \cdot 10^6$	799.20	6264.00
Lower Bounds	$2.5200 \cdot 10^7$	$3.1680 \cdot 10^6$	$1.6800 \cdot 10^6$	532.80	4176.00
Cp Optimized	$2.5200 \cdot 10^7$	$3.1680 \cdot 10^6$	$2.2964 \cdot 10^6$	654.80	4212.80
Ls Optimized	$3.7800 \cdot 10^7$	$3.1802 \cdot 10^6$	$2.5148 \cdot 10^6$	736.2	4218.90
	c_{1z} $[N/m/s]$	k_{2y} $[N/m]$	k_{2z} $[N/m]$	c_{2y} $[N/m/s]$	c_{2z} $[N/m/s]$
Nominal Values	9910.00	$1.9700 \cdot 10^5$	$6.8700 \cdot 10^5$	$4.270 \cdot 10^4$	$4.270 \cdot 10^4$
Upper Bounds	11892.0	$2.3640 \cdot 10^5$	$8.2440 \cdot 10^5$	$5.124 \cdot 10^4$	$5.124 \cdot 10^4$
Lower Bounds	7928.0	$1.5760 \cdot 10^5$	$5.4960 \cdot 10^5$	$3.416 \cdot 10^4$	$3.416 \cdot 10^4$
Cp Optimized	8300.0	$2.3410 \cdot 10^5$	$5.5713 \cdot 10^5$	$4.148 \cdot 10^4$	$3.586 \cdot 10^4$
Ls Optimized	10164.0	$2.3440 \cdot 10^5$	$7.7381 \cdot 10^5$	$5.115 \cdot 10^4$	$3.620 \cdot 10^4$

† See Table A.2 in Appendix A for the definitions of the symbols in the table.

7.5 Results and Discussion

7.5.1 Conflicting Requirements on Design Variables

In Chapters 4, 5, and 6, the corresponding vehicle models are optimized with respect to lateral stability, vertical ride quality, and curving performance, respectively. Since the design criteria for these optimization problems are different, they impose different or even conflicting requirements on the specific design variable or variables. Table 7.1 offers selected numerical results based on the optimization of lateral stability (L_s) and the optimization of curving performance (C_p). Notice that the vehicle model used for optimization of the lateral stability and that used for optimization of curving performance share the identical vehicle configuration. For both optimization problems, 10 design variables (i.e. the relevant stiffness and damping coefficients for the secondary and primary suspensions) are permitted to vary by ±20% from their nominal values.

Table 7.1 shows the optimized design variables from both optimization problems and

the corresponding nominal and bound values for the design variables. Since primary suspension parameters have much more significant effect on curving performance and lateral stability of rail vehicles than secondary suspension parameters [35, 139], the following analysis places emphasis on the primary suspension parameters. As shown in Table 7.1, for the curving performance optimization problem, among the optimized primary suspension parameters, the longitudinal, lateral, and vertical damping coefficients, c_{1x}, c_{1y}, and c_{1z}, take the values lower than the corresponding nominal values, and the longitudinal and lateral spring stiffness coefficients, k_{1x} and k_{1y}, take the corresponding lower bound values. However, for the lateral stability optimization problem, among the optimized primary suspension parameters, except for the lateral spring stiffness and damping coefficients, k_{1y} and c_{1y}, the other parameters either take the values higher than the corresponding nominal values or take the corresponding upper bound values. Thus, the lateral stability and curving performance have conflicting requirements on the primary suspension design variables. These optimization results are consistent with previous observations by Wickens[139] that suspensions that are soft in the lateral and longitudinal direction tend to hunt more readily on tangent track and become unstable even at low speeds. However, such suspensions allow the wheelsets to follow curved track with decreased wheel wear and flange forces. The exception, i.e. the lateral spring stiffness and damping coefficients (k_{1y} and c_{1y}), to the observation by Wickens may be interpreted by the fact that at values above certain values, the lateral stability becomes relatively insensitive to these parameters. This exception was once reported by Hedrick et al. [74].

Besides the above conflicting requirements on suspension parameters, the lateral stability and curving performance also have conflicting requirements on geometric, inertial, or even active design variables. We will see in the following subsection that the hybrid MDO optimization approach offers an effective way to resolve these conflicting requirements.

7.5.2 Results of the Hybrid MDO Optimization Problem

The combined vehicle model is optimized with respect to three criteria, lateral stability, curving performance, and vertical ride quality as shown in the objective function (7.5). The constants ξ and η are both set to 1.0. To facilitate the implementation of the optimization problem, $max(\mid \frac{A_{ani}}{\tilde{A}_{ani}} \mid) + max(\mid \frac{L_{vk}}{\tilde{L}_{vk}} \mid)$ ($i = 1, 2, 3, 4$, and $k = 1, 2, ..., 8$), $\frac{A_R}{\tilde{A}_R}$, and $\frac{\tilde{V}_c}{V_c}$ are

defined as curving performance index, lateral stability performance index, and vertical ride quality index, respectively. To obtain a whole picture of the EP-optimal set, three sets of weighting factors ($\{\nu_1, \nu_2, \nu_3\}$) are selected and the corresponding optimizations are carried out. The three selected sets of weighting factors take the values of $\{1, 1, 1\}$, $\{1, 1, 2\}$, and $\{1, 1, 4\}$, respectively. A total of 29 parameters including geometric parameters, inertial property parameters, and suspension stiffness and damping coefficients are chosen as design variables. These design variables are permitted to vary by $\pm 20\%$ from their nominal values.

Figures 7.5, 7.6 and 7.7 illustrate selected results from the hybrid MDO optimization combining IDF and A-i-O. The individual designs from the GA are represented by circles, which tend to cluster as the GA converges to the optimal design. Plotted in Figure 7.5 is the vertical ride quality performance index versus lateral stability performance index. The clustered data corresponding to the EP-optimal set is almost horizontal, which shows that the optimized vertical ride quality is mainly independent of lateral stability. This is also true, as shown in Figure 7.6, for the relationship between vertical ride quality and curving performance. The observation about the relationship between vertical ride quality and lateral stability and that between vertical ride quality and curving performance demonstrates the conclusion [54] that a relatively weak coupling exists between the vertical and lateral motions of a rail vehicle.

However, Figure 7.7 shows a distinct trade-off in the relationship between lateral stability and curving performance. The EP-optimal set in the densely-clustered region shows that lateral stability can only be improved at the expense of curving performance, and vice-versa. No one criterion is favored over another; instead, the designer obtains explicit information about the trade-offs between lateral stability and curving performance. By running several more optimizations with different sets of weighting factors, one can get an even clearer picture of the EP-optimal set. Although this is a computationally-expensive process, the results are of obvious importance to rail vehicle designers.

7.6 Summary

In this chapter, a hybrid MDO optimization method, which is a combination of the individual discipline feasible (IDF) method used at the discipline level and the All-i-One (A-i-O)